**Drug Discovery Using Deep Learning**

Project Report Submitted in Partial Fulfillment of the Requirements for the Degree of

**Bachelor of Technology**

***in***

**Computer Science and Engineering**

*Submitted by:*

Burhan Iqbal Khan: (CSE-20-06)

Urbeena Rashid: (CSE-20-55)

Zuha Noor: (CSE-20-47)

***Under the Supervision of***

**Dr Assif Assad  
Assistant Professor, Department of CSE, IUST**



Department of Computer Science and Engineering

Islamic University of Science and Technology, Kashmir - 192122

Summer, 2024



**Department of Computer Science and Engineering**

**Islamic University of Science & Technology, Kashmir – 192122**

**CERTIFICATE**

This is to certify that the work contained in this report entitled **“Drug Discovery Using Deep Learning**” is submitted by the group members Mr. Burhan Iqbal Khan (Roll. No: CSE-20-05) , Ms. Urbeena Rashid Raina (Roll No: CSE-20-55) and Ms. Zuha Noor (Roll. No: CSE-20-47) to the **Department of Computer Science & Engineering**, Islamic University of science & Technology, Kashmir for the partial fulfilment of the requirements for the degree of **Bachelor of Technology** in **Computer Science and Engineering**.

They have carried out their work under my/our supervision. This work has not been submitted else-where for the award of any other degree.

**Supervisor(s) I/C Head**

Department of CSE. Department of CSE.

## Table of Contents

Table of Contents

[Table of Contents i](#_Toc173738752)

[List Of Figures ii](#_Toc173738753)

[Abstract 3](#_Toc173738754)

[Chapter 1: Introduction 1](#_Toc173738755)

[1. Overview 1](#_Toc173738756)

[1.1 Rationale 1](#_Toc173738757)

[1.2 Objectives 1](#_Toc173738758)

[Chapter 2: Literature Survey 4](#_Toc173738760)

[Chapter 3: Methodologies 9](#_Toc173738761)

3.1 MPNN

3.2 GAT

3.3 ADMET Analysis

[3.4 Tools Required 16](#_Toc173738763)

[Chapter 4: Summary and Conclusion 19](#_Toc173738764)

## List Of Figures

|  |  |
| --- | --- |
| **Figure 2.1** | Feed forward neural network |
| **Figure 2.2** | Feed forward neural network(RNN) |
| **Figure 3.1** | MPNN |
| **Figure 3.2** | GAT |
| **Figure 4.1** | Molecule 1 |
| **Figure 4.2** | Molecule 2 |

## Abstract

One of the most important areas of research in the field of drug development is drug target interaction, or DTI.It describes the understanding of how substances interact with the protein targets in the human body.It costs money and takes time to find these connections through wet lab studies.Drug research has witnessed the emergence of novel technologies such as deep learning to tackle these difficulties.Deep learning is a branch of artificial intelligence that is used to model intricate biological systems, analyze large datasets, and forecast molecular features.It expedites the identification of possible candidates, predicts drug-target interactions, and helps with virtual screening of compounds, all of which streamline the initial phases of drug development.

The objective is to accelerate the discovery process, reduce failures, and ultimately provide more effective and reasonably priced drugs to treat a range of illnesses and medical problems by utilizing deep learning and other cutting-edge technology. Using a deep neural network architecture and an attention mechanism, DTI-GAT utilizes the characteristics of drug and protein sequences in addition to interaction patterns on graph-structured data. By giving each node with the self-attention mechanism a variable attention weight, DTI-GAT makes it easier to understand the DTI topological structure. The final results demonstrate that the GAT structure better describes the relationships and has the highest level of accuracy among the obtained results..

## Chapter 1: Introduction

### 1. Overview

Alzheimer's disease, marked by memory loss and significant daily life disruptions, stands as a complex health concern. The traditional path to discovering suitable medications has been difficult, marked by prolonged timelines. In response to this challenge, our project introduces a

breakthrough strategy: Graph Neural Networks (GNNs). These advanced tools offer a dynamic and expedited avenue for identifying potential medicines for Alzheimer's. By embracing GNNs, we aim to redefine the pace and efficiency of drug discovery, assisting in a new era of hope for those impacted by Alzheimer's disease.

### 1.1 Rationale

We employed a deep neural network architecture with attention mechanisms in our study, which processes graph-structured data. It first transforms the drug fingerprint and the position-specific scoring matrix (PSSM) into feature vectors for each target protein and medication. Second, the similarity of feature vectors is used to build a DTI graph.

Every node in the network symbolizes a protein or a medication, and the edges between the nodes show the interactions between drugs and proteins as well as between drugs and other drugs. After applying the graph attention network to the constructed graph, embeddings for every protein and medication are produced. A final decoder architecture is then used to forecast the outcome of the interaction.

### 1.2 Objectives

**Identifying a potential Alzheimer's drug:**

Drugs are frequently created and used to treat symptoms, halt the progression of the disease, or address specific components of cognitive decline.

**Experimental Therapies:**

* Ongoing research involves the development and testing of drugs targeting various aspects of Alzheimer's disease pathology, such as beta-amyloid plaques and tau protein tangles.

It's important to emphasize that while these drugs can help manage symptoms, there is currently no cure for Alzheimer's disease. Additionally, drug therapies are generally prescribed based on a clinical diagnosis rather than as part of a diagnostic process.

### Implementation of Graph Neural Networks (GNNs)

* + Our study will implement various models of Graph Neural Networks (GNNs). These models offer promising avenues for enhancing the performance of Alzheimer's drug detection through deep learning techniques. Below are the GNN models we intend to explore:
  + 1.2.1.1 Message Passing Neural Network (MPNN)
  + MPNNs are a class of neural networks designed to operate on graph-structured data. In the context of molecular property prediction, MPNNs treat molecules as graphs where atoms are nodes and bonds are edges. By passing messages between neighboring nodes, MPNNs can aggregate information about the molecular structure, enabling accurate property prediction. We will use Chemprop to implement MPNN in our study.
  + 1.2.1.2 Graph Attention Network (GAT)
  + GATs are another variant of GNNs that leverage attention mechanisms to weigh the importance of neighboring nodes during message passing, enabling better modeling of graph structures. We plan to integrate GATs into our deep learning pipeline to enable the model to focus on the most relevant nodes and edges in the molecular graph, thus improving the discriminatory power of our drug detection system.graph, thus improving the discriminatory power of our drug detection system.

Through systematic experimentation and evaluation, we seek to identify the most suitable GNN architecture for our specific application domain, thereby enhancing the accuracy, robustness, and interpretability of our Alzheimer's drug detection system

**ADMET Analysis**

To ensure the potential drug candidates identified through our models are viable, we conducted ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) analysis using SwissADME. This step is crucial for evaluating the pharmacokinetic and drug-like properties of the molecules. After this initial evaluation, we wrote code to filter molecules based on specific ADMET properties such as hydrogen bond count and molecular weight (MW). This process started with a library of 1 million molecules and resulted in the identification of 2 molecules for detailed analysis.

**Experimental Validation**

Finally, we aim to validate our computational predictions through experimental assays. This step will involve collaboration with laboratory researchers to test the biological activity of the predicted drug candidates against Alzheimer's disease targets.

## Chapter 2: Literature Survey

[Alzheimer's disease](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/alzheimer-disease)[1] (AD) is the most common cause of dementia worldwide. The etiology is multifactorial, and [pathophysiology](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/pathophysiology)[2] of the disease is complex. Data indicate an exponential rise in the number of cases of AD, emphasizing the need for developing an effective treatment. AD also imposes tremendous emotional and financial burden to the patient's family and community. The disease has been studied over a century, but [acetylcholinesterase inhibitors](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/cholinesterase-inhibitor)[3]and [memantine](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/memantine)[4] are the only drugs currently approved for its management. These drugs provide symptomatic improvement alone but do less to modify the disease process. The extensive insight into the molecular and cellular pathomechanism in AD over the past few decades has provided us significant progress in the understanding of the disease. A number of novel strategies that seek to modify the disease process have been developed. The major developments in this direction are the amyloid and tau based therapeutics, which could hold the key to treatment of AD in the near future. Several putative drugs have been thoroughly investigated in [preclinical studies](https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/preclinical-study)[5], but many of them have failed to produce results in the clinical scenario; therefore it is only prudent that lessons be learnt from the past mistakes. The current rationales and targets evaluated for therapeutic benefit in AD are reviewed in this article.

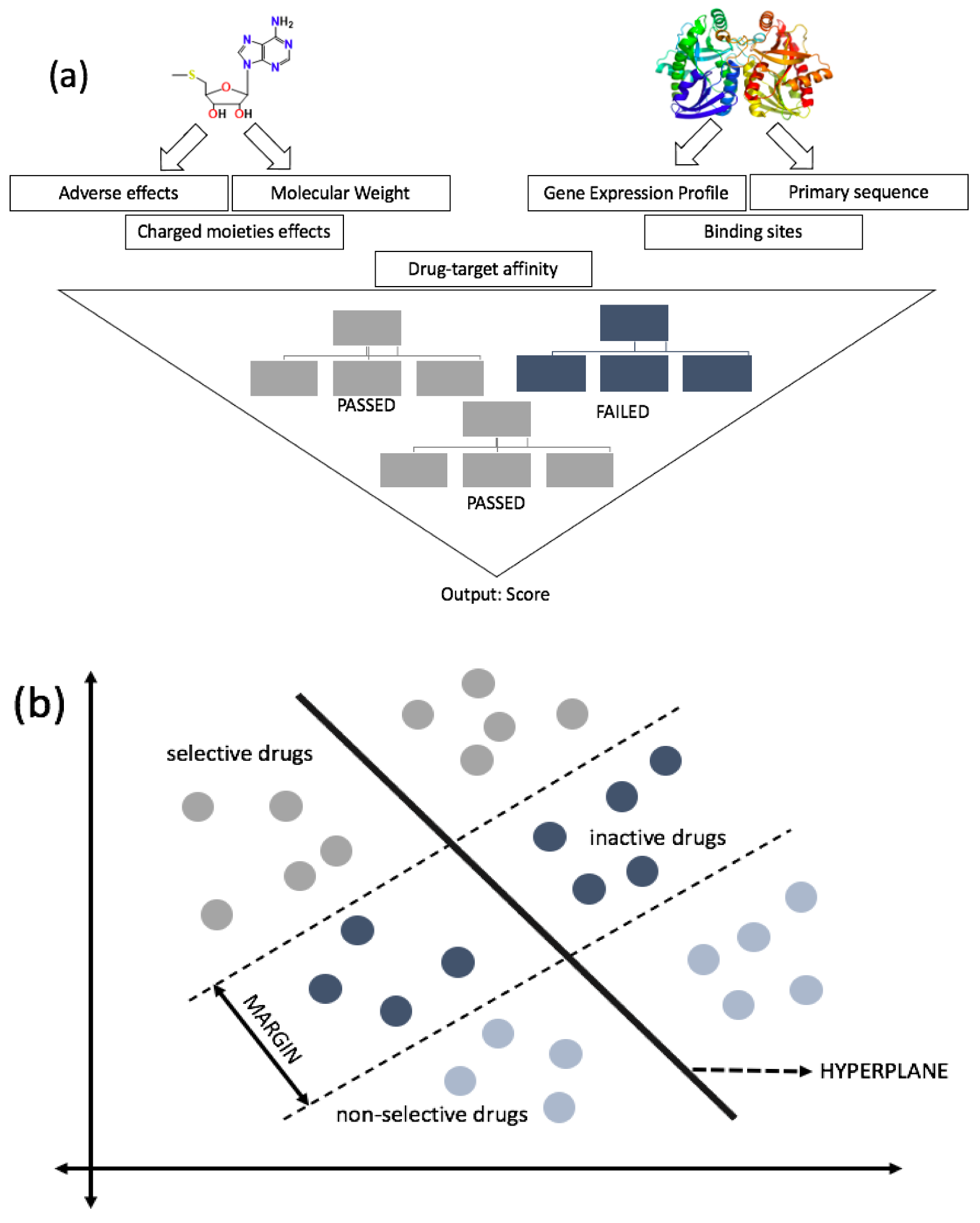
This article is part of the Special Issue entitled ‘The Synaptic Basis of Neurodegenerative Disorders’.

Alzheimer's disease (AD) is the most common cause of dementia, and with a new case occurring every seven seconds globally, the disease itself is becoming a slow pandemic (Ferriet al., 2005). One person for every 85 individuals can be expected to suffer from AD by the year 2050 (Brookmeyer et al., 2007)[6]. AD also imposes tremendous emotional and financial burden to the patient's family and community through the provision of care and loss of wages. The disease may be classified based on the age of onset into early-onset AD and late-onset AD. Early onset AD accounts for approximately 1%–6% of all cases and manifests roughly between 30 and 60 years. Late onset form accounting for around 90% of cases has an age at onset later than 60 years. Etiology of AD is multifactorial with genetic, environmental, behavioral and developmental components playing a role. The greatest risk factor is advancing age; others being a positive family history, head trauma, female gender, previous depression, diabetes mellitus, hyperlipidemia and vascular factors (Kivipelto et al., 2001)[7]. The understanding of the pathophysiology of AD is constantly changing; for instance the tangles, a well known pathological hallmark of AD, earlier thought to be responsible for the disease now rather seem to reflect the damage which the neurons have endured over a long time. The notion that amyloid beta peptide (Aβ) and phosphorylated tau are pathologic molecules is slowly changing, and it seems that they represent a cellular adaptive strategy to oxidative stress. Apart from them, various deranged mechanisms such as chronic oxidative stress, mitochondrial dysfunction, Aβ production, neurofibrillary tangles accumulation, hormone imbalance, inflammation, mitotic dysfunction, calcium mishandling, and genetic components play a role in the disease process. Although the mechanisms are diverse, neuronal death, the inevitable event occurs resulting in AD.

DL algorithms are considered one of the cutting-edge areas of development and study in almost all scientific and technological fields[8]. The renaissance of artificial NNs into workable algorithms from their former theorized and predicted applications, first developed in the 1950s, is an essential pillar of DL and the continued success brought by AI-based integration of standard techniques. DL algorithms give computational models the ability to learn a representation of multidimensional data through abstraction [9]. DL has allowed for resolving many challenges faced by standard ML algorithms, including image recognition and speech recognition. In the drug discovery process, DL techniques have become exemplary methods of drug activity prediction, target discovery, and lead molecule discovery [10,11]. The basis of DL is often implicated in NN systems, where they are used to create systems that have the capability to complete complex data recognition, interpretation, and generation. The main subsets of artificial NNs used in current drug discovery are deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs)[12]

The utilization of specific NNs from the variations that exist in the subset is dependent on multiple factors. DNNs, a specific type of feed forward neural networks, function with singular path data flow from the input layer through the hidden layer(s) reaching an output layer ([**Figure**](https://www.mdpi.com/1420-3049/25/22/5277#fig_body_display_molecules-25-05277-f003) 2a).[13] The outputs generated are typically identified using trained supervised learning algorithms. DL algorithms function through neural networks which can incorporate other ML techniques for training. Through supervised and reinforcement learning guided methods, a DNN can be trained to complete complex tasks. A generative DNN can create novel chemical compounds from existing libraries and training sets ([**Figure**](https://www.mdpi.com/1420-3049/25/22/5277#fig_body_display_molecules-25-05277-f003) 2a); while, a predictive DNN can predict the chemical attributes of the novel compounds [14,15]. QSAR models are currently being used to find the correlation between these compounds’ chemical structure and activity.

Once the initial work of target discovery is complete and better understanding is developed for target-molecule interaction, chemical synthesis and characterization become a priority in the pipeline. An important note in this process is using descriptive simplified molecular-input line-entry system (SMILES) nomenclature in many of the algorithms regarding de novo drug design and discovery[16]. RNNs, which are a type of NN that utilize a system of self-learning through generational processing of the inputs and developing hidden layers. The subset RNN-type long short-term memory has become a reliable, standardized method for generating novel chemical structures. RNNs are unique in their ability to use neurons connected in the same hidden layer to form a functioning cycle of processing inputs and outputs compared to DNNs and feedforward neural networks ([**Figure**](https://www.mdpi.com/1420-3049/25/22/5277#fig_body_display_molecules-25-05277-f003) 2b), which have no connections within the same layer and only push outputs. These generative RNNs have shown promising results in the generation of sensible, structurally correct, and feasible, novel SMILE structures that were not included in the original SMILE training sets [17,18]. A recent study by Segler et al. used generative RNN models to develop possible molecular structures that could have activity against *Staphylococcus aureus* (*S. aureus*) and *Plasmodium falciparum* (*P. falciparum*)[19]. Their models were given small sets of molecular structures that had known activity against these target organisms; from these inputs, the model generated 14% of the 6051 potential molecule candidates for *S. aureus* that has been developed by medicinal chemists. The model also generated 28% of the existing compounds developed for *P. falciparum* [20]. Traditionally, the generation and implementation of chemical synthesis routes have been the sole responsibility of chemists. However, this role is evolving to include more and more computational based synthesis, also known as computer-aided synthesis planning (CASP), with the emergence of AI [21,22]. The Monte Carlo tree search (MCTS) based on NN techniques have been used in current studies to generate CASP workflows.



**Figure 2.1:** Feed Forward Network

**Figure 2.2:** Feed Forward Network(RNN)

**Examples of Drug Discovery**

ML is already being used to develop novel molecules that could be used as future antibiotic candidates. In a recent, groundbreaking study conducted by Stokes et al., the researchers demonstrated the utility and capability of ML techniques in the drug discovery process [23]. They specifically capitalized on the use of DNNs to create novel molecules with broad-spectrum antibacterial activity. These discovered candidates were also identified to be structurally distinct from any known antibiotics. The researchers utilized a training set of 2335 molecules for a DNN model to predict the growth inhibition of Escherichia coli, followed by the running of the model on greater than 107 million molecules from several chemical libraries. This gave the researchers the ability to identify potential lead compound candidates that may have similar bioactivity. [24,25]Through scoring generated by the model, the researchers were able to identify a list of sensible candidates that meet a predetermined score threshold and various other eliminative criteria. The researchers’ efforts proved fruitful, and they were able to identify a c-Jun N-terminal kinase inhibitor, halicin, that is distinct from known antibiotics. This antibacterial candidate was also discovered to be a potent growth inhibitor of Escherichia coli, and had shown efficacy against Clostridioides difficile and Acinetobacter baumannii infections in murine models [26,27]. In a study conducted by Fields et al., ML algorithms, including NNs-based techniques and SVM models, were used to discover novel antimicrobial peptides, also known as bacteriocins, from bacteria that could ultimately be used as compelling antibiotic candidates [28].

Their model incorporated RF and elastic net (Enet) algorithms to evaluate the DNN model’s results. This framework was only tested on five patients; thus, not much coverage was obtained through this model; therefore, they expanded their study to a more massive sample size. They utilized response data for two drugs: Cisplatin and paclitaxel, and analyzed it with gene expression profiles and patients’ responses to those two drugs gathered from different clinical trials[29].The diseases discussed have been around for a long time, but the emergent need for a treatment for Coronavirus disease 2019 (COVID-19) has stirred up the research world. The pandemic outbreak has caused detrimental effects around the world, but the COVID-19 virus (SARS-CoV-2) is a novel strain of the same species of virus causing the 2003 Severe acute respiratory syndrome (SARS-CoV-1); thus, several studies are incorporating earlier information into supervised ML to quickly find a remedy for this virus [30]. Researchers worldwide are exhausting all available resources, and ML has helped narrow down the drug candidates and minimize clinical trial failure. Kowalewski and Ray developed ML models to help identify effective drugs against 65 human proteins (target) studied to interact with SARS-CoV-2 proteins. As the virus is known to target the respiratory tract, including nasal epithelial cells and upper airway and lungs, they deduce it from inhaling therapeutics to directly target the damaged cells.[31]

## Chapter 3: Methodologies

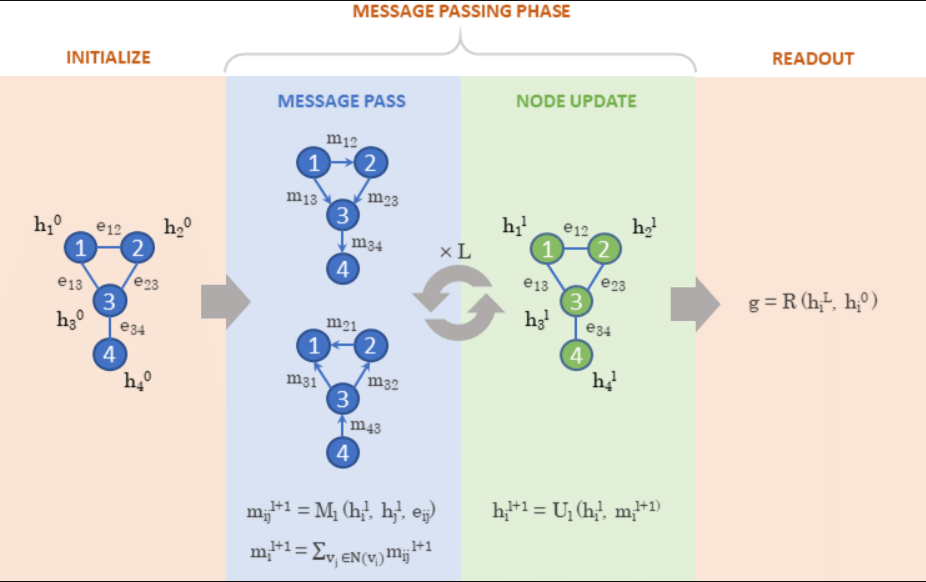
In this chapter, we describe the methodologies employed in our project on molecular property prediction for Alzheimer's disease. The combination of Chemprop's MPNN and the implementation of GAT provided a robust framework for accurately predicting properties related to Alzheimer's disease. These methods were chosen for their ability to effectively capture the complex relationships between molecular structures and their properties.

**The collaborators had supplied the datasets which we have considered for this project, each of which has a distinct target protein.**

Here's a general overview of GNN models that have been applied:

**3.1 Message Passing Neural Network (MPNN)**

MPNNs are a class of neural networks designed to operate on graph-structured data. In the context of molecular property prediction, MPNNs treat molecules as graphs where atoms are nodes and bonds are edges.

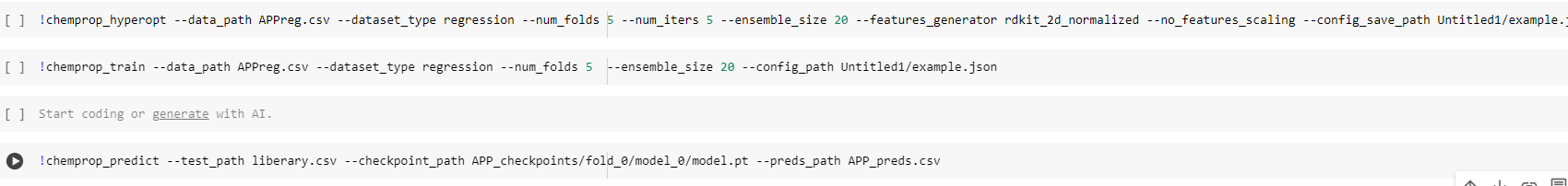
****

**Figure 3.1** Source:ResearchGate

uploaded by [Ola Engkvist](https://www.researchgate.net/profile/Ola-Engkvist?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ)

* Message Passing Phase: During this phase, messages are passed between neighboring nodes to aggregate information about their local environment.
* Update Phase: The aggregated messages are used to update the hidden states of the nodes.
* Readout Phase: Finally, the hidden states are combined to produce a graph-level output, which corresponds to the predicted molecular property.

**Implementation Details**

****

**3.2** **Graph Attention Network (GAT)**

**Overview of GAT**

Graph Attention Networks (GAT) extend the capabilities of traditional graph neural networks by incorporating attention mechanisms. This allows the model to weigh the importance of different nodes' contributions during the message passing process.

**Attention Mechanism**

The attention mechanism in GAT assigns different weights to neighboring nodes based on their features, enabling the model to focus on the most relevant parts of the graph.

**Architecture**

The GAT architecture includes:

* **Self-Attention Layers**: These layers compute attention coefficients for each edge, which are used to aggregate neighbor information.
* **Node Feature Update**: Node features are updated by combining the weighted messages from their neighbors.
* **Multi-Head Attention**: Multiple attention heads are used to stabilize the learning process and improve model performance.

# Computation Steps

1. **Input Features**: Each node in the graph has a feature vector. Let’s denote the feature vectors of the nodes as *x₁*, *x*₂, and *x*₃​.
2. **Linear Transformation**: The first step in GAT is a linear transformation of the input features. This is done using a weight matrix *W*. The transformed features are then *hᵢ* =*xᵢW*, for *i*=1,2,3.
3. **Self-Attention Mechanism**: The next step is to compute the attention coefficients, which indicate the importance of node m’s features to node i. This is done using the following equation:

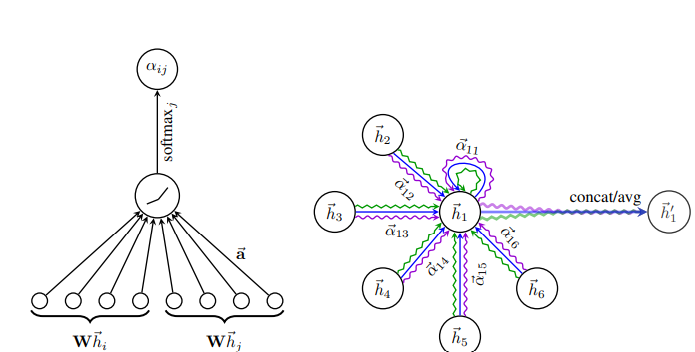
eᵢj = LeakyReLU(aT[Whᵢ​ ∣∣Whj])

where ∣∣ denotes concatenation, *a* is a learnable weight vector, and *LeakyReLU* is the activation function

1. **Normalization**: The attention coefficients are then normalized using the SoftMax function to make them sum up to 1. This is done for each node i across all its neighbors j, using the following equation:

αᵢj ​= SoftMaxₘ(eᵢj) = exp(eᵢj) / ∑ₖ exp(eᵢₖ​​)​

where k∈N(i)​ denotes the neighbors of node i.



Source: https://paperswithcode.com/method/gat

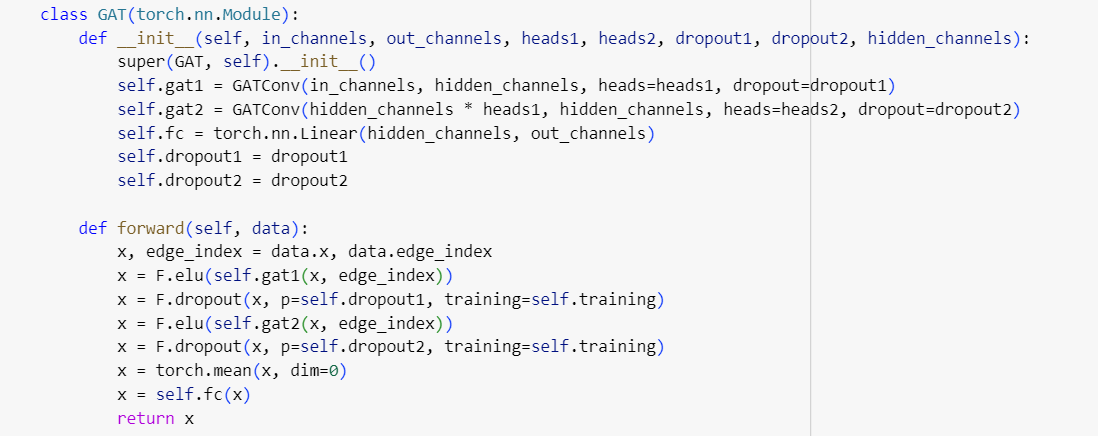
**Figure 3.2**

**5- Feature Aggregation**: The final step is to compute the output features of each node as a weighted sum of its neighbors’ features, using the normalized attention coefficients as weights. This is done using the following equation by aggregating over *m* index:

*hᵢ*′ ​= *σ*(∑*jαᵢjWhj*)

where *σ* is the activation function

**Implementation Details**



**3.3 ADMET Analysis Using SwissADME**

**Overview of SwissADME**

SwissADME is an online tool used for predicting the Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties of small molecules. It is widely used for evaluating the pharmacokinetic and drug-like properties of compounds.

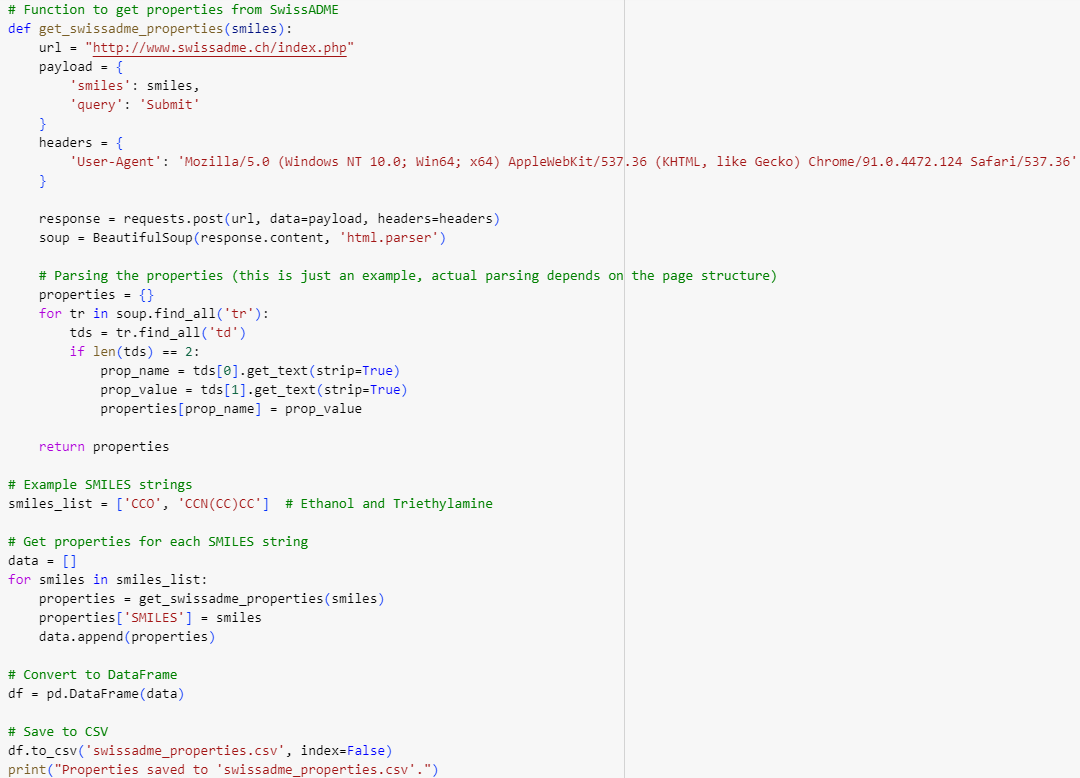
**Implementation Details**

For our project, we conducted ADMET analysis using SwissADME to assess the pharmacokinetic properties of the predicted molecules. The process involved:

Data Input: SMILES strings of the predicted molecules were input into the SwissADME tool.

Property Prediction: SwissADME generated predictions for various ADMET properties, including solubility, permeability, and potential toxicity.

**implementation**





Result Analysis: The results were analyzed to identify molecules with favorable ADMET profiles for potential drug development.

We identified the following two molecules for detailed ADMET analysis:

* **Molecule 1**: N#CC(C(C#N)(C)C)(C)C
* **Molecule 2**: N#CC12CCCC(C2)(C(=O)O1)

**Conclusion:**

Applying graph neural network methodologies to Alzheimer's disease drug discovery holds great promise. However, it's crucial to collaborate with experts in biology and medicine, ensure ethical considerations, and validate predictions through rigorous experimentation for the successful development of novel therapeutics.

### 3.4 Tools Required

In the pursuit of identifying potential drug molecules for Alzheimer's disease through the innovative lens of Graph Neural Networks (GNNs) and a deep learning approach, the success of our research is intricately linked to the strategic utilization of cutting-edge tools and frameworks. This section delineates the essential components that form the technological backbone of our investigative journey.

1. **Graph Neural Network Frameworks:**

* **DeepChem:** An essential tool for drug discovery, DeepChem is a deep learning framework specifically designed for cheminformatics and bioinformatics. It provides tools for molecular data handling and GNN training.
* **DGL (Deep Graph Library):** A dedicated library for building and training graph neural networks, crucial for the analysis of molecular structures.
* **PyTorch Geometric:** An extension library for PyTorch tailored for processing irregularly structured input data, particularly useful for working with graphs.

1. **Molecular Representation and Cheminformatics Libraries:**

* **RDKit:** This cheminformatics toolkit will be instrumental in handling and analyzing chemical informatics data.
* **ChemPy:** A library designed for chemical information processing, including molecular structure manipulation and chemical calculations.
* **SMILES and InChI Representation Tools:** Tools for encoding molecular structures using SMILES and InChI representations.

1. **Python Libraries**: There are several popular Python libraries that provide essential functionalities for implementing neural network models, including Graph Neural Networks.. These include:

* **TensorFlow:** An open-source deep learning library that offers a high-level API, called Keras, for building and training neural networks, including GNNs.
* **PyTorch:** Another popular deep learning library with a dynamic neural network framework that supports the construction and training of GNN models efficiently.
* **Keras:** A user-friendly, high-level neural networks library that provides a simple and intuitive API for building GNN models on top of TensorFlow, Theano, or CNTK.

1. **Data Preparation:** Python offers various libraries for data manipulation and preprocessing required for implementing neural network models, including Graph Neural Networks. These include:

* **NumPy:** A fundamental library for numerical computations, providing support for handling multidimensional arrays and mathematical operations essential for data preprocessing.
* **Pandas:** Pandas is widely used in data analysis, machine learning, and data science projects due to its versatility and efficiency in handling structured data. Its intuitive syntax and comprehensive set of functions make it a powerful tool for data manipulation and analysis tasks.
* **Scikit-learn:** A machine learning library that provides utilities for data preprocessing, including data splitting, feature scaling, and one-hot encoding.

1. **Model Training and Evaluation:** Python libraries provide tools for training and neural network models. These libraries offer functionality for defining loss functions, choosing optimization algorithms (e.g., stochastic gradient descent, Adam), setting batch sizes, and specifying the number of training epochs. You can also monitor the model's training progress and evaluate its performance using evaluation metrics like accuracy, precision, recall, and F1-score.
2. **Visualization and Analysis**: Python provides numerous visualization libraries that facilitate the interpretation and analysis of neural network models. These libraries include:

* **Matplotlib**: A powerful plotting library that allows the creation of various visualizations, such as line plots, scatter plots, and histograms, to analyze the model's performance, loss curves, and feature maps.
* **Seaborn**: A statistical data visualization library built on top of Matplotlib, offering additional functionalities and improved aesthetics.
* **TensorBoard**: A visualization toolkit provided by TensorFlow that enables visualizing model architectures, training curves, and embedding spaces.
* **RDKit Visualization Tools***:* RDKit includes visualization tools for molecular structures.

1. **Jupyter Notebooks:**

* **Jupyter:** An interactive computing environment, invaluable for creating, sharing, and documenting code and analyses.

1. **High-Performance Computing (HPC) Resources:**

* Depending on the project scale, access to high-performance computing resources may be sought for training complex models efficiently.

1. **Literature and Research Resources:**

* Access to scientific journals, articles, and research papers relevant to drug discovery and GNNs will be crucial for staying updated on the latest advancements.

1. **Collaboration Tools:**

* **GitHub:** These platforms will facilitate collaborative work on Git repositories, supporting teamwork and version control.

## Chapter 4: Summary and Conclusion

Various different graphical neural network architectures, like message passing neural networks, learns to predict molecular properties directly from the graph structure of the molecule, where atoms are represented as nodes and bonds are represented as edges. For every molecule, we reconstructed the molecular graph corresponding to each compound's SMILES string and determined the set of atoms and bonds using the open-source package RDKit (Landrum, 2006).

Next, we initialized a feature vector, for each atom and bond based on computable features:

**1. Atom features:** atomic number, number of bonds for each atom, formal charge, chirality, number of bonded hydrogens, hybridization, aromaticity, atomic mass.

**2. Bond features:** bond type (single/double/triple/aromatic), conjugation, ring membership, stereochemistry.

The model applied a series of message passing steps where it aggregated information from neighboring atoms and bonds to build an understanding of local chemistry. On each step of message passing, each bond's featurization is updated by summing the featurization of neighboring bonds, concatenating the current bond's featurization with the sum, and then

applying a single neural network layer with non-linear activation. After a fixed number of message-passing steps, the learned featurizations across the molecule are summed to produce a single featurization for the whole molecule.

Finally, this featurization is fed through a feed- forward neural network that outputs a prediction of the property of interest. Since the property of interest in our application was the REGRESSION to predict the pIC50 value of molecules. The model is trained to output pIC50 values for each compound.

**Results**

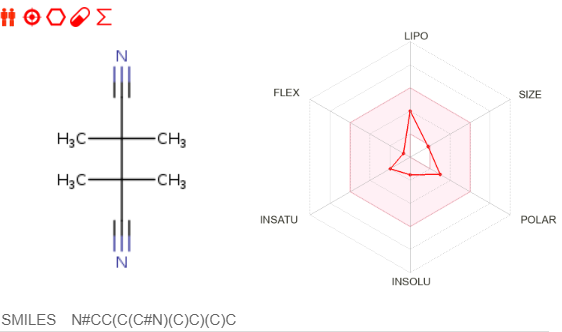
The molecular property prediction project for Alzheimer's disease achieved the following results:

* **MPNN Model (using Chemprop):**
  + Test RMSE: 0.67594
  + Validation RMSE: 0.673537
  + The model predicted the pIC50 values of a dataset of 6808 molecules, with the predicted values ranging from 1.431789 to 12.69.
  + Additionally, the model predicted the pIC50 values of a ChEMBL dataset consisting of 1,048,576 molecules
* **GAT Model:**
  + Test RMSE:
  + Validation RMSE:
* **Admet Analysis:**

From the ChEMBL dataset, molecules with higher pIC50 values were selected, and ADMET analysis was performed on them

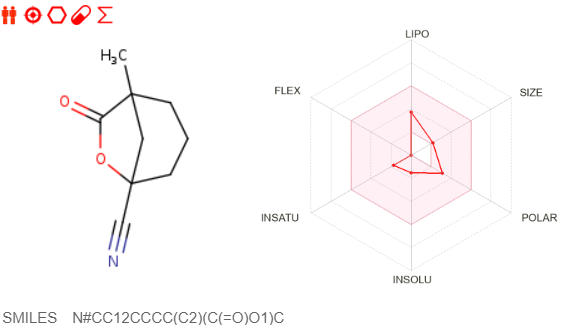
After performing ADMET analysis, we filtered the dataset and identified 2 molecules that met the specified criteria.

* **Molecule 1**: N#CC(C(C#N)(C)C)(C)C

****

**Figure 4.1**

* **Molecule 2**: N#CC12CCCC(C2)(C(=O)O1)C

****

**Figure 4.2**

**Range of pIC50 1.431789 to 12.69**

**Total number of molecules 6808**

## Chapter 5:

**References**

1. Fuhai Song, ... Hongxing Lei, in [International Review of Neurobiology](https://www.sciencedirect.com/bookseries/international-review-of-neurobiology), 2015

2.Demosthenes G. Papamatheakis, William R. Auger, in [Cardiac Intensive Care (Third Edition)](https://www.sciencedirect.com/book/9780323529938/cardiac-intensive-care), 2019

3.[Meyler's Side Effects of Drugs: The International Encyclopedia of Adverse Drug Reactions and Interactions (Fifteenth Edition)](https://www.sciencedirect.com/referencework/9780444510051/meylers-side-effects-of-drugs), 2006

4.M.A. Keiski, in [New Therapeutics for Traumatic Brain Injury](https://www.sciencedirect.com/book/9780128026861/new-therapeutics-for-traumatic-brain-injury), 2017

5.Shayne C. Gad, in [Encyclopedia of Tissue Engineering and Regenerative Medicine](https://www.sciencedirect.com/referencework/9780128137000/encyclopedia-of-tissue-engineering-and-regenerative-medicine), 2019

6. Hafiz A.M., Bhat G.M. A Survey of Deep Learning Techniques for Medical Diagnosis. In: Tuba M., Akashe S., Joshi A., editors. Information and Communication Technology for Sustainable Development. Volume 933 Springer; Singapore: 2020. Advances in Intelligent Systems and Computing. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Information+and+Communication+Technology+for+Sustainable+Development&author=A.M.+Hafiz&author=G.M.+Bhat&publication_year=2020&)]

7. Civit-Masot J., Luna-Perejón F., Domínguez Morales M., Civit A. Deep Learning System for COVID-19 Diagnosis Aid Using X-ray Pulmonary Images. Appl. Sci. 2020;**10**:4640. doi: 10.3390/app10134640. [[CrossRef](https://doi.org/10.3390%252Fapp10134640)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Appl.+Sci.&title=Deep+Learning+System+for+COVID-19+Diagnosis+Aid+Using+X-ray+Pulmonary+Images&author=J.+Civit-Masot&author=F.+Luna-Perej%25C3%25B3n&author=M.+Dom%25C3%25ADnguez+Morales&author=A.+Civit&volume=10&publication_year=2020&pages=4640&doi=10.3390/app10134640&)]

8. Yamanakkanavar N., Choi J.Y., Lee B. MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer’s Disease: A Survey. Sensors. 2020;**20**:3243. doi: 10.3390/s20113243. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7313699/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/32517304)] [[CrossRef](https://doi.org/10.3390%252Fs20113243)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Sensors&title=MRI+Segmentation+and+Classification+of+Human+Brain+Using+Deep+Learning+for+Diagnosis+of+Alzheimer%25E2%2580%2599s+Disease:+A+Survey&author=N.+Yamanakkanavar&author=J.Y.+Choi&author=B.+Lee&volume=20&publication_year=2020&pages=3243&doi=10.3390/s20113243&)]

9.LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Deep+learning&author=LeCun,+Y.&author=Bengio,+Y.&author=Hinton,+G.&publication_year=2015&journal=Nature&volume=521&pages=436%E2%80%93444&doi=10.1038/nature14539&pmid=26017442)] [[**CrossRef**](https://doi.org/10.1038/nature14539)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/26017442)]

10.Dana, D.; Gadhiya, S.V.; St Surin, L.G.; Li, D.; Naaz, F.; Ali, Q.; Paka, L.; Yamin, M.A.; Narayan, M.; Goldberg, I.D. Deep Learning in Drug Discovery and Medicine; Scratching the Surface. *Molecules* **2018**, *23*, 2384. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Deep+Learning+in+Drug+Discovery+and+Medicine;+Scratching+the+Surface&author=Dana,+D.&author=Gadhiya,+S.V.&author=St+Surin,+L.G.&author=Li,+D.&author=Naaz,+F.&author=Ali,+Q.&author=Paka,+L.&author=Yamin,+M.A.&author=Narayan,+M.&author=Goldberg,+I.D.&publication_year=2018&journal=Molecules&volume=23&pages=2384&doi=10.3390/molecules23092384)] [[**CrossRef**](https://doi.org/10.3390/molecules23092384)][[**Green Version**](https://www.mdpi.com/1420-3049/23/9/2384/pdf)]

11.Korotcov, A.; Tkachenko, V.; Russo, D.P.; Ekins, S. Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol. Pharm.* **2017**, *14*, 4462–4475. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Comparison+of+Deep+Learning+with+Multiple+Machine+Learning+Methods+and+Metrics+Using+Diverse+Drug+Discovery+Data+Sets&author=Korotcov,+A.&author=Tkachenko,+V.&author=Russo,+D.P.&author=Ekins,+S.&publication_year=2017&journal=Mol.+Pharm.&volume=14&pages=4462%E2%80%934475&doi=10.1021/acs.molpharmaceut.7b00578)] [[**CrossRef**](https://doi.org/10.1021/acs.molpharmaceut.7b00578)]

12.Ekins, S. The Next Era: Deep Learning in Pharmaceutical Research. *Pharm. Res.* **2016**, *33*, 2594–2603. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=The+Next+Era:+Deep+Learning+in+Pharmaceutical+Research&author=Ekins,+S.&publication_year=2016&journal=Pharm.+Res.&volume=33&pages=2594%E2%80%932603&doi=10.1007/s11095-016-2029-7)] [[**CrossRef**](https://doi.org/10.1007/s11095-016-2029-7)]

13.D’Souza, S.; Prema, K.V.; Balaji, S. Machine learning models for drug-target interactions: Current knowledge and future directions. *Drug Discov. Today* **2020**, *25*, 748–756. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Machine+learning+models+for+drug-target+interactions:+Current+knowledge+and+future+directions&author=D%E2%80%99Souza,+S.&author=Prema,+K.V.&author=Balaji,+S.&publication_year=2020&journal=Drug+Discov.+Today&volume=25&pages=748%E2%80%93756&doi=10.1016/j.drudis.2020.03.003)] [[**CrossRef**](https://doi.org/10.1016/j.drudis.2020.03.003)]

14.Baskin, I.I.; Winkler, D.; Tetko, I.V. A renaissance of neural networks in drug discovery. *Expert Opin. Drug Discov.* **2016**, *11*, 785–795. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=A+renaissance+of+neural+networks+in+drug+discovery&author=Baskin,+I.I.&author=Winkler,+D.&author=Tetko,+I.V.&publication_year=2016&journal=Expert+Opin.+Drug+Discov.&volume=11&pages=785%E2%80%93795&doi=10.1080/17460441.2016.1201262&pmid=27295548)] [[**CrossRef**](https://doi.org/10.1080/17460441.2016.1201262)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/27295548)]

14. Chen J., Bi S., Zhang G., Cao G. High-Density Surface EMG-Based Gesture Recognition Using a 3D Convolutional Neural Network. Sensors. 2020;**20**:1201. doi: 10.3390/s20041201. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070985/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/32098264)] [[CrossRef](https://doi.org/10.3390%252Fs20041201)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Sensors&title=High-Density+Surface+EMG-Based+Gesture+Recognition+Using+a+3D+Convolutional+Neural+Network&author=J.+Chen&author=S.+Bi&author=G.+Zhang&author=G.+Cao&volume=20&publication_year=2020&pages=1201&doi=10.3390/s20041201&)]

15.Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=The+rise+of+deep+learning+in+drug+discovery&author=Chen,+H.&author=Engkvist,+O.&author=Wang,+Y.&author=Olivecrona,+M.&author=Blaschke,+T.&publication_year=2018&journal=Drug+Discov.+Today&volume=23&pages=1241%E2%80%931250&doi=10.1016/j.drudis.2018.01.039&pmid=29366762)] [[**CrossRef**](https://doi.org/10.1016/j.drudis.2018.01.039)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/29366762)]

16.Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, eaap7885. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Deep+reinforcement+learning+for+de+novo+drug+design&author=Popova,+M.&author=Isayev,+O.&author=Tropsha,+A.&publication_year=2018&journal=Sci.+Adv.&volume=4&pages=eaap7885&doi=10.1126/sciadv.aap7885&pmid=30050984)] [[**CrossRef**](https://doi.org/10.1126/sciadv.aap7885)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/30050984)][[**Green Version**](https://advances.sciencemag.org/content/advances/4/7/eaap7885.full.pdf)]

17.Yuan, W.; Jiang, D.; Nambiar, D.K.; Liew, L.P.; Hay, M.P.; Bloomstein, J.; Lu, P.; Turner, B.; Le, Q.T.; Tibshirani, R.; et al. Chemical Space Mimicry for Drug Discovery. *J. Chem. Inf. Model.* **2017**, *57*, 875–882. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Chemical+Space+Mimicry+for+Drug+Discovery&author=Yuan,+W.&author=Jiang,+D.&author=Nambiar,+D.K.&author=Liew,+L.P.&author=Hay,+M.P.&author=Bloomstein,+J.&author=Lu,+P.&author=Turner,+B.&author=Le,+Q.T.&author=Tibshirani,+R.&publication_year=2017&journal=J.+Chem.+Inf.+Model.&volume=57&pages=875%E2%80%93882&doi=10.1021/acs.jcim.6b00754)] [[**CrossRef**](https://doi.org/10.1021/acs.jcim.6b00754)][[**Green Version**](http://europepmc.org/articles/pmc5802964?pdf=render)]

15. Alam M., Munia T.T.K., Tavakolian K., Vasefi F., MacKinnon N., Fazel-Rezai R. Automatic detection and severity measurement of eczema using image processing; Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Orlando, FL, USA. 16–20 August 2016; pp. 1365–1368. [[PubMed](https://pubmed.ncbi.nlm.nih.gov/28268579)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Proceedings+of+the+2016+38th+Annual+International+Conference+of+the+IEEE+Engineering+in+Medicine+and+Biology+Society+(EMBC)&title=Automatic+detection+and+severity+measurement+of+eczema+using+image+processing&author=M.+Alam&author=T.T.K.+Munia&author=K.+Tavakolian&author=F.+Vasefi&author=N.+MacKinnon&pages=1365-1368&)]

16. Mehdy M., Ng P., Shair E.F., Saleh N., Gomes C. Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. Comput. Math. Methods Med. 2017;**2017**:2610628. doi: 10.1155/2017/2610628. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5394406/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/28473865)] [[CrossRef](https://doi.org/10.1155%252F2017%252F2610628)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Comput.+Math.+Methods+Med.&title=Artificial+Neural+Networks+in+Image+Processing+for+Early+Detection+of+Breast+Cancer&author=M.+Mehdy&author=P.+Ng&author=E.F.+Shair&author=N.+Saleh&author=C.+Gomes&volume=2017&publication_year=2017&pages=2610628&pmid=28473865&doi=10.1155/2017/2610628&)]

17. Rathod J., Waghmode V., Sodha A., Bhavathankar P. Diagnosis of skin diseases using Convolutional Neural Networks; Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); Coimbatore, India. 29–31 March 2018; pp. 1048–1051. [[CrossRef](https://doi.org/10.1109%252FICECA.2018.8474593)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Proceedings+of+the+2018+Second+International+Conference+on+Electronics,+Communication+and+Aerospace+Technology+(ICECA)&title=Diagnosis+of+skin+diseases+using+Convolutional+Neural+Networks&author=J.+Rathod&author=V.+Waghmode&author=A.+Sodha&author=P.+Bhavathankar&pages=1048-1051&doi=10.1109/ICECA.2018.8474593&)]

18. Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. J. Biomed. Inform. 2018;**86**:25–32. doi: 10.1016/j.jbi.2018.08.006. [[PubMed](https://pubmed.ncbi.nlm.nih.gov/30103029)] [[CrossRef](https://doi.org/10.1016%252Fj.jbi.2018.08.006)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=J.+Biomed.+Inform.&title=Skin+lesion+classification+with+ensembles+of+deep+convolutional+neural+networks&author=B.+Harangi&volume=86&publication_year=2018&pages=25-32&pmid=30103029&doi=10.1016/j.jbi.2018.08.006&)]

19.Gupta, A.; Müller, A.T.; Huisman, B.J.H.; Fuchs, J.A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* **2018**, *37*, 1700111. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Generative+Recurrent+Networks+for+De+Novo+Drug+Design&author=Gupta,+A.&author=M%C3%BCller,+A.T.&author=Huisman,+B.J.H.&author=Fuchs,+J.A.&author=Schneider,+P.&author=Schneider,+G.&publication_year=2018&journal=Mol.+Inf.&volume=37&pages=1700111&doi=10.1002/minf.201700111)] [[**CrossRef**](https://doi.org/10.1002/minf.201700111)][[**Green Version**](https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/minf.201700111)]

20.Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4*, 120–131. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Generating+Focused+Molecule+Libraries+for+Drug+Discovery+with+Recurrent+Neural+Networks&author=Segler,+M.H.S.&author=Kogej,+T.&author=Tyrchan,+C.&author=Waller,+M.P.&publication_year=2018&journal=ACS+Cent.+Sci.&volume=4&pages=120%E2%80%93131&doi=10.1021/acscentsci.7b00512)] [[**CrossRef**](https://doi.org/10.1021/acscentsci.7b00512)][[**Green Version**](https://pubs.acs.org/doi/pdf/10.1021/acscentsci.7b00512)]

21.De Almeida, A.F.; Moreira, R.; Rodrigues, T. Synthetic organic chemistry driven by artificial intelligence. *Nat. Rev. Chem.* **2019**, *3*, 589–604. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Synthetic+organic+chemistry+driven+by+artificial+intelligence&author=De+Almeida,+A.F.&author=Moreira,+R.&author=Rodrigues,+T.&publication_year=2019&journal=Nat.+Rev.+Chem.&volume=3&pages=589%E2%80%93604&doi=10.1038/s41570-019-0124-0)] [[**CrossRef**](https://doi.org/10.1038/s41570-019-0124-0)]

22.Chan, H.C.S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* **2019**, *40*, 592–604. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Advancing+Drug+Discovery+via+Artificial+Intelligence&author=Chan,+H.C.S.&author=Shan,+H.&author=Dahoun,+T.&author=Vogel,+H.&author=Yuan,+S.&publication_year=2019&journal=Trends+Pharmacol.+Sci.&volume=40&pages=592%E2%80%93604&doi=10.1016/j.tips.2019.06.004&pmid=31320117)] [[**CrossRef**](https://doi.org/10.1016/j.tips.2019.06.004)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/31320117)]

23.Segler, M.H.S.; Preuss, M.; Waller, M.P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604–610. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Planning+chemical+syntheses+with+deep+neural+networks+and+symbolic+AI&author=Segler,+M.H.S.&author=Preuss,+M.&author=Waller,+M.P.&publication_year=2018&journal=Nature&volume=555&pages=604%E2%80%93610&doi=10.1038/nature25978&pmid=29595767)] [[**CrossRef**](https://doi.org/10.1038/nature25978)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/29595767)][[**Green Version**](http://arxiv.org/pdf/1708.04202)]

24.Segler, M.H.S.; Waller, M.P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. Eur. J.* **2017**, *23*, 5966–5971. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Neural-Symbolic+Machine+Learning+for+Retrosynthesis+and+Reaction+Prediction&author=Segler,+M.H.S.&author=Waller,+M.P.&publication_year=2017&journal=Chem.+Eur.+J.&volume=23&pages=5966%E2%80%935971&doi=10.1002/chem.201605499&pmid=28134452)] [[**CrossRef**](https://doi.org/10.1002/chem.201605499)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/28134452)]

25.Zhou, Z.; Li, X.; Zare, R.N. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sci.* **2017**, *3*, 1337–1344. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Optimizing+Chemical+Reactions+with+Deep+Reinforcement+Learning&author=Zhou,+Z.&author=Li,+X.&author=Zare,+R.N.&publication_year=2017&journal=ACS+Cent.+Sci.&volume=3&pages=1337%E2%80%931344&doi=10.1021/acscentsci.7b00492)] [[**CrossRef**](https://doi.org/10.1021/acscentsci.7b00492)][[**Green Version**](https://pubs.acs.org/doi/pdf/10.1021/acscentsci.7b00492)]

26.Coley, C.W.; Barzilay, R.; Jaakkola, T.S.; Green, W.H.; Jensen, K.F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3*, 434–443. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Prediction+of+Organic+Reaction+Outcomes+Using+Machine+Learning&author=Coley,+C.W.&author=Barzilay,+R.&author=Jaakkola,+T.S.&author=Green,+W.H.&author=Jensen,+K.F.&publication_year=2017&journal=ACS+Cent.+Sci.&volume=3&pages=434%E2%80%93443&doi=10.1021/acscentsci.7b00064)] [[**CrossRef**](https://doi.org/10.1021/acscentsci.7b00064)][[**Green Version**](https://www.mdpi.com/1420-3049/25/22/5277)]

27.Lee, A.A.; Yang, Q.; Sresht, V.; Bolgar, P.; Hou, X.; Klug-McLeod, J.L.; Butler, C.R. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chem. Commun.* **2019**, *55*, 12152–12155. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Molecular+Transformer+unifies+reaction+prediction+and+retrosynthesis+across+pharma+chemical+space&author=Lee,+A.A.&author=Yang,+Q.&author=Sresht,+V.&author=Bolgar,+P.&author=Hou,+X.&author=Klug-McLeod,+J.L.&author=Butler,+C.R.&publication_year=2019&journal=Chem.+Commun.&volume=55&pages=12152%E2%80%9312155&doi=10.1039/C9CC05122H)] [[**CrossRef**](https://doi.org/10.1039/C9CC05122H)]

28.Reher, R.; Kim, H.W.; Zhang, C.; Mao, H.H.; Wang, M.; Nothias, L.F.; Caraballo-Rodriguez, A.M.; Glukhov, E.; Teke, B.; Leao, T.; et al. A Convolutional Neural Network-Based Approach for the Rapid Annotation of Molecularly Diverse Natural Products. *J. Am. Chem. Soc.* **2020**, *142*, 4114–4120. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=A+Convolutional+Neural+Network-Based+Approach+for+the+Rapid+Annotation+of+Molecularly+Diverse+Natural+Products&author=Reher,+R.&author=Kim,+H.W.&author=Zhang,+C.&author=Mao,+H.H.&author=Wang,+M.&author=Nothias,+L.F.&author=Caraballo-Rodriguez,+A.M.&author=Glukhov,+E.&author=Teke,+B.&author=Leao,+T.&publication_year=2020&journal=J.+Am.+Chem.+Soc.&volume=142&pages=4114%E2%80%934120&doi=10.1021/jacs.9b13786)] [[**CrossRef**](https://doi.org/10.1021/jacs.9b13786)]

29. Noord N., Postma E. Learning scale-variant and scale-invariant features for deep image classification. Pattern Recognit. 2017;**61**:583–592. doi: 10.1016/j.patcog.2016.06.005. [[CrossRef](https://doi.org/10.1016%252Fj.patcog.2016.06.005)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Pattern+Recognit.&title=Learning+scale-variant+and+scale-invariant+features+for+deep+image+classification&author=N.+Noord&author=E.+Postma&volume=61&publication_year=2017&pages=583-592&doi=10.1016/j.patcog.2016.06.005&)]

30. Sumithra R., Suhil M., Guru D.S. Segmentation and classification of skin lesions for disease diagnosis. Procedia Comput. Sci. 2015;**45**:76–85. doi: 10.1016/j.procs.2015.03.090. [[CrossRef](https://doi.org/10.1016%252Fj.procs.2015.03.090)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Procedia+Comput.+Sci.&title=Segmentation+and+classification+of+skin+lesions+for+disease+diagnosis&author=R.+Sumithra&author=M.+Suhil&author=D.S.+Guru&volume=45&publication_year=2015&pages=76-85&doi=10.1016/j.procs.2015.03.090&)]

31. Zhang X., Wang S., Liu J., Tao C. Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. BMC Med. Inform. Decis. Mak. 2018;**18**:59. doi: 10.1186/s12911-018-0631-9. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069289/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/30066649)] [[CrossRef](https://doi.org/10.1186%252Fs12911-018-0631-9)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=BMC+Med.+Inform.+Decis.+Mak.&title=Towards+improving+diagnosis+of+skin+diseases+by+combining+deep+neural+network+and+human+knowledge&author=X.+Zhang&author=S.+Wang&author=J.+Liu&author=C.+Tao&volume=18&publication_year=2018&pages=59&pmid=30066649&doi=10.1186/s12911-018-0631-9&)]

32.Bielza, C.; Larrañaga, P. Discrete Bayesian Network Classifiers: A Survey. *ACM Comput. Surv.* **2014**, *47*, 43. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Discrete+Bayesian+Network+Classifiers:+A+Survey&author=Bielza,+C.&author=Larra%C3%B1aga,+P.&publication_year=2014&journal=ACM+Comput.+Surv.&volume=47&pages=43&doi=10.1145/2576868)] [[**CrossRef**](https://doi.org/10.1145/2576868)]

33.Gilboa, E.; Saatçi, Y.; Cunningham, J.P. Scaling Multidimensional Inference for Structured Gaussian Processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Scaling+Multidimensional+Inference+for+Structured+Gaussian+Processes&author=Gilboa,+E.&author=Saat%C3%A7i,+Y.&author=Cunningham,+J.P.&publication_year=2013&journal=IEEE+Trans.+Pattern+Anal.+Mach.+Intell.&doi=10.1109/TPAMI.2013.192&pmid=26353252)] [[**CrossRef**](https://doi.org/10.1109/TPAMI.2013.192)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/26353252)][[**Green Version**](http://mlg.eng.cam.ac.uk/pub/pdf/GilSaaCun13.pdf)]

34.Sun, H. A naive bayes classifier for prediction of multidrug resistance reversal activity on the basis of atom typing. *J. Med. Chem.* **2005**, *48*, 4031–4039. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=A+naive+bayes+classifier+for+prediction+of+multidrug+resistance+reversal+activity+on+the+basis+of+atom+typing&author=Sun,+H.&publication_year=2005&journal=J.+Med.+Chem.&volume=48&pages=4031%E2%80%934039&doi=10.1021/jm050180t&pmid=15943476)] [[**CrossRef**](https://doi.org/10.1021/jm050180t)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/15943476)]

35.Ratanamahatana, C.A.; Gunopulos, D. Feature selection for the naive bayesian classifier using decision trees. *Appl. Artif. Intell.* **2010**. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Feature+selection+for+the+naive+bayesian+classifier+using+decision+trees&author=Ratanamahatana,+C.A.&author=Gunopulos,+D.&publication_year=2010&journal=Appl.+Artif.+Intell.&doi=10.1080/713827175)] [[**CrossRef**](https://doi.org/10.1080/713827175)]

36.Kim, S.-B.; Han, K.-S.; Rim, H.-C.; Myaeng, S.-H. Some Effective Techniques for Naive Bayes Text Classification. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 1457–1466. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Some+Effective+Techniques+for+Naive+Bayes+Text+Classification&author=Kim,+S.-B.&author=Han,+K.-S.&author=Rim,+H.-C.&author=Myaeng,+S.-H.&publication_year=2006&journal=IEEE+Trans.+Knowl.+Data+Eng.&volume=18&pages=1457%E2%80%931466&doi=10.1109/TKDE.2006.180)] [[**CrossRef**](https://doi.org/10.1109/TKDE.2006.180)]

37.Anagaw, A.; Chang, Y.-L. A new complement naïve Bayesian approach for biomedical data classification. *J. Ambient Intell. Hum. Comput.* **2019**, *10*, 3889–3897. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=A+new+complement+na%C3%AFve+Bayesian+approach+for+biomedical+data+classification&author=Anagaw,+A.&author=Chang,+Y.-L.&publication_year=2019&journal=J.+Ambient+Intell.+Hum.+Comput.&volume=10&pages=3889%E2%80%933897&doi=10.1007/s12652-018-1160-1)] [[**CrossRef**](https://doi.org/10.1007/s12652-018-1160-1)]

38. Yadav S.S., Jadhav S.M. Deep convolutional neural network based medical image classification for disease diagnosis. J. Big Data. 2019;**6**:1–18. doi: 10.1186/s40537-019-0276-2. [[CrossRef](https://doi.org/10.1186%252Fs40537-019-0276-2)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=J.+Big+Data&title=Deep+convolutional+neural+network+based+medical+image+classification+for+disease+diagnosis&author=S.S.+Yadav&author=S.M.+Jadhav&volume=6&publication_year=2019&pages=1-18&doi=10.1186/s40537-019-0276-2&)]

39. Hameed N., Shabut A.M., Hossain M.A. Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine; Proceedings of the 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA); Phnom Penh, Cambodia. 3–5 December 2018; pp. 1–7. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Proceedings+of+the+12th+International+Conference+on+Software,+Knowledge,+Information+Management+&+Applications+(SKIMA)&title=Multi-Class+Skin+Diseases+Classification+Using+Deep+Convolutional+Neural+Network+and+Support+Vector+Machine&author=N.+Hameed&author=A.M.+Shabut&author=M.A.+Hossain&pages=1-7&)]

40.Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Random+Forests&author=Breiman,+L.&publication_year=2001&journal=Mach.+Learn.&volume=45&pages=5%E2%80%9332&doi=10.1023/A:1010933404324)] [[**CrossRef**](https://doi.org/10.1023/A:1010933404324)][[**Green Version**](https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf)]

41.Cano, G.; Garcia-Rodriguez, J.; Garcia-Garcia, A.; Perez-Sanchez, H.; Benediktsson, J.; Thapa, A.; Barr, A. Automatic selection of molecular descriptors using random forest: Application to drug discovery. *Expert Syst. Appl.* **2017**, *72*, 151–159. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Automatic+selection+of+molecular+descriptors+using+random+forest:+Application+to+drug+discovery&author=Cano,+G.&author=Garcia-Rodriguez,+J.&author=Garcia-Garcia,+A.&author=Perez-Sanchez,+H.&author=Benediktsson,+J.&author=Thapa,+A.&author=Barr,+A.&publication_year=2017&journal=Expert+Syst.+Appl.&volume=72&pages=151%E2%80%93159&doi=10.1016/j.eswa.2016.12.008)] [[**CrossRef**](https://doi.org/10.1016/j.eswa.2016.12.008)]

42.Bielza, C.; Larrañaga, P. Discrete Bayesian Network Classifiers: A Survey. *ACM Comput. Surv.* **2014**, *47*, 43. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Discrete+Bayesian+Network+Classifiers:+A+Survey&author=Bielza,+C.&author=Larra%C3%B1aga,+P.&publication_year=2014&journal=ACM+Comput.+Surv.&volume=47&pages=43&doi=10.1145/2576868)] [[**CrossRef**](https://doi.org/10.1145/2576868)]

43.Gilboa, E.; Saatçi, Y.; Cunningham, J.P. Scaling Multidimensional Inference for Structured Gaussian Processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Scaling+Multidimensional+Inference+for+Structured+Gaussian+Processes&author=Gilboa,+E.&author=Saat%C3%A7i,+Y.&author=Cunningham,+J.P.&publication_year=2013&journal=IEEE+Trans.+Pattern+Anal.+Mach.+Intell.&doi=10.1109/TPAMI.2013.192&pmid=26353252)] [[**CrossRef**](https://doi.org/10.1109/TPAMI.2013.192)] [[**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed/26353252)][[**Green Version**](http://mlg.eng.cam.ac.uk/pub/pdf/GilSaaCun13.pdf)]

44.Rathi, P.C.; Ludlow, R.F.; Verdonk, M.L. Practical High-Quality Electrostatic Potential Surfaces for Drug Discovery Using a Graph-Convolutional Deep Neural Network. *J. Med. Chem.* **2020**, *63*, 8778–8790. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Practical+High-Quality+Electrostatic+Potential+Surfaces+for+Drug+Discovery+Using+a+Graph-Convolutional+Deep+Neural+Network&author=Rathi,+P.C.&author=Ludlow,+R.F.&author=Verdonk,+M.L.&publication_year=2020&journal=J.+Med.+Chem.&volume=63&pages=8778%E2%80%938790&doi=10.1021/acs.jmedchem.9b01129)] [[**CrossRef**](https://doi.org/10.1021/acs.jmedchem.9b01129)][[**Green Version**](https://pubs.acs.org/doi/pdf/10.1021/acs.jmedchem.9b01129)]

45.Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D.R. Protein–Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Protein%E2%80%93Ligand+Scoring+with+Convolutional+Neural+Networks&author=Ragoza,+M.&author=Hochuli,+J.&author=Idrobo,+E.&author=Sunseri,+J.&author=Koes,+D.R.&publication_year=2017&journal=J.+Chem.+Inf.+Model.&volume=57&pages=942%E2%80%93957&doi=10.1021/acs.jcim.6b00740)] [[**CrossRef**](https://doi.org/10.1021/acs.jcim.6b00740)][[**Green Version**](http://europepmc.org/articles/pmc5479431?pdf=render)]

46.Mohanty, S.; Rashid, M.H.A.; Mridul, M.; Mohanty, C.; Swayamsiddha, S. Application of Artificial Intelligence in COVID-19 drug repurposing. *Diabetes Metab. Syndr.* **2020**, *14*, 1027–1031. [[**Google Scholar**](https://scholar.google.com/scholar_lookup?title=Application+of+Artificial+Intelligence+in+COVID-19+drug+repurposing&author=Mohanty,+S.&author=Rashid,+M.H.A.&author=Mridul,+M.&author=Mohanty,+C.&author=Swayamsiddha,+S.&publication_year=2020&journal=Diabetes+Metab.+Syndr.&volume=14&pages=1027%E2%80%931031&doi=10.1016/j.dsx.2020.06.068)] [[**CrossRef**](https://doi.org/10.1016/j.dsx.2020.06.068)]