e-ISSN: 2687-6698 **Review**

Turkish Journal of Analytical Chemistry
www.turkjac.org
2019
TurkJAC

Theoretical and applied potential of artificial intelligence and machine learning in analysing molecular data

Fatih Mehmet Avcu

Inonu University, Department of Informatics, 44280, Malatya, Turkey

Abstract

This manuscript serves as a comprehensive review paper examining the theoretical potential and practical applications of artificial intelligence (AI) and machine learning (ML) in molecular analysis. The purpose of this study is to synthesize existing knowledge, highlight advancements in AI/ML algorithms for molecular structure prediction, and identify future directions in the field. The significance of this review lies in its focus on bridging molecular data with AI/ML methodologies, aiming to accelerate chemical and biological research processes with enhanced accuracy. This article examines how AI and ML techniques allow accelerating and improving the accuracy of chemical and biological processes. In particular, these methods are used to predict the chemical structure, biological activity, and protein structure of molecules. Various data types such as molecular dynamics simulations, spectroscopy, and cheminformatics data are discussed in relation to AI and ML algorithms. Additionally, the revolutionary contributions of deep learning algorithms in areas such as molecular representations, drug design, and protein structure prediction are highlighted. The effectiveness of reinforcement learning and graph-based models in the prediction and optimization of chemical reactions is also discussed. In conclusion, the use of AI and ML in molecular analyses is expected to expand into broader areas of scientific and industrial research in the future.

Keywords: Artificial intelligence, molecular analysis, machine learning, molecular analysis, deep learning, chemical reactions

1. Introduction

1.1. Definition of artificial intelligence and machine learning

Definition of Artificial Intelligence and Machine Learning Artificial intelligence (AI) and machine learning (ML) have become the focus of scientific research in recent years and have brought about significant changes in complex and data-intensive processes such as molecular analysis. AI is defined as the imitation of complex tasks based on human intelligence by computer systems, while machine learning refers to the ability of these systems to improve themselves by learning from data [1]. ML, as a sub-branch of AI, enables computers to learn patterns and relationships from data sets, thus enabling them to continuously improve their performance on specific tasks. In particular, supervised and unsupervised learning methods are widely used in molecular analysis [2].

1.2. The power of machine learning algorithms

One of the advantages of ML algorithms is the ability to work effectively on large data sets. While traditional data analysis methods are usually successful with limited sample sizes, ML algorithms go beyond these limitations and can process huge data pools [3]. Especially in molecular analysis, this means that algorithms can analyze the chemical and physical properties of thousands or even millions of molecules. In addition, these algorithms play an active role in predicting the potential behavior and reactions of molecules [4].

However, ML algorithms continuously improve the learning processes of artificial intelligence systems. For example, deep learning techniques provide more accurate results by better understanding structural details in molecular image analysis [5]. This allows molecular analysis to be more predictable, fast and efficient.

Citation: F.M. Avcu, Theoretical and applied potential of artificial intelligence and machine learning in analysing molecular data, Turk J Anal Chem, 7(1), 2025, 61–70.

***Author of correspondence:** fatihavcu@gmail.com
Tel: N/A **Fax:** N/A

Received: December 25, 2024 **Accepted:** January 23, 2025
https://doi.org/10.51435/turkjac.1607205

2. Historical development of artificial intelligence and machine learning

2.1. Early studies

The concepts of artificial intelligence and machine learning were born in the mid-20th century at the intersection of mathematics, computer science, and neurology. First, Alan Turing's "Turing Test" and John von Neumann's theories of automatic computation introduced the idea that computers could exhibit human-like intelligence [6]. Simple algorithms developed in the 1950s and 1960s worked with limited data sets to perform specific tasks. Artificial intelligence studies at the time focused more on symbolic logic and rule-based systems [1].

Machine learning began to take shape in the late 1950s. Arthur Samuel first used the term "machine learning" in 1959 while working at IBM to explain that computers could learn from experience and improve [7]. Samuel's checkers-playing program embodied the basic principles of ML in that it learned from its own mistakes and improved its performance over time. This is considered the first example of the ability to "learn" which is the basis of today's machine learning algorithms [8].

2.2. Current practices

Today's artificial intelligence and machine learning technologies are quite different from the rule-based systems of the early days. In particular, new generation techniques such as deep learning and artificial neural networks have made great strides thanks to their capacity to handle large data sets. After the 2010s, deep learning has achieved significant success in areas such as image and voice recognition. In 2012, a deep learning model called AlexNet revolutionized the field by achieving great success in image recognition in the ImageNet competition [9].

Machine learning algorithms have become widely used, especially in bioinformatics and molecular analysis. Methods such as artificial neural networks, support vector machines, and random forest algorithms have made great advances in areas such as the prediction of properties of chemical compounds, protein structure prediction, and drug discovery [10,11]. At the beginning of the 21st century, the increase in computing power and the availability of large datasets enabled the rapid development of artificial intelligence and machine learning technologies [12].

3. Data types used in molecular analysis

3.1. Spectroscopy data

Spectroscopy is one of the most widely used techniques in molecular analysis, and spectral data is used to analyze the physical and chemical properties of molecules. These data contain important information to understand the molecule's energy, electron distribution, and bond structures [13]. Traditionally, analyzing this data can be time-consuming and error-prone. However, when artificial intelligence and machine learning algorithms have been used to process this data, the results have become much faster and more accurate [14].

For example, Raman spectroscopy data can be analyzed with machine learning algorithms and used for molecular structure prediction. Such algorithms save both time and cost by extracting meaningful patterns from complex data [15]. In addition, the capacity of ML algorithms to detect anomalies while analyzing spectral data is an important factor that increases the accuracy of molecular analysis.

3.2. Molecular dynamics data

Molecular dynamics is another important field that studies the motion of molecules over time. The data obtained in this field are of large size and complexity. Molecular dynamics simulations are often used to analyze the interactions of thousands of molecules, and these analyses are computationally intensive [16]. Machine learning algorithms stand out as an excellent tool for analyzing these complex data structures [17].

In particular, deep learning methods have been effective in processing molecular dynamics data and have provided important insights into new molecular structures and dynamic behaviors [18]. By processing this data, ML allows the prediction of molecular behaviors and the prediction of the properties of new molecules.

3.3. Cheminformatics data

Cheminformatics involves the use of computer technologies to store, process, and analyze chemical information for molecular analysis [19]. The data types in this field are usually based on the structural properties of molecules. Data representations such as ECFP (Extended-Connectivity FingerPrint) and Morgan FingerPrint are used to describe different structural features of molecules. These data representations are analyzed by artificial intelligence and machine learning algorithms and used in areas such as classification of molecular structures, drug discovery, and toxicity prediction [20].

ML algorithms are also used to identify similarities and differences between molecules in cheminformatics data. This is particularly important for identifying new drug candidates and predicting chemical reactions. By working with large cheminformatics datasets, AI systems can predict the behavior of new molecules and accelerate drug discovery processes [21].

4. Use of artificial intelligence and machine learning algorithms in molecular analysis

4.1. Support Vector Machines (SVM)

Support Vector Machines (SVM) is a powerful machine learning algorithm widely used for classification and regression tasks. In molecular analysis, SVM algorithms have been used with particular success in tasks such as classifying molecular structures and predicting molecular properties. Briefly, the algorithm draws a line to separate points on the training data placed on a plane, and this line is planned to be the maximum distance for the points of the two classes to be separated (Fig. 1). SVM works by creating a hyperplane to separate the data into two classes and can model complex relationships between molecules. It is a method frequently used in molecular biology, especially in areas such as prediction of protein-protein interactions and classification of drug candidates [22].

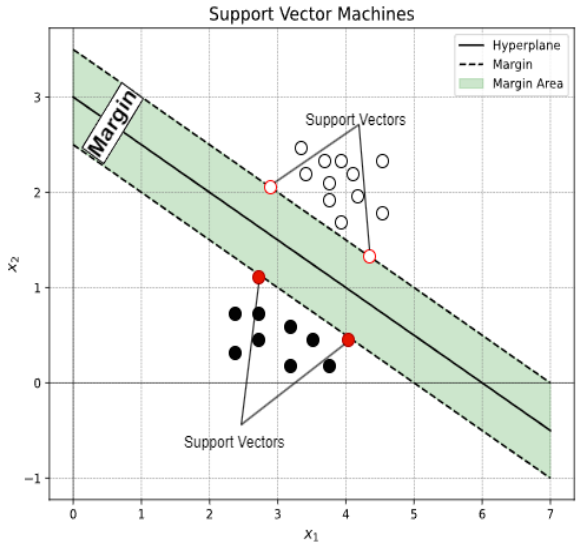

Figure 1. Support Vector Machine Algorithm

One of the biggest advantages of SVM in molecular analysis is that it can work effectively even with small data sets. While molecular analyses usually involve large datasets, in some rare cases there may be a limited number of data samples. In such cases, SVM minimizes the risk of overfitting and produces more accurate results [23]. Especially in drug discovery studies with small data sets, SVM stands out as an effective classification method [24].

4.2. Decision trees and random forests

Decision trees are an algorithm that extracts decision rules by partitioning data sets and thus performing classification or regression (Fig. 2). It is used in molecular analysis, especially in areas such as classification of compounds, toxicity prediction, and characterization of biomolecules [1]. Decision trees make distinctions between molecules based on each feature of the data set and thus classify the results.

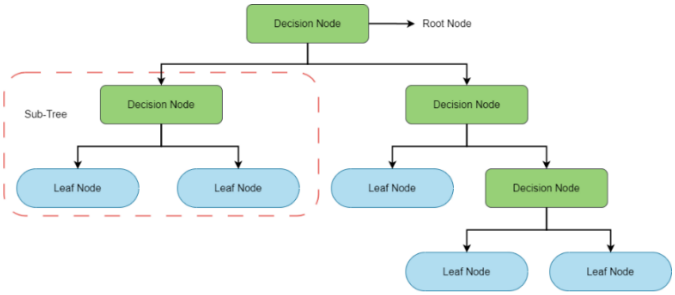

Figure 2. Decision Tree Algorithm

Random forests are a machine learning algorithm in which multiple decision trees work together to make more accurate predictions (Fig. 3). The random forest algorithm is widely used in molecular analysis, often applied in toxicity prediction, predicting the activity of new molecules and analyzing the dynamics of chemical reactions [25]. The random forest algorithm shows high performance on large datasets by analyzing a large number of molecules simultaneously. Since the data density in molecular analysis is quite high, the random forest algorithm stands out as an important tool that can work effectively with big data [26].

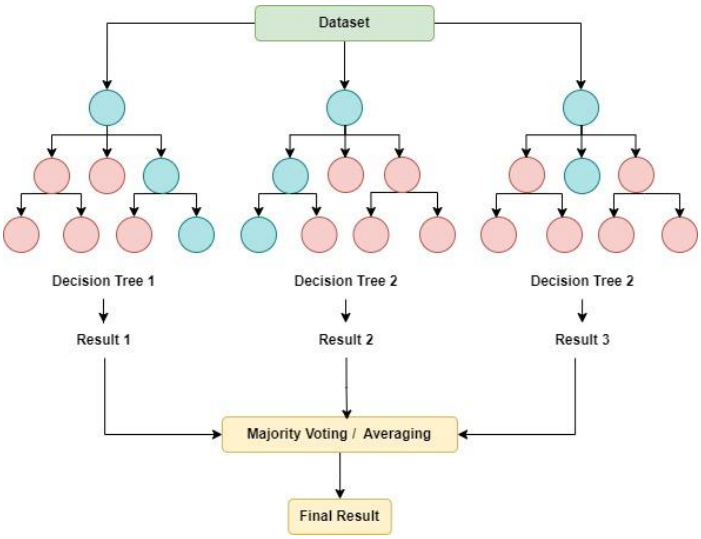

Figure 3. Random Forest Algorithm

4.3. Artificial neural networks (ANN)

Artificial neural networks are an algorithm based on data learning and classification that mimics the working principles of nerve cells in the human brain (Fig. 4). Artificial neural networks, which form the basis of deep learning methods, have become a powerful tool in

molecular analysis. In particular, multilayer perceptrons (MLP) and deep learning methods are used for modeling molecular structures and predicting complex relationships [5].

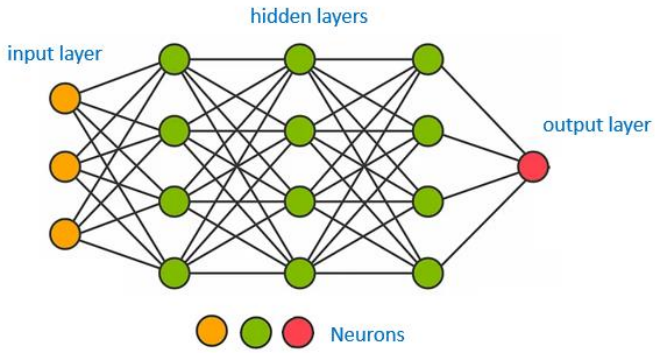

Figure 4. Artificial Neural Networks (ANN)

MLP and deep neural networks can learn patterns in molecular data through a large number of hidden layers. Therefore, they are effectively used to analyze chemical structures, protein interactions, and relationships between biomolecules. For example, neural networks are highly successful in modeling chemical reaction dynamics, drug discovery and predicting the behavior of biomolecules [27]. Moreover, the performance of deep neural networks in analyzing complex structures makes these algorithms indispensable for the future of molecular analysis [28].

4.4. Deep learning and convolutional neural networks (CNN)

Deep learning has achieved great success, especially in analyzing large data sets. Convolutional Neural Networks (CNN) are particularly effective deep learning algorithm for processing image data and are also used in molecular analysis (Fig. 5). CNN gives successful results in visualization of molecular structures, modeling of protein structures, and biomolecular image analysis [29].

CNN algorithms are widely used for 3D structure recognition and image analysis of biomolecules. In particular, CNN plays an important role in analyzing the folding process of proteins, molecular dynamics simulations, and studying the interactions of drug molecules with biomolecules [30]. Effective processing and analysis of such big data in molecular analysis has become possible with deep learning algorithms.

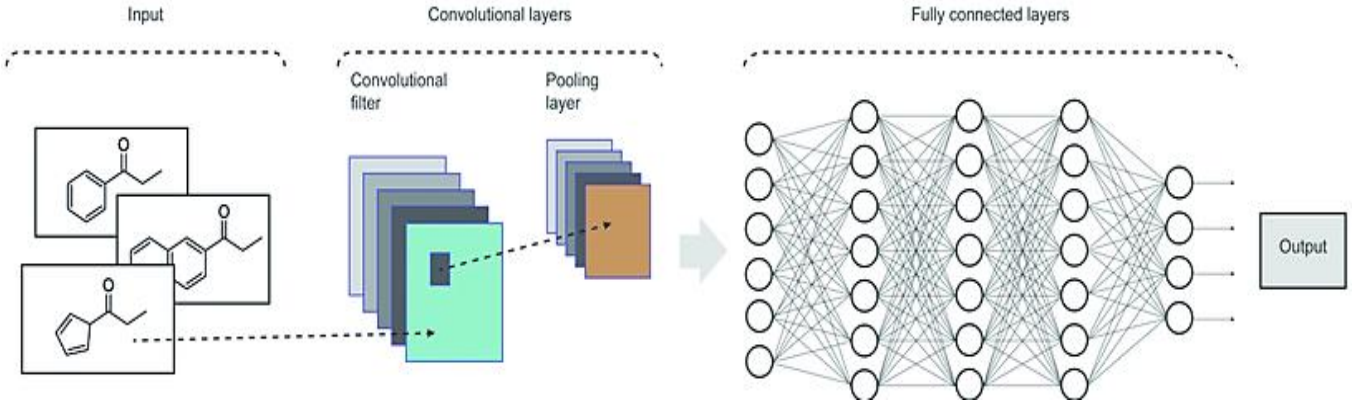

Figure 5. Convolutional Neural Networks

4.5. Reinforcement learning

Reinforcement learning is a machine learning methodology based on interactions between an agent (computer program) and its environment (Fig. 6). It is used in molecular analysis, especially in areas such as optimization of chemical reactions and design of new molecules [31]. Reinforcement learning algorithms can make better predictions about molecular structures and chemical reactions by working on a reward and punishment system. It is an effective method for finding optimal solutions, especially in chemistry and biomolecular processes.

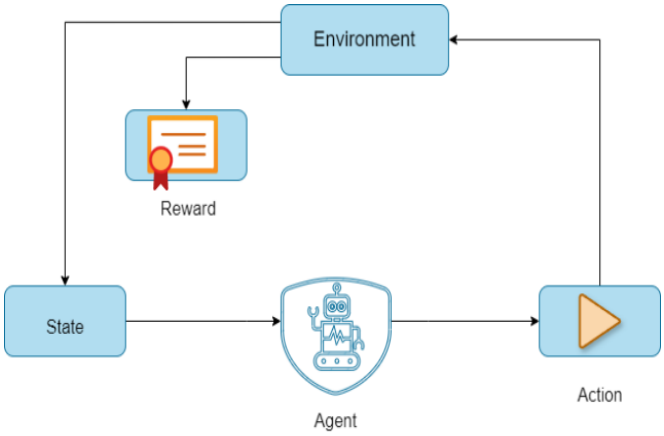

Figure 6. Reinforcement Learning

For example, during the discovery of new drug molecules, reinforcement learning algorithms can help identify the most efficient chemical structures by optimizing molecular interactions. They can also be used in molecular dynamics simulations to find the most efficient paths of molecules during a given reaction [32]. The flexible nature and learning capabilities of these algorithms can be widely applied in different areas of molecular analysis.

5. The role of artificial intelligence and machine learning in predicting molecular structures

5.1. Chemical structure prediction

Artificial intelligence and machine learning have an increasing importance in the prediction of chemical structures. In particular, the determination of molecular structures and the association of these structures with certain properties can be done more precisely and quickly thanks to artificial intelligence algorithms. For example, the high cost of calculations using quantum chemistry methods can be greatly reduced by artificial intelligence and machine learning algorithms.

In particular, graph-based machine learning algorithms are widely used to predict structural properties of molecules using graph representations of chemical compounds. By analyzing molecular graph representations by artificial intelligence models, the structure and properties of new molecules can be predicted. This offers great potential in areas such as drug discovery and materials science [33].

5.2. Protein structure prediction

Protein structure prediction is one of the common uses of artificial intelligence and machine learning algorithms in biomolecular research. Especially the folding processes of proteins and the prediction of 3D structures pose great challenges in this field. However, deep learning-based AI algorithms such as AlphaFold have revolutionized protein structure prediction [30]. AlphaFold has contributed to the advancement of biomolecular research by predicting the 3D structures of proteins with high accuracy.

Protein structure prediction plays a critical role, especially in drug discovery and understanding biological processes. By analyzing protein-protein interactions, artificial intelligence algorithms allow the discovery of new biomolecules and a better understanding of biological processes [30].

6. Applications of artificial intelligence and machine learning in molecular dynamics simulations

Accurate prediction and optimization of molecular properties is critical in drug discovery, materials science, and biomolecular processes. Artificial intelligence (AI) and machine learning (ML) techniques are used to predict the chemical and biological properties of molecules, providing effective results in these processes. Especially in the fields of molecular dynamics, quantum chemistry, and molecular design, AI and ML offer significant advantages in terms of speed and accuracy compared to conventional methods [17,33].

6.1. Molecular activity prediction

Prediction of molecular activity plays a major role, especially in drug discovery processes. Prediction of the effects of molecules in biological systems can be time-consuming and costly with traditional experimental methods. To overcome these challenges, AI and ML algorithms have been used to predict molecular properties and biological activities [27].

One of the common methods used in molecular activity prediction is to determine the electronic structures of the molecule by quantum mechanics-based calculations and analyze how these structures can be related to biological activity. Artificial neural networks and support vector machines (SVM) are powerful tools for understanding the complex relationships between molecular properties and biological activities [34]. For example, QSPR (Quantitative Structure-Activity Relationship) models are widely used to determine the relationships of molecular structures with biological activities [35].

6.2. Molecule design and optimization

The design and optimization of new molecules is of vital importance in fields such as drug discovery and materials science. ML algorithms enable the analysis of chemical properties and the generation of optimized chemical structures that can be used in the design of new molecules. Reinforcement learning and genetic algorithms are some of the most popular methods used in this field [32].

Reinforcement learning is used to optimize the properties of molecules and design more effective structures using a reward and punishment system. This method can be effective in predicting the effects of new molecules on biological systems. Reinforcement learning algorithms are frequently used to create new chemical structures and ensure that these structures exhibit specific biological activities [36].

Especially in molecule design processes, artificial intelligence and machine learning algorithms provide a better understanding and optimization of chemical and biological processes. Deep learning algorithms make important contributions to the discovery of new biomolecules by analyzing molecular structures [37].

6.3. Quantum chemistry and artificial intelligence

Quantum chemistry is a fundamental tool used to determine the electronic properties of molecular structures and to understand chemical reactions. However, since such calculations often require high computational power, AI and ML algorithms are used to

speed up quantum chemistry calculations and analyze larger datasets [17].

In particular, deep learning methods are an effective tool for optimizing quantum chemistry calculations of molecular structures. Deep learning algorithms can be used to predict molecular energy levels, orbital structures, and other electronic properties. Such algorithms reduce computational costs in quantum chemistry, allowing for faster and more precise predictions [18].

6.4. Molecular dynamics simulations

Molecular dynamics simulations are a technique used to understand the motions and interactions of biomolecules. These simulations play a major role in analyzing biomolecular processes. However, such simulations are often very time-consuming and require high computational power. AI and ML algorithms have been used to accelerate molecular dynamics simulations and analyze more complex biomolecular systems [38].

AI-based methods reduce simulation times by predicting the motion of biomolecules. Deep learning algorithms can predict the evolution of molecular structures over time by learning the dynamics of biomolecular processes [39]. Especially the simulation of protein-protein interactions and drug-protein interactions can be done faster and more precisely with such algorithms.

6.5. Optimization of chemical reactions

Optimization of chemical reactions is an important challenge in chemistry and materials science. Artificial intelligence and machine learning algorithms are used to make chemical reactions more efficient and optimize them. For example, ML algorithms can be used to predict the probability and conditions under which a given reaction will occur. Such algorithms optimize reaction rates, making chemical processes more efficient [40].

ML algorithms are used to discover and optimize new pathways, especially in chemical synthesis. In the analysis of chemical reactions, deep learning methods allow a better understanding of reaction mechanisms and can be used to discover new reaction pathways [41].

7. The role of artificial intelligence and machine learning in predicting chemical reactions

Accurate prediction of chemical reactions is of great importance in the fields of chemistry and biochemistry. AI and ML are increasingly being used in these prediction processes, offering new ways to better understand the kinetics, thermodynamics, and mechanisms of reactions.

7.1. Chemical reaction kinetics and thermodynamics

Chemical reaction kinetics studies the rate of a reaction and what factors influence this rate. While traditional methods need experimental data to predict reaction rates, artificial intelligence and machine learning can predict these processes without the need for experimental data. For example, deep learning models are used to predict rate constants and energy barriers of chemical reactions [40]. This greatly accelerates experimental processes and provides a great advantage for the discovery of new chemical reactions.

Thermodynamic properties of chemical reactions can also be predicted with machine learning algorithms. Free energy calculations and thermodynamic equilibria can be calculated faster and more accurately with ML models [42]. Thermodynamic equilibrium determines which products chemical reactions lead to and how stable these products are. Such calculations can be much less costly and time efficient than traditional methods.

7.2. Analysis of mechanisms of chemical reactions

The mechanisms of chemical reactions are processes that explain how bonds between molecules are broken and how new bonds are formed. Accurate modeling of these mechanisms is of great importance, especially in areas such as drug design and the discovery of new materials. Artificial intelligence and machine learning algorithms offer powerful tools for modeling and predicting the mechanisms of chemical reactions. Graph-based models and natural language processing (NLP) techniques can contribute to a better understanding of chemical reaction mechanisms [43].

By analyzing the steps of chemical reactions, graph-based machine learning algorithms can accurately predict which molecules will turn into which products. These algorithms have accelerated the understanding of organic reactions and the discovery of new chemical pathways. This provides a significant advantage in the discovery of new drugs and materials. At the same time, NLP techniques automatically analyze the mechanisms of chemical reactions found in the scientific literature, making the understanding of these processes faster and more efficient.

7.3. Optimization of chemical reactions

Optimization of chemical reactions involves determining the conditions necessary to make a given reaction more efficient. This includes optimizing temperature, pressure, catalyst usage, and other reaction conditions. Artificial intelligence and machine learning algorithms are becoming increasingly common in these

optimization processes. In particular, reinforcement learning algorithms are used to determine the optimal conditions for chemical reactions [44].

Reinforcement learning helps to identify the optimal conditions to increase the efficiency of a reaction using a reward-punishment mechanism. For example, it is possible to determine which catalyst performs best for a chemical reaction to occur with maximum efficiency and at what temperature the reaction proceeds fastest. These optimization processes can lead to huge economic gains by increasing efficiency, especially in areas such as industrial chemistry and pharmaceutical production.

8. Contributions of deep learning to molecular modeling

In recent years, deep learning (DL) algorithms have made great progress in the fields of artificial intelligence and machine learning and have revolutionized many scientific fields. Molecular modeling is one of these fields. Deep learning offers significant advantages in modeling, prediction, and simulation of complex molecular structures. In this chapter, the contributions of deep learning methods in molecular modeling and biomolecular processes will be discussed.

8.1. Use of deep learning in molecule representation

Accurately representing the structural properties of molecules is a critical step for the success of artificial intelligence and machine learning models. While traditional molecular representation methods deal with molecules in data formats such as simple line structures or atomic coordinates, deep learning algorithms can produce more complex and meaningful representations. In particular, graph-based deep learning methods represent the atomic and bond structure of molecules as graph nodes and edges [33].

Such graph-based representations allow for more accurate prediction of chemical reactions and biological activities of molecules. Methods such as Message Passing Neural Networks (MPNNs) can perform learning based on graph representations of molecules and model the properties of chemical structures. This is recognized as an important innovation in molecular simulations [18].

8.2. Prediction of molecular properties with deep learning

Deep learning algorithms achieve successful results in predicting the electronic and chemical properties of molecules. Especially in quantum chemistry calculations, deep learning methods can predict the energy levels, polarizations, and orbital structures of molecules. Such predictions are faster and less costly compared to traditional computational methods[12].

For example, predicting molecular energy levels using deep neural networks (DNNs) and convolutional neural networks (CNNs) replaces quantum chemistry calculations and accelerates the simulation of chemical reactions. Such prediction models allow to more accurately model the probability of chemical reactions and their outcomes [18].

8.3. Drug design with deep learning

Drug design processes rely on accurate analysis of chemical and biological interactions. Deep learning algorithms play an important role in the design of new drug molecules and the prediction of their interactions with biological targets. Especially in de novo drug design processes, deep learning methods accelerate and optimize the discovery of new biologically active molecules [32].

The genetic algorithms used in these processes can be integrated with deep learning models to predict how new molecules will function in biological systems. In particular, generative models are used to create new molecular structures suitable for biological targets. This has ushered in a new era in the design of biomolecules [37].

8.4. Protein structure prediction and deep learning

Protein structure prediction is an important field for understanding biomolecular processes. The three-dimensional structures of proteins determine their biological activity and function. Deep learning algorithms have made great advances in the field of protein structure prediction. In particular, deep learning models such as AlphaFold have the capacity to accurately predict the three-dimensional structure of proteins [30].

AlphaFold has revolutionized the field of protein structure prediction, achieving significant success in solving the protein folding problem. This model uses deep neural networks to predict the final three-dimensional structure of proteins by analyzing protein sequences. This has ushered in a new era in protein design and understanding of biomolecular interactions [30].

8.5. Accelerating molecular dynamics with deep learning

Molecular dynamics simulations are a powerful technique used to model the evolution of biological and chemical processes over time. However, these simulations often have high computational costs. Deep learning algorithms offer effective tools to speed up these simulations [39].

In particular, deep learning methods such as convolutional neural networks and message passing neural networks can predict the dynamic motions of biomolecules and allow molecular dynamics simulations to be made more efficient. This enables faster and more accurate analysis of biomolecular processes [18]. Deep learning also plays an important role in predicting biomolecular interactions and modeling chemical reactions more accurately [39].

9. Performance comparison of algorithms

Machine learning algorithms exhibit superior performances in molecular analysis for different data types and application areas. The advantages of each algorithm, the parameters used, and their success rates are evaluated in a wide range of applications from molecular dynamics simulations to protein structure prediction.

Table 1 summarizes the effectiveness of different algorithms in various applications and provides an opportunity to compare performance. For example, SVM work effectively on small data sets, while RF can achieve high accuracy rates on large data sets and minimize the risk of overlearning. Deep learning methods, on the other hand, stand out with high success rates, especially in the analysis of complex biomolecular structures.

This benchmark aims to guide researchers in algorithm selection and facilitate the identification of the most appropriate methods for molecular analysis. Table 1 summarizes the performance of the algorithms mentioned throughout the paper, providing the reader with an application-oriented perspective.

This table compares the performance of different machine learning algorithms in molecular analysis. It contains information about the application areas of the algorithms, the parameters used, and their success rates. Researchers can use this table to select the most appropriate algorithm for their own studies.

10. Conclusion

This article addresses the theoretical potential and current applications of AI and ML in molecular analysis. The examples reviewed show that AI and ML have enabled a significant transformation in predicting the chemical structures, biological activities, and protein structures of molecules. In particular, these technologies are characterized by increased speed and accuracy in experimental processes. The success of deep learning, reinforcement learning, and graph-based models in modeling the kinetics, thermodynamics, and mechanisms of chemical reactions indicates that these areas will become even more important in the future.

In the future, the impact of AI and ML in molecular analyses will deepen even further. The integration of quantum computing and AI models will offer a significant improvement, especially in the simulation of chemical processes. This integration will allow more complex molecules and reactions to be simulated with high accuracy, opening up new areas of discovery for researchers. The proliferation of quantum artificial intelligence algorithms could have groundbreaking consequences in the fields of chemistry and biochemistry.

Furthermore, improvements in data diversity and quality will increase the accuracy and generalizability of AI and ML algorithms. Diversification of data sets used in molecular analyses and the creation of larger data pools will strengthen the performance and adaptability of models. Especially in fields such as cheminformatics and bioinformatics, the creation of large open-access data sets will increase the wide applicability of AI.

In addition, advances in the modeling of biological systems will lead to significant innovations in areas such as drug design and personalized therapy. AI and ML algorithms will enable deeper analysis of the dynamic processes of biomolecules, and accelerating these processes will create significant opportunities in biotechnology.

Ethical aspects of the application of artificial intelligence and machine learning in molecular analyses are of great importance to ensure the responsible use of these technologies. Particularly in an era where molecular and biological data are increasingly digitised and shared, care should be taken to protect data confidentiality. Algorithmic bias during data preparation or model training can lead to systematic errors that may undermine the validity of research results. Furthermore, explainability, which refers to the understandability of the decision-making processes of AI models, is a critical element to increase the transparency of these technologies and build trust. Addressing these ethical challenges is essential for the sustainable development of AI and machine learning applications and their widespread acceptance in scientific research.

In conclusion, the role of AI and machine learning in molecular analysis will expand even further in the future, enabling new discoveries in chemistry and biochemistry. Areas such as quantum artificial intelligence, improving data quality, and more in-depth modeling of biological processes will be key factors that will determine the future development of these technologies.

Table 1. Performance Comparison of Machine Learning Algorithms

Ref.	Implemented Algorithm	Parameters Used	Metric Used	Success Rate
[4]	Molecular Docking	Ligand flexibility	RMSD	85% (binding accuracy)
[5]	Deep Learning	Deep neural networks	Accuracy	91% (Different data set tests)
[7]	Reinforcement Learning	Self-play algorithm	Win Rate	90% (in Checker game)
[8]	Decision Tree, Neural Networks	Different dataset	Accuracy	82% (Overall forecast rate)
[9]	Convolutional Neural Network (CNN)	Depth, dropout, learning rate	Top-1 Accuracy	84% (ImageNet dataset)
[10]	Support Vector Machine (SVM)	Kernel type (RBF), C value	Accuracy	97% (RBF Kernel)
[11]	Random Forest	Number of trees	F1-Score	90% (Complex data classification)
[12]	Ensemble Methods	Bagging and Boosting	Precision, Recall	88% (Trend analysis)
[15]	Raman Spectroscopy	Spectrum analysis	Sensitivity, Specificity	92% (Preliminary disease detection)
[16]	Particle Mesh Ewald	Ewald sums	Energy Deviation	85% (Energy calculations)
[17]	Kernel Ridge Regression (KRR)	Atomic properties	MAE	95.3% (MAE < 1 kcal/mol)
[18]	Deep Tensor Neural Networks (DTNN)	Atomic coordinates, bond energies	RMSE	98% (Quantum-chemical insights)
[19]	Chemoinformatics	Ligand basic parameters	R2	87% (QSAR estimate)
[20]	Extended-Connectivity Fingerprints	Molecular fingerprinting algorithms	Accuracy	89% (Chemical classification)
[21]	Virtual Screening	Molecular docking	Docking Score	82% (Chemical screening)
[22]	Support Vector Machine (SVM)	Subcellular localization	Accuracy	88% (Protein localization)
[23]	Support Vector Machine (SVM)	Kernel selection	Accuracy	87% (Classification accuracy)
[24]	Support Vector Machine (SVM)	Kernel selection	Accuracy	85% (Bioinformatics analysis)
[25]	Random Forest	Number of trees, maximum depth	MAE	93% (QSAR modeling)
[26]	Ensemble Learning	Boosting	Accuracy	91% (Prediction model)
[27]	Generative Models	SMILES-based learning	Validity	88% (De novo design)
[28]	Genetic Algorithm + DNN	Population size, number of layers	Accuracy	92% (Chemical classification)
[29]	U-Net	Segmentation parameters	Dice Coefficient	95% (Biomedical imaging)
[30]	Deep Learning (AlphaFold)	Protein sequence length	GDT_TS	92% (Protein structure prediction)
[31]	Reinforcement Learning	State and action space	Reward Function	90% (Learning optimization)
[32]	Deep Reinforcement Learning	Latent space size	Docking Score	86% (Molecular design)
[33]	Message Passing Neural Networks (MPNN)	Atomic and bond properties	MAE	96% (Quantum chemistry)
[34]	Quantum Machine Learning	Electronic structure	R2	89% (Molecular activity)
[35]	Neural Networks	QSAR	Accuracy	94% (Target identification)
[36]	Deep Learning	DDR1 inhibitor detection	Accuracy	94% (Target identification)
[37]	Generative Adversarial Networks (GANs)	Latent space size, learning rate	Validity	88% (Molecular design)
[38]	Neural Networks	High-dimensional energy surfaces	RMSE	95% (Quantum potential)
[39]	Machine Learning	Molecular simulation parameters	MAE	90% (Physical chemistry model)
[40]	Deep Neural Networks	Chemical synthesis planning	Success Rate	92% (Accurate synthesis prediction)
[41]	Random Forest	Chemical reaction outputs	Accuracy	89% (Organic reaction)
[42]	Quantum-based ML	Electronic features	R2	94% (Chemical compound space)
[43]	Random Forest	Reaction parameters	F1-Score	90% (Organic synthesis model)
[44]	Virtual Screening + Experimental	OLED design optimization	Yield	90% (Organic synthesis model)

References

[1] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach (4th ed.), 2021, USA, Pearson.

[2] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, 2016, USA, MIT Press.

[3] K.P. Murphy, Machine Learning: A Probabilistic Perspective, 2012, USA, MIT Press.

[4] D. Ramírez, Computational methods applied to rational drug design, Open Med Chem J, 10, 2016, 7–20.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature, 521(7553), 2015, 436–444.

[6] A.M. Turing, Computing machinery and intelligence, Mind, 59(236), 1950, 433–460.

[7] A.L. Samuel, Some studies in machine learning using the game of checkers, IBM J Res Dev, 3(3), 1959, 210–229.

[8] T.M. Mitchell, Machine Learning, 1997, USA, McGraw-Hill.

[9] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv Neural Inf Process Syst, 25, 2012, 1097–1105.

[10] C. Cortes, V. Vapnik, Support-vector networks, Mach Learn, 20(3), 1995, 273–297.

[11] L. Breiman, Random forests, Mach Learn, 45(1), 2001, 5–32.

[12] M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science, 349(6245), 2015, 255–260.

[13] D.A. Skoog, D.M. West, F.J. Holler, S.R. Crouch, Fundamentals of Analytical Chemistry (9th ed.), 2013, USA, Cengage Learning.

[14] H.A. Duran-Limon, A. Chavoya, M. Hernández-Ochoa, The role of machine learning in big data analytics: Current practices and challenges, Development Methodologies for Big Data Analytics Systems, Editors: M. Mora, F. Wang, J. Marx Gomez, H. Duran-Limon, 2024, USA, Springer, 15–28.

[15] Y. Oshima, T. Haruki, K. Koizumi, S. Yonezawa, A. Taketani, M. Kadowaki, S. Saito, Practices, potential, and perspectives for detecting predisease using Raman spectroscopy, Int J Mol Sci, 24(15), 2023, 12170.

[16] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, J Chem Phys, 98(12), 1993, 10089–10092.

[17] M. Rupp, A. Tkatchenko, K.R. Müller, O.A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Phys Rev Lett, 108(5), 2012, 058301.

[18] K. Schütt, F. Arbabzadah, S. Chmiela, et al., Quantum-chemical insights from deep tensor neural networks, Nat Commun, 8, 2017, 13890.

[19] F.K. Brown, Chemoinformatics: What is it and how does it impact drug discovery, Annu Rep Med Chem, 33, 1998, 375–384.

[20] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J Chem Inf Model, 50(5), 2010, 742–754.

[21] G. Schneider, Virtual screening: An endless staircase?, Nat Rev Drug Discov, 9(4), 2010, 273–276.

[22] S. Hua, Z. Sun, Support vector machine approach for protein subcellular localization prediction, Bioinformatics, 17(8), 2001, 721–728.

[23] B. Schölkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, 2002, USA, MIT Press.

[24] W.S. Noble, What is a support vector machine?, Nat Biotechnol, 24(12), 2006, 1565–1567.

[25] V. Svetnik, A. Liaw, C. Tong, T. Wang, Random forest: A classification and regression tool for compound classification and QSAR modeling, J Chem Inf Comput Sci, 43(6), 2003, 1947–1958.

[26] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.), 2009, USA, Springer.

[27] M. Cherti, B. Kégl, A.O. Kazakçi, De novo drug design using deep generative models: An empirical study, International Conference on Learning Representations, Toulon, France, 2017.

[28] M. Karakaplan, F.M. Avcu, Classification of some chemical drugs by genetic algorithm and deep neural network hybrid method, Concurr Comput Pract Exp, 33, 2021, e6242.

[29] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, 234–241.

[30] A.W. Senior, R. Evans, J. Jumper, et al., Improved protein structure prediction using potentials from deep learning, Nature, 577(7792), 2020, 706–710.

[31] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, 2018, USA, MIT Press.

[32] M. Popova, O. Isayev, A. Tropsha, Deep reinforcement learning for de novo drug design, Sci Adv, 4(7), 2018, eaap7885.

[33] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Neural message passing for quantum chemistry, Proc Int Conf Mach Learn, 34, 2017, 1263–1272.

[34] R. Xia, S. Kais, Quantum machine learning for electronic structure calculations, Nat Commun, 9, 2018, 4195.

[35] F. Ghasemi, A. Mehridehnavi, A. Pérez-Garrido, H. Pérez-Sánchez, Neural network and deep-learning algorithms used in QSAR studies: Merits and drawbacks, Drug Discov Today, 23(10), 2018, 1784–1790.

[36] A. Zhavoronkov, Y.A. Ivanenkov, A. Aliper, et al., Deep learning enables rapid identification of potent DDR1 kinase inhibitors, Nat Biotechnol, 37, 2019, 1038–1040.

[37] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science, 361(6400), 2018, 360–365.

[38] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys Rev Lett, 98(14), 2007, 146401.

[39] F. Noé, A. Tkatchenko, K.R. Müller, C. Clementi, Machine learning for molecular simulation, Annu Rev Phys Chem, 71, 2020, 361–390.

[40] M.H. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, Nature, 555(7698), 2018, 604–610

[41] C.W. Coley, R. Barzilay, T.S. Jaakkola, W.H. Green, K.F. Jensen, Prediction of organic reaction outcomes using machine learning, ACS Cent Sci, 3(5), 2017, 434–443.

[42] O.A. von Lilienfeld, K.R. Müller, A. Tkatchenko, Exploring chemical compound space with quantum-based machine learning, Nat Rev Chem, 4, 2020, 347–358.

[43] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. Hirzel, et al., Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, Nat Mater, 15, 2016, 1120–1127.

[44] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. Hirzel, et al., Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, Nat Mater, 15, 2016, 1120–1127.