RTET- 2K26", National Conference (June-2026)
Organized by Department of General Sciences and Engineering, AITRC Vita

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

A Comprehensive Review of MolBench: Benchmarking AI Models for Molecular Property Prediction

[1]Kale Anjali Rajesh, [2]Kale Shanur Dnyaneshwar, [3]Kambale Swapnil Sanjay, [4]Kamble Aryan Ravindra, [5]Kamble Riddhi Shrikrishna, [6]Kamble Yash Ajay, [7]Karande Priya Abaji, [8]Katkar Pratiksha Dnyandeo

[1-8]Students, General Scienceand and Engineering, AITRC, Vita

Abstract- The rapid advancement of artificial intelligence (AI) and deep learning has transformed molecular property prediction in areas such as drug discovery, material science, computational chemistry, and biotechnology. Accurate prediction of molecular properties reduces the dependence on expensive and time-consuming laboratory experiments while accelerating scientific innovation. The reviewed work, MolBench: A Benchmark of AI Models for Molecular Property Prediction, introduces a benchmark framework designed to evaluate AI models across multiple molecular prediction tasks using multidimensional metrics and standardized datasets. This review paper provides a detailed and original analysis of the MolBench framework, including its objectives, datasets, methodologies, evaluation metrics, experimental findings, strengths, limitations, and future implications. The paper critically examines the benchmark's contribution to molecular machine learning by comparing traditional machine learning methods, graph neural networks, and self-supervised pre-trained models. Furthermore, it discusses the significance of benchmark-driven research in ensuring fairness, reproducibility, and scientific progress in AI-based molecular property prediction.

Keywords: Molecular Property Prediction, Artificial Intelligence (AI), Deep Learning, Molecular Machine Learning, MolBench, Graph Neural Networks (GNNs), Self-Supervised Learning, Drug Discovery, Benchmark Framework, Computational Chemistry, Model Evaluation, Reproducibility.

I.INTRODUCTION

Molecular property prediction is one of the most important research areas in modern computational science. Predicting the physical, chemical, biological, and pharmacological properties of molecules plays a crucial role in pharmaceutical development, material engineering, toxicology analysis, and biomedical research. Traditional experimental approaches, although highly accurate, require substantial time, financial resources, and human effort. The emergence of AI and deep learning techniques has provided an efficient alternative capable of processing large-scale molecular data and generating predictions with remarkable speed.

In recent years, machine learning and graph-based neural network models have demonstrated significant success in learning molecular representations and predicting molecular behavior. However, the rapid growth of models has also introduced a major challenge: the lack of a unified and reliable benchmark for evaluating different AI approaches under standardized conditions.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

The MolBench framework addresses this issue by introducing a comprehensive benchmarking platform that evaluates molecular property prediction models using multiple datasets, evaluation metrics, and performance perspectives. Instead of focusing solely on predictive accuracy, MolBench evaluates model stability, task coverage, and generalization ability. This multidimensional approach enables a fair comparison among traditional machine learning methods, graph neural networks (GNNs), and self-supervised pre-trained models.

This review paper critically analyzes the MolBench benchmark and highlights its significance in advancing molecular AI research. The discussion includes the benchmark design, selected datasets, model categories, evaluation methodology, experimental results, and future research opportunities.

II. BACKGROUND AND NEED FOR MOLECULAR BENCHMARKS

Importance of Molecular Property Prediction

Molecular property prediction involves estimating the characteristics of molecules based on their structural and chemical information. These properties may include:

- ▪ Solubility
- ▪ Toxicity
- ▪ Bioactivity
- ▪ Lipophilicity
- ▪ Quantum mechanical properties
- ▪ Drug-target interactions
- ▪ Blood-brain barrier penetration

Accurate prediction of these properties supports numerous industrial and scientific applications:

Drug Discovery

AI-based molecular prediction accelerates the identification of promising drug candidates by estimating biological activity and toxicity before clinical testing.

Material Science

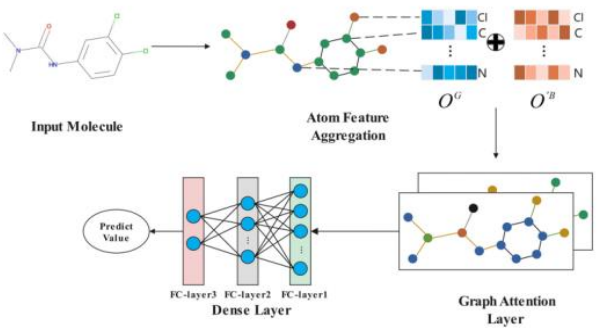
Predictive models assist in designing new materials with desired thermal, electrical, or mechanical properties.

Environmental and Toxicological Analysis

Molecular prediction models help assess chemical safety and environmental impact.

Personalized Medicine

AI systems may support customized treatment strategies through molecular-level analysis.



International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Challenges in Molecular Machine Learning

Although AI models have shown impressive results, molecular prediction tasks present several challenges:

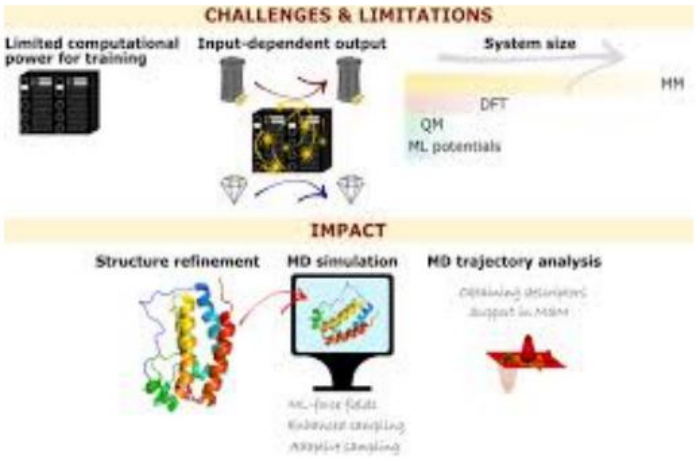
1. Complex molecular structures
2. Limited labeled datasets
3. High-dimensional feature spaces
4. Multi-task prediction requirements
5. Generalization across domains
6. Computational cost
7. Lack of standardized evaluation

Researchers often evaluate models using different datasets, preprocessing methods, random seeds, and hyperparameter configurations. Consequently, direct comparison among studies becomes unreliable.

A benchmark framework is therefore essential to:

- ▪ Ensure reproducibility
- ▪ Enable fair model comparison
- ▪ Standardize evaluation procedures
- ▪ Measure generalization ability
- ▪ Encourage innovation in model design

MolBench was developed to address these requirements.



III. OVERVIEW OF MOLBENCH

MolBench is a benchmark framework specifically developed for evaluating AI models used in molecular property prediction tasks. The benchmark integrates:

- ▪ Widely recognized molecular datasets
- ▪ Multiple model categories
- ▪ Standardized evaluation metrics
- ▪ Stability analysis
- ▪ Task coverage analysis
- ▪ Extensible API support

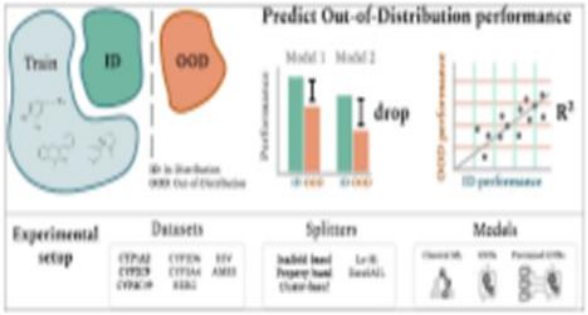
The benchmark evaluates three major categories of models:

1. Molecular fingerprint-based machine learning models

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

2. Graph neural network-based models
3. Self-supervised pre-trained molecular models

Unlike earlier benchmarks that focused mainly on supervised learning performance, MolBench introduces a multidimensional evaluation strategy that considers both predictive accuracy and robustness.



IV. RELATED BENCHMARK FRAMEWORKS

Before MolBench, several benchmark frameworks were proposed for molecular and material prediction tasks.

MoleculeNet

MoleculeNet is one of the earliest and most influential benchmark frameworks for molecular machine learning. It introduced:

- ▪ Standardized molecular datasets
- ▪ Common evaluation metrics
- ▪ Baseline implementations

However, MoleculeNet mainly focused on supervised learning models and did not include many modern self-supervised pre-trained architectures.

MatBench

MatBench focused primarily on material science applications. It provided benchmark tasks for predicting inorganic material properties using supervised learning techniques. Although valuable for material prediction, MatBench does not specifically address molecular representation learning or pre-trained molecular models.

MUBen

MUBen emphasized uncertainty quantification in pre-trained molecular models. It explored the confidence and reliability of molecular predictions. However, MUBen mainly concentrated on uncertainty estimation rather than broader multidimensional evaluation.

Distinctive Contribution of MolBench

MolBench differentiates itself by:

- ▪ Combining supervised and self-supervised models
- ▪ Using multidimensional evaluation metrics
- ▪ Including graph neural networks and pre-trained models
- ▪ Evaluating task coverage and stability
- ▪ Supporting extensibility through APIs

This comprehensive design makes MolBench highly relevant for modern molecular AI research.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

V. DATASETS USED IN MOLBENCH

The benchmark utilizes datasets from MoleculeNet along with large-scale pre-training datasets.

Pre-Training Datasets
ZINC15
ZINC15 is a large virtual screening database containing millions of purchasable chemical compounds. It provides extensive unlabeled molecular data suitable for self-supervised pre-training.

ChEMBL
ChEMBL is a curated bioactivity database containing molecular structures and biological activity information. It is widely used in drug discovery research.

These datasets enable pre-trained models to learn meaningful molecular representations before downstream fine-tuning.

MoleculeNet Benchmark Datasets
MolBench selects datasets spanning multiple scientific domains.
Quantum Mechanics Datasets
QM7
Contains molecular geometry and electronic property information.

QM8
Includes advanced quantum chemical calculations for molecules.

QM9
Provides extensive molecular properties related to geometry, energy, thermodynamics, and electronics.

Physical Chemistry Datasets
ESOL
Used for predicting aqueous solubility.

FreeSolv
Contains hydration free energy values.

Lipophilicity (Lipo)
Measures molecular lipophilic affinity.

Biophysics Datasets
HIV
Evaluates a molecule's ability to inhibit HIV replication.

BACE
Focuses on β-secretase inhibitors associated with Alzheimer's disease research.

Physiology and Toxicology Datasets
BBBP
Measures blood-brain barrier penetration.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Tox21
Predicts compound toxicity.

ToxCast
Provides high-throughput toxicology screening data.

SIDER
Contains drug side effect information.

ClinTox
Compares clinically approved drugs with toxic compounds.

Category	Dataset	Data Type	Task Type	# Tasks	# Compounds	Rec - Split[a]	Rec - Metric[b]
Quantum Mechanics	QM7	SMILES, 3D coordinates	Regression	1	7160	Stratified	MAE
	QM7b	3D coordinates	Regression	14	7210	Random	MAE
	QM8	SMILES, 3D coordinates	Regression	12	21786	Random	MAE
	QM9	SMILES, 3D coordinates	Regression	12	133885	Random	MAE
Physical Chemistry	ESOL	SMILES	Regression	1	1128	Random	RMSE
	FreeSolv	SMILES	Regression	1	642	Random	RMSE
	Lipophilicity	SMILES	Regression	1	4200	Random	RMSE
Biophysics	PCBA	SMILES	Classification	128	437929	Random	PRC-AUC
	MUV	SMILES	Classification	17	93087	Random	PRC-AUC
	HIV	SMILES	Classification	1	41127	Scaffold	ROC-AUC
	PDBbind	SMILES, 3D coordinates	Regression	1	11908	Time	RMSE
	BACE	SMILES	Classification	1	1513	Scaffold	ROC-AUC
Physiology	BBBP	SMILES	Classification	1	2039	Scaffold	ROC-AUC
	Tox21	SMILES	Classification	12	7831	Random	ROC-AUC
	ToxCast	SMILES	Classification	617	8575	Random	ROC-AUC
	SIDER	SMILES	Classification	27	1427	Random	ROC-AUC
	ClinTox	SMILES	Classification	2	1478	Random	ROC-AUC

VI. MODEL CATEGORIES IN MOLBENCH

MolBench evaluates three major categories of AI models.

ECFP-Based Machine Learning Models
Extended Connectivity Fingerprints (ECFPs) convert molecular structures into fixed-length binary vectors representing atom neighborhoods and topological information.

Support Vector Machine (SVM)
SVM is a supervised learning algorithm that identifies optimal decision boundaries for classification and regression tasks.

Random Forest (RF)
Random Forest is an ensemble learning technique based on multiple decision trees.

XGBoost
XGBoost is a gradient boosting framework known for high predictive accuracy and computational efficiency.

Multi-Layer Perceptron (MLP)
MLP is a feed-forward neural network consisting of input, hidden, and output layers.

International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

Graph Neural Network Models
Molecules naturally form graph structures where atoms represent nodes and chemical bonds represent edges. Graph neural networks directly learn from these graph structures.

Graph Convolutional Network (GCN)
GCN extends convolution operations to graph data, enabling local structural learning.

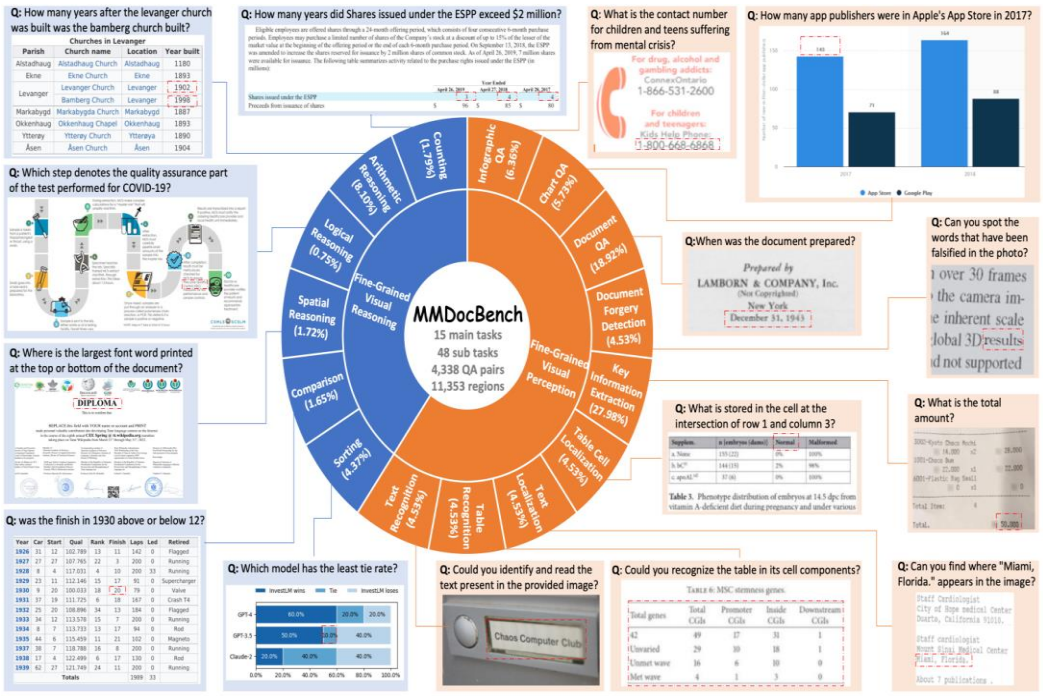
Graph Attention Network (GAT)
GAT introduces attention mechanisms that assign different importance weights to neighboring nodes.

Graph Isomorphism Network (GIN)
GIN improves graph discrimination power and effectively distinguishes structural patterns.

Acknowledgement
The author sincerely acknowledges the valuable contributions of researchers and scientists in the fields of Artificial Intelligence, Molecular Machine Learning, Computational Chemistry, and Bioinformatics whose pioneering work has greatly influenced the development of molecular property prediction systems. Special appreciation is extended to the authors of MolBench: A Benchmark of AI Models for Molecular Property Prediction for providing a comprehensive benchmarking framework that supports fair and multidimensional evaluation of AI models in molecular science.

The author is also grateful to the developers of benchmark datasets such as MoleculeNet, ZINC15, and ChEMBL, which have significantly contributed to advancements in molecular representation learning and predictive modeling. Their efforts in maintaining publicly accessible datasets and tools have played a crucial role in accelerating research in drug discovery, biotechnology, and material science.

Finally, sincere thanks are extended to the academic community and researchers working on graph neural networks, self-supervised learning, and deep learning techniques for their continuous innovation and support in advancing scientific machine learning research.



International Journal of Science, Engineering and Technology
ISSN: 2348-4098, P-ISSN: 2395-4752

REFERENCES

1. Jiang, X., Tan, L., Cen, J., & Zou, Q. MolBench: A Benchmark of AI Models for Molecular Property Prediction. Springer Nature Singapore, 2024.
2. Wu, Z., et al. MoleculeNet: A Benchmark for Molecular Machine Learning. Chemical Science, 2018.
3. Kipf, T. N., & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv, 2016.
4. Velickovic, P., et al. Graph Attention Networks. arXiv, 2017.
5. Xu, K., et al. How Powerful are Graph Neural Networks? arXiv, 2018.
6. Wang, Y., et al. MolCLR: Molecular Contrastive Learning of Representations via Graph Neural Networks. Nature Machine Intelligence, 2022.
7. Rong, Y., et al. Self-Supervised Graph Transformer on Large-Scale Molecular Data. NeurIPS, 2020.
8. Zhou, G., et al. Uni-Mol: A Universal 3D Molecular Representation Learning Framework, 2023.
9. Breiman, L. Random Forests. Machine Learning, 2001.