

# Leveraging Graph Neural Networks for Enhanced Prediction of Molecular Solubility via Transfer Learning

Dat P. Nguyen<sup>\*</sup>, Phuc T. Le

Ho Chi Minh City University of Technology and Education, Vietnam

<sup>\*</sup>Corresponding author. Email: [datnp@hcmute.edu.vn](mailto:datnp@hcmute.edu.vn)

## ARTICLE INFO

Received: 22/04/2024  
Revised: 17/05/2024  
Accepted: 20/05/2024  
Published: 28/06/2024

## KEYWORDS

Molecular property prediction;  
Machine learning;  
Deep learning;  
Graph neural network;  
Transfer learning.

## ABSTRACT

In this study, we explore the potential of graph neural networks (GNNs), in combination with transfer learning, for the prediction of molecular solubility, a crucial property in drug discovery and materials science. Our approach begins with the development of a GNN-based model to predict the dipole moment of molecules. The extracted dipole moment, alongside a selected set of molecular descriptors, feeds into a subsequent predictive model for water solubility. This two-step process leverages the inherent correlations between molecular structure and its physical properties, thus enhancing the accuracy and generalizability. Our data showed that GNN models with attention mechanism and those utilize bond properties outperformed other models. Especially, 3D GNN models such as ViSNet exhibited outstanding performance, with an  $R^2$  value of 0.9980. For the prediction of water solubility, the inclusion of dipole moments greatly enhanced the predictive power of various machine learning models. Our methodology demonstrates the effectiveness of GNNs in capturing complex molecular features and the power of transfer learning in bridging related predictive tasks, offering a novel approach for computational predictions in chemistry.

Doi: <https://doi.org/10.54644/jte.2024.1571>

Copyright © JTE. This is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purpose, provided the original work is properly cited.

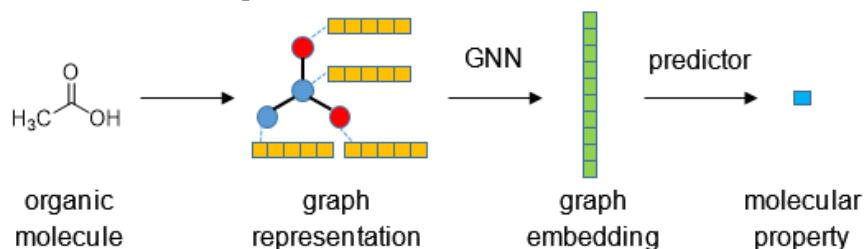
## 1. Introduction

Predicting molecular characteristics is crucial in computational chemistry, particularly in drug discovery and materials research [1]. The effectiveness and viability of molecules in various applications largely depend on properties such as solubility, lipophilicity, and reactivity [2]. Traditionally, experimental methods have been the primary means for determining these properties. However, the development of computational tools has substantially changed this process by providing faster and more affordable alternatives. Precise estimation of these properties is vital in the early stages of drug development, as it impacts the pharmacokinetics, pharmacodynamics, and overall feasibility of medicinal substances [3]. Similarly, in the field of materials research, anticipating molecular properties is essential for creating new materials with desired characteristics [4].

When computational methods were first developed, they primarily relied on quantum mechanical techniques like density functional theory (DFT), which offered detailed insights into the electronic structure of molecules. Although DFT and related methods have proved helpful in understanding the characteristics of compounds, they are frequently computationally demanding and may not always be practical for larger molecules or high-throughput screening [5]. This limitation led to the integration of machine learning (ML) algorithms in computational chemistry, striking a balance between accuracy and computational efficiency. ML models, particularly those based on molecular descriptors, have successfully predicted numerous molecular properties. However, these models often require extensive feature engineering and may not fully capture the complexity of molecular interactions [6].

Recent advancements have seen the emergence of graph neural networks (GNNs) as a powerful tool in the field of molecular property prediction (Figure 1). GNNs capture the essence of molecular structures more naturally by representing molecules as graphs [6]. This representation allows for a more

intuitive and detailed understanding of molecular interactions and properties [7]. GNNs have proven their usefulness in drug discovery and pharmaceutical analysis by accurately predicting the activity or effectiveness of medicinal compounds [8]. In addition, GNNs have shown remarkable performance in predicting the outcomes of chemical reactions, such as the Buchwald-Hartwig and Suzuki-Miyaura coupling reactions [9]. Beyond these specific applications, GNNs hold potential in areas such as organic synthesis [10], molecular docking [11], and material science [12]. Overall, the adoption of GNNs in these areas signifies a major shift towards more efficient, accurate, and predictive models in chemical research and pharmaceutical development.



**Figure 1.** GNN Model for Molecular Property Prediction.

The main goal of our research is to predict the solubility of molecules by first constructing a molecular dipole moment prediction model based on a GNN. Leveraging transfer learning, we then apply this trained model to predict the water solubility of molecules. By utilizing the extracted dipole moment features, along with other molecular descriptors, we aim to construct a more robust and accurate predictive model for solubility. This two-step method represents a significant departure from conventional approaches; it not only offers improved accuracy in molecular property prediction but also serves as an exemplar of transfer learning applied to computational chemistry, potentially setting a new precedent for future research in the field.

## 2. Computational Methods

### 2.1. Dataset Curation

The QM9 dataset, which was published by Ramakrishnan *et al.* in 2014 [13], offers a comprehensive collection of quantum mechanical properties for a wide range of small organic compounds. This dataset encompasses approximately 134,000 stable and synthetically accessible organic molecules along with their properties, computed using high-level quantum chemistry methods. The dataset was downloaded in comma-separated values (CSV) format from MoleculeNet [14]. For the construction of molecular graphs for GNN models, the ‘SMILES’ column, which contains the SMILES (Simplified Molecular Input Line-Entry System) representations of the compounds, were used. The ‘mu’ column, which contains the dipole moment of molecules (in Debye), was utilized as the target for the GNN models.

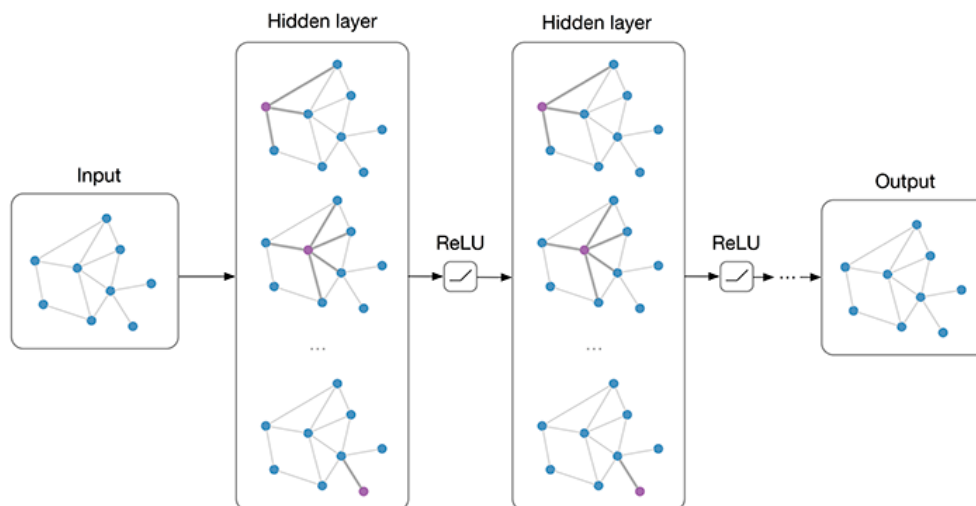
The water solubility dataset (also known as the ESOL dataset), originally published by Delaney *et al.* in 2004 [15], was obtained as a CSV file from MoleculeNet. This dataset is employed to predict compound solubility based on molecular structure and chemical features. The ‘SMILES’ column was utilized for feature extraction, while the ‘solubility’ column served as the target for our ML models.

### 2.2. GNN Models for Prediction of Dipole Moment

In constructing our GNN models, the first step involved transforming molecular structures into graph representations, utilizing RDKit and DeepChem, two prominent open-source cheminformatics and deep learning libraries. RDKit facilitates the conversion of SMILES notation into molecular structures, while DeepChem’s graph featurizer is used to extract graphical data from these structures. Subsequently, the dataset was divided into three subsets: training, validation, and testing, in a 60:20:20 ratio.

The molecular graphs generated from the training set were input into the GNN models (Figure 2). These models were developed using PyTorch-Geometric, a leading open-source library for geographical deep learning. The customized GNN models consist of 3 convolutional layers, each followed by a linear layer with a rectified linear unit (ReLU) activation function. The outputs of these GNN layers were

further passed through a multilayer perceptron (MLP) predictor that consists of 3 linear layers with ReLU activation function. During training, the models were periodically validated to monitor performance and make necessary parameter adjustments. Various types of convolution layers were employed in the GNN architecture, aiming to investigate different facets of graph-based learning and feature extraction. The GNN models were trained for 100 epochs (except 3D models, which were trained for 500 epochs) using a GeForce RTX 4090 with a specific set of training parameters for each model.



**Figure 2.** Structure of GNN Model.

After training, the test dataset was employed to assess the model's predictive performance. Regression metrics, including mean absolute error (MAE), root mean square error (RMSE) and the coefficient of determination  $R^2$ , were used for this evaluation.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Here,  $\hat{y}_i$  represents the predicted values,  $y_i$  the observed values,  $n$  is the number of observations, and  $\bar{y}$  is the mean of observed values.

### 2.3. ML Models for Prediction of Water Solubility

In the development of our water solubility prediction model, the process began with extracting a comprehensive list of molecular features using RDKit from the SMILES representations of the molecules. These features include molecular weight, number of rings, ratio of  $sp^3$ -hybridized carbon, number of hydrogen bond donors and acceptors. The extracted features were then combined with the dipole moments produced by the best-performing GNN model, creating a rich dataset for solubility prediction. Subsequently, the dataset was divided into two primary parts: training and testing, following an 80:20 ratio. In cases where model validation during training was required, a three-part split comprising training, validation, and testing was employed, following a 60:20:20 ratio.

The training set was used to train various machine learning (ML) models selected from well-established libraries such as Scikit-learn and PyTorch, each offering a range of algorithms suitable for regression tasks. Prior to training, the features were scaled using normalization to ensure that the transformed values were comparable and to prevent the dominance of features with larger scales. The training of ML models was conducted on a local computer with a selected set of parameters. Deep learning models were trained for 100 epochs using a GeForce RTX 4090.

Upon completion of the training phase, the test dataset was used for model evaluation. Similar to the GNN models, the effectiveness of each model was assessed using a range of metrics, including MAE, RMSE and  $R^2$ .

### 3. Results and Discussion

#### 3.1. Evaluation of GNN Models for Prediction of Dipole Moment

In assessing the performance of our GNN models, a pivotal factor considered was the structure of the convolutional layers. Models with and without an attention mechanism, as well as those incorporate or exclude bond attributes, were examined (Table 1). The attention mechanism in GNNs is designed to selectively weigh the significance of nodes during the feature aggregation process.

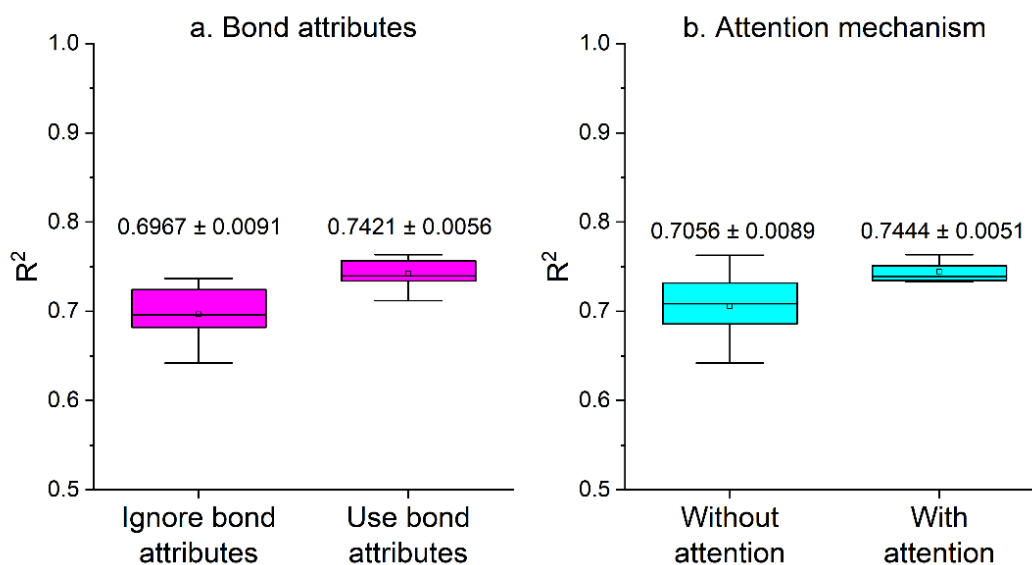
**Table 1.** Evaluation of Convolutional Layers for Prediction of Dipole Moment.

Entry	Convolutional layer	Attention	Bond attributes	MAE	RMSE	R <sup>2</sup>	Ref.
1	GCNConv	No	No	0.6332	0.9156	0.6478	[16]
2	SAGEConv	No	No	0.5736	0.8502	0.6963	[17]
3	SGConv	No	No	0.6350	0.9232	0.6419	[18]
4	ClusterGCNConv	No	No	0.5764	0.8372	0.7055	[19]
5	GraphConv	No	No	0.5904	0.8648	0.6857	[20]
6	LEConv	No	No	0.5903	0.8571	0.6913	[21]
7	EGConv	No	No	0.5214	0.7913	0.7369	[22]
8	MFConv	No	No	0.5373	0.7987	0.7319	[23]
9	TAGConv	No	No	0.6005	0.8699	0.6820	[24]
10	ARMAConv	No	No	0.5179	0.8102	0.7242	[25]
11	FilMConv	No	No	0.5347	0.8159	0.7203	[26]
12	PDNConv	No	Yes	0.5536	0.8281	0.7119	[27]
13	GENConv	No	Yes	0.5056	0.7865	0.7400	[28]
14	ResGatedGraphConv	No	Yes	0.4847	0.7516	0.7626	[29]
15	GATConv	Yes	Yes	0.5315	0.7963	0.7336	[30]
16	GATv2Conv	Yes	Yes	0.5062	0.7695	0.7512	[31]
17	TransformerConv	Yes	Yes	0.5169	0.7949	0.7345	[32]
18	SuperGATConv	Yes	Yes	0.5194	0.7876	0.7393	[33]
19	GeneralConv	Yes	Yes	0.4803	0.7503	0.7635	[34]

Convolutional layers such as GCNConv, SAGEConv, and SGConv, which do not utilize attention mechanisms or bond attributes, exhibited moderate degrees of performance (Entries 1–11). Specifically, SGConv showed the highest RMSE value of 0.9232, indicating less accurate predictions (Entry 3). Conversely, EGConv demonstrated the best performance among this group, with an R<sup>2</sup> value of 0.7369 (Entry 7).

When comparing the impact of incorporating bond attributes into the model, we observed a notable improvement in predictive accuracy (Entries 12–14). Models utilizing bond attributes yielded an average R<sup>2</sup> of 0.7421, in contrast to their counterparts that ignored these attributes, which exhibited an average R<sup>2</sup> of 0.6967 (Figure 3a). This improvement underscores the significance of bond information in

capturing the nuances of molecular interactions essential for the accurate prediction of properties such as solubility. Convolutional layers like PDNConv (Entry 12) and GENConv (Entry 13) demonstrated better predictive accuracy. Notably, ResGatedGraphConv emerged as the most accurate, with the lowest MAE of 0.4847 and the highest  $R^2$  of 0.7626 (Entry 14), indicating the benefits of including bond attributes.



**Figure 3.**  $R^2$  Values of GNN Models for Prediction of Dipole Moment.

Further results indicated that models with attention mechanisms (Entries 15–19) surpassed those without, evidenced by a lower MAE of 0.51084 compared to 0.56106 and an average  $R^2$  of 0.7444 versus 0.7056 (Figure 3b). The enhanced performance suggests that attention mechanisms enable models to focus on more relevant features for predicting the dipole moment, crucial for accurate solubility predictions. This is exemplified by the improved outcomes of GATConv and GATv2Conv, both employing attention and bond attributes, with  $R^2$  values of 0.7336 and 0.7512, respectively (Entries 15–16). Remarkably, GeneralConv, integrating both features, achieved the best overall result with an  $R^2$  of 0.7635 (Entry 19), highlighting the combined benefits of attention mechanisms and bond attributes.

In summary, GNN models enhanced with attention mechanisms and bond attributes consistently outshone those without. The data supports the assertion that these features are substantial contributors to model accuracy, signifying their crucial role in designing GNNs for molecular property prediction. Such advancements in model precision indicate a promising future for computational chemistry, offering scientists more reliable and sophisticated tools for their research.

We next evaluated the outputs of 3D GNN models, which consider the spatial arrangement of atoms and thus offer a more detailed depiction of molecular structures. The results demonstrated that 3D GNNs substantially outperform other GNN models in predicting the dipole moment, as shown in Figure 4. SchNet [35], with an RMSE of 0.0052 and an  $R^2$  of 0.9899, effectively captures the electronic environments of molecules through 3D convolution. DimeNet++ [36] shows further enhancement, with an RMSE of 0.0041 and an  $R^2$  of 0.9937. The most impressive performance is from ViSNet [37], which achieves an exceptional RMSE of 0.0023 and an  $R^2$  of 0.9980, underscoring the benefits of leveraging 3D spatial data.

Compared to standard GNN models that overlook spatial information, the advancements with 3D GNNs are noteworthy. These models transcend the limitations of simple structural and bonding details by effectively integrating the molecular geometry and orientations, which are critical for accurately predicting inherently three-dimensional properties such as the dipole moment. These findings suggest that for challenges demanding a nuanced comprehension of molecular geometry, 3D GNNs are not only more suitable but also significantly more precise in their predictions.

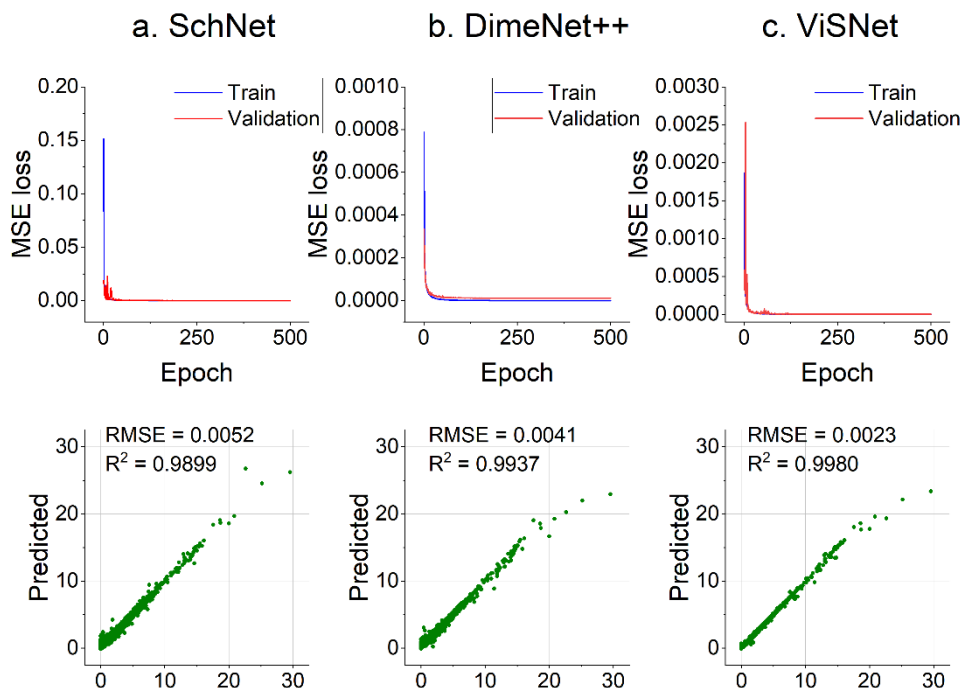


Figure 4. Evaluation of 3D GNN Models for Prediction of Dipole Moment.

### 3.2. Evaluation of ML Models for Prediction of Water Solubility

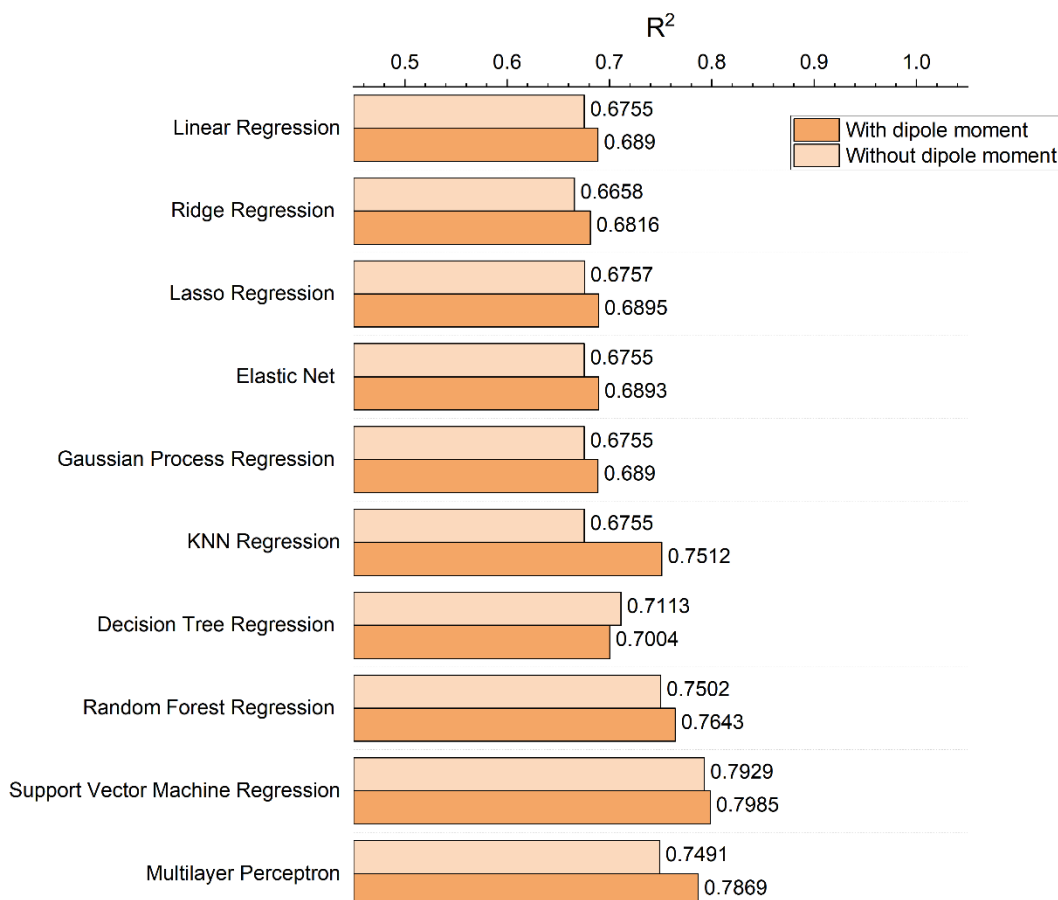


Figure 5. Evaluation of ML Models for Prediction of Water Solubility.

The evaluation of various machine learning (ML) models for water solubility prediction offers intriguing insights into the influence of molecular dipole moments on these models' predictive capabilities. Incorporating dipole moments predicted by ViSNet into the feature set generally led to improvements in the  $R^2$  scores for most models (Figure 5). These enhancements suggest that the dipole moment is a relevant feature for solubility prediction and that its accurate prediction by 3D GNNs like ViSNet can be effectively harnessed in other ML models.

For the linear models (linear regression, ridge regression, lasso regression, and elastic net), there was a modest increase in  $R^2$  values upon integrating the dipole moment feature (from 0.6755 to approximately 0.689 for linear regression, with similar marginal improvements observed for the others). This suggests that while there is an improvement, the linear relationship between the dipole moment and solubility is not strongly pronounced within these models. Among the non-linear models, KNN regression showed a marked enhancement with the inclusion of the dipole moment, with its  $R^2$  value increasing from 0.6755 to 0.7512. This indicates that the dipole moment imparts significant structural insights relevant to solubility prediction. Similarly, the  $R^2$  for random forest regression improved from 0.7502 to 0.7643, suggesting that this ensemble method effectively utilizes the additional feature.

In contrast, decision tree regression saw a minor decline in its  $R^2$  from 0.7113 to 0.7004, potentially indicating a propensity for overfitting when the dipole moment is added. Gaussian process regression did not exhibit a notable change with the new feature. However, the multilayer perceptron's performance increased from 0.7491 to 0.7869, underlining the model's adeptness at identifying complex patterns with the inclusion of features such as the dipole moment.

#### 4. Conclusions

In conclusion, our comprehensive investigation into the predictive modeling of molecular properties has demonstrated the substantial impact of employing advanced machine learning techniques, such as GNNs and transfer learning. Utilizing ViSNet, the best-performing GNN model, we achieved remarkable accuracy in predicting dipole moments, with an  $R^2$  value reaching up to 0.9980. This unprecedented precision in capturing the quantum mechanical properties of molecules paves the way for unbiased and accurate predictions in computational chemistry.

The application of transfer learning, utilizing the dipole moments predicted by ViSNet, further augmented our predictive models for water solubility. Although the integration of this feature yielded mixed results across various ML architectures, it was evident that non-linear models such as KNN improved significantly, with  $R^2$  values increasing from 0.6755 to 0.7512. This enhancement in model performance underscores the potential of combining complex molecular descriptors with ML algorithms to predict solubility, a critical parameter in drug development and materials science. Our findings highlight the potential for synergy between advanced computational methods and traditional chemical descriptors in predicting molecular properties. The integration of 3D molecular geometry with machine learning offers a faster and more cost-effective alternative to traditional experimental methods.

#### Acknowledgments

This work belongs to the project in 2024 funded by the Ho Chi Minh City University of Technology and Education, Vietnam.

#### Conflict of Interest

The authors declare no conflict of interest.

#### REFERENCES

- [1] A. Saunders and P. B. Harrington, "Advances in activity/property prediction from chemical structures," *Critical Reviews in Analytical Chemistry*, vol. 54, no. 1, pp. 135–147, 2024, pMID: 35482792.
- [2] J. Shen and C. A. Nicolaou, "Molecular property prediction: Recent trends in the era of artificial intelligence," *Drug Discovery Today: Technologies*, vol. 32-33, pp. 29–36, 2019, artificial Intelligence.
- [3] K. Ham, J. Yoon, and L. Sael, "Towards accurate and certain molecular properties prediction," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 1621–1624.
- [4] J. A. Keith *et al.*, "Combining machine learning and computational chemistry for predictive insights into chemical systems," *Chemical Reviews*, vol. 121, no. 16, pp. 9816–9872, 2021, pMID: 34232033.
- [5] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," *Journal of Computational Chemistry*, vol. 38,

- no. 16, pp. 1291–1307, 2017.
- [6] P. Reiser *et al.*, “Graph neural networks for materials science and chemistry,” *Communications Materials*, vol. 3, no. 1, p. 93, 2022.
  - [7] O. Wieder *et al.*, “A compact review of molecular property prediction with graph neural networks,” *Drug Discovery Today: Technologies*, vol. 37, pp. 1–12, 2020.
  - [8] J. Xiong, Z. Xiong, K. Chen, H. Jiang, and M. Zheng, “Graph neural networks for automated de novo drug design,” *Drug Discovery Today*, vol. 26, no. 6, pp. 1382–1393, 2021.
  - [9] Y. Kwon, D. Lee, Y. S. Choi, and S. Kang, “Uncertainty-aware prediction of chemical reaction yields with graph neural networks,” *Journal of Cheminformatics*, vol. 14, pp. 1–10, 2022.
  - [10] L. Ferretti, A. Cini, G. Zacharopoulos, C. Alippi, and L. Pozzi, “Graph neural networks for high-level synthesis design space exploration,” *ACM Trans. Des. Autom. Electron. Syst.*, vol. 28, no. 2, 2022.
  - [11] H. Jiang *et al.*, “Predicting protein–ligand docking structure with graph neural network,” *Journal of Chemical Information and Modeling*, vol. 62, no. 12, pp. 2923–2932, 2022, pMID: 35699430.
  - [12] W. Gong and Q. Yan, “Graph-based deep learning frameworks for molecules and solid-state materials,” *Computational Materials Science*, vol. 195, p. 110332, 2021.
  - [13] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. V. Lilienfeld, “Quantum chemistry structures and properties of 134 kilo molecules,” *Scientific data*, vol. 1, no. 1, pp. 1–7, 2014.
  - [14] Z. Wu *et al.*, “MoleculeNet: a benchmark for molecular machine learning,” *Chem. Sci.*, vol. 9, pp. 513–530, 2018.
  - [15] J. S. Delaney, “ESOL: Estimating aqueous solubility directly from molecular structure,” *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 3, pp. 1000–1005, 2004, pMID: 15154768.
  - [16] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks” 2017. arXiv:1609.02907.
  - [17] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs” 2018. arXiv:1706.02216.
  - [18] F. Wu, T. Zhang, A. H. de Souza Jr., C. Fifty, T. Yu, and K. Q. Weinberger, “Simplifying graph convolutional networks,” 2019. arXiv:1902.07153.
  - [19] W. L. Chiang *et al.*, “An efficient algorithm for training deep and large graph convolutional networks,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining, KDD 19, ACM*, 2019, doi:10.1145/3292500.3330925.
  - [20] C. Morris *et al.*, “Weisfeiler and leman go neural: Higher-order graph neural networks” 2021. arXiv:1810.02244.
  - [21] E. Ranjan, S. Sanyal, and P. P. Talukdar, “ASAP: Adaptive structure aware pooling for learning hierarchical graph representations,” 2020. arXiv:1911.07979.
  - [22] S. A. Tailor, F. L. Opolka, P. Lio, and N. D. Lane, “Do we need anisotropic graph neural networks?” 2022. arXiv:2104.01481.
  - [23] D. Duvenaud *et al.*, “Convolutional networks on graphs for learning molecular fingerprints,” 2015. arXiv:1509.09292.
  - [24] J. Du, S. Zhang, G. Wu, J. M. F. Moura, and S. Kar, “Topology adaptive graph convolutional networks,” 2018. arXiv:1710.10370.
  - [25] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, “Graph neural networks with convolutional ARMA filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, doi:10.1109/tpami.2021.3054830.
  - [26] M. Brockschmidt, “GNN-FiLM: Graph neural networks with feature-wise linear modulation,” 2020, arXiv:1906.12192.
  - [27] B. Rozemberczki, P. Englert, A. Kapoor, M. Blais, and B. Perozzi, “Pathfinder discovery networks for neural message passing,” 2021. arXiv:2010.12878.
  - [28] G. Li, C. Xiong, A. Thabet, and B. Ghanem, “DeeperGCN: All you need to train deeper GCNs,” 2020. arXiv:2006.07739.
  - [29] X. Bresson and T. Laurent, “Residual gated graph convnets,” 2018. arXiv:1711.07553.
  - [30] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” 2018. arXiv:1710.10903.
  - [31] S. Brody, U. Alon, and E. Yahav, “How attentive are graph attention networks?” 2022. arXiv:2105.14491.
  - [32] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, “Masked label prediction: Unified message passing model for semi-supervised classification,” 2021. arXiv:2009.03509.
  - [33] D. Kim and A. Oh, “How to find your friendly neighborhood: Graph attention design with self-supervision,” arXiv:2204.04879.
  - [34] J. You, R. Ying, and J. Leskovec, “Design space for graph neural networks,” 2021. arXiv:2011.08843.
  - [35] K. T. Schütt, P. J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K. R. Müller, “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions,” 2017.
  - [36] J. Gasteiger, S. Giri, J. T. Margraf, and S. Günnemann, “Fast and uncertainty-aware directional message passing for non-equilibrium molecules,” 2022.
  - [37] Y. Wang *et al.*, “ViSNet: An equivariant geometry-enhanced graph neural network with vector-scalar interactive message passing for molecules,” 2023.



**Nguyen Phat Dat** was born in 1992 in Ho Chi Minh City, Vietnam. In 2015, he earned B.Eng. in Chemical Engineering from the Ho Chi Minh City University of Technology. In 2016, he relocated to the USA and start doing photochemical research under the guidance of Prof. Oleg V. Larionov at the University of Texas at San Antonio. He then joined the research group of Prof. Loi H. Do at the University of Houston in 2017. His doctoral work encompassed studies on metal-catalyzed transfer hydrogenation and single-molecule detection using super-resolution fluorescence imaging (SRFM) within living cells. After earning the Ph.D degree in 2022, he joined the Faculty of Chemical and Food Technology at the HCMC University of Technology and Education. Email: [datnp@hcmute.edu.vn](mailto:datnp@hcmute.edu.vn). ORCID: <https://orcid.org/0000-0002-0634-8994>.



**Le Trong Phuc** was born in 2004 in Dong Nai Province, Vietnam. He is a 3<sup>rd</sup> year student at Ho Chi Minh City University of Technology and Education. His major is Chemical Technology and Engineering. In 2022, he won the 3<sup>rd</sup> prize in the National Chemistry Olympiad Competition. He is currently doing research under the supervision of Dr. Nguyen Phat Dat. Email: [lephuc270704@gmail.com](mailto:lephuc270704@gmail.com). ORCID: <https://orcid.org/0009-0007-6088-4767>