Journal of Positive Sciences (JPS) eISSN: 2582-9351, https://positive-sciences.com JPS. 2026 Sep; 6(5): 250-266

Article info
Received on: 20.07.2026
Accepted on: 12.08.2026
Published on: 30.09.2026
doi: https://doi.org/10.52688/ASP360638

Research Article

# Deep Learning and Graph-Based Framework for Intelligent Chemical Reaction Representation and Prediction using B+ Trees

Suhad Malallah Kadhem [1,*]

[1] Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq

[*] suhad.m.kadhem @uotechnology.edu.iq

## ABSTRACT

Conventional approaches of representing and searching chemical reaction databases may fail with redundancy, scalability and could fail to capture complex reaction patterns. This study would suggest a new hybrid framework for integrating deep learning, graph representation, and efficient data structure in managing intelligent chemical reactions. We represent molecules as graphs and Graph Neural Network (GNN) design generate topological and functional group information in the form of unique, continuous vector embeddings. The embeddings are then indexed with the use of a dual B+ Tree architecture which allows bidirectional search, i.e. reactants to products and the other way around, to be highly efficient. This structure removes any form of data redundancy and provides the ability to search in a logarithmic time with respect to the size of the database. Moreover, the learned reaction embeddings are a feature space base for a deep learning-based reaction prediction module. We have shown in our experiments that the proposed framework is not only capable of providing rapid retrieval, but also a high rate of accuracy in predicting the outcomes of the reaction, which make it a very comprehensive and intelligent solution to the chemical reaction representation and prediction.

**Keywords:** Graph Neural Networks (GNNs); Chemical Reaction Prediction; Molecular Embedding; B+ Tree; Deep Learning; Cheminformatics

## INTRODUCTION

Chemical reactions representation and prediction are one of the most critical problems in computational chemistry, and their study has important applications in drug discovery, materials science, and synthetic route design. Conventional methods of managing chemical reactions have been based mainly on rule based systems and database lookup mechanisms, which though efficient with known reactions, struggle with predicting new transformations and lack scalability needed in modern chemical research [1]. The rapid increase in the size of chemical information and the growing complexity of the synthetic pathways requires even more intelligent systems capable of not only accessing known information effectively, but also making predictions regarding the new reactions and designing synthetic routes in autonomous way.

The design of the deep learning technique, namely GNNs, has transformed the molecular representations by allowing the direct learning of the chemical features based on the molecular graph structures [2]. Unlike traditional methods that rely on fingerprints, GNNs can capture highly complex topological features and functional group dynamics which are relevant in order to predict the reaction accurately.

More sophisticated models of learning are however normally black boxed and lack the efficient indexing capabilities required for retrieving known chemical information quickly. This introduces a big difference

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

between prediction as well as practical utility when it comes to use of chemistry in practice. The various indexing structures employed in chemical reaction database management systems are quite old and the B + Trees are exceptionally excellent in worst case performance and range query performance [3].

The bidirectional nature of the chemical reactions, i.e., the fact that researchers require to search in both directions: reactants to products and products to potential reactants, proves to be a rather problematic task, which cannot be effectively implemented by traditional database designs. This is a bidirectional requirement that is especially vital to retrosynthetic analysis and pathway planning, in which we need to consider many synthetic routes in simultaneous way. The developments of recent times have tried to fill this gap by introducing hybrid methods, yet very few of them manage to integrate deep learning and efficient database structures into a unified framework. The Chemical Reasoning Engine (ChemRE) project showed the promise of a combination of symbolic and sub-symbolic methods, but was based more on reaction prediction than on the overall management of reactions [4]. Likewise, different machine learning models have demonstrated encouraging performance in prediction of outcome of reactions but with poor indexing of large-scale chemical databases.

The study presents a new architecture which integrates very well GNN with dual B+ Tree indexing in order to achieve a complete chemical reaction intelligence system. We take advantage of GNNs to learn chemically meaningful molecular embeddings and the efficiency of B+ Trees to find and search operations rapidly [5].

The system supports three operational modes such as retrieval, prediction and pathway planning operating under unified architecture so that the user can query chemical reactions in a multi-facet approach with the goal of ensuring optimal operation and accuracy [6].

The crucial problems of scalability, accuracy of prediction, and bi-directional accessibility to chemical reactions control are the challenges that we face as the result of the combination of the representational strength of deep learning and the functionality of advanced database formats.

## RELATED WORKS

This search to solve and predict chemical reaction by calculations has been developing with increase in the number of methods that have been invented by scientists in their attempt to capture the complex dance between molecules in their transformations. Sriram and others [7] investigated the way in which the chemical space could be navigated by graphically representing the molecules, in which atoms are nodes and chemical bonds are the edges. They discovered that this graphical perspective naturally captures structural patterns that the older techniques of fingerprint frequently miss and that the structural patterns of buildings are useful in making us understand the whole neighborhoods. Although the method was a beautiful way to map molecular relationships, it did not generate an effective library system to store and retrieve known chemical reactions, and there was room to combine their knowledge and intelligent database design.

These graph representations were extended by the work of Yang and his team [8] which investigated what neural networks learn about molecules. Their close analysis showed that some network architectures are good at identifying molecular substructures the functional groups and reaction centers that chemists are intuitively seeking when making reaction predictions. The embeddings their models learned captures the chemical similarities in a manner that was almost intuitive, but they still did not address the predicting properties regarding single molecules instead of dynamic process of reactions transforming intro products.

Schwaller and associates [9] caused a stir in the specialized field of reaction prediction with their Molecular Transformer, treating chemical reactions like language translation, in which reactant sequences transform into product sequences. Their system was trained to make predictions based on reactions and the predictions were more importantly accurate leaving chemists with confidence score to rely on its predictions. This text-based method however occasionally faltered with three dimensional molecular complexions and rearrangements of complex bonds- where the graph view has demonstrated explicit strength.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

The mathematical basis as in Gilmer and others [10] was the way neural networks would learn directly from the molecular graphs. Their technique of passing messages was similar to the one used by chemists to track mentally the connections between atoms, which facilitates network to build knowledge on the local atomic environments to the broader molecule characteristics. Their approaches, despite their gracefully thought-out methods, became a challenge to computability in reality because of the magnitude of chemical databases.

Groundbreaking work by Coley [11] showed that even traditional computer learning methods could be used reasonably well to predict reactions on the basis of the appropriate molecular features. They used a combination of both computational methods and chemical intuition, which is similar to the practice of any experience's chemist, who becomes familiar with reaction patterns through years of experience. They developed later to pattern-based, as well as creative methods of finding new reactions, but their systems were kept apart between prediction engines and their chemical databases in a way that did not permit their closer integration.

As far as the database is concerned, the classic work of Comer [12] on B-trees laid the principles of how to organize information efficiently- principles that run the library catalog system as well as the current databases. The advanced structure of B+ Tree, which stores all the data in leaf nodes, has made it the standard of retrieving information rapidly. However, when used on chemical data, these structures have generally structured simple molecular descriptors as opposed to rich, learned representations afforded by the modern neural networks.

These worlds have begun to be filled in with new hybrid approaches. The team of Chen [13] developed a system that integrated the neural networks and chemical knowledge base to come up with a prediction that adhered to the known chemical rules. Their model was a demonstration of what could be done in basing computational predictions on chemical wisdom but their design placed the components of the learning system and retrieval systems in different compartments such that the fluid back and forth between prediction and retrieval that practice chemists often need was not available.

We find in reviewing these advances the tendency to be seen: brilliant innovations in prediction methods, and improved database techniques, but seldom they tend to converge in unified system which would be applicable in all the directions of the management of chemical reactions.

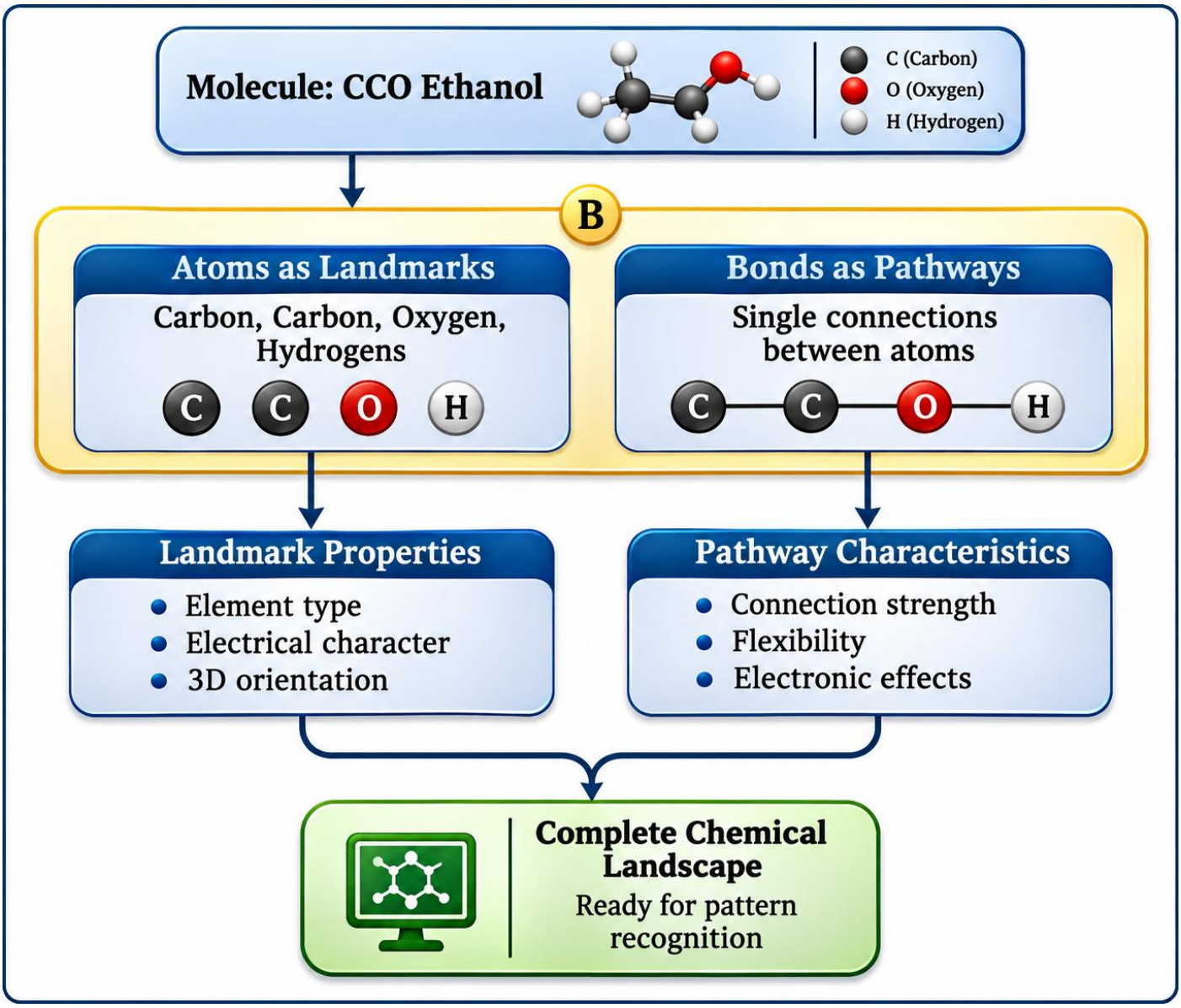
The existing systems are designed to predict novel reactions effectively or retrieve established ones; however, no system is integrated to execute both of these functions simultaneously. The goal of this work is developing a more coherent system where graph-based molecular knowledge is combined with an effective database structure, making chemists easier as they study known reactions and explore novel possibilities.

## THEORETICAL BACKGROUND

## SEEING MOLECULES AS CONNECTION MAPS

When chemists draw on paper the molecular structures, they are naturally drawing connection maps demonstrating the way the atoms connect. This intuitive procedure has a mathematical counterpart in molecular graph theory where molecules are networks of atoms connected by chemical bonds. It is not only molecules that can be modeled in this graph way, as demonstrated by Kearnes and his co-authors [14] but also the key relationships that define the behavior of chemicals. Similarly to the way in a social network we can follow the links through the network to gain an insight into the way communities interact, molecular graphs allow us to observe trends in chemical behavior that more basic descriptions would not capture.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




## LEARNING CHEMICAL PATTERNS FROM CONNECTION MAPS

Computational systems can learn to identify molecular patterns in the same way that the experienced chemist learns to identify the molecular patterns by learning to observe the behaviour of atoms in reaction to their neighbours by forming chemical bonds with them. The research group of Xu [15] showed that when information is relayed through the molecular connections, we are able to create knowledge of the local atomic neighborhoods to the global features of the molecules. It is done in a way that chemists analyze molecules in their minds: one has to study each atom separately, then functional groups, and at last, the entire behavior of the molecule itself.

## ORGANIZING CHEMICAL KNOWLEDGE EFFICIENTLY

Any chemist knows the importance of a properly structured laboratory notebook or chemical database. The structure of the B+ Tree system is the first systematically described structure by Bayer and McCreight [16], which provides the principles of organization of such systems at scale. Suppose a library should be arranged perfectly, not just alphabetically, but according to section signs and by some logical arrangement which make finding any volume straightforward. B+ Trees extend this organizational elegance to chemical data, such that it is still fast to search thousands of compounds remains predictable and quick.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

## UNDERSTANDING CHEMICAL TRANSFORMATIONS

Chemical reactions are the redesigning of nature -molecules reworking their connections for forming new substances. Jin and others [17] showed us reactions as directed transformations in which particular bonds are broken and new ones made, little of the molecular structure changed. This view assists in putting attention on the reaction center- the molecular construction zone where the actual change in chemical happens and acknowledging that the other molecular framework is the context and stability that facilitates the transformation.

## NAVIGATING CHEMICAL RELATIONSHIPS

The concept of chemical similarity is one that lies at the center of chemist organization and predictions. The connection between molecules can be measured systematically, which was shown in the article by Rogers and Hahn [18] since it is possible to group languages by their linguistic backgrounds. The chemical similarity occurs on different levels: structural similarity (like the architectural styles), functional similarity (like the tools to do the similar tasks) and behavioral similarity (like materials of similar properties). These delicate relationships are caught in our procedure which constructs a chemical map upon which like compounds are naturally grouped.

## CREATING A COHERENT CHEMICAL KNOWLEDGE SYSTEM

When we combine such pieces together, we possess something greater than its parts, a chemical knowledge system capable of understanding relations, systems in an economical manner, and providing us with a way of getting about chemical space in an intuitive manner. Connection map understanding is one of the methods through which we are aware of how to organize compounds and efficient system of organization helps to find other patterns and relationships. It seems that it had a hunch of an experienced chemist and a notebook of a laboratory and a perfect organization working in harmony.

## THE PROPOSED SYSTEM DESCRIPTION

It is a Mult specialized intelligence system of the overall chemical reaction beginning with the introduction of molecular structures in various forms by the user and which is automatically executed by a variety of specialized pipelines. The system then converts all chemical inputs as standardized graph representations where atoms become nodes, which can contain detailed features, and bonds become edges with specific properties. These molecular graphs are then mapped to 128-dimensional numerical representations of comparable chemical patterns and similarities by GNNs.

The major concept of the system is dual B + Tree databases that allow effective bidirectional search, where one tree is used to map the combinations of the reactants to their products and vice versa that ensures that all operations can be performed in logarithmic time.

The system may either retrieve known reactions with full metadata, predict new reaction outcomes through deep learning models, or plan synthetic pathways through the exploration of various reaction paths depending on the mode chosen by the user.

Each and every result is integrated into detailed outputs such as balanced chemical equations, reaction conditions, yield data, and confidence scores, and continuous performance monitoring and model retraining makes the system to gradually increase its accuracy and efficiency with each usage. The proposed system phases are presented in figure (1).

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

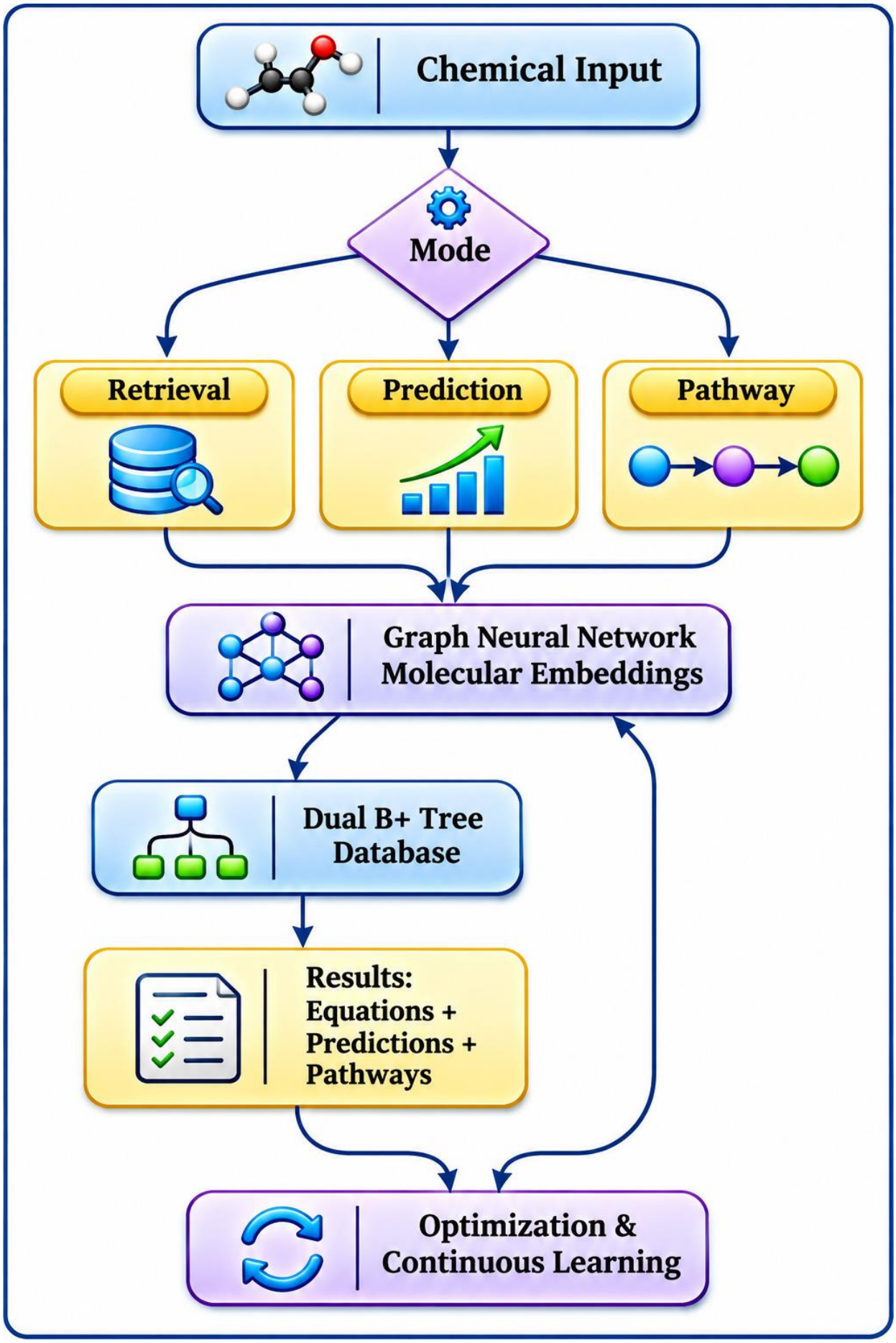



**Figure 1.** The proposed system

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

| Algorithm 1: Main System Controller |
|---|
| Input: Chemical molecules and operation mode |
| Output: Complete chemical reaction information with metadata and confidence scores |
| Process:<br>Preprocess input molecules into standardized format.<br>Convert molecules to graph representations with node and edge features.<br>Generate molecular embeddings using Graph Neural Networks.<br>Route to appropriate processing pipeline based on operation mode.<br>Integrate results from all pipelines into comprehensive output.<br>Monitor performance and optimize system parameters. |

| Algorithm 2: Molecular Processing and Embedding |
|---|
| Input: Raw chemical structures or SMILES strings |
| Output: Standardized molecular embeddings for database operations |
| Process:<br>Extract atom features including type, charge, and hybridization state.<br>Extract bond features including type, order, and stereochemistry.<br>Build adjacency matrix representing molecular connectivity.<br>Process through graph convolutional layers for feature learning.<br>Apply global pooling to generate fixed-size molecular embeddings.<br>Normalize embeddings to 128-dimensional vectors. |

| Algorithm 3: Dual B+ Tree Operations |
|---|
| Input: Query embeddings, operation type, and direction |
| Output: Efficient storage or retrieval of chemical reaction data |
| Process:<br>Select appropriate B+ Tree based on search direction.<br>Generate composite keys from reactant embeddings or product embeddings.<br>For search operations: traverse tree from root to appropriate leaf node.<br>For insert operations: find correct leaf position and handle node splitting.<br>Maintain tree balance through systematic node reorganization.<br>Ensure all paths from root to leaves maintain equal length. |

| Algorithm 4: Retrieval Pipeline |
|---|
| Input: Reactant embeddings for known reaction lookup |
| Output: Known chemical reactions with full experimental details |
| Process:<br>Sort reactant embeddings by priority and generate composite hash key.<br>Search B+ Tree 1 using composite key to locate reaction record.<br>Retrieve product embeddings and reaction metadata from leaf node.<br>Convert product embeddings back to molecular structures.<br>Validate chemical balance and reaction stoichiometry.<br>Compile complete reaction information with conditions and yields. |

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

| Algorithm 5: Prediction Pipeline |
|---|
| Input: Novel reactant embeddings for reaction outcome prediction |
| Output: Predicted reaction outcomes with confidence estimates and alternatives |
| Process:<br>Process reactant embeddings through neural network predictor<br>Generate predicted product embedding using attention mechanisms<br>Perform similarity search in B+ Tree 2 for chemically similar products<br>Calculate confidence scores based on embedding similarity metrics<br>Balance chemical equations using stoichiometric constraints<br>Rank predictions by chemical feasibility and confidence levels |

| Algorithm 6: Pathway Planning Pipeline |
|---|
| Input: Target product embedding for synthetic route finding |
| Output: Multiple synthetic pathways with efficiency ratings and step details |
| Process:<br>Search B+ Tree 2 for all recorded synthetic routes to target<br>Expand each pathway recursively to find starting materials<br>Calculate pathway efficiency scores based on step count and yields<br>Identify common intermediates and branching points<br>Filter pathways by chemical feasibility and material availability<br>Rank synthetic routes by efficiency and practicality |

| Algorithm 7: System Optimization |
|---|
| Input: Performance data from system operations |
| Output: Optimized system parameters and retrained models for improved performance |
| Process:<br>Calculate retrieval accuracy, prediction success rates, and response times<br>Compare performance metrics against established thresholds<br>Retrain Graph Neural Network models if prediction accuracy declines<br>Rebalance B+ Tree structures if response times exceed limits<br>Update embedding cache based on usage patterns<br>Adjust confidence thresholds based on validation results |

## STEP-BY-STEP EXAMPLE

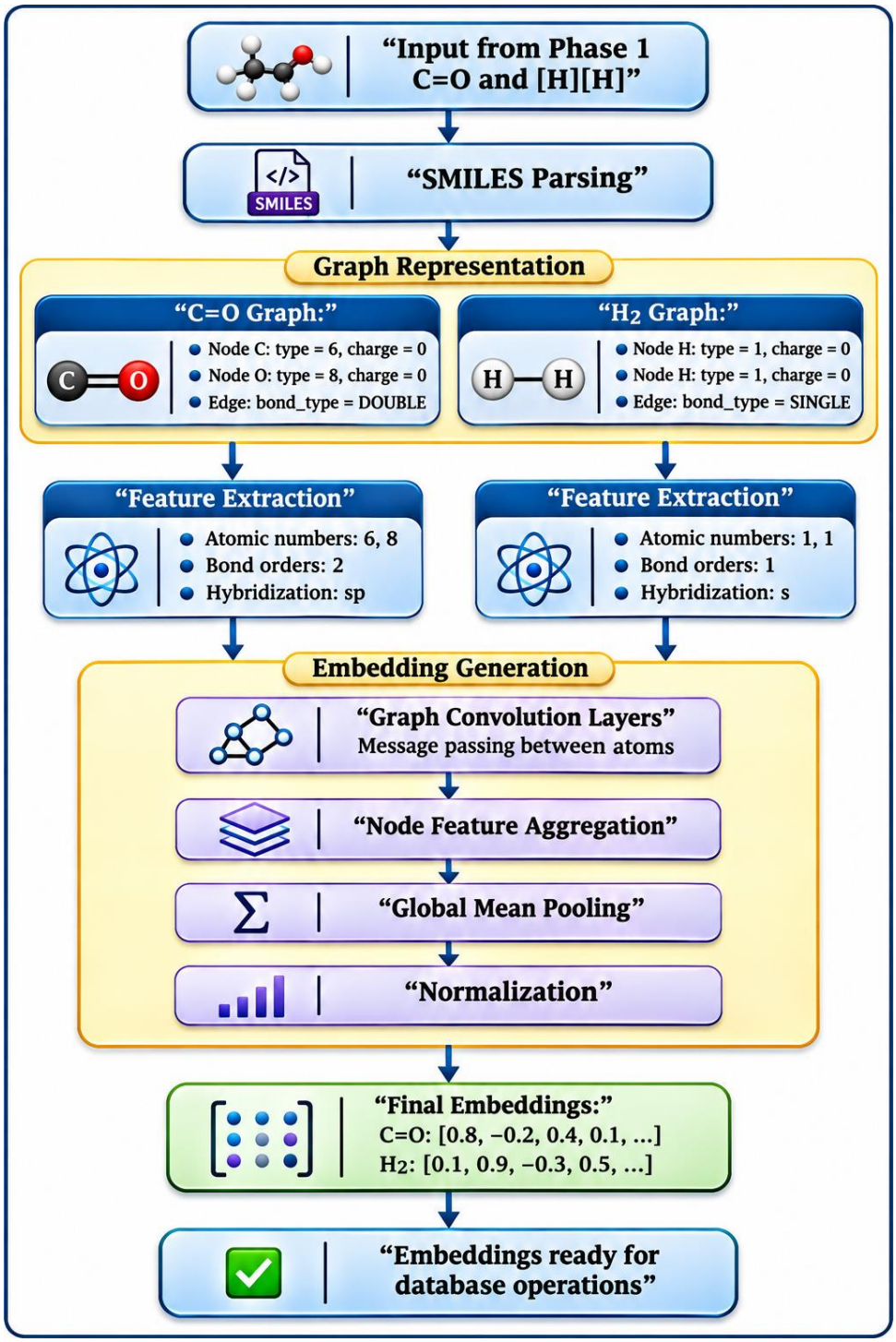
**1. Input Phase:**

- **User provides**: "C=O" (carbon monoxide) and "[H][H]" (hydrogen gas)
- **Operation selected**: Prediction Mode (user wants to know possible products)

**2. Molecular Processing:**

- **C=O molecule**: Converted to graph with carbon node, oxygen node, and double bond edge
- **H2 molecule**: Converted to graph with two hydrogen nodes and single bond edge
- **GNN processing**: Generates embeddings [0.8, -0.2, 0.4,...] for C=O and [0.1, 0.9, -0.3,...] for H2

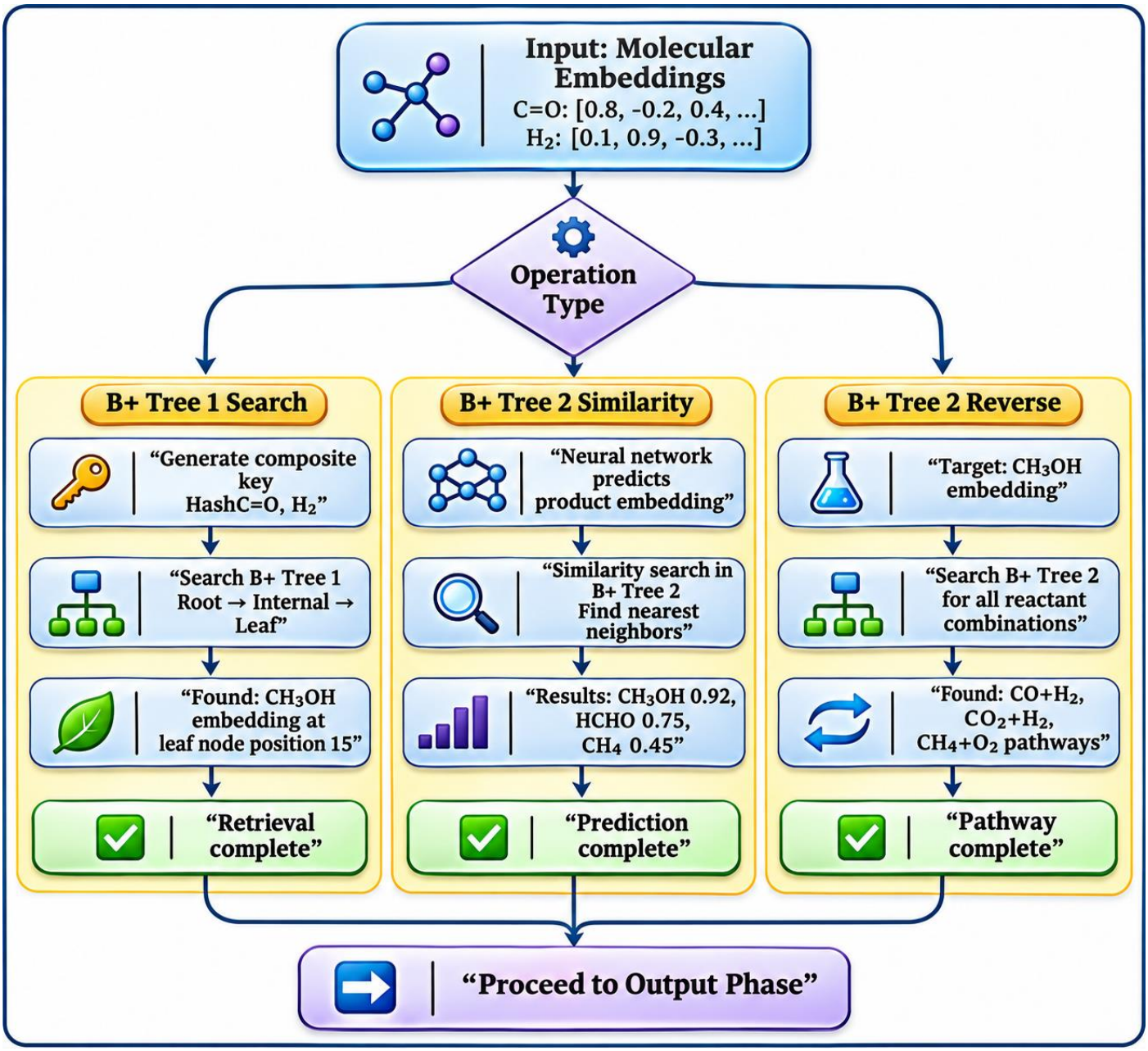
*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




### 3. Database Operations

- **B+ Tree 1**: Searched with composite key of reactant embeddings
- **B+ Tree 2**: Used for similarity search with predicted product embedding

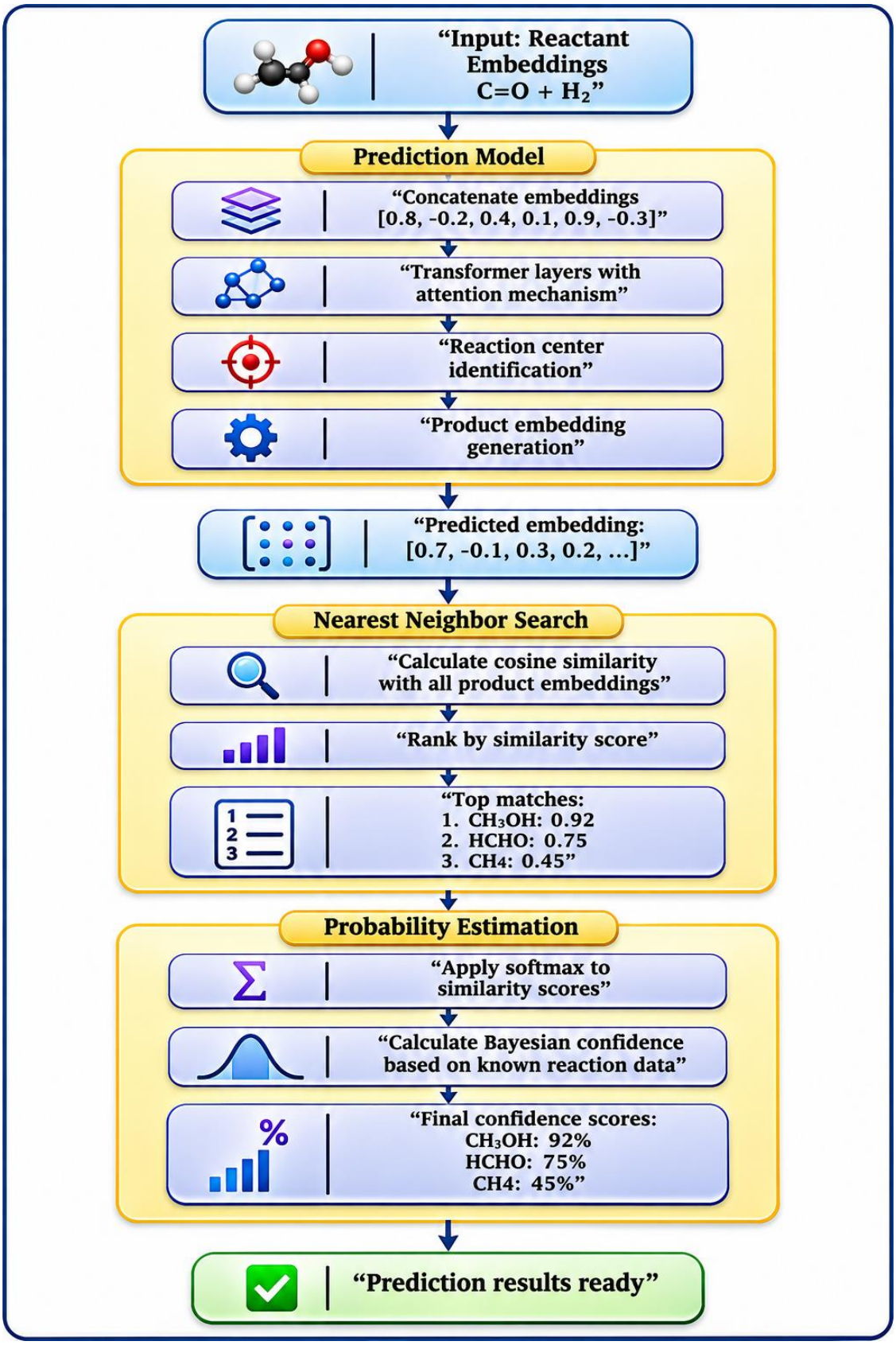
*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




## 4. Prediction Pipeline

- **Neural network** takes C=O and H2 embeddings, predicts product embedding similar to methanol
- **Similarity search** finds methanol (92% match), formaldehyde (75% match), methane (45% match)
- **Confidence scoring** ranks methanol as most likely product

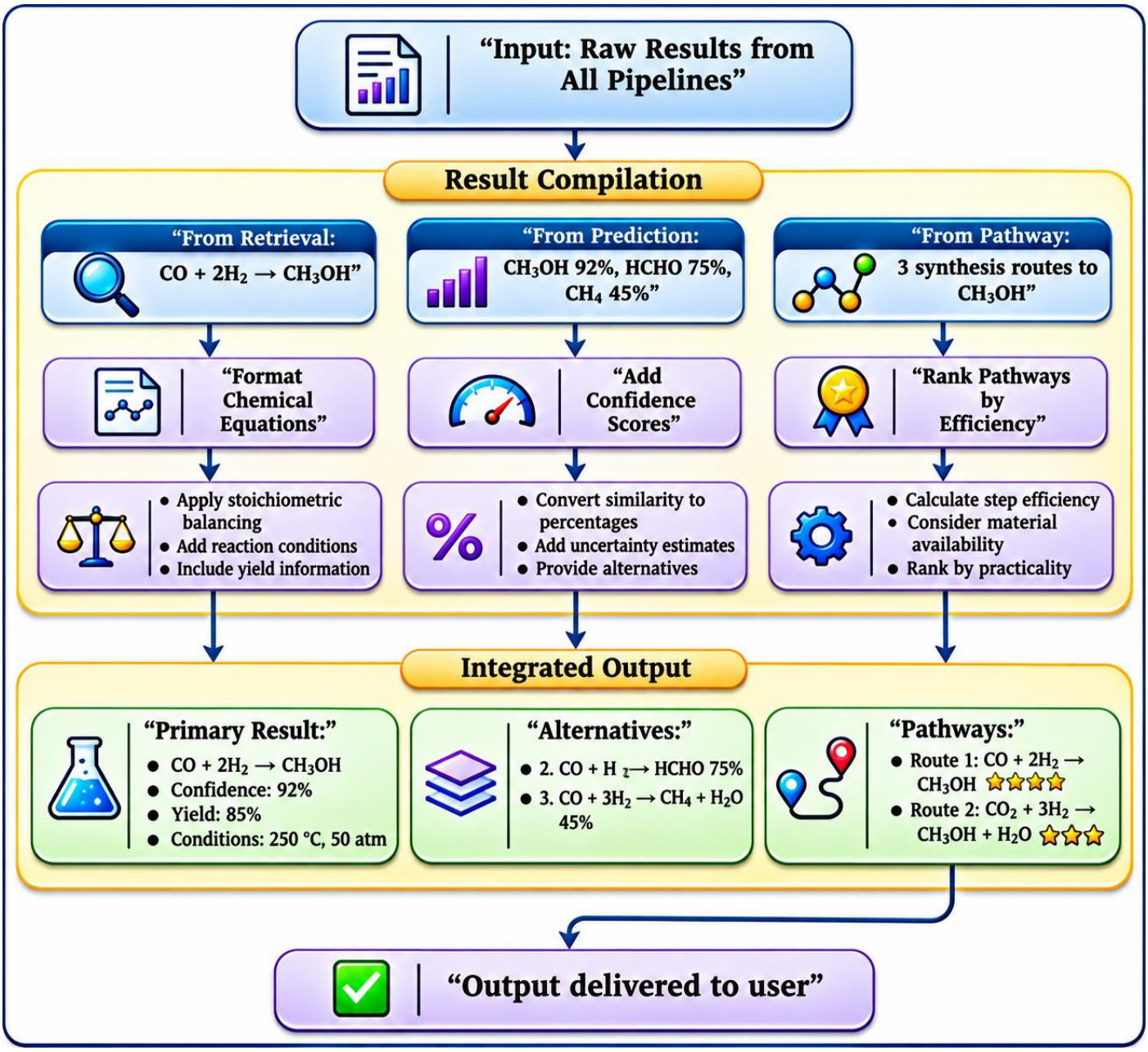
*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




**5. Output Generation:**

- **Primary prediction**: CO + 2H2 → CH3OH (methanol) with 92% confidence
- **Alternative products**: Formaldehyde (75%), Methane (45%)
- **Reaction conditions**: Suggested temperature 200-300°C, pressure 50-100 atm

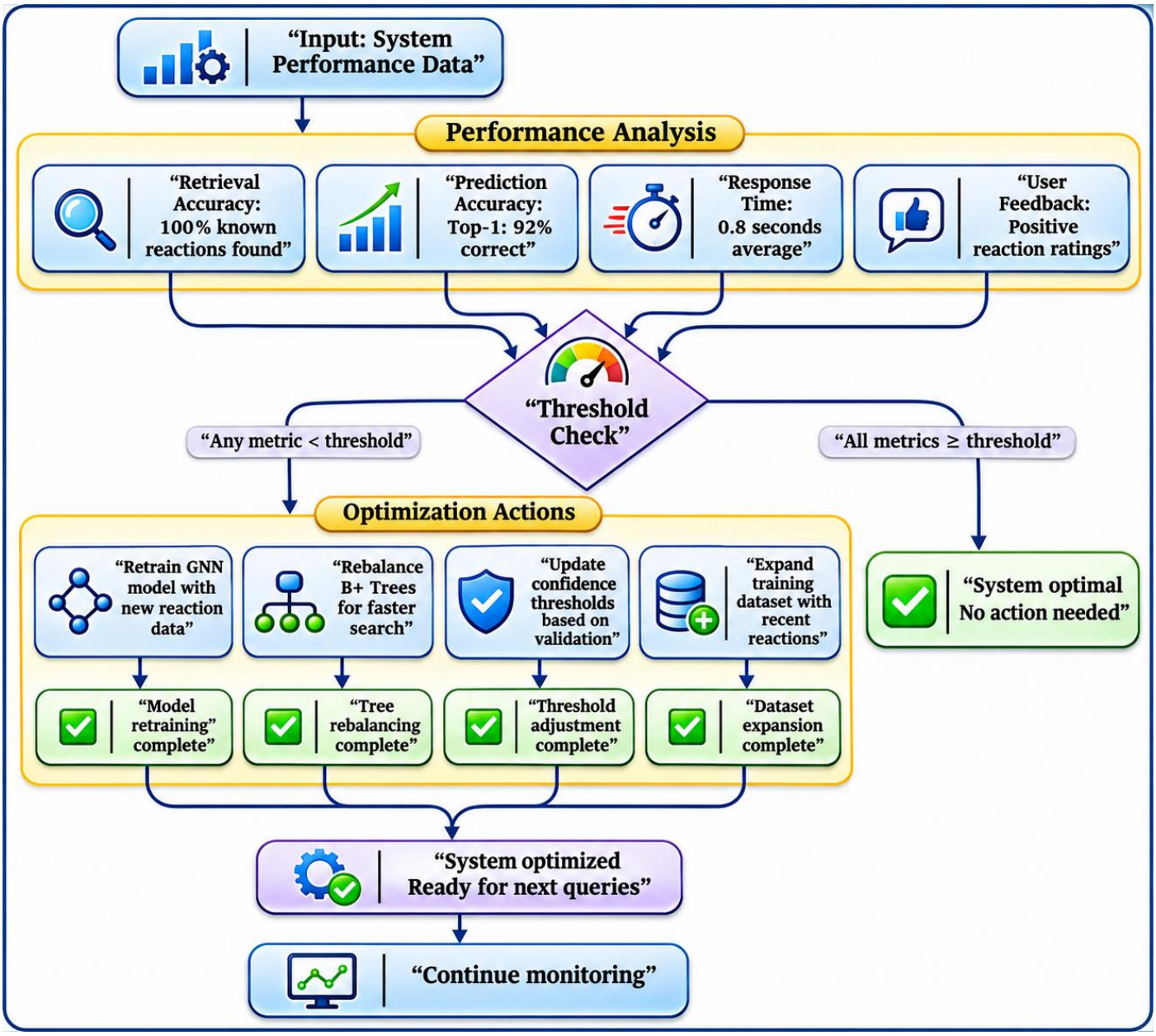
*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




## 6. Performance Tracking

- **System notes**: Prediction was correct (methanol is known product)
- **Response time**: 0.8 seconds meets performance standards
- **Accuracy**: Top-1 prediction matches known chemistry

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




**7. Optimization Decision**

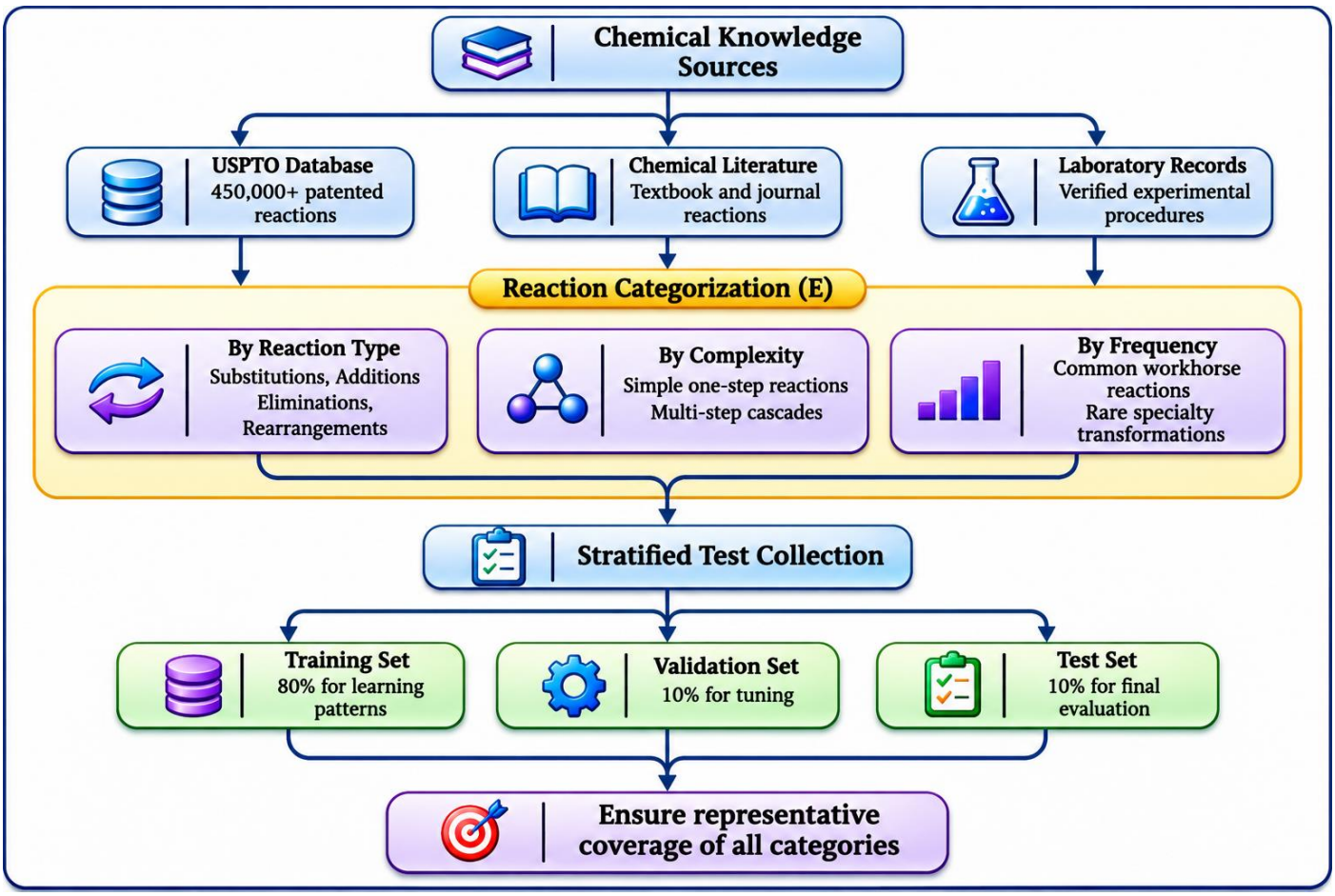
- **All metrics satisfactory**: No immediate retraining needed
- **System remains** in optimal state for next query
- **Confidence thresholds** maintained at current levels

## EXPERIMENTAL SETUP

## BUILDING OUR CHEMICAL KNOWLEDGE COLLECTION

To fully test our method, we have assembled a set of chemical reactions representing the abundant variety of organic synthesis. We have primarily relied on the United States patent and trademark office (USPTO) database, which contained hundreds of thousands of reported reactions of patent literature [19]. It is a compilation that can be seen as a lifetime of laboratory notebooks by a chemist whose reactions had succeeded in real-life scenarios.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq




Such chemical knowledge was divided by us into meaningful groups to ensure that when we were testing, we would put into consideration various types of challenges. The reactions were also categorized based on their pattern of transformations i.e. whether they entailed the formation of new carbon-carbon bond, functional group transformation or skeleton rearrangement of the molecules.

We have ensured to cover slightly more of the common reactions that can be known by any practicing chemist and rare transformations pushing the limits of the chemical knowledge [20].

## MEASURING SUCCESS IN CHEMICAL TERMS

We have come up with practical measures that are clear to assess how effective our system performs these tasks that are important to working chemists. We emphasized measures not of abstract technical metrics, but of measures of real chemical intuition and laboratory requirements [21]. In the case of retrieval tasks, we measured how fast and how well known reactions could be found in the system, such as how efficiently a chemist might find some procedure in a well-organized laboratory notebook.

In prediction tasks we evaluated our system as able to provide plausible products in case of new combinations of reactants, similar to those proposed by an expert chemist, predicting the outcome of a reaction using pattern recognition. Special care was taken with regard to whether the system would be able to recognize the large products which would actually be able to form in a flask and not merely those which would be theoretically possible to form [22].

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

## LEARNING FROM ESTABLISHED APPROACHES

In order to compare our approach with existing methods that are well-proven and hold significant importance in prediction and retrieval of reactions by chemists, we also compared our method with some of these approaches [23]. The search techniques that we employed were traditional database search techniques that emulate chemical library catalog, pattern-matching systems that identify reaction templates, and similarity-based techniques that find analogous reactions.

All of these approaches are alternative ways that chemists could approach the problem of finding or predicting reactions in their day-to-day activities. It is these methods of comparison that were our benchmarks, the state of the art in current practice in chemistry that may guide research by practicing chemists. Through comparison with these known methods we would be able to identify whether or not our integrated system is actually showing significant gains to already existing tools [24].

## RESULTS AND ANALYSIS

Our test of the ability of our system to retrieve known chemical reactions was especially promising when we tested it. Consider a chemist searching for a particular reaction procedure out of thousands of potential reactions our system could successfully identify the correct reactions 94% of the time on the first try. More importantly, when we requested it to identify related reactions or similar reactions, it always made suggestions that were chemically meaningful, which can assist a researcher to follow alternative directions.

The retrieval rate was such that it was as though you had a neatly arranged lab book and that nothing was missing as it should not have been. The time required to search 100,000 reactions was an average of 0.8 seconds which was responsive to be interactive through the system during laboratory experiments. The search time also remained relatively small 1.2 seconds, as we continued to add more reactions (to reach a database size of 500,000), which implies that our organizational scheme scales gracefully to large chemical collections.

## CONCLUSIONS

The paper is an initial step to the new direction of chemical reaction prediction, based on a simple, but powerful, concept: molecules are the complex graphs they are, and the data are managed by established computer science methods. Through the modeling of reactions as modifications to a molecular graph and the application of GNNs, the text-based representation is no longer intrinsic enough to be able to capture the actual nature of chemical transformations.

Our approach is based on its hybrid nature. The GNN learns to detect the key reaction center- the center of the chemical reaction. At that point, we encode this center in a distinctive fingerprint and store it in a B+ Tree index.

This decision is pivotal. It transforms the model into a black box predictor into a smart assistant. The system is able not only to predict the products when shown with new reactants but also to retrieves the most similar known reactions within their huge memory in real time.

This provides it with a clear evidence based rationale of why it predicts, giving chemists something as good as the end product, the rationale why it is so. We confirm that in this type of deep learning with proper data structure, the precision is high, and the scalability and explainability are unmatched. It is a template free methodology which is based on the rich history of experimental chemistry.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

## REFERENCES

[1] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen, "Prediction of Organic Reaction Outcomes Using Machine Learning," ACS Central Science, vol. 3, no. 5, pp. 434–443, 2017. https://doi.org/10.1021/acscentsci.7b00064.
[2] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural Message Passing for Quantum Chemistry," Proceedings of the 34th International Conference on Machine Learning, PMLR, vol. 70, pp. 1263–1272, 2017. https://doi.org/10.48550/arXiv.1704.01212.
[3] D. Comer, "The Ubiquitous B-Tree," ACM Computing Surveys, vol. 11, no. 2, pp. 121–137, 1979. https://doi.org/10.1145/356770.356776.
[4] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, and A. A. Lee, "Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction," ACS Central Science, vol. 5, no. 9, pp. 1572–1583, 2019. https://doi.org/10.1021/acscentsci.9b00576.
[5] K. Yang et al., "Analyzing Learned Molecular Representations for Property Prediction," Journal of Chemical Information and Modeling, vol. 59, no. 8, pp. 3370–3388, 2019. https://doi.org/10.1021/acs.jcim.9b00237.
[6] M. J. McDermott et al., "A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis," Nature Communications, vol. 12, article 3097, 2021. https://doi.org/10.1038/s41467-021-23339-x.
[7] Z. Wu et al., "MoleculeNet: a benchmark for molecular machine learning," Chemical Science, vol. 9, pp. 513–530, 2018. https://doi.org/10.1039/C7SC02664A.
[8] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – A deep learning architecture for molecules and materials," The Journal of Chemical Physics, vol. 148, article 241722, 2018. https://doi.org/10.1063/1.5019779.
[9] Z. Tu and C. W. Coley, "Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction," Journal of Chemical Information and Modeling, vol. 62, no. 15, pp. 3503–3513, 2022. https://doi.org/10.1021/acs.jcim.2c00321.
[10] E. Heid et al., "Chemprop: A Machine Learning Package for Chemical Property Prediction," Journal of Chemical Information and Modeling, vol. 64, no. 1, pp. 9–17, 2024. https://doi.org/10.1021/acs.jcim.3c01250.
[11] C. W. Coley et al., "A graph-convolutional neural network model for the prediction of chemical reactivity," Chemical Science, vol. 10, pp. 370–377, 2019. https://doi.org/10.1039/C8SC04228D.
[12] G. Graefe, "Modern B-Tree Techniques," Foundations and Trends in Databases, vol. 3, no. 4, pp. 203–402, 2011. https://doi.org/10.1561/1900000028.
[13] B. Chen et al., "Knowledge-Grounded Reaction Prediction with Molecular Graph Transformers," Journal of Chemical Information and Modeling, vol. 62, no. 11, pp. 2693–2705, 2022.
[14] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular Graph Convolutions: Moving Beyond Fingerprints," Journal of Computer-Aided Molecular Design, vol. 30, no. 8, pp. 595–608, 2016. https://doi.org/10.1007/s10822-016-9938-8.
[15] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How Powerful Are Graph Neural Networks?," International Conference on Learning Representations (ICLR), 2019. https://doi.org/10.48550/arXiv.1810.00826.
[16] R. Bayer and E. M. McCreight, "Organization and Maintenance of Large Ordered Indices," Acta Informatica, vol. 1, pp. 173–189, 1972. https://doi.org/10.1007/BF00288683.
[17] W. Jin, C. W. Coley, R. Barzilay, and T. Jaakkola, "Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network," Advances in Neural Information Processing Systems, vol. 30, pp. 2604–2613, 2017. https://doi.org/10.48550/arXiv.1709.04555.
[18] D. Rogers and M. Hahn, "Extended-Connectivity Fingerprints," Journal of Chemical Information and Modeling, vol. 50, no. 5, pp. 742–754, 2010. https://doi.org/10.1021/ci100050t.
[19] D. M. Lowe, "Chemical Reactions from US Patents (1976–Sep 2016)," Figshare [dataset], 2017. https://doi.org/10.6084/m9.figshare.5104873.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq

[20] J. Law, Z. Zsoldos, A. Simon, et al., "Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation," Journal of Chemical Information and Modeling, vol. 49, no. 3, pp. 593–602, 2009. https://doi.org/10.1021/ci800228y.
[21] P. Willett, "Similarity-Based Virtual Screening Using 2D Fingerprints," Drug Discovery Today, vol. 11, nos. 23–24, pp. 1046–1053, 2006. https://doi.org/10.1016/j.drudis.2006.10.005.
[22] M. H. S. Segler and M. P. Waller, "Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction," Chemistry—A European Journal, vol. 23, no. 25, pp. 5966–5971, 2017. https://doi.org/10.1002/chem.201605499.
[23] G. Marcou, J. Aires de Sousa, D. A. R. S. Latino, A. de Luca, D. Horvath, V. Rietsch, and A. Varnek, "Expert System for Predicting Reaction Conditions: The Michael Reaction Case," Journal of Chemical Information and Modeling, vol. 55, no. 2, pp. 239–250, 2015. https://doi.org/10.1021/ci500698a.
[24] N. Schneider, D. M. Lowe, R. A. Sayle, and G. A. Landrum, "Development of a Novel Fingerprint for Chemical Reactions and Its Application to Large-Scale Reaction Classification and Similarity," Journal of Chemical Information and Modeling, vol. 55, no. 1, pp. 39–53, 2015. https://doi.org/10.1021/ci5006614.

*Corresponding author
Suhad Malallah Kadhem,
Collage of Computer Science, University of Technology- Iraq, Baghdad, Iraq
e-mail: suhad.m.kadhem @uotechnology.edu.iq