Communication in Physical Sciences, 2024, 11(1): 47-75

Machine Learning in Thermochemistry: Unleashing Predictive Modelling for Enhanced Understanding of Chemical Systems

Humphrey Sam Samuel, Emmanuel Edet Etim*, John Paul Shinggu. Bulus Bako
Received: 23 December 2023/Accepted: 18 March 2024 /Published:29 March 2024

Abstract: Machine Learning (ML) has become a game-changing tool in many scientific sectors, altering research and spurring progress in a wide range of fields. The incorporation of ML approaches has created new predictive modelling opportunities in the context of thermochemistry, enabling more accurate and efficient prediction of the thermodynamic parameters of chemical systems. The article emphasizes the use of machine learning techniques in thermochemistry, highlighting the potential advantages and difficulties encountered in this quickly expanding field. The application of these algorithms helps in the prediction of fundamental thermodynamic quantities, including enthalpy, entropy, heat capacity, and free energy, allowing researchers to learn more about the energetics of chemical reactions and the stability of intricate molecular systems. The article also discusses openness, accountability, and the appropriate use of these formidable tools to ensure scientific integrity and prevent potential biases. These issues are related to the ethical problems linked with the application of ML in thermochemistry. As a result of the application of machine learning to thermochemistry research, a new era of predictive modelling has begun, offering a variety of opportunities to understand the intricate workings of chemical systems. ML provides enormous promise for expediting scientific discovery and improving our comprehension of thermodynamics in chemistry by eliminating obstacles and incorporating moral principles.

Keywords: *Machine learning, thermochemistry, artificial intelligence*

Humphrey Sam Samuel
Computational Astrochemistry and Bio-Simulation Research Group, Federal University Wukari
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: humphreysedeke@gmail.com
Orcid: 0009-0001-7480-4234

Emmanuel Edet Etim
Computational Astrochemistry and Bio-Simulation Research Group, Federal University Wukari
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: emmaetim@gmail.com
Orcid: 0000-0001-8304-9771

John Paul Shinggu
Computational Astrochemistry and Bio-Simulation Research Group, Federal University Wukari
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: Johnshinggu@gmail.com
Orcid: 0009-0005-2216-3155

Bulus Bako
Computational Astrochemistry and Bio-Simulation Research Group, Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: bakobulus01@gmail.com

Orcid: 0009-0001-3946-0712

1.0 Introduction

The study of the heat energy changes that take place during chemical reactions and other physical processes is the focus of the discipline of physical chemistry known as thermochemistry. Chemical engineering, material science, pharmaceuticals, and environmental chemistry are some scientific and commercial fields that rely on thermochemistry to a good extent. Thermochemistry, a subfield of physical chemistry, studies the energy changes that take place during physical and chemical processes. Understanding the stability, reactivity, and behaviour of chemical systems depends on the precise prediction of thermodynamic parameters (Oliveira, *et al.*, 2022). The exploration of huge chemical regions was previously constrained by the time-consuming and complicated experimental procedures required to achieve these properties. With its invaluable insights into the stability, reactivity, and viability of chemical reactions, thermochemistry is crucial to both scientific inquiry and industrial applications (Agúndez, *et al.*, 2015). Researchers can refine procedures, create effective reactions, and create environmentally friendly technology by looking at the energy changes that take place during chemical transformations. However, the development of machine learning (ML) has fundamentally changed how scientists approach thermochemistry. Thermodynamic characteristics can be predicted more effectively and accurately using machine learning (ML), a potent computational technology that makes it possible to identify patterns and insights from enormous amounts of data (Samuel, *et al.*, 2023). Researchers can now unleash the potential to speed discoveries, improve chemical processes, and spur innovation in a variety of sectors like material science and medicines by utilizing ML algorithms. A branch of artificial intelligence known as machine learning (ML) involves the creation of statistical models and techniques that allow computers to learn from data without explicit programming. ML provides a potent and cutting-edge method for predicting thermodynamic parameters and comprehending energy changes in chemical systems in the context of thermochemistry (Gúndez , *et al.*, 2018). Although machine learning has the potential to be transformative, there are still some ethical issues that are associated with it concerningthermochemistry research. To preserve scientific integrity, reduce biases, and uphold ethical norms in scientific pursuits, transparency, accountability, and responsible usage of ML algorithms are crucial. A paradigm change in the field of thermochemistry research has been achieved with the incorporation of machine learning. Predictive modelling and data-driven insights made possible by machine learning enable researchers to comprehend and manipulate chemical systems more quickly. ML offers the key to opening up new thermochemical frontiers, promoting innovation, and creating a more sustainable future by resolving obstacles and embracing ethical behaviours (Hirota, *et al.*, 2002). ML algorithms can be trained on enormous datasets of experimental and theoretical thermodynamic data. Then, compared to conventional techniques, these models can be utilized to estimate the thermodynamic properties of complex chemical systems (Janet, *et al.*, 2020). Also, machine learning makes it possible to extract significant correlations and patterns from huge datasets. (Lee, *et al.*, 2021). Discoveries are made far more quickly thanks to ML's high-throughput chemical system screening capabilities, which hasten the development of new materials, catalysts, and reaction pathways. By doing so, it is quicker to identify candidates for additional experimental validation. Furthermore, Complex systems with lots of variables are frequently used in thermochemistry research (Mattioda, *et al.*, 2020). In contrast to conventional linear regression methods, ML techniques like neural networks can handle this complexity and describe nonlinear interactions between variables. ML can be

used to develop valuable surrogate models based on available data, boosting the knowledge base for researchers, in circumstances when experimental thermodynamic data is scarce or expensive to gather (McGuire, 2018). Understanding energy changes and reactions in chemical systems through thermochemistry is crucial for gaining a fundamental understanding of a variety of scientific and practical applications. A transformative technique for improving our understanding of energy changes in chemical systems, the incorporation of machine learning in thermochemistry research has major benefits for predictive modelling, data-driven insights, and speeding up scientific discovery (Zhao, *et al.,* 2020). The aim of the study examine how machine learning is used in thermochemistry research and demonstrate its potential for predictive modelling and data-driven discoveries. The scope explores the many machines learning (ML) approaches frequently used in this setting. Researchers may create models that precisely predict crucial thermodynamic variables like enthalpy, entropy, heat capacity, and free energy using these techniques.

1.1 Role of Machine Learning in Thermochemistry

Machine Learning (ML) has become a game-changing technology in many fields of science, including chemistry. ML that are applied in thermochemistry have revolutionized how scientists examine energy fluctuations and reactions in chemical systems. The study of chemical spaces, the identification of novel materials, and the prediction of thermodynamic properties are all made possible by machine learning approaches (Ori, *et al.,* 2023). The machine learning's function in thermochemistry includes:

i. Predictive Thermodynamic Property Modelling: Large datasets of experimental and theoretical thermodynamic data can be used to train machine learning algorithms to create prediction models. These models can then be used to calculate complex chemical systems' enthalpy, entropy, heat capacity, and free energy, among other thermodynamic parameters (Etim, *et al.,* 2020). Traditional approaches can be greatly outperformed by ML models, which also produce predictions that are more precise and dependable. An example is predicting the OER catalytic activities of hOER by comparing DFT and machine learning. The best performing ML models obtained in figure 1.0 is then used to predict the OER catalytic activities hOER of a SAC of the remaining 14 transition metal species (that is, Sc, Y, Zr, Tc. The most stable isotopes of Tc have a half-life ranging from 211,000 years to 4.21 million years.], Nb, Rh, Cd, Hf, Ta, W, Re, Os, Ir, and Hg) on an SV site and a DV site on a carbon surface, respectively (Chen, *et al.,* 2020).

ii. High-Throughput Screening and compounds Discovery: Thermochemistry research sometimes entails sifting through a sizable chemical space in search of compounds with particular features. High-throughput screening powered by ML enables researchers to quickly evaluate a huge pool of possible candidates. ML expedites the development of new chemicals, catalysts, and materials for a variety of applications, including batteries, catalysts, and solar cells, by predicting the thermodynamic properties of a wide range of materials (Jia, *et al.,* 2018).

iii. Mechanistic Studies and reaction prediction: Mechanistic studies and the prediction of reaction pathways are two areas where machine learning (ML) can be helpful. ML models can pinpoint the most likely reaction routes and transition states by examining the energy landscapes of chemical reactions. For comprehending reaction mechanisms and creating effective synthetic pathways, this knowledge is important (Zhang, *et al.,* 2022).

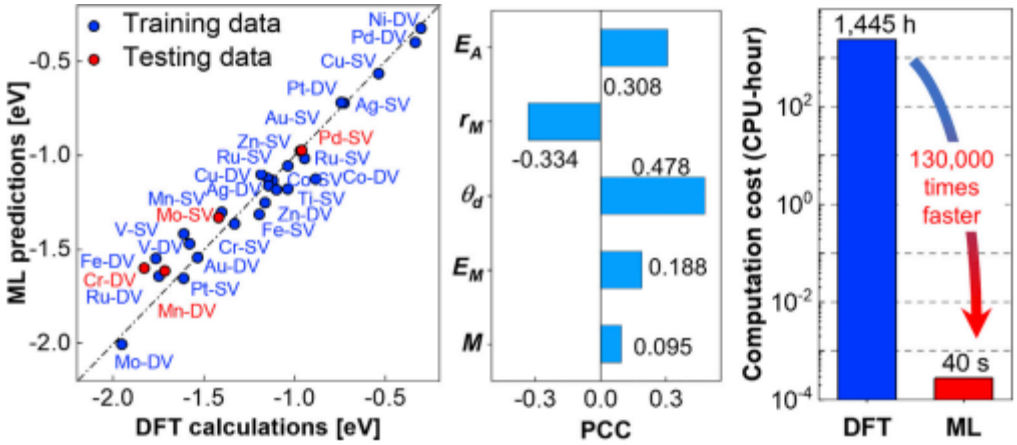

Fig.1: Machine learning in predicting the OER catalytic activity (Chen, *et al.*, 2020).

iv. Solvent and catalyst selection: The best catalysts and solvents for a given chemical reaction can be chosen with the help of machine learning. Researcher decisions on the settings of reactions can be guided by ML models, which examine the connections between solvent or catalyst parameters and thermodynamic outcomes. This leads to higher yields and selectivity (Zhao, *et al.*, 2020).

v. Data Augmentation and Surrogate Models: ML can be used to produce useful surrogate models based on available data in situations when experimental thermodynamic data is scarce or expensive to collect. Researchers may explore broader chemical spaces and get beyond data constraints thanks to these surrogate models, which serve as approximate representations of complicated thermodynamic phenomena (Tessarini, *et al.*, 2022).

vi. Density Functional Theory (DFT) and Quantum Chemistry: To speed up computations and lower expenses, ML techniques have been combined with quantum chemistry approaches like Density Functional Theory (DFT). Quantum chemical simulations are more accessible and useful for large-scale studies when combined with machine learning (ML), which can be a useful tool for predicting DFT-calculated properties. Researchers can use ML to get data-driven insights into the connections between molecular architectures, thermodynamic features, and chemical reactivity (Etim, *et al.*, 2023). Advanced interpretability techniques also assist researchers in comprehending how ML models generate certain predictions, hence increasing the models' dependability and trustworthiness (Etim, *et al.*, 2023).

vii. Rational drug development: The rational design of pharmaceutical compounds can benefit from the use of ML in thermochemistry in pharmaceutical research. When choosing lead compounds and improving drug prospects, predictive models can calculate solubility, other thermodynamic parameters, and drug-target binding affinities (Whitley, *et al.*, 2012).

viii. Impact on the Environment and Sustainable Energy: ML in thermochemistry can advance the study of sustainable energy by enhancing materials for energy conversion and storage systems. ML can direct the creation of more effective and sustainable energy solutions by predicting the thermodynamic properties of materials (Kuz'min, *et al.*, 2021).

Machine learning has a significant and varied role in thermochemistry. Predictive modelling, data-driven insights, and high-throughput screening are made possible by ML approaches, which promote faster discovery, better response design, and the creation of environmentally friendly solutions (Komp, *et al.*, 2022). The study of energy changes and reactions in chemical systems has the potential to be revolutionized by the incorporation of ML

with thermochemistry, advancing both academic and commercial endeavours.

1.2 Complex thermodynamic datasets using machine learning

1.3 Data Processing and Analysis

1.3.1 Preprocessing and cleaning of thermochemical data for machine learning applications

Thermochemical data must first be preprocessed and cleaned before being used in machine-learning applications (Khan, *et al.,* 2022). The ability of the ML models to properly learn from the data and produce precise predictions is ensured by proper data preparation. The primary steps in preprocessing and cleansing thermochemical data are listed below:

i. Data collection and compilation: Gather information on thermochemistry from a variety of sources, including experimental databases, literature, and computer simulations. Make sure the information is pertinent to the particular thermodynamic properties or reactions you are interested in (Barto, *et al.,* 2022).

ii. Data Formatting and Representation: Format the gathered data so that ML algorithms may use it. Make sure the data is presented consistently, with each sample (such as a chemical reaction or compound) and its related aspects (such as the molecular properties and reaction conditions) being correctly organized.

iii. Handling Missing Data: Due to mistakes in the experiments or insufficient measurements, thermochemical data may have missing values. Missing data can negatively impact the training and predictions of ML models. Consider deleting samples with a considerable amount of missing data, or handling missing data using imputation approaches like mean, median, or regression-based imputation (Ushie, *et al.,* 2017).

iv. Outlier Removal and Detection: Outliers are data points that dramatically vary from the pattern of the data as a whole. They might be the outcome of anomalous data or flawed experiments. Find and eliminate outliers to stop them from biasing predicts and impacting model training (Baskes, *et al.,* 1997).

v. Feature selection: The most important characteristics that affect the prediction of thermodynamic properties should be chosen as features (such as molecular descriptors and reaction conditions). Feature selection aids in lowering the data's dimensionality and preventing over fitting.

vi. Scaling of Features: The features should be normalized or scaled to a common range, such as 0 to 1 or -1 to 1. A feature cannot dominate the learning process by itself; hence feature scaling ensures that all features have an equivalent impact on model training (Behler, *et al.,* 2017).

vii. Encoding Categorical Variables: If the data includes categorical variables (such as the kind of chemical substance or type of reaction), encode them into numerical representations using methods such as one-hot encoding or label encoding so that they can be used successfully in ML algorithms (Behler, & Parrinello, 2007).

viii. Train-Test Split: Separate the training and test sets from the preprocessed data. The ML model is trained using the training set, and its performance is assessed using the testing set. For training and testing, typical splits are 70-30 and 80-20, respectively.

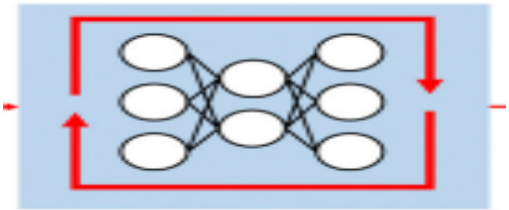

Fig. 2: schematic representation of train-test split (Behler, & Parrinello, 2007).

ix. Adjusting for Class Imbalance: The data may occasionally be unbalanced, with samples for particular thermodynamic properties or reactions being much smaller than samples for others. Utilize methods like oversampling, under-sampling, or SMOTE (Synthetic Minority Over-sampling Technique) to

correct for class imbalance (Behn, *et al.*, 2011).

x. Cross-Validation: Use cross-validation methods to check the model's generalization performance and lower the likelihood of over fitting, such as k-fold cross-validation.

1.3.2 Feature extraction and selection techniques for thermochemistry datasets

In thermochemistry datasets, feature extraction and selection approaches are crucial for locating the most pertinent and informative features that help to properly predict thermodynamic properties. These methods aid in lowering the number of dimensions in the data, enhancing model effectiveness, and preventing over fitting (Angulo, *et al.*, 2022). The following list of frequent feature extraction and selection methods for thermochemistry datasets:

i. Principal Component Analysis (PCA): A popular method for feature extraction is PCA. The initial features are changed into a fresh collection of uncorrelated variables (principal components) that fully account for the volatility in the data. PCA decreases the dimensionality of the data while maintaining the most crucial information by choosing a selection of the most important principle components (Velasco, *et al.*, 2022).

ii. Autoencoders: This class of neural networks is utilized for feature extraction and unsupervised learning. They are made to learn compressed versions of the input data, which are subsequently used as the basis for the features that are extracted. To find underlying structures and patterns in thermochemical data, autoencoders can be useful (Etim *et al.*, 2018).

iii. Molecular Descriptors: Molecular descriptors are frequently utilized as features in thermochemistry datasets involving molecular systems. Chemical structures and characteristics, such as atom counts, bond lengths, and electrical properties, are represented numerically by molecular descriptors (Neumann, *et al.*, 2005). They are suitable characteristics for predicting thermodynamic properties because they offer useful information about the chemical make-up and molecular structure of molecules (Ori *et al.*, 2024).

iv. Reaction fingerprinting: Reaction fingerprints can be utilized as characteristics for datasets including chemical reactions. In a binary format, reaction fingerprints represent the presence or absence of particular reactants, products, and intermediates. They encapsulate the essence of the reaction and can be used to pinpoint important elements that have an impact on the process's thermodynamics (Madden, *et al.*, 2002).

v. Techniques based on information theory: For feature selection, information theory-based techniques like mutual information and entropy can be applied. These methods assess how relevant and redundant features related to the desired attribute. For model training, features with a high degree of relevance and little duplication are chosen (Giambagli, *et al.*, 2021).

vi. Recursive Feature Elimination (RFE): RFE is a feature selection technique that operates by repeatedly eliminating the dataset's least significant features. It entails repeatedly training the ML model while excluding the least significant feature each time until the necessary number of features is obtained. The most important features for predicting thermodynamic properties are found via RFE.

vii. Least Absolute Shrinkage and Selection Operator (LASSO): A regularization method known as LASSO can be used to both select features and shrink model parameters. It does feature selection by introducing a penalty term into the regression model that encourages some of the feature coefficients to be absolutely zero (Chowdhury, *et al.*, 2021).

viii. Feature selection based on correlation: The correlation between each feature and the desired thermodynamic attribute is

measured by correlation-based approaches. High correlation coefficient features are thought to be pertinent.

The size, scope, and intended level of complexity of the ML model all influence the choice of feature extraction and selection approaches for the thermochemistry dataset. Researchers can improve the effectiveness and interpretability of ML models for predicting thermodynamic characteristics in chemical systems by successfully extracting and choosing pertinent features (Wei, *et al.,* 2023).

1.4 Property Prediction and Modelling

1.4.1 Prediction of thermodynamic properties using machine learning techniques

A hot topic in thermochemistry research is the prediction of thermodynamic properties using machine learning methods. For chemical systems, ML models may precisely estimate a range of thermodynamic characteristics, including enthalpy, entropy, heat capacity, and free energy (Etim, *et al.,* 2020). Here are some pertinent instances of the application of ML approaches to the prediction of thermodynamic property:

i. Prediction for Enthalpy of Formation: A key thermodynamic property used to evaluate the stability and energy content of chemical compounds is the standard enthalpy of formation (H°f). A collection of experimentally observed enthalpies of formation for distinct chemicals can be used to train ML models like support vector machines (SVM), random forests, or neural networks (Leonard, *et al.,* 2023). The trained model may then accurately predict the production enthalpies of novel molecules. For instance, the enthalpy of formation for a newly synthesized material can be predicted using an SVM model trained on a dataset of known H°f values, assisting in the discovery and design of new materials.

ii. Reaction energy prediction: The energy changes resulting from chemical reactions can be predicted using ML approaches. Regression models like random forests or gradient boosting can estimate the enthalpy change (ΔH) or the Gibbs free energy change (ΔG) of a reaction by using chemical descriptors and reaction fingerprints as features. These models are used in chemical synthesis for comprehending reaction feasibility and maximizing reaction conditions (Ding, *et al.,* 2021).

iii. Prediction of solubility: Machine learning (ML) models can be trained on experimental solubility data to predict the solubility of chemical compounds in various solvents or environments. For instance, a neural network can be trained to learn the connections between solubility and chemical characteristics to predict how a compound's solubility would change with temperature or solvent polarity (Wu, *et al.,* 2021).

iv. Vapour Pressure Prediction: A crucial thermodynamic factor affecting the volatility and phase behaviour of substances is vapour pressure. To predict the vapour pressure of various substances at various temperatures, machine learning (ML) models, such as support vector regression or decision trees, can be trained on experimental vapour pressure data. These models are useful for developing vapour-liquid equilibrium processes and comprehending the behaviour of volatile substances.

v. Heat Capacity Prediction: ML methods, such as deep neural networks and Gaussian process regression, can be used to predict heat capacity at various temperatures. To estimate heat capacities for compounds for which there are insufficient experimental data, models can be trained on heat capacity data obtained using calorimetric techniques (Longqiang, *et al.,* 2023).

vi. Phase diagram prediction: For multi-component systems, phase diagrams can be created using ML models. The phase

borders and coexistence zones in phase diagrams can be predicted by ML models using Gibbs free energy data and taking into account phase stability requirements (Osigbemhe, *et al.,* 2022; 2022a).

vii. Drug Design Thermodynamic Property Prediction: By foreseeing thermodynamic features important to drug-target interactions, machine learning (ML) plays a significant role in drug design (Cheng, *et al.,* 2023). Models that have been developed using data on solubility and known drug-target binding affinities can help identify new drug candidates with advantageous thermodynamic profiles. Employing machine learning approaches to predict thermodynamic parameters has considerable efficiency and accuracy benefits. These models help progress a variety of scientific and commercial applications by enabling researchers to explore enormous chemical regions, create new materials and reactions, and optimize chemical processes (Zaw-Myo, *et al.,* 2023).

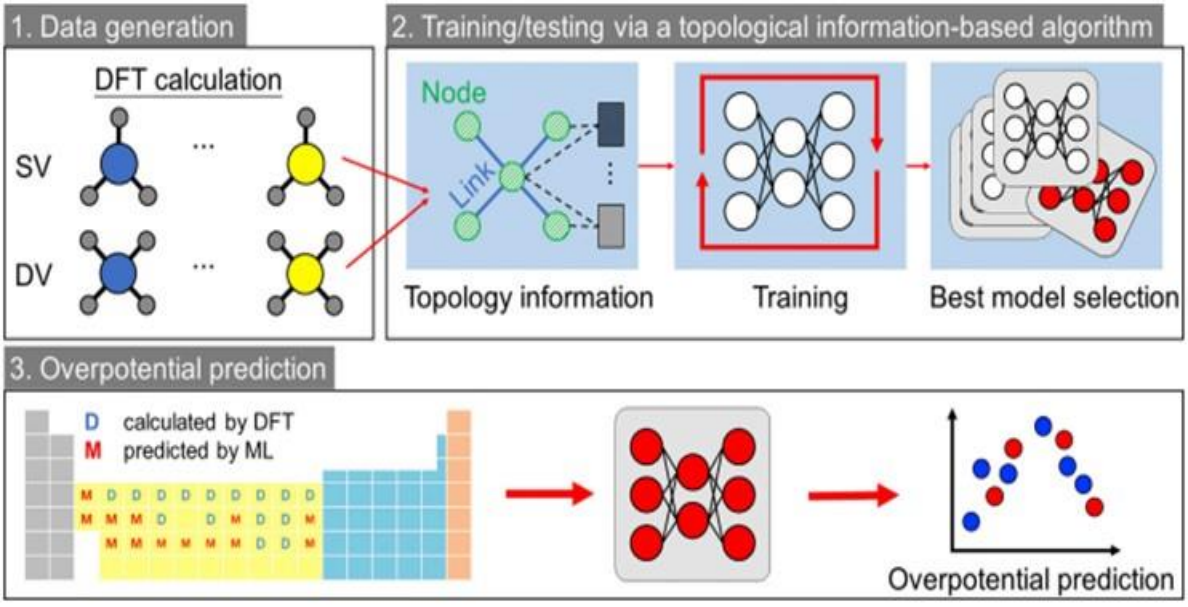

Fig. 3: diagram describing the machine learning training model (Zaw-Myo, *et al.*, 2023).

1.4.2 Quantitative structure-property relationship modelling for thermochemical data

A potent method for connecting the structural features of chemical compounds to their thermodynamic properties is quantitative structure-property relationship (QSPR) modelling. The foundation of QSPR models is the idea that a molecule's molecular structure and make-up dictate its physicochemical characteristics. QSPR models can precisely predict a variety of thermodynamic characteristics for novel chemical compounds using a dataset of molecular descriptors and associated thermodynamic property values (Weimin, *et al.,* 2023). A few pertinent examples of QSPR modelling for thermochemical data are provided below:

i. Prediction of Enthalpy of Formation: QSPR models can be created to predict the typical enthalpy of formation (H°f) of chemical compounds. The chemical characteristics (such as atomic composition, bond lengths, and electronegativity) and experimental H°f values for a variety of compounds are included in the dataset used to train the model. The QSPR model can predict the enthalpy of formation for novel compounds by learning the correlations between molecular characteristics and H°f (Etim, *et al.,* 2022).

ii. Theoretical Heat Capacity: The heat capacity (Cp) of chemical compounds at various temperatures can also be predicted using QSPR models. Molecular descriptors and experimentally determined heat capacity values for various substances over a variety of temperatures make up the dataset used to train the model (Minjie, *et al.,* 2022). The QSPR model then discovers the relationships between Cp and molecular characteristics.

iii. Boiling point prediction: The QSPR modelling technique can be used to estimate the boiling points of organic molecules. The dataset contains empirically determined boiling points and chemical characteristics for a wide range of substances. The QSPR model can predict the boiling points of novel compounds because it learns the correlations between molecular characteristics and boiling points (Guella, *et al.,* 2015).

iv. Prediction of Solubility: Foretelling a chemical compound's solubility in multiple solvents, QSPR models are useful. The chemical descriptors and experimentally determined solubility values for various substances in various solvents make up the dataset utilized for model training. The QSPR model can predict the solubility of novel compounds in certain solvents by learning the correlations between chemical characteristics and solubility (Tu, *et al.,* 2012).

v. Vapour Pressure Prediction: At various temperatures, chemical compounds' vapour pressures can be predicted using QSPR models. Molecular descriptors and experimentally determined vapour pressure values for different compounds are included in the dataset used to train the model. The QSPR model can calculate the vapour pressure for new compounds at various temperatures because it learns the relationships between molecular characteristics and vapour pressure (Weili, et al., 2010).

vi. Prediction for the Partition Coefficient (LogP): The partition coefficient (LogP), a metric of a compound's lipophilicity, can be predicted by QSPR models. The dataset comprises experimentally determined LogP values and chemical characteristics for a wide range of substances. The QSPR model can estimate the LogP values for novel compounds because it learns the correlations between molecular characteristics and LogP.

A useful approach for predicting different thermodynamic properties based on molecule structures and descriptors is QSPR modelling for thermochemical data. These models give scientists the tools they need to efficiently explore the chemical universe, create novel molecules with specified thermodynamic features, and enhance chemical reactions for a variety of uses (Weili, *et al.,* 2009). Thermochemistry research has made strides thanks to the effective method known as QSPR modelling, which has applications in many different areas of science and industry.

2.0 Machine learning in predicting reaction energies, enthalpies, and free energies

In thermochemistry, reaction energies, enthalpies, and free energies are all predicted using machine learning (ML) approaches. These ML models assess the energy changes brought on by chemical processes by utilizing molecular descriptors, reaction fingerprints, and other pertinent information. Examples of how machine learning can be used to predict reaction energies, enthalpies, and free energies include the following:

1. Reaction energy prediction: A collection containing experimentally determined or computed quantum chemical reaction energies can be used to train machine learning (ML) models such kernel ridge regression, random forests, or support vector regression. The dataset includes reaction fingerprints, related reaction energies, and molecular descriptors. The ML model can predict the energy changes

(E) of novel chemical reactions since it learns the correlations between the characteristics and reaction energies. For example, a dataset of reaction energies for a number of organic reactions is used to build a random forest regression model (Hong, *et al.,* 2007). The link between molecular descriptors, reaction fingerprints, and reaction energies is learned by the model. As a result, the trained model can predict the energy changes for novel reactions, assisting in the design and optimization of reactions.

2. Enthalpy Change Prediction: The enthalpy change (H) of chemical reactions can be predicted using ML approaches like gradient boosting or neural networks. For a wide range of reactions, the collection includes chemical descriptors, reaction fingerprints, and experimentally measured or DFT-calculated enthalpies (Palomba, *et al.,* 2012). The ML model can precisely estimate the enthalpy change for novel reactions after learning the patterns in the data. For example, a dataset of reaction energies for several organic reactions is used to build a random forest regression model. The link between molecular descriptors, reaction fingerprints, and reaction energies is learned by the model. As a result, the trained model can predict the energy changes for novel reactions, assisting in the design and optimization of reactions (Etim, *et al.,* 2022a).
3. Enthalpy Change Prediction: The enthalpy change (H) of chemical reactions can be predicted using ML approaches like gradient boosting or neural networks. For a wide range of reactions, the collection includes chemical descriptors, reaction fingerprints, and experimentally measured or DFT-calculated enthalpies. The ML model can precisely estimate the enthalpy change for novel reactions after learning the patterns in the data. As an illustration, a dataset of Gibbs free energy changes for a variety of chemical reactions is used to build a Gaussian process regression model. The link between molecular characteristics, reaction fingerprints, and Gibbs free energies is taught to the model. The model can then predict the changes in free energy for novel reactions, assisting thermodynamic analyses and reaction enhancement (Rabbani, *et al.,* 2021).

Thermochemical reaction energies, enthalpies, and free energies may all be predicted using machine learning. These ML models make it possible to develop new compounds and processes as well as to optimize reactions and efficiently explore chemical space. ML approaches help advance our understanding of how energy changes in chemical systems and spur innovation in a variety of scientific and industrial fields by utilizing pertinent features and training on a variety of datasets.

2.1 Database Development and Exploration

Building and improving thermochemical databases with machine learning can significantly increase their breadth, accuracy, and usability. For scientists, engineers, and researchers working in a variety of disciplines such as material science, medicine and energy, thermochemical databases are essential tools. Data curation, prediction, and optimization can be aided by ML approaches, creating better databases with insightful information (Menke, *et al.,* 2021). The following are some methods for developing and improving thermochemical databases using machine learning:

i. Data cleaning and Curation: Data on thermochemistry from multiple sources can be cleaned and curated with the use of machine learning methods. ML models can manage missing values, identify errors, and eliminate outliers from the data automatically, guaranteeing that the database contains accurate and trustworthy data (Edet and Samuel 2023).
ii. Data Fusion and Integration: ML approaches can be used to combine data from many sources and in a variety of formats. Utilizing information from many researches and combining experimental and theoretical data, data fusion techniques can ensure a more complete depiction of

thermochemical parameters (Maji, *et al.*, 2015).

iii. Data imputation and predictive modelling: Based on previously collected thermochemical data, machine learning models, such as random forests, neural networks, or Gaussian processes, can be trained to predict missing values or calculate the thermodynamic properties of novel substances or reactions. This predictive feature enables data imputation, expanding the database's coverage and filling up any gaps.

iv. Reaction and Property Prediction: For unstudied chemical systems, ML models can be created to predict reaction results and thermodynamic properties. ML can predict reaction energies, enthalpies, and free energies for a variety of chemical reactions by using molecular descriptors, reaction fingerprints, and other pertinent information, enhancing the database with useful predictions (Kong, and Yu, 2018).

v. Active Learning: In active learning contexts, machine learning can be utilized to direct the choice of fresh data points for verification through experimentation. Data collection is made more efficient and the amount of experimental work necessary for database improvement is decreased when ML models are used to determine which samples to include in the database are the most informative.

vi. Engineering features: ML methods can help in locating pertinent molecular descriptors and reaction fingerprints that significantly influence thermochemical characteristics. By creating more informative and discriminative features, feature engineering can improve the database's representation of chemical systems and give ML models more useful information (Kang, *et al.*, 2003).

vii. Design and optimization: Machine learning can be used to build novel compounds with desired thermodynamic properties and optimize chemical processes. Based on the available thermochemical data, ML algorithms can look for the best reaction conditions or materials, improving performance and efficiency.

viii. Estimating Uncertainty: To show how confidently the predictions were made, ML models can offer estimates of uncertainty for projected thermodynamic parameters. This uncertainty estimation improves the database's dependability and enables users to use the data to make wise judgments (Ertl, *et al.*, 2019).

Thermochemical databases can be improved and built with the aid of machine learning, allowing researchers to produce more thorough and trustworthy resources for the scientific community. These improved databases speed up and improve the accuracy of predictions, help with the identification of novel substances and reactions, and hasten technological advancement in some thermochemistry-related sectors (Rasmussen, 2004). One of the most important areas of thermochemistry study is the creation of machine learning prediction models for heat capacity, enthalpies of formation, and other thermodynamic parameters. The ability to precisely estimate thermodynamic parameters provided by ML approaches empowers researchers to investigate enormous chemical regions, enhance processes, and create novel materials. Let's see how these prediction models were created:

i. Collection of data and preprocessing: The collection of a comprehensive and top-notch thermochemical dataset is the initial step as shown in Fig. 4. For a variety of chemical substances and reactions, this dataset should include pertinent molecular descriptors, reaction fingerprints, and the appropriate thermodynamic property values (such as heat capacities, enthalpies of formation). By addressing missing values, eliminating outliers, and guaranteeing consistency in the representation of features, the data should be preprocessed (Pedregosa, *et al.*, 2012).

ii. Feature selection strategies aid in locating the most pertinent molecular descriptors and reaction fingerprints that have a major impact on the desired thermodynamic parameters. By developing new features or transformations, feature engineering helps ML models perform better and better at representing chemical systems (Suwarno, *et al.,* 2022).
iii. Model selection: Prediction models can be created using a variety of ML methods. Support vector machines (SVM), random forests, gradient boosting, and neural networks are examples of frequently used methods. The intricacy of the data and the specific thermodynamic property being predicted determine which model is used (Eqwuatu, *et al.,* 2023).
iv. Model training: The dataset is divided into training and testing sets. The ML model is trained using the training set, and its performance is assessed using the testing set. To reduce prediction errors, the ML model tunes its internal parameters during training to learn the correlations between the input features and the target thermodynamic properties (Sarkar, *et al.,* 2021).
v. Validation and evaluation of the model: The testing set is used to evaluate the trained model's generalization capability. The precision with which the model predicts thermodynamic properties is measured using a variety of metrics, including mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R-squared).
vi. Tuning of hyperparameters: To attain the optimum performance, hyperparameters in ML models frequently need to be tuned. The best settings for the model are found by exploring various combinations of hyperparameters using methods like grid search or random search (Finkelmann, *et al.,* 2016).
vii. Cross-Validation: Cross-validation methods, such as k-fold cross-validation, are used to confirm the model's robustness further. Cross-validation minimizes overfitting and aids in estimating the model's performance on new data.
viii. Model Use and Deployment: The thermodynamic properties of new chemicals or reactions can be predicted using the ML model after it has been created and validated. To acquire the estimated heat capacities, enthalpies of formation, or other thermodynamic parameters, users can enter molecular descriptors and reaction fingerprints (Etim, *et al.,* 2022b, 2022c).
ix. Model Interpretability: Depending on the application, efforts can be taken to understand the key factors impacting the thermodynamic properties and comprehend the ML model's predictions. Tools like feature importance analysis can shed light on the factors that underlie thermodynamic behaviour (Xiangyu, *et al.,* 2019).

These processes enable researchers to create machine-learning prediction models for heat capacities, enthalpies of formation, and other thermodynamic features. In numerous fields of science and industry, these models aid in the improvement of thermochemistry research, material discovery, and reaction optimization (Onen, *et al.,* 2017).

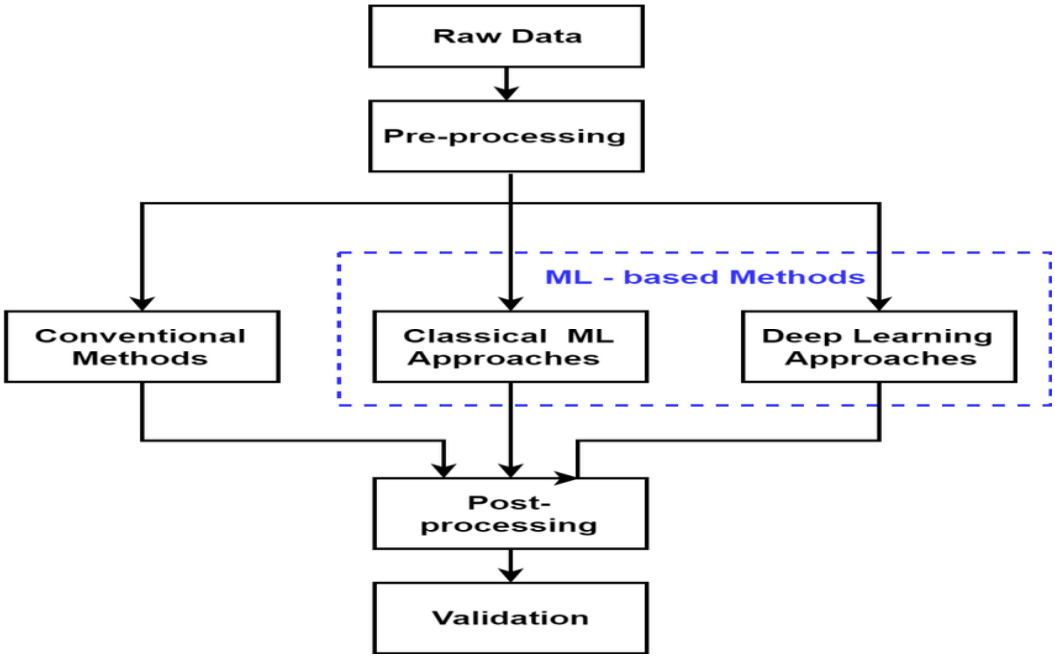

Fig. 4: Stages in machine learning prediction models (Xiangyu, *et al.*, 2019)

2.2 Integration of thermochemical databases for rapid access and analysis of thermochemical data

In terms of quick access to and interpretation of thermochemical data, integrating machine learning techniques into thermochemical databases offers several benefits. It improves the databases' usefulness and usability, making it possible for researchers to quickly access, investigate, and analyze chemical systems' thermodynamic properties (Stein,, *et al.*, 2019). We go into great detail about how machine learning methods are incorporated into thermochemical databases below.

i. Data cleaning and curation: The database's thermochemical data can be curated and cleaned using machine learning methods. These methods are capable of handling missing numbers, outlier removal, and error detection and correction automatically. By utilizing ML approaches, the database's quality and dependability are increased, guaranteeing that researchers have access to precise and reliable thermochemical data (Osigbemhe, *et al.*, 2022a; 2022b).

ii. Predictive Models for Missing Data: Machine learning models, including neural networks or regression, can be trained using the database's existing thermochemical data. The missing thermodynamic property values for compounds or reactions with little to no experimental data can then be estimated using these predictive models. The database is expanded and researchers have access to a wider variety of thermochemical data via imputed missing values (Onen, *et al.*, 2004).

iii. Fusion and Integration of Data: Data from many sources can be combined and integrated using machine learning into the thermochemical database. The seamless integration of experimental and computational data is made possible by ML approaches' ability to handle a variety of formats, data structures, and units. Data from many studies are combined to create a more thorough and varied database (Dai, *et al.*, 2019).

iv. Effective Search and Retrieval: Machine learning methods can enhance the database's search and retrieval capabilities. ML models can improve the retrieval of thermochemical data and prioritize relevant search results by learning from user interactions. This improves user convenience and cuts down on the time needed to get particular

thermodynamic parameters (Etim, *et al.*, 2016).

v. Predictability analysis: The database can use machine learning models to provide predictive analysis for different thermodynamic parameters. These models enable the quick screening of chemical systems and the study of chemical spaces by allowing researchers to predict thermodynamic values for new compounds or processes.

vi. Interactive Data Visualization: Machine learning methods can improve thermochemical database data visualization. Researchers can investigate the connections between chemical characteristics, reaction fingerprints, and thermodynamic variables using sophisticated ML-based visualization tools. Plots and visuals that can be interacted with promote data-driven insights and a deeper knowledge of the data (Jochen, *et al.*, 2019).

vii. Continuous learning and real-time updates: As fresh data becomes available, machine learning models in the database can continuously learn and update. Researchers can access the most recent thermochemical data because to the database's ability to stay current with the most recent experimental and computational data (Edet and Samuel, 2024).

viii. Active Learning and Data Suggestion: Within a database, active learning can be facilitated by machine learning. To increase data coverage and fill in any gaps in the database, the ML model can recommend further tests or calculations. The database can expand and improve its information more quickly by automatically choosing additional data points for validation (Robert, *et al.*, 2012).

ix. Computational tool integration: The computational tools used by researchers can easily incorporate machine learning models. During computer simulations, this integration permits on-the-fly predictions of thermodynamic parameters, expediting the process and granting quick access to pertinent thermochemical data.

2.3 Reaction network analysis and kinectics

2.3.1 Application of machine learning algorithms for reaction network analysis and prediction

In computational chemistry and chemical engineering, the use of machine learning algorithms for reaction network analysis and prediction is a potent strategy. Complex reaction networks may be understood, outcomes of reactions can be predicted, and reaction conditions can be optimized with the help of ML approaches (Shinggu, *et al.*, 2023). Here are a few significant uses of machine learning in this situation:

i. Reaction Pathway Prediction: To predict reaction pathways for particular starting materials, machine learning algorithms can be trained using reaction databases, such as graph-based models or recurrent neural networks. ML models can predict the most likely reaction sequences by examining the interactions between reactants, intermediates, and products, which helps with comprehending complex reaction mechanisms (Qianyi, *et al.*, 2009).

ii. Reaction Rate Prediction: For a variety of chemical reactions, ML approaches can be used to predict reaction rates and rate constants. Using experimental or quantum chemical data, models can be trained to predict reaction rates at various temperatures and pressures. Understanding reaction kinetics and improving reaction conditions depend on accurate rate estimates (Ekpan *et al.*, 2024).

iii. Multi-Step Reaction Product Prediction: Given the beginning ingredients and reaction conditions, machine learning models, particularly sequence-to-sequence models or transformer-based architectures, can predict the end products of multi-step reactions. Accurate product prediction is made possible by these models, which make

use of sizable reaction datasets and learn the transitions between chemical structures (John & Ryszard 2009).

iv. Reaction Optimization: To optimize reaction conditions for desired results, ML algorithms like Bayesian optimization or reinforcement learning can be used. These models can direct the search for the best reaction parameters, such as temperature, pressure, and catalysts, to produce improved yields and selectivity by learning from experimental data or quantum chemical simulations (Yang, 2020).

v. Designing a catalyst: By foreseeing the activity and selectivity of catalysts for particular processes, machine learning approaches can aid in the design of catalysts. To find potential candidates for catalytic processes, models can be trained on data from catalytic reactions and descriptors of catalyst qualities, which eliminates the need for laborious experimental screening.

vi. Mechanism Clarification: By using experimental data, ML models can help clarify response mechanisms. ML algorithms may infer possible reaction paths and mechanistic stages by examining kinetic data, reaction intermediates, and product distributions, offering insights into the underlying chemistry (Jiménez-Martínez, *et al.*, 2022).

vii. Prediction of Adverse Reactions and By-products: The occurrence of side effects and by-products in chemical reactions can be predicted by machine learning. These predictions assist chemists in locating probable impurities and formulating plans to reduce unfavourable side effects during synthesis (Francis-Dominic *et al.*, 2024).

viii. Enantioselectivity Prediction: The enantioselectivity of chiral catalysts or processes can be predicted using ML models. ML algorithms can calculate the enantiomeric excess of products by learning from data on chiral ligands and substrates, aiding in the development of asymmetric processes (Yaroslava, *et al.*, 2014)

ix. Biochemical Pathway Analysis: ML algorithms in biochemistry are capable of analyzing intricate biochemical pathways like metabolic networks or enzyme-catalyzed reactions. These models aid in understanding cellular functions and foretelling how biological systems would behave (Saeed, *et al.*, 2013).

Research in chemistry and chemical engineering can be sped up by using machine learning techniques for reaction network analysis and prediction. These models increase the effectiveness of reaction design, optimization, and mechanism elucidation, allowing scientists to decide more wisely and effectively explore the chemical space.

2.3.2 Machine learning in determining reaction kinetics, rate constants, and activation energies

The determination of reaction kinetics, rate constants, and activation energies in chemical reactions has shown tremendous promise when using machine learning (ML). Understanding reaction pathways, improving reaction conditions, and projecting reaction results all depend on these thermodynamic features. To create precise predictions, ML approaches can use experimental data, quantum chemical computations, and molecular descriptors. Here is how ML is used in this situation:

i. Prediction for Reaction Rate: To predict reaction rates for novel reactions or situations, machine learning models can be trained on experimental reaction rate data. To estimate the reaction rate, these models take as inputs molecular descriptors, reaction circumstances, and other pertinent information. For example, regression models like random forest or neural networks are frequently employed in predicting the reaction rate in 3D porous media as shown in figure 5.0 (Jahan, *et al.*, 2013).

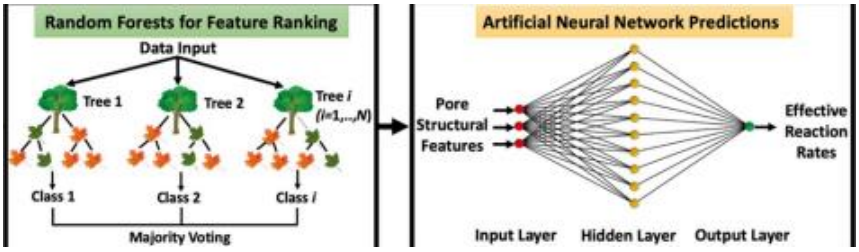

Fig.5: schematic of the ML-based framework of 3D pore-scale Random Forests (RF), and Artifcial Neural Network (ANN) models (Jahan, *et al.*, 2013)

ii. Rate Constant Estimation: By analyzing experimental kinetic data, ML can calculate rate constants for chemical reactions. ML models can predict rate constants for new reactions or situations by learning from a wide set of reactions and their rate constants. The effectiveness and quickness of the reaction must be determined using this information (Mani, *et al.*, 2012).

iii. Predicted Activation Energy: A crucial thermodynamic factor that affects a reaction's pace is activation energy. To estimate activation energies, machine learning (ML) models can be trained on data from Arrhenius plots, which connect reaction rates to temperature. These models can be used to predict activation energies for various reactions and reveal information about how temperature affects the reaction. The highest-energy regions along the chemical route are called transition states, and they can be predicted using machine learning (ML) approaches. Understanding reaction mechanisms requires the prediction of transition states. Quantum chemical data can be used to train machine learning models to identify the structural characteristics and energy landscapes connected to transition states (Mani, *et al.*, 2013).

iv. Analysis of Reaction Mechanisms: ML models are capable of deciphering the workings of chemical reactions and pinpointing their crucial steps. These models can identify chemical intermediates and transition states by using graph-based models or recurrent neural networks, giving important information on reaction pathways.

v. Kinetic Isotope Effect Prediction: By learning from experimental data including isotopically labelled reactants, machine learning (ML) can predict kinetic isotope effects (KIEs). KIEs offer details on the reaction mechanism and rate-determining step. Mechanistic studies are aided by the ability of ML models to estimate KIEs for various isotope substitutions (Tetyana, *et al.*, 2011).

vi. Rate Constant QSPR Modelling: Chemical reaction rate constants can be predicted using quantitative structure-property relationship (QSPR) models. These models estimate rate constants for new reactions by establishing links between the reaction's structure and its rate constant using molecular descriptors.

2.4 Identification of reaction trends using machine learning

In big datasets of chemical reactions, machine learning can be used to find and analyze reaction trends. These patterns offer insightful information about the variables affecting reaction outcomes, selectivities, and efficiency (Nikolay *et al.*, 2011). To aid researchers in making decisions and directing experimental and computational efforts, machine learning (ML) models can reveal patterns and correlations between reaction circumstances, chemical structures, and thermodynamic parameters. How machine learning is used to spot reaction tendencies is as follows:

i. Classification of Reactions: Reactions can be categorized using machine

learning models into many groups based on their types, such as substitution, addition, or elimination reactions. ML models may accurately categorize new reactions by learning the characteristics that distinguish various response classes through training on a labelled dataset (Anikina, *et al.,* 2011). Structure-Activity Relationships (SAR) can be established between reactants and reaction outcomes using machine learning approaches like regression models or neural networks. These models examine molecular descriptors and reaction fingerprints to pinpoint structural characteristics that result in the production of desired products or certain reaction pathways (Egwuatu *et al.,* 2024).

ii. Design and screening of catalysts: In the creation and screening of catalysts for particular reactions, machine learning is useful. The links between catalyst characteristics, reaction circumstances, and reaction results can be examined using ML models. ML algorithms can find potential catalyst candidates with desired selectivities or activity by learning from a variety of catalytic processes (Semenov, *et al.,* 2011).

iii. Solvent Effects and Reaction Conditions: ML may show how the characteristics of the solvent and the circumstances of the reaction affect the results of the reaction. ML models can predict how these factors will affect reaction rates, selectivity, and yields by training on reaction data under various solvent and condition circumstances.

iv. Predicting outcomes of Reactions: Based on reactant structures and reaction conditions, machine learning algorithms can predict the products or outcomes of chemical reactions. Understanding the reactivity and selectivity of various functional groups and reaction types is made easier thanks to these predictions (Mohammad, *et al.,* 2010).

v. Mechanistic Understandings: Mechanistic insights into chemical reactions can be obtained using ML algorithms, particularly graph-based models or understandable models like decision trees. ML models can offer insight into the crucial steps and intermediates in reaction pathways by assessing the significance of various molecular characteristics (Mani, *et al.,* 2010).

2.5 Advancements in machine learning for thermochemistry

Thermochemistry research has greatly benefited from developments in machine learning, which have helped predict thermodynamic parameters, comprehend reaction pathways, and optimize chemical processes. Some significant developments include:

i. Multiple datasets with big data: Large, varied, and carefully curated databases are now readily available, which has improved machine learning in thermochemistry. A large range of chemical compounds and processes are represented in these databases, allowing ML models to gain knowledge from a variety of thermodynamic features and trends (Etim, *et al.,* 2017a; 2017b).

ii. Deep learning architectures: Many fields of science, including thermochemistry, have seen radical change as a result of the rise of deep learning. Convolutional neural networks (CNNs) and transformer-based models are two examples of deep learning architectures that have impressively shown the ability to predict chemical characteristics and reaction outcomes more accurately and effectively (Etim, *et al.,* 2021).

iii. Reactive Learning: ML models can use transfer learning approaches to use knowledge from one domain to perform better in a related area. Transfer

learning has been used in thermochemistry research to share knowledge among many chemical systems, hence lowering the demand for vast volumes of data for particular tasks.

iv. Active Learning: ML models can actively choose the most instructive data points for training thanks to active learning methodologies, which optimize the learning process. Active learning has been used in thermochemistry to direct experiments or quantum chemical computations, lowering the amount of experimental work necessary to improve the database.

v. Estimating uncertainty: Because computational calculations might be inaccurate and experimental data can be noisy, it is essential to quantify uncertainty in thermochemistry predictions. To increase the trustworthiness of outcomes, advanced ML techniques like Bayesian neural networks and Gaussian processes have been used to provide uncertainty estimates alongside predictions.

vi. Reasonable AI: In scientific research, ML models' interpretability is crucial. Researchers in thermochemistry are increasingly employing explainable AI methods to determine why a model predicts certain outcomes (Onen, *et al.*, 2017).

vii. Identification of Active Sites: Understanding reactivity and binding interactions in catalysis and drug design requires accurate identification of the active sites on catalysts or proteins. o build effective catalysts and drug candidates, ML models have been used to predict active sites based on chemical characteristics.

viii. Screening with high throughput: High-throughput screening methods have combined machine learning to quickly explore broad chemical space and find interesting candidates for particular thermodynamic features or reaction outcomes.

ix. Precision Machine Learning: Complex thermochemistry problems can be solved and quantum systems can be simulated by utilizing quantum machine learning techniques, such as quantum neural networks and vibrational quantum circuits (Etim, et al., 2017).

3.0 Conclusion

The analysis, prediction, and comprehension of thermodynamic parameters and chemical reactions by researchers have been revolutionized by machine learning (ML), which has emerged as a transformational force in thermochemistry. The blending of data-driven methodologies with conventional theoretical and experimental approaches has created new opportunities for chemical exploration, material discovery, reaction optimization, and the development of a profound understanding of complicated molecular interactions. Numerous uses of ML in thermochemistry have had a substantial impact on many areas of research, and as new, more sophisticated methods are created, their potential keeps expanding. The capacity of machine learning to predict thermodynamic parameters with high accuracy is one of the most important contributions of ML in thermochemistry. For a variety of chemical systems, ML models can estimate enthalpies, free energies, heat capacities, and other thermodynamic parameters. These models range from conventional regression approaches to cutting-edge deep learning architectures. These predicting tools are priceless for researchers because they give quick access to thermochemical data, particularly for substances with scant experimental evidence or when in-depth quantum chemical computations might not be possible. Additionally, thermochemical databases have grown and improved because to ML approaches. The reliability and completeness of thermochemical data repositories have increased thanks to machine learning (ML) through data integration,

cleaning, and imputation. Experimental efforts have been directed by active learning methodologies, which have effectively filled in data gaps and fueled the ongoing growth of these databases. The efficiency and precision of computational studies have been improved by the real-time thermodynamic property predictions made possible by the integration of ML-driven tools with computational software during simulations. Explainable AI methods have made ML predictions transparent, allowing researchers to decipher and believe in the underlying mechanisms generating the results. This interpretability is essential for directing additional experimental validations, discovering novel chemical insights, and encouraging multidisciplinary cooperation between ML specialists and domain-specific thermochemistry researchers. The difficulties in integrating and scaling these approaches to handle more challenging thermochemistry problems increase as machine learning (ML) continues to develop. Research is still being done on how to handle scarce or noisy data, properly handle ML models across a variety of chemical domains, and measure uncertainty in predictions.

A new era of chemistry research and discovery has begun because of the combination of machine learning and thermochemistry. Our knowledge of chemical phenomena has increased as a result of the interaction between data-driven methods and conventional techniques. This has also sped up the creation of novel materials and paved the way for environmentally friendly and effective chemical processes. The subject of thermochemistry stands to gain from transformational improvements as researchers continue to harness the power of ML, stimulating innovation and propelling development across numerous scientific and industrial applications. As machine learning solidifies its position as a crucial tool for directing the course of thermochemistry research, the future is filled with promise.

4.0 References

Agúndez, M., Cernicharo, J., & de Vicente P. (2015). Discovery of HC3O+ in space: The chemistry of O-bearing species in TMC-1. *Astronomy & Astrophysics*, 642, doi:10.1051/0004-6361/201526650.

Andrew, C., Etim E. E, Ushie, O. A. & Khanal. G. P. (2018). Vibrational-Rotational Spectra of Normal Acetylene and Doubly Deuterated Acetylene: Experimental and Computational Studies. *Chemical Science Transactions, 7, 1, pp.* 77-82. DOI:10.7598/cst2018.1432.

Angulo, A., Yang, L., Aydil, E.,& Modestino, M. A. (2022). Machine learning enhanced spectroscopic analysis: towards autonomous chemical mixture characterization for rapid process optimization. *Digital Discovery.*, 1, 1, pp. 5-44. doi:10. -1039/D1DD00027F

Anikina, O. Ya. Krivuschenko, D. V. Schur, S. Yu. Zaginaichenko, E. A. & Kamenetskaia. (2011). *Special features and regularities of interaction between fullerene molecules and aromatic solvents*. Pp. 53-74. https://doi.org/10.1007/978-94-007-0899-0_4

Baskes, M. I. (1997). Determination of modified embedded atom method parameters for nickel. *Mater. Chem. Phys.* 50, pp. 152–158.

Behler, J. (2017). First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed*. 56, pp. 12828–12840.

Behler, J. & Parrinello, M. (2007). Generalized neural-network represent - ation of highdimensional potential-energy surfaces. *Phys. Rev. Lett.* 98, 146401, https://doi.org/10.1103/PhysRevLett.98.146401

Behn, A., Zimmerman, P.M., Bell, A.T., & HeadGordon, M. (2011). Efficient exploration of reaction paths via a

freezing string method. *J. Chem. Phys*. 135, 224108, doi: 10.1063/1.3664901.

Chen, Z., Alexandre, M., Weihua, Li. & Konstantinos G,. (2020). A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks." Mechanical Systems and Signal Processing, 140, 106683. https://doi.org/10.1016/j.ymssp.2020.106683.

Chowdhury, M, Rice, T. & Oehlschlaeger, M. A. (2021). Evaluation of machine learning methods for classification of rotational absorption spectra for gases in the 220–330 GHz range. *Appl. Phys.* B 127, 34 https://doi.org/10.1007/s00340-021-07582-0.

Ding, T., Readshaw, T., Rigopoulos, S. & Jones, W P. (2021). Machine learning tabulation of thermochemistry in turbulent combustion: An approach based on hybrid flamelet/random data and multiple multilayer perceptrons. *Combustion and Flame.* 231, 111493. doi:10.1016/j.combustflame.2021.111493

Edet, P. I., & Samuel, H. S. (2023). A review of antioxidant applications and phytochemical constituents of Anacardium Occidentale leaf extract. Faculty of Natural and Applied Sciences *Journal of Scientific Innovations*, 5, 1, pp. 15-20. https://www.fnasjournals.com/index.php/FNAS-JSI/article/view/198

Egwuatu, O .P, Ori, M. O, Samuel, H.S., & Ekpan, F. M. (2024). AI-enabled Diagnostics and Monitoring in Nanomedicine. *Eurasian Journal of Science and Technology*, 4(3), 208-229. doi: 10.48309/ejst.2024.426725.1116

Ekpan, F. M., Ori, M. O, Samuel, H.S., & Egwuatu, O. P. (2024). Emerging technologies for eco-friendly production of bioethanol from lignocellulosic waste materials. *Eurasian Journal of Science and Technology*, (), 179-194. doi: 10.48309/ejst.2024.429106.1119

Ertl P. (2019). An algorithm to identify functional groups in organic molecules. *J Cheminform* 9, 1, pp. 1-7. https://doi.org/10.1186/s13321-017-0225-z

Etim, E. E., Andrew, C., Lawal, U., Udegbunam, I. Etiowo G. (2020). Protonation of Carbonyl Sulfide: *Ab initio* Study. *Journal of Applied Sciences*, 20, pp. : 26-34. doi: 10.3923/jas.2020.26.34

Etim, E. E., Asuquo, J. E., Ngana, O.C., & Ogofotha, G. O. (2022). Investigation on the thermochemistry, molecular spectroscopy and structural parameters of pyrrole and its isomers: a quantum chemistry approach. *J. Chem. Soc*. Nigeria, 47, 1, pp. 129 - 138.

Etim, E. E., Gorai, P., Das, A., Chakrabarti, S. & Arunan, E. (2018). Interstellar Hydrogen Bonding. *Advances in Space Research*, 61, 11, pp. 2870-2880, https://doi.org/10.1016/j.asr.2018. 03.003.

Etim, E. E., Oko, E. G., Sulaiman, A. O. (2020). Protonation in Noble Gas Containing Molecular Systems: Observing Periodic Trends in CF_3Cl, CF_3Br, CH_3F, CH_3Cl. *International Journal of Advanced Research in Physical Science (IJARPS)* 7(6): 14-19

Etim, E.E., Ugo Nweke-Maraizu., Samuel, H.S, (2023). A Review of Theoretical Techniques in Corrosion Inhibition Studies. Communication in Physical Sciences, 9, 4, pp. 394-403

Etim, E.E., Ashu, H. A., Mbakara, I.E., Inyang, E. J., Ukafia, O. P., Sambo, I. F. (2017). Quantum Chemical Calculations on Oxygen Monofluoride (OF) and its Protonated Analogues: Comparison of Methods. *Elixir Computational Chemistry*, 111, pp. 3-48827.

Finkelmann, A. R, Göller, H., Schneider, G. (2016). Robust molecular representations for modelling and

design derived from atomic partial charges. *Chem Commun* 52, pp. 681–684. https://doi.org/10.1039/c5cc07887c

Francis-Dominic, M., Ekpan, M. O., Samuel, H. S. & Egwuatu, O. P. (2024). The synergy of AI and Drug delivery: A Revolution in Healthcare. International *Journal of Advanced Biological and Biomedical Research*, 12, 1, pp. 45-67. https://www.ijabbr.com/article_709788.html

Giambagli, L., Buffoni, L. & Carletti, T. (2021). Machine learning in spectral domain. *Nat Commun* 12, 1330**,** https://doi.org/10.1038/s41467-021-21481-0

Guella, S., Argoub, K. & Benkouider, A. M. (2015). Artificial Neural Network-Group Contribution Method for Predicting Standard Enthalpy of Formation in the Solid State: C–H, C–H–O, C–H–N, and C–H–N–O Compounds. *Int J Thermophys* 36, pp. 2820–2832 https://doi.org/10.1007/s-10765-015-1928-x

Gúndez, M., Marcelino, N. & Cernicharo, J. (2018). Tentative detection of HC5NH+ in TMC-1. Astronomy & Astrophysics. 861,. doi:10.1051/0004-6361/201833657.

Hirota, T., Ito, T. & Yamamoto, S. (2002). A Study of the Physical and Chemical Properties of the Quiescent Cores L1521B and L1521E. *The Astrophysical Journal*. 565, pp. 359-372. doi:10.1086/324588.

Hong, Huo., Ngai, T. & Suat, H. (2007). Self-Organization of Double-C60 End-Capped Poly(ethylene oxide) in Chloronaphthalene and Benzene Solvent Mixtures. Langmuir 23, 24, pp. , 12067-12070. https://doi.org/10.1021/la701762e

Jahan, B., Ghasemi, M. & Rofouei, M. (2013). Alignment Independent 3D-QSAR Modeling of Fullerene (C 60) Solubility in Different Organic Solvents. Fullerenes, *Nanotubes and Carbon Nanostructures* 21, 5, pp. 367-380. https://doi.org/10.1080/1536383X.2011.629751

Janet, J., Kulik, H., Morency, Y., Caucci, M. (2020). *Machine Learning in Chemistry. ACS* In Focus (Washington, DC: American Chemical Society).

Jia, F., Lei, Y., Guo, L., Lin, J. & Xing. S. (2018). A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neuro -computing.* 272, pp. 619-628. doi:10.1016/j.neucom.2017.07.032

Jiménez-Martínez, J. (2020). Homogenization of dissolution and enhanced precipitation induced by bubbles in multiphase fow systems. *Geophys. Res. Lett.* https://doi.org/10.1029/2020GL087163

Jochen, S., Florian, F. & Matthias, R. (2019). in need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. *Journal of Chemical Information and Modeling* 59, 3, pp. 947-961. https://doi.org/10.1021/acs.jcim.8b00712

John, M. & Ryszard, C. (2008). SAMFA: Simplifying Molecular Description for 3D-QSAR. *Journal of Chemical Information and Modeling*, 48, 6, pp. 1167-173. https://doi.org/10.1021/ci80-0 009u

Kang, Q., Zhang, D. & Chen, S. (2003). Simulation of dissolution and precipitation in porous media. J. Geophys. *Res. Solid Earth.* https://doi.org/10.1029/2003JB002504

Khan, M., Naqvi, S., Ullah, Z., Taqvi, S., Khan, M., Farooq, W., Mehran, M. & Juchelková Štěpanec L. (2023). Applications of machine learning in thermochemical conversion of biomass-A review. *Fuel.* 332(Part 1):126055. doi:10.1016/j.fuel.2022. -126055

Komp, E., Janulaitis, N. & Valleau, S. (2022). Progress towards machine learning reaction rate constants.

Physical Chemistry Chemical Physics. https://doi.org/10.1039/D1CP04422B

Kong, Y., & Yu, T. (2018)/ A Deep Neural network model using random forest to extract feature representation for gene expression data classifcation. *Sci. Rep.* https://doi.org/10.1038/s41598-018-34833-6

Kuz'min, V., Artemenko, A. & Ognichenko, L. (2021). Simplex representation of molecular structure as universal QSAR/QSPR tool. Strutural Chemistry, 32, pp. 1365–1392 https://doi.org/10.1007/s11224-021-01793-z

Lee, K., Loomis, R. & Burkhardt AM. (2021). Discovery of Interstellar trans-cyanovinylacetylene (HC # CCH = CHC # N) and vinylcyanoacetylene (H 2 C = CHC 3 N) in GOTHAM Observations of TMC-1. *Astrophysical Journal Letters.* 908:L11. doi:10.3847/2041-8213/abdbb9.

Leonard, T., Chun, C. & Wilson, W. (2023). Emotional Variance Analysis: A new sentiment analysis feature set for Artificial Intelligence and Machine Learning applications, *PLOS ONE*, 18, 1, (e0274299).https://doi.org/10.1371/journal.pone.0274299

Longqiang, L., Zhou, L., Guixia, L., Yun, T. & Weihua, L. (2023). Machine Learning Models to Predict Cytochrome P450 2B6 Inhibitors and Substrates. *Chemical Research in Toxicology* Article ASAP.

Madden, M. & Ryder A. (2002). Machine learning methods for quantitative analysis of Raman Spectroscopy data. *In Proceedings of SPIE*, 4876, pp. 1013-1019

Maji, D., Santara, A., Ghosh, S., Sheet, D. & Mitra, P. (2015). Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) IEEE, https://doi.org/10.1109/EMBC.2015.7319030.

Mani, S.i, Hamid, M. & Mohsen, M. (2010). Modeling the hydrogen solubility in methanol, ethanol, 1-propanol and 1-butanol. *Fluid Phase Equilibria* 289 , 1, pp. 32-39. https://doi.org/10.1016/j.fluid.2009.10.012

Mani, S.i, Hamid, M. & Mohsen, M. (2012). Application of neural network molecular modeling for correlating and predicting Henry's law constants of gases in [bmim][PF6] at low pressures. *Fluid Phase Equilibria* 332 , 165-172. https://doi.org/10.1016/j.fluid.2012.07.009

Mani, S. I., Hamid, M. & Mohsen, M. (2012).Correlating and predicting low pressure solubility of gases in [bmim][BF4] by neural network molecular modeling. *Thermochimica Acta*, 545 , pp. 125-130. https://doi.org/ 10.1016/j.tca.2012.07.005

Mattioda, A., Hudgins, D. & Boersma, C. (2020). The NASA Ames PAH IR Spectroscopic Database: The 2019 Release. *Astrophysical Journal Supplement Series.* 25, doi:10.3847/15 -38 -4365/abb3db.

McGuire B. (2018). 2018 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules. *Astrophys J Suppl Ser*. 239(, 1, 17, doi:10.3847/1538-4365/ -aae5d2

Menke, P., Maes, J. & Geiger, S. (2021). Upscaling the porosity–permeability relationship of a microporous carbonate for Darcy-scale fow with machine learning. *Sci. Rep.* https://doi.org/10.1038/s41598-021-82029-2

Minjie, M., Ziqi, P., Mingkun, L., Huaicheng, S., Yunxia, W., Yongchao, L. & Feng, Z. (2022). Application of Machine Learning in Spatial Proteomics. *Journal of Chemical Information and Modeling*, 6, 23, pp.

5875-5895. https://doi.org/10.1021/acs.jcim.2c01161

Mohammad, G., Pablo, R., Duchowicz, P., Freitas, F. & Fernández, F. (2010). Prediction of the Hildebrand parameter of various solvents using linear and nonlinear approaches. *Fluid Phase Equilibria*, 2932, 2, pp. 130-136. https://doi.org/10.1016/j.fluid.2010.02.025

Neumann, J., Christoph, S., Gabriele, C., Schn¨orr & Steidl, G. (2005). "Combined SVMBased Feature Selection and Classification," Machine Learning, vol. 61, no. 1-3, pp. 129– 150, issn: 0885-6125. doi: 10.1007/s10994-005-1505-9.

Nikolay, O. & Mchedlov-Petrossyan. (2011). Fullerenes in molecular liquids. Solutions in "good" solvents: Another view. *Journal of Molecular Liquids* 161, 1, pp. 1-12. https://doi.org/10.1016/j.molliq.2011.04.001

Oliveira, J., Frey, J., Zhang, S., Xu, L., Li, X., Li, S., Hong, X. & Ackermann L. (2022). When machine learning meets molecular synthesis. Trends in Chemistry. 4, 10, pp. 863-885. doi:10.1016/j.trechm.2022.07.005.

Onen, A. I., Joseph, J., Etim, E. E. & Eddy, N. O.(2017). Quantum Chemical Studies on the Inhibition Mechanism of Ficus carica, FC and Vitellaria paradoxa, VP Leaf Extracts. *Journal of Advanced Chemical Sciences*, 3(3):496-498. http://jacsdirectory.com/ journal-of-advanced-chemical-sciences/articleview.php?id=155

Ori, M. O., Edet, P. I., Ekpan, F. M., Samuel, H. S., Egwuatu, O. P. & Ajor, E. J. (2024). Revisiting on Applications of Industrial Filters in Enhancing Polymer Product Quality and Performance. *Eurasian Journal of Science and Technology*, 4, 2, pp. 116-132. doi: 10.48309/ejst.2024.423429.1107

Osigbemhe, I. G., Emmanuella, E. O., Hitler L., Emmanuel, M. K., Emmanuel E. E., Henry O., .Edet, Onyinye J. I., Amoawe P. O. & Obuye, F. . (2022). Antibacterial potential of N-(2-furylmethylidene)-1, 3, 4-thiadiazole-2-amine: Experimental and theoretical investigations. *Journal of the Indian Chemical Society*, 99, 9, 100597. https://www.sciencedirect.com/science/article/abs/pii/S0019452222000259X

Osigbemhe, I.G., Louis, H., Khan, E.M., Etim, E. E., Odey, D. O., Oviawe, A. P., Edet, H. O. & Obuye, F. (2022a). Synthesis, characterization, DFT studies, and molecular modeling of 2-(-(2-hydroxy-5-methoxyphenyl)-methylidene)-amino) nicotinic acid against some selected bacterial receptors. *J Iran Chem. Soc*, https://doi.org/10.1007/s13738-022-02550-7

Palomba, D., Martínez, M., Ponzoni, I., Díaz, M. F., Vazquez, G. E. & Soto, A. J. (2012). QSPR Models for Predicting Log Pliver Values for Volatile Organic Compounds Combining Statistical Methods and Domain Knowledge. Molecules 17, pp. 14937-14953.https://doi.org/10.3390/molecules171214937

Pedregosa F., Varoquaux G. & Gramfort (2012). A Scikit-learn: machine Learning in Python. *J Mach Learn Res* 12:2825–2830

Qianyi Z., Jacqueline M., Hughes-Oliver & Raymond T. (2009). A Model-Based Ensembling Approach for Developing QSARs. *Journal of Chemical Information and Modeling*, 49, 8, pp. 1857-1865. https://doi.org/10.1021/ci900080f

Rabbani, A. (2021). Review of data science trends and issues in porous media research with a focus on image-based techniques. *Water Resour. Res.* https://doi.org/10.1029/2020WR029472

Rasmussen C., (2004). Gaussian Processes in Machine Learning. In: Bousquet O, von Luxburg U, Rätsch G (eds)

Advanced Lectures on Machine Learning: ML Summer Schools *Springer*, pp 63–71 Berlin Heidelberg,

Robert P., (2012). Three Useful Dimensions for Domain Applicability in QSAR Models Using Random Forest. *Journal of Chemical Information and Modeling 52, 3, pp.* 814-823. https:// doi. org/10.1021/ci300004n

Samuel, H. S., E. E. Etim, U. & Nweke-Maraizu. (2023). Approaches for Special Characteristics of Chalcogen Bonding: A mini Review. *J. Appl. Organomet. Chem*. 3(3), 199-212. https://doi.org/10.22034/jaoc.2023.405432.1089

Samuel, H. S., Etim, E. E., Nweke-Maraizu, U. & Andrew, C. (2023). Computational Electrochemistry Techniques Used In Corrosion Inhibition Studies. *FUW Trends in Science & Technology Journal, 8, 3, pp.* 033-039 https://www.ftstjournal.com/uploads/docs/83%20Article%207.pdf

Samuel, H. S., Etim, E. E., Oladimeji E.O., Shinggu J. P. & Bako B. (2023). Machine Learning in Characterizing Dipole-Dipole Interactions. *FUW Trends in Science & Technology Journal*, 8, 3, pp. 070-082

Samuel, H.S. U. Nweke-Maraizu, E.E. Etim (2023). Supercritical Fluids: Properties, Formation and Applications. J. Eng. Ind. Res. 4, 3, pp. 176-188. https://doi.org/10 .48309/jeires.2023.3.5

Samuel, H. S., Nweke-Maraizu, U., Johnson, G., & Etim, E. E. (2023). nonelectrochemical techniques in corrosion inhibition studies: analytical techniques. *Communication in Physical Sciences*, 9(3): 383-393

Samuel, H. S., Nweke-Maraizu, U. & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. Faculty of Natural and Applied Sciences Journal of Scientific Innovations, 5, 1, pp. 103–115. https:// www.fnasjournals.com/index.php/FNAS-JSI/article/view/208

Samuel, H. S., U. Nweke-Mariazu, E. & Etim., E. (2023). Experimental and Theoretical Approaches for Characterizing Halogen Bonding. *J. Appl. Organomet. Chem.*, 3, 3, pp. 169-183. https://doi.org/10.22034/jaoc.2023.405412.1088

Samuel, H. S., Etim, E. E. & Ugo Nweke-Maraizu., (2023). Understanding the experimental and computational approach in characterizing intermolecular and intramolecular hydrogen bond, *Journal of Chemical Review*, https://doi.org/10.48309/JCR.2023.407989.1235

Samuel, H. S., Etim, E. E., Ugo Nweke-Maraizu., Shinggu, J. P. & Bako B (2023). Machine learning of rotational spectra analysis in interstellar medium. *Communication in Physical Sciences*, 10, 1, pp. 172-203.

Sarkar, S., Chakraborty, S. & Das S. (2021). Machine learning enabled quantification of the hydrogen bonds inside the polyelectrolyte brush layer probed using all-atom molecular dynamics simulations. *J Chem Phys*. 155, 14, 144902. doi:10.1063/5.00626-59

Semenov, N., Charykov, V., Postnov, O. & Krokhina, E. G. (2010). Gruzinskaya. isothermal solubility of individual light fullerenes in the homologous series of n-alkanes, n-alkanols, n-alkylcarboxylic acids, and arenes. Russian *Journal of General Chemistry* 80 (12), 2443-2449. https://doi.org/10.1134/S1070363210120078

Stein, H. S., Guevarra, D., Newhouse, P. F., Soedarmadjia, E. & Gregoire, J. M. (2019). Machine learning of optical properties of materials – predicting spectra from images and images from spectra. *Chemical Science*. https://doi.org/10.1039/C8SC03077D

Suwarno, S., Dicky, G., Suyuthi, A., Effendi, M., Witantyo, W., Noerochim, L., and Ismail M. (2022). Machine learning analysis of alloying element effects on hydrogen storage properties of AB2 metal hydrides. *Int J Hydrogen Energy*. 47, 23, pp. 11938-11947. doi:10.1016/j.ijhydene.2022.01.210

Tessarini, S., Fix, M. K. & Manser, P. (2022). Semi-classical Monte Carlo algorithm for the simulation of X-ray grating interferometry. *Sci Rep* 12, 2485 https://doi.org/10.1038/s41598-022-05965-7.

Tu, L., Chandana E., Frank R., Burden, & David A. (2012). Water quantitative structure–property relationship modeling of diverse materials properties. *Chemical Reviews*, 112, 5, pp. 2889-2919. https://doi.org/10.1021/cr200066h

Ushie, O. A., Etim, E. E., Onen, A. I., Andrew, C., Lawal, U. & Khanal, G.P. (2019). Computational Studies of β-amyrin acetate (C32H52O2) Detected in Methanol Leaf Extract of *Chrysophyllum albidium. J. Chem Soc. Nigeria,* 4, 3, , pp 561 -581.

Ushie, O. A., Etim, E. E., Adamu, H.,' Chindo, I., Andrew, C., & Khanal, G.P. (2017). Quantum Chemical Studies on Decyl Heptadecanoate ($C_{27}H_{54}O_2$) Detected in Ethyl Acetae Leaf Extract of *Chrysophyllum albidium*. *Elixir Applied Chemistry,* 111, pp. 48828-48838.

Velasco, L., Ruiz, M., Shariati, B. & Vela, AP. (2022). Chapter Eight - Machine Learning for optical spectrum analysis. In: Lau APT, Khan FN, eds. Machine Learning for Future Fiber-Optic Communication Systems. *Academic Press;* 225-279. doi:10.1016/B978-0-32-385227-2.00015-2

Wei, Y., Carol, H., Christopher, C. & Saima H. (2023). A Machine Learning Approach for Early Diagnosis of Cognitive Impairment Using Population-Based Data, *Journal of Alzheimer's Disease*, 91, 1, pp. 449-461.https://doi.org/10.3233/JAD-220-776

Weili, L. & Liu Y.(2010) Tailoring the Morphology of the Poly(3-hexylthiophene)/C60 Films and Charge Carrier Mobility. https://doi.org/10.10-21/bk-2010-1034.ch009

Weili, L., Ruigang, L., Wen, W., Weiwei, L., Wenyong, L., Kai, Z., Lin, M., Ye T., Zhishan B. & Yong H. (2009). Tailoring Nanowire Network Morphology and Charge Carrier Mobility of Poly(3-hexylthiophene)/C60 Films. *The Journal of Physical Chemistry C,* 113 , 26, pp. 11385-11389. https://doi.org/10.1021/jp900042h

Weimin, Z., Yi, Z., Duancheng, Z., Jianrong, X. & Ling, W. (2023). HiGNN: A Hierarchical Informative Graph Neural Network for Molecular Property Prediction Equipped with Feature-Wise Attention. *Journal of Chemical Information and Modeling* , 63, 1, pp. , 43-55. https://doi.org/10.1021/acs.jcim.2c01099

Whitley, D., Sutton, A.M.(2012). Genetic Algorithms — A Survey of Models and Methods. In: Rozenberg, G., Bäck, T., Kok, J.N. (eds) Handbook of Natural Computing. Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-540-92910-9_21

Wu, L., Guo, T. & Li, T. (2021). Machine learning-accelerated prediction of overpotential of oxygen evolution reaction of single-atom catalysts. iScience, 24, 5, https://doi. -org/10.1016/

Xiangyu, Z., Jing, C., Kexin, Z., Jiasheng, W. & Yongjin, L. (2019). Machine Learning Prediction on Properties of Nanoporous Materials Utilizing Pore Geometry Barcodes. *Journal of Chemical Information and Modeling 59, 11, pp.* 4636-4644. https://doi.org/10.1021/acs.jcim.9b00623

Yang, Y. (2020). Dynamic pore-scale dissolution by CO_2-saturated brine in

carbonates: Impact of homogeneous versus fractured versus vuggy pore structure. *Water Resour. Res.* https://doi.org/10.1029/2019WR026112

Yaroslava, P., Yuriy, K. (2014). A procedure for meaningful unsupervised clustering and its application for solvent classification. *Open Chemistry* 12, 5, pp. 594-603. https://doi.org/10.2478/s11532-014-0514-6

Zavitsas A. (1998). Energy barriers to chemical reactions. Why, how, and how much? Non-Arrhenius behavior in hydrogen abstractions by radicals. *J Am Chem Soc*. 120, 26, pp. 6578-6586. doi:10.1021/ja973698y.

Zaw-Myo, W., Allen, M., Cheong, W., and Scott H. (2023). Using Machine Learning To Predict Partition Coefficient (Log P) and Distribution Coefficient (Log D) with Molecular Descriptors and Liquid Chromatography Retention Time. *Journal of Chemical Information and Modeling 63* (7), 1906-1913. https://doi.org/10.1021/acs.jcim.2c01373

Zhang, Z., Huang, W., Liao, Y., Song, Z., Shi, J,, Jiang, X., Shen, C., and Zhu Z. (2022). Bearing fault diagnosis via generalized logarithm sparse regularization. *Mechanical Systems and Signal Processing.* 167(Part B):108576. doi:10.1016/j.ymssp.2021.108576

Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen X. (2020). Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. ISA Transactions. 107, pp. 224-255. doi:10.1016/j.isatra.2020.08.010

Compliance with Ethical Standards

Declarations

The authors declare that they have no conflict of interest.

Data availability

All data used in this study will be readily available to the public.

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public.

Competing interests

The authors declared no conflict of interest.

Funding

The authors declared no source of fundig

Authors' Contributions

H.S. Samuel., J.P. Shinggu and B. Bako were involved in literature review, writing and drafting, revision and editing while E.E. Etim was involved in conceptualization, revision and drafting