Faculty of Natural and Applied Sciences Journal of Scientific Innovations
Print ISSN: 2814-0877 e-ISSN: 2814-0923
www.fnasjournals.com
Volume 5; Issue 1; December 2023; Page No. 103-115.

MACHINE LEARNING FOR CHARACTERIZING HALOGEN BONDING INTERACTIONS

[1]Samuel, H. S., [2]Nweke-Maraizu, U., & *[1]Etim, E. E.
[1]Department of Chemical Sciences, Federal University PMB, 1020, Wukari
[2]Department of Chemistry, Rivers State University, Nkpolu-Oroworukwo, Port Harcourt.

***Corresponding author email:** emmaetim@gmail.com

Abstract
Halogen bonding, a non-covalent interaction involving halogen atoms as electron acceptors, plays a crucial role in various chemical and biological processes. Understanding the subtleties of halogen bonding interactions is essential for advancing fields such as drug design, materials science, and supramolecular chemistry. This article explores the application of machine learning techniques in characterizing halogen bonding interactions, providing a novel and efficient approach to deciphering complex molecular relationships. Machine learning models, ranging from classical methods to advanced deep learning architectures, are employed to analyze large datasets of molecular structures and their corresponding halogen bonding properties. These models are trained on diverse sets of halogen-containing compounds, enabling them to recognize patterns and correlations that may be challenging for traditional computational methods. The article discusses the potential of machine learning in predicting halogen bond strength, identifying key molecular features influencing interaction strength, and elucidating structure-activity relationships. The presented findings underscore the potential of machine learning as a powerful tool for unravelling the intricacies of molecular interactions and guiding the design of functional molecular systems.

Keywords: Machine Learning, Artificial Intelligence, Halogen Bonding, Non-Covalent Interactions

Introduction

Non-covalent interactions are interactions between molecules that do not involve the sharing of electrons. They are more dispersed variations of electromagnetic interactions between molecules or within a molecule. These interactions are critical in maintaining the three-dimensional structure of large molecules, such as proteins and nucleic acids (Fourmigué, 2009). Some common types of non-covalent interactions in proteins include hydrogen bonds, salt bridges, van-der-Waals forces, π-cation interactions, sandwich or π-π stacking or T stacking. Other types of non-covalent interactions include the Aerogen bond, Chalcogen bond, Pnicogen bond, Tetrel bond, Triel bond and Halogen bond (Costa, 2017). Due to its essential function in a variety of chemical and biological processes, halogen bonding, a kind of noncovalent contact involving halogen atoms (such as fluorine, chlorine, bromine, or iodine) as electron acceptors, has drawn substantial interest in several scientific fields (Szell et al., 2016). Non-covalent interactions, essential in maintaining the structural integrity of large molecules like proteins and nucleic acids, encompass diverse electromagnetic forces. Among these interactions, halogen bonding, involving halogen atoms as electron acceptors, has gained considerable attention due to its pivotal role in various chemical and biological processes. Traditional methods have contributed to understanding halogen bonding, yet the complex and diverse nature of these interactions across molecular systems poses challenges (Clark et al., 2007). The burgeoning availability of experimental and theoretical data has prompted the exploration of data-driven approaches, with machine learning emerging as a potent tool in chemistry. Machine learning, a subset of artificial intelligence, facilitates the identification of patterns and relationships in large datasets, providing predictive models and data-driven insights (Samuel et al., 2023). In the context of halogen bonding, machine learning serves as a versatile and powerful tool, complementing traditional theoretical and experimental approaches. Machine learning enables a data-driven analysis of halogen bonding interactions through feature engineering, extracting pertinent features from comprehensive databases (Auffinger et al., 2004). Various molecular descriptors, quantum-chemical traits, and geometrical dimensions serve as input data, elucidating crucial variables influencing halogen bonding. Predictive modelling, facilitated by supervised machine learning algorithms, offers rapid and precise forecasts of halogen interactions, aiding in the development of advanced materials and pharmaceuticals (Auffinger et al., 2004). Unsupervised learning approaches, such as clustering

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

algorithms, identify patterns and novel classes of halogen bonding, contributing to a deeper understanding of its distribution in chemical systems. In situations with limited labelled data, semi-supervised learning techniques enhance prediction models by incorporating unlabeled data (Auffinger et al., 2004). Deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), exhibit promise in capturing complex patterns in chemical bonding, providing a comprehensive understanding of halogen bonding processes. Machine learning not only sheds light on the influencing variables but also facilitates interpretability through model interpretability methodologies. Chemists and researchers gain insights into how specific characteristics or atoms impact the overall geometry and bonding strength of halogens (Andrew et al., 2018). The integration of machine learning in halogen bonding research accelerates molecular chemistry investigations, offering predictive models, data-driven insights, and enhanced comprehension of these crucial interactions (Metrangolo & Resnati, 2008). The article aims to explore the various machine learning models in the characterization of halogen bonding. The scope of the study is to provide more insights into halogen bonding and its applications in numerous scientific fields and the advancement of machine learning over conventional theoretical and practical approaches.

Machine learning algorithms used in characterizing halogen bonding

Machine learning algorithms are pieces of code that help people explore, analyze, and find meaning in complex data sets. Each algorithm is a finite set of unambiguous step-by-step instructions that a machine can follow to achieve a certain goal1. Some popular machine learning algorithms include linear regression, logistic regression, decision tree algorithm, support vector machine algorithm, and many more (Guo et al., 2018). By utilizing massive datasets and discovering patterns from the data, machine learning techniques significantly contribute to the characterisation of halogen bonding interactions (Mas et al., 2003). We will go into great detail about some of the popular machine learning methods here that are applied in this situation:

i. Support Vector Machines (SVM): SVM is a well-known supervised learning technique in computational chemistry and cheminformatics. It helps determine whether or not a halogen bonding connection occurs between two molecules since it is good for binary classification tasks. Finding the best hyperplane to distinguish the positive and negative occurrences in the feature space is the goal of SVM. SVM can distinguish between compounds that have and don't have halogen bonding interactions using molecular descriptors and quantum-chemical characteristics (Metrangolo & Resnati, 2008). The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed a Support Vector Machine. Consider the below diagram in Figure 1.0 which two different categories are classified using a decision boundary or hyperplane:

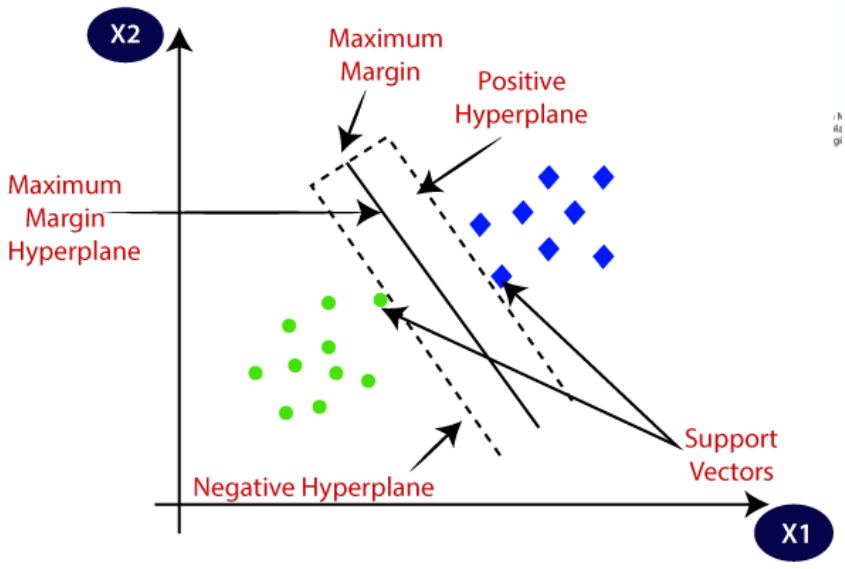

Figure 1.0: Schematic diagram of SVM (Metrangolo & Resnati, 2008).

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, if we want a model that can accurately identify whether it is a cat or dog, such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as the support vector creates a decision boundary between these two data (cat and dog) and chooses extreme cases (support vectors), (it will see the extreme case of cat and dog. Based on the support vectors, it will classify it as a cat. Consider the below diagram:

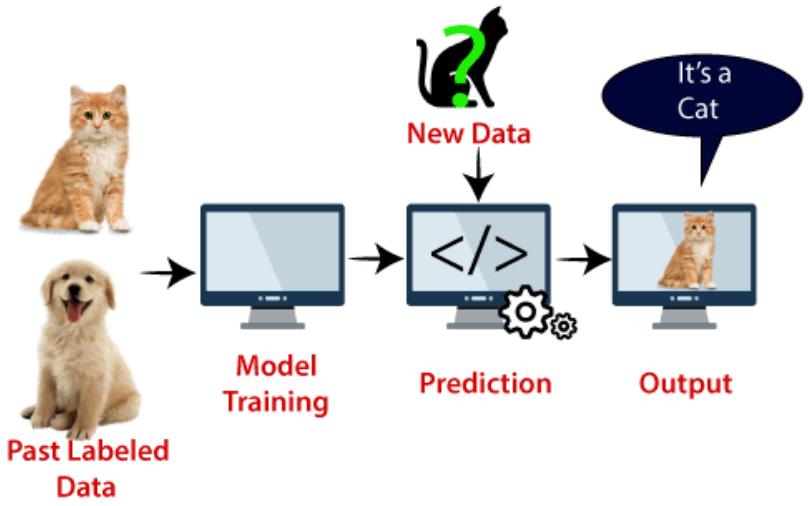

Figure 2.0: Representation of SVM using KNN classifier (Metrangolo & Resnati, 2008).

ii. Random Forest (RF): Random Forest is an ensemble learning technique that blends various decision trees as shown in Figure 3.0 to increase prediction resilience and accuracy. It may be applied to both classification and regression problems. The strength or energy of halogen bonding interactions between molecules can be predicted using RF when characterizing halogen bonding. The complicated and non-linear correlations between chemical characteristics and halogen bonding qualities can be captured by RF by taking into account several decision trees (

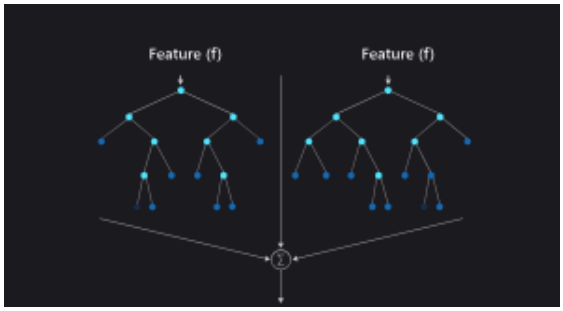

Figure 3.0: Diagram showing random forest (Li et al., 2019).

iii. Gradient Boosting Machines (GBM): GBM is a different ensemble learning method that constructs several weak learners (usually decision trees) progressively, with each one concentrating on the flaws of the preceding one. Gradient boosting can handle complicated datasets and is frequently more accurate than individual decision trees. Compared to standalone decision trees, GBM has a greater predictive performance when used to forecast halogen bonding parameters including strength, shape, and energy (Li et al., 2019).

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

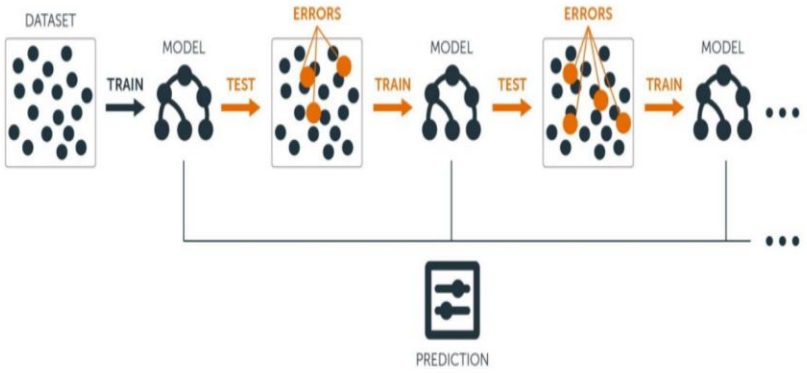

Figure 4.0: graphical representation of the GBM algorithm (Li et al., 2019).

iv. K-Nearest Neighbors (KNN): For classification and regression problems, KNN is a straightforward and understandable approach. The majority class among a sample's k-nearest neighbours in the feature space is used to classify it. KNN may be used to determine if a molecule shows a halogen bonding interaction based on the similarities with other compounds that have known interactions to characterize halogen bonding as shown in Figure 5.0 (Novak & Stilinović, 2021).

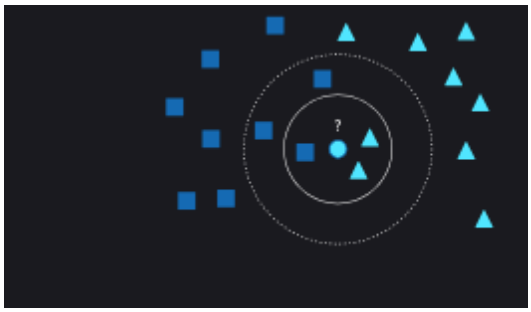

Figure 5.0: graphical representation of KNN (Novak, & Stilinović 2021).

v. Neural Networks (NN): Deep learning architectures, in particular, have achieved outstanding results in several domains, including chemistry. Halogen bonding interactions have been studied using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) (Aubert et al., 2017). While RNNs can handle time-series data relating to halogen bonding interactions, molecular dynamics trajectories, and other sequential data, CNNs can evaluate molecular structures and discover patterns within the 3D coordinates of atoms.

vi. Gaussian Process Regression (GPR): A probabilistic model that can include uncertainty in predictions is Gaussian Process Regression. When handling tiny datasets or noisy data, it is very helpful. By giving both point estimates and confidence ranges for the forecasts, GPR may be used to anticipate the strengths and characteristics of halogen bonds (Beale et al., 2013).

vii. Clustering Algorithms: Molecules with comparable halogen bonding characteristics can be grouped using unsupervised clustering methods such as K-Means, Hierarchical Clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Clustering can shed light on different kinds of halogen bonding interactions and how they vary in various chemical environments (Bondi et al., 1964).

Machine learning models

Halogen Bonding Supervised Learning Model

Models of Supervised Learning Training and Validation: In the process of understanding the underlying patterns and correlations between chemical characteristics and halogen bonding qualities, the model is trained using Labelled

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

data—that is, data containing known halogen bonding interactions (Catalano et al., 2020). The following stages are involved in training and evaluating supervised learning models for predicting halogen bonds:

i. Data Splitting: The labelled dataset is split into two groups: the training set and the validation set. The validation set is used to evaluate the model's performance and generalizability whereas the training set is used to train the model (Cavallo et al., 2016).
ii. Feature Extraction: From the molecular structures, pertinent molecular features are extracted, including geometric metrics, quantum-chemical characteristics, and molecular descriptors. The supervised learning algorithm receives these characteristics as input (Christopherson et al., 2018).
iii. Model selection: To forecast halogen bonding, a variety of supervised learning methods, including Support Vector Machines (SVM), Random Forest (RF), and Neural Networks (NN), can be used. The particulars of the task and the properties of the dataset will determine which algorithm is used.
iv. Model Training: Using the training set, the chosen algorithm is trained to discover the connections between the input features and the desired halogen bonding characteristics (such as halogen bond strength or shape) (Turunen et al., 2017a).
v. Tuning the hyperparameters: To get the greatest performance, the hyperparameters of many supervised learning algorithms must be tuned. The ideal set of hyperparameters is found using methods like grid search or random search.
vi. Model Validation: The performance of the trained model is tested on the validation set. The goal of this stage is to prevent overfitting (when a model performs well on training data but badly on fresh data) and to make sure the model generalizes well to samples that have not yet been observed (Clark et al., 2007).

Applications of Supervised Learning in Halogen Bonding Characterization

i. Halogen Bond Strength Prediction: Supervised learning methods have been used to estimate the strength of halogen bonding interactions in chemical complexes. The models can precisely predict the interaction energies between halogen atoms and electron donors by training on datasets with empirically established halogen bond strengths. Such forecasts are useful for rational drug design and material discovery (Turunen, et al., 2017b).
ii. Halogen bonding classification: Classifying compounds based on the presence or lack of halogen bonding interactions has been done using supervised learning techniques. The models can detect compounds with halogen bonding and discriminate them from non-halogen bonding systems by training on labelled datasets (Vanderkooy et al., 2019).
iii. Virtual Screening for Halogen Bond Acceptors: To find probable halogen bond acceptors in huge chemical datasets, supervised learning methods have been utilized for virtual screening. The models can prioritize substances with comparable structural characteristics that can be further investigated in experimental experiments by training on known halogen bond acceptors (Tu & Laaksonen, 2000).
iv. Prediction of halogen bond geometrics: Halogen bond geometries have been predicted using supervised learning techniques, which include bond angles and lengths. The models may examine the structural characteristics of the interfacing molecules and offer information on the ideal halogen bonding geometries (Walsh et al., 2001).

Unsupervised learning for pattern recognition

Unsupervised learning is a method of machine learning where the model learns from data that has not been labelled to find patterns, structures, or groupings. Unsupervised learning approaches may be extremely helpful for discovering and displaying halogen bonding patterns in big datasets without the requirement for predetermined labels when it comes to describing halogen bonding (Nazemzadeh et al., 2021). Some typical halogen bonding analysis uses of unsupervised learning include

2.2.1 Clustering Methods for Identifying Halogen Bonding Patterns in Large Datasets

In unsupervised learning, clustering techniques are frequently used to group comparable data points based on their inherent characteristics. Clustering techniques may be used to categorize molecules with similar halogen bonding patterns and to distinguish between different types of halogen bonding interactions (Thoeny et al., 2022). In halogen bonding studies, the following clustering techniques are frequently employed

i. K-Means Clustering: K-Means is a well-liked and uncomplicated approach for clustering data into k clusters, where k is a user-defined parameter. The algorithm's objective is to reduce the distance between individual data points and the centroids of their clusters as shown in figure 6.0 (Politzer et al., 2007). Halogen bonding

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

complexes with comparable geometric characteristics can be clustered together by using K-Means to analyze a collection of molecular structure data.

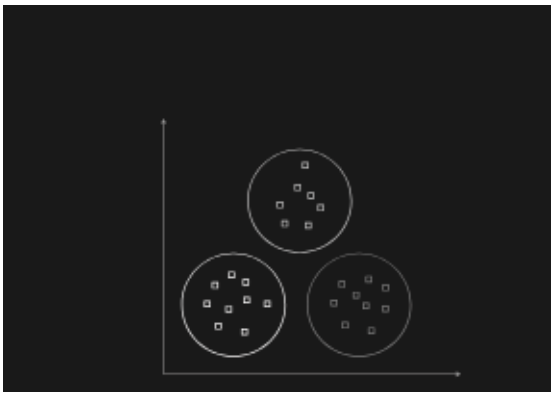

Figure 6.0: graphical representation of K-mean clustering (Politzer et al., 2007).

ii. DBSCAN (Density-Based Spatial Clustering of Applications with Noise): DBSCAN is a density-based clustering technique that divides data points into clusters according to their connectedness and density. Even in the face of noise, it can successfully locate clusters of various forms and densities in halogen bonding datasets. When the quantity of clusters is unknown in advance, DBSCAN is especially helpful (Politzer et al., 2007).

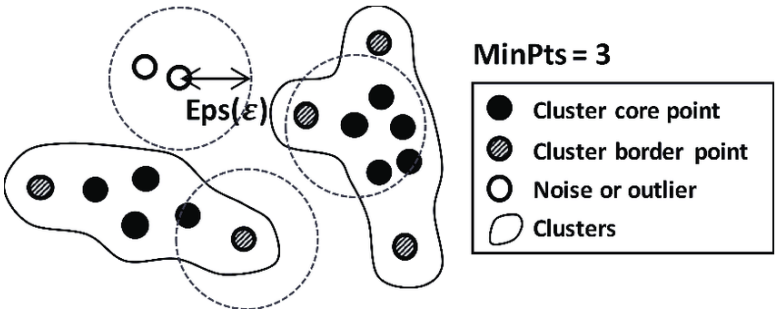

Figure 7.0: graphical representation of DBSCAN (Politzer et al., 2007).

iii. Hierarchical Clustering: To describe the link between data points, hierarchical clustering creates a tree-like structure (dendrogram). It can be either agglomerative—starting with each data point as a distinct cluster and gradually merging ones that are similar—or divisive—starting with all the data points in one cluster and repeatedly splitting them. The hierarchical structure of halogen bonding patterns is shown through hierarchical clustering as shown in Figure 8.00 (Etim et al., 2021).

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

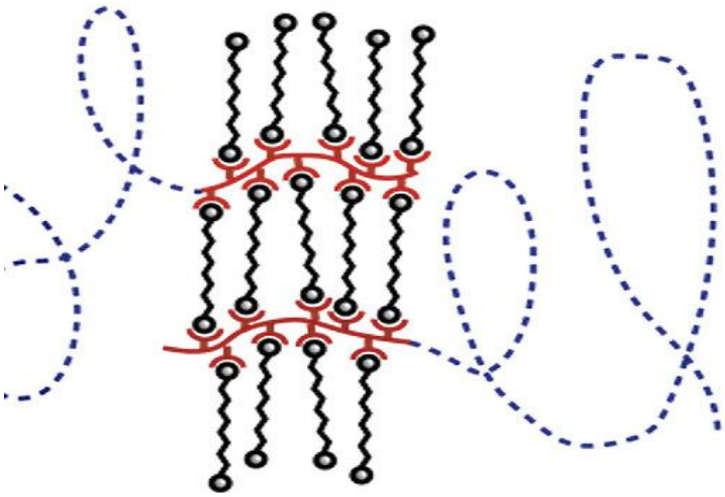

Figure 8.0: hierarchical clusters of halogen bonding patterns (Mohammadi et al., 2023).

Visualization of Principal Component Analysis and t-distributed Stochastic Neighbor Embedding

Principal Component Analysis (PCA): With the use of Principal Component Analysis (PCA), which is a popular dimensionality reduction approach, high-dimensional data may be projected onto a lower-dimensional space while retaining the majority of its original variance (Khan et al., 2021). The most important components (principal components) can be found by using PCA to a dataset of molecular properties pertinent to halogen bonding (Mohammadi et al., 2023). The significant structural changes in the dataset may then be deduced by visualizing these components.

i. t-Distributed Stochastic Neighbor Embedding (t-SNE): t-SNE is a nonlinear dimensionality reduction technique that emphasizes maintaining local pairwise similarities in the high-dimensional space. When displaying complicated datasets with comparable data points clustered together in a lower-dimensional space, it is especially helpful. To see the variety and grouping of halogen bonding interactions in sizable molecular datasets, t-SNE has been employed in halogen bonding analysis (Beale et al., 2013).

Semi-Supervised Learning for Limited Data Scenarios

Semi-supervised learning is a machine learning technique that blends labelled (with known results) and unlabelled (with unknown results) data to enhance model performance, particularly when Labelled data is difficult to come by or is expensive to acquire (Nakajima et al., 2021). Semi-supervised learning can be very helpful for assessing halogen bonding because it can be difficult and expensive to gather experimental data for big datasets. By combining the large quantity of unlabelled data with the already available Labelled data, semi-supervised learning can improve prediction models and provide important new information about halogen bonding interactions. Let's talk about a few case examples that show how semi-supervised learning is advantageous for halogen bonding analysis (Humphrey et al., 2023). For example, Halogen Bond Classifier with Partial Supervision. Let's say we have a tiny dataset that has been tagged and has data on halogen bonding interactions (positive occurrences) and non-halogen bonding interactions (negative instances). The supplied Labelled data in this case study is insufficient to train a reliable classifier. On the other hand, we have access to a sizable collection of unlabelled data that contains different chemical configurations (Higuchi & Iijima 1985). Co-training or self-training are examples of semi-supervised learning methods that can be used. In co-training, two distinct models are trained on various sets of features, with the predictions from one model on the unlabelled data serving as additional labels for the training of the second model and vice versa. The models use the unlabelled data to enhance their performance as they update and label each other repeatedly. A single model is first trained on the Labelled data in self-training. After making predictions on the unlabelled data, the Labelled dataset is then updated with the high-confidence predictions. On the new Labelled dataset, the model is retrained, and this procedure is performed repeatedly (Takamura et al., 2021). The performance of the halogen bond classifier may be significantly enhanced by the semi-supervised learning strategy by including the unlabelled data during the training phase. The model gains knowledge from the rich and varied information in the unlabelled data, improving generalization and improving prediction accuracy for halogen bonding interactions (Bayon et al., 2020).

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

2.5 Deep Learning Models for Halogen Bonding

A group of machine learning algorithms called neural networks are modelled after the structure and operation of the human brain. They are made up of layer-organized networks of linked nodes known as neurons. Each neuron analyses the incoming data, performs a mathematical change, and then transfers the result to the following layer, continuing until the desired result is obtained. A subclass of neural networks known as "deep learning" refers to models with several hidden layers that allow them to learn sophisticated patterns and representations from difficult input (Wang et al., 2018).

i. Feedforward Neural Networks (FNN): The simplest type of neural network, feedforward neural networks only allow data to flow in one way, from the input layer to the output layer. They excel in classification and regression-related tasks.

ii. Convolutional Neural Networks (CNNs): CNNs were developed to handle problems requiring spatial and visual data. Convolutional, pooling and fully linked layers make up these layers. CNNs are useful for halogen bonding analysis when taking into account chemical structures and visuals since they are particularly good at identifying spatial patterns and characteristics in pictures (Ester et al., 1996).
Recurrent neural networks (RNNs): RNNs are made for sequential data, such as time series or data sequences. They include loops that let data endure through time and let the model account for temporal dependencies. When examining molecular dynamics trajectories or other sequential data about halogen bonding interactions, RNNs are helpful.

iii. LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit): The vanishing gradient problem is addressed by the specialized RNN versions LSTM and GRU, which improve the handling of long-range dependencies in sequences. When collecting long-term trends in molecular dynamics data or other sequential data pertinent to halogen bonding, LSTM and GRU designs are frequently utilized (Bakshi et al., 1998).

Convolutional neural networks and recurrent neural networks in characterizing halogen bonding

i. Utilizing CNNs for Molecular Structure Analysis: CNNs have been used to analyse three-dimensional molecular structures and locate halogen bonding connections inside of molecules (Bakshi et al., 1998). CNNs may learn to detect halogen bond shapes and recognize the interacting atoms by feeding in the 3D coordinates of the atoms. This makes it possible to automatically identify and view halogen bonding patterns in molecular systems (Shedrach et al., 2023).

ii. RNNs for Molecular Dynamics Analysis: RNNs, particularly LSTM and GRU, are useful for deciphering the temporal development of halogen bonding interactions and evaluating molecular dynamics trajectories. RNNs can capture the dynamic behaviour of halogen bonding and give insights into its stability, variation, and duration by feeding them sequential data reflecting atomic locations and pressures over time (Wilcken, 2017).

iii. Combination CNN-RNN Models: CNN and RNN architecture-based hybrid models are effective at describing halogen bonding. For instance, a molecular dynamics simulation's various snapshots may be used to train a CNN to extract characteristics from molecular structures, which can then be input into an RNN to examine the temporal development of halogen bonding (Onen et al., 2023).

iv. Deep Learning in Virtual Screening: CNNs and other deep learning techniques have been used in virtual screening to find prospective acceptors or donors of halogen bonds in sizable chemical compound databases. CNNs can rank potential compounds for additional experimental testing by analyzing the 3D structures of molecules and predicting their likelihood to create halogen bonding bonds (Onen et al., 2017).

v. Many facets of halogen bonding studies have benefited from deep learning approaches. Researchers may better understand the complexity of halogen bonding interactions and speed up the development of new materials and medications by utilizing the capabilities of neural networks, especially CNNs and RNNs.

vi. It is crucial to remember that deep learning models frequently need a lot of data for training and careful hyperparameter adjustment to prevent overfitting. Although attempts are being made to enhance model interpretability and comprehend the underlying causes driving the predictions in halogen bonding analyses, the interpretability of deep learning models continues to be a problem

Visualization and Interpretability of Machine Learning Models

Techniques for Analyzing Machine Learning Models in Halogen Bonding Studies

Understanding how machine learning models work and how they make predictions is crucial for understanding the variables affecting halogen bonding interactions. Model interpretability can be a useful tool in halogen bonding

investigations for learning more about the molecular details and structural traits that affect the stability, shape, and strength of halogen bonds. Following are some techniques for analyzing machine learning models in halogen bonding studies:

i. Feature Importance: Techniques for determining the most important features in a model's decision-making process include permutation importance and SHAP (SHapley Additive exPlanations). The most important molecular descriptors or quantum-chemical features might be ranked to help researchers determine the essential variables influencing halogen bonding interactions.
ii. Convolutional neural networks (CNNs) employed in halogen bonding research may display the portions of chemical structures that activate particular neurons in the model using a technique called activation mapping. This clarifies which molecular components are essential for the CNN to detect halogen bonding patterns (Bulat & Hobza, 2013).
iii. Attention Mechanisms (RNNs): Recurrent neural networks (RNNs) with attention mechanisms allow scientists to pinpoint the precise timesteps in a molecular dynamics trajectory when the model concentrates most of its attention on halogen bonding interactions. As a result, during the simulation, the temporal evolution and significance of halogen bonding may be shown (Shinggu et al., 2023).
iv. Layer-wise Relevance Propagation (LRP): LRP is a method that links model predictions to the features that were used to create them. LRP can shed light on how certain atoms or molecular areas affect the overall shape or strength of halogen bonds in halogen bonding research (Riley et al., 2003).

Visualizing Halogen Bonding Interactions Using Machine Learning Insights

Several methods of visualization include:

i. Molecular visualization: Highlighting halogen bonding connections within a molecule may be done by visualizing molecular structures using programs like PyMOL or VMD. The selection of atoms or residues important to halogen bonding can be guided by machine learning insights, such as feature significance or attention weights, improving the interpretability of the image (Bissantz, 2007).
ii. 3D Heatmaps: The distribution of halogen bonding contacts in a chemical complex may be seen using 3D heatmaps. Researchers can locate hotspots and areas of high interaction energy by utilizing colour gradients to indicate the intensity of halogen bonding at various locations in space.
iii. Time-Series Plots: For molecular dynamics simulations, time-series plots can visualize the changes in halogen bonding interactions over time. The selection of timesteps to concentrate on certain instances when halogen bonding interactions are most prominent can be guided by machine learning insights, such as attention processes.
iv. Interaction Networks: The network of interactions between halogen bond acceptors and donors in a chemical system may be shown using interaction networks. The network may show the connectedness and significance of various halogen bonding interactions by displaying the strength and shape of halogen bonding as edge weights (Doherty, 2007)

A thorough understanding of halogen bonding interactions and their contributions to molecule identification and stability may be obtained by researchers by integrating machine learning interpretability tools with visualization methodologies. Designing novel materials and medications based on halogen bonding interactions is made easier with the aid of these visualizations, which assist in closing the gap between sophisticated machine learning algorithms and insightful chemical understandings (Kupka et al., 2015).

Prospects of machine learning in halogen bonding

Halogen bonding has very bright future potential for machine learning, and these changes will greatly increase our knowledge of this crucial chemical interaction. Here are some important halogens bonding applications of machine learning in the future:

1. Halogen Bonding Predictive Models:
   As machine learning algorithms continue to get more accurate and effective, they will make it possible to create halogen bonding predictions that are quite accurate. These models can help with the quick selection and creation of compounds with certain halogen bonding characteristics, which can lead to the development of new substances and medications (Paraschiv, 2020).
2. Handling Larger and More Complicated Systems: Machine learning algorithms will improve as they develop, making it easier for them to handle larger and more complicated molecular systems.

Cite this article as:
Samuel, H. S., Nweke-Maraizu, U., & Etim, E. E. (2023). Machine learning for characterizing halogen bonding interactions. *FNAS Journal of Scientific Innovations*, *5*(1), 103-115.

3. Quantum chemistry and machine learning together: Combining quantum chemistry and machine learning techniques has a lot of potential (Bulat & Hobza, 2015). To analyze halogen bonding interactions in vast systems with great accuracy, hybrid techniques that combine the precision of quantum mechanics with the effectiveness of machine learning are often used (Zheng, 2018).
4. Unsupervised Learning for finding: The finding of novel halogen bonding patterns and interactions will be aided by unsupervised learning approaches like clustering and dimensionality reduction. By identifying fresh types of halogen bonding, these techniques can shed light on the various ways that this interaction might take place in various chemical contexts (Etim et al., 2018).
5. Insight into Biological Halogen Bonding: Machine learning can offer important insights into halogen bonding interactions in biological systems, such as protein-ligand and protein-protein interactions. Understanding how halogen bonding affects biomolecular recognition and function can have a big impact on medication development and customized treatment (Humphrey et al., 2023).
6. Accelerating Materials and Drug Discovery: Machine learning algorithms can make it easier to find potential halogen bonding interactions in huge chemical and compound datasets. Concentrating on potential candidates with certain halogen bonding characteristics can hasten the identification of new materials and the development of new drugs (Etim et al., 2020).
7. Integrating Experimental and Computational Data: Machine learning can make it easier to integrate experimental data with computational findings. Examples of such experimental data include crystallographic structures and NMR data. Halogen bonding interactions in actual systems may be better understood and characterized as a result of this integration (Humphrey et al., 2023).

Machine learning has a bright future in halogen bonding research. Machine learning will be primary to expanding our understanding of halogen bonding and its applications in numerous scientific fields, from predictive modelling to comprehending complicated biomolecular interactions and expediting materials and medicine discovery (Mondal et al., 2021). Innovative findings and advancements in the realm of halogen bonding interactions will surely result from continued research and development in this area (Jeffrey et al., 1994).

Conclusion

Machine learning has become a potent and adaptable method for the analysis of halogen bonding interactions. Its uses in this area are numerous and include tasks like forecasting halogen bond strengths and geometries as well as grouping and visualizing various halogen bonding patterns. Researchers have been able to learn more about the nature and importance of halogen bonding in different chemical systems by combining machine learning with experimental and theoretical data. The creation of predictive models that can precisely forecast the features of halogen bonding has been made possible using supervised learning approaches. These models allow for the quick screening of various compounds, which aids in the identification of novel materials with particular halogen bonding properties. Additionally, the fusion of quantum chemical techniques with machine learning has created new opportunities. Unsupervised learning techniques have opened up new insights into the variety of halogen bonding interactions. To better understand the subtle differences in these interactions across various chemical environments, clustering and dimensionality reduction algorithms have discovered new classes of halogen bonding patterns. A more thorough knowledge of the temporal and angular dynamics of halogen bonding has been achieved by the interaction of machine learning with more established theoretical techniques, such as molecular dynamics simulations. We can better understand the stability and variations of halogen bonding thanks to such multiscale investigations that provide light on its dynamic character. The ability to interpret machine learning models has enhanced model transparency and allowed for the extraction of valuable chemical insights. Halogen bonding interactions may now be explored more easily thanks to visualization tools, which also give scientists useful visual representations of molecular structures and dynamic behaviours. Machine learning will continue to reveal the subtleties of halogen bonding interactions by utilizing the enormous quantity of data currently available, enabling advancements in medication design, the development of new materials, and other scientific fields. In the years to come, we may expect new developments and advances that will influence our understanding and use of this crucial chemical interaction as researchers further explore the potential of machine learning in describing halogen bonding.

References

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*, *7*(1), 77-82.

Aubert, E., Espinosa, E., Nicolas, I., Jeannin, O., & Fourmigué, M. (2017). Toward a reverse hierarchy of halogen bonding between bromine and iodine. *Faraday Discussions*, *203*, 389-406.

Auffinger, P., Hays, F. A., Westhof, E., & Ho, P. S. (2004). Halogen bonds in biological molecules. *Proceedings of the National Academy of Sciences*, *101*(48), 16789-16794.

Bakshi, B. R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *AIChE journal*, *44*(7), 1596-1610.

Bayon, A., de la Calle, A., Ghose, K. K., Page, A., & McNaughton, R. (2020). Experimental, computational and thermodynamic studies in perovskites metal oxides for thermochemical fuel production: A review. *International Journal of Hydrogen Energy*, *45*(23), 12653-12679.

Beale, T. M., Chudzinski, M. G., Sarwar, M. G., & Taylor, M. S. (2013). Halogen bonding in solution: thermodynamics and applications. *Chemical Society Reviews*, *42*(4), 1667-1680.

Bissantz, C. (2017). Halogen bonding: a powerful tool for modulating drug–protein interactions. Chemical Reviews, 117(8), 6159-6214.

Bondi, A. V. (1964). van der Waals volumes and radii. *The Journal of physical chemistry*, *68*(3), 441-451.

Bulat, F. A., & Hobza, P. (2013). Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics*, 15(21), 8107-8119.

Bulat, F.A., & Hobza, P. (2015). Advances in describing non-covalent interactions involving proteins: theory and experiment. *Chemical Society Reviews, 44*(11), 3897-3907.

Catalano, L., Perez-Estrada, S., Terraneo, G., Pilati, T., Resnati, G., Metrangolo, P., & Garcia-Garibay, M. A. (2015). Dynamic characterization of crystalline supramolecular rotors assembled through halogen bonding. *Journal of the American Chemical Society*, *137*(49), 15386-15389.

Cavallo, G., Metrangolo, P., Milani, R., Pilati, T., Priimagi, A., Resnati, G., & Terraneo, G. (2016). The halogen bond. *Chemical reviews*, *116*(4), 2478-2601.

Christopherson, J. C., Topić, F., Barrett, C. J., & Friščić, T. (2018). Halogen-bonded cocrystals as optical materials: Next-generation control over light–matter interactions. *Crystal Growth & Design*, *18*(2), 1245-1259.

Clark, T., Hennemann, M., Murray, J. S., & Politzer, P. (2007). Halogen bonding: The σ-hole: Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th–9th, 2005. *Journal of molecular modeling*, *13*, 291-296..

Costa, P. J. (2017). The halogen bond: nature and applications. *Physical Sciences Reviews*, *2*(11), 20170136.

Doherty, M. F., & Martínez, T. J. (2007). Probing halogen bonds in solution. *Journal of the American Chemical Society, 129*(4), 870-871.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* 96(34), 226-231.

Etim, E. E., Khan, M. E., Godwin, O. E., & Ogofotha, G. O. (2021). Quantum chemical studies on C4H4N2 Isomeric Molecular Species. *Journal of the Nigerian Society of Physical Sciences*, 429-445.

Etim, E. E., Olagboye, S. A., Godwin, O. E., & Atiatah, I. M. (2020). Quantum Chemical studies on Silicon tetrafluoride and its protonated analogues. *Int. J. Modern Chem*, *12*(1), 26-45.

Etim, E. E., Onen, A. I., Andrew, C., Lawal, U., Udegbunam, I. S., & Ushie, O. A. (2018). Computational studies of C5H5N Isomers. *Journal of Chemical Society of Nigeria*, *43*(2).

Fourmigué, M. (2009). Halogen bonding: recent advances. *Current Opinion in Solid State and Materials Science*, *13*(3-4), 36-45.

Guo, N., Maurice, R., Teze, D., Graton, J., Champion, J., Montavon, G., & Galland, N. (2018). Experimental and computational evidence of halogen bonds involving astatine. *Nature Chemistry*, *10*(4), 428-434.

Higuchi, A., & Iijima, T. (1985). DSC investigation of the states of water in poly (vinyl alcohol) membranes. *Polymer*, *26*(8), 1207-1211.

Jeffrey, G. A., Saenger, W., Jeffrey, G. A., & Saenger, W. (1991). The role of hydrogen bonding in the structure and function of the nucleic acids. *Hydrogen Bonding in Biological Structures*, 394-422.

Khan, M. E., Etim, E. E., Anyam, V. J., Abel, A., Osigbemhe, I. G., & Agber, C. T. (2021). Computational studies on Emodin (C15H10O5) from Methanol extract of Pteridium acquilinum leaves. *Journal of the Nigerian Society of Physical Sciences*, 360-384.

Kupka, T. (2015). Understanding the halogen bond: The σ-hole and beyond. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *5*(5), 360-373.

Mas, E. M., Bukowski, R., & Szalewicz, K. (2003). Ab initio three-body interactions for water. II. Effects on structure and energetics of liquid. *The Journal of chemical physics*, *118*(10), 4404-4413.

Metrangolo, P., & Resnati, G. (Eds.). (2008). *Halogen bonding: fundamentals and applications* (Vol. 126). Springer.

Mohammadi, M. D., Abdullah, H. Y., Louis, H., Etim, E. E., Edet, H. O., & Godfrey, O. C. (2023). Hexachlorobenzene (HCB) adsorption onto the surfaces of C60, C59Si, and C59Ge: Insight from DFT, QTAIM, and NCI. *Chemical Physics Impact*, *6*, 100234.

Mondal, S. K., Gorai, P., Sil, M., Ghosh, R., Etim, E. E., Chakrabarti, S. K., ... & Das, A. (2021). Is there any linkage between interstellar aldehyde and alcohol?. *The Astrophysical Journal*, *922*(2), 194.

Nakajima, M., & Nemoto, T. (2021). Machine learning enabling prediction of the bond dissociation enthalpy of hypervalent iodine from SMILES. *Scientific Reports*, *11*(1), 20207.

Nazemzadeh, N., Nielsen, R. F., Gernaey, K. V., Mansouri, S. S., & Andersson, M. P. (2021). Implementation of first-principles surface interactions in a hybrid machine learning assisted modelling of flocculation. In *Computer Aided Chemical Engineering*, *50*, 845-850. Elsevier.Novak, M., & Stilinović, V. (Eds.). (2021). Halogen Bonding in Supramolecular Chemistry. Wiley-VCH.

Onen, A. I., Etim, E. E., Victor, A. A., & Oghenekeno, O. G. (2023). Improved photochemical performance of dye-sensitized solar cell using tectona grandis and magnifera indica dye mix. *Communication in Physical Sciences*, *9*(2).

Onen, A. I., Joseph, J., Etim, 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*, 496-498.

Paraschiv, V.-R. (2020). Halogen bonding in rational drug design. *Journal of Medicinal Chemistry*, *63*(1), 31-69.

Politzer, P., Lane, P., Concha, M. C., Ma, Y., & Murray, J. S. (2007). An overview of halogen bonding. *Journal of molecular modeling*, *13*, 305-311.

Riley, K. E., & Bader, B. P. (2003). Halogen bonding: an AIM analysis of the σ-hole. *Journal of Physical Chemistry* A, 107(48), 10155-10163.

Samuel, H. S., Nweke-Mariazu, U., & Etim, E. E. (2023). Understanding intermolecular and intramolecular hydrogen bonds: spectroscopic and computational approaches. *Journal of Chemical Reviews*, *5*(4), 439-465. https://doi.org/10.48309/JCR.2023.407989.1235

Samuel, H. S., Nweke-Maraizu, U., Johnson, G., & Etim, E. E. (2023). A review of theoretical techniques in corrosion inhibition studies. *Communication in Physical Sciences*, *9*(4), 394-403.

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

Samuel, H. S., Etim, E. E., 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.

Shinggu, J. P., Etim, E. E., & Onen, A. I. (2023). Protonation-induced structural and spectroscopic variations in molecular species: A Computational Study on N2, H2, CO, CS, and PH3. *Communication in Physical Sciences*, *9*(4).

Shinggu, J. P., Etim, E. E., & Onen, A. I. (2023). Quantum chemical studies on C2H2O isomeric species: Astrophysical implications, and comparison of methods. *Communication in Physical Sciences*, *9*(2), 93-105.

Szell, P. M., & Bryce, D. L. (2016). Solid-state NMR studies of halogen bonding. *Modern Magnetic Resonance*, 1-18.

Szell, P. M., Zablotny, S., & Bryce, D. L. (2019). Halogen bonding as a supramolecular dynamics catalyst. *Nature communications*, *10*(1), 916.

Takamura, A., Tsukamoto, K., Sakata, K., & Kikuchi, J. (2021). Integrative measurement analysis via machine learning descriptor selection for investigating physical properties of biopolymers in hairs. *Scientific Reports*, *11*(1), 24359.

Thoeny, A. V., Parrichini, I. S., Gasser, T. M., & Loerting, T. (2022). Raman spectroscopy study of the slow order–order transformation of deuterium atoms: Ice XIX decay and ice XV formation. *The Journal of Chemical Physics*, *156*(15).

Tu, Y., & Laaksonen, A. (2000). The electronic properties of water molecules in water clusters and liquid water. *Chemical Physics Letters*, *329*(3-4), 283-288.

Turunen, L., Peuronen, A., Forsblom, S., Kalenius, E., Lahtinen, M., & Rissanen, K. (2017a). Tetrameric and dimeric [N··· I+··· N] Halogen-Bonded Supramolecular Cages. *Chemistry–A European Journal*, *23*(48), 11714-11718.

Turunen, L., Warzok, U., Schalley, C. A., & Rissanen, K. (2017b). Nano-sized I12L6 molecular capsules based on the [N··· I+··· N] halogen bond. *Chem*, *3*(5), 861-869.

Vanderkooy, A., Gupta, A. K., Földes, T., Lindblad, S., Orthaber, A., Pápai, I., & Erdélyi, M. (2019). Halogen bonding helicates encompassing iodonium cations. *Angewandte Chemie*, *131*(27), 9110-9114.

Walsh, R. B., Padgett, C. W., Metrangolo, P., Resnati, G., Hanks, T. W., & Pennington, W. T. (2001). Crystal engineering through halogen bonding: complexes of nitrogen heterocycles with organic iodides. *Crystal Growth & Design*, *1*(2), 165-175.

Wang, X., Shi, D., Wang, Z., Xu, C., Zhang, Q., Zhang, X., & Yu, Z. (2017). Online calibration of phasor measurement unit using density-based spatial clustering. *IEEE Transactions on Power Delivery*, *33*(3), 1081-1090.

Wilcken, R(2017). Halogen bonding and other σ-hole interactions: A perspective. *Chemical Reviews*, *117*(16), 9421-9481.

Yakubu, S., & Etim, E. E. (2023). Modern advances in click reactions and applications. *Journal of Chemical Reactivity and Synthesis*, *13*(2), 123-145.

Zheng, Y.-Z. (2018). Halogen bonding interactions in ligand recognition and catalysis. Chemical Reviews, 118(12), 6051-6107.