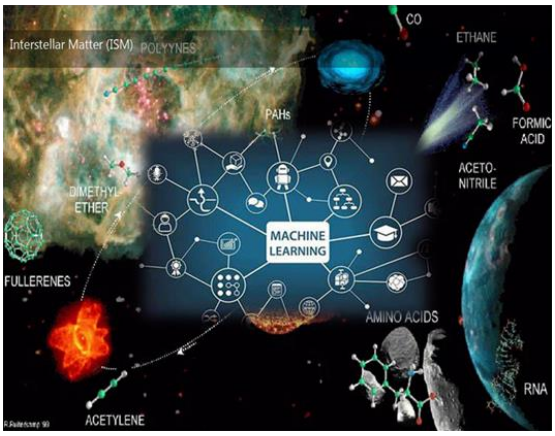
Communication in Physical Sciences, 2023, 10(1): 172-203

Machine learning of Rotational spectra analysis in interstellar medium

Humphrey Sam Samuel, *Emmanuel Edet Etim, John Paul Shinggu and Bulus. Bako

***Received:* 14 February 2023/Accepted 20 November 2023/Published 25 November 2023**

Abstract: In the investigation of rotating spectra concerning the interstellar medium, machine-learning approaches have been documented as effective instrument. The understanding of molecular rotational transitions in space and can be a significant source of information on the dynamics, physical properties, and chemical make-up of interstellar spaces. Traditional analytical techniques are however confronted with difficulties when dealing with the enormous and complicated information produced by telescopic observations. The handling of these massive datasets and the extraction of useful data from rotating spectra can be accomplished using machine learning methods, which are a promising approach. This article gives a general overview of the developments of machine learning in the analysis of rotational spectra in the interstellar medium. It goes over how to recognize and describe molecular transitions using supervised and unsupervised learning algorithms, deep learning architectures, and spectral line fitting methods. Also, machine learning algorithms can aid detection of spectral lines that are weak or infrequent but may contain important data regarding the chemical complexity of interstellar areas. They help make new molecular discoveries and enable the research of previously undiscovered spectral regions in the electromagnetic spectrum. Despite these developments, there are still problems to be solved, such as handling data noise, uncertainty, and over fitting. By enabling effective and automatic extraction of chemical information from complicated datasets, machine learning in rotational spectra analysis revolutionizes the study of interstellar chemistry. It enables scientists to learn about the chemical diversity and development of interstellar regions, making crucial contributions to our comprehension of the genesis and development of the universe.



Keywords: *Machine learning, artificial intelligence, interstellar molecules, rotational spectroscopy*

Humphrey Sam Samuel
Computational Astrochemistry and Bio-Simulation Research Group, Federal University Wukari
Email: humphreysedeke@gmail.com
Orcid id: 0009-0001-7480-4234

Emmanuel Edet Etim*
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email*:* *emmaetim@gmail.com*
Orcid id: 0000-0001-8304-9771

John Paul Shinggu
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: johnshinggu@gmail.com
Orcid id: /0009-0005-2216-3155

Bulus. Bako
Department of Chemical Sciences, Federal University Wukari, Taraba State
Email: bakobulus01@gmail.com
Orcid id: 0009-0001-3946-0712

1.0 Introduction

Molecular spectroscopy is a effective and flexible combination of methods that is useful in the investigation of the interaction of electromagnetic radiation with molecules. was Molecular spectroscopy can enables the identification, characterization, and investigation of chemical compounds and their properties by providing essential insights into the energy levels, vibrational states, rotational motion, and electronic transitions of molecules (Samuel *et al.,* 2023). Numerous scientific fields, including chemistry, physics, astronomy, biochemistry, environmental science, and materials science, rely on the application of molecular spectroscopy for various sectors. A subfield of spectroscopy called molecular spectroscopy investigates how electromagnetic radiation affects molecules. It offers useful details on the energy levels, vibrational states, rotational motion, and electronic transitions of molecules, which in turn aids in classifying molecules, figuring out their structures, and examining their chemical properties (Oliveira, *et al.,* 2022). The quantization of a molecule's energy levels forms the foundation of the fundamental ideas that guide molecular spectroscopy. Molecules consist of electrons orbiting nuclei, and transitions between different electronic states involve changes in the arrangement of electrons. Electronic transitions occur at specific energy levels, representing the energy required for an electron to move from one orbit to another. Transitions between these energy levels are brought about by the absorption or emission of photons by molecules in response to the incident electromagnetic radiation (Agúndez *et al.,* 2015). The generated spectra reveal a plethora of knowledge on the make-up and behaviour of the molecules. Molecular spectroscopy comes in a variety of forms, each based on a particular portion of the electromagnetic spectrum which includes vibrational, rotational, and electronic spectroscopy. Molecular vibrations, which are particular rhythmic motions of atoms within a molecule, are the subject of the field of vibrational spectroscopy, which examines them. Since vibrations in molecules are quantized, they take place at specific energy levels (Gúndez, *et al.,* 2018). Two methods that are primarily used to conduct vibrational spectroscopy are Infra red spectroscopy and Raman spectroscopy. The study of electronic transitions within molecules is the focus of electronic spectroscopy. Molecules and ultraviolet (UV), visible (VIS), or near-infrared (NIR) radiation interact in this process. When electrons take in energy and shift between various energy levels, electronic transitions take place. When examining the electronic energy levels, states, and transitions in molecules, which might reveal information about their electronic configuration and chemical reactivity, electronic spectroscopy is crucial. The study of molecule rotations is the main subject of rotational spectroscopy (Cernicharo, *et al.,* 1991). Changes in rotational energy levels are the result of molecules interacting with microwave radiation. Diatomic and straightforward polyatomic compounds are particularly amenable to rotational spectroscopy. For studying molecular structures, it gives exact data on molecule geometry, bond lengths, and moments of inertia (Etim *et al.,* 2017). Rotational spectroscopy offers special insights into the characteristics and dynamics of interstellar gas and molecules, making it a useful instrument for researching the interstellar medium (ISM). Understanding the molecular composition, physical characteristics, and kinematics of interstellar gas clouds and regions is particularly crucial

to the use of rotational spectroscopy in the ISM. Astronomers can discover and describe numerous compounds in the ISM using rotational spectroscopy (Samuel *et al.,* 2023). Several basic diatomic and polyatomic species, including formaldehyde (H_2CO), ammonia (NH_3), and carbon monoxide (CO), show conspicuous rotational transitions that can be seen in the millimeter wavelength ranges (Hirota, *et al.,* 2002). Astronomers can determine the abundances of these chemicals by examining these rotational lines, which gives them important knowledge about the chemical make-up of various interstellar environments. Rotational spectroscopy is useful in exploring the molecular clouds that give rise to the stars. These regions' physical characteristics, such as their temperature, density, and kinematics, can be learned through observations of rotational transitions in molecules (Janet, *et al.,* 2020). This information ability to enhance the understanding of the dynamics and structure of molecular clouds, which are where young stars are formed (Kim *et al.,* 2021). Rotational spectroscopy is crucial to astrochemistry, the study of chemical reactions in space. Astronomers can investigate the chemical interactions, ionization processes, and the creation and annihilation of molecules in the ISM by examining rotating spectra. Understanding the chemical development of interstellar gas and its function in star formation requires knowledge of this information (Lee *et al.,* 2021). Rotational spectroscopy is very helpful for finding molecular ions since they have distinctive rotational transitions. It is possible to learn more about the ionization processes and magnetic fields in interstellar gas clouds by observing molecular ions, such as HCO^+, HCN^+, and $NH^{+\cdot}$ Additionally, complex organic compounds, which are relevant to astrobiology and prebiotic chemistry investigations, can be found via rotational spectroscopy (Etim *et al.,* 2023). The interstellar medium (ISM) functions as the universe's cosmological research facility to solve cosmic mysteries. There are molecular clouds within this large area, where intricate chemical reactions produce a profusion of molecules, including diatomic and polyatomic species (Shinggu *et al.,* 2023). Understanding the molecular composition, physical settings, and kinematics of interstellar gas clouds is made possible by studying these molecules via rotational spectroscopy. With improvements in observational equipment, the amount of spectral data in the ISM is increasing exponentially, making conventional analysis techniques laborious and time-consuming. As a result, machine learning has begun to take off as a revolutionary technique for rotational spectra analysis (Mattioda *et al.,* 2020). By offering effective and automated methods for data processing, categorization, and interpretation, machine learning techniques have the potential to transform the study of rotating spectra in the ISM. Researchers can handle the complexity of enormous datasets and gain useful insights from the rich and varied rotational spectra emitted or absorbed by interstellar molecules by utilizing the computational capacity of machine learning methods (McGuire 2018). The use of machine learning in rotational spectra analysis has numerous significant benefits in this context. First off, it makes it possible for molecular species to be automatically identified and categorized based on their spectral signatures, speeding up the process of molecular identification and providing a thorough database of detected species (McGuire *et al.,* 2020). This study examines the numerous ways that machine learning can be used to analyze rotating spectra in the interstellar medium. We

examine the foundations of machine learning algorithms frequently applied to spectroscopic analysis, talk about the difficulties and opportunities in managing large and complex spectral datasets, and look at how machine learning can improve molecular identification, abundance measurements, and physical parameter estimation in the ISM (Zhao *et al.,* 2020). The combination of machine learning with rotational spectroscopy promises to open up new vistas in interstellar study as we travel farther into space and collect ever-increasing volumes of rotational spectra data (Etim *et al.,* 2020). The present article aims to explore machine learning in rotational spectra analysis in the interstellar medium. Additionally, we explore the potential of machine learning in investigating interstellar dynamics, such as gas kinematics and turbulent motions, and we emphasize its function in astrochemistry, which sheds light on the intricate chemical processes taking place in space.

2.0 Machine learning algorithm in rotational spectra analysis

In rotational spectroscopy, machine learning methods are frequently used to automate data processing, detect spectral patterns, and extract useful information from large datasets of rotating spectra. Researchers may generate predictions, categorize data, and interpret it based on the taught models thanks to these algorithms, which use mathematical and statistical techniques to understand patterns and relationships within the data (Chen *et al.,* 2020). The following machine-learning techniques are widely used in rotational spectroscopy in the interstellar medium (ISM):

1. Supervised Learning

Supervised learning algorithms are trained on labeled datasets, where each data point is associated with a known target or label. In rotational spectroscopy, supervised learning is frequently used for molecular identification, abundance measurements, and predicting physical parameters from rotational spectra.

i. Support Vector Machines (SVM): SVM is a classification-related supervised learning method. SVM can be used in rotational spectroscopy to categorize various molecular species according to their rotational spectral patterns. The method looks for the best hyperplane to divide various classes in the feature space. When working with complicated and non-linear spectrum data, SVM is especially helpful (Jia, *et al.,* 2018).

ii. Random Forests: Random Forests is an ensemble learning technique that brings together various decision trees to increase accuracy and decrease overfitting. Random Forests can be utilized in rotational spectroscopy for both classification and regression tasks. For example, it may categorize different molecular species according to their rotational spectra or forecast physical variables like temperature or density based on the strengths of rotational lines (Wang, *et al.,* 2020).

iii. Gradient Boosting Machines (GBM): GBM is another ensemble method that sequentially constructs a several weak learners (usually decision trees), each of which attempts to fix the mistakes of the previous one. Rotational spectra can be used to predict chemical characteristics and abundances using GBM.

iv. Linear Regression: In rotational spectroscopy, linear regression is employed for regression problems where the objective is to predict continuous variables (for example, chemical abundances or physical temperatures) based on rotational spectra features. It allows researchers to estimate quantities from spectral data by fitting a linear relationship between the input features and the goal value (Huang, *et al.,* 2019)

2. Unsupervised Learning Algorithms

On unlabeled datasets, unsupervised learning algorithms are applied with the aim of discovering patterns and structures in the data without the use of explicit target labels. In rotational spectroscopy, these techniques are particularly beneficial for problems involving clustering and dimensionality reduction

i. K-Means Clustering: K-means clustering is used to group rotational spectrum data into K clusters according to how similar they are as shown in fig 1. To identify molecular families or chemically related species, it can be helpful to group rotational spectra with comparable characteristics (Clarke *et al.*, 2008).

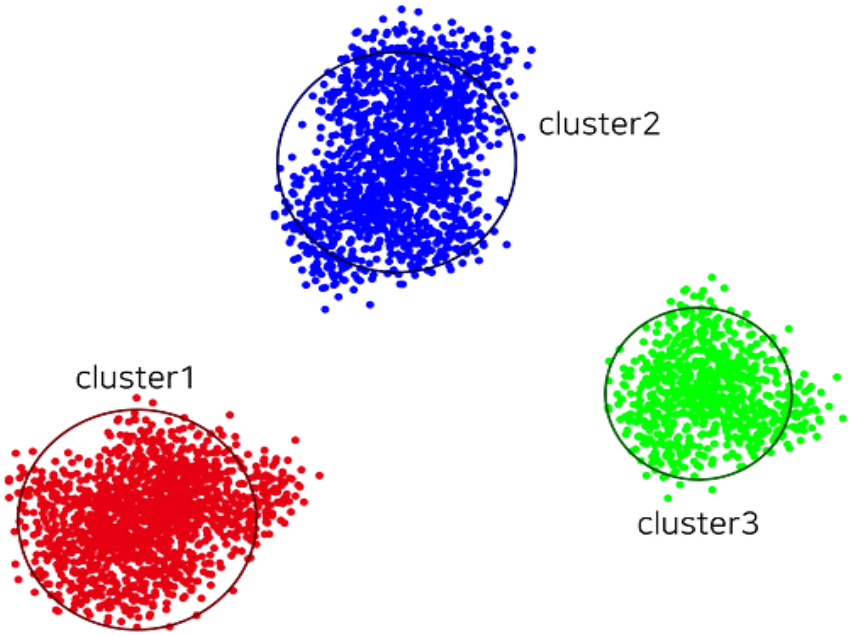


Fig. 1.0: Structures of K-mean clustering (Clarke, *et al.*, 2008)

\

ii. Principal Component Analysis (PCA): PCA is a method for reducing the dimensions of rotational spectrum data while maintaining the data's important variability. PCA makes the data easier to visualize and analyze by breaking it down into a new set of uncorrelated variables (principal components) (Provost, and Fawcett, 2013)

3. Deep Learning Algorithms

Due to their capacity to automatically develop hierarchical representations from raw data, deep learning algorithms, in particular Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have become increasingly prominent in rotational spectroscopy.

i. Convolutional Neural Networks (CNNs): CNNs are typically employed for image analysis tasks, but they may be modified to analyze rotational spectra by treating the spectra as one-dimensional signals. CNNs are effective for molecular identification and classification tasks because they can learn to recognize spectral characteristics and patterns. The structure of the convolutional Neural Network is shown in Fig. 2 (Liu, 2021).

i. Recurrent neural networks (RNNs): are effective for sequential data, such as time series or spectrum data. RNNs can recognize temporal patterns in the data and capture the sequential character of rotational spectra in rotational spectroscopy, allowing predictions based on previous spectral measurements.

2.1 Training, validation, and testing of ML models for rotational spectra analysis

To construct and assess machine learning (ML) models for rotating spectra analysis in the interstellar medium, training, validation,

and testing are essential procedures. By using these procedures, you can make sure that the models can accurately learn from the data, generalize to new data, and make accurate predictions for tasks like molecular identification and abundance calculations, among others (Sun, *et al.*, 2021). The procedure entails segmenting the rotational spectra data available into several subsets for training, validation, and testing as shown Fig. 3, each fulfilling a particular function in model creation and evaluation.

i. Training Data: The largest subset of rotational spectra utilized to train the ML model is the training dataset. It includes tagged rotational spectra examples and the target labels that correlate to them, such as molecular species or physical properties. By making iterative adjustments to its internal parameters during training, the model discovers the underlying patterns and relationships in the data (Zhao *et al.*, 2020). The objective is to maximize the model's performance by reducing the discrepancy between the predicted target labels and the actual target labels.

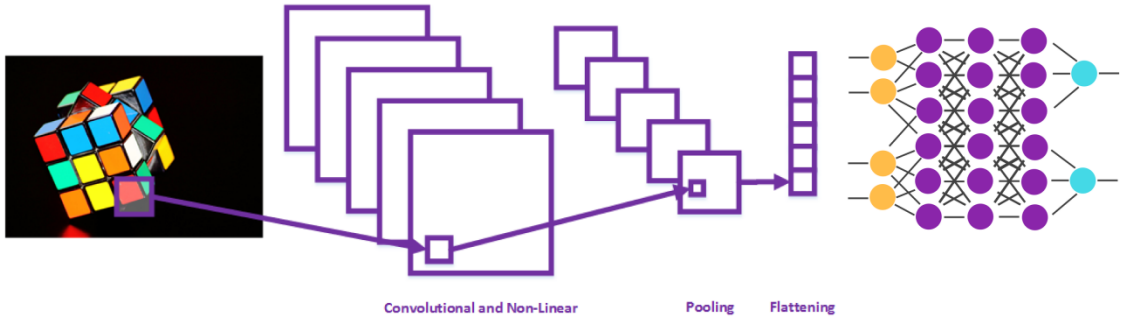


Fig. 2.0: Structure of Convolutional Neural Networks (Liu, 2021)

ii. Validation data: To adjust the hyperparameters of the ML model, a smaller subset of the rotational spectra is used for the validation dataset. Hyperparameters are settings that affect the behaviour and complexity of the model but are not learned during training (Zhang, *et al.*, 2022). These hyperparameters have a substantial impact on the model's performance, and hyperparameter tuning is frequently used to determine their ideal values. The validation dataset aids in evaluating how well the model works on untested data and guards against overfitting, which occurs when the model performs well on training data but badly on fresh, untested data (Neumann, *et al.*, 2005).

iii. Dataset for Testing: The testing dataset consists of a distinct subset of rotational spectra that is not used for training or hyperparameter tuning. It acts as an objective evaluation set to judge how well the model generalizes. The model is assessed on the testing data after training and hyperparameter tuning to determine an estimate of its performance in the actual world. This process enables scientists to assess how well the model will function on brand-new, unobserved rotating spectra in the ISM (Zhang, *et al.*, 2020).

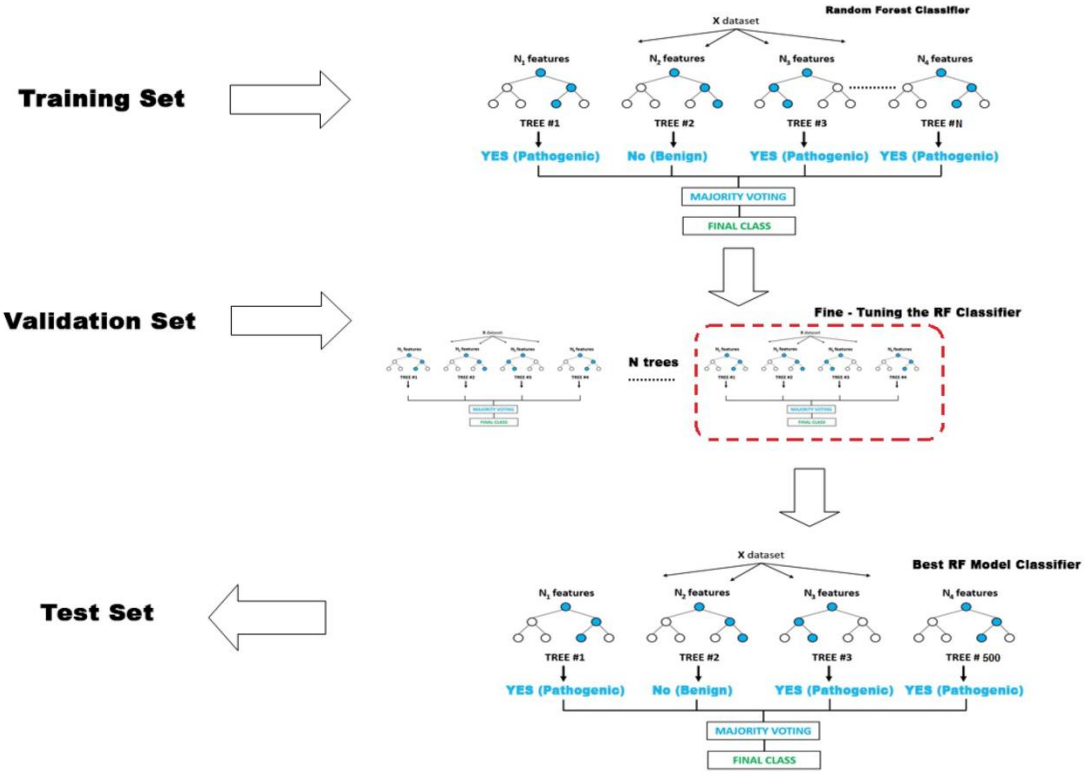


Fig. 3.0: Training, validation and testing set model (Zhang, *et al.*, 2020).

The following is a typical approach for developing, validating, and testing ML models for the study of rotational spectra:

i. Data preprocessing: To guarantee consistency and quality, it is crucial to preprocess the rotational spectra before splitting the data. Normalizing the spectra, eliminating noise, adjusting for instrument effects, and handling missing or incomplete data are some examples of preprocessing techniques.

ii. Data Splitting: Training, validation, and testing sets of the available rotational spectra data are created. Depending on the size of the dataset, the split percentages can change, but typical splits are 60–80% for training, 10–20% for validation, and 10–20% for testing (Tetko *et al.*, 1995).

iii. Model Training: Using the training set of data, the ML model is trained. By changing its internal parameters during training, the model learns to map the input rotational spectra to their corresponding target labels. Using optimization algorithms like gradient descent, the training procedure passes the data through the model, computing the prediction error, and adjusting the parameters (Krizhevsky *et al.*, 2012).

iv. Tuning of hyperparameters: The validation results are used to tune the hyperparameters. The combination that performs the best on the validation data is chosen after various combinations of hyperparameter values have been examined.

v. Model Evaluation: Using the testing data, the model's performance is assessed after it has been trained and its hyperparameters have been fine-tuned. Based on the testing data, the model produces predictions, and its generalization performance is evaluated by computing its accuracy, precision, recall, F1-score, or other pertinent metrics (Zhang *et al.*, 2017)

vi. Model Selection and Deployment: The best-performing ML model is chosen for rotating spectra analysis in the ISM based on the evaluation findings. The chosen model can then be used for tasks like identifying molecules, measuring

abundances, and estimating physical parameters.

vii. Model Monitoring and Iteration: As new data becomes available or when the underlying patterns in the ISM change, machine learning models may need periodic monitoring and retraining (Etim *et al.,* 2018a).

Researchers can create solid and trustworthy models for the investigation of rotating spectra in the interstellar medium by employing this methodical technique of training, validating, and testing ML models. Astronomers can learn more about the molecular make-up and physical characteristics of the ISM using the combination of rotational spectroscopy and machine learning, which advances our understanding of the intricate processes that form our universe (Claesen and Moor, 2015).

2.2 Sources of rotational spectra data in the ISM

The interstellar medium (ISM) has a variety of sources for rotational spectra data, including both ground- and space-based observatories with specialized equipment for spotting and measuring interstellar molecule rotational transitions. The rotational spectral lines of numerous compounds are prominent in the radio, microwave, and far-infrared frequency ranges, which are covered by these observatories. The important sources of data on rotating spectra in the ISM are listed below:

1. Radio telescopes: For observing rotating spectra in the ISM, radio telescopes are an essential tool. They operate at radio frequencies, which are excellent for detecting rotational transitions of various diatomic and polyatomic molecules. These frequencies are typically in the range of a few gigahertzes to several hundred gigahertzes (Raissi *et al.,* 2019). Several well-known radio telescopes are as follows:

i. The Atacama Large Millimeter Array (ALMA): is a group of radio telescopes that are situated in Chile as shown in Fig. 4 and have grown to be a premier resource for researching rotational spectra in the millimeter wavelength range. Molecular clouds and star-forming regions can be observed in great detail thanks to ALMA's exceptional sensitivity and resolution.

Fig. 4: Atacama Large Millimeter (Chen, *et al.*, 2006).

ii. Observatories in Space: Space-based observatories have the benefit of observing rotating spectra without air interference, giving them access to particular frequency ranges that are not accessible from the ground (Liu, *et al.*, 2017). Several well-known space observatories are:

a. Herschel Space Observatory: Launched in 2009 by the European Space Agency (ESA), Herschel was an observatory designed to study the far-infrared and submillimeter spectrums. It was essential for spotting rotational transitions in interstellar clouds and determining how chemicals affect star

formation as shown in Fig. 5 (Pathak, *et al.*, 2022).

Fig. 5: Herschel Space Observatory (Pathak, *et al.*, 2022).

b. James Webb Space Telescope (JWST): NASA will launch this cutting-edge space telescope. In addition to completing the observations made by earlier observatories like Spitzer, it will have improved infrared capabilities that will allow rotating spectra measurements in the mid-infrared area as seen in Fig. 6 (Yang, *et al.*, 2003).

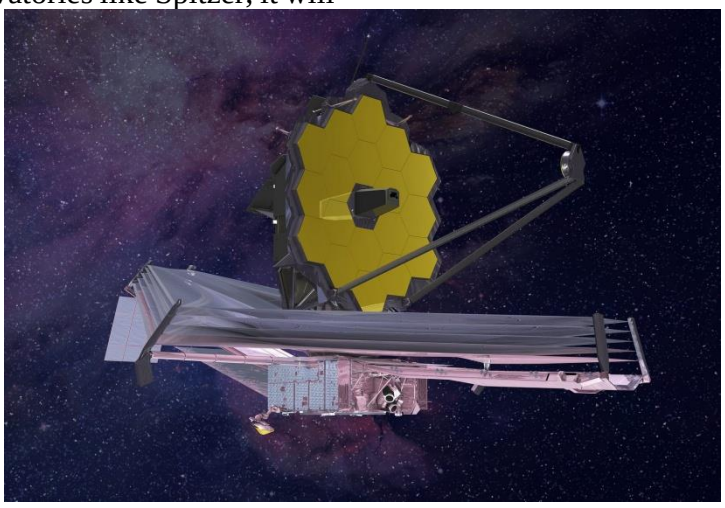


Fig. 6 James Webb Space Telescope (JWST) (Yang, *et al.*, 2003).

iii. Airborne Observations: Mounted on aircraft or high-altitude balloons, airborne observatories give users access to particular spectral bands without the atmospheric interference that ground-based telescopes face. To observe certain chemical transitions with precision, these platforms can be fitted with rotational spectroscopy equipment.

2. Suborbital telescopes: Suborbital telescopes are astronomical instruments sent into space aboard suborbital vehicles, either sounding rockets or high-altitude balloons. These platforms allow rotational spectra studies in certain frequency ranges but have very short observation intervals due to their ability to ascend to heights above a sizeable portion of the Earth's atmosphere.

3. Data Archives and Surveys: The scientific community has access to a variety of rotating spectra data from the ISM since they have been gathered and archived in public data repositories. The

Spectral Line Atlas of Interstellar Molecules (SLAIM) and the Cologne Database for Molecular Spectroscopy (CDMS) are two examples of such archives (Zou, *et al.,* 2007)

Researchers can access a multitude of rotational spectra data from these sources, as well as from the instruments and data archives that go along with them, to study the chemical make-up, physical properties, and dynamics of the interstellar medium. Astronomers improve our understanding of the universe by using data from several observatories and combining observations made at various frequencies to provide a more complete picture of the intricate chemistry and physics of the ISM (Luinge, et al., 1995).

2.3 Machine learning models for Molecular Identification and Abundance Measurements

The method of automating molecule identification in rotational spectra investigation of the interstellar medium (ISM) shown considerable promise when using machine learning (ML) models. Understanding the chemical composition and physical circumstances in various interstellar settings depends on being able to identify molecular species. Effective and precise chemical identification is made possible by ML models, which are excellent at identifying patterns and extracting pertinent information from complicated spectrum data. Using supervised learning algorithms for molecular identification is a typical strategy. A labeled dataset of rotational spectra and the matching chemical species is used to train the machine learning algorithm (Madden, and Ryder 2002). During training, the model links particular spectral features with particular molecules. Once trained, the model can identify unknown rotational spectra by inferring their molecular identities from their spectral signatures. For the analysis of rotating spectra, convolutional neural networks (CNNs) have been effectively modified. Rotational spectra can be represented as one-dimensional images, making CNNs particularly well-suited for applications involving images. The model gains the ability to identify distinctive spectrum transitions and patterns linked to various molecular species. Additionally, decision tree-based techniques for molecular identification, like Random Forests and Gradient Boosting, are frequently employed. These models enable the classification of rotational spectra into various chemical categories by recursively separating the spectral data based on distinct properties. Quantifying molecule abundances in the ISM is a crucial task for machine learning approaches. Understanding the chemical composition and the function of diverse molecules in various astrophysical processes requires an accurate assessment of molecular abundances (Giambagli, *et al.,* 2021). For abundance measurements, regression models, a kind of supervised learning method, are frequently utilized. The rotational spectra and accompanying chemical abundances from labeled datasets that include other observational or laboratory measurements are used to train these models. The molecule abundance values for fresh rotational spectra can be predicted thanks to the ML model's ability to learn the correlation between spectral properties and molecular abundances. Since Support Vector Regression (SVR) can handle continuous target variables, it is frequently used for abundance measurements. SVR determines a regression function that balances the trade-off between accuracy and generalization and best fits the data (Chowdhury, *et al.,* 2021). For example, the effectiveness of machine learning in molecular identification and abundance measurements in the ISM is demonstrated by several successful case studies:

Carbon monoxide (CO) and other diatomic molecules have been identified and their abundance in molecular clouds has been measured using ML models. These investigations have shed important light on the existence of particular molecular species and their contributions to the ISM's

chemistry. The identification and characterization of complex organic molecules (COMs) in the ISM have been made possible via ML-driven rotational spectroscopy. COMs have been discovered by ML models, including glycolaldehyde (CH_2OHCHO) as shown in Fig. 7, methyl formate (CH_3OCHO), and others, offering light on the complex organic chemistry taking place in space (Etim *et al.*, 2020). ML models have been used to calculate the molecular abundances in star-forming regions, which has aided in understanding their chemistry and the procedures that result in the creation of new stars (Etim *et al.*, 2018a).

Rotational spectra have been automatically classified into various molecular categories using machine learning techniques, which streamlines the molecular identification procedure and enables high-throughput study of massive datasets. These case studies show the effectiveness of machine learning in the interpretation of rotating spectra and its potential to fundamentally alter our knowledge of the chemistry and chemical composition of the interstellar medium. As a result of the efficient processing of enormous amounts of spectrum data made possible by machine learning and rotational spectroscopy, astrochemistry and interstellar studies have significantly advanced (Leonard, *et al.*, 2023).

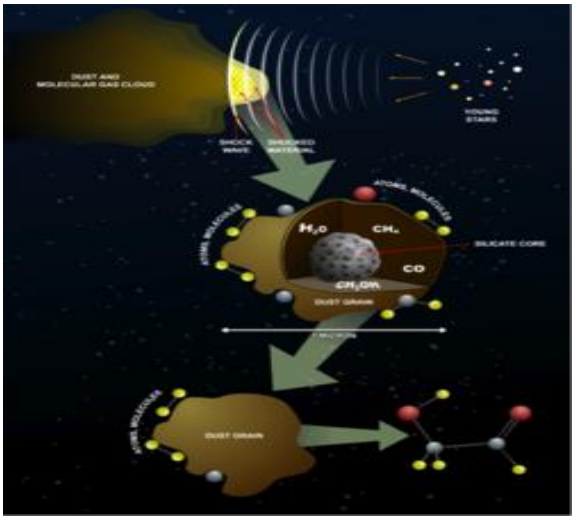


Fig. 7 Formation of glycolaldehyde in star dust (Wei *et al.*, 2023)

2.4 Machine learning in Probing Interstellar Clouds and Star-Forming Regions

Huge areas of gas and dust are found between stars in interstellar gas clouds, which are present in galaxies. These clouds are essential for comprehending the process of star formation because they act as the sites of star birth. Analysis of rotational spectra is essential for understanding the physical settings of these interstellar gas clouds (Pellegrino *et al.*, 2021). Astronomers can learn more about the temperature, density, and chemical make-up of the gas by watching the rotational transitions of different molecules.

i. Temperature estimation: The thermal energy of the molecules in the gas cloud is revealed by the rotational spectra. The line widths and intensities of rotational transitions are temperature-dependent. Astronomers can estimate the kinetic temperature of the gas by examining the patterns and intensities of the spectral lines, which gives them important details about the energy distribution and heating mechanisms in the cloud (Kempema *et al.*, 2021).

ii. Calculating Density: The gas density in interstellar clouds has a big impact on collisional processes and the speed of molecular interactions. The gas density can be calculated using the examination of rotational spectra, particularly for molecules having many rotational transitions. The spectral line intensities alter with density because of increased collisional broadening, allowing astronomers to determine the gas density in various parts of the cloud (Prasanta *et al.*, 2017).

iii. Machine learning techniques have become effective resources for deriving the physical parameters of interstellar gas clouds from rotating spectra. These models are capable of learning the intricate connections between spectral patterns and physical factors, allowing

for precise predictions and the extraction of important data.

iv. Temperature and Density Estimation: Labelled datasets comprising rotational spectra and matching temperature and density measurements from various observational techniques can be used to train supervised learning algorithms, such as regression models. The prediction of temperature and density values from rotating spectra is made possible by the machine learning model, which learns to link the spectral features to the physical parameters. For these tasks, neural networks and support vector regression (SVR) are frequently used (Etim *et al.,* 2022).

v. Kinematic Analysis: Machine learning approaches can be used to analyze the kinematics of interstellar clouds, including their movements and velocity structures. Convolutional neural networks and clustering algorithms, for example, can be utilized to discover spectral properties connected to various velocity components in the cloud. This makes it possible to analyze cloud dynamics like rotation, expansion, and infall motions, which sheds light on the cloud's general structure and evolution. Interstellar clouds contain complicated, dense regions called star-forming zones where new stars are formed. Studying the processes of star formation and the evolution of young stellar objects requires an understanding of the dynamics of these areas (Suwarno *et al.*, 2022).

vi. Analysis of Velocity Field: Star-forming areas' velocity fields can be examined using machine learning algorithms. The Doppler changes in the rotational spectra of the molecules in the cloud allow us to determine their velocities. Astronomers can map the overall kinematic structure of the star-forming region by applying ML algorithms to these velocity measurements to uncover patterns of gas motion, locate turbulent areas, and identify turbulent regions (Meduri and Nandanavanam, 2023).

vii. Protostellar Object Identification: ML models can be used to identify protostellar objects in star-forming regions. The rotational spectra of known protostars are included in labeled datasets that can be used to train these models. The machine learning algorithms pick up on the spectral cues that indicate protostellar emission, enabling the automatic detection of new protostellar candidates in the area (Sarkar *et al.,* 2019).

Machine learning can be used to investigate the chemical outflows and jets that come from newborn stars. The rotating spectra of these discharges have distinctive spectral characteristics. Astronomers can automatically find and examine these outflows by using ML algorithms, giving them insights into the launching mechanisms and feedback loops from young star objects. Rotational spectroscopy is an effective method for exploring star-forming regions and interstellar clouds. Astronomers can study star formation processes and the development of newborn stellar objects by analyzing rotational spectra to deduce the physical properties, kinematics, and dynamics of interstellar gas clouds. (Finkelmann *et al*., 2016). The combination of rotational spectroscopy and machine learning improves the effectiveness and precision of data processing, making it a potential strategy for expanding our understanding of the interstellar medium and its function in the genesis of stars and planetary systems.

2.5 Machine learning in Astrochemistry and Chemical Evolution

The study of chemical processes taking place in space, notably in the interstellar medium (ISM) and other celestial settings, is a multidisciplinary field known as astrochemistry. It is crucial to comprehend how molecules, atoms, and ions create and change throughout the cosmos.

Astrochemists now have a potent tool for exploring and analyzing the intricate chemical processes that form the ISM in the form of machine-learning (Bauer *et al.,* 2019). Some of the chemical processes in the ISM using machine learning in astrochemistry:

i. Chemical Reaction Networks: The ISM contains numerous chemical reaction networks that involve interactions between atoms, ions, and molecules. Different species are created, destroyed, and interconverted as a result of these reactions. The simulation and exploration of chemical reaction networks using machine learning models, such as neural networks and genetic algorithms, can reveal information on the reaction kinetics, reaction pathways, and the function of various species as intermediates or catalysts in chemical processes (Von, 2019). The analysis and interpretation of spectroscopic lines found in the rotational, vibrational, and electronic spectra of interstellar molecules is done using machine learning methods. Researchers can recognize and classify spectral lines, assisting in molecular identification, by training models using labeled datasets containing known spectral signatures of various molecular species. Also, molecule abundances from spectral data can be estimated using ML models, which are essential for comprehending the chemical composition of the ISM (Shinggu *et al.,* 2023).
ii. Construction of spectrum databases: Extensive spectrum databases of interstellar molecules are built and maintained using machine learning. These databases keep track of molecules' relevant physical and chemical properties as well as their rotational, vibrational, and electronic spectra. The enormous volumes of spectroscopic data gathered from numerous observatories and experiments are organized and cataloged with the aid of machine learning (ML) techniques like data mining and pattern recognition (Ertl 2019).
iii. Exoplanet Atmospheres: Beyond the ISM, machine learning is used in astrochemistry to explore exoplanet atmospheres. To predict the chemical compositions of exoplanet atmospheres from their spectroscopic signals, machine learning models can be trained using laboratory data and theoretical models. This helps in the analysis of the atmospheres of exoplanets as well as the hunt for biosignature chemicals or indications of habitability.
iv. Complex Molecule Detection: The detection of complex organic compounds (COMs) in the ISM has been made possible using machine learning techniques. COMs are multi-carbon atom compounds that are important for comprehending primordial chemistry and the origins of life. Based on spectral data, ML models may be taught to recognize the distinctive properties of COMs, making it easier to find them and investigate compounds of astrobiological significance (Grimme *et al.,* 2017).
v. Chemical Evolution Modelling: To simulate the chemical enrichment of the ISM over time, chemical evolution modeling employs machine learning approaches. These models take into account some variables, including grain-surface chemistry, gas-phase processes, and star nucleosynthesis. To better comprehend chemical development, ML algorithms can be used to refine the model parameters and compare the simulated results with observed abundances (Halgren, 1996).

2.5.1 Machine Learning-driven insights into the chemical evolution of the ISM

The abundance and distribution of atoms, ions, and molecules within the ISM are shaped by a variety of physical and chemical processes throughout cosmic timescales. Astronomers have never before been able to simulate and evaluate the chemical evolution of the ISM thanks to ML-driven investigations. The following are some

crucial areas where machine learning has improved our comprehension of the ISM's chemical evolution:

Large-scale chemical reaction networks, which control the ISM's chemical evolution, can be simulated by ML models. These networks contain tens of thousands of atomic and molecular species-based processes (Rasmussen 2004). Based on observational data, laboratory tests, and theoretical calculations, machine learning algorithms can optimize the rate coefficients of these reactions. Researchers can use ML to study how various reactions affect the chemical complexity of the ISM and model the evolution of abundances for a variety of species. In a variety of astrophysical contexts, ML can be used to estimate the initial chemical conditions of the ISM. For instance, ML models can infer the initial abundances of elements and molecules present during the early phases of star formation using measured abundances of molecules and ions. This reveals details about the chemistry of the molecular cloud that gives rise to stars (Ushie 2018a).

Supernovae, stellar winds, and nucleosynthetic activities all contribute to stellar feedback, which affects the chemical development of the ISM. Modeling the effect of star feedback on the chemical enrichment of the ISM can be aided by machine learning.

2.6 Machine learning approaches for identifying molecular ions in rotational spectra

As molecular ions are important participants in the chemical evolution of the interstellar medium (ISM) and are involved in a variety of interstellar processes, it is imperative in astrochemistry to identify them in rotational spectra. Identifying chemical ions from rotational spectra has been successfully automated and improved using machine-learning techniques (Pedregosa *et al.*, 2012). We go into great detail about the various machine learning approaches utilized for this purpose below:

2.6.1. Supervised learning

Identifying chemical ions in rotational spectra is frequently done using supervised learning. ML models are trained on labeled datasets in supervised learning, where rotational spectra are linked to the labels of the relevant chemical ions. The ML model gains the ability to identify spectral data patterns and features that are suggestive of particular chemical ions.

i. Classification Models: Support Vector Machines (SVM), Random Forests, and Neural Networks are some of the most popular classification algorithms. Based on the distinctive spectral fingerprints of various chemical ions, these models may learn to differentiate between them (Angulo*et al.*, 2022).
ii. Engineering of Features: In supervised learning, feature engineering is a crucial stage. Peak intensities, line widths, and frequency shifts are a few of the pertinent characteristics that researchers may derive from the rotational spectra. The ML model uses these features as input, which aids in its ability to anticipate outcomes correctly.

2.6.2 Unsupervised Learning

Unsupervised learning is a different machine learning method that can be used to detect chemical ions in rotational spectra, particularly when labeled data is scarce or nonexistent.

i. Clustering: Rotational spectra can be grouped into clusters using clustering methods like K-Means and DBSCAN based on similarities in their spectral patterns. It may be possible to identify molecular ions without knowing their names by comparing the spectra within the same cluster, which may belong to the same molecular ion. Spectral lines that do not follow the conventional patterns found in rotational spectra can be found using anomaly detection methods like Isolation Forest and One-Class SVM. These anomalies might be caused by rare species or unidentified molecular ions (Etim *et al.*, 2015).

ii. Pre-trained models and transfer learning: Transfer learning uses the information gained from one task to enhance performance on a related one. To locate molecular ions in rotational spectra, researchers can utilize pre-trained models that have already been trained on vast datasets of molecular spectra. These models' precision and effectiveness can be considerably increased by fine-tuning them using particular rotational spectrum data (Etim *et al.*, 2016).

2.6.3 Deep learning

Due to its capacity to automatically generate hierarchical representations from raw data, deep learning, a subset of machine learning has proven considerable potential in the identification of chemical ions from rotational spectra (Velasco, *et al.*, 2022).

i. Convolutional Neural Networks (CNNs): By considering the spectra as one-dimensional signals, CNNs—which are frequently used to identify patterns in images—can be modified for the analysis of rotational spectra. CNNs are useful for identifying chemical ions because they can learn to recognize distinctive patterns in the spectrum data.
ii. RNNs (recurrent neural networks): RNNs function well with sequential data, such as rotational spectra. RNNs can recognize distinctive patterns connected to certain chemical ions and capture temporal relationships between spectral properties (Mercier and Lennon, 2003).

The identification of new molecular species in the ISM may be sped up by machine learning techniques for locating molecule ions in rotating spectra. They facilitate the evaluation of enormous volumes of spectrum data and further knowledge of the chemistry and development of interstellar environments. However, it is essential to remember that the effectiveness of these ML techniques relies on the availability of diverse, high-quality training data as well as careful consideration of feature selection and data preprocessing. Our knowledge of the cosmic chemistry in the ISM is expected to grow as ML techniques develop and are combined with rotational spectroscopy (Bandos *et al.*, 2009). Astrochemistry as an interesting field of complex organic molecule (COM) and prebiotic chemistry in the interstellar medium (ISM) attempts to comprehend the creation and distribution of organic compounds in space. Complex organic molecules are fundamental components of life as we know it, and investigating their abundance in the ISM can shed light on prebiotic chemistry's potential as well as the universe's early history (Yuxuan, *et al.*, 2023). Astronomers can now examine enormous and complex datasets of rotational and vibrational spectra using machine learning (ML) techniques, and they may also investigate the complex chemistry of the ISM. ML methods in examining COMs and prebiotic chemistry in the ISM are given below:

i. Automated COMs Detection: Because there are so many spectral lines and there may be spectral overlaps, it can be difficult to find COMs in rotational and vibrational spectra. Clustering and pattern recognition are two ML-driven automated detection strategies that aid in locating and classifying spectral data connected to COMs. These algorithms can quickly sort through the data and identify possibilities for additional research.
ii. Analysis of Spectral Lines: Spectral lines in COMs are analyzed using ML methods to identify their locations, intensities, and other distinctive characteristics (Longqiang *et al.*, 2023) Researchers can anticipate the spectral properties of novel molecules and identify previously unidentified species by training ML models on the known spectral data of COMs.
iii. Modeing Chemical Reaction Networks: Prebiotic chemistry in the ISM involves intricate networks of chemical reactions involving atoms, ions, and molecules. ML-driven models may simulate and optimize chemical reaction networks,

giving information about the processes through which COMs are formed and the distribution of their abundance. These models explore the circumstances under which COMs can arise by taking into account elements like temperature, density, and radiation fields (Cheng *et al.*, 2023).

iv. Prediction of Prebiotic Molecules: It is possible to predict the presence of additional prebiotic compounds in the ISM using machine learning algorithms that have been trained on spectroscopic data of known prebiotic molecules. Astronomers can determine the possibility of prebiotic chemistry occurring in particular parts of the ISM by finding the spectrum signatures of these molecules (Zaw-Myo *et al.*, 2023).

v. Exoplanet Habitability: By examining the makeup of their atmospheres, ML approaches can be extended to analyze the possible habitability of exoplanets. ML models can evaluate the possibility of prebiotic chemistry on exoplanets by comparing spectroscopic data from the atmospheres of such planets with the known spectra of COMs in the ISM (Weimin *et al.*, 2023).

vi. Big Data Analysis: The investigation of COMs and primordial chemistry requires a substantial amount of observational data from both terrestrial and planetary telescopes. Astronomers can make sense of the enormous amount of spectral data by using ML methods, like as deep learning, to quickly evaluate and identify patterns from huge data (Onen *et al.*, 2000). Astronomers may now investigate the chemistry of the ISM in ways that were not previously conceivable by utilizing ML techniques. Our knowledge of the origins of life and the potential habitability of other planets has the potential to change as a result of ML-driven discoveries into complex organic compounds and primordial chemistry. As ML develops, its incorporation with astrochemistry has the prospect of revealing even more about the interesting chemistry taking on in the cosmos (Xia *et al.*, 2022).

2.7 Advanced ML Techniques in Rotational Spectroscopy

2.7.1 Deep learning applications in rotational spectra analysis

The analysis of rotational spectra in astrochemistry has been completely transformed by modern machine learning (ML) methods, particularly deep learning. Rotational spectroscopy applications using deep learning have produced encouraging results, delivering increased precision, effectiveness, and the capacity to automatically extract intricate features from spectrum data (Minjie *et al.*, 2022). We go into great detail about the numerous deep learning uses for rotating spectra analysis below:

1. Convolutional Neural Networks (CNNs) for Spectral Line Identification: Rotational spectroscopy uses CNNs, which were initially developed for image recognition, to identify one-dimensional spectral lines. CNNs are capable of automatically picking up on spectrum patterns and features related to particular chemical transitions. The models can precisely identify and classify spectral lines, enabling effective molecular identification, by training CNNs on labeled spectral datasets (Thereza *et al.*, 2022).
2. Denoising Spectral Data using Autoencoders: Autoencoders are unsupervised deep learning models that are employed in the reduction of dimensionality and the reconstruction of data. Autoencoders can be used in rotational spectroscopy to denoise spectral data, reducing noise and artefacts from observed spectra. This procedure improves the data's quality and increases the precision of following analysis activities (Osigbemhe *et al.*, 2022a; 2022b; 2022c).
3. Synthetic spectral generation using generative adversarial networks (GANs): To create synthetic spectrum data that

closely mimics actual observational spectra, GANs are used. Researchers can produce synthetic spectra with predetermined qualities by training GANs on datasets of observed spectral data. These artificial spectra provide useful datasets for ML model validation and comprehension of data constraints (Yajuan *et al.*, 2022).

4. Time-Series Spectral Analysis Using Recurrent Neural Networks (RNNs): RNNs are appropriate for time-series spectrum analysis because they excel at processing sequential data. RNNs are particularly effective for analyzing dynamic phenomena including outflows, turbulence, and changing spectral characteristics because they can capture temporal correlations in rotational spectra (Karteek *et al.*, 2022).
5. Attention Mechanisms for Spectral Feature Selection: Attention mechanisms are used to suppress noise or unimportant information and concentrate on pertinent spectral features. These processes enable the ML model to identify critical features for molecule identification and abundance measurements by giving different weights to various components of the spectral data.
6. Transfer Learning for Abundance Estimation: Transfer learning makes use of models that have already been pre-trained on large datasets to enhance the performance of ML models on related tasks using smaller datasets. Transfer learning can be used in rotational spectroscopy to refine pre-trained models from other spectroscopic datasets to estimate molecule abundances in the ISM (Wendy *et al.*, 2021).
7. Quantifying Prediction Uncertainty with Bayesian Deep Learning: Bayesian deep learning techniques make it possible to quantify prediction uncertainty. To accurately estimate molecule abundances and other physical characteristics from observed spectrum data in rotational spectroscopy while accounting for measurement errors and model uncertainties, uncertainty quantification is crucial (Daiguo *et al.*, 2021).
8. Deep Reinforcement Learning for Spectral Line Fitting: Deep reinforcement learning is a powerful tool for enhancing spectral line fitting to observed data. These models provide the ability to iteratively modify the spectral line profile parameters to reduce fitting errors, allowing for more precise modeling of rotational spectra.

Our understanding of the chemistry, physics, and dynamics of the interstellar medium has greatly improved as a result of the use of deep learning methods in rotating spectra research. These cutting-edge ML methods provide astronomers with strong tools to investigate and comprehend the rich and intricate information contained in rotating spectra, paving the way for advances in astrochemistry and our comprehension of the chemical evolution of the universe (Giambagli *et al.*, 2021).

2.8 Transfer learning and domain adaptation for interstellar rotational spectroscopy

In the context of interstellar rotational spectroscopy, transfer learning and domain adaptation are potent techniques because they allow for the efficient and effective use of knowledge from one spectral dataset (source domain) to enhance the analysis and modeling of another spectral dataset (target domain). These methods enable researchers to use data from other datasets to improve the analysis and interpretation of rotational spectra, which is particularly useful when working with sparse or limited data in the target domain (Kenya *et al.*, 2019).

i. Transfer learning for molecule Identification: Rotational spectra analysis frequently uses transfer learning for molecule identification. The goal is to use a large dataset of labeled rotational spectra from a source domain, where known molecular species are known, to pre-train a deep learning model, such as a convolutional neural network (CNN). The target domain, which might only include a

small amount of labeled data, is then given the information gained from the source domain. Even with a tiny target domain dataset, researchers can dramatically increase the accuracy and efficiency of molecular identification by fine-tuning the pre-trained model on the target domain (Xiangyu *et al.,* 2019).

ii. Domain Adaptation for Abundance Estimation: In abundance estimation tasks where the distribution of spectral data in the target domain (for example, a particular interstellar region) may differ from that in the source domain (for example, a laboratory or well-characterized region), domain adaptation techniques are used. The objective is to minimize the consequences of domain shift while adapting the model developed on the source domain to the target domain. By taking into account variations in observational settings and environmental factors, domain adaptation contributes to the robust estimation of molecule abundances in the ISM (Jesse, *et al.,* 2019).

iii. Data Augmentation for Improving Generalization: Data augmentation is a type of transfer learning that entails producing fake data to increase the dataset for the target domain. Researchers can construct extra training samples that capture various facets of the underlying data distribution by applying various changes to the current spectral data, such as adding noise, moving frequencies, or creating spectral variations. The capacity of ML models to generalize is improved through data augmentation, which improves their performance on spectrum data that has not yet been observed.

iv. Transfer Learning for Spectral Line Fitting: In rotational spectroscopy, extracting physical parameters like line intensities, line widths, and velocities requires the use of spectral line fitting, which is a crucial task (Stein, *et al.*, 2019). Models can be fine-tuned on the target domain using transfer learning after being pre-trained on a source domain with well-characterized spectral lines to speed up the fitting procedure. Even when the target domain data have various observational conditions or spectral resolutions, this method increases the accuracy and efficiency of spectral line fitting.

v. Domain Adaptation for Interstellar Cloud Analysis: Domain adaptation techniques can be used to modify machine learning (ML) models trained on one cloud to assess rotational spectra from different clouds in investigations of interstellar clouds with varying physical characteristics, such as temperature, density, and turbulence. Domain adaptation aids in capturing the unique qualities of each cloud while accounting for the differences in physical properties (Dai, *et al.,* 2019).

vi. Transferring Learned Representations: Transfer learning is not just about changing complete models; it can also mean transferring learned features or representations from a model that has already been trained. This may entail leveraging the intermediate layers of a deep learning model that has already been trained in rotational spectroscopy as feature extractors for other tasks like molecule classification, abundance estimate, or chemical evolution modeling.

Powerful methods that improve the analysis and interpretation of rotating spectra in the interstellar medium include transfer learning and domain adaptation. Researchers can improve the study of another spectral dataset by using information from one to create predictions, model physical parameters, and gain important insights into the intricate chemistry and evolution of the ISM. These methods are essential for maximizing the information we do have while also expanding our knowledge of the chemical complexity of the cosmos (Jochen, *et al.,* 2019).

2.9 Integrating ML with traditional spectral analysis in rotational spectroscopy

Rotational spectroscopy can produce better results and a deeper comprehension of the interstellar medium (ISM) by combining

machine learning (ML) with conventional spectral analysis techniques. Astronomers can extract more data, spot intricate characteristics, and increase the precision of molecule identification and abundance measurements by using machine learning (ML) techniques to supplement and improve conventional methods (Hiromasa *et al.*, 2018). Here are some examples of how rotational spectroscopy might incorporate ML with conventional spectral analysis techniques:

i. Automated Line Identification: The manual inspection and comparison of observable spectral lines with theoretical or experimental data are the traditional approaches for line identification. This process can be automated using ML tools like convolutional neural networks (CNNs), which learn the distinctive patterns of chemical transitions from labelled datasets. The ML model can accurately recognize and categorize spectral lines while requiring less human effort than previous methods (Robert and Sheridan 2013).
ii. Spectral denoising and deblending: Noise and blending of various chemical transitions can have an impact on spectral data from observational sensors. To eliminate noise and separate overlapping spectral lines, machine learning (ML) algorithms like autoencoders and deep denoising models can be used. Traditional approaches for line fitting and abundance estimate can function better and yield more trustworthy findings by deblending and denoising the data (Todd, *et al.*, 2012).
iii. Enhancing Spectral Line Fitting: ML methods can help to streamline the spectral line fitting procedure. ML models can learn the parameters that best represent the line profiles of various chemical transitions by being trained on simulated or existent spectral datasets. ML models can be used to predict uncertainty and errors in the study of rotational spectra. Bayesian deep learning, for example, can provide probabilistic predictions, allowing for the quantification of uncertainty in abundance measurements and other derived parameters (Bin, *et al.*, 2012).
iv. Building and Maintaining spectrum Databases: ML approaches can help build and maintain thorough spectrum databases of interstellar molecules. Astronomers will be able to access and use the data more easily for additional investigation by using ML algorithms to analyze and cluster spectral data from diverse sources.
v. Quality control and outlier detection: ML techniques can help find outliers or suspicious data points in the rotational spectra. This can assist astronomers in identifying anomalies, instrument artifacts, or odd spectral features that need additional research and quality assurance (Christian, *et al.*, 2010).

In rotational spectroscopy, combining ML with conventional spectrum analysis techniques has many advantages, such as automatic line identification, denoising, improved fitting, and increased generalization. These combined methods can result in a more precise and effective examination of rotating spectra in the ISM, expanding our knowledge of astrochemical processes and paving the way for fresh insights into the study of the interstellar medium (Robert, 2012).

Rotational spectroscopy in the interstellar medium (ISM) holds promising potential for the future, and it is anticipated to see new trends that will help us comprehend astrochemistry and the cosmos' chemical development. The following are some of the major rising trends and future opportunities in this industry:

i. Big Data and High-Resolution Spectroscopy: Larger and higher-resolution datasets of rotating spectra will be made available as observational capabilities advance. It will be necessary to modify machine learning approaches to effectively manage these enormous datasets. The development of scalable ML algorithms that can handle and interpret

massive data will be a key area of interest shortly. These algorithms will allow astronomers to investigate more intricate chemical networks and find unusual molecular species (Ushie, *et al.,* 2022).

ii. Deep Learning Architectures: ML techniques for rotational spectroscopy will continue to be led by deep learning. We will investigate advanced deep learning architectures for molecule identification, abundance estimates, and chemical evolution modeling, such as transformer networks and graph neural networks (John and Ryszard, 2008)

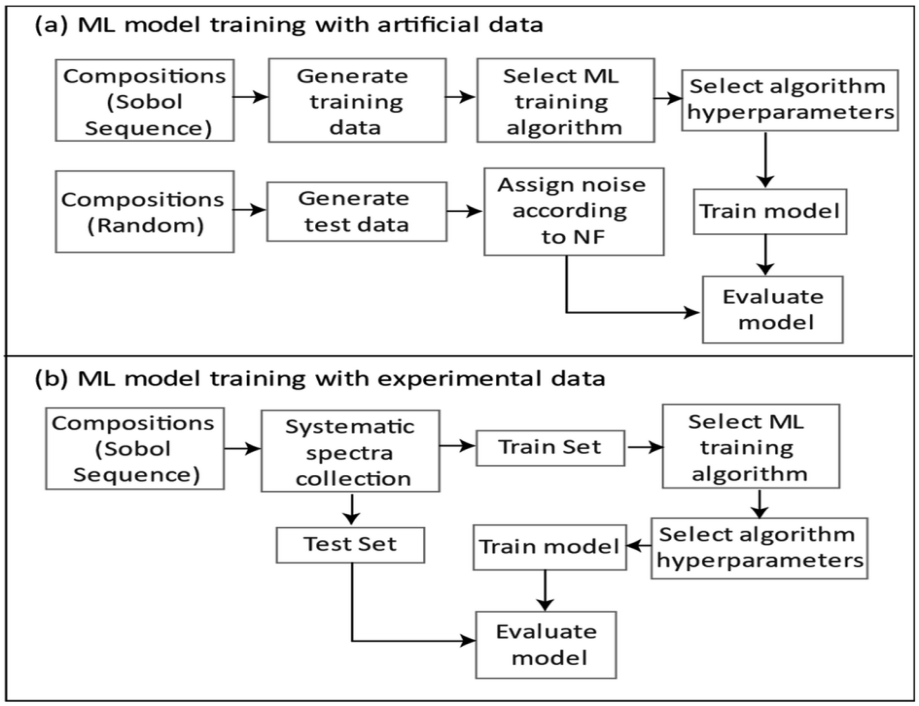


Fig. 8: A flowchart for the general approach to ML model development with (a) simulated generated and (b) experimentally collected data.

3.0 Prospects in machine learning for rotational spectroscopy in the ISM

iii. Bayesian Deep Learning for Uncertainty Quantification: To quantify uncertainty in abundance observations and model predictions, Bayesian deep learning techniques will be further integrated into rotational spectroscopy analysis. This will make it possible to interpret the results in a more solid and trustworthy manner (Qianyi and Jacqueline, 2009).

iv. Integrated Analytical Pipelines: A developing trend is the creation of integrated analytical pipelines that combine conventional techniques with spectrum analysis powered by machine learning. This will offer a thorough method to investigate rotating spectra and gain useful information more quickly.

v. Interpretable ML Models: There will be a greater demand for interpretable ML models as the use of ML models grows. Astronomers will be able to comprehend the physical foundation of their forecasts by developing models that offer concise explanations of their choices (Prachi, *et al.*, 2023).

vi. Hardware Acceleration and Quantum Computing: New developments in hardware acceleration and quantum computing may be used to increase the effectiveness and speed of machine learning (ML) algorithms for rotational spectroscopy and allow for real-time or almost real-time analysis of spectrum data.

With new trends emphasizing scalability, deep learning architectures, uncertainty quantification, and interdisciplinary

collaborations, the future of machine learning in rotational spectroscopy looks bright. These developments will enable astronomers to understand the chemical makeup and evolution of the ISM in greater detail and open up new vistas for the study of astrochemistry in the context of the constantly growing corpus of spectral data (Qinghua, *et al.,* 2023).

4.0 Conclusion

The study of the interstellar medium (ISM) and our understanding of astrochemistry have both been revolutionized by the development of machine learning as a transformative tool in rotational spectroscopy. Astronomers may now examine the intricate chemical processes that generate the ISM thanks to ML methods' previously unheard-of capabilities for analyzing, interpreting, and retrieving useful data from rotating spectra. Researchers can gain better outcomes, more efficiency, and deeper insights into cosmic chemistry by combining ML with conventional spectral analysis techniques. Scalable ML algorithms, deep learning architectures, domain adaptation techniques, and interdisciplinary collaborations will shape the next generation of rotational spectroscopy analysis, leading to a deeper understanding of astrochemical processes, the origins of complex molecules, and the chemical evolution of the universe. Rotational spectroscopy in the interstellar medium (ISM) has been significantly impacted by machine learning, altering how astronomers investigate the chemistry and development of the cosmos. By automating molecule identification, abundance estimate, and physical parameter determination from rotational spectra, ML approaches have decreased human error and improved accuracy. Researchers have been able to study larger, higher-resolution datasets thanks to the capabilities of ML algorithms to handle huge data, which has made it possible to explore more intricate chemical networks. The primordial chemistry of the ISM has been revealed thanks to the discovery and identification of complex organic molecules (COMs) and uncommon species using ML-driven rotational spectroscopy. The effectiveness and dependability of spectral line fitting and uncertainty quantification have been improved by the incorporation of ML with conventional spectral analysis techniques.

The importance of machine learning in advancing interstellar molecular analysis cannot be overemphasized. The obstacles faced by enormous volumes of observational data have been overcome by ML approaches, which have automated, efficient, and scalable rotational spectrum analysis. Deep learning architectures are used in ML models to enable autonomous learning and recognition of spectrum patterns, which improves molecular identification and abundance estimates. A more detailed knowledge of the chemical processes in the ISM is made possible by the uncertainty quantification offered by ML-driven Bayesian techniques, which improves the robustness of conclusions. Our understanding of astrochemical processes and the possibility of primordial chemistry in the universe has been revolutionized by the identification of complex and uncommon molecular species by ML-driven spectral analysis.

In the coming years, machine learning has the potential to completely transform interstellar spectroscopy. As machine learning methods develop, their combination with rotational spectroscopy will yield even more ground-breaking results. Future developments in the subject will be fuelled by trends including scalable algorithms for huge data handling, domain adaption methods for multi-region investigations, and deep learning architectures for exoplanet atmosphere study. Interprofessional partnerships involving astrophysicists, chemists, and ML specialists will promote creative solutions adapted to the special difficulties of rotational spectroscopy in the ISM. To unveil the mysteries of astrochemistry and provide a greater understanding of the chemical processes, machine learning must be able to extract useful insights from enormous amounts of spectral data. Astronomers are prepared to

enter a new era of discovery by utilizing the power of ML, unravelling the secrets of the interstellar medium and its part in the grand scheme of cosmic evolution.

5.0 References

Kirk S., John, K., Aaron, O., Daniel A. K., Karen, B., Dave A., Alicia W. & Ganesh, V. (2008). Practical Outcomes of Applying Ensemble Machine Learning Classifiers to High-Throughput Screening (HTS) Data Analysis and Screening. *Journal of Chemical Information and Modeling 48,* (11), 2196-2206. https://doi.org/10.1021/ci800164u

Agúndez M., Cernicharo J., De Vicente P. (2015). Discovery of HC3O+ in space: The chemistry of O-bearing species in TMC-1. *Astronomy & Astrophysics.* ;579:L10. 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), 77-82. DOI:10.7598/cst2018.1432.

Angulo A., Yang L., Aydil E. S, & Modestino M.A. (2022). Machine learning enhanced spectroscopic analysis: towards autonomous chemical mixture characterization for rapid process optimization. *Digital Discovery*.;1(1):35-44. doi:10.1039/D1DD00027F

Bandos, T. V., Bruzzone L. & G. Camps-Valls, (2009). "Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis," in IEEE Transactions on *Geoscience and Remote Sensing,* vol. 47, no. 3, pp. 862-873, doi: 10.1109/TGRS.2008.2005729.

Bauer, C. A., Schneider, G. & Göller, A. H. (2029). Gaussian process regression models for the prediction of hydrogen bond acceptor strengths. *Mol Inform* 38, 1800115, . doi.org/10.1002/minf.201800115

Bin C., Robert P., Sheridan, V. H. & Johannes H. Voigt.(2012). Comparison of Random Forest and Pipeline Pilot Naïve Bayes in Prospective QSAR Predictions. *Journal of Chemical Information and Modeling 52* (3) , 792-803. https://doi.org/10.1021/ci200615h

Cernicharo J., Gottlieb C.A, & Guelin M. (1991). Detection of HC5NH+ in TMC-1. The *Astrophysical Journal Letters*.;368: 39. doi:10.1086/185846.

Chen, R. X., Liu, S., Jin, J. Lin, & J. Liu, (2006). "Machine Learning for Drug Target Interaction Prediction,"**,** doi: 10.3390/molecules23092208. Available:www.mdpi.com/journal/molecules.

Chen, Z., Alexandre M., Weihua L., & 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*, vol. 140, 106683. https://doi.org/10.1016/j.ymssp.2020.106683.

Cheng F., Ye W., Richard G., Sudarshan K., Cheryl B., Patrick T. & Simone S. (2023). Prospective Validation of Machine Learning Algorithms for Absorption, Distribution, Metabolism, and Excretion Prediction: An Industrial Perspective. *Journal of Chemical Information and Modeling 63* (11) , 3263 3274. https://doi.org/10.1021/acs.jcim.3c00160

Chowdhury, M.A., Rice, T.E. & 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

Christian K., Bernd B., & Timothy C. (2010). Insolubility Classification with Accurate Prediction Probabilities Using

a MetaClassifier. *Journal of Chemical Information and Modeling* , *50* (3) , 404 414. https://doi.org/10.1021/ci900377e

Claesen M. & B. De Moor, (2015). "Hyperparameter search in machine learning," *arXiv preprint* arXiv:1502.02127, arXiv: 1502 . 02127. https://arxiv.org/abs/2302.05911

Clarke, R. H. W. Ressom, A. T. Wang, J. H. Xuan, M. C. Liu, E. A. Gehan, & Y. W. (2021). "The properties of high-dimensional data spaces: implications for exploring gene and protein expression data," *Nature Reviews Cancer*, vol. 8, no. 1, pp. 37–49, issn: 1474-175X. doi: 10.1038/nrc2294.

Dai F., Vladimir S., Andy L., Matthew P., & Robert P. S. (2019). Building Quantitative Structure–Activity Relationship Models Using Bayesian Additive Regression Trees. *Journal of Chemical Information and Modeling 59* (6) , 2642-2655. https://doi.org/10.1021/acs.jcim.9b00094

Daiguo D., Xiaowei C., Ruochi Z., Zengrong L., Xiaojian W. & Fengfeng Z. X. (2021). GraphBoost: Extracting Graph Neural Network-Based Features for a Better Prediction of Molecular Properties. *Journal of Chemical Information and Modeling 61* (6) , 2697-2705. https://doi.org/10.1021/acs.jcim.0c01489

Ertl P. (2021). An algorithm to identify functional groups in organic molecules. *J Cheminform* 9:1–7. https://doi.org/10.1186/s13321-017-0225-z

Etim E.E., & E. Arunan (2015). Rotational Spectroscopy and Interstellar Molecules. Planex News letter, 5 (2): 16-21. Invited mini-review article.

Etim, E. E., Adelagun, R.O.A; Andrew, C; Oladimeji, E. (2021). Optimizing the Searches for Interstellar Heterocycles. *Advances in Space Research Journal*, https://doi.org/10.1016/j.asr.2021.06.003.

Etim, E. E., (2015). Benchmark Studies on the Isomerization Enthalpies for Interstellar Molecular Species *J. Nig. Soc. Phys. Sci.* 5, 527. https://doi.org/10.46481/jnsps.2023.527

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

Etim, E. E., Gorai, P., Das,A., Chakrabarti, S. K & Arunan, E. (2018). Interstellar Hydrogen Bonding. *Advances in Space Research*, 61(11): 2870-2880, https://doi.org/10.1016/j.asr.2018.03.003.

Etim, E. E., Gorai, P., Das,A., Chakrabarti, S. K & Arunan, E. (2018a). Interstellar Hydrogen Bonding. *Advances in Space Research*, 61(11): 2870-2880, https://doi.org/10.1016/j.asr.2018.03.003.

Etim, E. E., Inyang, E. J., Ushie, O. A., Mbakara, I. E., Andrew, C. & Lawal., U. (2017). Is ESA Relationship the tool for searching for Interstellar Heterocycles? *FUW Trends in Science and Technology Journal*, 2(2): 665-678.

Etim, E. E., J. E. Asuquo, O. C. Ngana & 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):129 - 138.

Etim, E. E., Lawal, U., Andrew, C., & Udegbunam, I. S. (2018). Computational Studies on $C_3H_4N_2$ Isomers. *International Journal of Advanced Research in Chemical Science (IJARCS)* 5 (1) 29-40. DOI: http://dx.doi.org/10.20431/2349-0403.0501005

Etim, E. E., Magaji, A. & Ogofotha, G. O. (2022). Pipeline corrosion and its preventions in the oil and gas sector: a review. *International Journal of Environment and Bioenergy* 17 (1), 1-11.

Etim, E. E., Mbakara, I. E., Khanal, G. P., Inyang, E. J., Ukafia, O. P. & Sambo, I. F. (2017). Coupled Cluster Predictions of Spectroscopic Parameters for (Potential) Interstellar Protonated Species. *Elixir Computational Chemistry*, 111: 48818-48822.

Etim, E. E., Oko Emmanuel, Godwin, Ifiok F. Sambo, Sulaiman Adeoye Olagboye. (2020a). Quantum Chemical Studies on Furan and Its Isomers. *International Journal of Modern Chemistry*, 12(1): 77-98.

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., Oko Godwin E., Ogofotha, G. O. (2021). Quantum Chemical Studies on $C_4H_4N_2$ Isomeric Molecular Species. *J. Nig. Soc. Phys.* Sci. 3 429–445. https://doi.org/10.46481/jnsps.2021.282

Etim, E. E., Oko, G. E., Onen, A. I., Ushie, O. A., Andrew, C., Lawal., U, & Khanal, G. P. (2018). Computational Studies of Sulphur Trioxide (SO_3) and its Protonated Analogues. *J. Chem Soc. Nigeria*, 43 (2): 10 – 17.

Etim, E. E., Onen, A.I, Andrew,C., Lawal, U., Udegbunam, I. S. & Ushie, O. A. (2018). Computational Studies of C_5H_5N Isomers. *J. Chem Soc. Nigeria*, 43(2):1 – 9.

Etim, E. E., Onudibia, M. E., Asuquo, J. E., Ukafia, O. P., Andrew, C. & Ushie, O.A. (2017). Interstellar C_3S: Different Dipole Moment, Different Column Density, Same Astronomical Source. FUW Trends in Science and Technology Journal, 2 (1B): 574-577.

Etim, E. E., Sulaiman Adeoye Olagboye, Oko Emmanuel Godwin, Irene Mfoniso Atiatah. **(**2020b). Quantum Chemical Studies on Silicon Tetrafluoride and Its Protonated Analogues. *International Journal of Modern Chemistry*, 12(1): 26-45

Etim, E. E., Sulaiman A. O., Oko E. G., & Irene M. A. (2020). Quantum Chemical Studies on Silicon Tetrafluoride and Its Protonated Analogues. *International Journal of Modern Chemistry*, 12(1): 26-45

Etim, E. E., (2023). Benchmark Studies on the Isomerization Enthalpies for Interstellar Molecular Species. J. Nig. Soc. *Phys. Sci.* 5, 527. https://doi.org/10.46481/jnsps.2023.527 https://arxiv.org/abs/2302.05911

Etim, E.E, Akpan N. I., Ruth O. A., & Usman L. (2020). Deuterated Interstellar and Circumstellar Molecules: D/H Ratio and Dominant Formation Processes. *Indian Journal of Physics* https://doi.org/10.1007/s12648-020-01747-x

Etim, E.E. & E. Arunan. (2020). Interstellar Isomeric Species: Energy, Stability and Abundance Relationship. *European Physical Journal Plus*, 131:448. DOI 10.1140/epjp/i2016-16448-0

Etim, E.E, Prsanta Gorai, Ankan Das, & E. Arunan (2018). Theoretical investigation of interstellar C–C–O and C–O–C bonding backbone molecules. Astrophysics and Space Science, 363:6. DOI 10.1007/s10509-017-3226-5

Etim, E.E., Mbakara, I.E., Inyang, E.J., Ushie, O.A., Lawal, U. & Andrew, C. (2017). Spectroscopy of Linear Interstellar Carbon Chain Isotopologues: Meeting Experimental Accuracy. Trop. *J. Appl. Nat. Sci.*, 2(1): 11-16. *Doi: https://doi.org/10.25240/TJANS.2017.2.1.03*

Etim, E.E., Abah, B.S., Mbakara, I.E., Inyang, E.J., & Ukafia, O.P. (2017). Quantum Chemical Calculations on Silicon Monoxide (SiO) and its Protonated Analogues. . Trop. J. Appl. Nat. Sci., 2(1): 61-68. Doi: https://doi.org/10.25240/TJANS.2017.2.1.10

Etim, E.E., Onudibia, M. E., Asuquo, J. E., Ukafia, O. P., Andrew, C., Ushie, O.A.(2017). Interstellar C_3S: Different Dipole Moment, Different Column Density, Same Astronomical Source, *FUW Trends in Science and Technology Journal*, 2 (1B): 574-577.

Etim, E.E., Prsanta G., Ankan D., & E. Arunan. (2017). C_5H_9N Isomers: Pointers to Possible Branched Chain Interstellar Molecules. *European Physical Journal D,* 71:86. DOI: 10.1140/epjd/e2017-70611-3

Etim, E.E., Ugo Nweke-Maraizu., Samuel, H.S, **(**2023**).** Techniques used in corrosion inhibition studies: Modelling/computational techniques *Communications in Physical Sciences***.**

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): 394-403.

Etim, E.E., Ashu, H. A, Mbakara, I.E, Inyang, E. J., Ukafia, O. P, & Sambo, I. F. (2017b). Quantum Chemical Calculations on Oxygen Monofluoride (OF) and its Protonated Analogues: Comparison of Methods. *Elixir Computational Chemistry,*111: 48823-48827.

Finkelmann A.R., Göller A.H. & Schneider G. (2016). Robust molecular representations for modelling and design derived from atomic partial charges. *Chem Commun* 52:681–684. https://doi.org/10.1039/c5cc07887c

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

Grimme S., Bannwarth C. & Shushkov P. (2017). A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (*Z* = 1-86). J *Chem Theory Comput* 13:1989–2009. https://doi.org/10.1021/acs.jctc.7b00118

Gúndez M, Marcelino N, Cernicharo J. (2018). Tentative detection of HC5NH+ in TMC-1. *Astronomy & Astrophysics*. 861:L22. doi:10.1051/0004-6361/201833657.

Halgren T.A. (1996). Molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J Comput Chem* 17:553–586. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6%3c553:AIDJCC3%3e3.0.CO;2-T

Herbst E, & Klemperer W. (1973). The Formation and Depletion of Molecules in Dense Interstellar Clouds. *The Astrophysical Journal.*;185:505-533. doi:10.1086/152436.

Hiromasa Kaneko (2018). Discussion on Regression Methods Based on Ensemble Learning and Applicability Domains of Linear Submodels. *Journal of Chemical Information and Modeling 58* (2) 480-489. https://doi.org/10.1021/acs.jcim.7b00649

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*.;565:359-372.doi:10.1086/324588.

Huang W, Cheng J, Yang Y, Guo G. (2019). An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing*. 359:77-92. doi:10.1016/j.neucom.2019.05.052

Janet JP, Kulik HJ, Morency Y, Caucci MK. Machine Learning in Chemistry. ACS In Focus (Washington, DC: American Chemical Society). 2020.

Jesse G. Meyer, Shengchao Liu, Ian J. Miller, Joshua J. Coon, & Anthony Gitter (2019). Learning Drug Functions from Chemical Structures with Convolutional

Neural Networks and Random Forests. *Journal of Chemical Information and Modeling 59* (10), 4438 4449. https://doi.org/10.1021/acs.jcim.9 b00236

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. *Neurocomputing.* 272:619-628. doi:10.1016/j.neucom.2017.07.032

Jochen Sieg, 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) , 947-961. https://doi.org/10.1021/acs.jcim.8b 00712

John M. & Ryszard C. (2008). SAMFA: Simplifying Molecular Description for 3D-QSAR. *Journal of Chemical Information and Modeling 48* (6) , 1167 1173. https://doi.org/10.1021/ci800009 u

Karteek K. Bejagam, J. L., Carl N. Iverson, B. L. & Marrone, G. P. (2022). Machine Learning for Melting Temperature Predictions and Design in Polyhydroxyalkanoate-Based Biopolymers. *The Journal of Physical Chemistry B 126* (4), 934-945. https://doi.org/10.1021/acs.jpcb.1c 08354

Kempema, N. J., Sharpe, C., Wu, X., Shahabi, M., & Kubinski, D. (2021). Machine-Learning-Based Emission Models in Gasoline Powertrains Part 2: Virtual Carbon Monoxide. SAE *International Journal of Engines*, 16(6). https://doi.org/10.4271/ 03-16-06-0045

Kenya Tanaka, Kengo Hachiya, Wenjin Zhang, Kazunari Matsuda, & Yuhei Miyauchi.(2019). Machine-Learning Analysis to Predict the Exciton Valley Polarization Landscape of 2D Semiconductors. *ACS Nano 13* (11), 12687 12693. https://doi.org/10.1021/acsnano. 9b04220

Kim S, Chen J, & Cheng T. (2021). New data content and improved web interfaces. Nucleic Acids Res. 49:D1388. *PubChem* in 2021 doi:10.1093/nar/gkaa971.

Krizhevsky, A. I. Sutskever, & G. E. Hinton, (2012). "ImageNet Classification with Deep Convolutional Neural Networks," *Tech. Rep.*, pp. 1097–1105

Lee, K. L., Loomis R. A. & Burkhardt A.M. (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 Tan, Ooi Kiang Tan, Chun Chau Sze, Wilsonm & Wen Bin Goh. (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.02 74299

Liu, Youlin. (2021). Machine learning methods for spectral analysis. *Purdue University Graduate School. Thesis*. https://doi.org/10.25394/PGS.15040332 .v1

Liu,W. Z. Wang, X. Liu, N. Zeng, Y. Liu, & Alsaadi, F. E. (2017). "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11–26, Apr. issn: 18728286. doi: 10.1016/j.neucom.2016.12.038.

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

Luinge, H.J., van der Maas, J.H. & Visser, T. (1995). Partial least squares regression as a multivariate tool for the interpretation of infrared spectra.

Chemometrics and intelligent laboratory system, 28, 125–138. 15

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

Mattioda AL, Hudgins DM, Boersma C. (2020). The NASA Ames PAH IR Spectroscopic Database: The 2019 Release. *Astrophysical Journal Supplement Series.*;251(2):22. doi:10.3847/1538-4365/abb3db.

McGuire BA, Burkhardt AM, Loomis R. (2020). 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.*;900(1):L10. doi:10.3847/2041-8213/abafaf

McGuire B.A. (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

Meduri S, & Nandanavanam J. (2023). Prediction of hydrogen uptake of metal organic frameworks using explainable machine learning. *Energy and AI.* 12:100230. doi:10.1016/j.egyai.2023.100230

Mercier G. & M. Lennon, (2003). "Support vector machines for hyperspectral image classification with spectral-based kernels," IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. *Proceedings (IEEE Cat. No.03CH37477), Toulouse*, pp. 288-290 vol.1, doi: 10.1109/IGARSS.2003.1293752.

Minjie Mou, Ziqi Pan, Mingkun Lu, Huaicheng Sun, Yunxia Wang, Yongchao Luo, & Feng Zhu. (2002). Application of Machine Learning in Spatial Proteomics. *Journal of Chemical Information and Modeling 62* (23) , 5875-5895. https://doi.org/10.1021/acs.jcim.2c01161

Neumann, J. S. Christoph, S. Gabriele, C. Schn¨orr, & G. Steidl, (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.

Oliveira JCA, Frey J, Zhang SQ, Xu LC, Li X, Li SW, Hong X, & Ackermann L. (2022). When machine learning meets molecular synthesis. Trends in Chemistry. 4(10):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

Osigbemhe, I. G., Emmanuella E.O., HitlerLouis, E. M. Khan, E. E. Etim., Henry O. E., Onyinye J. I., Amoawe P. O.,& Faith O. (2022c). 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/S001945222200259X

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

Osigbemhe, I.G., Louis, H., Khan, E.M., Etim, E. E., Odey, D. O., Oviawe, A. P., Edet, H. O., & Obuye, F. (2022b).

Antibacterial Potential of 2-(-(2-Hydroxyphenyl)-methylidene)-amino)nicotinic Acid: Experimental, DFT Studies, and Molecular Docking Approach. *Appl Biochem Biotechnol* https://doi.org/10.1007/s12010-022-04054-9

Pathak, D.K., Kalita, S.K. & Bhattacharya, D. K. (2022). Hyperspectral image classification using support vector machine: a spectral spatial feature based approach. *Evol. Intel.* 15, 1809–1823. https://doi.org/10.1007/s12065-021-00591-0

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

Pellegrino, E., Jacques, C., Beaufils, N. (2021). Machine learning random forest for predicting oncosomatic variant NGS analysis. *Sci Rep* 11, 21820. https://doi.org/10.1038/s41598-021-01253-y

Prachi Garg, Scott Broderick, Baishakhi Mazumder. (2023). Machine learning-based accelerated design of fluorphlogopite glass ceramic chemistries with targeted hardness. *Journal of the American Ceramic Society 106* (8), 4654-4663. https://doi.org/10.1111/jace.19133

Prasanta Gorai, Ankan Das, Amaresh Das, Bhalamurugan Sivaraman, Emmanuel E. Etim, & Sandip K. (2017). A Search for Interstellar Monohydric Thiols. The Astrophysical Journal, 836, 70. DOI 10.3847/1538-4357/836/1/70

Provost, F. & T. Fawcett, (2017). "Data Science and its Relationship to Big Data and DataDriven Decision Making," *Big Data*, vol. 1, no. 1, pp. 51–59, issn: 2167647X. doi: 10.1089/big.2013.1508.

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

Qinghua Wang, Zhe Wang, Qirui Deng, Sutong Xiang, Rongfan Tang, Haiping Hao, & Huiyong Sun. (2023). Discriminating functional and non-functional nuclear-receptor ligands with a conformational selection-inspired machine learning algorithm. *Cell Reports Physical Science 4* (7), 101466. https://doi.org/10.1016/j.xcrp.2023.101466

Raissi, P. Perdikaris, & G. E. Karniadakis (2019). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, issn: 10902716. doi: 10.1016/j.jcp.2018.10.045.

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

Robert P. S. (2013). Using Random Forest To Model the Domain Applicability of Another Random Forest Model. *Journal of Chemical Information and Modeling 53* (11), 2837-2850. https://doi.org/10.1021/ci400482e

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, Etim, E.E., & Ugo Nweke-Maraizu., (2023). Understanding the experimental and computational

approach in characterizing intermolecular and intramolecular hydrogen bond, *Journal of Chemical Review*, 5(4), 439-465. https://doi.org/10.48309/JCR.2023.407989.1235

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

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

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:10.1063/5.0062659

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.

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

Sun, Y., Brockhauser, S., Hegedűs, P. (2021). Machine Learning Applied for Spectra Classification. In: Gervasi, O., *et al*. Computational Science and Its Applications – ICCSA 2021. ICCSA 2021. Lecture Notes in Computer Science, vol 12957. *Springer,* Cham. https://doi.org/10.1007/978-3-030-87013-3_5

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

Tetko, D. J. Livingstone, & A. I. Luik, (1995). "Neural Network Studies. 1. Comparison of Overfitting and Overtraining," *Journal of Chemical Information and Computer Sciences*, 35, 5, pp. 826–833, issn: 00952338. doi: 10.1021/ci00027a006.

Thereza A. Soares, Ariane Nunes-Alves, Angelica Mazzolari, Fiorella Ruggiu, Guo-Wei Wei, & Kenneth Merz. (2022). the (Re)-Evolution of Quantitative Structure–Activity Relationship (QSAR) Studies Propelled by the Surge of Machine Learning Methods. *Journal of Chemical Information and Modeling*, *62* (22) , 5317-5320. https://doi.org/10.1021/acs.jcim.2c01422

Todd M. Martin, Paul Harten, Douglas M. Young, Eugene N. Muratov, Alexander Golbraikh, Hao Zhu, & Alexander Tropsha. (2012). Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?. *Journal of Chemical Information and Modeling 52* (10) , 2570-2578. https://doi.org/10.1021/ci300338w

Ushie O.A, Etim, E.E, Adamu, H.M, Chindo, I.Y.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: 48828-48838.

Ushie, O. A., Etim, E. E., Onen, A. I., Andrew, C., Lawal, U., & Khanal, G.P., (2019). Computational Studies of β-amyrin

acetate ($C_{32}H_{52}O_2$) Detected in Methanol Leaf Extract of *Chrysophyllum albidium*. *J. Chem Soc. Nigeria*, Vol. 44, No. 3, pp 561 - 581.

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

Von Lilienfeld O. A. (2014). Quantum machine learning in chemical compound space. Angew Chemie Int Ed 57:4164–4169. https://doi.org/10.1002/anie.201709686

Wang X, Shen C, Xia M, Wang D, Zhu J, Zhu Z. (2020). Multi-scale deep intra-class transfer learning for bearing fault diagnosis. Reliability Engineering & System Safety. 202:107050. doi:10.1016/j.ress.2020.107050

Wei Ying Tan, Carol Hargreaves, Christopher Chen, & Saima Hilal, (2023). A Machine Learning Approach for Early Diagnosis of Cognitive Impairment Using Population-Based Data, *Journal of Alzheimer's Disease*, 91, 1, (449-461). https://doi.org/10.3233/JAD-220776

Weimin Zhu, Yi Zhang, Duancheng Zhao, Jianrong Xu, & Ling Wang. (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) , 43-55. https://doi.org/10.1021/acs.jcim.2c01099

Wendy L. Williams, Lingyu Zeng, Tobias Gensch, Matthew S. Sigman, Abigail G. Doyle, Eric V. Anslyn.(2021). The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Central Science 7* (10) , 1622-1637. https://doi.org/10.1021/acscentsci.1c00535

Xia Zhao, Yuhao Sun, Ruiqiu Zhang, Zhaoyang Chen, Yuqing Hua, Pei Zhang, Huizhu Guo, Xueyan Cui, Xin Huang, & Xiao Li.(2022). Machine Learning Modeling and Insights into the Structural Characteristics of Drug-Induced Neurotoxicity. *Journal of Chemical Information and Modeling 62* (23) , 6035-6045. https://doi.org/10.1021/acs.jcim.2c01131

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

Yajuan Shi, Jiang Wang, Qiang Wang, Qingzhu Jia, Fangyou Yan, Zheng-Hong Luo, & Yin-Ning Zhou. (2022). Supervised Machine Learning Algorithms for Predicting Rate Constants of Ozone Reaction with Micropollutants. *Industrial & Engineering Chemistry Research 61* (24) , 8359-8367. https://doi.org/10.1021/acs.iecr.1c04697

Yang, H., Griffiths, P.R. & Tate, J.D. (2003). Comparison of partial least squares regression and multi-layer neural networks for quantification of non-linear systems and application to gas phase fourier transfrom infrared spectra. *Analytica Chimica Acta*, 489, 125–136. 13

Yuxuan Hu, Qiuhan Ren, Xintong Liu, Liming Gao, Lecheng Xiao, & Wenying Yu. (2003). In Silico Prediction of Human Organ Toxicity via Artificial Intelligence Methods. *Chemical Research in Toxicology*, *36* (7), 1044-1054. https://doi.org/10.1021/acs.chemrestox.2c00411

Zaw-Myo Win, Allen M. Y. Cheong, W. & 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, & 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

Zhang, R. H. Xie, S. Cai, Y. Hu, G.-k. Liu, W. Hong, & Z.-q. Tian, (2020). "Transfer-learningbased Raman spectra identification," Journal of Raman Spectroscopy, vol. 51, no. 1, pp. 176–186, issn: 0377-0486. doi: 10.1002/jrs.5750.

Zhang,C. S. Bengio, M. Hardt, B. Recht, & O. Vinyals, (2017). "Understanding deep learning requires rethinking generalization," 5th International Conference on Learning Representations, ICLR - Conference Track Proceedings

Zhao Z, Li T, Wu J, Sun C, Wang S, Yan R, & Chen X. (2023). Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. ISA Transactions. 107:224-255. doi:10.1016/j.isatra.2020.08.010

Zou, T., Dou, Y., Mi, H., Ren, Y. & Ren, Y.(2007). Support vector regression for determination of component of compound oxytetracycline powder on near-infrared spectroscopy. *Analytical Biochemistry*, 355, 1–7. 14.

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

There is no source of external funding

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.