International Journal of Applied Pharmaceutics

ISSN- 0975-7058 Vol 16, Issue 3, 2024

Review Article

AUTOMATION IN ANALYTICAL CHEMISTRY: THE ROLE OF AI IN CHROMATOGRAPHY

SHISHIR KUMAR PRASAD, DIVEKAR KALPANA*

[1]Department of Pharmaceutical Chemistry, College of Pharmaceutical Sciences, Dayananda Sagar University, Bengaluru-562112, India
*Corresponding author: Kalpana Divekar; Email: kalpana-sps@dsu.edu.in

Received: 03 Jan 2024, Revised and Accepted: 09 Mar 2024

ABSTRACT

Artificial Intelligence (AI) has facilitated significant breakthroughs in drug discovery, the design of materials, and organic synthesis. The advancements in the latter group are especially remarkable due to the abilities of the latest computational methods (molecular design algorithms) that enable the exploration of extensive chemical spaces and enhance research in fields such as predicting molecule properties, designing molecules, retrosynthesis, predicting reaction conditions, and predicting reaction outcomes. A literary review was conducted following PRISMA guidelines. This study aimed to review existing data on the application of AI in separation chromatography. The evolution and utilization of AI in the pharmaceutical industry and its future aspects were articulated in this study. The utilization of AI can completely transform the field of chromatography analysis by facilitating expedited, more precise, and more effective data processing. By automating chromatography analysis, AI can enhance efficiency and minimize the potential for human mistakes. This advancement enables scientists to dedicate their efforts towards addressing intricate and demanding analytical issues. With the evolution of technology and the increasing adoption, we can anticipate more progress in chromatography analysis and analytical chemistry.

Keywords: Artificial intelligence, Chromatography, Pharmaceutical sciences, Separation and Purification

DOI: https://dx.doi.org/10.22159/ijap.2024v16i3.50290 Journal homepage: https://innovareacademics.in/journals/index.php/ijap

INTRODUCTION

Chromatography is a crucial separation technique for high purities in food, soil, water and pharmaceutical samples. Preparative chromatography is an established technology in biopharmaceutical manufacturing that is crucial in achieving high-quality separation and purification [1]. Historically, preparative chromatography was established through laborious laboratory experiments that required significant amounts of time and materials [2, 3]. However, the manual operational process is tedious, time-consuming, and susceptible to human error [4]. Further, more than these methods are needed to meet the current demands of efficient and customized pharmaceutical manufacturing within shorter timeframes for development [5]. Hence, there is a demand for more expeditious approaches to process development, which enhance the pharmaceutical industry's understanding of processes [6, 7]. A practical method to address this goal is to employ process modelling for the development. Models are constructed using mathematical equations that accurately represent physical and biological aspects of the process, which significantly enhances comprehension [8, 9]. Therefore, there is a need for rapid and precise methods to determine model parameters with minimal effort [10, 11].

AI is an emerging discipline focused on utilizing computer systems to solve issues by executing algorithms that imitate the cognitive functions of the human brain. AIis emulating human cognitive processes via machines, particularly computer systems. Most contemporary AI algorithms can establish connections between inputs and outputs, adjust their behaviour based on environmental cues, and subsequently make decisions, enhancing the likelihood of delivering precise responses. AI's primary benefit includes extracting significant and impartial information from datasets that are either extraordinarily huge or highly intricate, beyond the analytical capabilities of humans, and giving better and more precise predictions [12]. In addition, the increased processing capacity of computers, coupled with the advancement of robust algorithms and their availability through open-source platforms (such as APIs, frameworks, and training data), has facilitated the utilization of AIin various scientific domains [13-17]. AI has facilitated significant breakthroughs in drug discovery [18, 19], drug safety [20], the design of materials [21-23], and organic synthesis [24]. The advancements in the latter group are especially remarkable due to their ability to apply new computational methods (molecular design algorithms) that enable the exploration of extensive chemical spaces and enhance research in fields such as predicting molecule properties [25], designing molecules [26], retrosynthesis [27], predicting reaction conditions [28], and predicting reaction outcomes [29].

Additionally, it can facilitate the automation of analyzing extensive datasets, enhance the precision and uniformity of data analysis, detect and rank areas for further investigation, optimize the configuration of separation experiments, gain a deeper comprehension of the intricate connections among various components in a mixture, and expedite the advancement of novel separation techniques and technologies. Despite notable progress in education, the development of easy-to-use frameworks, and the availability of pre-trained neural networks, applying AI for analytical methods has yet to be thoroughly studied and remains poorly understood. The issues highlighted can be attributed to the discrepancy between the current academic training and the complex nature of modern algorithms used in data science. Utilizing machine learning algorithms with less intricate data can potentially overcome challenges in analytical chemistry [18, 21-24]. Preparative chromatography is a well-established biopharmaceutical manufacturing technology that provides high-quality separation and purification. An effective strategy to address this goal is employing process modelling for development. Models are constructed using mathematical representations of physical and biological phenomena to enhance understanding of the process. This leads to reduced development time and the ability to use model-based process control methods and optimization [30]. Given the context, a comprehensive literary analysis was undertaken to comprehend the incorporation of AI in separation chromatography.

Methodology

Reputed databases, such as PubMed and Google Scholar, were searched for the research articles published in chromatography and AI using a search strategy.

Search strategy

The secondary data was obtained using a search strategy developed with keywords such as “Separation chromatography” and "Analytical chemistry and artificial intelligence”. “Chromatography and AI” were used. The study approach adhered to the PRISMA principles, prioritising transparency and reproducibility. This ensured that every step, from the search strategy to data synthesis

and reporting, was clearly and thoroughly documented. The obtained publications underwent a thorough screening process, where their titles and abstracts were carefully examined to discover potentially relevant studies. The eligibility of full-text articles was evaluated based on predetermined criteria for inclusion and exclusion. Systematic data extraction involves gathering relevant information from each chosen study, such as study design, sample size, methodology, outcomes, and significant findings. The quality of the included papers, particularly in systematic reviews and meta-analyses, was evaluated using quality assessment methods or checklists. The inclusion criteria included studies comprised of research and review articles published between 2018 and 2023 and articles published in English Language and Peer-reviewed journals. Articles not available in full text or required payment, those published before the selected time, and in languages other than English were excluded.

Evolution of AI in chromatography

Over the last fifty years, analyzing and comparing large amounts of data from chromatography of natural and complex products, including essential oils, flavours and fragrances, pharmaceuticals, and petroleum products, have required manual methods due to unpredictable and nonlinear variations in retention times (RT). These datasets typically contain 20 to 1000 or more peaks. The issue has been resolved using software that employs neural algorithms, enabling the automated processing of intricate chromatograms [31].

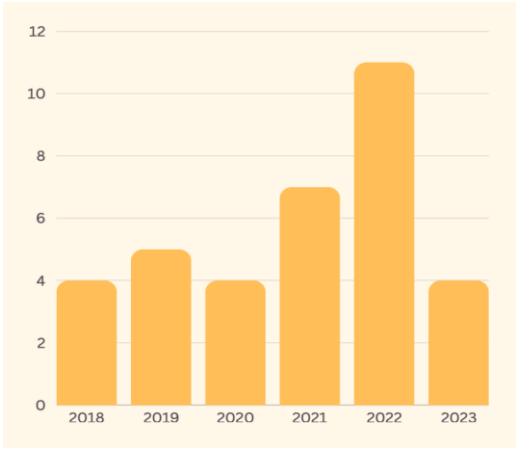

Fig. 1: Bar chart showing the number of publications related to the application of AI in chromatography
Source: Pubmed, keywords used: chromatography and AI, year: 2018-2023

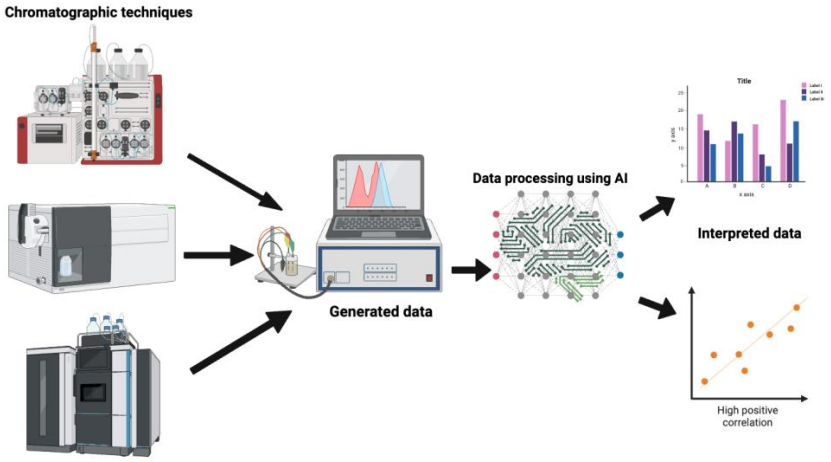

Fig. 2: Automatic data interpretation using AI in different chromatographic techniques
Source: Author generated

Evolution of high throughput process development (HTPD)

Introduction to the concept of model-based approaches. HTPD facilitated expedited and comprehensive screening of conditions, hence augmenting knowledge. Model-based HTPD has played a crucial role in the (bio)pharmaceutical business, specifically in chromatography. Chromatography is the primary method used for purifying protein subunit vaccines. The majority of vaccine purification methods rely on heuristics. For instance, when purifying hepatitis A virus from mammalian cell cultures, the initial step involves using low-cost anion-exchange chromatography to capture the product and eliminate significant impurities. The final step in the downstream process involves a polishing and desalting step using size-exclusion chromatography. Currently, there are several commercially available chromatographic mechanistic models software, such as GoSilico (now part of Cytivia, formerly known as ChromX), Aspen Chromatography, DelftChrom, CADET, and ChromaTech. While equilibrium and binding capabilities of membrane chromatography are typically restricted, membrane chromatography surpasses conventional packed bed chromatography in terms of productivity and bed utilization at high flow rates and short residence times [32].

A crucial chromatography element is the capacity to create a chromatograph using artificial intelligence. This allows for the automatic development of an analytical method for High-Performance Liquid Chromatography (HPLC). This means that the ideal composition of the mobile phase can be chosen from scouting tests, and the optimal operating parameters can be adjusted to achieve the desired analytical results. The chromatographic approach can also provide qualitative information about the peaks, particularly for an unidentified mixture, based on the chromatogram. To achieve this objective, essential equations for the retention of undissociated solutes, weak organic bases, weak acids, and amphoteric substances in liquid-solid chromatography were derived by considering variations in the composition of the mobile phase. The software RVPKLC-83 was designed to compute the parameters of these equations based on empirical data, and the accuracy of the equations was experimentally confirmed. The OMPCLC-83 programme was created to forecast the most favourable composition of the mobile phase. A HPLC equipment fitted with a fast-scanning Ultra Voilet (UV) detector system was utilized. The ChgrA-83 programme is currently under development to determine peak purity and detect unresolved peaks [33].

Preparative and process chromatography is a flexible procedure used to separate, purify, and refine a wide range of molecules, particularly those that are highly similar and complex, such as sugars, diastereomers, isomers, plant extracts, enantiomers, and rare earth metal ions. Bio-chromatography is an ever-expanding area of application that involves a wide range of complex molecules, including peptides, proteins, Monoclonal Antibodies (mAbs), fragments, Virus Like Particles (VLPs), and even mRNA vaccines. In addition to chemical diversity, separation processes encompass selective affinity ligands, hydrophobic contact, ion exchange, and mixed modes. Bio-chromatography ranges from a few kilogrammes to 100,000 tonnes per year, with column diameters typically ranging from 20 to 250 cm. Therefore, there is a requirement for a multifunctional and efficient tool that can be used for both process design and operation optimization, as well as process control [30].

Extensive experimentation is frequently required to ascertain the best solvent system in mixed solvent extraction. Centrifugal Partition Chromatography (CPC) is a liquid-liquid preparative chromatographic separation technique commonly employed in pharmaceutical and natural product purifications. It often necessitates using a solvent system including three or more components. To get the desired results, it is necessary to use multi-stage hybrid solutions with different components when dealing with complicated feedstocks, such as lignin depolymerization products. AIoffers significant potential for improving the complex process of selecting solvents. The training dataset can be obtained and organized from numerous sources, including academic publications, printed handbooks, and online archives. Machine learning can be utilized to create quantitative structure-property relationship (QSPR) models. These models establish a connection between the molecular structure of solvents and solutes and their physicochemical properties and extraction performance. They can predict the behaviour of untested combinations of solvents and solutes, providing valuable information on the most favourable solvents for specific extraction tasks [34].

AI approaches have significantly increased the accuracy of predicting retention in chromatographic procedures. AI can effectively analyze large data sets and simplify the identification and separation of substances. Multiple methodologies have been documented for the prediction of retention in various chromatographic techniques. Consistent findings have shown that deep learning models surpass linear machine learning models in terms of accuracy and efficacy, particularly in liquid and gas chromatography. The most commonly used method for predicting retention factors of various substances in thin-layer chromatography is Support Vector Machine-based neural networks. Cheminformatics, chemometrics, and hybrid techniques were utilized for the modelling and proved more dependable in retention prediction than traditional models. The Quantitative Structure Retention Relationship (QSRR) is a promising approach for predicting the retention of analytes in various chromatographic methods and identifying the optimal separation procedure. By integrating QSRR with AI-driven methodologies, these methods showcased the benefits of achieving more accurate retention predictions [35].

Used in the food industry

Recently, Phyto-control, a French company, has collaborated with Fujitsu (a Japanese company) to automate chromatographic techniques using AI. AI-enhanced chromatography offers quick sample analysis without any human error. Food products must be accurately analyzed to prevent contaminants from being introduced into the supply chain. If the contaminant-detecting process is weak, entire populations could be affected. As stated above, chromatography is one of the most reliable techniques for analyzing food samples, i.e., processed and raw products.

According to a recent World Health Organization (WHO) study, approximately 0.4 million people die annually due to the ingestion of contaminated food. Besides Phyto-control, Virtual Control, a Hong Kong-based company, has developed AI technology and machine learning-based software to provide analytical solutions in laboratory testing. In the case of Virtual Control, AI has been integrated with gas chromatography and mass spectrometry (GC/MS) platforms. The integrated product is ACIES, which has enhanced laboratory testing accuracy, efficiency, and productivity. This technology can benefit various industries, including agriculture, food, the environment, and applied materials [36].

A study conducted by Aghili *et al.* in 2022 presented a method to determine the odour characteristics of edible vegetable oils by analyzing their volatile aromas using an electronic olfactory device. This investigation collected odour profiles for eight different concentrations of sunflower and canola oil combined with sesame oil. The samples were analyzed simultaneously using GC-MS. The chemometric approaches, such as Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Support Vector Machine (SVM), Quadratic Discriminant Analysis (QDA), and Artificial Neural Networks (ANN), were used to analyze the data collected from the electronic nose. The electronic olfactory system effectively identified a subtle deception involving a mixture of 25% sunflower oil and 75% sesame oil despite the difficulty of detecting it through the gas chromatography-mass spectrometry method. This implies that the existing technique can detect and quantify occurrences of fraudulent activities related to edible oil to improve efficiency and monitoring and ensure the safety of eating edible vegetable oils [36].

Leite *et al.*, 2019, created two models, Radial Basis Function (RBF) and MIP, using MATLAB and integrated them into High-Performance Liquid Chromatography, which is employed to detect the lactose concentration following the absorption process. The RBF and MIP models offer superior efficiency, speed, and simplicity. When comparing RBF with MIP, it is shown that RBF requires more neurons in each layer for various tasks. However, RBF requires more hidden layers of neurons [37].

Further, Viejo *et al.*, 2022, developed two ANN models to evaluate the quality of beer and to forecast: i) the peak area (PA) of 17 distinct volatile aromatic compounds (Model 1) obtained through GC–MS, and ii) the intensity of ten sensory descriptors collected from a sensory session involving 12 trained panellists. The ANOVA results indicated significant disparities among the utilized samples, demonstrating the e-nose's ability to differentiate between them. The ANN models produced highly accurate results, with correlation values of R = 0.97 (Model 1) and R = 0.93 (Model 2) [36]. Another study by Warren-vega *et al.*, 2023 developed a novel AIapproach to explore the relationship between the physicochemical profile and colour gained during the 100% agave Tequila maturation phase. The findings demonstrate using artificial intelligence-based techniques as a supplementary approach for assessing quality control in aged beverages [39].

Uses in healthcare

The application of three AI techniques, namely Hammerstein-Wiener (HW), multilayer perceptron (MLP), and SVM, in qualitative properties prediction of an anti-Alzheimer agent using high-pressure liquid chromatography technique, demonstrated the promising capability of AI-based models in modelling the qualitative properties of the anti-Alzheimer agent. By observing the varying outputs of AI-based models over different time intervals, it became clear that combining the outputs of these models, known as ensembling, is necessary. Thus, the simple average ensemble and support vector machine ensemble (SVM-E) were utilized to improve the performance capabilities of the basic models [40].

In a recent publication, an AI-based solution for chromatographic data processing in the pharmaceutical industry was developed and applied using a "Digital by design" managerial approach. The authors, from Merck Serono (Italy) and Bosch Global Software Technologies Private Limited (India), proposed a potential GxP framework for using AI across the healthcare industry. The project was executed under a Digital Innovation Management framework, ensuring the involvement of stakeholders and decision-makers from proof of concept through proof of feasibility and finally to proof of value [41].

Further, De Vooght-Johnson, 2021, developed an artificial intelligence-based model for improving the prediction of peak perfection of an anti-oxidant Isoquercetin, where the RT was predicted using two individual nonlinear AI models, namely ANN and Adaptive Neuro-Fuzzy Inference System (ANFIS), along with Multi Linear Regression (MLR) Analysis, a traditional linear model.

In addition, the models were improved by using different ensemble techniques, specifically the simple average ensemble (SAE) and two types of ANFIS ensembles: the adaptive neuro-fuzzy inference system grid-partitioning ensemble (ANFIS-GPE) and the adaptive neuro-fuzzy inference system subclustering ensemble (ANFIS-SCE) [42].

Usman *et al.*, 2020 used four models in a different study to predict the RT and PA of isoquercitrin (extracted from various plant species) using HPLC: ANFIS, ANN, SVM, and MLR. The simulation uses the standard concentration, the composition of the mobile phases (MP-A and MP-B), and the pH as the input variables. The performance efficacy of the models was evaluated using the relative mean square error (RMSE), determination coefficient (DC), mean square error (MSE), and correlation coefficient (CC). The findings of this investigation demonstrate that all four models can precisely forecast the qualitative and quantitative attributes of the bioactive chemical. Through a predictive comparison, it was ascertained that M3 demonstrated the utmost prediction accuracy among the three models.

Further examination of the findings indicated that ANFIS–M3 outperformed the other models and is the most efficient model for forecasting PA. Nevertheless, ANN–M3 proved its value. It emerged as the superior model for tR simulation due to its high projected accuracy, establishing it as a dependable tool for qualitative and quantitative determination [43].

The results showed that the developed AI architecture could automate the chromatographic peak integration process with high accuracy and efficiency. The AI model learned the analytical variations in the chromatographic profiles related to the peak shapes, baseline drift, operator variations, and RT shifts. This allowed the algorithm to predict new chromatographic profiles' RT and peak shape, integrating them with high accuracy. The results of this study suggest that AI has the potential to revolutionize chromatographic peak integration. AI algorithms can be used to automate the process of peak integration, significantly improving the accuracy and efficiency of the process. This could lead to significant benefits for the biopharmaceutical industry, including improved patient safety and reduced cost.

Future of AI in Chromatography

Shortly, membrane materials with increased binding capabilities will be developed, potentially resolving the limitation on surface area per unit volume of resin. The biopharmaceutical business finds the progress in membrane chromatography technology highly intriguing. Chromatography uses two phases that do not mix to extract and separate components from mixtures. Combining HT methods with statistical or mathematical/thermodynamic models is a convenient approach for characterizing these systems [44].

Current process modelling methods have the challenge of requiring complex laboratory experiments to determine and validate model parameters. To have a broader range of uses in everyday project tasks, the technique must be more efficient and demand less exertion from individuals who are not experts in chromatography. Due to significant advancements in artificial intelligence, novel approaches have been developed to meet this requirement. Once the ANN has undergone training, it can be utilized to predict the isotherm parameters of unfamiliar components, provided that they fall within the limits of its training data. This provides the opportunity to significantly decrease the amount of experimental and computational work required, allowing those without expertise to complete model parameter estimations accurately.

Moreover, this approach provides the chance to obtain real-time parameter estimations for controlling the chromatographic process. This is possible because of its fast computation times and outstanding accuracy. Subsequent research will explore the expansion of the artificial neural network's capabilities to estimate more model parameters and isotherms. This will be done to achieve a model-based autonomous process operation in conjunction with the process analytical technology approaches [30].

Continuing with this pursuit results in the complete conversion of the plant into a digital format, sometimes called the digital twin. Therefore, there is a need for a rapid and precise method to determine model parameters with minimal effort [30]. When used with optimization algorithms, AI can aid in examining solution spaces with several dimensions and determine the most advantageous compromises between conflicting objectives, such as extraction efficiency, selectivity, environmental effect, and cost. The rapid progress in AI and its use in this crucial domain present promising prospects, including enhanced sorbent material design, refined extraction solvent selection, and optimized process operating conditions. Using artificial intelligence, scientists and engineers can fundamentally restructure separation processes, revolutionize several industries, and significantly contribute to achieving a more sustainable future [45].

Comparison between traditional and AI-based approach

AI has revolutionized chromatography by streamlining method development, improving data analysis, enhancing accuracy, and increasing the speed and efficiency of analysis. It has opened up new possibilities for various industries for advanced research, process optimization, and quality control. However, it is essential to note that while AI-based methods offer advantages, they should be validated and optimized using traditional methods to ensure reliability and accuracy [23-25].

Table 1: Comparison between traditional and AI-based approach

Aspect	Traditional analytical method development	Ai-added method development	Source of the data
Time	Time-consuming	Faster	[11]
Expertise required	High skilled analysts	Less expertise required	[11]
Trial and error	Iterative process	Reduced trial and error	[11]
Cost	Expensive	Cost-effective	[11]
Sample size	Limited sample size	Larger sample size	[11]
Optimization	Manual optimization	Automated optimization	[11]
Flexibility	Less flexible	More flexible	[11]
Data analysis	Manual interpretation	Automated data analysis	[11]
Accuracy	Human error-prone	Improved accuracy	[11]
Scalability	Limited scalability	Scalable	[11]

Use of AI models in different chromatographic techniques

ANN in process chromatography

In the chromatography process, numerous approaches for modelling and optimizing this technique have been proposed and put into practice [46]. Instead of solely focusing on established techniques, researchers investigate novel approaches, such as applying machine learning algorithms. These tools include the Partial Least Squares (PLS) method and ANN [47], which are recognized as universal approximators [48].

ANN have diverse applications

They can be predictors for process analytical technology (PAT) sensor data or substitute approaches such as the PLS algorithm. ANN can also be employed to ascertain model parameters rather than minimizing the sum of most minor square errors between

experimental and simulated datasets. In addition, ANNs can be employed to adjust the process and model parameters of the digital twin based on real-time operational data. ANN can also be applied in process models, including hybrid models. An all-encompassing framework for comprehensive process design and operation has already been devised [49, 50].

Many studies on ANNs utilized as regressors are employed for process optimization. An ANN is initially constructed to forecast specific values. Next, a conventional optimization algorithm is employed to optimize using the inputs and outputs of the ANN. Some examples include the studies conducted by Golubović *et al.* [51], Nagrath *et al.* [52], and Pirrung *et al.* [53]. In their study, Golubović *et al.* [51] employed ANNs to optimize the retention factor of mycophenolate mofetil (MFM) and its breakdown products. The ANN utilized buffer composition, flow rate, and column temperature to predict the retention factors. The dataset utilized was experimental and consisted of 33 samples. The dataset was based on the Central Composite Design (CCD), which can detect both linear and quadratic effects. The ANN surpassed the performance of the normal MLR and allowed for a decrease in the time of the experiment from 6.2 min to 5.2 min. The limits of this approach may include constraints on the area and information of the CCD design space, as well as the required experimental effort. Conducting these tests on a large-scale preparative chromatography may potentially be unfeasible. Conversely, this strategy does not require knowledge or modelling of the process. No prior knowledge of artificial neural networks is required, as the network was optimized through iterative experimentation.

Nagrath *et al.* [52] employed simulated data, as opposed to Golubović *et al.* [51], to address the limitations of experimental data. This approach is particularly advantageous when dealing with increasingly intricate tasks. According to Nagrath *et al.* [52], the increasing number of parameters has become a significant problem in optimization processes that rely purely on mechanical models, as suggested by Narayanan *et al.* [54]. The factors contributing to this issue are local minima and the total computational time. Hence, it is suggested that ANNs be employed to predict the target variable to optimize preparative chromatography. The proposed method for separating three components involved manipulating the simulations' gradient slope, feed load, flow rate, and column length. The impact of these factors on the desired outcomes, such as yield, production rate, and maximum concentration, was assessed to establish a training dataset.

Furthermore, an additional dataset was generated to account for less stringent conditions, mitigating the potential bias of zero productivity on the middle component due to its overlap with the left and right components. Moreover, this approach demonstrated favourable optimization outcomes and significantly profited from reduced calculation durations. However, prior knowledge is essential for modelling the process and generating appropriate data for training the ANN, as demonstrated by the additional dataset for the middle component. Furthermore, while the ANN may exhibit adaptability for the taught system, it necessitates a thorough retraining process when applied to other systems or different numbers of components. Hence, it is imperative to consider the additional exertion required for data formulation, creation, and ANN training to make a well-informed choice between the ANN methodology and the traditional approach.

Deep Neural Network (DNN) for chromatography

Deep learning is a crucial aspect of chemistry's most advanced AI technologies. It is relied upon due to its capacity to process vast amounts of data effectively. Hence, deep learning significantly advantages from extensive and varied datasets (mainly when the connections between the input and output data are intricate); thus, it must also be appropriate for our datasets. Aside from the abundance and diversity of data, it is crucial to consider the appropriate modification of the DNN method (such as type, topology, and hyperparameter values), as it directly impacts the accuracy. Various specialized adaptations have been effectively created for neural networks. Consequently, our attention is directed towards the following options that possess potential and are well-suited for our problem-solving objectives:

- A Feedforward Neural Network (FFNN) is a DNN that transmits information from input to output nodes without feedback loops. Although FFNNs are less complex than their descendants, they have proven effective in investigating and representing chemical space [55]. Nevertheless, FFNNs are specifically designed to acquire knowledge about the connections among independent variables, rendering them potentially inappropriate for addressing specific problems. Nevertheless, we will use FFNN in our experiments as the reference method to facilitate comparison with more advanced architectures.

- Convolutional neural networks (CNNs) [51] are a specific kind of deep learning network initially designed for image recognition. They are capable of identifying repetitive spatial patterns within the data. CNNs, or Convolutional Neural Networks, have become widely employed in various applications to handle data that exhibit grid patterns effectively. Graph CNNs have been utilized to forecast drug-target interactions [56]. Widely utilized CNN architectures VGG-19, ResNet512, AlexNet, and DenseNet-201[57,58,59,60] were employed to predict cytotoxicity for eight cancer cell lines.

- Moreover, CNNs have demonstrated their efficacy in accurately detecting steroids using deep learning RT modelling, as observed in gas chromatography analysis [61]. It is postulated that molecules possessing comparable chemical substructures or functional groups can exhibit similar behaviour and establish patterns that convolutional layers can identify and process. It elucidates the reasons why CNNs are appropriate for addressing our problems. Given that we are working with sequences of symbols, we utilize models that employ 1D convolutions.

- Long short-term memory (LSTM) neural networks, as described in reference [62], are well-suited for training on sequences because they incorporate a memory cell and feedback connections between cells. In contrast to Simple Recurrent Neural Networks (RNNs), LSTMs do not experience the issue of vanishing gradients. Consequently, the task of learning longer sequential dependencies is now easier. LSTMs employ supplementary internal mechanisms, known as gates, to regulate the flow of information and retain pertinent information as they analyze the sequence. The transmission of information from previous time points in a sequence holds great significance in understanding the meaning of chemical symbol sequences. For instance, the symbols O and H, when considered individually, are less informative compared to the compound (OH), which represents a hydroxyl group. Similarly, the symbol c1 in aromaticity notation is less significant than the complete sequence cc1ccccc1, where c1 indicates the beginning and end of an aromatic ring. Hence, the sequential nature of the input data justifies the decision to use LSTMs for our research.

DISCUSSION

Developing algorithms (of varying complexities) to analyze large volumes of data and extract meaningful information and patterns from even minute differences in individual measurements has been a prevalent trend in the literature. The primary catalyst for the early advancements in AI was likely its application in image recognition, vibrational spectroscopy, and mass spectrometry [63]. Analyzing chromatography data can be laborious, monotonous, and error-prone, posing difficulties in consistently obtaining precise results. AI can significantly enhance the chromatography analysis process. AI can carry out tasks that usually necessitate human intelligence, such as sensing, reasoning, and learning. Using AI methodologies, chromatography analysis can be mechanized, optimized, and enhanced in precision and effectiveness.

An application of AI in chromatography analysis involves the development of machine learning models capable of predicting the characteristics of unidentified samples using existing data. For instance, when employing chromatography equipment to segregate a blend of substances, a machine-learning model can be used to identify the peaks associated with each component and forecast its characteristics, such as molecular weight, polarity, and solubility.

Implementing this approach can substantially decrease the duration and exertion needed to analyze chromatography data, enhancing the outcomes' precision and dependability. AI can also be advantageous in chromatography analysis by facilitating the creation of automated systems that can enhance the efficiency of the chromatography process. AI algorithms can determine the most effective separation conditions, including the selection of stationary phase, mobile phase, and gradient elution parameters, to attain the utmost resolution and sensitivity. AI can be employed for real-time monitoring and control of the chromatographic process, allowing for timely alterations to the circumstances to achieve optimal performance.

Furthermore, AI may be utilized to create data processing algorithms capable of extracting significant insights from the vast volumes of chromatography data produced in contemporary analytical laboratories. AI algorithms can detect patterns and trends in data, such as the correlations between the characteristics of various chemicals or the alterations in the composition of a mixture over time. Such analysis can yield significant knowledge about the fundamental chemistry of the sample and aid in detecting possible contaminants or impurities.

Various AI algorithms, such as ANN, DNN models like FFNN's [55], CNNs [51, 56-61], LSTM [61], and RNNs [64], can be widely used to enhance the efficiency and accuracy of chromatography. Additional research is required to enhance the AI models to adapt them to specific chromatographic methods and substances utilized in analysis.

CONCLUSION

To summarise, AI can transform chromatography analysis by facilitating expedited, more precise, and more effective data processing. AI can streamline the chromatographic analysis process, saving time and minimizing the potential for human error. This enables scientists to dedicate their attention to intricate and demanding analytical issues. With the ongoing evolution and increasing technology adoption, we can anticipate additional progress in chromatography analysis and the broader realm of analytical chemistry.

FUNDING

Nil

AUTHORS CONTRIBUTIONS

SKP-Conceptualization, Data collection, reviewing, manuscript writing, DK-Planning, Supervision, Reviewing Manuscript. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTERESTS

Declared none

REFERENCES

1. Guiochon G. Preparative liquid chromatography. J Chromatogr A. 2002 Aug 2;965(1-2):129-61. doi: 10.1016/s0021-9673(01)01471-6, PMID 12236522.
2. Altenhoner U, Meurer M, Strube J, Schmidt Traub H. Parameter estimation for the simulation of liquid chromatography. J Chromatogr A. 1997 May 2;769(1):59-69. doi: 10.1016/S0021-9673(97)00173-8.
3. Eisele P, Killpack R. Ullmann's encyclopedia of industrial chemistry Wiley: Chichester, UK. Vol. 101. Propene; 2010.
4. Zobel Roos S, Schmidt A, Mestmacker F, Mouellef M, Huter M, Uhlenbrock L. Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? Processes. 2019 Feb 13;7(2):94. doi: 10.3390/pr7020094.
5. Helgers H, Schmidt A, Lohmann LJ, Vetter FL, Juckers A, Jensch C. Towards autonomous operation by advanced process control-process analytical technology for continuous biologics antibody manufacturing. Processes. 2021 Jan 18;9(1):172. doi: 10.3390/pr9010172.
6. European Medicines Agency. EU guidelines for good manufacturing practice for medicinal products for human and veterinary use-annex 15: Qualification and validation. Available online: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/Quality/Quality–Quality-design-qbd.
7. Guidance for industry PAT–A framework for innovative pharmaceutical development, manufacturing, and quality assurance. Washington DC: United States Department of Health and Human Services; 2004.
8. Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the Digital Twin: A systematic literature review. CIRP J Manuf Sci Technol. 2020 May 1;29:36-52. doi: 10.1016/j.cirpj.2020.02.002.
9. Schmidt A, Helgers H, Vetter FL, Juckers A, Strube J. Digital twin of mRNA-based SARS-COVID-19 vaccine manufacturing towards autonomous operation for improvements in speed, scale, robustness, flexibility and real-time release testing. Processes. 2021 Apr 23;9(5):748. doi: 10.3390/pr9050748.
10. Gerogiorgis DI, Castro Rodriguez D. A digital twin for process optimisation in pharmaceutical manufacturing. In: Computer aided chemical engineering. Vol. 50. Elsevier; 2021 Jan 1. doi: 10.1016/B978-0-323-88506-5.50041-3.
11. Wang G, Briskot T, Hahn T, Baumann P, Hubbuch J. Estimation of adsorption isotherm and mass transfer parameters in protein chromatography using artificial neural networks. J Chromatogr A. 2017 Mar 3;1487:211-7. doi: 10.1016/j.chroma.2017.01.068, PMID 28159368.
12. La Porta CA, Zapperi S. Explaining the dynamics of tumor aggressiveness: at the crossroads between biology, AIand complex systems. In: Seminars in cancer biology. Vol. 53. Academic Press; 2018.
13. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. Clin Radiol. 2019 May 1;74(5):357-66. doi: 10.1016/j.crad.2019.02.006, PMID 30898381.
14. Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci. 2019 Aug 1;40(8):592-604. doi: 10.1016/j.tips.2019.06.004, PMID 31320117.
15. Hippe Z. Problems in the application of artificial intelligence in analytical chemistry. Anal Chim Acta. 1983 Jan 1;150:11-21. doi: 10.1016/S0003-2670(00)85455-0.
16. Brady M. AIand robotics Artif. Intell. 1985.
17. Bahiraei M, Heshmatian S, Moayedi H. Artificial intelligence in the field of nanofluids: a review on applications and potential future directions. Powder Technol. 2019 Jul 15;353:276-301. doi: 10.1016/j.powtec.2019.05.034.
18. Richardson A, Signor BM, Lidbury BA, Badrick T. Clinical chemistry in higher dimensions: machine-learning and enhanced prediction from routine clinical chemistry data. Clin Biochem. 2016 Nov 1;49(16-17):1213-20. doi: 10.1016/j.clinbiochem.2016.07.013, PMID 27452181.
19. Kalayil NV, D'Souza SS, Khan SY, Paul P. AIin pharmacy drug design. Artif Intell. 2022;15(4).
20. Sujith T, Chakradhar T, Marpaka S, Sowmini K. Aspects of utilization and limitations of AIin drug safety. Asian J Pharm Clin Res. 2021;14(8):34-9.
21. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. Transl Res. 2018 Apr 1;194:36-55. doi: 10.1016/j.trsl.2017.12.004, PMID 29360430.
22. Nayak J, Vakula K, Dinesh P, Naik B, Pelusi D. Intelligent food processing: journey from artificial neural network to deep learning. Comput Sci Rev. 2020 Nov 1;38:100297. doi: 10.1016/j.cosrev.2020.100297.
23. Engkvist O, Norrby PO, Selmi N, Lam YH, Peng Z, Sherer EC. Computational prediction of chemical reactions: current status and outlook. Drug Discov Today. 2018 Jun 1;23(6):1203-18. doi: 10.1016/j.drudis.2018.02.014, PMID 29510217.
24. Panteleev J, Gao H, Jia L. Recent applications of machine learning in medicinal chemistry. Bioorg Med Chem Lett. 2018 Sep 15;28(17):2807-15. doi: 10.1016/j.bmcl.2018.06.046, PMID 30122222.
25. Szymanska E. Modern data science for analytical chemical data-a comprehensive review. Anal Chim Acta. 2018 Oct 22;1028:1-10. doi: 10.1016/j.aca.2018.05.038, PMID 29884345.

26. Jimenez Carvelo AM, Gonzalez Casado A, Bagur Gonzalez MG, Cuadros Rodriguez L. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity-a review. Food Res Int. 2019 Aug 1;122:25-39. doi: 10.1016/j.foodres.2019.03.063, PMID 31229078.
27. Reichenbach SE, Zini CA, Nicolli KP, Welke JE, Cordero C, Tao Q. Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition. J Chromatogr A. 2019 Jun 21;1595:158-67. doi: 10.1016/j.chroma.2019.02.027, PMID 30833025.
28. Zhong S, Zhang K, Wang D, Zhang H. Shedding light on "Black Box" machine learning models for predicting the reactivity of HO radicals toward organic compounds. Chem Eng J. 2021 Feb 1;405:126627. doi: 10.1016/j.cej.2020.126627.
29. Zhang Y, Li A, Deng B, Hughes KK. Data-driven predictive models for chemical durability of oxide glass under different chemical conditions. NPJ Mater Degrad. 2020 May 26;4(1):14. doi: 10.1038/s41529-020-0118-x.
30. Mouellef M, Vetter FL, Zobel Roos S, Strube J. Fast and versatile chromatography process design and operation optimization with the aid of artificial intelligence. Processes. 2021 Nov 25;9(12):2121. doi: 10.3390/pr9122121.
31. Raymond L. The introduction of AIto help in the analysis of the chromatographic big data. J Chromatogr Sep Tech. 2019;10.
32. Keulen D, Geldhof G, Le Bussy OL, Pabst M, Ottens M. Recent advances to accelerate purification process development: a review with a focus on vaccines. J Chromatogr A. 2022 Aug 2;1676:463195. doi: 10.1016/j.chroma.2022.463195, PMID 35749985.
33. Peichang L, Xiaoming L. Development of a high-performance liquid chromatograph with artificial intelligence. J Chromatogr A. 1984 May 25;292(1):169-88. doi: 10.1016/S0021-9673(01)96200-4.
34. Liu D, Sun N. Prospects of artificial intelligence in the development of sustainable separation processes. Front Sustain. 2023;4:1210209. doi: 10.3389/frsus.2023.1210209.
35. Singh YR, Shah DB, Maheshwari DG, Shah JS, Shah S. Advances in AI-driven retention prediction for different chromatographic techniques: unraveling the complexity. Crit Rev Anal Chem. 2023 Aug 31:1. doi: 10.1080/10408347.2023.2254379.
36. Aghili NS, Rasekh M, Karami H, Azizi V, Gancarz M. Detection of fraud in sesame oil with the help of artificial intelligence combined with chemometrics methods and chemical compounds characterization by gas chromatography-mass spectrometry. LWT. 2022 Sep 15;167:113863. doi: 10.1016/j.lwt.2022.113863.
37. Leite MS, Santos MA, Costa EM, Balieiro A, Lima ÁS, Sanchez OL. Modeling of milk lactose removal by column adsorption using artificial neural networks: MLP and RBF. Chem Ind Chem Eng Q. 2019;25(4):369-82. doi: 10.2298/CICEQ180606015L.
38. Viejo CG, Fuentes S, Godbole A, Widdicombe B, Unnithan RR. Development of a low-cost e-nose to assess aroma profiles: an aIapplication to assess beer quality. Sens Actuators B. 2020 Apr 1;308:127688.
39. Warren Vega WM, Contreras Atrisco ZA, Ramirez Quezada MF, Romero Cano LA. A novel approach of artificial intelligence for the study of the relation of physicochemical profile and color acquired by Tequila 100% agave in its maturation process. J Food Compos Anal. 2023 Oct 1;123:105533. doi: 10.1016/j.jfca.2023.105533.
40. Ghali UM, Usman AG, Chellube ZM, Degm MA, Hoti K. Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: an ensemble machine learning approach. SN Appl Sci. 2020 Nov;2:1-2.
41. Satwekar A, Panda A, Nandula P, Sripada S, Govindaraj R, Rossi M. Digital by design approach to develop a universal deep learning AI architecture for automatic chromatographic peak integration. Biotechnol Bioeng. 2023 Apr 22;120(7):1822-43. doi: 10.1002/bit.28406, PMID 37086414.
42. De Vooght Johnson R. AI produces peak prediction perfection. Wiley Anal Sci. 2021.
43. Usman AG, Isik S, Abba SI, Mericli F. Artificial intelligence-based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method. Turk J Chem. 2020;44(5):1339-51. doi: 10.3906/kim-2003-6, PMID 33488234.
44. Keulen D, Geldhof G, Le Bussy OL, Pabst M, Ottens M. Recent advances to accelerate purification process development: a review with a focus on vaccines. J Chromatogr A. 2022 Aug 2;1676:463195. doi: 10.1016/j.chroma.2022.463195, PMID 35749985.
45. Liu D, Sun N. Prospects of artificial intelligence in the development of sustainable separation processes. Front Sustain. 2023;4:1210209. doi: 10.3389/frsus.2023.1210209.
46. Beal LD, Hill DC, Martin RA, Hedengren JD. Gekko optimization suite. Processes. 2018 Jul 31;6(8):106. doi: 10.3390/pr6080106.
47. Mouellef M, Vetter FL, Zobel Roos S, Strube J. Fast and versatile chromatography process design and operation optimization with the aid of artificial intelligence. Processes. 2021 Nov 25;9(12):2121. doi: 10.3390/pr9122121.
48. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989 Jan 1;2(5):359-66. doi: 10.1016/0893-6080(89)90020-8.
49. Uhl A, Schmidt A, Hlawitschka MW, Strube J. Autonomous liquid–liquid extraction operation in biologics manufacturing with aid of a digital twin including process analytical technology. Processes. 2023 Feb 10;11(2):553. doi: 10.3390/pr11020553.
50. Zobel Roos S, Schmidt A, Mestmacker F, Mouellef M, Huter M, Uhlenbrock L. Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? Processes. 2019 Feb 13;7(2):94. doi: 10.3390/pr7020094.
51. Golubovic J, Protic A, Zecevic M, Otasevic B, Mikic M. Artificial neural networks modeling in ultra performance liquid chromatography method optimization of mycophenolate mofetil and its degradation products. J Chemom. 2014 Jul;28(7):567-74. doi: 10.1002/cem.2616.
52. Nagrath D, Messac A, Bequette BW, Cramer SM. A hybrid model framework for the optimization of preparative chromatographic processes. Biotechnol Prog. 2004;20(1):162-78. doi: 10.1021/bp034026g, PMID 14763840.
53. Pirrung SM, van der Wielen LAM, van Beckhoven RFWC, van de Sandt EJAX, Eppink MHM, Ottens M. Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks. Biotechnol Prog. 2017 May;33(3):696-707. doi: 10.1002/btpr.2435, PMID 28054462.
54. Natarajan V, Bequette BW, Cramer SM. Optimization of ion-exchange displacement separations. I. Validation of an iterative scheme and its use as a methods development tool. J Chromatogr A. 2000 Apr 21;876(1-2):51-62. doi: 10.1016/s0021-9673(00)00138-2, PMID 10823501.
55. Karlov DS, Sosnin S, Tetko IV, Fedorov MV. Chemical space exploration guided by deep neural networks. RSC Adv. 2019;9(9):5151-7. doi: 10.1039/c8ra10182e, PMID 35514634.
56. Indolia S, Goswami AK, Mishra SP, Asopa P. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia Comput Sci. 2018 Jan 1;132:679-88. doi: 10.1016/j.procs.2018.05.069.
57. Guerra E, de Lara J, Malizia A, Diaz P. Supporting user-oriented analysis for multi-view domain-specific visual languages. Inf Softw Technol. 2009 Apr 1;51(4):769-84. doi: 10.1016/j.infsof.2008.09.005.
58. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2016. p. 770-8. doi: 10.1109/CVPR.2016.90.
59. Krizhevsky A, Sutskever I, Hinton GE. Image net classification with deep convolutional neural networks. Commun ACM. 2017 May 24;60(6):84-90. doi: 10.1145/3065386.
60. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2261-9. doi: 10.1109/CVPR.2017.243.
61. Randazzo GM, Bileck A, Danani A, Vogt B, Groessl M. Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry. J Chromatogr A. 2020 Feb 8;1612:460661. doi: 10.1016/j.chroma.2019.460661, PMID 31708215.

62. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997 Nov 15;9(8):1735-80. doi: 10.1162/neco.1997.9.8.1735, PMID 9377276.
63. Ayres LB, Gomez FJV, Linton JR, Silva MF, Garcia CD. Taking the leap between analytical chemistry and artificial intelligence: a tutorial review. Anal Chim Acta. 2021 May 29;1161:338403. doi: 10.1016/j.aca.2021.338403, PMID 33896558.
64. de Araujo Padilha CE, de Araujo Padilha CA, de Santana Souza DF, de Oliveira JA, de Macedo GR, dos Santos ES. Recurrent neural network modeling applied to expanded bed adsorption chromatography of chitosanases produced by Paenibacillus ehimensis. Chem Eng Res Des. 2017 Jan 1;117:24-33. doi: 10.1016/j.cherd.2016.09.022.