International Journal of Pharmaceutical Research and Applications
Volume 10, Issue 4 July-Aug 2025, pp: 425-428 www.ijprajournal.com ISSN: 2456-4494

A Review on Artificial Intelligence in Gas Chromatography Ms Generation

Aneesh M A[(1*)] , Alan Antony[(1)] , Dr Uma Nath U[(2)],Dr . Prasobh G.R[(3)]
[1]*Student, Seventh Semester B.Pharm ,Sree Krishna college of pharmacy and Research centre , Parassala,Thiruvananthapuram ,Kerala ,India-695502*
[2]*Professor,Department of pharmaceutical chemistry and analysis,,Sree Krishna college of pharmacy Parassala ,Thiruvananthapuram,kerala.-695502*
[3,]*Principal, Sree Krishna college of pharmacy Parassala ,Thiruvananthapuram ,kerala -695502*

Date of Submission: 20-07-2025 Date of Acceptance: 30-07-2025

ABSTRACT: One particularly powerful technique, gas chromatography–mass spectrometry (GC-MS) was done by AI is widely used for its high sensitivity and precise quantification. Recent studies have leveraged AI to enhance GC-MS data analysis, particularly in pattern recognition and predictive modeling. However, the performance of such AI approaches heavily depends on access to comprehensive and diverse training datasets—something that can be limited when experimental data is sparse or constrained to specific chemicals(2,3). In this work, we explore the use of AI-driven generative approaches, including Variational Autoencoders (VAE) and Generative Adversarial Networks (GANs), to synthesize high-quality data for GC-MS analysis. We also conduct exploratory data analysis (EDA) to investigate the complexity and features of real-world GC-MS datasets, establishing a foundation for evaluating the effectiveness of generated data in downstream tasks.(4)

KEYWORDS:,Artificialintelligence, Chromatography, Gas Chromatography,(GC-MS), Exploitary data analysis.

I. INTRODUCTION

In Artificial intelligence (AI) has brought significant advancements to the chemical sciences, especially in substance detection and analysis(1)**.** One particularly powerful technique, gas chromatography–mass spectrometry (GC-MS), is widely used for its high sensitivity and precise quantification. Recent studies have leveraged AI to enhance GC-MS data analysis, particularly in pattern recognition and predictive modeling. However, the performance of such AI approaches heavily depends on access to comprehensive and diverse training datasets—something that can be limited when experimental data is sparse or constrained to specific chemicals(2,3)

Gas chromatography (GC), a powerful analytical technique, emerged in the early 1950s, largely credited to the work of A. T. James and A. J. P. Martin. Building upon earlier chromatography research, their 1952 work on gas-liquid partition chromatography revolutionized chemical analysis. This technique, now a mainstay in many industries, allows for the separation and analysis of complex mixtures of volatile and semi-volatile compounds

Gas chromatography (GC) separates volatile substances based on their different interactions with a stationary phase while being carried by a mobile gas phase (carrier gas). The principle involves vaporizing a sample and passing it through a column containing the stationary phase. Different components of the sample interact with the stationary phase to varying degrees, leading to different retention times and allowing for their separation

TYPES OF GAS CHROMATOGRAPHY

1) Gas-Liquid Chromatography (GLC)
a) Gaseous Mobile Phase
b) Liquid Stationary Phase

2) Gas-solid Chromatography (GSC)
a) Gaseous mobile phase
b) Solid Stationary phase(5)

International Journal of Pharmaceutical Research and Applications
Volume 10, Issue 4 July-Aug 2025, pp: 425-428 www.ijprajournal.com ISSN: 2456-4494

II. METHODOLOGY

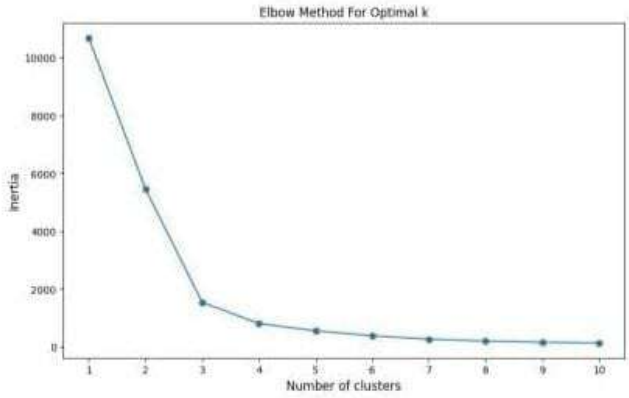

Fig. 2: Using the elbow method to find the optimal number of k-mean clusters.

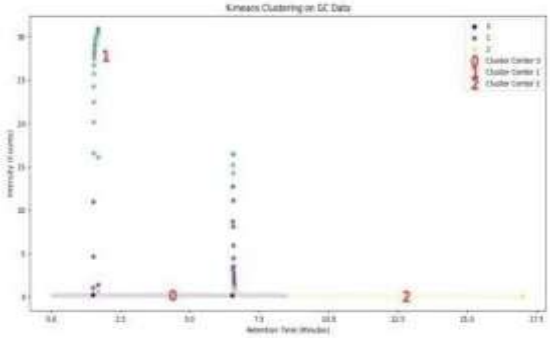

Fig. 3: Visualize K-means clustering on the original GC data feature space.

Exploratory Data Analysis

Before applying complex AI algorithms, we perform exploratory data analysis (EDA) on the GC-MS data to uncover underlying patterns, detect outliers, and understand the data's structure. The EDA encompasses various statistical and machine learning techniques as follows: • K-means clustering: K-means clustering effectively segments data into groups based on similarities as an unsupervised learning algorithm. It identifies patterns or (a)Generator of LSTM-CNN GAN. (b) Discriminator of LSTM-CNN GAN.

groups in chemical signatures, which may correspond to interactions with various types of CWA or different solvents [6]. Utilizing the elbow method as shown in Fig.2, we determine the optimal number of clusters to be K = 3. These clusters allow us to explore data patterns without prior labelling, which is crucial in analyzing the unstructured nature of GC-MS data. • Cluster Interpretation: Cluster 0 corresponds to baseline values in zones characterized by pronounced peaks, suggesting regions of lower chemical activity or baseline noise. Cluster 1 consists of data points with intensity values significantly above the baseline, indicative of prominent chemical peaks. Cluster 2 comprises baseline values where no significant peaks are identified, representing areas of no or minimal chemical activity.

AI-based Generative Models

We apply a pre processing technique to address the inherent scaler deviation of GC-MS data identified through EDA before models suitable for time series data augmentation based on previous studies [7]. The normalization pre processing technique of peak to overcome the deviation of GC-MS data scale is as follows and we

International Journal of Pharmaceutical Research and Applications
Volume 10, Issue 4 July-Aug 2025, pp: 425-428 www.ijprajournal.com ISSN: 2456-4494

use it for all generative models that are applied later. C. Variational Autoencoders (VAE) We utilize the pre processing and fully connected layer to apply the VAE which is known to be effective in learning time series data to GC-MS and construct it The model also balances the possibility of data with the complexity of latent representation and uses the Evidence Lower Bound (ELBO), which combines reconstruction loss with Kullback-Leibler divergence

III. EVALUATION

GC-MS Experimental Setup In this paper, data obtained by GC-MS analysis with an ethanol solvent are used for dimethyl methylphosphonate (DMMP), a similar agent of CWAs with limited acquisition. We conduct GC-MS experiments using standard protocols optimized for detection of CWA. We develop a customized 3D visualization tool to enhance interpretability of GC-MS data as an additional contribution to this study This facilitates the identification of patterns and correlations in GC-MS data that may not appear clearly in existing twodimensional images. Visual qualitative evaluation with these tools is complemented by quantitative measurements such as the time it takes to identify the peak and the accuracy of the analysis, which shows a significant improvement over traditional 2D visualization methods.

Performance evaluations We implement the proposed generative models to learn the dataset composed of GC and MS data using parameters as (a) In the epoch 100. (b) In the epoch 500. (c) In the epoch 900. (d) Generated synthetic data using LSTM-CNN GAN. Fig. 10: Visualization results of learning DMMP data using LSTM-CNN GAN. shown in Table I. Synthetic data generated through learning is compared with the original GC-MS data using the following metrics. • Pearson Correlation Coefficient (PCC) The PCC measures the linear relationship between the composite dataset and the features of the original dataset. It ranges from -1 to +1, and the closer to zero, the less correlation. • Jaccard Similarity (Jaccard) Jacquard similarity evaluates the similarity between two sets. This metric provides insight into how well the synthetic data captures the overall structure and pattern found in the original data. • Peak Difference (PD) The peak difference evaluates the diversity of spectral peaks using local maxima by comparing adjacent values [8]. When evaluating the generated synthetic datasets for GC and MS data, PCC recorded values close to zero for all models. This represents a weak linear relationship between the generated synthetic data and the original data as shown in Fig.8d, Fig.9d, Fig.10d. An increase in the number of peaks was observed in the synthetic dataset compared to the original data for all models. This trend suggests the propensity of models to generate new chemical signal peaks. This suggests that new features can be introduced into the data synthesized by AI-based generation models to secure diversity of data different from simple modulation. Overall, low PCC and Jaccard, and the process by which models have significantly difficulty learning GC-MS data,, indicates that the generative models studied for similar data forms are inadequate for the GC-MS field. Therefore, we propose an approach that fits the characteristics of the data, as shown in Fig.3, for the progress of GC-MS data generation beyond the limitations of these existing models. In order to properly learn GC-MS data analyzed through EDA

it is necessary to focus the input data on a specific part of the data. We implement and apply an attention mechanism layer for each model to selectively focus on important parts of GC-MS data [9]. As a result, it is confirmed that the performance is improved by 2% to 5%, but the influence of the attention mechanism on the model is insignificant. Through this evaluation, we propose that an innovative approach that goes beyond the paradigm of previously studied models is needed for successful generation of GC-MS data

IV. CONCLUSION

This study highlights the pressing need for generative model research focused on limited datasets, such as those involving chemical warfare agents (CWAs), and introduces a comprehensive framework for applying artificial intelligence to the synthesis of GC-MS (Gas Chromatography–Mass Spectrometry) data. Through machine learning-based exploratory data analysis (EDA), we investigate the intrinsic properties of GC-MS profiles using surrogate agents that resemble actual CWAs. Furthermore, we demonstrate the capability of existing generative models to produce synthetic GC-MS-like instances. However, our experimental results reveal significant challenges in training these models due to the complex patterns and high variability inherent in real GC-MS data.Ultimately, this work underscores the necessity for novel generative approaches specifically designed to accommodate the unique characteristics of GC-MS datasets. It represents a pivotal step toward scalable AI integration in chemical data synthesis, paving

International Journal of Pharmaceutical Research and Applications
Volume 10, Issue 4 July-Aug 2025, pp: 425-428 www.ijprajournal.com ISSN: 2456-4494

the way for broader applications of generative models in analytical chemistry and beyond.

REFERENCES

[1]. Nils Krone, Beverly A Hughes, Gareth G Lavery, Paul M Stewart, WiebkeArlt, and Cedric HL Shackleton. Gas chromatography/mass spectrometry (gc/ms) remains a preeminent discovery tool in clinical steroid investigations even in the era of fast liquid chromatography tandem mass spectrometry (lc/ms/ms). The Journal of steroid biochemistry and molecular biology, 121(3- 5):496–504, 2010

[2]. Giacomo Baccolo, Beatriz Quintanilla-Casas, Stefania Vichi, DillenAugustijn, and Rasmus Bro. From untargeted chemical profiling to peak tables–a fully automated ai driven approach to untargeted gc-ms. TrAC Trends in Analytical Chemistry, 145:116451, 2021

[3]. Yuji Roh, Geon Heo, and Steven Euijong Whang. A survey on data collection for machine learning: a big data-ai integration perspective.IEEETransactions on Knowledgeand Data Engineering, 33(4):1328–1347.

[4]. Carlos A Valdez, Roald N Leif, SaphonHok, and Bradley R Hart. Analysis of chemical warfare agents by gas chromatography-mass spectrometry: methods for their direct detection and derivatization approaches for the analysis of their degradation products. Reviews in Analytical Chemistry, 37(1):20170007, 2017Schechter, M.; and Levin, M., 1998, "Camless Engine," SAE Paper No. 960581

[5]. P.D Sethi; quantitative analysis of drugs ; 3rd edition

[6]. TR Noviandy, A Maulana, NR Sasmita, R Suhendra, GM Idroes, M Paristiowati, Z Helwani, E Yandri, S Rahimah, R Idroes, et al. The implementation of k-means clustering in kovats retention index on gas chromatography. In IOP Conference Series: Materials Science and Engineering, volume 1087, page 012051. IOP Publishing, 2021

[7]. Guillermo Iglesias, Edgar Talavera, Angel Gonz ´ alez- ´ Prieto, Alberto Mozo, and Sandra Gomez-Canaval. Data ´ augmentation techniques in time series domain: a survey and taxonomy. Neural Computing and Applications, 35(14):10123–10145, 2023.

[8]. DzmitryBahdanau, Kyunghyun Cho, and YoshuaBengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014

[9]. DzmitryBahdanau, Kyunghyun Cho, and YoshuaBengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014