Title

Machine learning and signal processing assisted differential mobility spectrometry (DMS) data analysis for chemical identification

Journal

Analytical Methods, 14(34)

ISSN

1759-9660

Authors

Chakraborty, Pranay
Rajapakse, Maneeshin Y
McCartney, Mitchell M
et al.

Publication Date

2022-09-01

DOI

10.1039/d2ay00723a

Peer reviewed

HHS Public Access

Author manuscript

Anal Methods. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Anal Methods. ; 14(34): 3315–3322. doi:10.1039/d2ay00723a.

Machine Learning and Signal Processing Assisted Differential Mobility Spectrometry (DMS) Data Analysis for Chemical Identification

Pranay Chakraborty[1], **Maneeshin Y. Rajapakse**[1,2], **Mitchell M. McCartney**[1,2,3], **Nicholas J. Kenyon**[2,3,4], **Cristina E. Davis**[1,2,3,*]

[1]Department of Mechanical and Aerospace Engineering, University of California Davis, Davis, CA, USA

[2]UC Davis Lung Center, One Shields Avenue, Davis, CA, USA

[3]VA Northern California Health Care System, 10535 Hospital Way, Mather, CA, USA

[4]Department of Internal Medicine, University of California Davis, Davis, CA, USA

Abstract

Differential mobility spectrometry (DMS)-based detectors are being widely studied to detect chemical warfare agents, explosives, chemicals, drugs and analyze volatile organic compounds (VOCs). The dispersion plots from DMS devices are complex to effectively analyze through visual inspection. In the current work, we adopted machine learning to differentiate pure chemicals and identify chemicals in a mixture. In particular, we observed the convolutional neural network algorithm exhibits excellent accuracy in differentiating chemicals in their pure forms while also identifying chemicals in a mixture. In addition, we propose and validate the magnitude-squared coherence (msc) between the DMS data of known chemical composition and that of an unknown sample can be sufficient to inspect the chemical composition of the unknown sample. We have shown that the msc-based chemical identification requires the least amount of experimental data as opposed to the machine learning approach.

1. Introduction

Differential mobility spectrometry (DMS) [1–7], capable of detecting both positive and negative ions, is one of the most critical technologies for developing small-scale and portable chemical sensing devices. Due to the less constrained operating conditions (no vacuum or temperature constraint) [7], DMS gets attention for various applications, e.g., diagnosing pulmonary diseases and respiratory infections, detecting explosives and

*Corresponding author: cedavis@ucdavis.edu.

narcotics, and chemical signature detection [2, 7]. DMS is a powerful tool that distinguishes ions based on their divergent mobilities under oscillating high and low electric fields as applied field strength alters ion mobility. Several research groups have focused on understanding the working phenomena [1], improving data analysis techniques [8], and enhancing the efficiency of DMS [3, 6]. From a practical point of view, DMS has been studied for detecting biological and chemical agents [9], ion filtration [5], water [10] and air [11] quality monitoring, disease diagnostics [12–15], and differentiating chemicals [16]. Significant ongoing research focuses on formulating effective DMS data analysis strategies [1, 17, 18]. There are also reports showing a wide range of applications of hyphenated DMS devices, e.g., gas chromatography-DMS [19], DMS-mass spectrometry [5], DMS-DMS [20] for rapid separation and detection. Moreover, DMS has a simple design structure making it possible to be miniaturized [5, 6].

The dual separation voltage (SV) and compensation voltage (CV) scanning in DMS creates 3-dimensional dispersion plots. These dispersion plots portray the fingerprints of ions that originate as the samples flow through the DMS device under a changing electric field. Several research groups have focused on single chemical identification [9, 21] and developing predictive models [17] to analyze the dispersion plots from DMS devices. In particular, the machine learning strategy has been greatly successful in interpreting and differentiating DMS data, extracting fundamental chemical properties, and even optimizing DMS performance. For example, there is a report showing the applicability of machine learning for determining molecular properties from the DMS data [22]. It is worth noting that machine learning has also been applied to successfully predict the DMS behavior [23], which would be helpful for optimizing the DMS operating parameters. Machine learning-assisted differentiation of DMS data has been studied for differentiating chemicals from food [24, 25], controlled chemical sources [17], and diagnosing diseases [12–14]. In a pioneering work, Li et al. [26] adopted deep learning to identify specific substances in the DMS spectra and reported excellent accuracy.

The complex dispersion plots are almost impossible to analyze and differentiate based on pure human visual inspection. Robust software with data visualization, data processing, and user-friendly machine learning implementation capability will significantly help chemists, engineers, and researchers to analyze the data quickly. In Peirano et al. [27], our team published custom software, AnalyzeIMS (“AIMS”) to visualize raw DMS data, apply noise reduction techniques like Savitzky-Golay smoothing and baseline removal and included principal components analysis (PCA) and partial least squares (PLS) analysis. Building from that work, Rajapakse et al. [17] conducted the partial least squared discriminant analysis to distinguish chemicals in their pure form and mixtures. Later on, Yeap et al. [18] adopted machine vision methods, natural language processing, and machine learning algorithms to identify chemicals from complex mixtures. AIMS was most recently updated to include automated peak detection algorithms for GC-DMS data and included random forests classification algorithms [28]. Both Rajapakshe et al. [17] and Yeap et al. [18] reported that incorporating chemical mixture data in training samples increases the classification accuracy of machine learning models while identifying the chemical composition of a mixture.

In the current work, we implement the convolutional neural network (CNN) algorithm to our AIMS software and study its ability to distinguish chemicals and accurately identify chemicals in a mixture. We observe the CNN model outperforms all the previously reported algorithms in differentiating DMS data of pure chemicals and identifying chemical compounds of a mixture. We also demonstrate that magnitude-squared coherence (msc) can be an excellent tool to check the chemical composition of a sample. As opposed to the machine learning approach, the msc-based chemical composition detection requires a very small set of DMS data for chemical classification.

2. Methods

Dispersion plot data of individual chemicals and mixtures were generated from a MicroAnalyzer DMS (Sionex Corp, Bedford, MA) with a 5 mCi, ^{63}Ni ionization source. We adopted the same experimental methodology as reported by Rajapakshe et al. [17] for collecting the dispersion plots of the pure chemicals and chemical mixtures. We introduced the chemical samples with a concentration of 500 ppb into the inlet of the DMS device. We conducted a series of dilution processes to feed the DMS with chemicals of a concentration of 500 ppb. Initially, a stock concentration of 1000 ppm was prepared by injecting the required volume of analyte into a Tedlar bag (SKC Tedlar Sample bag, SKC Inc. Eight Four, PA) with 3 L of nitrogen balance gas. Chemicals were allowed to equilibrate within the bags at room temperature for 10 minutes. A volume of the stock solution was injected into a second Tedlar bag with 3 L nitrogen balance with water content of 1 ppm to dilute the concentration to 100 ppm, The 100 ppm gas sample were loaded into a 1 μL glass syringe (Hamilton Co., Reno, NV), and we used a syringe pump to inject the samples to dilution nitrogen gas flow directed to the inlet of the DMS such that the chemicals entering the DMS have a concentration of 500 ppb. Chemical standards were obtained from MilliporeSigma (Missouri, USA). We used ultra-pure nitrogen with ~1 ppm humidity as the carrier gas (200 mL/min) for the device, and the carrier gas temperature was maintained at 80 °C.

We varied the compensation voltage (separation voltage) from −10 V to +30 V (500 V to 1490 V) range to scan the dispersion plots of the samples. The separation voltage of the MicroAnalyzer ranges from 0 to 1500 V with a frequency of 1.2 ± 0.1 MHz, and the duty cycle of the separation field is 30%. The filter gap of the micro analyzer 500 μm and the maximum field strength is about 110 Townsends. The separation voltage scanning was performed with a step size of 10 V, and the number of separation voltage steps was 100. The CV step was 0.4 V, and the CV scanning was performed with step duration, scan duration, and step settle time of 10 ms, 1000 ms, and 3 ms, respectively. We only considered the DMS dispersion data of positive polarity; however, the approach of this study can be easily modified or extended for negative polarity.

For differentiating DMS dispersion plots of chemicals, we implement the convolutional neural network (CNN) within an updated AnalyzeIMS (AIMS) [27] software for differentiating the DMS plots of different chemicals. In the current work, we complete all the data visualization and analyses with this updated version of AIMS that operates in MATLAB r2021a. Before training and testing the CNN model, we perform noise reduction (through Savitzky-Golay filter) and baseline correction (through asymmetric least squares

smoothing) of the raw DMS data with the AIMS software [27]. The basic working principle of CNN algorithm can be found in any machine learning textbook. However, we briefly discuss the most important features of the CNN algorithm for a self-contained discussion.

A convolutional neural network (CNN) [29] employs mathematical convolution kernels or filters that slide along input features to map the original features. The output of the convolutional and pooling layers in a CNN network are flattened and fed to a regular neural network for classification purposes [29, 30]. A simplified CNN structure is shown (Figure 1). Since the DMS dispersion data are 100-by-100 dimensional for our study, the input layer for our CNN structure is 100-by-100, and we choose three convolutional layers, each of which is followed by a normalization layer, a nonlinear activation function layer, and a max-pooling layer, to build the CNN structure. Each of the convolutional layers consists of 8–32 filters, and the filter size is 3-by-3. We incorporate a max-pooling layer (pool size = 2 and stride = 2) after each convolutional layer to reduce the spatial extent of the feature map and remove the redundant information. We adopt rectified linear unit (ReLU) for the convolutional and down-sampling layers as it learns faster and performs better for a deep neural network. A fully connected layer follows the convolutional and down-sampling layers. The output size of the fully connected layers equals the number of classes of our data set. The final layer of the network is the classification layer that uses probabilities returned by the preceding activation layer to assign the input the most plausible class. We randomly split the total data into three segments: training data (60 %), validation data (20 %), and testing data (20%). We select all the hyperparameters of the machine learning models through careful experimentation such that the validation accuracy is higher than 95%. The nature of DMS data makes the CNN algorithm an obvious choice to differentiate the DMS data. However, to have a quantitative comparison, we have compared the performance of the CNN model and several previously reported models in identifying chemicals in our supplementary document [31].

In addition to applying a machine learning model to classify the DMS dispersion data of different chemicals, we study the effectiveness of the magnitude-squared coherence (msc) index of dispersion data to identify chemicals and mixtures of chemicals. Specifically, we calculate the msc index of two DMS signals. The msc of two signals X and Y is a function of their power spectral densities ($P_{XX}(\omega)$ and $P_{YY}(\omega)$) and their cross power spectral density, $P_{XY}(\omega)$; and msc is expressed as

$$msc_{XY}(\omega) = \frac{|P_{XY}(\omega)|^2}{P_{XX}(\omega)P_{YY}(\omega)}, \quad (1)$$

in which ω is the frequency, and $msc_{XY}(\omega)$ represents the frequency dependent msc of the two signals: X and Y. If the msc of the new (unknown chemical composition) DMS data set and one DMS data of a known chemical (or a known chemical mixture) is higher than a threshold value in a certain frequency range, we can say that the sample corresponding to the new data set consists of the known chemical (or a chemical mixture).

3. Results and Discussions

3.1 Dispersion plots of pure chemicals:

We report the dispersion plots of reactant ion peak (RIP), Ethyl Acetate (EA), Methanol (M), 2-Propanone (2P), 2-Butanone (2B), and Ethanol (E) in Figure 2. Protonation of moisture in the carrier gas results in the formation of hydronium $[(H_2O)_nH^+]$, which appears as the RIP in the dispersion plot of carrier gas (nitrogen in our study). It is noteworthy that the DMS plots of all five chemicals have a RIP signal in the background, resulting from the same carrier gas used throughout. We observe the signs of proton-bound monomer, IH^+ (I = EA, M, 2P, 2B or E) and dimer, I_2H^+ in all five chemicals (Figure 2). The generic reactions for the monomer and dimer formation can be portrayed as below:

$$H^+(H_2O)_n + I \rightarrow IH^+(H_2O)_{n-x} + xH_2O \quad (2)$$

$$IH^+(H_2O)_m + I \rightarrow I_2H^+(H_2O)_{m-y} + yH_2O \quad (3)$$

It is evident from Figure 2 that prominent fragment ions appear next to the RIP of the dispersion plots of Ethyl Acetate, 2-Butanone, and Ethanol. The trace of fragment ions in 2-Propanone and Methanol seems obscured. Due to the complex nature of the dispersion plots, it is painstaking, if not impossible, to check visually even if two dispersion plots are alike. When it comes to comparing numerous DMS dispersion plots or identifying chemical signatures in the dispersion plots of a mixture, the complexity increases even further. Therefore, we rely on machine learning to distinguish chemicals and find chemical traces in a mixture.

3.2 Differentiation of chemical compounds with convolutional neural network model

In this section, we first study the feasibility of convolutional neural network (CNN) algorithm to distinguish the dispersion plots of different chemicals. In particular, we train a CNN model with the dispersion plot data of 2-Butanone, Ethanol, 2-Proponone, Ethyl Acetate, and Methanol samples and then predict the presence of these chemicals in some unknown (not used for training) samples containing only a single chemical. We note that our training, validation, and testing data consists of 87, 29, and 29 dispersion plots, respectively. We choose the CNN structure through a careful experimentation method (described in the methods section) such that the model's prediction accuracy for validation data set is more than 95%. We report the accuracy and loss of both training and validation data as a function of epoch number as the training progresses in Figure 3. It is evident that 25 epochs are sufficient to attain almost 100% accuracy and almost zero loss for training the CNN model. We then use the trained CNN model to identify the chemicals in testing DMS data set. We report the confusion metrics of the CNN model for the testing data in Figure 4, and the CNN model differentiates the dispersion plots of five distinct chemicals with 100% accuracy.

Since the CNN model has successfully differentiated pure chemicals, we now test the capability of the CNN model to identify the chemicals in binary and ternary mixtures. The dispersion plots of chemical mixtures can be found in our supplementary document [31]. We

adopt the same approach as that explained for pure chemicals to choose the hyperparameters and epoch numbers to train a CNN model for studying the chemical mixtures. In other words, the trained CNN models for mixture identification have almost 100% accuracy and almost zero loss. In Figure 5a, we report the performance of a CNN model to check the presence of 2-Butanone and Ethyl Acetate in a binary mixture, and it is evident that the CNN model can identify the presence of 2-Butanone and Ethyl Acetate in a binary mixture with 100% accuracy (for both validation and testing data). To recheck the generality of our findings, we report the validation and testing accuracy of the CNN model for identifying the existence of 2-Butanone and Hexanone in another binary mixture, and again we observe 100% validation and testing accuracy (Figure 5b). It should be noted that none of the previously studied algorithms [17, 18] have provided this high accuracy of prediction for identifying the chemicals in a mixture. More specifically, the previous machine learning studies [17, 18] have reported the highest accuracy below 95%, while the CNN model, reported in this work, can reach 100% accuracy while identifying chemicals in a mixture.

To increase the problem complexity, we now develop a new CNN model to identify the chemicals in a ternary mixture (2-Butanone + Ethyl Acetate + Methanol). The validation and accuracy of the CNN model for identifying the chemicals in a ternary mixture also reach 100% (Figure 6). As such, we believe that a CNN model can be a very useful tool to identify chemical (chemicals) in pure (mixture) samples. It is worth noting that the training data set in our study has been sufficient to get excellent prediction accuracy from the CNN model. However, the required training samples can vary from case to case.

While we have shown that CNN models successfully predict the chemical composition of samples, we should appreciate that collection of sufficient experimental data takes a significant amount of time and effort. It is essential to devise an alternative approach to identify chemicals in unknown samples with the least possible experimental data. In the following section, we compare the magnitude-squared coherence index of DMS plots to identify the chemicals in unknown samples.

3.4 Signal comparison to detect the chemical composition:

To predict the existence of certain chemicals in a sample, we propose and validate a simple but effective way in this section. Initially, we experimentally collected the dispersion data of several pure chemicals and several mixtures of chemicals, and later, we used those dispersion plots as standard to compare with other unknown samples. Then we calculate the magnitude-squared coherence (msc) index of the new data set and that of the previously collected standard (with known chemical composition) dispersion plot. If the msc of the new data set and one of the standard dispersion plots is higher than a threshold value in a certain frequency range, we can say that the sample corresponding to the new data set consists of chemicals pertaining to that standard dispersion plot. We study the msc for pure chemicals, binary, and ternary mixtures to validate our approach.

In Figure 7, we report the msc of two dispersion plots of 2-Butanone, two dispersion plots of Ethyl Acetate. We observe that the msc of two dispersion plots of 2-Butanones is more than 0.7 in the normalized frequency range of 0.15 to 0.45. We observe the same trend for the msc of two dispersion plots of Ethyl Acetate. Therefore, we claim that if the msc

between a chemically unknown sample and a chemically known sample is very high, i.e., above a threshold in a certain frequency range, the unknown sample will have the same chemical composition as the known sample. It is noteworthy that the threshold msc value and frequency range should be chosen using the DMS data of our interest. We emphasize that the msc of dispersion plots of 2B and EA is extremely low in the normalized frequency range of 0.15 to 0.45, and this is expected because 2B and EA are two different chemicals, and their dispersion plots are very distinct (Figure 7a). Then we compare the msc of two dispersion plots of a binary mixture of 2B and EA, and observe that the msc of 2B+EA mixture is more than 0.7 in the normalized frequency range of 0.15 to 0.45. To reinforce our findings for binary mixture, we report the msc for another binary mixture of 2-Butanone (2B) and Hexanone (H). Similar to that of the 2B+EA mixture, we observe very high msc in the normalized frequency range of 0.15 to 0.45. We thus conclude that one dispersion plot of a known binary mixture is sufficient to check if a new unknown sample also has the composition of that known binary mixture.

To demonstrate the effectiveness of msc to identify the chemical composition of a more complex mixture, we now report the msc of two dispersion plots of a ternary mixture (2B+EA+M) in Figure 8. We observe that the msc of the two dispersion plots of a ternary mixture exhibit the same trend as that observed for binary mixture. In particular, the msc of two dispersion plots of 2B+EA+M mixture has a magnitude of more than 0.7 in the normalized frequency range 0.15 to 0.45. Based on our findings of the msc trends of pure chemicals, binary mixture, and ternary mixture, we can use a few standard dispersion plots to identify the chemical compositions of a new sample. We want to emphasize that only one dispersion plot of a standard chemical (pure or mixture) is sufficient to find out if a new sample has the same standard chemical. The msc calculation approach will significantly reduce the experimental data collection effort, which is essential to adopt a machine learning approach to identify chemicals in a system.

3.5 Discussions

In the current work, we did not consider the variation of chemical concentration while identifying chemicals. However, we can normalize the DMS plots while training and testing the machine learning model or adopting the magnitude-squared coherence approach if we have to deal with samples with varying chemical concentrations. We also note that the number of samples has been sufficient to get excellent prediction accuracy in our work due to measurements of pure chemical standards. In other studies, the required number of samples can largely vary based on the type (e.g., biological and non-biological) and source of training and testing samples. As such, we must consider the variability of testing data the machine learning model has to deal with before we deploy the model to a device for further prediction. If there is a concentration difference of chemicals among the samples, we recommend normalizing the data for implementing machine learning or msc analysis for identifying chemicals. We believe the msc approach would be an excellent substitute for machine learning in several situations, e.g., differentiating the VOCs of biological samples as they are challenging to collect, and this is our future study. While our approach of using machine learning and signal processing to identify chemicals in a controlled environment

has been successful, developing a generic library will require further investigation with changing relevant factors (e.g., the performance variance of devices of the same genre).

4 Conclusions

In this work, we have shown that the convolutional neural network (CNN) model shows excellent accuracy in differentiating differential mobility spectrometry (DMS) data of pure chemicals. We also observed that the CNN model demonstrates unprecedented accuracy consistently while identifying chemicals in binary and ternary mixtures. We updated the custom AnalyzeIMS ("AIMS") software with a user-friendly implementation of the CNN model. Moreover, we have shown that the calculation of magnitude-squared coherence (msc) of the DMS data is very effective in identifying the chemical identity of a sample. Chemical library building with standard DMS plots and msc analysis can significantly reduce the burden of collecting substantial DMS data to train machine learning models for constituent chemical identification in a sample.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by: NIH NCATS 1U18TR003795-01, 4U18TR003795-02 [CED, NJK] and UL1 TR001860 [CED, NJK]; NIH award UG3-OD023365 [CED, NJK]; NIH award 1P30ES023513-01A1 [CED, NJK]; the Department of Veterans Affairs award I01 BX004965-01A1 [CED, NJK]; and the University of California Tobacco-Related Disease Research Program award T31IR1614 [CED, NJK]. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

References:

1. Krylov E, Nazarov E, and Miller R, Differential mobility spectrometer: Model of operation. International Journal of Mass Spectrometry, 2007. 266(1–3): p. 76–85.
2. Campbell JL, Le Blanc JY, and Kibbey RG, Differential mobility spectrometry: a valuable technology for analyzing challenging biological samples. Bioanalysis, 2015. 7(7): p. 853–856. [PubMed: 25932519]
3. Anttalainen O, et al. , Possible strategy to use differential mobility spectrometry in real time applications. International Journal for Ion Mobility Spectrometry, 2020. 23(1): p. 1–8.
4. Miller RA, et al. , A MEMS radio-frequency ion mobility spectrometer for chemical vapor detection. Sensors and Actuators A: Physical, 2001. 91(3): p. 301–312.
5. Schneider BB, et al. , Planar differential mobility spectrometer as a pre-filter for atmospheric pressure ionization mass spectrometry. International journal of mass spectrometry, 2010. 298(1–3): p. 45–54. [PubMed: 21278836]
6. Schneider BB, et al. , Maximizing ion transmission in differential mobility spectrometry. Journal of The American Society for Mass Spectrometry, 2017. 28(10): p. 2151–2159. [PubMed: 28664477]
7. Morgan J and Davis C. Differential mobility spectrometry applications in homeland security, clinical diagnostics and drug discovery. in ASME International Mechanical Engineering Congress and Exposition. 2006.
8. Kondratev A, et al. Clustering of Alpha Curves in Differential Mobility Spectrometry Data. in 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN). 2022. IEEE.
9. Krebs MD, et al. , Detection of biological and chemical agents using differential mobility spectrometry (DMS) technology. IEEE Sensors Journal, 2005. 5(4): p. 696–703.

10. Coy SL, et al. , Differential mobility spectrometry with nanospray ion source as a compact detector for small organics and inorganics. International Journal for Ion Mobility Spectrometry, 2013. 16(3): p. 217–227. [PubMed: 23914140]
11. Safaei Z, et al. , Differential Mobility Spectrometry of Ketones in Air at Extreme Levels of Moisture. Scientific reports, 2019. 9(1): p. 1–13. [PubMed: 30626917]
12. Arasaradnam RP, et al., A novel tool for noninvasive diagnosis and tracking of patients with inflammatory bowel disease. 2013. 19(5): p. 999–1003.
13. Martinez-Vernon AS, et al., An improved machine learning pipeline for urinary volatiles disease detection: Diagnosing diabetes. 2018. 13(9): p. e0204425.
14. Covington JA, et al., Application of a novel tool for diagnosing bile acid diarrhoea. 2013. 13(9): p. 11899–11912.
15. Covington J, et al., The application of FAIMS gas analysis in medical diagnostics. 2015. 140(20): p. 6775–6781.
16. Fabianowski W, et al. , Detection and Identification of VOCs Using Differential Ion Mobility Spectrometry (DMS). Molecules, 2021. 27(1): p. 234. [PubMed: 35011466]
17. Rajapakse MY, et al. , Automated chemical identification and library building using dispersion plots for differential mobility spectrometry. Analytical Methods, 2018. 10(35): p. 4339–4349. [PubMed: 30984293]
18. Yeap D, et al. , Machine vision methods, natural language processing, and machine learning algorithms for automated dispersion plot analysis and chemical identification from complex mixtures. Analytical chemistry, 2019. 91(16): p. 10509–10517. [PubMed: 31310101]
19. Lambertus GR, et al. , Silicon microfabricated column with microfabricated differential mobility spectrometer for GC analysis of volatile organic compounds. Analytical chemistry, 2005. 77(23): p. 7563–7571. [PubMed: 16316163]
20. Menlyadiev M, Stone J, and Eiceman G, Tandem differential mobility spectrometry with chemical modification of ions. International Journal for Ion Mobility Spectrometry, 2012. 15(3): p. 123–130.
21. Krylov EV, et al. , Selection and generation of waveforms for differential mobility spectrometry. Review of Scientific Instruments, 2010. 81(2): p. 024101. [PubMed: 20192506]
22. Walker SW, et al., Determining molecular properties with differential mobility spectrometry and machine learning. 2018. 9(1): p. 1–7.
23. Ieritano C, Campbell JL, and Hopkins WS, Predicting differential ion mobility behaviour in silico using machine learning. Analyst, 2021. 146(15): p. 4737–4743. [PubMed: 34212943]
24. Jin J, et al. , Identification of soy sauce using high-field asymmetric waveform ion mobility spectrometry combined with machine learning. Sensors and Actuators B: Chemical, 2022. 365: p. 131966.
25. Sinha R, et al., FAIMS based sensing of Burkholderia cepacia caused sour skin in onions under bulk storage condition. 2017. 11(4): p. 1578–1585.
26. Li H, et al., Identification of Specific Substances in the FAIMS Spectra of Complex Mixtures Using Deep Learning. 2021. 21(18): p. 6160.
27. Peirano DJ, Pasamontes A, and Davis CE, Supervised semi-automated data analysis software for gas chromatography / differential mobility spectrometry (GC/DMS) metabolomics applications. International Journal for Ion Mobility Spectrometry, 2016. 19(2): p. 155–166. [PubMed: 27799845]
28. Yeap D, et al. , Peak detection and random forests classification software for gas chromatography/ differential mobility spectrometry (GC/DMS) data. Chemometr Intell Lab Syst, 2020. 203.
29. Murphy KP, Machine learning: a probabilistic perspective. 2012: MIT press.
30. Alpaydin E, Introduction to machine learning. 2020: MIT press.
31. Chakraborty Pranay, M.Y.R., McCartney Mitchell M., Kenyon Nicholas J., Davis Cristina E.*, Supplementary Document: Machine Learning and Signal Processing Assisted Differential Mobility Spectrometry (DMS) Data Analysis for Chemical Identification

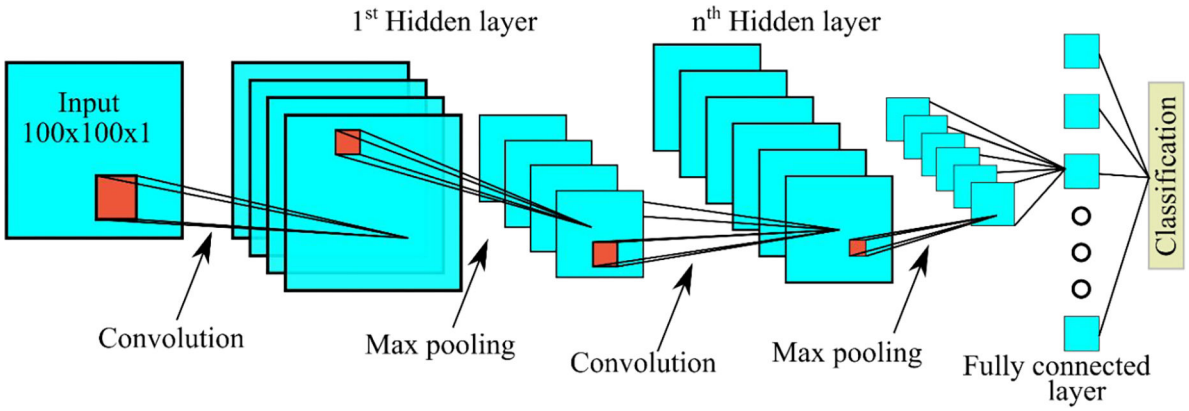

Figure 1:
Simplified schematic of a convolutional neural network (CNN) structure. The CNN structure consists of several convolution and max pooling layers. The nodes in fully connected layer equal the number of classes.

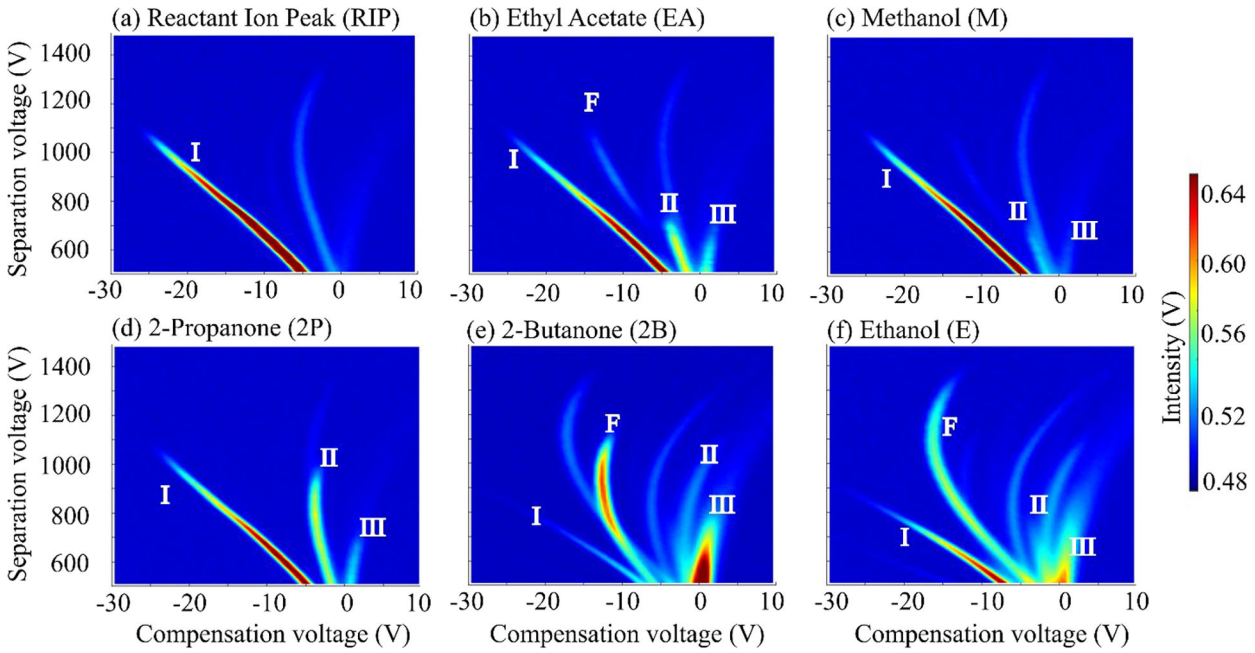

Figure 2:

Dispersion plots of (a) Reactant Ion Peak, RIP, (b) Ethyl Acetate, EA (c) Methanol, (M), (d) 2-Propanone, 2P, (e) 2-Butanone, 2B, and (f) Ethanol, E. Here, I, II, III, F represents reactant ion, proton-bound monomer, proton-bound dimer, fragment ion, respectively. The samples have a concentration of 500 ppb and the drift tube temperature is maintained at 80°C.

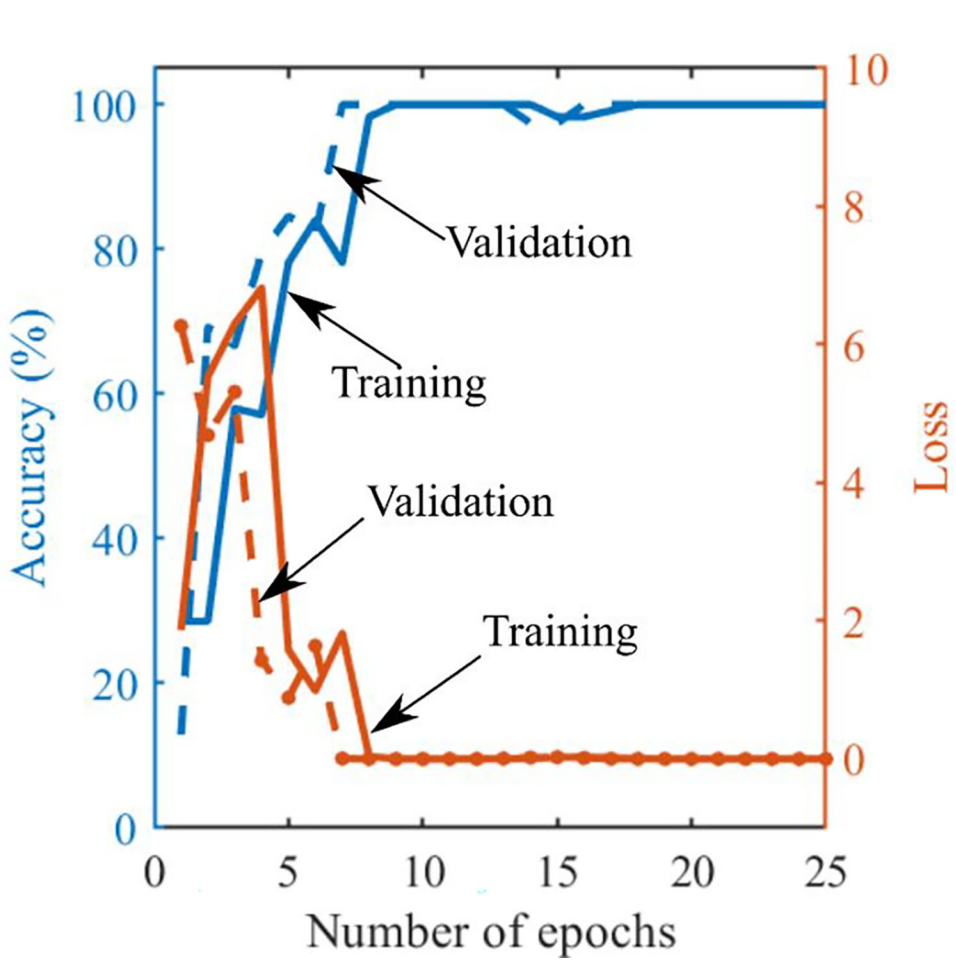

Figure 3:

Variation of accuracy and loss for the training and validation data as the training of the convolutional neural network model progresses. *Training and validation data consist of the dispersion plots of 2-Butanone, 2-Propanone, Methanol, Ethanol, and Ethyl acetate.*

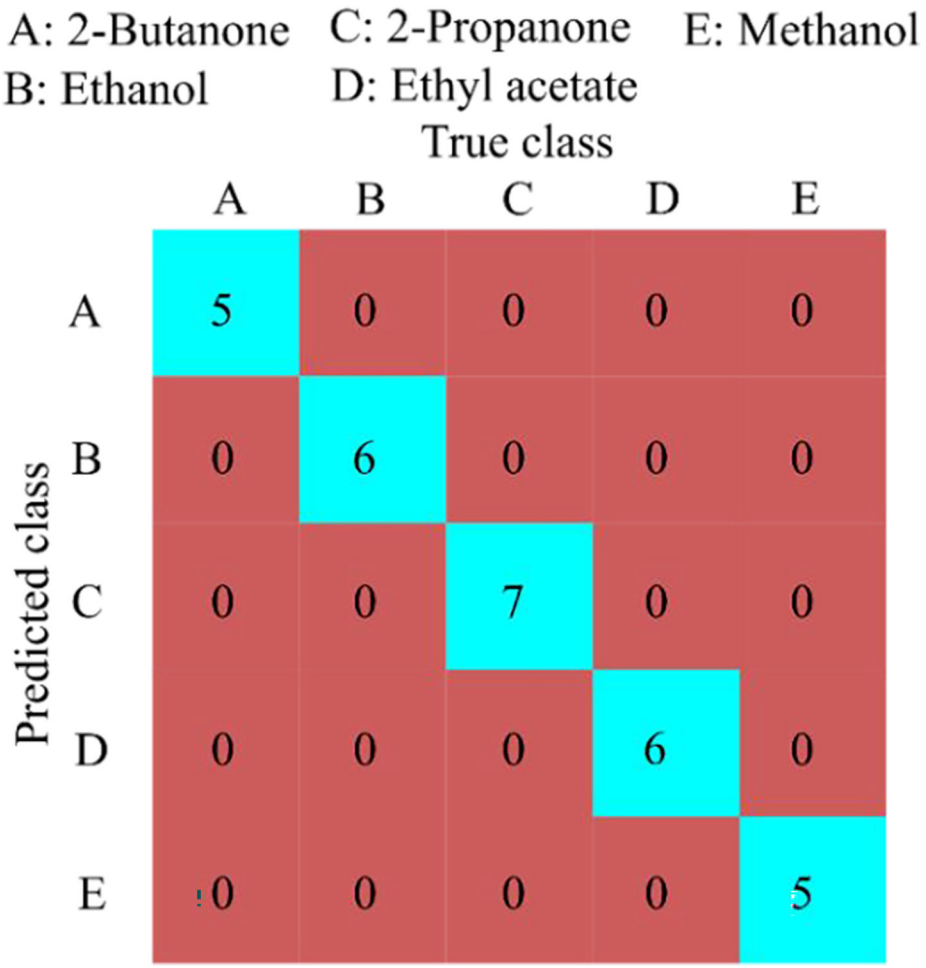

Figure 4:

Confusion matrix of the convolutional neural network (CNN) machine learning model for differentiating the pure chemicals in test data. Testing data consists of arbitrability chosen 5 number of 2-Butanone, 6 number of Ethanol, 7 number of 2-Propanone, 6 number of Ethyl acetate, and 5 number of Methanol samples. *True class and the model-predicted class match exactly for the test data set of samples containing single pure chemicals. The confusion matrix corresponds to the trained model reported in* Figure 3.

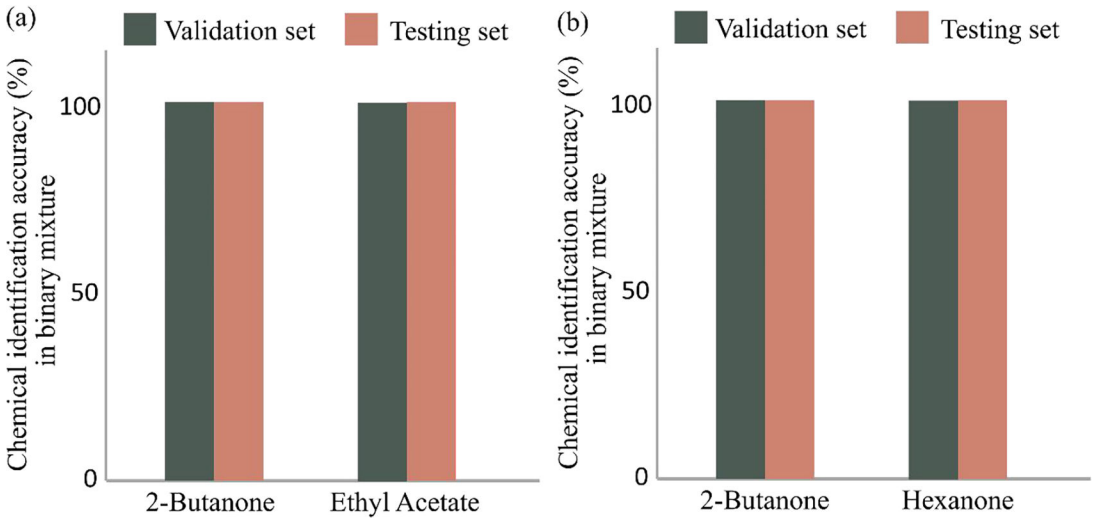

Figure 5:

a) Validation and testing accuracy of the convolutional neural network model for identifying the existence of 2-Butanone (2B), and Ethyl Acetate (EA) in a binary mixture of 2B and EA, b) Validation and testing accuracy of the convolutional neural network model for identifying the existence of 2-Butanone (2B) and Hexanone (H) in a binary mixture of 2B and H. Training, validation data for the CNN model of consists dispersion plot data of pure chemicals and their respective mixtures.

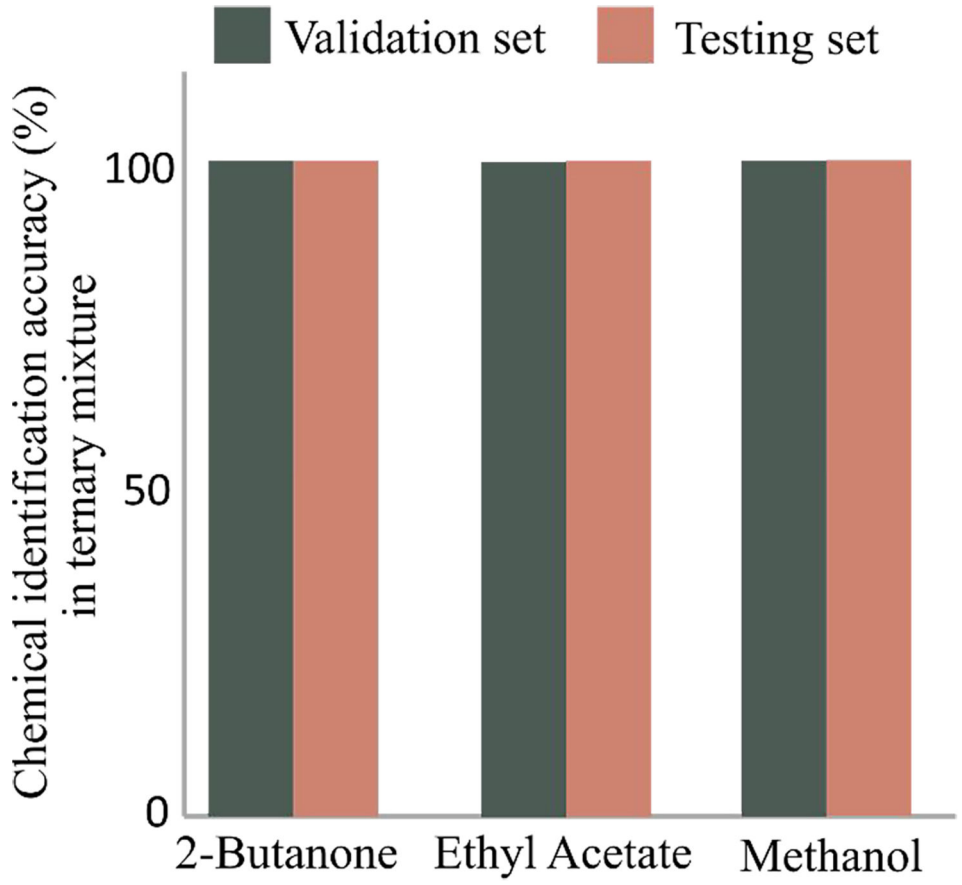

Figure 6:

Validation and testing accuracy of the convolutional neural network model for identifying 2-Butanone (2B), Ethyl Acetate (EA), and Methanol (M) in the ternary mixtures of 2B, EA, and M. Training, validation data for the CNN model of consists DMS dispersion data and their respective mixtures.

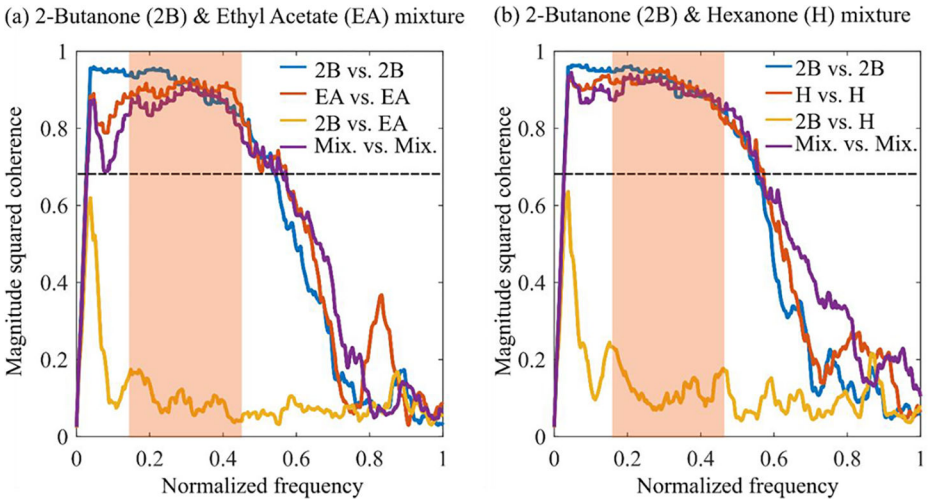

Figure 7:

Calculated magnitude-squared coherence (msc) of different dispersion plots. a) Mix. Vs. Mix. represents the msc for two dispersion plots of a binary mixture of 2-Butanone (2B) and Ethyl Acetate (EA). 2B vs. 2B, EA vs. EA, 2B vs. EA correspond to two 2B dispersions, two EA dispersions, and one 2B and one EA dispersions, respectively. b) Mix. Vs. Mix. represents the msc for two dispersion plots of a binary mixture of 2-Butanone (2B) and Hexanone (H). 2B vs. 2B, H vs. H, 2B vs. H correspond to two 2B dispersions, two H dispersions, and one 2B and one H dispersions, respectively.

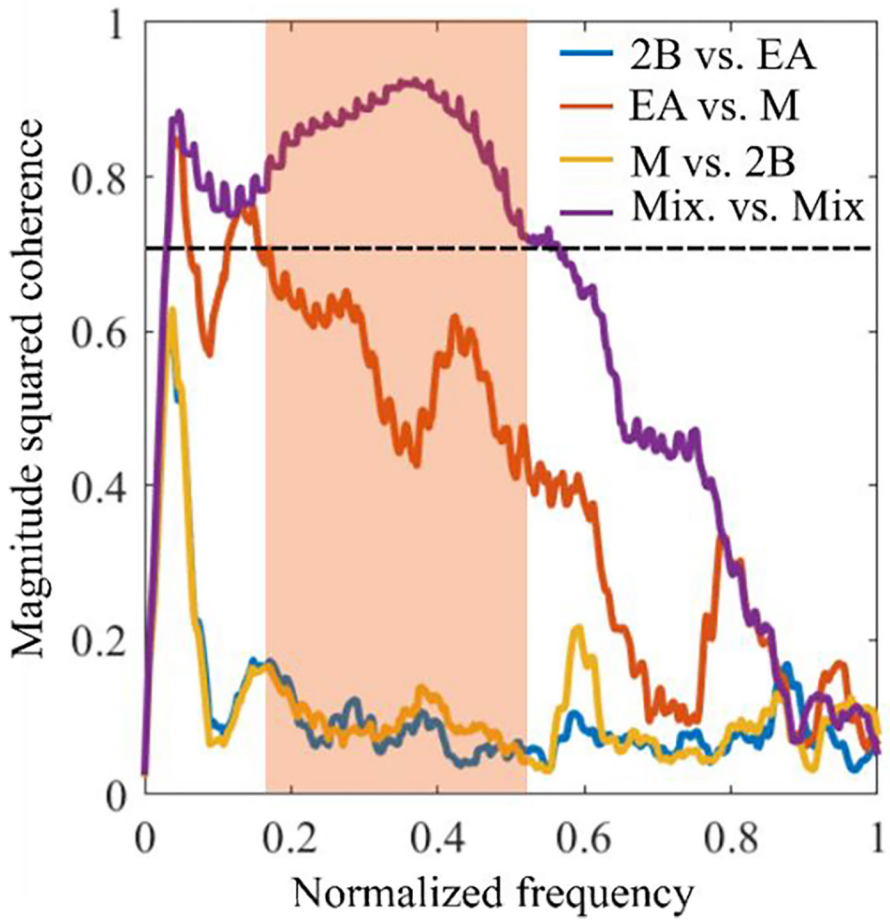

Figure 8:

Calculated magnitude-squared coherence (msc) of different dispersion plots. Mix. Vs. Mix. represents the msc for two dispersion plots of a ternary mixture of 2-Butanone (2B) and Ethyl Acetate (EA), and Methanol (M). 2B vs. EA, EA vs. M, M vs. 2B correspond to one 2B and one EA dispersions, one EA and one M, and one M and one 2B dispersions, respectively.