# ACGANs Improve Chemical Sensors for Challenging Distributions

Alexander M. Moore
*Worcester Polytechnic Institute*
Worcester MA, USA
ammoore@wpi.edu

Randy C. Paffenroth
*Worcester Polytechnic Institute*
Worcester MA, USA
rcpaffenroth@wpi.edu

Ken T. Ngo
*Chemical/Biological Innovative Material and Ensemble Development Team*
*U.S. Army DEVCOM Solder Center*
Natick MA, USA
ken.a.ngo.civ@army.mil

Joshua R. Uzarski
*Chemical/Biological Innovative Material and Ensemble Development Team*
*U.S. Army DEVCOM Solder Center*
Natick MA, USA
joshua.r.uzarski.civ@army.mil

***Abstract*—Fast and accurate chemical sensors play vital roles in medical, military, and home safety applications. Artificial olfactory systems utilize arrays of chemiresistive sensors to maximize discriminability of chemical analytes given sensor responses. At deployment time, exposures to analytes may differ from training in concentrations, analyte types, and environmental factors. Training machine learning models to remain accurate on distribution-shifted testing data requires performing many diverse, costly experiments in controlled laboratory settings to create a well-rounded training data set. In practice even expensive, large data sets may be insufficient for generalization of a trained model to a real-world testing distribution. It is possible but costly to create large, diverse datasets of single-analyte exposures, but experimentation with multi-analyte combinations quickly becomes intractable. With this in mind, we benchmark and propose machine learning and deep learning approaches to the detection of obscured chemical analytes given only single-analyte training. We significantly improve upon baseline classification of a particular chemical analyte borne in vapor mixed with an obscurant chemical agent by inducing multitask learning through adversarial data synthesizing models. Given chemical analyte exposure times between 0.75 and 5 seconds and across three unique experiment data sets comprised of different sensors and experimental controls, we find approaches utilizing Auxiliary Classifier Generative Adversarial Networks (ACGANs) outperform machine learning and comparable neural network approaches. We corroborate with multitask learning to find that across domains the utilization of generative tasks in supervised learners may increase data efficiency and model robustness to out-of-distribution samples without additional data annotation.**

***Index Terms*—Chemical sensors, deep learning, multitask learning**

## I. Introduction

Contemporary methods of machine learning for chemical sensing classify chemical analytes given a set of sensor responses for a variety of possible hardware olfactory designs [1], [2]. It is vital to the health and safety of civilians, scientists, and warfighters that hazardous chemicals are rapidly and accurately detected [1]. We optimize machine learning approaches to the classification of hazardous chemicals of interest ("Analyte A") in the presence of additional analytes ("obscurant" analytes B, C, and D) for a variety of sensor designs. We find that ACGANs (Section III) outperform machine learning and comparable neural network approaches for the classification of Analyte A in the presence of obscurant analytes without access to multi-analyte training data (Section V).

Many sensors may fail to respond to hazardous chemicals in the presence of obscurants [3]. In addition, gathering data for a chemical sensing data set in order to train olfactory-based classifiers requires expert end-users, specialized hardware design and maintenance, and time-consuming experimental trials [1], [4], [5].

Limitations in experimental data volume lead to investigations in improving model performance on a challenging chemical sensor problem - particularly when the space of possible analyte exposures is highly complex and expensive to sample from. Generalizing from laboratory to real-world data incurs additional complexity due to distributional differences which may lead to detection failures in machine learning tools [6]. In classifying the presence of Analyte A borne in vapor by an olfactory system, it is vital to measure the performance of the classifier for unseen obscurant analytes. Since it is prohibitive to perform exhaustive experimentation of double-analytes in a laboratory setting, we turn to the optimization of classifiers in order to handle unseen challenging distributions.

Figure 1 demonstrates an observation of Analyte A unobscured by additional analytes with a characteristic sensor response. The introduction of three other obscurant analytes each modifies the resistance curves in different ways. Though Analyte A is still present, the relationships of sensor responses from the unobscured curve have been altered. A supervised model must account for this change in distribution without access to these double-analyte responses during training.

In order to measure model response to a likely change in real-world deployment distribution caused by obscurant analytes, we utilize single-analyte exposures of four chemical analytes of interest (Analyte A, B, C, or D). Models trained with these single-analyte exposures are tested on unseen double-analyte exposures (combinations A-B, A-C, A-D, B-

978-1-6654-6283-9/22/$31.00 ©2022 IEEE
DOI 10.1109/ICMLA55696.2022.00047

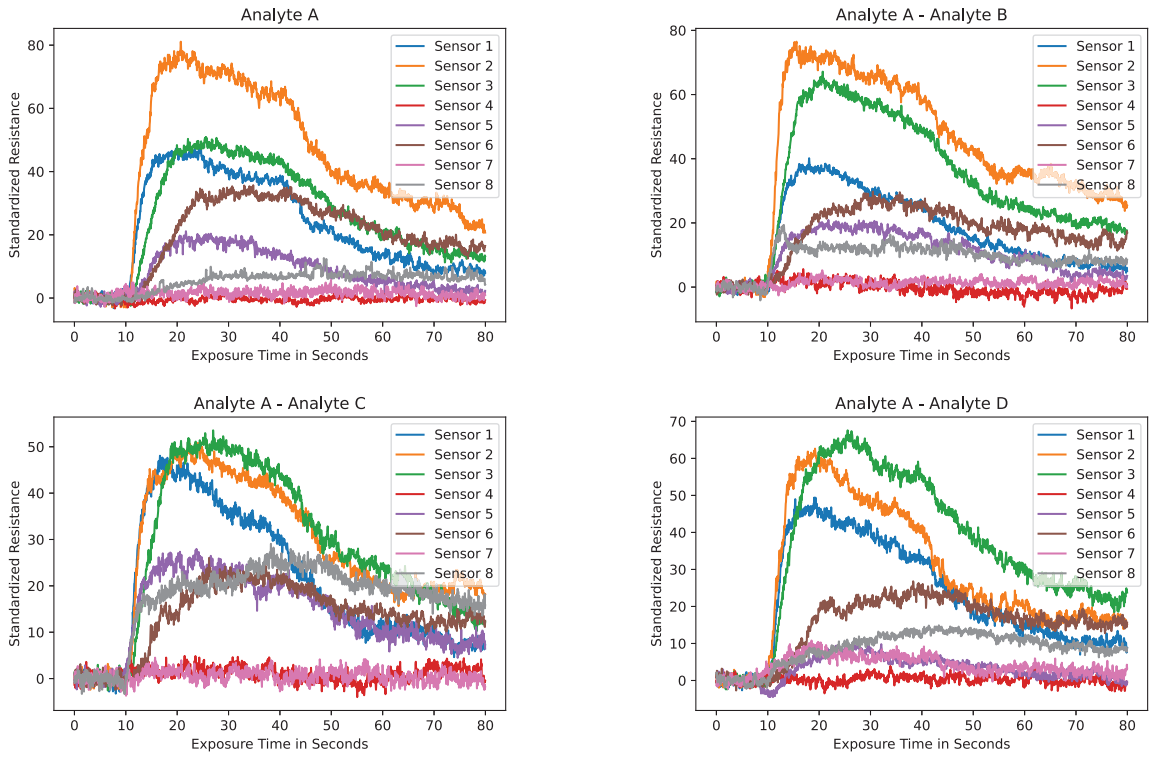


Fig. 1: Sensor response curves to Analyte A are shown unobscured (top left) and obscured in three different ways. Obscurants change the characteristic sensor resistance response of Analyte A alone.

C, C-D) in which Analyte A may be present but masked by some unknown obscurant analyte. Section II discusses the complications of gathering exhaustive experimentation beyond single-analyte exposures. The classification of Analyte A in the presence of obscurant analytes emulates the vital task of hazardous analyte detection for real-world applications where environmental factors are unpredictable for experimentation.

We propose the utilization of Auxiliary Classifier Generative Adversarial Networks (ACGANs) as supervised learners which make use of adversarial and multitask training to improve out-of-distribution performance [7]. Though generative deep learning has been applied to molecular synthesis and drug discovery [8], [9], the inclusion of adversarial and generative training to chemical analyte discrimination may be beneficial to chemical discrimination with sensor devices and to our knowledge are yet unexplored [10]–[12]. We hypothesize that simultaneous adversarial training of a data-synthesizing model improves generalizability of the discriminator-classifier model through the parameterization bias imparted by multitask learning [13]–[15]. Results in Section V demonstrate inducing additional data-synthesizing tasks (Section III) without additionally annotated data significantly improves testing outcomes for distribution-shifted data (Section II).

### A. Summary of Contributions

- We propose the utilization of ACGANs as generalizable multi-task learners which improve testing classification outcomes on challenging distribution shifts. We leverage multi-task learning without requiring additional experimental labels or annotations (Section III).
- We record optimized benchmarks for four baseline machine learning models as well as a comparison between feedforward neural network classifiers and ACGANs given optimized hyperparameters for discrimination of Analyte A as well as detection in the presence of obscurant analytes (Section IV, V).
- We perform model comparisons between comparable ACGAN and feedforward neural networks to diagnose how inducing additional tasks may benefit generalizability of deep learners for chemical sensing under distribution shifts (Section V).
- We study the scaling behavior of classifier models with respect to the combination of single- and double-analyte training samples to improve future chemical sensor data set development (Section V-A).

## II. Data

Three experimental data sets are gathered using three arrays of eight chemical sensors which react differently to chemical analyte exposures using a process described in detail in [1], [16], [17] and [5]. The unique coatings on the chemiresistive sensor cause change in resistances to facilitate discrimination of the gas exposures [4], [5], [17]. Figure 2 shows an example of the contrast in sensor responses to 17.5% Analyte A

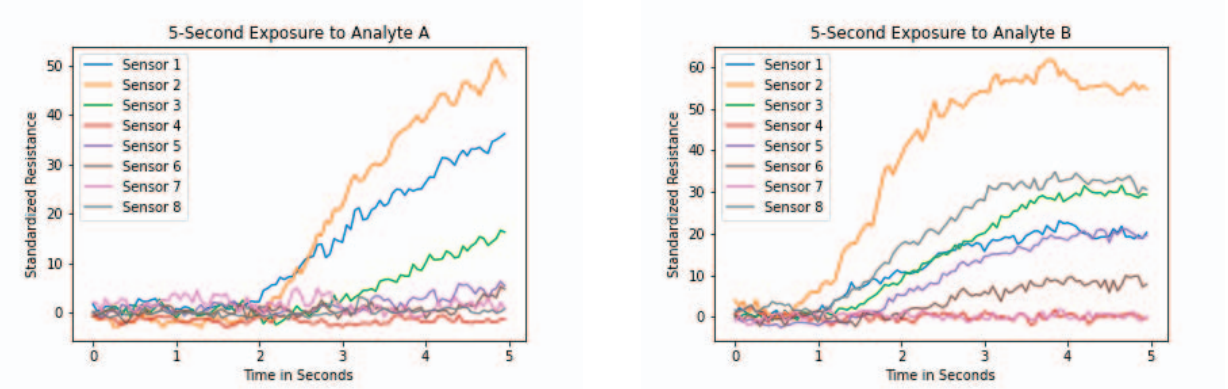


Fig. 2: Two five second single-analyte exposures to different analytes at the same concentration. Discrepancies in sensor resistance are explained by adsorption interactions between analytes and sensor coatings.

and $17.5\%$ Analyte B vapors given the same set of eight sensors. From the 8-channel signal the supervised learners must classify the presence of Analyte A for unseen testing samples.

In Figure 3, we see separation between Analyte A-containing and non-Analyte-A containing signals in the space dependent on the coatings of the sensor set given in the experiment. Experiments One and Two may have sensor coatings more conducive to the discrimination of Analyte A from other analytes as judged by the distance and linear separability and distance of Analyte A samples under one ISOMAP transformation.

To collect one experimental trial (Figure 4), the sensors first rest for ten seconds. This period is used to calculate the mean shift in sensor resistance to standardize the data during preprocessing. A controlled flow of analyte vapor is released into the sensor chamber for a thirty second response time before the exposure valve is shut. Finally, a desorption time begins and lasts forty seconds in which the sensor resistances recover and no further analyte vapor is released, indicated in Figure 4 by the lengthy period of gradual decline in resistance. Between trials, an eight minute recovery period allows the sensors to fully recover and any lingering analyte vapor to dissipate. Triplicate trials for each analyte concentration are performed for data integrity.

The data are split into two categories for the experiments in Section V. Single-analyte exposures compose the training set which is used to train machine and deep learning classifiers to predict the presence of Analyte A. The single-analyte exposure data are also used for validation in which unseen holdout samples are used to optimize over hyperparameter grid searches (Algorithm 1) in Section IV. Triplicate experiments are performed for each of four analytes at each of four concentrations$\{6.25\%, 12.5\%, 17.5\%, 25\%\}$. We retain an unseen holdout set of double-analyte exposures. This small data set ($n \approx 30$ for each of three experimental data sets) of two simultaneous analyte exposures are used as the holdout set to measure how well methods are able to generalize to unseen distributions of obscured analytes.

Exhaustive experimentation of double-analyte exposures for use as training data would be intractable given even a limited number of analyte concentrations. Current experiments produce samples using a small subset of the many analytes and analyte mixtures of interest. Given a number of chemical analytes of interest $x_1, ..., x_n$ at various concentrations $c_1, ..., c_k$, the number of experiments necessary to perform exhaustive experimentation of all mixtures at all concentrations is given by

$$\sum_{p=1}^{n} \binom{n}{p} k^p \tag{1}$$

where $p$ represents the number of analytes in the mixture. Therefore four analytes at four concentration levels would make for 16 single-analyte experiments and 96 double-analyte experiments. If six analytes were under consideration, however, these numbers would rise to 36 single-analyte and 540 double-analyte experiments. Accordingly, a complete set of training data could only be produced given a limited number of analytes and exposure magnitudes. By utilizing ACGANs as multitask adversarial learners [7] with single-analyte exposures, supervised learning outcomes are improved and constraints to experimental hardware, budget, and time limit may be mitigated.

## III. Generative Multitask Models

Data generation is the process of sampling from an approximation of a data distribution. ACGANs are one approach to induce a multitask training paradigm with no additional labels or exterior tasks. Rather than including additional tasks with extra labels on training samples, the model utilizes additional *induced* tasks via the adversarial data synthesis process which optimize the model with respect to the same training data and labels but with additional loss terms [13]. We hypothesize the representation bias phenomena [19] from the additional tasks improves model parameterizations. Deeper representations may improve supervised learning generalizability over comparable feedforward classifiers, particularly for low-data experiments with a shifted testing distribution.

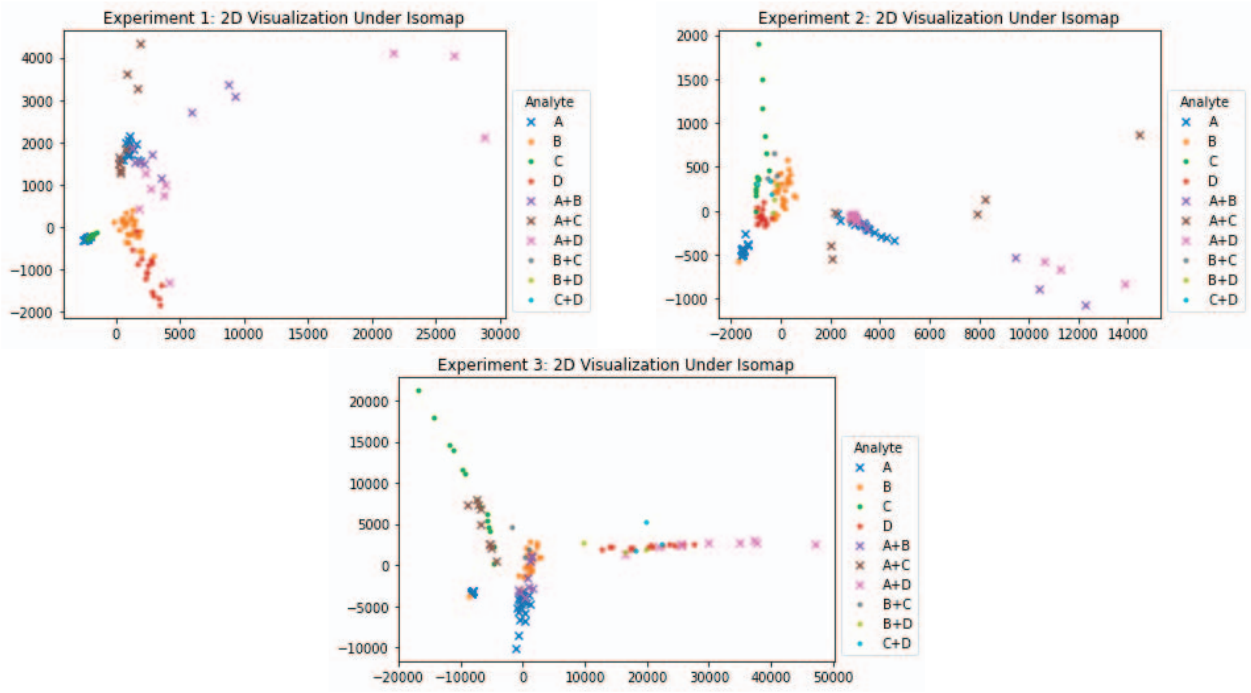


Fig. 3: Exploratory data analysis demonstrating ISOMAP dimensionality reduction [18] from the space of sensor response curves to two dimensions. In this visualization, an ISOMAP embedding trained on single-analyte exposures then encodes both the single and double-analyte sets. Samples containing the analyte of interest Analyte A are marked with crosses to differentiate the positive classification label.

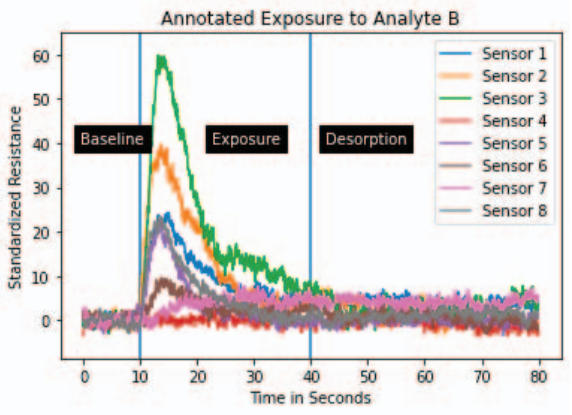


Fig. 4: One experimental sample including baseline, absorption, and desorption for an exposure of Analyte B. The sensor resistances rapidly increase during the analyte exposure period before returning to baseline levels during desorption. Classifiers use discrepancies in sensor resistances responses to discriminate between analytes.

For Generative Adversarial Networks (GANs, [20]), a generative function learns parameters mapping from a low-dimensional standard normal prior to an estimate of the training distribution in the data space based on a game-theoretic min-max game. Two deep learning agents, the generator $G$ and discriminator $D$, alternate weight updates to maximize the adversary agent error.

The min-max optimization uses expectation over samples from the training set $x \sim p_{data}$ of the error of the discriminator given by $\log D(x)$. The discriminator error on real samples is balanced against that of error on generated samples given by the expectation over samples $z \sim p_z(z)$ decoded by the generator $G(z)$, where $z$ is the multivariate standard normal prior. The agents update in turn, where the generator tries to minimize the objective against a maximizing discriminator solved by

$$L_{GAN} = \min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log 1 - D(G(z))]. \quad (2)$$

The Auxiliary Classifier GAN (ACGAN, interested readers may refer to [7] on adversarial supervised training) modifies the discriminator of the GAN to predict the corresponding class of input training and synthesized samples. ACGANs account for conditional probabilities approximating the joint data distribution $p(x, c)$ by conditioning the generation and discrimination of samples on the analyte label $c$. In training, the generator $G$ uses a condition $c$, here representing a chemical exposure label, to conditionally generate from a random latent sample $z$ converging to an approximation of the conditional training distribution. The ACGAN alters GAN training by incorporating an auxiliary classifier head on the

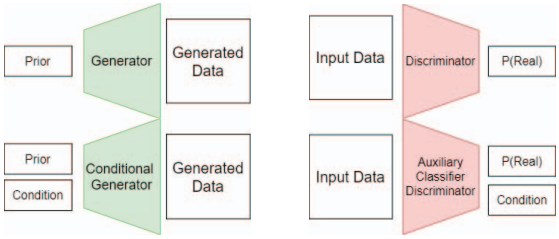


Fig. 5: The ACGAN modifies the GAN architecture to include an auxiliary classifier head on the discriminator output layer. A condition classification task extends a data synthesizing model to a conditional data synthesizing model. We utilize the classifier output which shares parameter optimization with the discriminator as a regularized supervised learner.

output of the discriminator, with a conditional component given by:

$$L_C = \mathbb{E}_{x\sim p_{data}}[\log P(C=c|x)] + \mathbb{E}_{z\sim p_z(z)}[\log P(C=c|G(z))]. \quad (3)$$

Using the conditional loss, the Discriminator is trained to maximize $L_C + L_{GAN}$ (2) while the Generator is trained to maximize $L_C - L_{GAN}$ [7]. Odena (2017) finds the inclusion of the conditional adversarial training quantifiably stabilizes training and improves discriminability of generated samples alongside advantages in representation bias through multitask learning by balancing adversarial and supervised losses [13]. Utilizing the conditional discrimination of the ACGAN to classify samples from a shifted testing distribution with results from Section V demonstrates that adversarial training improves upon feedforward networks for three distribution-shifted chemical sensing data sets.

## IV. Methods

Results in Section V use three separate real chemical sensing data sets with different sensor materials, analyte concentrations, and analyte exposures. Each experiment is separated into a single-analyte data set and a double-analyte data set. Single-analyte data sets are utilized as training and validation data for hyperparameter optimization. Double-analyte exposure experiments are utilized as testing data.

For each set of experiments, we train the following supervised learners which represent a diverse set of baseline classifiers. Each model is optimized over a broad swathe of hyperparameters selected by cross-validation given the experiment, data set, and exposure time period $t \in \{1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, 5\}$ seconds. The hyperparameter grid searches utilized for each model are as follows:

1) Auxiliary Classifier GANs [7]
   - Model width: 8, 16, 32, 64, 128
   - Learning rate: 0.001, 0.002, 0.01, 0.05
   - Training epochs: 4, 8, 16, 32, 64, 128
   - Batch size: 4, 8, 16
   - Latent Dimension: 32
2) Feedforward Neural Networks [21]
   - Model width: 8, 16, 32, 64, 128
   - Learning rate: 0.001, 0.002, 0.01, 0.05
   - Training epochs: 4, 8, 16, 32, 64, 128
   - Batch size: 4, 8, 16
3) Baseline models (Implemented via [22]):
   a) Decision Tree Classifier [23]
      - Max depth: 1-8
      - Minimum samples per split: 2-8
      - Max features considered: 1, 2, 4, 8, $\sqrt{vars}$
   b) Random Forest Classifier [24]
      - Number of estimators: 2, 5, 25, 100, 200, 400, 800
      - Splitting criterion: gini, entropy
      - Max depth: 2, 4, 8, 16, 32
   c) K-Neighbors Classifier [25]
      - Number of neighbors: 1-20
      - Weights: Uniform weighting, distance weighting
      - Distance power: 1, 2
   d) Logistic Regression [26]
      - Parameter Lasso penalty: 0.0001, 0.001, 0.01, 0.05, 0.1, 0.15, 0.25, 0.5, 0.75, 1.0

For each of the six models the following optimization method is used. For each experiment and for each time $t \in \{1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, 5\}$ seconds, the model is trained on half of the training data using a stratified sampling by analyte concentration. The other half is the validation set used to select the optimal hyperparameter configurations given 5-fold cross validation [27]. The optimal model from the exhaustive hyperparameter search is taken to be the highest-performing model based on validation set F1 [22] score for that time period. The F1 score is given by the harmonic mean of the precision and recall of the model classifications. The validation process is repeated 5 times to find optimal hyperparameters, and the corresponding testing F1 score is recorded. The mean and standard error of testing performance is provided for each data set, experiment, and model in Section V.

## V. Results

We report the improvements to rapid and accurate classification of an analyte of interest in the presence of potential obscurant analytes. Throughout, the F1 score is used as the metric to quantify success.

Results from Table I indicate a lack of certainty in the superior model for the classification of Analyte A without the presence of obscurants in the testing set. Multiple models perform well on a single-analyte training set when optimized on single-analyte training data. ACGANs perform best on Experiment Two single-analyte classification data, and are second best for Experiments One and Three.

Results from Table II indicate that the usage of the ACGAN discriminator as a supervised learner dramatically improves testing outcomes across data sets compared to comparable neural networks and baseline machine learning models. Changing the testing distribution significantly reduces F1 metric outcomes compared to results in Table I, but the ACGAN loses less ground when compared to peer models which demonstrate a failure to generalize to the testing set.

TABLE I: Mean ± standard error testing F1 score of classification of Analyte A given single-analyte training and testing sets. Models are cross-validated for hyperparameter selection on a validation set before being tested on a holdout set of single-analytes. For each experimental data set, the best overall model is highlighted in dark green and the second best overall model is highlighted in light green.

| Model | Experiment One | Experiment Two | Experiment Three |
|---|---|---|---|
| ACGAN | $0.7540 \pm 0.0121$ | $\mathbf{0.7968} \pm 0.0114$ | $0.8684 \pm 0.0130$ |
| Neural Network | $\mathbf{0.8504} \pm 0.0113$ | $0.6980 \pm 0.0287$ | $0.6186 \pm 0.0140$ |
| Decision Tree Classifier | $0.2679 \pm 0.0536$ | $0.4075 \pm 0.0815$ | $0.7529 \pm 0.1506$ |
| K-Neighbors Classifier | $0.4254 \pm 0.0851$ | $0.5271 \pm 0.1054$ | $0.8448 \pm 0.1690$ |
| Logistic Regression | $0.3661 \pm 0.0732$ | $0.4610 \pm 0.0922$ | $\mathbf{0.8843} \pm 0.1769$ |
| Random Forest Classifier | $0.2969 \pm 0.0594$ | $0.4875 \pm 0.0975$ | $0.7949 \pm 0.1590$ |

TABLE II: Mean ± standard error testing F1 score of classification of Analyte A in the presence of an additional analyte given training on single-analyte exposures. Models are cross-validated for hyperparameter selection on a validation set of single-analytes before being tested on an unseen set of double-analyte exposures to measure the generalizability of classifiers to new distributions. For each experimental data set, the best overall model is highlighted in dark green and the second best overall model is highlighted in light green.

| Model | Experiment One | Experiment Two | Experiment Three |
|---|---|---|---|
| ACGAN | $\mathbf{0.4436} \pm 0.0493$ | $\mathbf{0.5525} \pm 0.0614$ | $\mathbf{0.4655} \pm 0.0517$ |
| Neural Network | $0.1804 \pm 0.0497$ | $0.1804 \pm 0.0497$ | $0.0691 \pm 0.0402$ |
| Decision Tree Classifier | $0.2169 \pm 0.0434$ | $0.2098 \pm 0.0420$ | $0.1119 \pm 0.0224$ |
| K-Neighbors Classifier | $0.2755 \pm 0.0551$ | $0.2287 \pm 0.0446$ | $0.0906 \pm 0.0181$ |
| Logistic Regression | $0.2339 \pm 0.0468$ | $0.2231 \pm 0.0446$ | $0.1170 \pm 0.0234$ |
| Random Forest Classifier | $0.1474 \pm 0.0295$ | $0.1797 \pm 0.0359$ | $0.0111 \pm 0.0022$ |

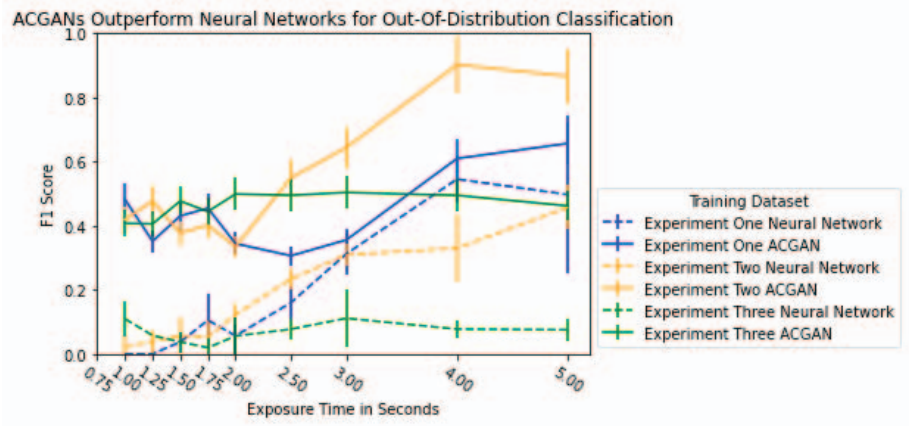


Fig. 6: Across data sets and exposure times, ACGANs trained with single-analyte exposures outperform comparable feedforward neural networks when tested on double-analyte exposures. These performances account for a cross-validation optimized model over a large gridsearch of parameters for each possible exposure window and experimental data set.

Figure 6 visualizes the difference between comparable feedforward neural network and ACGAN models for a change in testing distribution for all three data sets. Though optimized on comparable hyperparameter sets, the addition of the adversarial data synthesis task for ACGANs has substantially improved the performance on the more challenging double-analyte distribution for all time windows. Though F1 improvement trends of neural networks and ACGANs are comparable when increasing exposure time, ACGANs perform substantially better on generalization to an out-of-distribution testing set of obscured double-analytes across all exposure windows with the exception of a near-time for three- and four-second exposures in Experiment One.

### A. Scaling Behavior

Figure 7 explores the possibility of including double-analyte training data to improve learning outcomes for double-analyte classification. As the number of double-analyte experiments included in training increases, we find decreased variability in F1 outcomes with respect to variable numbers of single-analyte training data. In addition, including 16 double-analyte samples significantly improves learning outcomes for all numbers of

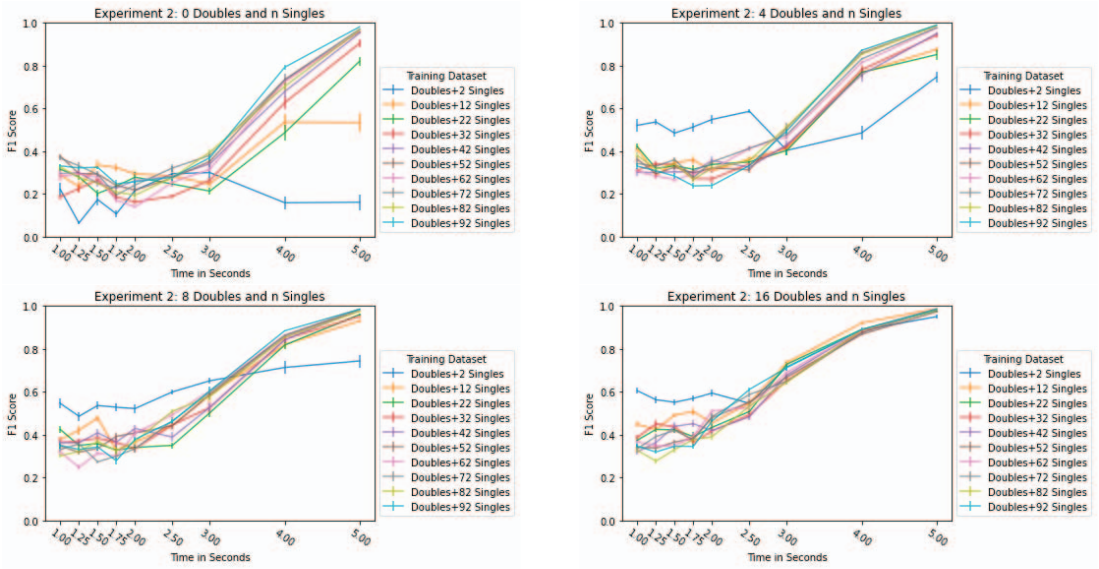


Fig. 7: For Experiment Two, we demonstrate F1 score outcomes for the classification of double-analyte samples for variable time windows. For each combination of a random sample of $k$ double-analyte experiments and $n$ single-analyte experiments, the resulting F1 outcome on holdout double-analytes is reported for each exposure window time. Moving from top left to bottom right, the classifier is trained with incrementally more double-analyte exposures.

**Algorithm 1** A hyperparameter gridsearch algorithm for one model, experiment, and time period.

**Require:** Supervised learning model $m$. Exposure time $t$. A number $k$ of folds for cross-validation. A set of hyperparameters to consider $\{h_0, ..., h_p\}$. A number of trials to reperform the experiment for standard error. Training single-analyte dataset ($n \approx 100$) from one of three experiments. Testing double-analyte dataset ($n \approx 30$).

Declare a split $K$ of the training data into $k$ disjoint partitions using label-stratified sampling.

**for** each trial **do**

  **for** $h \in H$ **do**

    **for** each train-validation split $i \in K$ **do**

      Remove $i^{th}$ partition of training data as validation.

      Train a classifier on the training partition using the hyperparameter configuration.

      Record the model's F1 score on the validation set.

    **end for**

  **end for**

  Select the model hyperparameter configuration with the lowest mean validation score.

  Retrain this model on the entire training set.

  Record this model's testing F1 performance.

**end for**

Summarize mean and standard error of testing F1 performances across trials given optimal models.

additional single-analytes. There exists an inflection point between 0 and 4 double-analyte samples which causes the addition of further single-analyte samples to detract from training, particularly in small time window detection problems. The variance in classification outcomes is highly dependent on time and number of single-analytes for zero double-analytes (top-left), and consistently low with 16 double-analyte training samples, regardless of the number of singles-analyte samples.

The addition of double-analyte training samples improves classification outcomes for these chemical sensing data sets. In addition, we have demonstrated little improvement beyond 32 single-analyte training samples on the learning outcomes for double-analyte classification. These two results together indicate that given the experimental design used in this paper with four analytes. further experiments may be required to sample from the space of double-analyte exposures. Though this effect is consistent in three chemical sensing experiments using four analytes of interest, it may not hold for higher numbers of analytes as the scaling behavior of experimentation with high numbers of analytes. We may expect similar scaling behavior in the number of double-analyte experiments used as training data for differing analytes, analyte combinations, or sensor designs. However, exhaustively sampling from double-analytes rises quadratically with the number of analytes of interest, whereas single-analyte sampling used in this research remains linear.

## VI. Discussion

Though GANs are known to suffer collapse in low-data training, we demonstrate that the ACGAN may be trained such that its supervised generalizability is significantly improved beyond that of comparable neural networks or baseline models.

This finding may be corroborated by ongoing research into the representations learned by multitask synthesizing-predicting models and investigations into the apparent data efficiency and robustness to distribution shifts of these adversarial-trained models. It is possible that deeper, more generalizable representations are learned as a function of additional optimization terms induced from data synthesis.

Data synthesizing models represent a highly flexible function from a latent prior to a complex manifold approximating the training distribution. For this reason further analysis of the role of generators as adaptive augmentations for the training of classifying discriminators is necessary. Auxiliary classifier discrimination evaluation of synthetic samples may serve as a form of training data augmentation factoring into the increased generalizability seen in Section V. Conditional generators could play a similar role to augmentations in the training of auxiliary discriminators, and further research is needed into the relationship between augmenting data sets with synthetic samples versus perturbed ones particularly in the adversarial learning paradigm at various training data volumes and complexity.

For sensor tasks where the testing distribution is known to depart from training, adversarial-trained classifier models may be appropriate for tasks in which gathering exhaustive data on the new distribution may be intractable. The representation bias from multitask models may benefit supervised learners for low-data paradigms, even in the absence of additionally-annotated data.

## VII. Acknowledgements

This manuscript has been authored with funding provided by the Defense Threat Reduction Agency (DTRA). The publisher acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes.

Approved for public release. Distribution unlimited.

## References

[1] M. S. Wiederoder, E. C. Nallon, M. Weiss, S. K. McGraw, V. P. Schnee, C. J. Bright, M. P. Polcha, R. Paffenroth, and J. R. Uzarski, "Graphene nanoplatelet-polymer chemiresistive sensor arrays for the detection and discrimination of chemical warfare agent simulants," *ACS sensors*, vol. 2, no. 11, pp. 1669–1678, 2017.

[2] W. Zhang, L. Wang, J. Chen, W. Xiao, and X. Bi, "A novel gas recognition and concentration detection algorithm for artificial olfaction," *IEEE Transactions on Instrumentation and Measurement*, 2021.

[3] Y. Seto, M. Kanamori-Kataoka, K. Tsuge, I. Ohsawa, K. Matsushita, H. Sekiguchi, T. Itoi, K. Iura, Y. Sano, and S. Yamashiro, "Sensing technology for chemical-warfare agents and its evaluation using authentic agents," *Sensors and Actuators B: Chemical*, vol. 108, no. 1, pp. 193–197, 2005, proceedings of the Tenth International Meeting on Chemical Sensors. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400505000377

[4] M. Weiss, M. S. Wiederoder, R. C. Paffenroth, E. C. Nallon, C. J. Bright, V. P. Schnee, S. McGraw, M. Polcha, and J. R. Uzarski, "Applications of the kalman filter to chemical sensors for downstream machine learning," *IEEE Sensors Journal*, vol. 18, no. 13, pp. 5455–5463, 2018.

[5] M. S. Wiederoder, M. Weiss, B. Yoon, R. C. Paffenroth, S. K. McGraw, and J. R. Uzarski, "Impact of graphene nanoplatelet concentration and film thickness on vapor detection for polymer based chemiresistive sensors," *Current Applied Physics*, vol. 19, no. 9, pp. 978–983, 2019.

[6] K. Bousmalis and S. Levine, "Closing the simulation-to-reality gap for deep robotic learning," Oct 2017. [Online]. Available: https://ai.googleblog.com/2017/10/closing-simulation-to-reality-gap-for.html

[7] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," 2017.

[8] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, "Deep learning for molecular generation and optimization - a review of the state of the art," *CoRR*, vol. abs/1903.04388, 2019. [Online]. Available: http://arxiv.org/abs/1903.04388

[9] N. W. A. Gebauer, M. Gastegger, S. S. P. Hessmann, K.-R. Müller, and K. T. Schütt, "Inverse design of 3d molecular structures with conditional generative neural networks," *Nature Communications*, vol. 13, no. 1, feb 2022. [Online]. Available: https://doi.org/10.10382Fs41467-022-28526-y

[10] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017.

[11] A. C. Mater and M. L. Coote, "Deep learning in chemistry," *Journal of Chemical Information and Modeling*, vol. 59, no. 6, pp. 2545–2559, 2019. [Online]. Available: https://doi.org/10.1021/acs.jcim.9b00266

[12] J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," *InfoMat*, vol. 1, no. 3, pp. 338–358, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/inf2.12028

[13] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[14] Y. Shinohara, "Adversarial multi-task learning of deep neural networks for robust speech recognition," in *INTERSPEECH*, 2016.

[15] R. Child, "Very deep vaes generalize autoregressive models and can outperform them on images," 2021.

[16] J. Park, W. A. Groves, and E. T. Zellers, "Vapor recognition with small arrays of polymer-coated microsensors. a comprehensive analysis," *Analytical chemistry*, vol. 71, no. 17, pp. 3877–3886, 1999.

[17] E. C. Nallon, V. P. Schnee, C. J. Bright, M. P. Polcha, and Q. Li, "Discrimination enhancement with transient feature analysis of a graphene chemical sensor," *Analytical chemistry*, vol. 88, no. 2, pp. 1401–1406, 2016.

[18] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.290.5500.2319

[19] J. Baxter, "A model of inductive bias learning," *Journal of artificial intelligence research*, vol. 12, pp. 149–198, 2000.

[20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.

[24] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[25] B. W. Silverman and M. C. Jones, "E. fix and jl hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951)," *International Statistical Review/Revue Internationale de Statistique*, pp. 233–238, 1989.

[26] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 20, no. 2, pp. 215–232, 1958.

[27] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1974.tb00994.x