Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Modeling of Physics Based UVActivated Acetone Sensor with Machine Learning Analysis

Shefali Mamataj

smamataj@gmail.com

National Institute of Technology Durgapur / MCET

Bhaskar Roy

Asansol Engineering College

Deepshikha Datta

University of Engineering and Management

Sunipa Roy

Krishnagar Women's College

Bimal Das

National Institute of Technology Durgapur

Research Article

Keywords: Acetone sensor, Virtual calibration, gallium oxide, machine learning, Schottky diode, ultraviolet activation

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11533961/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Acetone in exhaled breath is a recognised marker for diabetes. However, most gallium oxide sensors need to operate at high temperatures to detect acetone. In this study, we developed a simplified computational model of a proposed UVactivated Ni/$(Al_{0.17}Ga_{0.83})_2O_3$/$Ga_2O_3$ acetone sensor. The model describes how electrical current flows through the sensor, how UV light lowers the energy barrier, how acetone molecules attach to the sensor surface, how temperature affects the surface properties, and how the sensor response changes over time. We generated 2000 simulated observations to study whether acetone concentration could be predicted from the sensor's response. Acetone concentrations ranging from 10 to 500 ppm were selected independently. Seven machine-learning methods were tested. The models were trained and tuned using nested five-fold cross validation on 1,600 observations. The remaining 400 observations were kept separate and used only for final testing. The multilayer perceptron (MLP) produced the lowest cross validation RMSE of 50.92 ± 3.94 ppm. On the final test set, it achieved an RMSE of 55.63 ppm, an MAE of 40.87 ppm, and an R^2 value of 0.8475. When one wavelength was left out of the training data, the RMSE ranged from 62.71 to 95.53 ppm. When the model was used the concentrations above 400 ppm, the RMSE increased to 99.92 ppm. The results show that the physical sensor model can be combined with machine learning to estimate acetone concentration. However, the study is based only on simulated data. Experiments with a real sensor are therefore needed to confirm the findings.

I. INTRODUCTION

There are several reasons to monitor acetone. It is a workplace hazard, it affects indoor air quality, it is monitored in process control and it is a non-invasive marker in exhaled breath that has been correlated with blood glucose in healthy people and in patients with type 1 diabetes [1, 2]. Metal oxide semiconductor sensors are approachable for these applications. However many conventional metal oxide sensors must operate at elevated temperatures to activate oxygen adsorption and the associated surface reactions. This requirement increases power consumption. It can cause thermal drift, packaging difficulties, and long term stability issues. Ultraviolet photoactivation may be a possible alternative. Because at lower substrate temperatures, UV illumination can change carrier populations, surface oxygen, and adsorption desorption processes. β-Ga_2O_3 is an ultrawide bandgap semiconductor with high chemical and thermal stability. It also has a large breakdown field and surface electrical properties that are sensitive to oxygen vacancies and adsorbed species [3–7]. Because of these properties it can be used in power devices, photodetectors and gas sensors. Ga2O3 gas sensors basically need elevated operating temperatures and may be affected by humidity, limited selectivity and slow recovery [3–7].

UV illumination has also been used to improve oxygen sensing in β Ga_2O_3 nanowires. That shows photoactivation can influence the surface adsorption and charge transfer processes of this material [8]. Earlier experimental studies on acetone and other volatile organic compounds (VOCs) provide a useful comparison. Dyndal et al. studied CuO– Ga_2O_3 composite films and found better response to acetone at

elevated temperatures. It shows that heterostructure devices can improve gas sensitivity [9]. Krawczyk et al. examined β Ga_2O_3 nanowires exposed to some VOCs. It shows the highest responses to 100 ppm ethanol and acetone at approximately 760 and 690°C, respectively [10]. These results verify the sensitivity of Ga_2O_3 surfaces to VOCs, and the need of reducing the operating temperature.

The change of interfacial charge in Schottky contacts can modify the effective barrier height and the measured current. Hydrogen-dependent shifts can be seen in Pt/ Ga_2O_3/SiC Schottky structures in forward voltage at elevated temperatures [11]. Reversible and crystallographic-orientation-dependent hydrogen responses [12] have also been reported for Pt contacts on bulk β Ga_2O_3. These studies support barrier modulation as a gas-sensing mechanism. Kalarickal et al. reported modulation-doped heterostructures with mobile sheet-charge density above 1013 cm^{-2} and an effective room-temperature mobility of 150 $cm2V^{-1}$ s^{-1} [13].

Al fraction (x = 0.17) used here and it shows that an AlGaO barrier can confine and redistribute carriers. Hao et al. measured an Urbach tail near the β Ga_2O_3 absorption edge and a defect-related absorption band from 2.5 to 4.5 eV [14]. Recent calibration studies also provide useful examples for linking simulation to machine learning. Zhuang et al. combined a surface-state model with a gated recurrent unit and evaluated the theory guided regression on experimental pulse heated gas sensor data [15]. Koziel et al. calibrated low cost NO_2 sensors against independent reference stations over a five-month field campaign including environmental covariates [16]. This study checks consistency within the simulations. It distinguishes simulation based calibration from experimental calibration. It investigates the numerical behaviour of a UV-activated Ni/ $(Al_{0.17}Ga_{0.83})_2O_3$/ Ga_2O_3 junction based acetone sensing framework.

The main contributions are: (i) a reduced-order formulation that combines Schottky junction transport with surface adsorption and reaction kinetics; (ii) simulations of voltage, temperature, concentration, wavelength, and time-dependent behaviour; and (iii) a reproducible machine learning analysis to test the strengths and limitations of physics-based virtual-calibration data. Here, the machine-learning inputs are generated from the adsorption, barrier-modulation, and Schottky transport equations used for the sensor simulations. This links the physical simulations to machine learning. The simulated range of 10–500 ppm is also a safety and process-monitoring range, well above the low-ppm and sub-ppm levels typical of breath, so biomarker detection motivates this work

II. MODEL AND COMPUTATIONAL METHODS

A. **Device and Junction Representation**

The proposed sensor Ni/$(Al_{0.17}Ga_{0.83})_2O_3$/$Ga_2O_3$ consists of a lateral metal semiconductor metal (MSM) heterostructure. The Al mole fraction fixed at (x = 0.17). It includes two Ni electrodes, each 0.5 mm wide and 200 nm thick, are deposited on the top surface and separated by a 1 mm gap. The semiconductor stack comprises of a 25 nm thick n type $(Al_{0.17}Ga_{0.83})_2O_3$ barrier and sensing layer, a 1.2 μm thick n type

Ga_2O_3transport layer, and a 50 nm thick Al_2O_3 buffer layer grown on a 500 μm thick Si(111) substrate. The device has lateral dimensions of 5 mm × 5 mm. Using an out of plane width of 5 mm, the contact area at each Ni/semiconductor interface is 2.5 mm^2 (0.025 cm^2). Uniform donor concentrations of 1.0 × 10^{16} and 1.0 × 10^{17} cm^{-3} are assigned to the $(Al_{0.17}Ga_{0.83})_2O_3$ barrier and Ga_2O_3 transport layers, respectively. Both Ni electrodes are assumed to form rectifying contacts with the underlying semiconductor. The resulting MSM structure can therefore be represented as two coupled Schottky junctions connected in a back to back configuration. The present analysis uses a geometry informed, reduced order transport model. The device dimensions and contact area are used for the thermionic emission calculation. Ni work function of 5.15 eV is used, while the semiconductor electron affinity is assumed to lie between 4.0 and 4.3 eV. Table 1 shows Geometrical, material, and doping parameters of the proposed sensor.

Table 1
Geometrical, material, and doping parameters specified for the proposed sensor.

Component	Specified value	Role in the model
Ni contacts	Two top contacts; 0.5 mm wide; 200 nm thick	Schottky electrodes and adsorption-active surfaces
$(Al_{0.17}Ga_{0.83})_2O_3$	25 nm; N_D = 1.0 × 10^{16} cm^{-3}	Barrier and sensing layer
Ga_2O_3	1.2 μm; N_D = 1.0 × 10^{17} cm^{-3}	Lateral transport layer
Al_2O_3	50 nm	Buffer/passivation layer
Si(111)	500 μm	Mechanical substrate
Device footprint	5 mm × 5 mm	Lateral geometry
Contact separation	1 mm	Nominal lateral conduction length
Out-of-plane width	5 mm	Conversion from 2D current per width to terminal current
Area per Ni contact	2.5 mm^2 = 0.025 cm^2	Thermionic-emission area

The ideal Schottky barrier height was calculated using the Schottky–Mott relation as follows:

$$\Phi_{B,ideal} = \Phi_m - \chi \quad (1)$$

Assuming thermionic-emission, the reverse-saturation current is given by

$$I_0 = A\, A^{*}\, T^2 \exp[-q\Phi_B/(k_B T)] \quad (2)$$

Here, A is the contact area, A^* is the effective Richardson constant, q is the elementary charge, K_B is Boltzmann's constant, and Φ_B is the barrier height. Table 2 gives the barrier heights, ideality factors and saturation currents at 373 K that we used for the dark and UV-illuminated cases.

Table 2
Effective Schottky-barrier heights, ideality factors, and calculated saturation currents used for the dark and UV-illuminated conditions at 373 K.

Condition	**Φ_B (eV)**	**n**	**I_0 at 373 K (A)**
Dark	1.05	1.18	9.29×10^{-10}
UV, 365 nm	1.00	1.15	4.40×10^{-9}
UV, 300 nm	0.95	1.12	2.09×10^{-8}
UV, 254 nm	0.90	1.10	9.88×10^{-8}

The ideal forward-current relation is

$$I = I_0 \{\exp[q(V - I R_S)/(n K_B T)] - 1\} + (V - I R_S)/R_{sh}, \quad (3)$$

Where, n is the ideality factor, R_S is the effective series resistance, and R_{Sh} is the effective shunt resistance. Eq. (3) must be solved when R_S is nonzero. These resistances prevent the ideal thermionic emission current from increasing without bound and represent the combined series and leakage paths of the reduced order device.

For the Figs. 4–5, the thermionic emission area was set equal to the geometric area of one Ni contact: A = (0.5 mm)(5 mm) = 2.5 mm² = 0.025 cm^2. The remaining parameters were $A^* = 41.1$ A cm^{-2} K^{-2}, R_S = 500 Ω, and $R_{Sh} = 1.0 \times 10^9$ Ω. For each applied voltage, Eq. (3) was solved for current with Brent's bracketing algorithm on $I \in [-0.02, 0.02]$ A, absolute and relative current tolerances of 1.0×10^{-12}, and a maximum of 200 iterations. The exponential argument was clipped to [− 100,100] to prevent numerical overflow. These are summarised in Table 3.

Table 3
Electrical and numerical parameters used to calculate the current voltage characteristics

Assumed quantity	Value	Unit / numerical setting
Area per Ni contact, A	0.025	Cm^2 (2.5 mm^2)
Richardson constant, A*	41.1	A cm^{-2} K^{-2}
Series resistance, R_S	500	Ω
Shunt resistance, R_{sh}	1.0×10^9	Ω
Root solver	Brent bracketing	I bracket: −0.02 to 0.02 A
Convergence	xtol = rtol = 10^{-12}	maximum 200 iterations

A physical MSM structure consists of two Schottky contacts facing each other and a complete process would simultaneously solve the voltage partitioning and current continuity of both junctions. But we used a simpler effective junction approximation here. Rs and Rsh are the series and the leakage paths of the lateral device. The reverse biased contact controls the barrier limited current. This is appropriate for comparison of relative changes due to temperature, wavelength and gas induced barrier modulation. It is not a spatially resolved two contact solution and the currents it yields should not be considered as absolute predictions for a fabricated device. Within the approximation, forward current at high bias is limited by R_S, and R_{sh} sets the low reverse bias.

B. **Ultraviolet and Surface Model**

UV activation is represented by a wavelength dependent carrier generation factor and an effective lowering of the barrier. We ran calculations at 254, 300 and 365 nm which correspond to photon energies of about 4.88, 4.13 and 3.40 eV. Photons at 300 and 365 nm lie below the nominal β-Ga_2O_3 bandgap. So any response at these wavelengths has to come from defect assisted, surface state or other sub bandgap absorption; both energies fall within the 2.5–4.5 eV defect band reported in [14]. The effective barrier is

$$\Phi_{B,eff} = \Phi_{B,dark} - \Delta\Phi UV(\lambda) - \Delta\Phi_{gas}(C,T,\lambda) \quad (4)$$

At 373 K, 0.15 eV barrier reduction at 254 nm raises I_0 by a factor of about 106 (Table 2). Equilibrium acetone coverage is described by a Langmuir form,

$$\theta(C,T) = K(T)\, C / [1 + K(T)\, C], \quad (5)$$

and an effective surface-reaction coefficient by

$$k(T) = k_0 \exp[-E_a/(k_B T)] \quad (6)$$

These are reduced order relations. The reducing-gas reaction is

$CH_3COCH_3 + 8O^-(ads) \rightarrow 3CO_2 + 3H_2O + 8e^-$ .....(7)

Figure 1 shows how UV illumination changes the band structure at one $Ni/(Al_{0.17}Ga_{0.83})_2O_3/Ga_2O_3$ interface of the lateral device. In the dark (Fig. 1a) upward band bending near the Ni contact produces a depletion region and an effective Schottky barrier Φ_B. There are few mobile carriers in that region that limit charge transport across the interface. Under UV (Fig. 1b) absorbed photons create extra electrons and holes through intrinsic or defect assisted excitation. These carriers can redistribute the charge at the interface and partly relax the band bending.

The sequence of surface and interfacial processes of the sensing mechanism are laid out in Fig. 2. First, UV photons generate electrons and holes in the semiconductor through the photoactivation mechanism. Some of the photogenerated electrons take part in the adsorption and ionisation of oxygen molecules at the surface. The adsorbed oxygen captures electrons from the semiconductor, which creates a surface depletion region and raises the effective energy barrier. When acetone reaches the activated surface it reacts with the adsorbed oxygen species (Eq. 7), forming carbon dioxide and water and releasing the trapped electrons back into the semiconductor. The returned electrons raise the mobile carrier concentration, reduce the surface depletion and the effective barrier height, and so increase the device current.

C. Numerical Sweeps and Transient Response

We ran a set of deterministic sweeps to see how the response depends on acetone concentration, operating temperature, UV wavelength and applied voltage. The concentrations were 10, 25, 50, 100, 200, 300, 400 and 500 ppm. Temperatures were 298, 348, 373 and 398 K (25, 75, 100 and 125°C), wavelengths were 254, 300 and 365 nm, and the voltage was swept from − 2 to + 2 V. The lowest simulated concentration 10 ppm should not be taken as a limit of detection or quantification of the sensor. The percentage response to acetone is

$S = 100\ (I_{gas} - I_{air})/I_{air}$ .....(8)

Here I_{gas} is the current in acetone and I_{air} is the baseline current without acetone, both evaluated at the same voltage, temperature and UV wavelength. A positive S means that acetone increases the current; a negative S means it decreases it. Transients were described with first-order exponentials. During acetone exposure the response approaches its steady state as

$S_{rise}(t) = S_o + (S_{ss} - S_o)\{1 - \exp[- t/\tau_{rise}]\}$, ...(9)

Where, S_0 is the response before exposure, S_{ss} is the steady state response in acetone, and τ_{rise} is the rise time constant. After acetone was removed recovery toward the baseline was given by

$S_{recovery}(t) = S_{base} + (S_{initial} - S_{base}) \exp[-t/\tau_{recovery}]$..... (10)

In Eq. (10), $S_{initial}$ is the response at the start of recovery, S_{base} is the recovered baseline, and $\tau_{recovery}$ is the recovery time constant. For a first-order process, the time needed to complete 90% of the total signal change is given as

$t_{90} = -\tau \ln(0.10) = \ln(10)\tau \approx 2.303\tau$ ..... (11)

So τ_{rise} = 5.21 s corresponds to a 90% response time of about 12.0 s and $\tau_{recovery}$ = 7.82 s to a 90% recovery time of about 18.0 s. The adsorption, UV, barrier-modulation and transient parameters assigned in the model are listed in Table 4.

Table 4
Adsorption, UV-activation, barrier-modulation, and transient parameters used in the reduced-order sensing model

Process	Parameter	value	Unit/interpretation
Adsorption	Effective Langmuir coefficient at 298 K, K_{298}	0.010063	ppm^{-1}
Adsorption	Effective Langmuir coefficient at 348 K, K_{348}	0.012767	ppm^{-1}
Adsorption	Effective Langmuir coefficient at 373 K, K_{373}	0.014735	ppm^{-1}
Adsorption/desorption	Effective Langmuir coefficient at 398 K, K_{398}	0.014556	ppm^{-1}
Surface coverage	$\theta(C,T)$	$K(T)C/[1 + K(T)C]$	Dimensionless
Reaction and charge transfer	Effective coverage-to-barrier coefficient at 298 K, β_{298}	0.017227	eV
Reaction and charge transfer	Effective coverage-to-barrier coefficient at 348 K, β_{348}	0.035713	eV
Reaction and charge transfer	Effective coverage-to-barrier coefficient at 373 K, β_{373}	0.041302	eV
Reaction/desorption balance	Effective coverage-to-barrier coefficient at 398 K, β_{398}	0.041534	eV
Gas-induced barrier modulation	$\Delta\Phi_{gas}(C,T)$	$\beta(T)\theta(C,T)$	eV
Dark condition	$\Phi_{B,dark}$	1.05	eV
UV activation at 365 nm	$\Delta\Phi_{UV}$	0.05	eV
UV activation at 300 nm	$\Delta\Phi_{UV}$	0.10	eV
UV activation at 254 nm	$\Delta\Phi_{UV}$	0.15	eV
Effective barrier at 365 nm	$\Phi_{B,UV}$	1.00	eV

Process	Parameter	value	Unit/interpretation
Effective barrier at 300 nm	$\Phi_{B,UV}$	0.95	eV
Effective barrier at 254 nm	$\Phi_{B,UV}$	0.90	eV
Response transient	τ_{rise}	5.21	s
Recovery transient	$\tau_{recovery}$	7.82	s
Calculated response time	$t_{90,rise}$	12.0	s
Calculated recovery time	$t_{90,recovery}$	18.0	s
Transient concentration	C	100	ppm
Transient temperature	T	373	K
Transient illumination	λ	254	nm
Number of simulated cycles	N_{cycles}	3	Deterministic cycles
Virtual-calibration UV input	Photon flux, *F*	0.5×10^{15} to 1.5×10^{15}	photons cm^{-2} s^{-1}; calibrated simulation input

D. **Dataset and Machine-Learning Protocol**

We built a physics generated dataset of 2000 observations with the NumPy generator default_rng, seeded with 42. Acetone concentration was drawn independently from a uniform distribution, C ~ U (10,500) ppm. Temperature was sampled between 298 and 398 K, wavelength was taken from 254, 300 and 365 nm, photon flux was sampled between 0.5×10^{15} and 1.5×10^{15} photons cm^{-2} s^{-1}, and forward bias between 0.05 and 0.35 V. Equilibrium coverage came from Eq. (5), with K(T) and β(T) interpolated linearly between the four temperature specific values in Table 4. To represent modest device to device or surface state variability, an independent lognormal perturbation with a logarithmic standard deviation of 0.02 for K and β is considered. Each independently sampled value was passed through the forward sensor model and generates the electrical observables. The UV induced barrier reduction scales with photon flux as

$$\Delta\Phi UV(\lambda,F) = \Delta\Phi UV,ref(\lambda)\ F/Fref \quad (12)$$

with $F_{ref} = 1.0 \times 10^{15}$ photons cm^{-2} s^{-1}. The reference reductions were 0.15, 0.10 and 0.05 eV at 254, 300 and 365 nm respectively. The air state barrier is $\Phi_{B,air} = \Phi_{B,dark} - \Delta\Phi_{UV}$ and the gas state barrier is $\Phi_{B,gas} = \Phi_{B,air} - \beta(T)\theta(C, T)$. A normally distributed perturbation of 0.003 eV was applied to $\Phi_{B,dark}$ and 1% multiplicative lognormal readout uncertainty was applied independently to the calculated air and gas currents. The concentrations in the samples were between 10.25 and 499.72 ppm and the simulated

responses were between 2.84% and 222.78%. Seven predictors were used. They are temperature, UV wavelength, calibrated photon flux, bias, log10(I_{air}), log10(I_{gas}) and ln(I_{gas}/I_{air}). We used log currents because Schottky transport is exponential in barrier height and current spans several orders of magnitude. Each observation was assigned a stable identifier OBS0001 to OBS2000. A shuffled 80:20 split with random state = 42 produced 1600 development observations and 400 independent holdout observations. The model selection and hyperparameter optimization were only done on the development set of 1,600 rows. We used nested cross validation for estimating development performance with five shuffled outer folds (seed 43) and for selecting the hyperparameters with five shuffled inner folds (seed 42). Nesting prevents the optimization from leaking into the performance estimate [18]. For elastic-net regression [17], support vector regression (SVR) and the multilayer perceptron (MLP), all of the operations were fitted within the corresponding training folds and features were standardized within each fold. Hyperparameters were selected by minimizing the root mean square error (RMSE) on the inner fold using GridSearchCV in scikit-learn [19]. After nested cross validation each selected pipeline was refitted on the complete development set. Then it was evaluated once on the untouched 400 observation holdout set. Uncertainty intervals for the selected model's holdout metrics came from 5000 row level bootstrap resamples that were drawn with default_rng (42). We compared the Paired absolute holdout errors with two sided Wilcoxon signed rank tests. Holm adjustment across the seven comparisons was done also. The barrier lowering (0.15 eV) increases I_0 by a factor of 106 (approx) at 373K. At high bias the forward current is limited by Rs and the low reverse bias is controlled by R_{sh}. The calculation steps used in this study is shown in Fig. 3. The method includes six linked stages. Those are device specification, junction transport, surface kinetics, numerical simulation, dataset and machine learning analysis and performance evaluation.

III. RESULTS AND DISCUSSION

A. **Simulated Current–Voltage Characteristics**

Figure 4 shows the simulated current–voltage characteristics under dark conditions and UV illumination at wavelengths of 365, 300 and 254 nm. The calculations used a thermionic-emission area of A = 0.025 cm^2 per Ni contact, an effective Richardson constant of $A^* = 41.1$ A cm^{-2} K^{-2}, a series resistance of R_S = 500 Ω and a shunt resistance of R_{sh} = 1.0 GΩ. In the reverse bias region the current approaches the saturation current associated with each illumination condition together with the leakage contribution through the shunt resistance. Differences among the reverse bias currents arise mainly from the effective barrier heights. Under forward bias the current increases rapidly as the applied voltage approaches the turn on region. The curves are clearly separated near turn on because the thermionic emission current depends exponentially on the effective Schottky barrier height and ideality factor. At higher forward voltages the curves gradually converge because the voltage drop across the 500 Ω series resistance limits the current. The 254 nm condition produces the highest simulated current whereas the dark condition produces the lowest.

B. **Temperature Dependence**

Figure 5 shows the influence of temperature on the simulated electrical characteristics. Figure 5(a) presents current voltage curves under 254 nm UV illumination at 298, 348 and 373 K. For these calculations the effective barrier height and ideality factor were 0.90 eV and 1.10, respectively. Increasing temperature raises the thermionic emission saturation current and shifts the apparent turn on region toward a lower forward voltage. At higher forward bias the influence of temperature becomes less pronounced. Figure 5(b) shows the corresponding Richardson representation in which $\ln(I_0/T^2)$ is plotted against 1000/T over 250 to 450 K. The slope of the resulting straight line is related to the effective barrier height. Linear regression of the simulated Richardson plots gives barrier heights of 1.05 eV for the dark condition and 0.90 eV under 254 nm illumination.

C. **Concentration and Temperature Response**

Figure 6 shows the simulated response as a function of acetone concentration and operating temperature. At every temperature the response increases as acetone concentration rises from 10 to 500 ppm. At low concentration there are still many free adsorption sites. Therefore the response increases rapidly with acetone molecules. As the sites fill up the response becomes sublinear and gradually approaches saturation. Temperature also affects the response. The response increases with the increase of temperature from 298 to 373 K and the model gives maximum response at 373 K (Table 5). At 500 ppm the response ranges from around 75% at 298 K to 180% at 348 K and 210% at 373 K, then decreases slightly to 190% at 398 K.

Table 5
Simulated acetone responses at concentrations of 100 and 500 ppm

Temperature	Response at 100 ppm (%)	Response at 500 ppm (%)
298 K (25°C)	40	75
348 K (75°C)	95	180
373 K (100°C)	115	210
398 K (125°C)	105	190

D. **Ultraviolet Gain**

The UV gain is defined as $G_{UV} = (I_{UV} - I_{dark})/I_{dark}$. So a positive value means that UV exposure raises the current above the dark baseline. Figure 7 shows G_{UV} at 254 nm and 373 K. It rises from 46 at the lowest concentrations to 355 at 500 ppm, and the rise is steeper at the high-concentration end. The two shaded regions in the figure separate a low concentration range, where the gain is set mostly by UV photocarrier generation and initial surface activation from a higher range where coverage dependent barrier lowering dominates. The gain depends strongly on the dark current and optical coefficient.

E. **Transient Response**

Figure 8 shows the simulated response over three consecutive exposures to 100 ppm acetone at 373 K and 254 nm. After acetone is introduced the response approaches its steady state with τ_{rise} = 5.21 s. When acetone is removed it returns toward baseline with $\tau_{recovery}$ = 7.82 s. Using Eq. (11), $t_{90,rise}$ = 2.303 × 5.21 ≈ 12.0 s and $t_{90,recovery}$ = 2.303 × 7.82 ≈ 18.0 s. The simulated signal therefore reaches about 90% of its steady state value within 12 s of exposure and after removal about 90% of the signal change has disappeared within 18 s leaving 10% of the initial deviation.

F. **Regression Results**

Among the seven tuned regression algorithms, the MLP produced the lowest mean outer cross validation RMSE. The selected network contained hidden layers of 64 and 32 units, α = 0.001 and learning rate 0.01. After refitting on all 1600 development observations it achieved holdout RMSE 55.63 ppm, MAE 40.87 ppm and R^2= 0.8475. The corresponding 95% row bootstrap intervals were from 50.75 to 60.49 ppm, 37.29–44.61 ppm and 0.8171–0.8738 respectively. SVR, at 55.76 ppm, was practically level with it on the holdout set. Figure 9 compares predicted and reference concentrations for the holdout observations and shows the residuals, which fan out toward higher concentration. That is where the Langmuir response flattens, so a given change in current corresponds to a larger change in concentration.

Table 6
Outer-cross-validation and independent holdout performance of the evaluated regression models

Model	Outer CV RMSE	Outer CV MAE	Outer CV R^2	Holdout RMSE
Elastic net	106.02 ± 4.35	85.80 ± 2.21	0.4427 ± 0.0465	111.36
SVR	55.48 ± 1.82	40.23 ± 0.94	0.8477 ± 0.0061	55.76
Extra trees	56.47 ± 3.96	40.16 ± 2.10	0.8418 ± 0.0196	58.16
Gradient boosting	57.49 ± 2.39	43.05 ± 1.74	0.8362 ± 0.0128	60.10
Random forest	57.92 ± 2.42	40.65 ± 1.10	0.8339 ± 0.0102	60.15
MLP	50.92 ± 3.94	37.76 ± 3.60	0.8714 ± 0.0167	55.63
Histogram boosting	56.32 ± 2.26	40.13 ± 1.53	0.8428 ± 0.0108	57.74

Permutation analysis identified $\ln(I_{gas}/I_{air})$ as the strongest predictor. It increases holdout RMSE by 337.26 ± 11.15 ppm after permutation. Temperature, bias, UV wavelength, log current levels and photon flux followed with mean RMSE. It increases of 184.69, 85.15, 57.56 approximately 52 and 35.69 ppm respectively.

The all feature ordinary least squares baseline gave a holdout RMSE of 111.36 ppm, an MAE of 90.07 ppm and R^2 = 0.3888. The MLP beat this baseline significantly after Holm correction (adjusted $p < 0.001$)

which supports the nonlinear character of the inverse problem. It was not significantly different from any of the other nonlinear models (all adjusted $p \geq 0.138$). These results support the use of nonlinear models. An RMSE of 56 ppm over a 10–500 ppm range can be called modest. This error reflects the variability included in the simulation. A target built directly from the predictors would give near perfect scores. Here the residual error comes from adsorption saturation, coupled operating condition effects, the imposed parameter variability and the current readout uncertainty. Experimental calibration is still needed before any detection limit, deployment accuracy, selectivity or real world generalisation.

G. **Out-of-Distribution and Correlated-Feature Analysis**

Out of distribution tests used the selected MLP architecture without retuning. In the leave one wavelength out tests, the RMSE was 76.56, 62.71 and 95.53 ppm when 254, 300 and 365 nm respectively were excluded from training. Training below 373 K and testing from 373 to 398 K gave an RMSE of 72.95 ppm and $R^2 = 0.7357$. The most difficult case was concentration extrapolation: training up to 400 ppm and testing above it gave an RMSE of 99.92 ppm and $R^2 = -10.8086$. The negative R^2 is partly an artefact of the narrow test range, since concentrations between 400 and 500 ppm have a standard deviation of only about 29 ppm if spread uniformly. RMSE is more informative here and it shows that the model cannot be trusted to extrapolate into the strongly saturated region. The random holdout result should therefore be read as interpolation within the simulated design space and not as evidence of unrestricted generalisation.

Table 7
Out-of-distribution and correlated-feature ablation results for the selected MLP.

Evaluation	***n***	**RMSE (ppm)**	**MAE (ppm)**	**R^2**
Random holdout, all seven features	400	55.63	40.87	0.8475
Leave out 254 nm	656	76.56	60.62	0.7071
Leave out 300 nm	687	62.71	49.31	0.8040
Leave out 365 nm	657	95.53	85.37	0.5580
Test at $T > 373$ K	490	72.95	57.41	0.7357
Test at $C > 400$ ppm	423	99.92	82.80	−10.8086
Without log-current ratio	400	59.35	46.31	0.8264
Without absolute log currents	400	59.09	42.55	0.8279
Operating conditions only	400	143.11	124.20	−0.0094
Current-derived features only	400	102.37	82.88	0.4835

The current derived predictors are correlated by construction. The Pearson correlation between $\log_{10}(I_{air})$ and $\log_{10}(I_{gas})$ was 0.9976, whereas $\ln(I_{gas}/I_{air})$ correlated at − 0.0775 and − 0.0084 with the two absolute

log current features. Permuting all three current derived variables together raised the holdout RMSE by 264.69 ± 11.82 ppm, so their combined information is essential. In the ablation tests, removing the log current ratio raised the RMSE from 55.63 to 59.35 ppm and removing the two absolute log current features raised it to 59.09 ppm. Operating conditions alone gave an RMSE of 143.11 ppm, while the three current derived features alone gave 102.37 ppm. The model uses both the operating conditions and the electrical signal.

IV. CONCLUSION

We developed a reduced order computational framework for a UV activated Ni/$(Al_{0.17}Ga_{0.83})_2O_3$/$Ga_2O_3$ lateral MSM acetone sensor and used it for simulation-based calibration. Acetone concentration was sampled independently and propagated through Langmuir adsorption, temperature dependent surface parameters, UV induced barrier modulation and Schottky transport. The resulting electrical quantities and operating conditions were then used to estimate the concentration linking the physical model to the machine learning analysis. Of the seven regression methods, the multilayer perceptron gave the lowest mean outer crossvalidation RMSE (50.92 ± 3.94 ppm) and the highest outer cross validation R^2 (0.8714 ± 0.0167). Refitted on the full development set it reached a holdout RMSE of 55.63 ppm an MAE of 40.87 ppm and R^2 = 0.8475. Ordinary least squares did much worse which confirms that the inverse problem is not linear. After Holm correction though the MLP, was not statistically different from SVR, extra trees, random forest, gradient boosting or histogram gradient boosting. The results support nonlinear regression in general and do not show that the MLP is better than the other nonlinear options. For this dataset it is the selected model with the best cross validation performance and a competitive holdout result. Permutation analysis put the logarithmic gas to air current ratio first, followed by temperature, bias, UV wavelength, the absolute current levels and photon flux. This pattern agrees with the model equations. The concentration estimation depends on adsorption, optical activation, junction transport and the combined effects of the operating conditions. Performance dropped in the out of distribution tests, most of all for extrapolation above 400 ppm. So the model should be seen as an interpolator within the simulated design space. The method combines physical simulations with machine learning to study nonlinear virtual calibration. The accuracy reflects numerical consistency under the parameters and simulated measurement variability and it is not an experimental sensor calibration. Future work should replace the effective junction approximation with a coupled MSM treatment where needed, estimate the uncertain physical parameters from measurements, and validate the models with repeated acetone exposures, calibrated illumination, humidity and interfering gases, device to device variability, long term drift, and temporally or physically independent test sets. Measurements at breath relevant concentrations would be needed before the sensor can be considered for biomarker detection.

Declarations

Conflict of Interest

—The authors declare no conflict of interest.

Ethics, Consent to Participate, and Consent to Publish declarations

not applicable.

Funding—

This research received no external funding

Author Contribution

Ms. Mamataj drafted the paper and carried out the research work. Dr Bhaskar Roy conceptualised the work.Dr Deepshikha Datta analysed the data.Dr Sunipa Roy reviewed, supervised and edited the manuscript.Dr Bimal Das reviewed, supervised and edited the manuscript.

Acknowledgment—

The authors acknowledge institutional support from the National Institute of Technology Durgapur.

Data Availability Statement:

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request

References

1. Righettoni M, Tricoli A, Pratsinis SE. Si:WO_3 sensors for highly selective detection of acetone for easy diagnosis of diabetes by breath analysis. *Anal Chem*. 2010;82(9):3581–7. https://doi.org/10.1021/ac902695n
2. Rydosz A. A negative correlation between blood glucose and acetone measured in healthy and type 1 diabetes mellitus patient breath. *J Diabetes Sci Technol*. 2015;9:881–4. https://doi.org/10.1177/1932296815572366
3. Zhu J, Xu Z, Ha S, Li D, Zhang K, Zhang H, et al. Gallium oxide for gas sensor applications: a comprehensive review. *Materials*. 2022;15(20):7339. https://doi.org/10.3390/ma15207339
4. Pearton SJ, Yang J, Cary PH IV, Ren F, Kim J, Tadjer MJ, et al. A review of Ga_2O_3 materials, processing, and devices. *Appl Phys Rev*. 2018;5:011301. https://doi.org/10.1063/1.5006941
5. Higashiwaki M, Sasaki K, Kuramata A, Masui T, Yamakoshi S. Gallium oxide (Ga_2O_3) metal-semiconductor field-effect transistors on single-crystal β-Ga_2O_3 (010) substrates. *Appl Phys Lett*.

2012;100:013504. https://doi.org/10.1063/1.3674287
6. Zhai H, Wu Z, Fang Z. Recent progress of Ga_2O_3-based gas sensors. *Ceram Int*. 2022;48(17):24213–33. https://doi.org/10.1016/j.ceramint.2022.06.066
7. Afzal A. β-Ga_2O_3 nanowires and thin films for metal oxide semiconductor gas sensors: sensing mechanisms and performance enhancement strategies. *J Materiomics*. 2019;5(4):542–57. https://doi.org/10.1016/j.jmat.2019.08.003
8. Cheng Y, Wang W, Zhu F, Yu T, Zhang X. UV-enhanced O_2 sensing using β-Ga_2O_3 nanowires at room temperature. *Sens Actuators A Phys*. 2024;379:115997. https://doi.org/10.1016/j.sna.2024.115997
9. Dyndal K, Zarzycki A, Andrysiewicz W, Grochala D, Marszalek K, Rydosz A. CuO–Ga_2O_3 thin films as a gas-sensitive material for acetone detection. *Sensors*. 2020;20(11):3142. https://doi.org/10.3390/s20113142
10. Krawczyk M, Suchorska-Woźniak P, Szukiewicz R, Kuchowicz M, Korbutowicz R, Teterycz H. Morphology of Ga_2O_3 nanowires and their sensitivity to volatile organic compounds. *Nanomaterials*. 2021;11(2):456. https://doi.org/10.3390/nano11020456
11. Trinchi A, Wlodarski W, Li YX. Hydrogen sensitive Ga_2O_3 Schottky diode sensor based on SiC. *Sens Actuators B Chem*. 2004;100(1–2):94–8. https://doi.org/10.1016/j.snb.2003.12.028
12. Jang S, Jung S, Kim J, Ren F, Pearton SJ, Baik KH. Hydrogen sensing characteristics of Pt Schottky diodes on (−201) and (010) Ga_2O_3 single crystals. *ECS J Solid State Sci Technol*. 2018;7(7):Q3180–2. https://doi.org/10.1149/2.0261807jss
13. Kalarickal NK, Xia Z, McGlone JF, Liu Y, Moore W, Arehart AR, et al. High electron density β-$(Al_{0.17}Ga_{0.83})_2O_3/Ga_2O_3$ modulation doping using ultra-thin (1 nm) spacer layer. *J Appl Phys*. 2020;127:215706. https://doi.org/10.1063/5.0005531
14. Hao S, et al. Sub-bandgap optical spectroscopy of epitaxial β-Ga_2O_3 thin films. *Appl Phys Lett*. 2020;116:092102. https://doi.org/10.1063/1.5143393
15. Zhuang Y, Yin D, Wu L, Niu G, Wang F. A deep learning approach for gas sensor data regression: incorporating surface state model and GRU-based model. *APL Mach Learn*. 2024;2:016104. https://doi.org/10.1063/5.0160983
16. Koziel S, Pietrenko-Dabrowska A, Wojcikowski M, et al. High-performance machine-learning-based calibration of low-cost nitrogen dioxide sensor using environmental parameter differentials and global data scaling. *Sci Rep*. 2024;14:26120. https://doi.org/10.1038/s41598-024-77214-y
17. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x
18. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11:2079–107.
19. Pedregosa F, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.

Figures

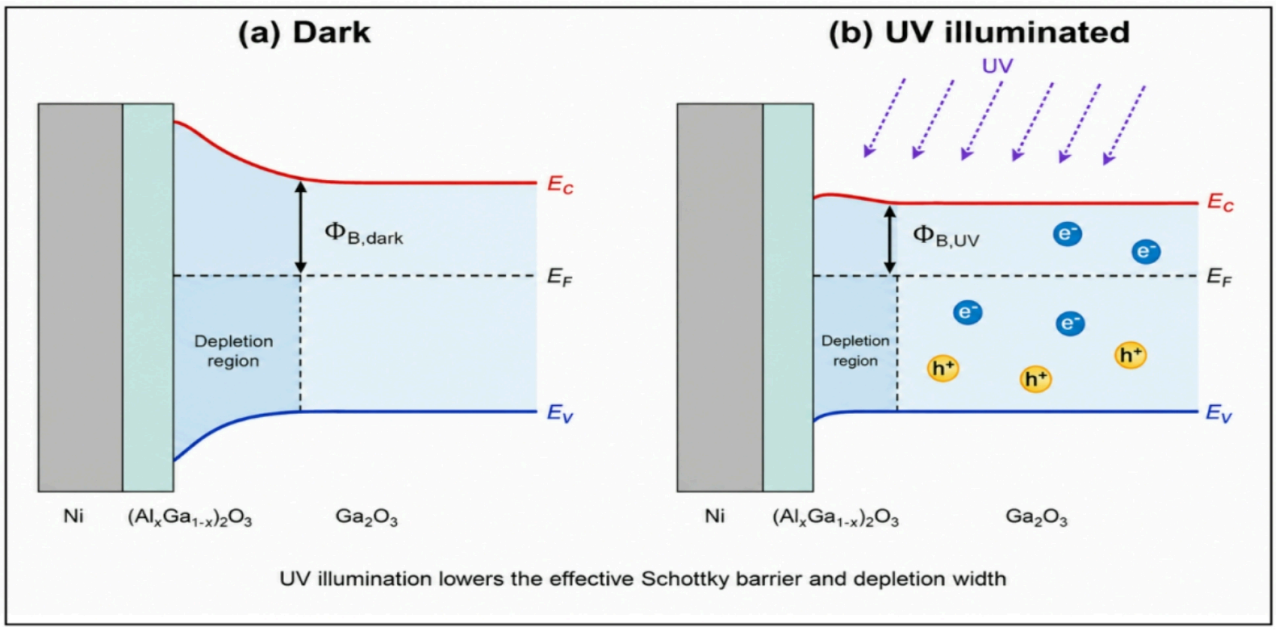

Figure 1

Conceptual energy-band representation of one Ni/$(Al_{0.17}Ga_{0.83})_2O_3$/$Ga_2O_3$ interface in the lateral MSM device under (a) dark and (b) UV-illuminated conditions.

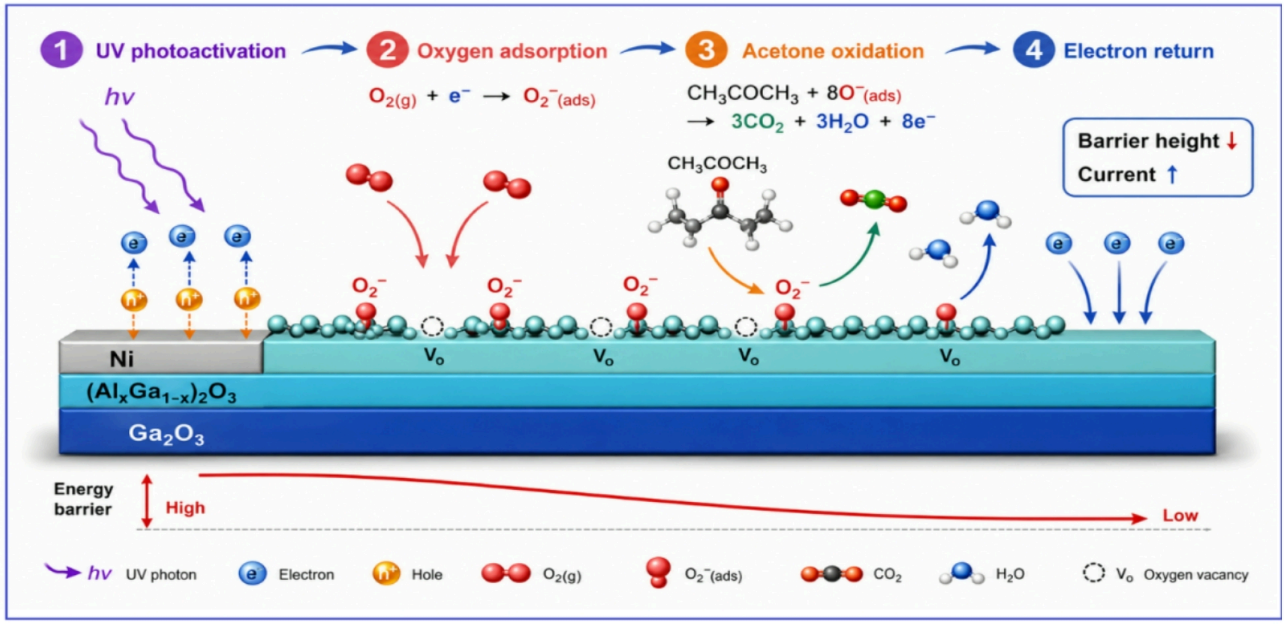

Figure 2

Ultraviolet-assisted acetone-sensing mechanism. Photocarriers and surface oxygen chemistry are represented through effective carrier-generation, coverage, and barrier-modulation terms.

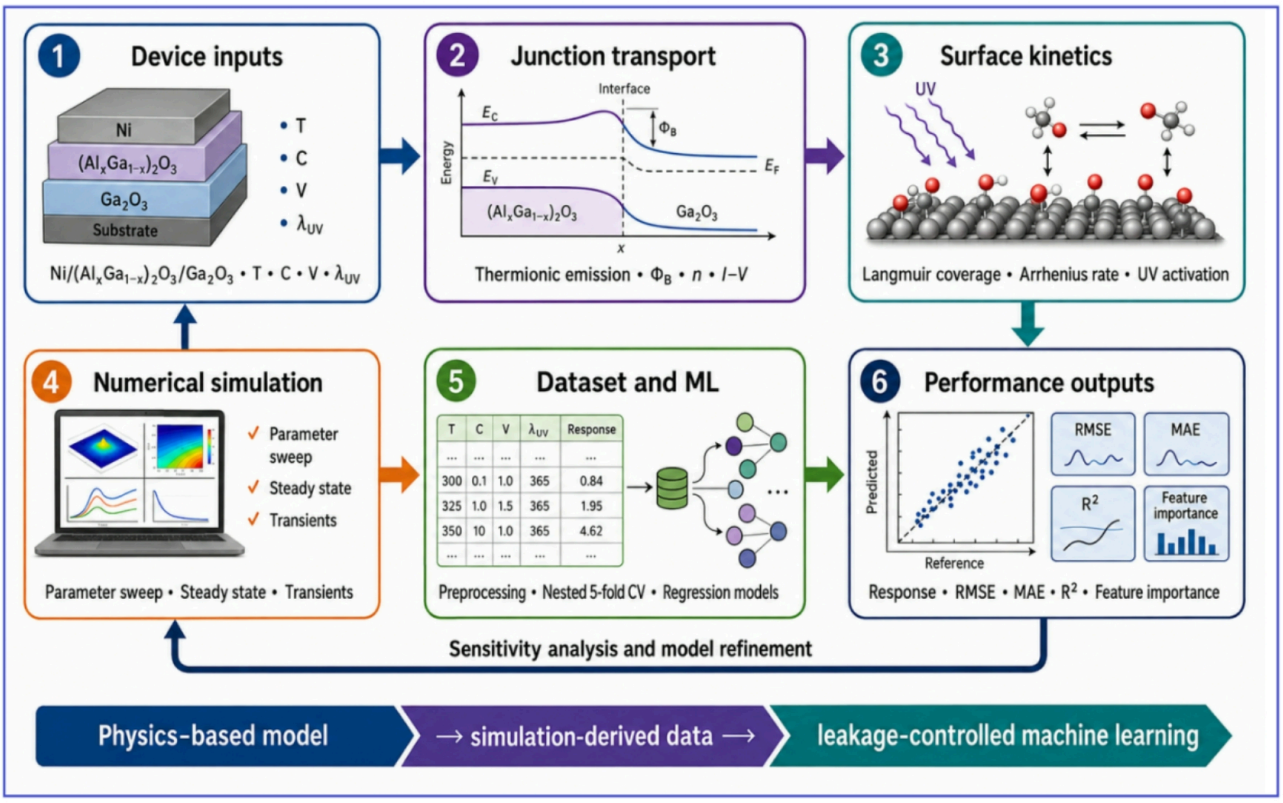

Figure 3

Computational workflow linking model inputs, Schottky transport, phenomenological surface kinetics, dataset generation, regression, and evaluation.

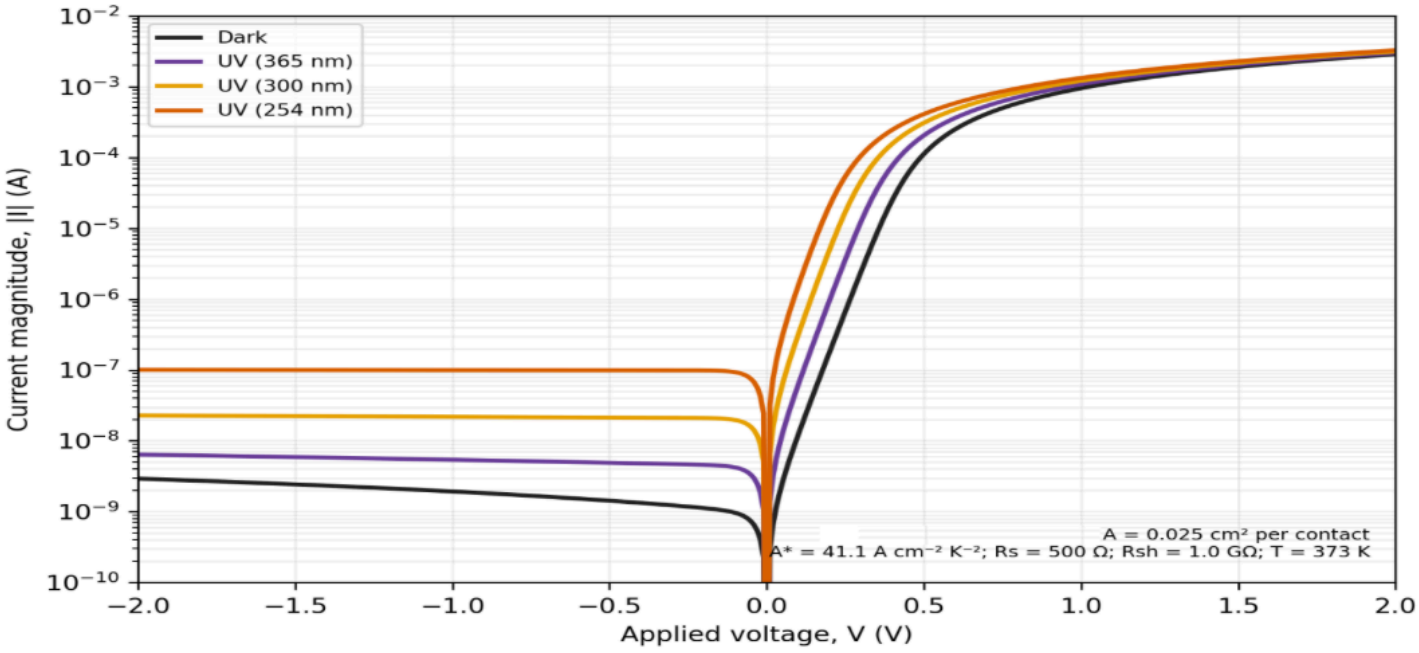

Figure 4

Simulated current voltage characteristics of the effective Ni/$(Al_{0.17}Ga_{0.83})_2O_3$/$Ga_2O_3$ Schottky junction under dark conditions and UV illumination at 365, 300, and 254 nm at 373 K.

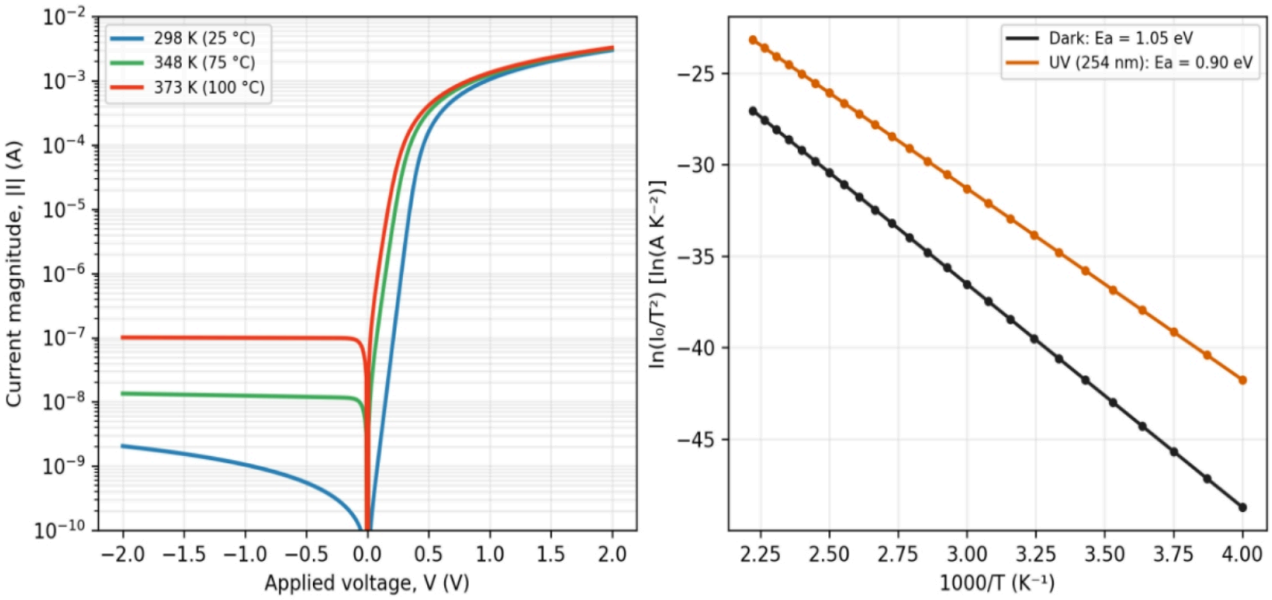

Figure 5

Temperature analysis: (a) 254-nm current–voltage curves at 298, 348, and 373 K and (b) Arrhenius representation of the assumed dark and 254-nm saturation currents.

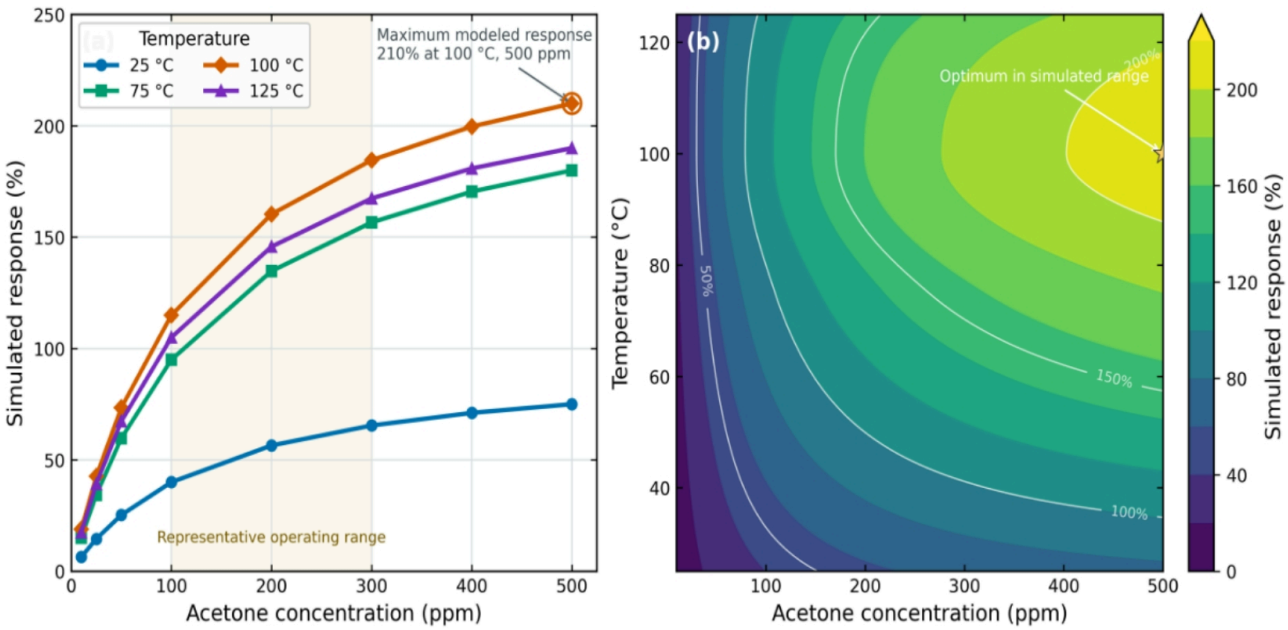

Figure 6

Simulated acetone response versus concentration and temperature: (a) Langmuir-type curves and (b) response contour. The 10-ppm point is the lowest simulated concentration and is not a statistical limit of detection.

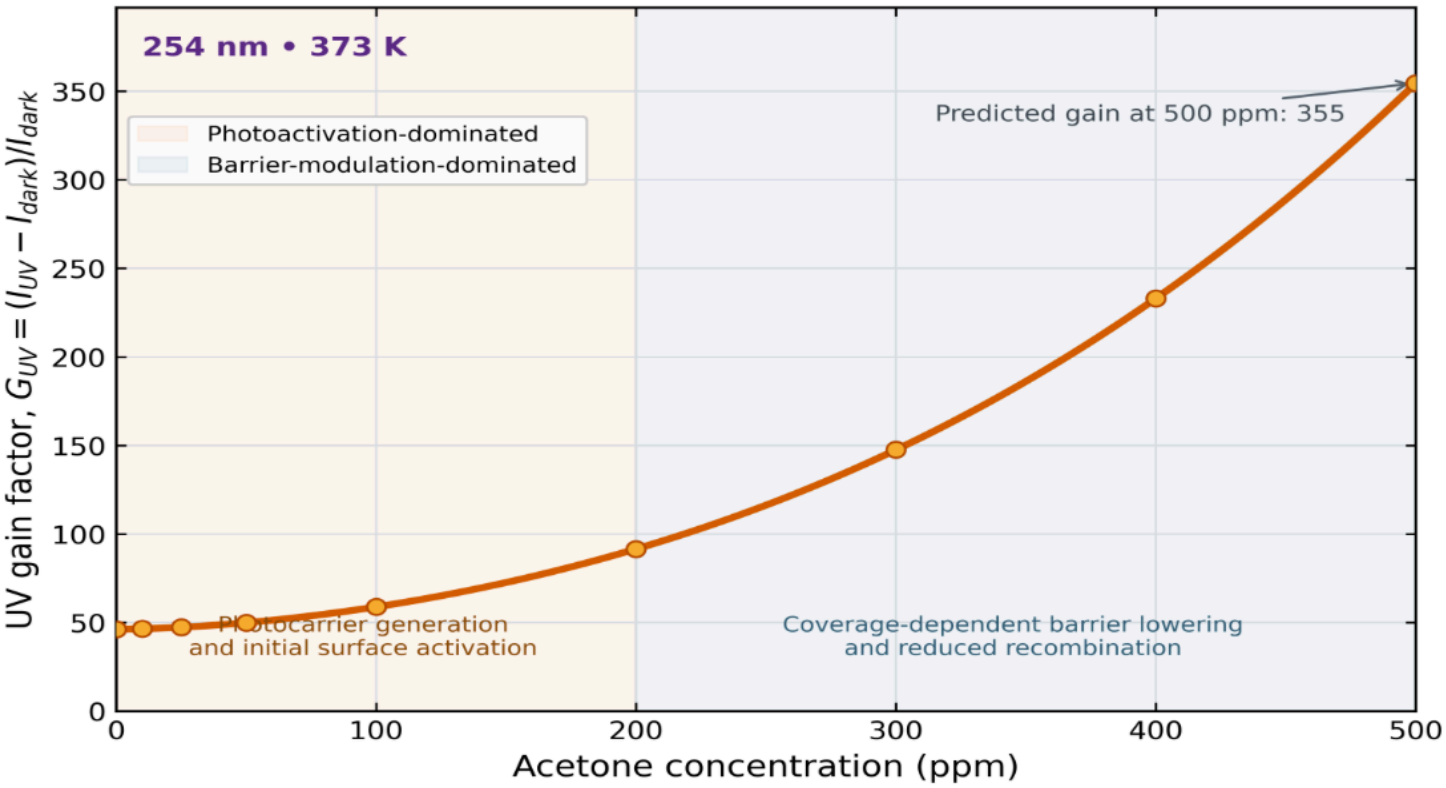

Figure 7

Simulated ultraviolet gain versus acetone concentration at 373 K under 254-nm illumination

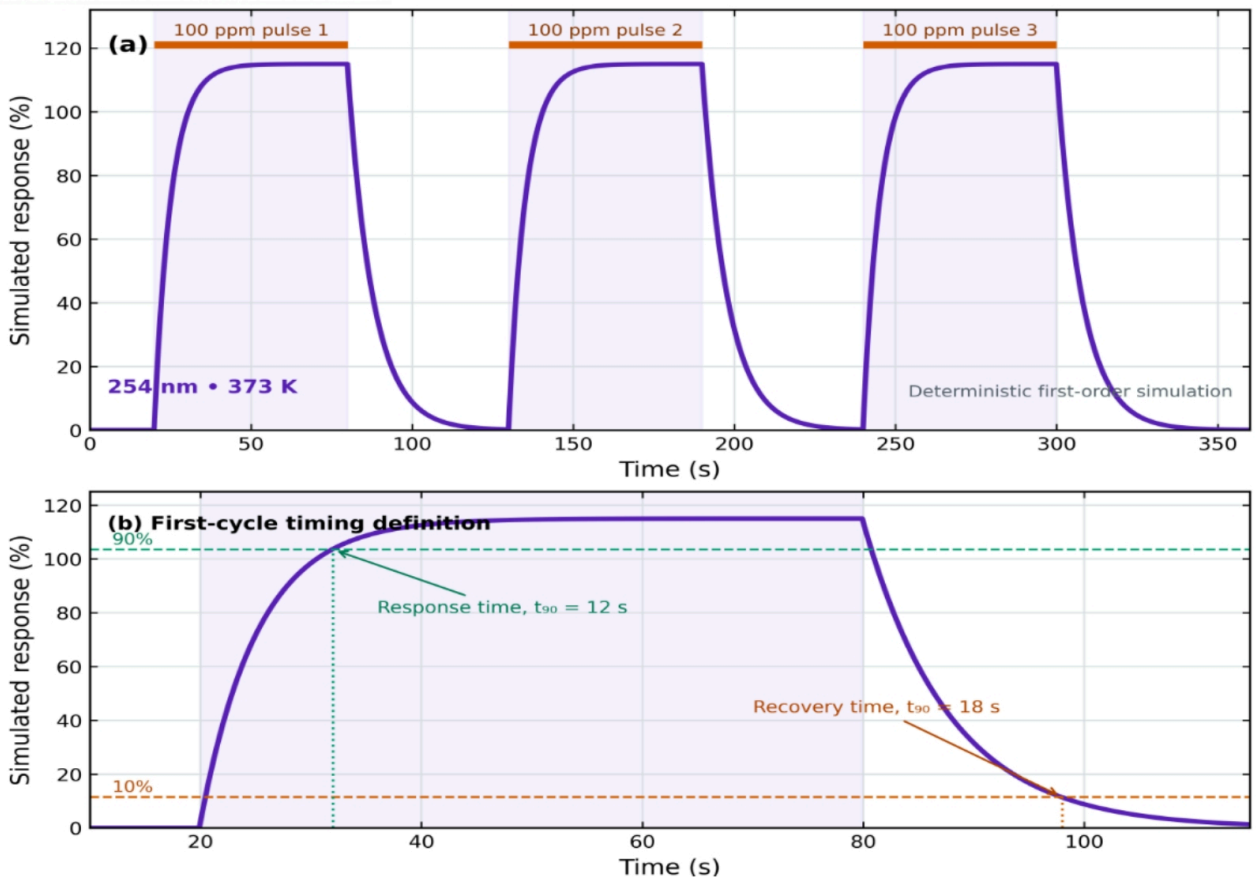

Figure 8

Deterministic first order response and recovery simulation for three 100 ppm acetone exposures at 373 K and 254 nm: (a) three consecutive cycles and (b) timing definitions for the first cycle

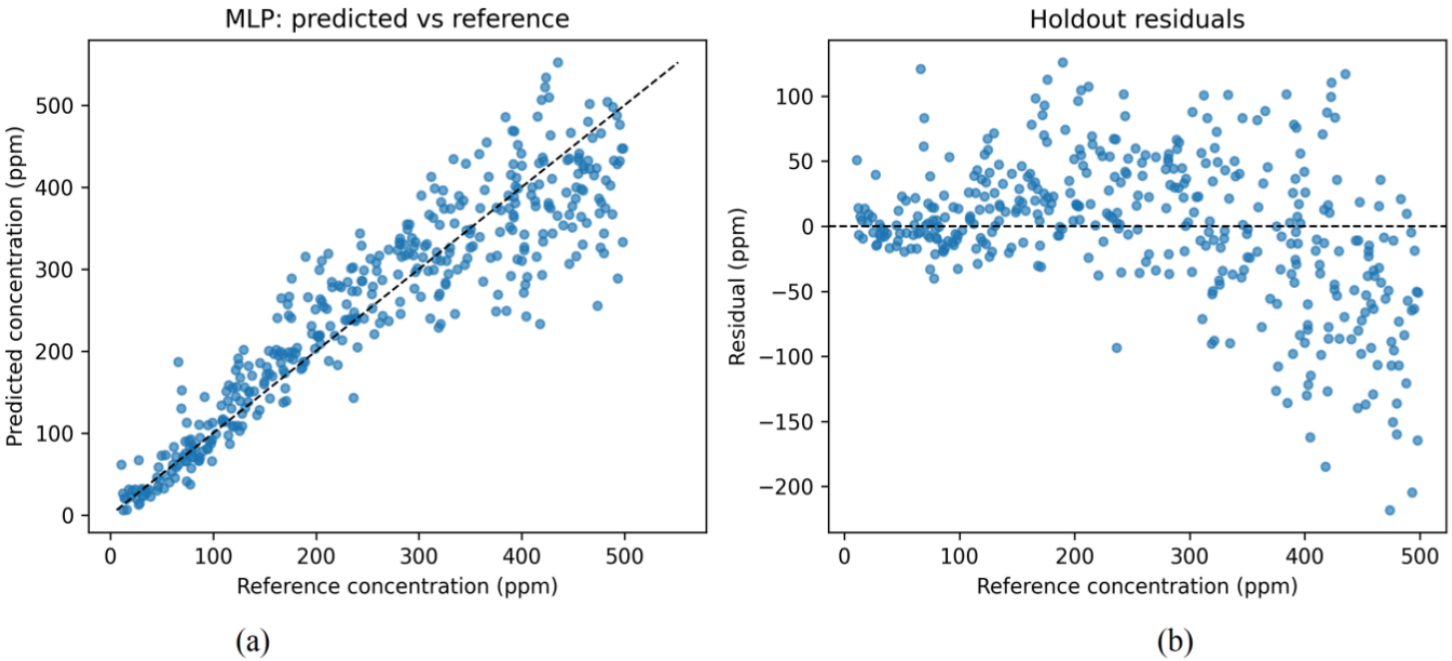

Figure 9

Holdout performance of the selected MLP pipeline for 400 physics-generated virtual calibration observations: (a) predicted versus reference concentration and (b) residuals versus reference concentration.

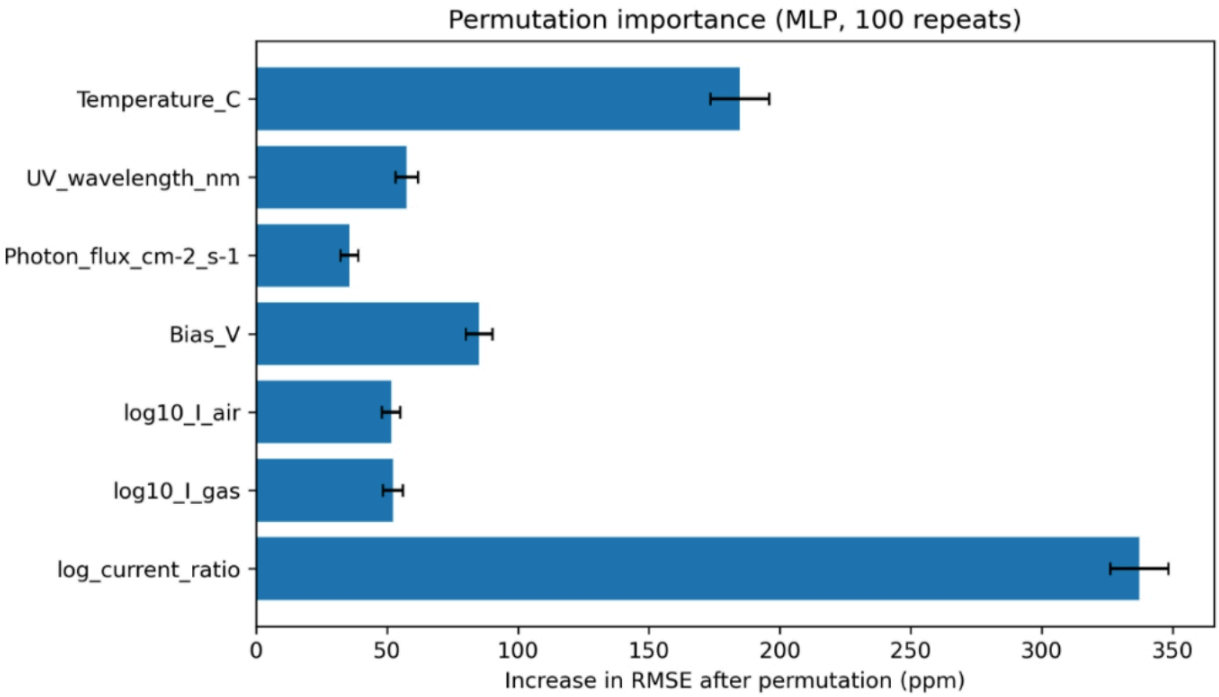

Figure 10

Holdout permutation importance for the selected MLP, expressed as the increase in RMSE over 100 permutations of each physical operating or sensor-derived feature.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- APPENDIX.docx