Opto-Electronic Advances

CN 51-1781/TN ISSN (Print) 2096-4579 ISSN (Online) 2097-3993

Rayleigh-driven ethanol cluster tracking based on non-contact deep optical molecular diagnosis

Geon Mo Kim, Yun Ji Hwang, Chengyi Li, Teajong Hwang, In-Sung Hwang, James Hone and Seong Chan Jun*

Citation: Kim GM, Hwang YJ, Li CY et al. Rayleigh-driven ethanol cluster tracking based on non-contact deep optical molecular diagnosis. *Opto-Electron Adv* **9**, 250278 (2026).

https://doi.org/10.29026/oea.2026.250278

Received: 18 October 2025; Accepted: 27 March 2026; Published online: 7 May 2026

OEA | Opto-Electronic Advances

Article

June 2026, Vol. 9, No. 6

Open Access

Rayleigh-driven ethanol cluster tracking based on non-contact deep optical molecular diagnosis

Geon Mo Kim[1†], Yun Ji Hwang[1†], Chengyi Li[1], Teajong Hwang[1], In-Sung Hwang[2], James Hone[3] and Seong Chan Jun[1*]

Abstract: Investigating the composition ratio of ethanol–water molecular clusters in air, which are responsible for anisotropic behavior, is a significant challenge, primarily owing to the difficulty of detecting Rayleigh scattering, an inherently weak signal highly susceptible to external interference. This study overcame these limitations by integrating laser diffraction focusing with artificial neural networks. We demonstrate a fully non-contact sensing system that circumvents the direct measurement of difficult-to-distinguish Rayleigh scattering signals. Our approach infers ethanol content by utilizing a self-fabricated multi-layered graphene Fresnel lens. An analysis of gaseous ethanol content in the 0.01%–0.1% range revealed that while a wavelength of 405 nm exhibited high sensitivity with up to a 7% intensity change corresponding to ethanol content, a wavelength of 638 nm provided superior stability for deep-learning analysis as its intensity parameter remained fixed. Ultimately, by combining the 638 nm laser with our self-developed self-aware assembly network model, we successfully inferred ethanol content with an R^2 of 0.884, even under varied power conditions. This study creates a new path for recognizing weak Rayleigh scattering signals that were previously difficult to measure and may facilitate the development of rapid and durable sensing systems, such as for the non-invasive diagnosis of gas molecules.

Keywords: Rayleigh scattering; ethanol-cluster; Fresnel lens; deep learning; ethanol content inference

DOI: 10.29026/oea.2026.250278 | CSTR: 32247.14.oea.2026.250278

Citation: Kim GM, Hwang YJ, Li CY et al. Rayleigh-driven ethanol cluster tracking based on non-contact deep optical molecular diagnosis. *Opto-Electron Adv* **9**, 250278 (2026).

1 Introduction

Rayleigh scattering, first described by Lord Rayleigh in 1871, predominantly occurs when light interacts with gas molecules that are much smaller than its wavelength, providing a key means for probing light–matter interactions and obtaining molecular information[1,2]. The phenomenon is based on the principle of forming a unique "optical fingerprint", in which the intrinsic average polarizability of each molecule induces isotropic scattering that determines the overall intensity, whereas its polarizability anisotropy induces anisotropic scattering that alters the polarization characteristics[1,3–5]. Furthermore, the structural asymmetry of molecular clusters, which are more likely to form at higher gas concentrations and pressures, amplifies this anisotropic scattering, making the optical fingerprint even more complex and rich[5–7].

The concentration range investigated in this study (0.01%–0.1%, ~100–1000 ppm) overlaps with ethanol vapor levels relevant to breath analysis, environmental monitoring, and industrial VOC detection[8–10] (Table S1). This interval also provides a stable measurement regime in which scattering-induced modulation remains detectable without severely degrading the focal profile. Ethanol–water vapor mixtures constitute a particularly relevant and physically demanding system for optical fingerprinting based on Rayleigh scattering. Although ethanol molecules possess a larger average polarizability and greater polarizability anisotropy due to their asymmetric molecular structure, water molecules are comparatively more isotropic[4,11]. Under such dilute

Received: 18 October 2025

Accepted: 27 March 2026

Published online: 7 May 2026

[1]School of Mechanical Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea; [2]Sentech GMI Co. Ltd, Seoul 07548, Republic of Korea; [3]Department of Mechanical Engineering, Columbia University, 500 West 120th Street, Mudd 220, New York 10027, USA.

[†]These authors contributed equally to this work.

*Correspondence: SC Jun, E-mail: scj@yonsei.ac.kr

conditions, however, the difference in scattering cross-section between ethanol and water becomes intrinsically subtle[1,4]. In addition, hydrogen-bond-mediated cluster dynamics introduce concentration-dependent and non-linear modulation of the anisotropic scattering component, further complicating the optical response[6,12]. Consequently, the modulation imposed on the transmitted beam remains extremely small and is difficult to resolve using conventional intensity-based measurements, rendering ethanol–water vapor mixtures a stringent test case for Rayleigh-scattering-based sensing strategies.

However, despite its long history and theoretical potential, the practical implementation of high-precision sensors based on Rayleigh scattering has been persistently challenged by three fundamental and interconnected problems. Practical application has been extremely difficult owing to the inherent limitations of the Rayleigh scattering signal being extremely weak[13], difficult to separate from other light sources[14], and vulnerable to external interference[15]. This study overcame these limitations by adopting an inverse approach that analyzed the main laser beam, which was minutely modulated by scattering, instead of directly measuring the scattered light. Through a custom-developed deep-learning model, we have successfully detected and quantified with high precision the minute changes in this "fingerprint," which were previously difficult to distinguish.

Precision measurement technology for ethanol content is essential in various fields, from medical diagnostics to industrial process control and environmental monitoring[16]. The primary methods for analyzing ethanol content include precision techniques such as gas chromatography (GC)[17,18] and contact-based sensors such as semiconductor[19], electrochemical[20], and biosensors[21]. Among these, GC is recognized as the gold standard for analyzing ethanol concentration owing to its high sensitivity, selectivity, and accuracy[17,18]. However, the lack of portability and long analysis times limit its application in rapid, on-site scenarios[22]. Contact-based sensors are some of the most widely researched devices owing to their convenience, cost-effectiveness, and rapid detection capabilities[17,23]. Nevertheless, because these sensors mostly rely on direct chemical reactions between the target molecules and the sensor material, they are prone to performance degradation and shortened lifespans over time owing to material deterioration[24]. Consequently, demand for real-time, high-sensitivity, non-contact ethanol inference technologies that can overcome the limitations of existing analytical methods is continuously increasing[25]. Therefore, by utilizing a non-contact optical principle, the system proposed in this paper offers advantages such as portability through a micro-lens, reduced analysis time via fast deep-learning inference, and resolutions to the problems of material degradation and limited lifespan of conventional contact-based sensors.

The introduction of two-dimensional (2D) materials has resulted in advancements in high-functionality material systems based on ultrathin structures to overcome the physical limitations of conventional three-dimensional (3D) materials[26–30]. Since the discovery of single-layer graphene, it has established itself as a revolutionary material in the field of optics[31–34]. With its outstanding optoelectronic properties, including high optical transparency, low reflectance, and high carrier mobility at room temperature, graphene is considered an ideal material for optoelectronic devices[35]. These characteristics make graphene a highly promising candidate for ultrathin optical components such as Fresnel lenses[33,36–41].

In recent decades, coupled with revolutionary advancements in computing power, machine learning and deep learning algorithms have been actively applied to the fields of optics and photonics. They have contributed to the enhanced precision of non-contact sensing technologies[42,43]. In particular, research on alcohol detection using artificial intelligence complements existing chemical sensor-based approaches, opening up possibilities for more sophisticated and non-invasive measurements[44,45]. Machine learning applications in optical sensing have traditionally focused on two main tasks: qualitative state identification and quantitative regression. For state identification, algorithms have been successfully deployed to compensate for optical aberrations and restore signals[46,47]. Conversely, single-stage regression models[48–51] have been widely adopted to directly predict quantitative metrics like chemical concentrations. A fundamental limitation of these conventional single-stage models, however, is their vulnerability to complex physical fluctuations (e.g., varying laser power). This often leads to severe performance degradation and significant uncertainties over broad dynamic ranges[52]. To systematically decouple such non-linearities, recent studies have increasingly adopted cascaded, hierarchical, or multi-stage architectures[53]. Motivated by these multi-stage approaches, we propose a two-stage, self-aware assembly network (SAAN) architecture tailored for optical ethanol content measurement. Unlike single unified networks, which are prone to severe errors over broad dynamic ranges, our model first operates as a macro-classifier to determine the specific concentration regime. The optical data is then routed to one of three regime-specific regression models. This hierarchical design allows each regression model to be optimized solely for a narrow concentration sub-range, effectively mitigating the global non-linearities of the optical signal and achieving high quantitative precision.

This paper represents the first experimental implementation of a precise, non-invasive, and fully non-contact technique for ethanol-content detection and inference by combining a graphene-based optical device with artificial intelligence (AI)-based data analysis. Thus, it presents new technological possibilities and application directions in the field of optics-based bio-analysis. Therefore, this technology is expected to be applied in the future to high-precision sensing systems, such as non-invasive real-time blood

alcohol content (BAC) estimation for sobriety testing, breath gas analysis for disease diagnosis, and hazardous gas monitoring in industrial settings.

2 Results and discussion

2.1 Mechanism of the optical ethanol-cluster content inference system

This paper proposes a fully non-contact, optics-based analysis system capable of precisely detecting ethanol content within the range of 0.01%–0.1%. The overall architecture of the system is illustrated in Fig. 1(a), while the underlying physical inference mechanism is schematically summarized in Fig. 1(b). The proposed platform operates based on the principle of Fresnel diffraction using a Fresnel lens fabricated from multilayer graphene. The core mechanism relies on detecting variations in the focal point formed when a visible-wavelength laser beam—whose angular spectrum has been perturbed by Rayleigh scattering through interaction with airborne ethanol molecules and clusters—is diffracted by the lens. The intensity and spatial distribution extracted from the captured micro-focus constitute a molecule-dependent optical fingerprint and are used as input features for a deep-learning model, which analyzes these features to

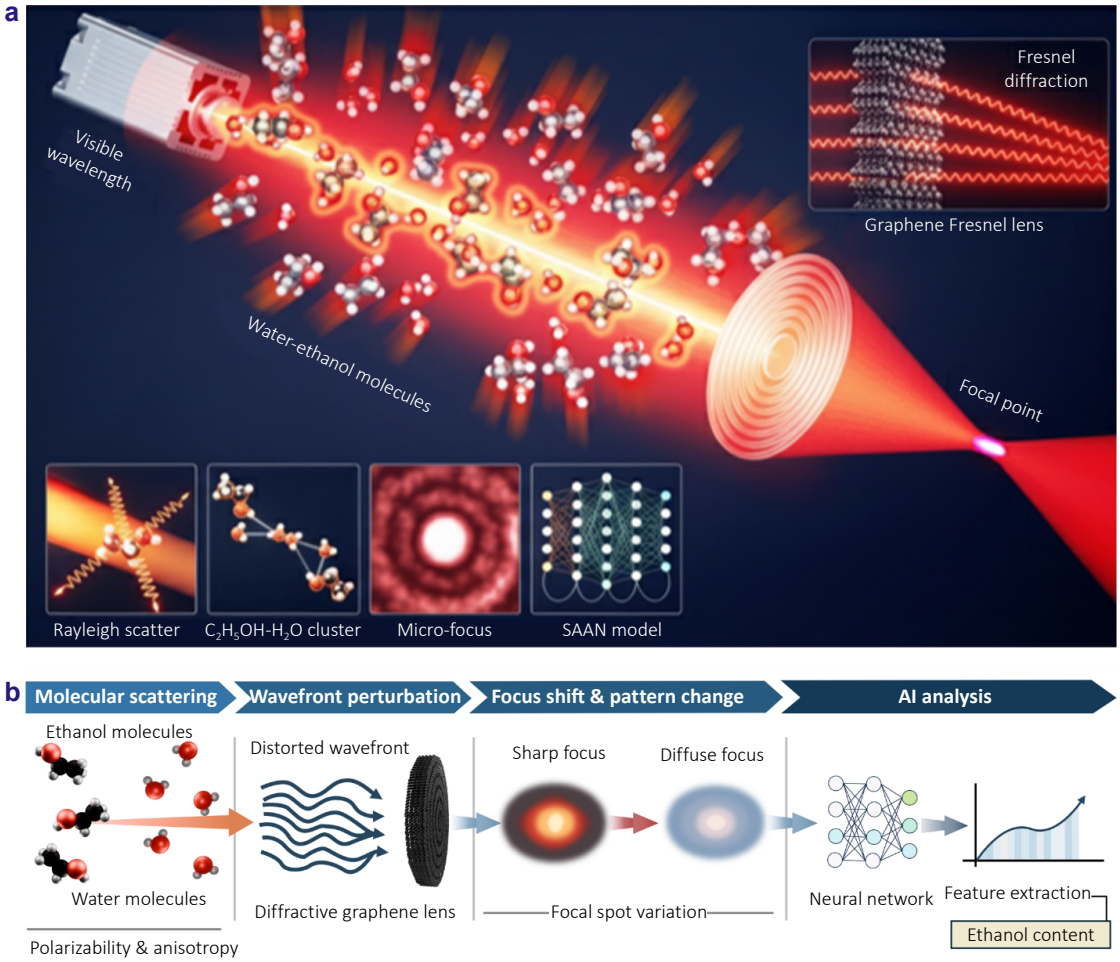

Fig. 1 | Rayleigh scattering-based non-contact ethanol-cluster content analysis system using a multi-layer graphene Fresnel lens and deep learning. (a) Conceptual illustration of the optical sensing system. The non-contact optical method for high-precision detection of gas molecular content is achieved by leveraging the unique optical properties of molecules. The principle is based on the distinct polarizability and concentration of gas molecules, which modulate the intensity of Rayleigh scattering when interrogated by a laser beam. The unscattered transmitted portion of the beam is focused using a diffraction-based graphene Fresnel lens. Consequently, variations in gas molecular content induce minute yet detectable alterations in the optical characteristics—such as intensity and geometry—of the resulting focal point. These subtle focal alterations are then analyzed using a deep-learning model to quantify the gas molecular content. (b) Stepwise physical mechanism of optical fingerprint formation and inference. Molecular polarizability and anisotropy determine molecule-specific Rayleigh scattering behavior, which modifies the angular spectrum of the transmitted beam. The resulting wavefront perturbation changes the focal-spot geometry (e.g., intensity profile and FWHM) after diffractive focusing. These concentration-dependent focal distributions constitute an optical fingerprint that is decoded by a neural network to infer ethanol content. This schematic visualizes how molecular-scale scattering properties are transduced into measurable macroscopic optical features suitable for learning-based analysis.

quantitatively classify and infer ethanol content. This framework establishes a direct physical pathway linking molecular scattering properties to measurable optical observables suitable for learning-based analysis.

2.2 Principle of ethanol molecular detection using Rayleigh scattering

The thickness of single-layer graphene is approximately 0.2–0.35 nm, which is theoretically equivalent to the size of a single carbon atom. The optical transmittance of such a single layer in the visible spectrum is about 97.7%, as light passes through the thin carbon atomic layer with minimal obstruction. However, when graphene is stacked with fewer than five layers for use in an optical device such as a graphene Fresnel lens, limitations occur, such as reduced focusing efficiency, an elongated focal length, and an increase in unintended transmission. Conversely, when the number of layers exceeds five, variations in the full width at half maximum (FWHM), focal length, and intensity become relatively stable, but the transmittance is reduced by more than 20%[37]. Therefore, in this study, we overcame these limitations by employing a five-layer stacked graphene structure as shown in Fig. 2(a). Specifically, five-layer graphene maintains an optical transmittance of approximately 80%–82% while achieving focusing efficiencies exceeding 60%, whereas four layers exhibit noticeably weaker focal contrast and six layers introduce more than 20% transmission loss[37,39] (Table S2). As shown in Fig. 2(b), the five-layer graphene Fresnel lens is composed of multiple concentric rings and forms a single focal point based on the principle of diffraction from each zone slit. The resulting focus is characterized by its Rayleigh range, which defines the axial distance from the beam waist (the focal point) to the point at which the beam's cross-sectional area doubles. In this study, this property was utilized to detect ethanol content.

Figure 2(c) shows the overall process wherein a laser beam, after passing through gaseous ethanol and water molecules in the air, is focused to a single point by the graphene Fresnel lens. As the laser beam encounters ethanol molecules in the air, optical scattering occurs. Consequently, the total amount of light incident on the graphene lens is reduced, which in turn alters the optical properties of the focal point. A schematic illustration of this mechanism is provided in Fig. 1(b) to clarify the stepwise physical pathway linking molecular scattering properties to wavefront perturbation, focal-spot modulation, and subsequent inference. Here, the optical fingerprint is defined as the molecule-dependent spatial intensity distribution of the focused beam that arises from scattering-induced perturbations of the incident wavefront. This distribution originates from the combined effects of molecular polarizability, anisotropy, and concentration, which jointly determine the magnitude and angular distribution of Rayleigh scattering and thereby modify the effective wavefront reaching the diffractive lens. Because diffractive focusing relies on phase-coherent interference, even small scattering-induced angular or wavefront distortions upstream of the lens manifest as measurable focal broadening and profile deformation downstream (Fig. S1). These concentration-dependent focal variations constitute the observable features used for ethanol detection.

Rayleigh scattering is a phenomenon that occurs when particles are smaller than the wavelength of light[1–5]. This scattering is categorized into isotropic Rayleigh scattering, which occurs uniformly in all directions, and anisotropic Rayleigh scattering, which varies with direction, as depicted in Fig. 2(d). In the gaseous state, isotropic scattering is determined by the average polarizability of the randomly rotating molecules. Anisotropic scattering results from the anisotropy of polarizability, which becomes more significant for asymmetric molecules or clusters with complex structures.

The total Rayleigh scattered light is determined by the sum of the isotropic and anisotropic scattered light components. Furthermore, the ethanol content has a decisive impact on the total scattered light. Ethanol molecules have a higher average polarizability than water molecules owing to their larger size and greater number of electrons. Additionally, ethanol (C_2H_5OH) has a larger and more complex molecule than water (H_2O), resulting in a much greater anisotropy of polarizability. Therefore, as the ethanol content increases, the total Rayleigh scattered light intensifies, resulting in a corresponding increase in transmitted laser loss. This results in a strongly attenuated light, as shown in Fig. 2(e). In this study, we focused the attenuated light with a graphene Fresnel lens to form a focal point and sensitively detected its properties to track ethanol content.

2.3 Design of a graphene Fresnel lens and ethanol sensing performance

Graphene Fresnel lenses are designed according to the specific wavelength of use, and the lens zones are derived from the Fresnel zone plate radius r_m formula Eq. (1), which maximizes constructive interference.

$$r_m \approx \sqrt{m\lambda f}\,(\Delta\psi = 0)\,, \tag{1}$$

where m is the number of rings, λ is the wavelength, f is the focal length, $\Delta\psi$ is the phase difference, and x_m is the path length from the m-th ring to the focal point (Supplementary Note S1)[37,39]. The graphene Fresnel zone plate created under these principles maximizes constructive interference. Consequently, the spacing between the zones decreases as the distance from the lens center increases. Additionally, an increase in wavelength leads to wider spacing between the zone slits, consequently resulting in a larger lens size. In this study, the lenses were designed for wavelengths of 638, 512, and 405 nm, as shown in Fig. 3(a). We fabricated Fresnel lenses composed of five layers of graphene according to the

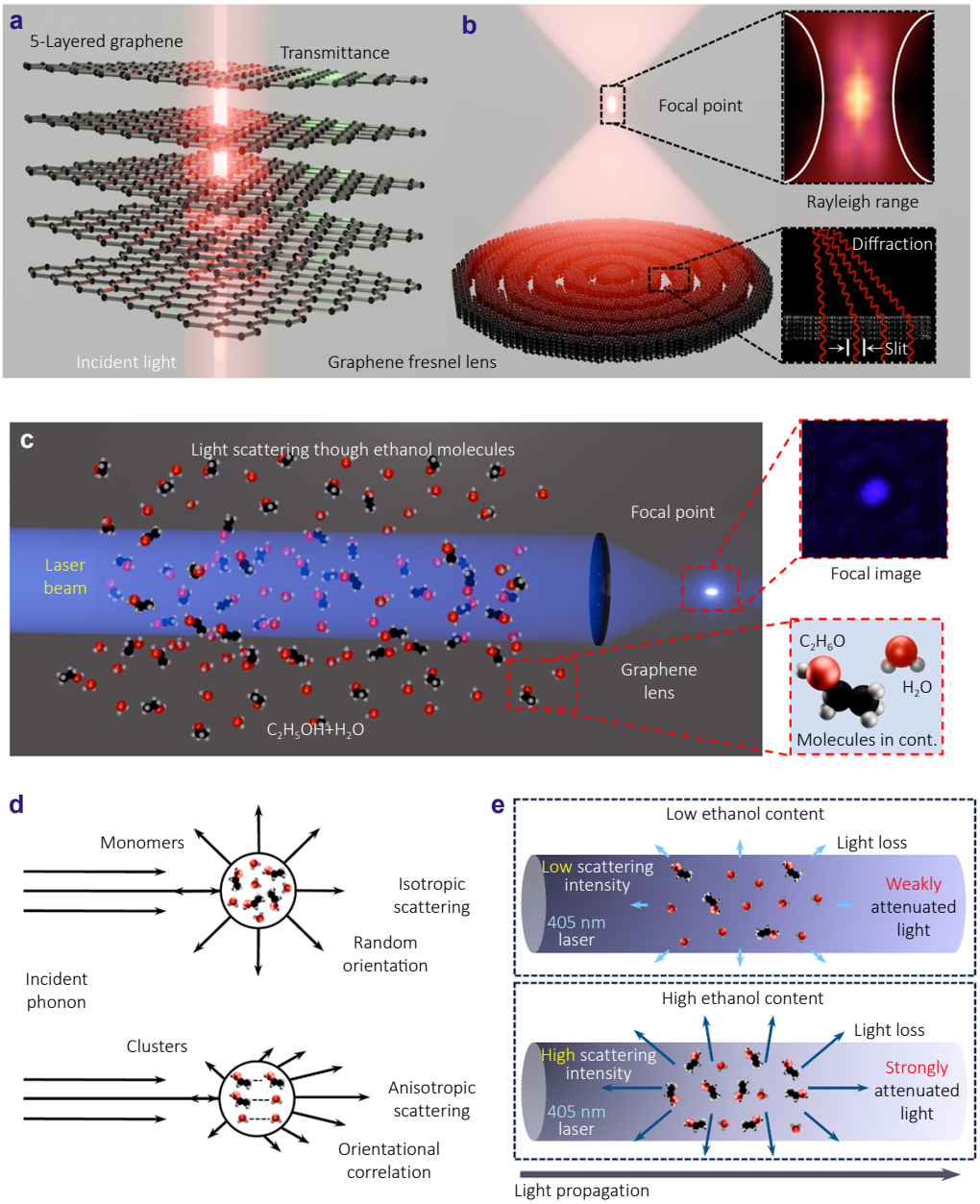

Fig. 2 | Principle of ethanol-content analysis using a multi-layer graphene lens and Rayleigh scattering. (a) Structure and optical characteristics of the five-layer graphene Fresnel lens. (b) Principle of laser focusing via Fresnel lens diffraction and the structure of the resulting single focal point. (c) Schematic of the laser beam focusing process after Rayleigh scattering and optical loss induced by ethanol and water vapor molecules. (d) Conceptual diagram of isotropic and anisotropic Rayleigh scattering depending on the molecular state (monomer vs. cluster). (e) Visualization of the variation in isotropic scattering intensity and optical loss with changing ethanol content under a constant total molecular density.

specified design (Fig. S2). f was uniformly designed to be 220 μm for all three lenses, with m set to 24 (Fig. S3 and Table S3).

Furthermore, as shown in Fig. 3(b), an analysis of the optical focusing efficiency of the lens designed in this study confirmed a trend of increasing efficiency with higher laser source power. Figure 3(c) shows that as the ethanol content increased from 0.01% to 0.1%, scattering increased, resulting in an increase in the normalized scattered light intensity. Accordingly, the transmitted light intensity decreased, with the degree of reduction being greater at a wavelength of 405 nm than at 638 nm, according to the Rayleigh scattering formula Eq. (2), where I represents the intensity of the scattered light, and λ is the wavelength.

$$I \propto \frac{1}{\lambda^4} . \tag{2}$$

In this study, we measured the change in the size of the

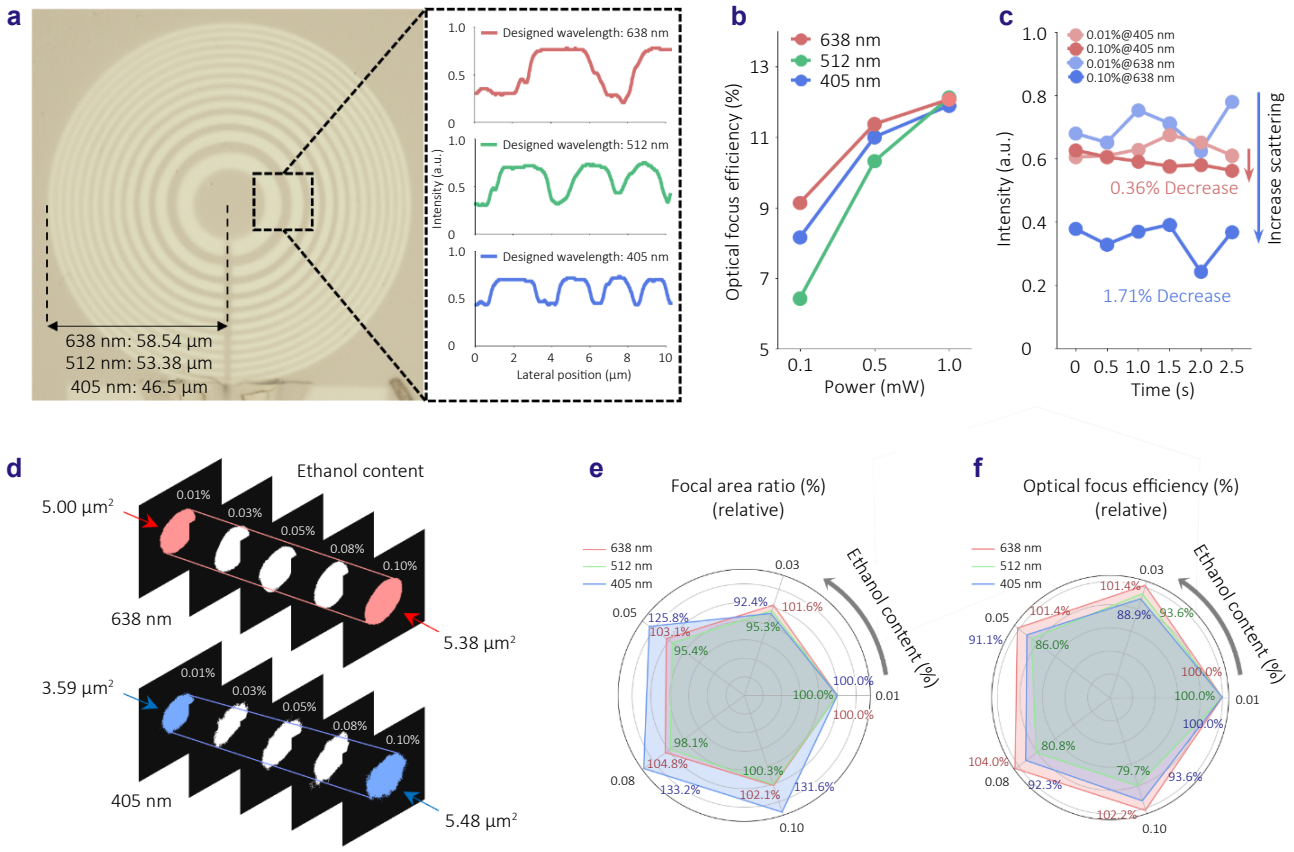

Fig. 3 | Analysis of the optical characteristics of a graphene Fresnel lens and its performance in ethanol-content sensing. (a) Optical microscope (OM) image of the graphene Fresnel lens with a concentric ring structure and transmittance performance of lenses according to each design wavelength. (b) Variation in optical focusing efficiency depending on the laser wavelength and power of the incident laser. (c) Comparison of the temporal responses of normalized scattering intensity during an ethanol-content change (from 0.01% to 0.1%) at wavelengths of 638 and 405 nm. (d) Comparison of the focal area with increasing ethanol content at wavelengths of 638 and 405 nm. (e) Relative focal spot size as a function of wavelength and ethanol content. (f) Relative focusing efficiency as a function of wavelength and ethanol content.

focal spot corresponding to the increase in ethanol content and compared it by laser wavelength. As shown in Fig. 3(d), when the ethanol content increased from 0.01% to 0.1%, the focal area at 638 nm increased by approximately 0.38 μm², from 5 to 5.38 μm². In contrast, at 405 nm, the focal area increased by approximately 1.89 μm², from 3.59 to 5.48 μm². This indicates that the beam spread at the focus increased owing to the intensification of Rayleigh scattered light at shorter wavelengths.

Figure 3(e) and 3(f) show the comparison of the relative rates of change for the focal spot size and focusing efficiency by wavelength as a function of ethanol content. The relative rate of change was calculated as a percentage by dividing the value at each concentration by the value at the lowest ethanol content and multiplying by 100. The results showed that, for most concentrations, the relative focal area ratio was greatest at 405 nm. Conversely, a higher relative optical focusing efficiency was observed at the 638 nm wavelength. These results imply that the increase in light loss owing to intensified Rayleigh scattering at shorter wavelengths reduces the amount of light that converges to a single focal point.

2.4 Parameter analysis of ethanol sensing performances

In this study, focal spot images, focused using a graphene Fresnel lens, were captured via a complementary metal–oxide–semiconductor (CMOS) camera (Fig. S4). As shown in Fig. 4(a), the three-channel (RGB) data from the captured focal images were compressed into one-channel (grayscale) data, thereby removing color information and retaining only luminance (brightness) information (Note S2). Figure 3(c) confirms that differences in scattered light due to wavelength and ethanol content cause variations in light intensity. Therefore, the removal of channel (RGB) data offers a significant advantage in terms of data processing while maintaining sufficient discriminatory power for detecting wavelength and ethanol content. Figure 4(b) shows the results of converting the image data into 1D light

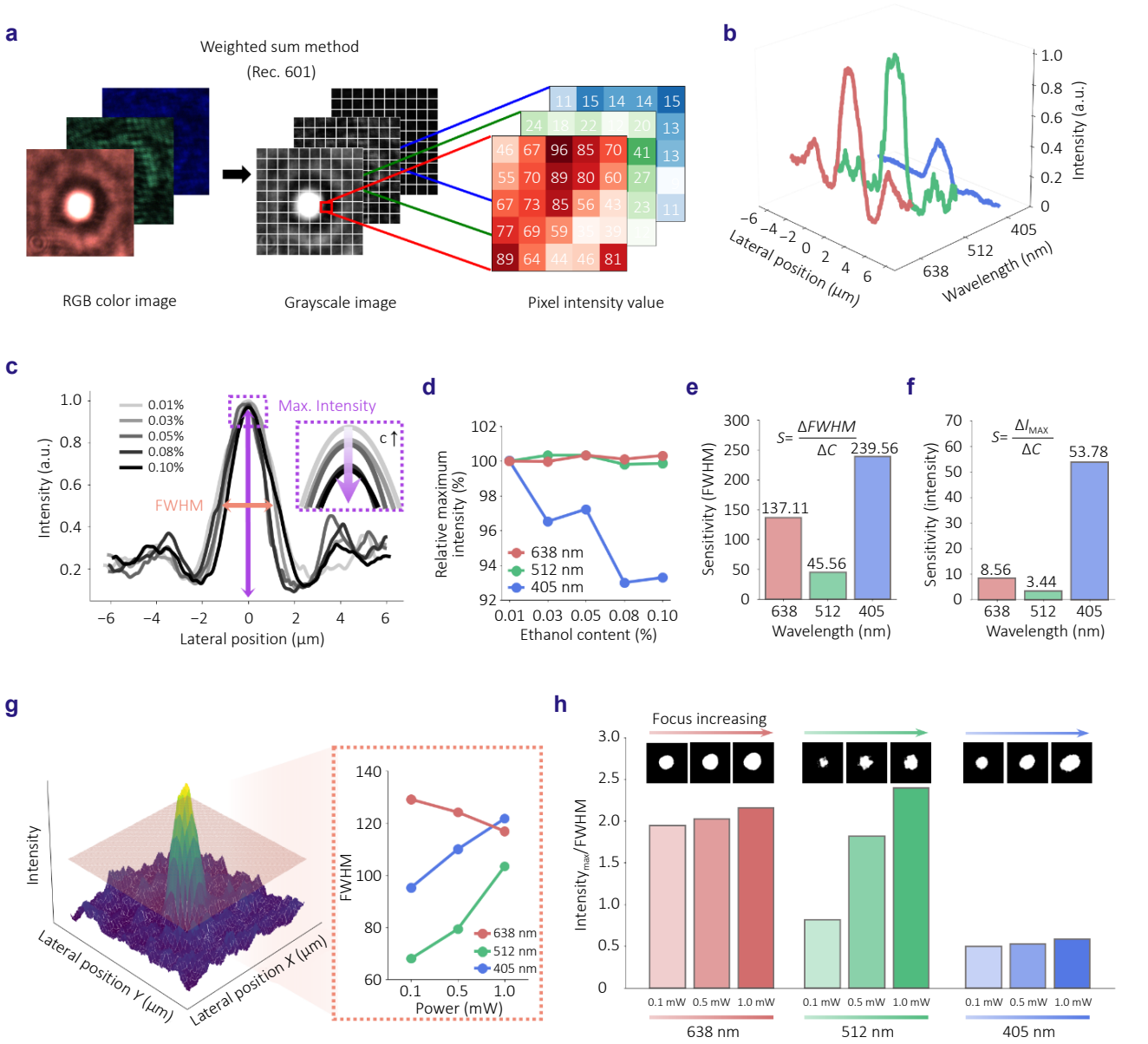

Fig. 4 | Quantification of ethanol-content sensing performance via focal profile analysis. (a) Schematic of the process for extracting a 1D intensity profile from a 2D focal image. (b) Comparison of 1D intensity profiles as a function of wavelength. (c) Variation in maximum intensity and FWHM with ethanol content. (d) Variation in relative maximum intensity as a function of wavelength and ethanol content. (e) Wavelength-dependent comparison of the ethanol-content sensing sensitivity using FWHM as a metric. (f) Wavelength-dependent comparison of the ethanol-content sensing sensitivity using maximum intensity as a metric. (g) 3D intensity distribution plot of the focal spot and a comparison of FWHM values at different wavelengths. (h) Analysis of the maximum intensity/FWHM ratio as a function of incident laser wavelength and power.

intensity data, confirming that different light intensity values and graph shapes appeared for each wavelength.

The results of comparing 1D light intensity data as a function of ethanol content are shown in Fig. 4(c). The data showed variations in the FWHM according to ethanol content, and the maximum intensity tended to decrease as the ethanol content increased.

$$\text{Relative } I_{\text{MAX}} = \frac{I_{\text{MAX}}}{(I_{\text{MAX}})_0} * 100 \ . \tag{3}$$

Figure 4(d) shows that the rate of change in maximum intensity at the 405 nm wavelength was significantly higher than at other wavelengths with varying ethanol content. This rate, calculated as shown in Eq. (3) represents the relative reduction calculated by comparing the maximum intensity at each concentration I_{MAX} to the baseline value at the lowest concentration of 0.01% $(I_{\text{MAX}})_0$. In contrast to the 638 nm wavelength, where the relative rate of change in maximum light intensity was negligible, the 405 nm

wavelength exhibited a significant change, reaching a maximum of over 9%. Similarly, we examined the sensitivity, which enabled the confirmation of the change in the FWHM and maximum intensity between the minimum and maximum ethanol-content values.

$$Sensitivity\ of\ FWHM = \frac{\Delta FWHM}{\Delta C}\ , \tag{4}$$

$$Sensitivity\ of\ Intensity = \frac{\Delta I_{MAX}}{\Delta C}\ . \tag{5}$$

A quantitative comparison of the sensitivity of the FWHM and maximum intensity, calculated according to Eqs. (4) and (5), revealed the highest sensitivity at the shortest wavelength of 405 nm, as shown in Fig. 4(e) and 4(f). The sensitivity was calculated by dividing the change in measured values ($\Delta FWHM$, ΔI_{MAX}) between the highest (0.1%) and lowest (0.01%) concentrations by the change in concentration (ΔC). A comparison revealed that the sensitivities for FWHM and maximum intensity at 405 nm were approximately 2 and 6.8 times greater, respectively, than those at the 638 nm wavelength.

In this study, we confirmed that the amount of light reaching the lens varies owing to Rayleigh scattering by ethanol molecules. Similarly, we investigated the reactivity of the lens focus by varying the power of the input laser source. As shown in Fig. 4(g), as the laser source power was increased, the amount of light reaching the graphene Fresnel lens increased, thereby increasing the size of the focal spot. This resulted in a corresponding change in the FWHM. Additionally, the maximum intensity/FWHM ratio, an indicator of focal sharpness, exhibited the most consistent and least variable trend at 405 nm, but it also exhibited the lowest value (Fig. 4(h)). In contrast, the 638 nm wavelength exhibited a relatively consistent trend while simultaneously maintaining a high level of focal sharpness. Regarding this, Fig. 4(h) shows that the FWHM for a given laser source power increased linearly at 512 and 405 nm, whereas it decreased linearly at 638 nm (Fig. S5).

2.5 Analysis of the trade-off between sensitivity and stability of optical sensing

In this study, we performed ethanol-content prediction and classification using the optical features of the focal spot, as shown in Fig. 5(a). Figure 5(b) presents a comparison of the changes in focal spot geometry at two wavelengths, 638 and 405 nm, as a function of ethanol content (Fig. S6). We confirmed that as the ethanol content increased, the instability of the focal point was greater at 405 nm than at 638 nm. Furthermore, Fig. 5(c) shows the aspect ratio as a function of ethanol content for each wavelength. The average error in the aspect ratio was found to be 2.6% at 638 nm, 6.6% at 512 nm, and 6.4% at 405 nm, confirming that the most stable values were obtained at 638 nm. This indicates that the shape is more circular at 638 nm. As shown in Fig. 5(d), we measured the temporal variation in the aspect ratio of the focal spot at ethanol contents of 0.01% and 0.1% for the 638 and 405 nm wavelengths. These results, measured at 0.5 s intervals, show that the aspect ratio remained relatively constant without significant deviation. However, the average error of the aspect ratio was 3.0% at 638 nm and 6.1% at 405 nm, confirming that the temporal stability of the aspect ratio was also greater at 638 nm. At 638 nm, the shape remained circular over time. Figure 3(d) confirms that an increase in ethanol content results in an increase in Rayleigh scattered light, which in turn causes a beam spread at the focus. This focal broadening arises because Rayleigh scattering introduces small angular deviations in the propagating wavefront. As scattering increases, the incident beam deviates from an ideal plane wave and acquires a finite angular spread before reaching the lens. Since diffractive focusing relies on phase-coherent interference across the aperture, such angular perturbations partially disrupt constructive interference at the focal point. This results in a broadened focal intensity distribution, observed experimentally as an increase in FWHM (Fig. S7). Figure 5(e) shows the results of measuring the normalized focal area over time at ethanol contents of 0.01% and 0.1% at 638 and 405 nm. As the ethanol content increased, the normalized focal area increased at both wavelengths, with the rate of increase being 2.04% at 638 nm and 31.56% at 405 nm, a difference of approximately 15.47 times. To exclude the possibility that the observed focal broadening originates from photothermal expansion of the graphene lens rather than scattering-induced wavefront modulation, we conducted an independent thermal stability test under continuous 405 nm irradiation. Although slow focal drift was observed under prolonged continuous exposure (~13 h), short acquisition windows (~3 s) corresponding to the actual sensing protocol exhibited negligible variation (<0.42 μm^2), confirming that the concentration-dependent FWHM increase arises from Rayleigh-scattering-induced angular redistribution rather than thermal effects (Fig. S8).

Figure 5(f) compares the 1D light intensity graphs for each ethanol content at 638 and 405 nm. As the ethanol content increased, the changes in the maximum intensity and FWHM were greater at 405 nm than at 638 nm. Figure 5(g) shows the variation and dispersion of the maximum intensity as a function of wavelength and ethanol content. The maximum intensity decreased with increasing ethanol content owing to the increase in Rayleigh scattered light, and its variance remained nearly constant at 638 nm. Figure 5(h) shows the change in the FWHM and relative maximum intensity/FWHM ratio. The magnitude of the change in FWHM was highest at 405 nm, the wavelength with the greatest reactivity owing to Rayleigh scattering, but the sharpness of the focus was simultaneously the lowest. In contrast, the 638 nm wavelength exhibited distinct changes in the FWHM by concentration while maintaining a higher focal sharpness compared to the other wavelengths.

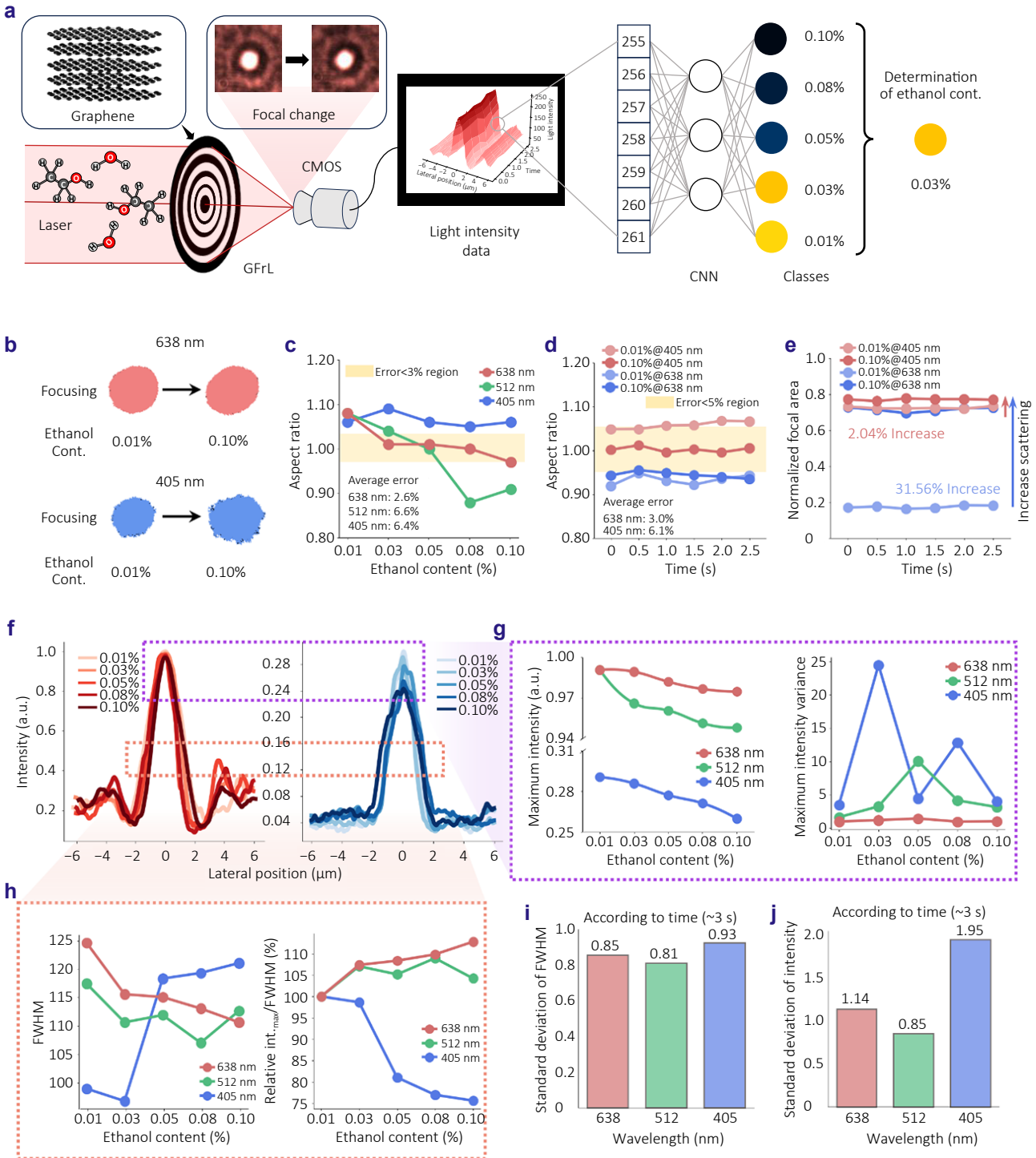

Fig. 5 | Analysis of the key parameters of the ethanol content inference. (a) Schematic of the ethanol-content prediction and classification process using optical features of the focal spot. (b) Visualization of the focal spot geometry variation with ethanol content (0.01% vs. 0.1%). (c) Variation in the aspect ratio of the focal spot as a function of wavelength and ethanol content. (d) Comparison of the temporal response of the aspect ratio during an ethanol-content change (from 0.01% to 0.1%) at 638 nm vs. 405 nm. (e) Comparison of the temporal variation in the normalized focal area during an ethanol-content change at 638 nm vs. 405 nm. (f) Comparison of 1D intensity profiles as a function of ethanol content at 638 and 405 nm. (g) Comparison of maximum intensity and its standard deviation as a function of wavelength and ethanol content. (h) Comparison of the FWHM and maximum intensity/FWHM ratio as a function of wavelength and ethanol content. (i) Comparison of the standard deviation of FWHM by wavelength. (j) Comparison of the standard deviation of maximum intensity by wavelength.

Figure 5(i) and 5(j) show the standard deviations of the FWHM and maximum intensity. Both were higher at the 405 nm wavelength over continuous time than at other wavelengths. This indicates that the 405 nm wavelength is disadvantageous in terms of data stability owing to its high reactivity with surrounding molecules.

In conclusion, the largest concentration-dependent variations were observed at 405 nm owing to the stronger Rayleigh scattering cross-section at shorter wavelengths. Since the Rayleigh scattering cross-section scales approximately as λ^{-4}, the scattering-induced attenuation at 405 nm is significantly greater than at 638 nm. While this enhances sensitivity, it also amplifies temporal fluctuations and instability in both peak intensity and focal geometry. In contrast, at 638 nm, the scattering-induced attenuation is substantially smaller (Fig. S9). Consequently, the peak intensity remains within the quasi-linear dynamic range of the CMOS detector and does not exhibit large concentration-dependent variation. Importantly, this behavior does not arise from saturation of the scattering process, but rather from operation within a stable detector regime. Under these conditions, concentration-dependent information is primarily encoded in the spatial broadening of the focal spot rather than intensity attenuation, making FWHM the dominant discriminative parameter. This separation —stable peak intensity combined with distinct FWHM modulation —provides a favorable condition for deep-learning-based inference. These results indicate that shorter wavelengths improve sensitivity but also increase temporal fluctuation in the focal features, whereas 638 nm provides more stable profiles for learning-based inference.

2.6 Optimization of a deep-learning-based ethanol-content classification model

In this study, a 1D CNN model was designed, trained, and tested to classify five levels of ethanol content. Figure 6(a) shows the structure of the six-channel light intensity profile of the focal spot (Figs. S10–S12), which is input into the designed 1D CNN model architecture shown in Fig. 6(b). The model architecture consists of three 1D CNN layers and two fully connected layers. It extracts features from the input data during training and subsequently produces results classifying the five ethanol-content classes.

First, we conducted experiments by randomly mixing three laser power condition outputs: 0.1, 0.5, and 1 mW. The receiver operating characteristic (ROC) curve resulting from the classification prediction performed by the 1D CNN model is shown in Fig. 6(c). The AUC calculated from the obtained ROC curve and the mean absolute error (MAE) results, which measured the error between the model's predicted values and the actual ground truth values, are shown in Fig. 6(d). The mean AUC values for all three wavelengths were 0.987 at 638 nm, 0.984 at 512 nm, and 0.974 at 405 nm, recording the highest value at 638 nm and the lowest at 405 nm. The MAE values also confirmed that the error was lowest at 638 nm and highest at 405 nm. This implies that while the sensitivity from a sensor perspective was high at 405 nm, as shown in Fig. 6(e), the sensitivity from a deep-learning model was higher at 638 nm. Therefore, in this study, 638 nm was strategically selected as the primary operating wavelength because it provides a more stable optical response for feature extraction, enabling more robust and reliable deep-learning inference. This choice reflects a deliberate design trade-off favoring model stability over raw scattering sensitivity.

Furthermore, when the AUC values were analyzed by power level at the 638 nm wavelength, high values exceeding 0.988 were achieved at all three power levels (Fig. 6(f)). The confusion matrix, which shows the ethanol-content classification prediction results for the randomly mixed-power data at the 638 nm wavelength, is presented in Fig. 6(g) (Fig. S13). The 1D CNN-based model achieved an accuracy of 93.33%. Notably, this performance was obtained under mixed-power conditions, indicating robust classification capability under realistic signal variability rather than idealized measurement conditions. This level of accuracy is comparable to, or exceeds, that reported for existing machine-learning-based gas sensing systems evaluated under practical operating conditions[54–56] (Table S4). Furthermore, when the prediction accuracies for each power level were compared, consistently higher and more stable performance was observed at 1 mW (Fig. 6(h), Fig. S14, and Table S5).

Therefore, this study also conducted experiments under the single-laser power condition, and the ROC curve for the prediction results is shown in Fig. 6(i). This experiment, conditioned only on 1 mW, was also conducted for the five ethanol-content levels from 0.01% to 0.1% at the three wavelengths: 638, 512, and 405 nm. The confusion matrices and AUC values corresponding to the ethanol-content classification prediction results at each wavelength are shown in Fig. 6(j) (Fig. S15 and Table S6). Accuracies of 100% at 638 nm, 100% at 512 nm, and 88% at 405 nm were obtained. While AUC values close to 1 were achieved at both 638 and 512 nm, a lower AUC value was achieved at 405 nm.

Based on these results, this study compared the training times required to derive each result (Fig. 6(k) and Table S7). At 638 nm, because the light intensity parameter at the focal center was fixed, the training time did not significantly differ from that of the dataset composed of single-laser power data, despite the higher complexity of the randomly mixed-laser power dataset, and the accuracy also remained high. However, as the wavelength shortened, the accuracy for the mixed dataset decreased, and notably, the training time for the 405 nm wavelength increased by more than 2.5 times.

Regarding this, we compared the probability values of being correct for each ethanol content at 638 and 405 nm (Fig. 6(l)) and calculated the standard deviation. The mean standard deviation was 0.2312 at 638 nm and 0.3454 at 405

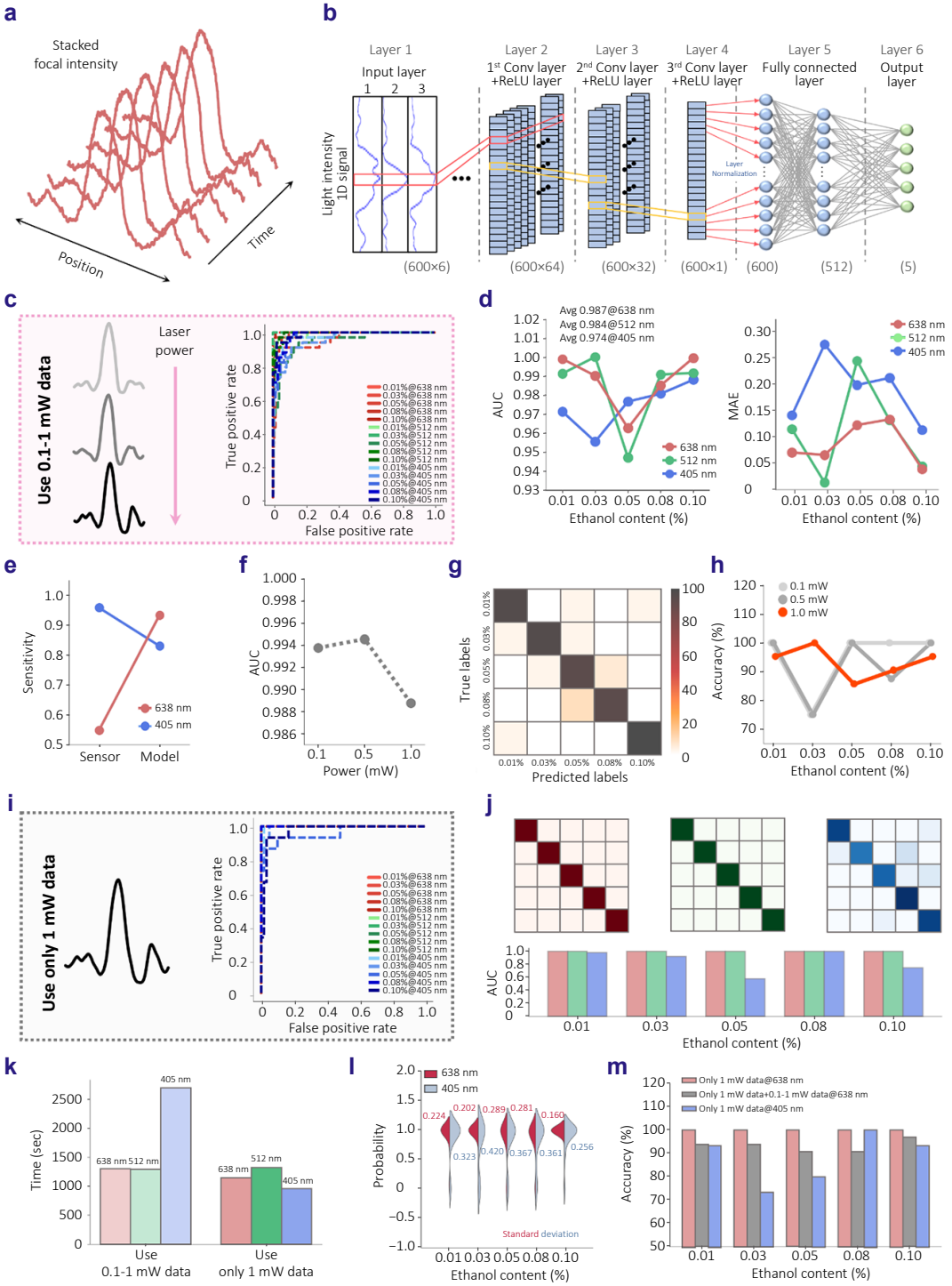

Fig. 6 | Performance and feature parameter analysis of the 1D CNN classification model. (a) Structure of the six-channel 1D intensity profile used as input data for the deep-learning model. (b) Architecture of the 1D CNN model for classifying five levels of ethanol content (0.01%–0.1%). (c) Receiver operating characteristic (ROC) curve of the classification model for the mixed-laser power (0.1–1 mW) dataset. (d) Comparison of area under the curve (AUC) and mean absolute error (MAE) values as a function of ethanol content at three wavelengths. (e) Comparison of the system's ethanol-content sensing sensitivity at wavelengths of 638 and 405 nm. (f) Comparative analysis of the model's AUC values according to incident laser power. (g) Confusion matrix for the ethanol-content classification results using the mixed-laser power dataset. (h) Classification accuracy as a function of incident laser power and ethanol content. (i) ROC curve of the classification model under the single-power condition. (j) Comparison of confusion matrices by wavelength and AUC values by ethanol content. (k) Wavelength-dependent comparison of model training times using the mixed-laser power dataset versus the single-laser power dataset. (l) Comparison of prediction probability and its standard deviation at wavelengths of 638 and 405 nm. (m) Comparison of classification accuracy by concentration under various training dataset conditions: 638 nm single-laser power, 638 nm mixed-laser power, and 405 nm single-laser power.

nm, indicating that the model had more difficulty correctly classifying the data from the 405 nm wavelength. In conclusion, because the data from the 405 nm wavelength was not fixed and varied sensitively, the results were not as high as those at 638 nm, even when using single-laser power. In contrast, at the 638 nm wavelength, owing to the fixation of the light intensity parameter, high accuracies of over 90% were recorded for the 0.01%–0.1% ethanol content when using both single-laser power and randomly mixed three laser power conditions of 0.1, 0.5, and 1 mW (Fig. 6(m)).

2.7 Inference of the ethanol content and evaluation of the SAAN model

For sensing systems to be implemented effectively, the mere detection of differences in ethanol content in the water should be transcended, and the inference of precise content values should be targeted. Consequently, we involved the formulation of a self-aware assembly network (SAAN) model, designed to accurately ascertain minimal ethanol-content values from 638 nm laser measurements (Fig. 7(a)). This model sequentially arranges classification and regression models to determine the level of ethanol content.

Stacked 2D focal data are then entered into the initial guide model. The guide model, depicted in Fig. 7(b), categorizes input data into low, medium, or high levels based on ethanol content. The classification results of the guide model are displayed in the confusion matrix in Fig. 7(c); a classification accuracy of 95.45% was achieved for the three levels. Depending on the level determined by the guide model, the model was automatically matched with one of the low-, medium-, or high-content models. The original 2D data were subsequently entered into the matched inference.

The ethanol-content inference models consisted of a combination of 1D convolution and linear operations, as shown in Fig. 7(d). The low model predicted below 0.03%, the medium model predicted concentrations between 0.03% and 0.08%, and the high model predicts concentrations of 0.08% or higher. Each model was trained individually, and the individual regression results of the models are shown in Fig. 7(e). The inference results of the entire SAAN model, which combined the guide and ethanol-content inference models, are shown in Fig. 7(f). The inferred values were compared with the actual values under each condition, and the overall R^2 value was 0.884. The MAE according to the ethanol-content conditions was 0.00518, thereby confirming the efficacy of the model in performing regression on the ethanol-content values at a high level (Fig. 7(g)). All component models of the SAAN model were reviewed during training to ensure no overfitting occurred, based on loss calculations for the training batch and performance metrics including classification accuracy exceeding 95% on the validation set and nMAE below 0.2%. (Table S8, Fig. S16)

The SAAN model was subjected to training and testing across a range of laser powers, specifically 0.1, 0.5, and 1 mW (Fig. 7(h), and Fig. S17). The mean value of the AUC calculated from the ROC curve was 0.969 for the original model, which was higher than the 0.950 for the model trained using only data from the 1 mW laser power condition. Furthermore, as shown in Fig. 7(i), when the prediction distributions for each ethanol-content condition in the test data were compared, the original model exhibited an average standard deviation of 0.00940, which was considerably smaller than the 0.0155 of the model trained using single-laser power data. These results suggest that the training diversity was enhanced, resulting in improved training performance, and the model demonstrated more stable and higher performance in the test than the single-laser power condition. Furthermore, Fig. 7(j) presents a comparison of the R^2 values according to the laser power conditions (Fig. S18). The data at 0.1 mW exhibited relatively modest values owing to the low intensity, a limitation that can be addressed by augmenting the training data. The SAAN model learned all experimental data under laser power conditions ranging from 0.1 to 1 mW, compensating for variations between data points. As shown in Fig. S19(a), the t-SNE analysis of the original data forms clusters based on laser power, regardless of ethanol content. Notably, 0.1 mW and 0.5 mW form continuous bands, limiting common inferences about ethanol content. However, t-SNE analysis of the feature values obtained during the SAAN model computation process shows clusters forming across three distinct ethanol content regions (0.01%, 0.03–0.05%, and 0.08–0.1%) regardless of laser power (Fig. S19(b)). This result indicates that the guide model successfully suppressed the effect of laser power and performed classification. Furthermore, after reducing the data feature weight derived from laser power in the low, medium, and high-content models, ethanol content quantification was efficiently inferred.

The channel-wise feature maps for the test results of the SAAN model under the 0.1% ethanol-content condition are shown in Fig. 7(k) (Fig. S20). In a manner analogous to the characteristics of the actual input data, the focal maximum intensity region manifested the most intense response, whereas other regions had lower values. Furthermore, all channels manifested analogous patterns. To compare the performance of this model with that of a general model, we designed a unified model (Tables S9 and S10). The unified model exhibited a structural resemblance to the ethanol-content model of the SAAN model, with input data undergoing direct regression operations to infer ethanol content without the necessity of model segmentation. The unified model was designed with the exclusion of data characterized by noise or intensity-wide anomalies, as plotted in Fig. 7(l) (Fig. S21). After extracting features from the original data, a unified model that quantifies ethanol content in a single step may struggle to simultaneously recognize data differences due to ethanol content and data variations due to laser power. This is because the input data requires quantifying float values within a 10-fold range of 0.01%–0.1%, while also

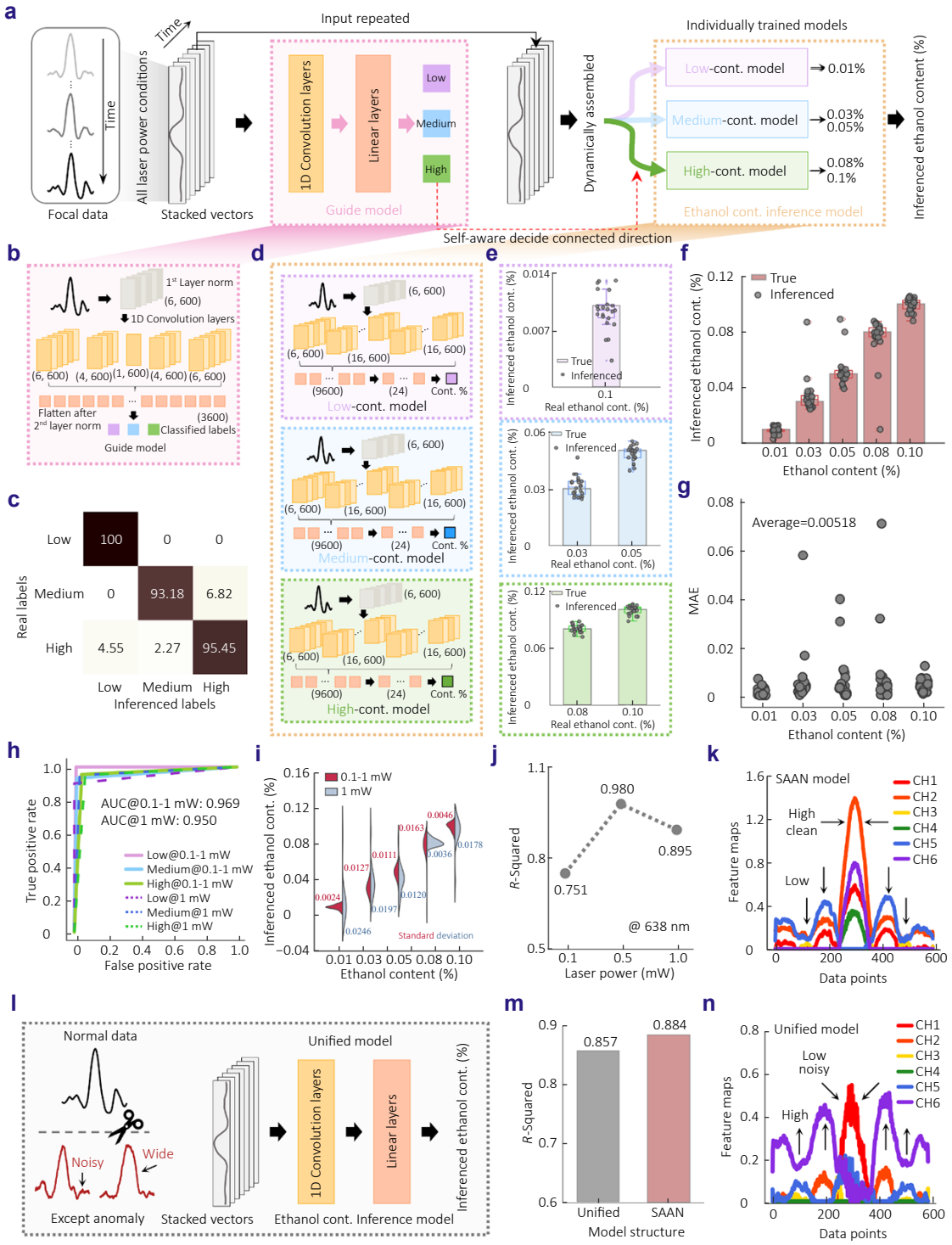

Fig. 7 | Principle of operation of SAAN models and ethanol-content inference results. (a) Mechanism for inferring ethanol content of the self-aware assembly network (SAAN) model. (b) Structure of the guide model. (c) Confusion matrix of the ethanol-content level classification results of the guide model. (d) Structure of the low-, medium-, and high-ethanol-content models. (e) Ethanol-content inference results for each level ethanol-content models. Ethanol-content inference results of the SAAN model: (f) comparison of real values and inferred values, (g) MAE for each ethanol-content condition. Comparison of inference results from single- or mixed-laser power condition datasets: (h) ROC curve comparison of the guide model and (i) comparison of the distributions and standard deviations of inferred ethanol-content values in the test set. (j) Comparison of model inference performances (R^2) for each laser power condition. (k) Channel-wise plot of the feature maps from the last 1D convolutional layer when the SAAN model operated. (l) Anomaly removal process and model structure of the unified model, which was the comparison model. (m) Comparison of the ethanol-content inference performance R^2 between the unified and SAAN models. (n) Channel-wise plot of the feature maps from the last 1D convolutional layer of the unified model.

compensating for data fluctuations due to laser power, increasing the difficulty for the model to recognize features related to ethanol content during training. However, the SAAN model effectively addresses this by limiting the inference range of ethanol content through its initial guide model, thereby reducing the computational scope for regression models and enabling more accurate feature recognition. Furthermore, by limiting the inference range, the guide model in the SAAN model reduces the impact of laser power on data during regression computations. This maximizes the ethanol content feature, resulting in a stable model architecture for ethanol content quantification. Notwithstanding the data selection, the unified model attained an R^2 value of 0.857, whereas the SAAN model achieved 0.884, as shown in Fig. 7(m) (Fig. S22). Furthermore, a test of the data under the ethanol content of 0.1% revealed that the channel-wise feature maps of the final 1D convolution layer of the model exhibited low feature values in proximity to the focal maximum intensity, as shown in Fig. 7(n). Feature values at other locations were found to be higher, suggesting an absence of consistent channel-wise features. Optical changes due to ethanol content occur primarily in the focal maximum intensity region. The SAAN model aims to perform scientific ethanol content inference that can be linked to hardware responses, evaluating the deep learning black box through feature maps. As data passes through 1D convolution layers while maintaining its length, parts judged as important features are emphasized, while noise components are weakened or simplified. The SAAN model exhibits a feature map clearly concentrated in the focal region, exhibiting a shape nearly identical to the raw data. This indicates the SAAN model reads scientific characteristics related to ethanol content. However, the unified model displays a dispersed pattern indicating a noisy state, suggesting the model was unstable in finding features. It particularly struggled to accurately locate feature positions, making accurate inference difficult when new data is input. Furthermore, the SAAN model's feature map (Fig. 7(k)) shows a feature value of approximately 1.4 at the focal region. This concentration of features in the focal region results in a value up to three times higher than in other areas. This strongly indicates that the model identifies the focal region as the main feature of the data. Furthermore, values ranging from 0.3 to 1.4 are distributed across the channels (time conditions), suggesting that each channel plays a distinct role during the model's final inference operations. However, the feature map of the unified model (Fig. 7(n)) shows feature values below 0.6, which is 2.3 times smaller than the SAAN model. Furthermore, specific key regions within the entire data length are not well-represented, indicating lower model stability.

In this study, we developed the SAAN model to quantify ethanol content based on the principle of collective scattering-induced wavefront perturbation in ethanol-water mixtures. This wavefront perturbation systematically varies with ethanol content, forming unique optical characteristics that clearly distinguish it from other nonpolar or weakly interacting molecular species. The SAAN model is designed not merely to measure a single variable, but to learn these multidimensional spatial features, encoding the ensemble's unique optical signature. Notably, even under ultra-trace conditions where the ethanol-to-water ratio is $1:10^4$, the SAAN model achieved excellent regression performance (R^2 = 0.8719) (Fig. S23). This demonstrates the ability to recognize ethanol molecules with high selectivity even amidst an overwhelmingly high proportion of H_2O molecules. Consequently, the combination of the molecular detection mechanism proposed in this study and the SAAN model suggests successful extension to future real-world applications.

The temporal performance of the proposed system is determined by both acquisition and computation processes. In the present implementation, focal-spot videos were recorded at 2 fps for 3 s, producing six frames for each inference. This corresponds to an effective sensing rate of approximately 0.33 Hz. The inference time of the trained deep-learning model is approximately 0.036 s, indicating that the overall response time (~3.036 s) is primarily limited by data acquisition rather than computation. Because this constraint originates from the imaging configuration rather than the sensing mechanism itself, the response speed can be substantially improved by employing higher-frame-rate detectors or reduced frame sampling, suggesting strong potential for real-time sensing in optimized systems.

3 Conclusions

In this study, we developed and validated a high-precision, non-contact ethanol-content inference system by combining an optical system based on a multi-layer graphene Fresnel lens with a self-aware assembly network (SAAN) model.

The core sensing principle of this system is based on the phenomenon of Rayleigh scattering between laser light and gaseous ethanol molecules. Rayleigh scattering is the scattering of light by molecules smaller than its wavelength, and the total amount of scattered light is directly dependent on the optical properties (polarizability) and density (concentration) of the molecules. Notably, ethanol molecules induce stronger scattering than water molecules owing to their larger average polarizability and anisotropy of polarizability. Therefore, as the ethanol content changes, the degree of light loss due to scattering also changes, which in turn causes minute variations in the optical properties (e.g., intensity, geometry) of the final focal spot that passes through the lens.

The short wavelength of 405 nm exhibited the strongest scattering interaction and highest sensitivity to changes in ethanol content but was limited by stability owing to the high variability of the scattering signal. In contrast, the longer wavelength of 638 nm exhibits substantially weaker scattering-induced attenuation owing to the λ^{-4} dependence of the Rayleigh scattering cross-section. As a result, the

detected signal remains within the quasi-linear dynamic range of the detector, and the peak intensity varies only minimally with concentration. Importantly, this stability does not indicate saturation, but rather operation in a weak-attenuation regime. Under these conditions, concentration-dependent information is primarily encoded in spatial features such as FWHM broadening rather than peak intensity variation. This separation of stable intensity and variable spatial features provides a favorable representation for deep-learning-based inference, leading to improved prediction stability and accuracy.

The proposed SAAN model adopts a hierarchical structure that first classifies the ethanol-content range and then connects to a regression model optimized for that specific range. This approach achieved significantly higher classification and regression accuracy (R^2=0.884, *MAE*=0.00518) and stability under mixed-laser power conditions when compared with single-laser conditions or unified regression models. These results indicate that the SAAN architecture can robustly handle the complexity and noise of optical data.

The measurable concentration range of the proposed sensing approach is fundamentally bounded by optical and instrumental constraints. At the lower limit, detectability is restricted by the signal-to-noise ratio of the imaging system, as scattering-induced modulation becomes comparable to or smaller than the detector noise floor at extremely low concentrations. At the upper bound of the measurable range, excessive scattering can introduce strong wavefront perturbations that distort the focal profile. This behavior reflects a fundamental physical constraint associated with strong scattering rather than a limitation of the proposed method. Therefore, the practical sensing range is determined by a balance between sufficient scattering contrast and preservation of stable focal-spot structure.

In conclusion, this paper presents a new optical sensing paradigm that converges Rayleigh scattering, diffractive optics, and deep learning to overcome the limitations of conventional contact-based sensors, such as material degradation, response delays, and maintenance requirements. Furthermore, the sensing principle of this system is based on the unique optical properties of each gas molecule. Because each molecule possesses distinct polarizability characteristics, it gives rise to molecule-dependent variations in the scattered light and the resulting focal profile. Therefore, while the present study focuses on ethanol, the proposed sensing framework is not inherently limited to a single molecular species. Its applicability to other gases depends on the optical properties of the target species, including their average polarizability, polarizability anisotropy, and the resulting Rayleigh scattering cross-section. In particular, the method requires that scattering-induced wavefront modulation remains detectable within the dynamic range of the imaging system. Gas species with sufficiently distinct anisotropic scattering characteristics may therefore produce discriminative focal-profile signatures. When the optical contrast is weak, multi-wavelength excitation may further enhance separability by exploiting the λ^{-4} dependence of Rayleigh scattering.

4 Materials and methods

4.1 CVD graphene

The single-layer graphene in this study was fabricated using chemical vapor deposition (CVD) within a quartz chamber. A 50 mm-diameter Goodfellow Cu foil (100 μm, CU000565) was used as the growth substrate, onto which an annealing process was first performed at 1010 °C for 30 min. Graphene synthesis was then conducted in a 1010 °C environment by flowing a gas mixture of methane (CH_4) and hydrogen (H_2) at flow rates and partial pressures of 20 sccm (290 mTorr) (1 mTorr = 0.133 Pa) and 8 sccm (100 mTorr), respectively. Subsequently, for the transfer process, a polymethyl methacrylate (PMMA) (MicroChem Corp., 950 K A6) support layer was spin-coated onto the graphene/Cu substrate at 4200 rad/min for 32 s, followed by 2 min of baking at 180 °C to enhance adhesion.

4.2 Fabrication of the Fresnel lens

To fabricate the multi-layer graphene film, we removed the Cu layer from the PMMA-supported graphene/Cu structure via wet etching with an ammonium persulfate ($(NH_4)_2S_2O_8$, Sigma-Aldrich) solution. Subsequently, the process of washing the film with water and transferring it onto a glass substrate was repeated five times to form a five-layer structure. Finally, a focused ion beam (FIB) process was utilized to create the Fresnel lens pattern on the five-layer graphene film. This procedure was conducted in the following sequence: deposition of titanium as a mask layer, milling of the titanium with a Crossbeam 540 instrument equipped with a gallium ion source (30 kV, 3 nA, dose factor 0.8), removal of the exposed graphene with oxygen plasma, and then removal of the residual titanium mask with an etchant.

4.3 Ethanol–water solution

The ethanol aqueous solutions for this study were prepared in volumes of 500 mL, with final ethanol content of 0.01%, 0.03%, 0.05%, 0.08%, and 0.1% in deionized (DI) water.

4.4 Ethanol-content optical sensor

Aqueous ethanol solutions with concentrations ranging from 0.01% to 0.1% were vaporized using a temperature-controlled heating system. The vaporization temperature was actively regulated by a resistive heater equipped with a feedback temperature controller and maintained at an average of 33.997 °C, with a measured fluctuation range of 33.974–34.021 °C during operation. Nitrogen (N_2) was used

as a carrier gas at a constant flow rate of 20 standard liters per minute (SLM). The carrier gas was passed through the heated ethanol-solution chamber to promote vapor saturation prior to injection into the optical path. The vapor delivery conditions—including flow rate, injection geometry, and the distance between the vapor outlet and the laser beam —were kept constant throughout all experiments to maintain steady-state operation and measurement repeatability. Therefore, the concentration levels investigated in this study are reported as nominal (labeled) conditions based on the prepared solution concentrations, rather than independently calibrated absolute vapor concentrations in the interaction region. A Cobolt Skyra multi-line laser was utilized as the light source, emitting beams at wavelengths of 638, 512, and 405 nm. The laser beam passed through a graphene Fresnel lens to form a focal point, which was captured using a Teledyne FLIR Blackfly S USB3 (12.3 MP) CMOS camera. For each experimental condition, a 3 s video of the focal spot was recorded at 2 fps. The number of repeated experiments for each of the five ethanol contents varied with laser power: 70 repetitions at 1 mW and 20 repetitions at 0.5 and 0.1 mW.

4.5 Classification model

In this study, a 1D CNN model was designed to classify ethanol-content levels from 1D intensity profiles. The model was designed using Python 3.11.13 and PyTorch 2.6.0+cu124 packages, and data visualization was performed using Matplotlib 3.10.0 and Seaborn 0.13.2 packages. A detailed description of the classification model structure is provided in Supplementary Note S3.

4.6 SAAN model

The SAAN model was designed using Python 3.11.13 and PyTorch 2.6.0+cu124 packages, and data visualization was performed using Matplotlib 3.5.1 and Seaborn 0.11.2 packages. A detailed description of the SAAN model structure is provided in Note S4.

Average of all batch losses within each epoch, formulated as:

$$L_{\text{epoch}} = \frac{1}{N}\sum_{i=1}^{N} L_{\text{batch},i} \ , \tag{6}$$

where N is the total number of batches.

Mean Absolute Error (nMAE) was calculated as follows:

$$nMAE = \frac{MAE}{y_{\max} - y_{\min}} \times 100(\%) \ . \tag{7}$$

References

1. Bohren CF, Huffman DR. *Absorption and Scattering of Light by Small Particles* (John Wiley & Sons, New York, 1998).
2. van de Hulst HC. *Light Scattering by Small Particles* (Dover Publications, New York, 1981).
3. Strutt HJW. XV. On the light from the sky, its polarization and colour. *London, Edinburgh, Dublin Philos Mag J Sci* **41**, 107–120 (1871).
4. Bucholtz A. Rayleigh-scattering calculations for the terrestrial atmosphere. *Appl Opt* **34**, 2765–2773 (1995).
5. Eckbreth AC. *Laser Diagnostics for Combustion Temperature and Species* (CRC Press, London, 2022).
6. Buck U, Huisken F. Infrared spectroscopy of size-selected water and methanol clusters. *Chem Rev* **100**, 3863–3890 (2000).
7. Hartweg S, Yoder BL, Garcia GA et al. Size-resolved photoelectron anisotropy of gas phase water clusters and predictions for liquid water. *Phys Rev Lett* **118**, 103402 (2017).
8. Dumas-Campagna J, Tardif R, Charest-Tardif G et al. Ethanol toxicokinetics resulting from inhalation exposure in human volunteers and toxicokinetic modeling. *Inhal Toxicol* **26**, 59–69 (2014).
9. Nazaroff WW, Weschler CJ. Methanol and ethanol in indoor environments. *Indoor Environ* **1**, 100049 (2024).
10. Fenske JD, Paulson SE. Human breath emissions of VOCs. *J Air Waste Manage Assoc* **49**, 594–598 (1999).
11. Haynes WM. *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, 2016).
12. Dixit S, Crain J, Poon WCK et al. Molecular segregation observed in a concentrated alcohol-water solution. *Nature* **416**, 829–832 (2002).
13. Kristensson E, Ehn A, Bood J et al. Advancements in Rayleigh scattering thermometry by means of structured illumination. *Proc Combust Inst* **35**, 3689–3696 (2015).
14. Hu QH, Shang F, Ma L et al. Revolutionizing underwater sensor performance: tackling rayleigh scattering challenges by pseudo random noise. *Adv Sci* **12**, 2411967 (2025).
15. Zhang L, Costa LD, Yang ZS et al. Analysis and reduction of large errors in Rayleigh-based distributed sensor. *J Lightwave Technol* **37**, 4710–4719 (2019).
16. Stetter JR, Li J. Amperometric gas sensors-a review. *Chem Rev* **108**, 352–366 (2008).
17. Ozoemena KI, Musa S, Modise R et al. Fuel cell-based breath-alcohol sensors: innovation-hungry old electrochemistry. *Curr Opin Electrochem* **10**, 82–87 (2018).
18. De Martinis BS, Martins Ruzzene MA, Santos Martin CC. Determination of ethanol in human blood and urine by automated headspace solid-phase microextraction and capillary gas chromatography. *Anal Chim Acta* **522**, 163–168 (2004).
19. Wang X, Yee S, Carey P. An integrated array of multiple thin-film metal oxide sensors for quantification of individual components in organic vapor mixtures. *Sens Actuators B Chem* **13**, 458–461 (1993).
20. Zagal-Padilla CK, Diaz-Gómez C, Gamboa SA. Electrochemical characterization of a plasmonic effect ethanol sensor based on two-dimensional ZnO synthesized by green chemistry. *Mater Sci Semicond Process* **137**, 106240 (2022).
21. Thungon PD, Kakoti A, Ngashangva L et al. Advances in developing rapid, reliable and portable detection systems for alcohol. *Biosens Bioelectron* **97**, 83–99 (2017).
22. Wang L, Cheng Y, Naidu R et al. Application of portable gas chromatography-mass spectrometer for rapid field based determination of TCE in soil vapour and groundwater. *Environ Technol Innov* **21**, 101274 (2021).
23. Zhang J, Liu L, Yang Y et al. A review on two-dimensional materials for chemiresistive- and FET-type gas sensors. *Phys Chem Chem Phys* **23**, 15420–15439 (2021).
24. Korotcenkov G. Metal oxides for solid-state gas sensors: what determines our choice. *Mater Sci Eng B* **139**, 1–23 (2007).
25. Baek JW, Shin E, Lee J et al. Present and future of emerging catalysts in gas sensors for breath analysis. *ACS Sens* **10**, 33–53 (2025).
26. Ahn EC. 2D materials for spintronic devices. *npj 2D Mater Appl* **4**, 17

(2020).
27. Fan FR, Wang RX, Zhang H et al. Emerging beyond-graphene elemental 2D materials for energy and catalysis applications. *Chem Soc Rev* **50**, 10983–11031 (2021).
28. Zhang ZB, Forti S, Meng WQ et al. Growth and applications of two-dimensional single crystals. *2D Mater* **10**, 032001 (2023).
29. Jeong GH, Sasikala SP, Yun T et al. Nanoscale assembly of 2D materials for energy and environmental applications. *Adv Mater* **32**, 1907006 (2020).
30. Sabattini L, Coriolano A, Casert C et al. Towards AI-driven autonomous growth of 2D materials based on a graphene case study. *Commun Phys* **8**, 180 (2025).
31. Bonaccorso F, Sun Z, Hasan T et al. Graphene photonics and optoelectronics. *Nat Photonics* **4**, 611–622 (2010).
32. Bao QL, Loh KP. Graphene photonics, plasmonics, and broadband optoelectronic devices. *ACS Nano* **6**, 3677–3694 (2012).
33. Shao LD, Zhang ZP, Wang X et al. Graphene-based ultralow-profile microwave Fresnel lens. *Carbon* **217**, 118599 (2024).
34. Novoselov KS, Geim AK, Morozov SV et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
35. Geim AK, Novoselov KS. The rise of graphene. *Nat Mater* **6**, 183–191 (2007).
36. Lin H, Sturmberg BC, Lin KT et al. A 90-nm-thick graphene metamaterial for strong and extremely broadband absorption of unpolarized light. *Nat Photonics* **13**, 270–276 (2019).
37. Park S, Lee G, Park B et al. Electrically focus-tuneable ultrathin lens for high-resolution square subpixels. *Light Sci Appl* **9**, 98 (2020).
38. Kim TT, Kim HD, Zhao R et al. Electrically tunable slow light using graphene metamaterials. *ACS Photonics* **5**, 1800–1807 (2018).
39. Park S, Park C, Hwang YJ et al. Focus-tunable planar lenses by controlled carriers over exciton. *Adv Opt Mater* **9**, 2001526 (2021).
40. Zheng XR, Jia BH, Lin H et al. Highly efficient and ultra-broadband graphene oxide ultrathin lenses with three-dimensional subwavelength focusing. *Nat Commun* **6**, 8433 (2015).
41. Koppens FHL, Mueller T, Avouris P et al. Photodetectors based on graphene, other two-dimensional materials and hybrid systems. *Nat Nanotechnol* **9**, 780–793 (2014).
42. Carleo G, Cirac I, Cranmer K et al. Machine learning and the physical sciences. *Rev Mod Phys* **91**, 045002 (2019).
43. Barbastathis G, Ozcan A, Situ GH. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
44. Liu SS, Hu NL, Hu JW et al. DNA-mediated bioinspired MXene gas sensor array with machine learning for noninvasive cancer recognition. *ACS Nano* **19**, 25363–25384 (2025).
45. Zhu JX, Ren ZH, Lee C. Toward healthcare diagnoses by machine-learning-enabled volatile organic compound identification. *ACS Nano* **15**, 894–903 (2020).
46. Guo YM, Zhong LB, Min L et al. Adaptive optics based on machine learning: a review. *Opto-Electron Adv* **5**, 200082 (2022).
47. Guo M, Wu YC, Hobson CM et al. Deep learning-based aberration compensation improves contrast and resolution in fluorescence microscopy. *Nat Commun* **16**, 313 (2025).
48. Zuo C, Qian JM, Feng SJ et al. Deep learning in optical metrology: a review. *Light Sci Appl* **11**, 39 (2022).
49. Li ZS, Sun JS, Fan Y et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electron Sci* **2**, 220023 (2023).
50. Krasikov S, Tranter A, Bogdanov A et al. Intelligent metaphotonics empowered by machine learning. *Opto-Electron Adv* **5**, 210147 (2022).
51. Li SW, Xing J. DeepSAT4D: deep learning empowers four-dimensional atmospheric chemical concentration and emission retrieval from satellite. *Innov Geosci* **2**, 100061 (2024).
52. Penfold TJ, Rankine CD. A deep neural network for valence-to-core X-ray emission spectroscopy. *Mol Phys* **121**, e2123406 (2023).
53. Del Bono V, Capaldi E, Kelshiker A et al. Multi-modal sensing in colonoscopy: a data-driven approach. *IEEE Robot Autom Lett* **11**, 2018–2025 (2026).
54. Capman NSS, Zhen XV, Nelson JT et al. Machine learning-based rapid detection of volatile organic compounds in a graphene electronic nose. *ACS Nano* **16**, 19567–19583 (2022).
55. Palacín J, Clotet E, Rubies E. Assessing over time performance of an eNose composed of 16 single-type MOX gas sensors applied to classify two volatiles. *Chemosensors* **10**, 118 (2022).
56. Heng SY, Yap KZ, Lim WY et al. AI-assisted sensor system for the acetone and ethanol detection using commercial metal oxide-based sensor arrays and convolutional neural network. *Sens Imaging* **25**, 51 (2024).

Acknowledgements

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (RS-2024-00339770) and Korea Environment Industry & Technology Institute through the Technology Development Project for Biological Hazards Management in Indoor Air Program (or Project), funded by the Korea Ministry of Environment (ARQ202101038001).

Author contributions

G.M.K., Y.J.H., S.C.J. designed the research. G.M.K. performed the experiments. G.M.K. and Y.J.H. analyzed data and wrote the manuscript. C. L. fabricated the Fresnel lens. T.H. assisted data analyzation and visualization. J.H. assisted with data analysis and discussion. S.C.J. supervised the research.

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary information for this paper is available at
https://doi.org/10.29026/oea.2026.250278