Artificial Intelligence & Environment

https://doi.org/10.66178/aie-0026-0006

Research

Open Access

Exploring emerging exposure scenarios via intelligent non-target screening: Chemical space characterization of pet hair contaminants

Liwei He[1,2,3], Fei Cheng[1*], Jingwen Shi[2], Huizhen Li[2], Zhiqiang Yu[1] and Jing You[2]

Citation: He LW; Cheng F; Shi JW; et al. Exploring emerging exposure scenarios via intelligent non-target screening: Chemical space characterization of pet hair contaminants. *AI Environ.* 2026, 1(1): 40-51.

Received: 15 November 2025

Revised: 30 December 2025

Accepted: 21 January 2026

Published online: 11 February 2026

Abstract

Pet ownership in urban households has created unique indoor exposure scenarios where humans and their companion animals cohabit intensively. Pet hair has become an emerging source of contaminants that may affect human health, yet the chemical composition and accumulation patterns in pet hair remain poorly characterized. Herein, an integrated data-to-laboratory framework combining text mining, machine learning, and high-resolution mass spectrometry-based non-target screening was developed to investigate the chemical space in pet hair and indoor dust systematically. A total of 16,692 publications were mined to curate 3,684 indoor-relevant chemicals from which we established a molecular ion library containing 2,661 compounds. Complementarily, a diagnostic ion library with 30 fragment ions was generated using an artificial neural network coupled with a convolutional capsule neural network. Together, the two newly established libraries were employed for non-target screening of 14 pet hair and 10 dust samples. These scenario-specific libraries substantially enhanced the coverage of chemical identification, increasing the number of annotated compounds in major chemical categories up to 28-fold and reducing analysis time by 47-fold compared with conventional target analysis. Non-target screening revealed high similarity between matrices, with over 50% of shared molecular features and consistent distributions of functional and structural categories. Artificial neural network analysis identified key physicochemical descriptors governing contaminant accumulation, including log P, pK_a, aromatic atom count, molecular flexibility, and polarizability. Further analysis of non-targeted compounds confirmed that these parameters drive the observed similarity between pet hair and dust.

Keywords: Indoor air pollution, Pet hair, Non-target screening, Chemical space characterization, Scenario-specific spectral library

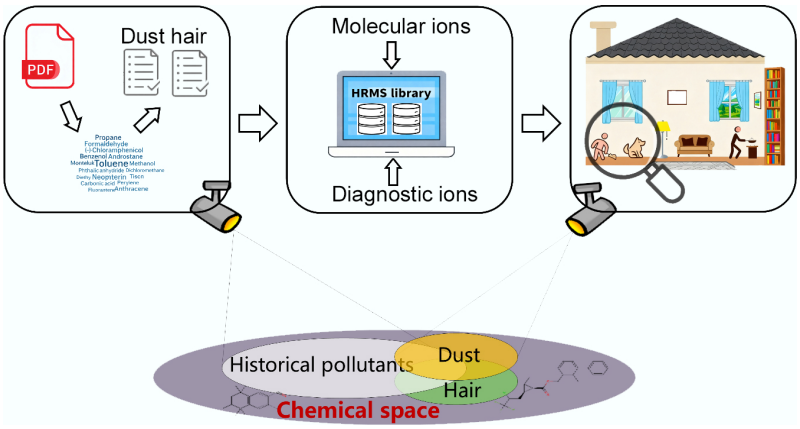
Highlights

- Integrated text mining, machine learning, and HRMS for indoor contaminant screening.
- Scenario-specific libraries increased annotation by 2–28-fold and reduced analytical time by 47-fold.
- More than 50% of contaminants were shared between pet hair and indoor dust.
- An ANN identified log P, pK_a, aromaticity, flexibility, and polarizability as key enrichment drivers.

* Correspondence: Fei Cheng (chengfei@gig.ac.cn)

Full list of author information is available at the end of the article.

Graphical abstract



Introduction

With rapid urbanization and increasingly diverse lifestyles, pet ownership has become a defining feature of modern urban living, giving rise to a large emerging demographic of pet-owning households. In the United States, the average household owns 1.5 dogs and 1.8 cats[1], and similar trends have been observed in China, where more than 100 million urban residents now keep pets indoors. Unlike outdoor pets, most companion animals in urban settings spend nearly all of their lifetimes inside homes, where residents also spend about 90% of their daily lives[2]. Such intensive human-pet cohabitation dramatically alters the indoor environment by introducing new sources of biological residues (e.g., dander, hair, and skin oils) and influencing the environmental behaviors of chemical contaminants. Consequently, the household microenvironment becomes a complex interface where emissions from building materials, furniture, and consumer products interact with those from human and pet activities[3,4], such as cooking[5] and pet movement[6], and are redistributed throughout the indoor environment via air or settled dust[7]. Understanding how these processes reshape the distribution and accumulation of indoor pollutants is therefore critical for characterizing emerging exposure pathways in a human-pet shared living environment.

Indoor environmental media such as airborne particles and dust can effectively accumulate organic pollutants, and their distribution is largely determined by their physicochemical properties. Over the past decade, indoor dust has been widely used as a representative matrix for evaluating plasticizers, pharmaceuticals and personal care products, and pesticides[8]. However, most indoor sampling studies have inadvertently collected pet hair together with dust, treating it as part of the solid matrix rather than as a distinct biological medium. In reality, pet hair is fundamentally different from dust. It originates from living organisms, contains lipid- and protein-rich surfaces, and was in direct and prolonged contact with skin and household air. These features likely enabled pet hair to selectively accumulate more bioavailable and biologically relevant contaminants, reflecting both the chemical compositions of the indoor environment and the potential for internal exposure. Recent studies have therefore proposed pet hair as a potential "sentinel" matrix for indoor exposure[9,10], supported by the physiological similarities between humans and pet animals, as domestic dogs shared more than 360 diseases with humans, including several cancers[11]. Developing analytical approaches specially tailored to pollutants retained in pet hair is thus essential for comprehensively characterizing indoor chemical mixtures and for establishing a new perspective on exposure assessment in pet-owning households.

Despite growing awareness of indoor exposure, emerging exposure scenarios such as those involving pet hair remained poorly characterized, largely due to the lack of targeted analysis frameworks and suitable reference standards. Conventional targeted analysis was inherently limited by its dependence on predefined compound lists and prior expert knowledge, making it inadequate for capturing the chemical diversity of complex indoor environments[12]. The advent of high-resolution mass spectrometry (HRMS) has greatly expanded the capabilities of non-target screening (NTS), enabling the detection of unknown features across diverse matrices[13,14]. However, this methodological capability came with the challenge of managing enormous datasets, where distinguishing true contaminants from false positives remained a bottleneck. To fully harness the potential of NTS, it was essential to clearly define an environmentally relevant chemical space, representing the subset of molecular features likely to occur in a given exposure scenario[15]. Scenario-specific HRMS libraries would have provided an effective way to constrain this chemical space, improving the relevance of detected compounds to their matrices and enhancing screening efficiency[15]. Constructing such libraries increasingly relied on artificial intelligence techniques, which enable the systematic extraction of pollutant information from historical literature and the analysis of molecular features to guide spectral database development[16]. In previous studies, we developed a big-data-based framework integrating literature-mined chemical information[17] and HRMS database curation to generate molecular and diagnostic ion libraries[15].

Building upon this foundation, the present study examined physicochemical parameters that governed contaminant retention in certain matrices and established HRMS-based analytical strategies tailored to emerging exposure scenarios, thereby elucidating mechanisms of contaminant enrichment in pet hair and indoor dust. First, chemical information on indoor pollutants was systematically extracted from over 16,000 publications through large-scale text mining of literature, enabling a comprehensive overview of the indoor air-related contaminant inventory. Second, molecular and diagnostic ion libraries were constructed for scenario-specific NTS by combining machine learning-based ion prediction with curated literature datasets, thereby establishing an HRMS strategy tailored to pollutant accumulation in pet hair. Third, chemical space characterization was performed to differentiate contaminants enriched in

https://doi.org/10.66178/aie-0026-0006

dust and pet hair, to identify structural profiles of compounds prevalent in the pet hair scenario, and to determine physicochemical parameters that influence their fate. Collectively, these efforts developed a framework for applying intelligent NTS to emerging exposure matrices and for evaluating the scientific relevance of pet hair in future environmental monitoring.

Materials and Methods

Experimental design

The present study was designed following an integrated data-to-laboratory framework (Fig. 1). First, literature related to indoor air pollution was systematically collected, comprising more than 10,000 publications, from which full-text information was extracted for subsequent analysis. In parallel, approximately 900,000 chemical entries were retrieved from the CompTox Chemicals Dashboard (CCD) and supplemented with compound names obtained from the PubChem. Second, the literature corpus was segmented and tokenized, and word embeddings were matched against a curated chemical lexicon using an optimized named entity recognition (NER) model, as described previously[17]. This procedure generated a comprehensive inventory of chemicals reported in indoor air pollution-related studies. Third, an artificial neural network (ANN) was applied to identify key physicochemical parameters governing the accumulation behavior of indoor pollutants. Based on the chemical inventory derived from text mining, diagnostic ion fragments were subsequently predicted using a cascaded convolutional neural network (CCNN) model to enable the identification of previously unreported contaminants. On this basis, two HRMS libraries, a molecular ion library and diagnostic ion library, were constructed. Finally, indoor dust and pet hair samples were collected and analyzed using NTS to characterize the chemical space of indoor pollutants under the defined exposure scenario.

Data collection

The data were obtained from three major sources: the CCD, the Web of Science (WoS), and PubChem. In the WoS, English-language publications indexed in the Science Citation Index Expanded (SCI-EXPANDED) between 1990 and 2021 were retrieved using the search string: TS = ((Occurrence OR Concentration) OR (Hazard OR Risk) OR (Monitoring)) AND TS = ((Indoor air) OR (Indoor dust) OR (Indoor PM2.5)). Relevant metadata, including publication year, abstracts, DOIs, and other bibliographic information, were exported for all records. The search was restricted to the WoS Core Collection and the SCI-EXPANDED databases. A total of 16,692 publications with available PDFs were collected. The full text of each PDF was extracted and saved as CSV files, forming a raw text corpus for subsequent NER.

The lexicon was developed based on previous studies[17] and constructed through a two-stage process. In brief, we accessed the CCD for the DSSTox Identifiers file in January 2022. The file contained the DSSTox Substance Identifiers (DTXSID) of 906,511 chemicals, along with their Chemical Abstracts Service Registry Numbers (CASRN) and Preferred Names. In March 2022, each CASRN was subsequently queried in PubChem to retrieve the corresponding PubChem compound identifier (CID). Using these CIDs, a list of synonyms for each chemical was extracted using the PubChemPy program (https://github.com/mcs07/PubChemPy). In total, 7.85 million names were extracted for these 906,511 chemicals, which were used as the lexicon for the subsequent NER. As this lexicon is restricted to compounds included in existing databases, it is inherently biased toward existing chemicals and may underrepresent other, unlisted substances.

Text mining was conducted based on previously established methods[17]. Briefly, sentence segmentation and word vectorization were performed using NLTK and Word2Vec based on the curated lexicon. Bi-LSTM+CRF and Bi-GRU neural network models, in combination with dictionary-matching strategies, were applied for text cleaning and NER, resulting in a comprehensive list of contaminants reported in the context of indoor air pollution. For each identified chemical entity, the corresponding DTXSID was used to retrieve the CID from PubChem, and the related Medical Subject Headings (MeSH) categories were assigned. During corpus processing, rule-based text classification was also performed to categorize literature according to environmental matrices, including dust and pet hair.

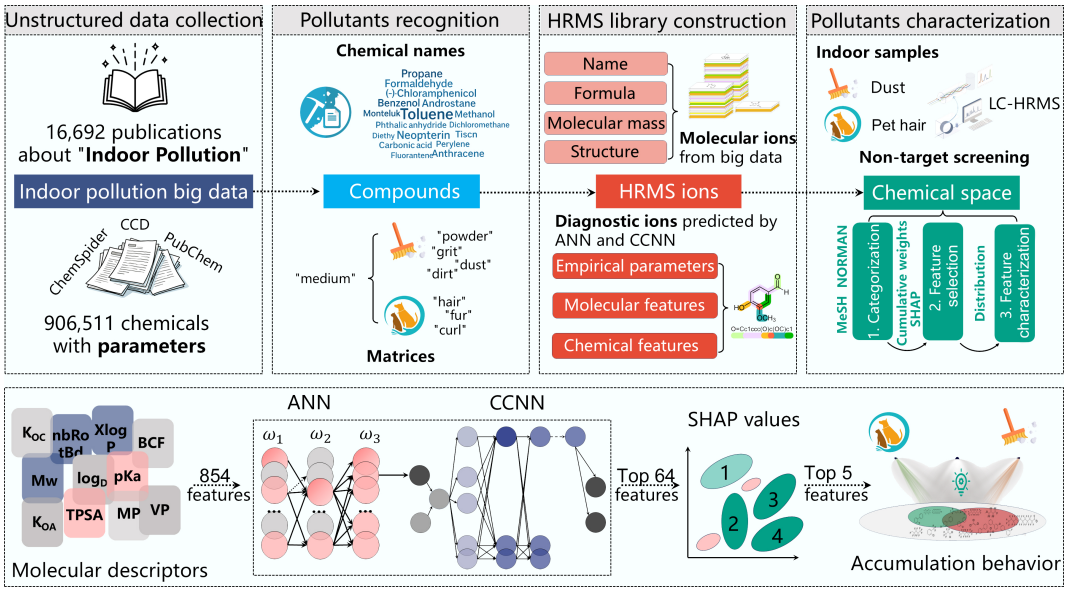


Fig. 1 Workflow of the established framework integrating text mining, machine learning, and non-target screening for identifying contaminants in dust and pet hair samples. CCD: CompTox Chemical Dashboard; WOS: Web of Science.

More details on matrix labelling procedures are provided in the Supporting Information.

Indoor environmental scenario-specific HRMS library

The HRMS library developed for indoor exposure scenarios (pet hair and dust) consisted of two components, including a molecular ion library constructed from known scenario-relevant pollutants collected through text mining, and a diagnostic ion library predicted via an ANN coupled with a CCNN model for unknown pollutants. The molecular ion library was designed for rapid, suspect identification of pollutants within the defined scenario, whereas the diagnostic ion library was intended to facilitate the detection of additional compounds based on structural and fragment-ion information.

A molecular ion library containing 2,661 substances was established for suspect screening (Supplementary Sheet 02). Initially, 3,684 compounds were retrieved from the text mining results (Supplementary Sheet 01). As not all compounds are ionizable during HPLC-HRMS analysis, substances were curated based on their molecular formulas, IUPAC names, exact masses (reported to five decimal places), and ionization data from the ChemSpider database. The curated dataset was subsequently filtered to remove species with charge states unsuitable for import, ultimately retaining 2,661 substances. The remaining entries were manually verified and supplemented with additional metadata from ChemSpider. Corresponding MOL files were then generated and imported into the Mass List section of the Compound Discoverer 3.3. The resulting library included comprehensive information, encompassing molecular identity, formula, exact mass, and structural representation.

In addition, a diagnostic ion library was also constructed using a cascaded deep learning framework comprising two sequentially linked models adapted from previous studies[18], which were used for predicting characteristic fragments (ANN) and diagnostic ions (CCNN), respectively. The dataset of 3,684 indoor air-relevant substances was partitioned into a training set (90%) and a test set (10%). To address class imbalance, the training procedure incorporated appropriate weighting strategies, and a 10-fold cross-validation strategy was applied.

The reliability of the diagnostic ion library developed in the present study was interpreted within a multi-level validation framework. The underlying model architecture, learning strategy, and fragment-to-ion mapping logic were not newly introduced, but inherited from and previously validated in our earlier studies[18]. In particular, the methodological principle was systematically verified through stacked reverse validation using QSAR-Ready and MS-Ready datasets, ensuring consistency between structure-based predictions and MS-interpretable evidence. In subsequent work, the practical reliability of the resulting spectral libraries was further confirmed using real environmental samples and reference standards in HRMS-based non-target screening. Building on this established validation, the present study focused on extending the framework to an indoor exposure-specific chemical space and assessing its performance at a certain scenario level. Accordingly, reliability was assessed here by model performance metrics (cross-validation accuracy, precision, recall, F1 score, and ROC-AUC) and by empirical screening outcomes, including reference-standard coverage, annotation efficiency, and comparative performance against conventional libraries. This evaluation strategy emphasizes transferability and practical applicability, rather than re-validation of the underlying methodological framework (Supplemental Fig. S1).

To mitigate overfitting in the high-dimensional descriptor space, the ANN architecture relied on activation-induced sparsity and dropout regularization rather than explicit feature elimination or extensive hyperparameter optimization. This design follows a previously validated strategy in which weakly informative descriptors are implicitly suppressed through network activation behavior, while interpretability is achieved via post hoc contribution ranking. The justification for the three-hidden-layer feed-forward ANN architecture has been thoroughly addressed in our previous studies[15,18]. The characteristic fragment prediction model (first stage) was trained as a three-hidden-layer feed-forward ANN, with the first hidden layer explicitly designed to include twelve nodes representing empirical parameters derived from atmospheric chemistry, including vapor pressure (P), Henry's law constant (H), octanol-air partition coefficient (log K_{oa}), octanol-water partition coefficient (log K_{ow}), water solubility (log S), molecular dipole moment (Di), ionization state (pK_a), molecular weight (Mw), molar refractivity, molecular flexibility, sp^2 carbon proportion, and organic carbon partition coefficient (log K_{oc}). The second and third hidden layers corresponded to molecular and chemical descriptors, respectively, with each node representing a single descriptor obtained from the OPERA and TEST databases. The ANN-predicted characteristic fragments were subsequently used as inputs for the diagnostic ion prediction model (second stage), which was implemented using the LC-MS/MS compatible version of the CCNN model to map these fragments to experimentally observable ions[15]. A mass spectral database was subsequently established, and MOL files were batch-generated using the RDKit library, resulting in a diagnostic ion library specific to indoor air contaminants.

Sample preparation

Paired samples comprising 14 pet hair samples and 10 indoor dust samples were collected from ten households located in Jiangsu, Chongqing, Guangxi, Fujian, and Guangdong Provinces in South China between November and December 2021. Detailed descriptions of sampling sites, pet species, and sample pretreatment procedures are available in our previous study[7]. Pet hair samples (approximately 2 g each, 1–4 cm in length) were trimmed near the abdominal skin of the pets using stainless-steel scissors[9], ensuring sufficient sample mass and representative cumulative exposure while minimizing the influence of transient or episodic contamination[19,20]. Indoor dust samples were collected with a cleaned vacuum cleaner. At each household, approximately 20 g of dust particles were obtained after removing large debris, such as food residues and coarse hair, by sieving through an 80-mesh stainless-steel screen. Chemical extraction was performed using an Accelerated Solvent Extraction (ASE)-350 system (Thermo Fisher Scientific). After the addition of surrogate standards, approximately 0.2 g of hair or 1 g of dust sample was mixed thoroughly with diatomaceous earth and loaded into stainless-steel extraction cells. Extraction was conducted at 1500 psi and 100 °C with a solvent mixture of hexane, dichloromethane, and acetone (2:2:1, v/v/v) under two static cycles of 5 min each. Following ASE, the residual solids were removed from the extraction cells and subjected to a second extraction with 30 mL of acetonitrile by sonication for 30 min to recover hydrophilic contaminants. The extracts were centrifuged, and the supernatant was decanted. Fresh acetonitrile was added, and the sonication was repeated twice. Extracts from the ASE and sonication steps were combined, concentrated, and solvent exchanged to hexane. The combined extracts were purified using glass columns packed with 4 cm of silica gel (deactivated with 3% water, w/w) and 1 cm of anhydrous Na_2SO_4. Due to the limited sample mass and low contaminant concentration in individual samples, eluates from 14 pet hair and 10 dust samples were pooled separately and were concentrated to 0.5 mL, filtered through a 0.22 μm PTFE membrane, and analyzed by LC-HRMS.

https://doi.org/10.66178/aie-0026-0006

Chemical analyses

Chemical analyses of pet hair and indoor dust samples were carried out using an Orbitrap Exploris 240 LC-HRMS operated in full scan mode and data independent acquisition (DIA) modes. Data were acquired in both positive and negative electrospray ionization (ESI^+/ ESI^-) modes. Chromatographic separation was achieved on a Waters ACQUITY UPLC HSS C18 column (2.1 × 100 mm, 1.8 μm) maintained at 40 °C, with a sample injection volume of 10 μL. For ESI^+ mode, the mobile phase consisted of 0.1% formic acid in water (A) and methanol (B), whereas in ESI^- mode, 2 mM ammonium formate in water (A) and methanol (B) were used. The flow rate was maintained at 0.15 mL/min. The gradient elution program was optimized based on over 150 representative compounds to ensure broad chemical coverage. For DIA acquisition, the MS^1 and MS^2 data were collected across the *m/z* range of 100–1,000. Full-scan MS^1 data was acquired at a mass resolution of 60,000, and MS^2 data were obtained at 15,000. The DIA method employed sequential isolation windows of *m/z* 200–300, 300–400, 400–500, 500–600, and 600–700. Additional instrumental parameters for HRMS-based NTS are provided in the Supporting Information.

Data processing was performed using Compound Discoverer 3.3 with the molecular ion libraries and diagnostic ion libraries integrated into the workflow. The workflow was optimized to automatically activate both library searches and apply an appropriate scoring algorithm. Detailed data-processing procedures and parameter settings are described in the Supporting Information and Supplemental Fig. S2. The major analytical steps included feature extraction and feature annotation. Feature extraction was performed to identify chromatographic peaks from LC-HRMS raw data and to extract associated parameters such as *m/z*, intensity, and retention time. Feature annotation involved molecular formula prediction and MS/MS spectra matching against mzCloud, the molecular ion library, and the diagnostic ion library. Following annotation, suspect compounds were manually filtered based on the following criteria: (i) high-quality peak shapes as evaluated by peak rating and manual inspection, and (ii) full MS spectrum matching with mzCloud or either of the ion libraries.

Chemical space characterization

Chemical identifiers and physicochemical properties of the detected contaminants, including CASRN, InChIKey, SMILES, refractive index, LogK_{ow}, and molar refractivity were retrieved from the CCD. Subsequently, SMILES strings of all compounds were batch-submitted to the SwissADME platform to predict aqueous solubility (LogS), lipophilicity (LogP), and topological polar surface area (TPSA). To compare the differences in functional groups between the two environmental matrices, molecular structure processing and functional group annotation were performed using RDKit and the Pandas library.

To further characterize the chemical space and elucidate parameters governing contaminant accumulation in different matrices, a feature selection based on ANN (as described in 2.3) was used. The input layer consisted of over 800 molecular descriptors derived from the database. The ANN was trained to predict pollutant accumulation behavior using a feed-forward architecture with three hidden layers and ReLU activation. The model output $\hat{y}$ was expressed as:

$$\hat{y} = f(W_3\sigma(W_2\sigma(W_1x + b_1) + b_2) + b_3)$$

where x denotes the descriptor vector, W_i and b_i represent the weight matrices and bias vectors of layer i, and $\sigma(\cdot)$ is the ReLU activation function. Model optimization was performed via backpropagation to minimize the mean squared error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

Feature importance was quantified using cumulative contribution ranking, and top 64 descriptors at the cumulative plateau (> 95% variance explained) were retained for further analysis.

To interpret the ANN output, Shapley additive explanations (SHAP) were applied to assess the contribution of each descriptor to the prediction outcome. The SHAP value for feature j was calculated as:

$$\Phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!}\left(f_{S\cup\{j\}}\left(x_{S\cup\{j\}}\right) - f_S\left(x_S\right)\right)$$

where F represents the set of all features, S is any subset excluding feature j, and $f_S(\cdot)$ is the model output using the subset S. The resulting SHAP values quantified both magnitude and direction for each descriptor's impact on the predicted accumulation, revealing which parameters were the dominant predictors.

Quality assurance and quality control

For the text-mining dataset, uncertainty analysis was conducted to ensure data reliability. Data robustness and reproducibility were assessed using a meta-analysis framework. Corpus variables were converted to binary form, odds ratios were calculated using a random Mantel–Haenszel model, effect sizes were expressed as Hedges' g, and heterogeneity was evaluated using an instrumental variable model, with $I^2 > 50\%$ considered significant. A sensitivity analysis was then conducted to remove heterogeneous samples.

The accuracy and precision of the text-mining results were validated using bootstrapping and manual inspection. Based on a manual verification of 100 randomly selected literature entries, the NER accuracy exceeded 99.9%, confirming the robustness of chemical identification in the corpus.

For chemical analysis, each batch included one procedural blank, one matrix blank (dust or pet hair), and one matrix spike with a duplicate. Surrogate standards were added to all samples before extraction to assess sample preparation efficiency. Instrument stability was monitored using a mid-level calibration standard after every ten injections, and deviations below 20% were considered acceptable. Recoveries of surrogate standards ranged from 48.5% to 115%[7], indicating satisfactory analytical reproducibility and accuracy.

Results and discussion

Historical overview of indoor air contaminants

A total of 16,692 publications related to indoor environments were compiled through large-scale text mining, from which 3,684 distinct chemical substances were identified (Fig. 2a and Supplementary Sheet 01). Among these, 11,399 studies focused on indoor dust, 1,818 investigated pet hair, and 1,357 examined both matrices, while the remaining studies addressed other environmental media. The results revealed a pronounced research bias accompanied by substantial compositional convergence. Dust accounted for approximately 68% of all studies, whereas pet-hair investigations represented less than 10%; however, nearly two-thirds of the identified compounds were shared between the two matrices. This finding suggested that contaminants detected in hair largely originate from the same indoor sources that govern dust composition. Such overlap indicated that most previous hair-related studies did not treat pet hair as an independent exposure matrix, but rather as a component of dust, reinforcing the dominant role of dust as the principal contaminant reservoir in indoor settings[21]. By contrast, the present analysis extended literature coverage to the broader indoor domain and explicitly distinguished between dust and

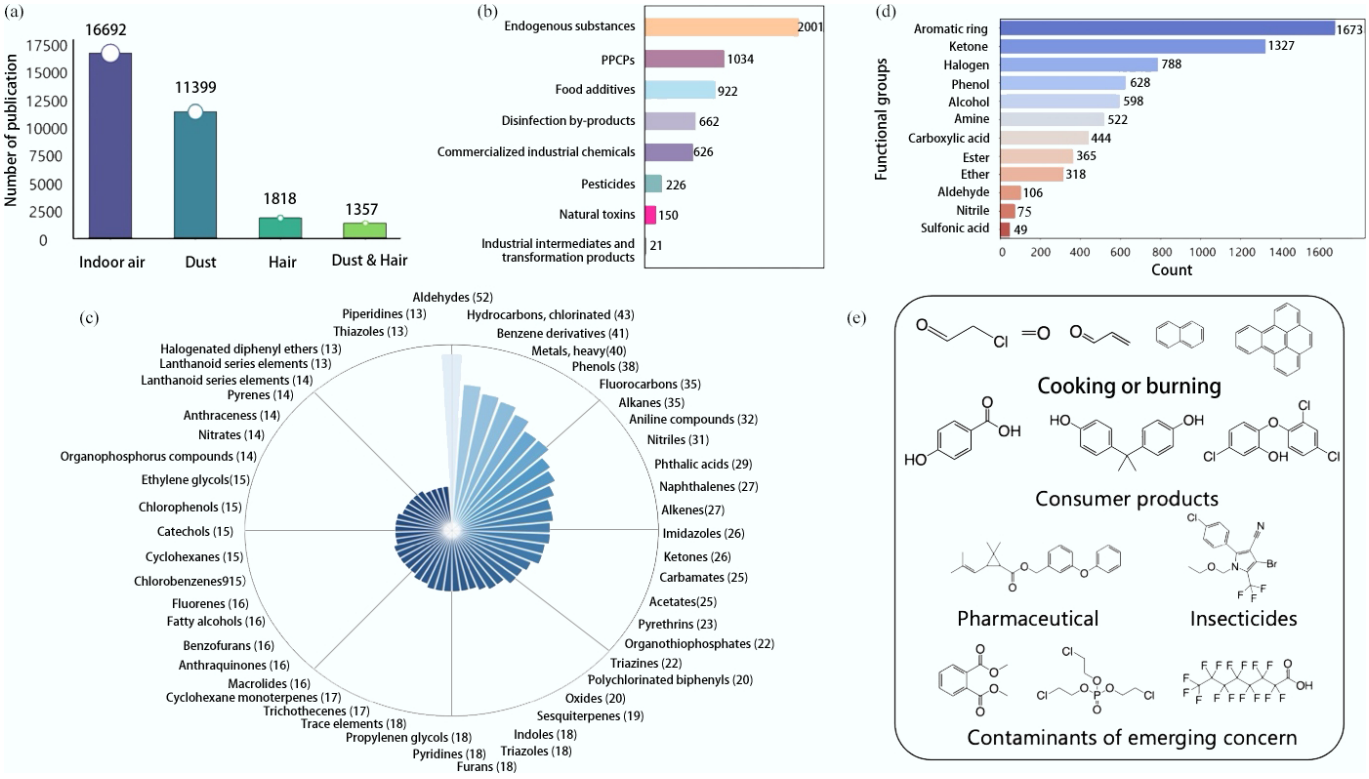


Fig. 2 Historical landscape of indoor contaminants revealed by large-scale text mining. (**a**) Number of publications associated with indoor air, dust, and pet hair matrices, illustrating the overall research distribution and matrix overlap; (**b**) NORMAN Network strategy-based categories; (**c**) Top 5% MeSH categories; (**d**) Major functional groups identified across all reported compounds; (**e**) Representative molecular structures and emission sources.

pet hair, providing the first large-scale overview of both chemical composition and research focus across these matrices.

The compiled dataset was categorized into functional use groups to visualize the dominant chemical domains under the NORMAN Network strategy (Fig. 2b). As shown in Fig. 2b, the constructed spectral library captures not only legacy compounds but also emerging contaminants that have recently gained attention, such as industrial intermediates and transformation products (e.g., 2-iodoacetamide), per- and polyfluoroalkyl substances (PFASs), flame retardants (e.g., triphenyl phosphate, tris (2-butoxyethyl) phosphate), and phthalates (e.g., dibutyl phthalate, diethyl phthalate). Endogenous substances and pharmaceuticals and personal-care products (PPCPs) were the most frequently reported categories, followed by food additives, disinfection by-products, and commercialized industrial chemicals. Pesticides[22], natural toxins, and industrial intermediates appeared less frequently but still appreciably contributed to the overall chemical diversity. A radial bar chart (Fig. 2c) further summarized the top 5% MeSH categories, highlighting aldehydes (52), chlorinated hydrocarbons (43), benzene derivatives (41), heavy metals (40), and phenols (38) as the most intensively studied pollutants. These compounds were representative of both volatile and semi-volatile organic compounds (VOCs/SVOCs) typically associated with indoor activities and materials[23]. To capture structural features, the major functional groups identified across the full contaminant inventory are shown in Fig. 2d. Aromatic rings, ketones, halogens, phenols, and alcohols dominated, reflecting the prevalence of aromatic and oxygenated species in indoor environments. Such structural traits correspond to compounds with moderate volatility and high persistence, favoring accumulation in indoor solid reservoirs such as dust and pet hair[11].

Representative molecular structures and their probable emission sources are illustrated in Fig. 2e, encompassing compounds originating from cooking or combustion processes, consumer products, pharmaceuticals, insecticides, and contaminants of emerging concern. Aldehydes and chlorinated hydrocarbons (such as chloroacetaldehyde) are typical indoor pollutants produced during cooking or cleaning activities, and their concentrations are often higher indoors than outdoors[24,25]. Formaldehyde released from cooking has been identified as carcinogenic to humans[24], while exposure to acrolein has been associated with renal impairment and other adverse effects[26]. Aromatic hydrocarbons such as benzene and naphthalene represent additional high-risk entities linked to carcinogenic outcomes[27]. Consumer products-related chemicals were equally prominent. Phenolic compounds including parabens, bisphenol A, and triclosan occur ubiquitously due to their widespread use in personal-care and household products[28]. Heavy metals such as lead and cadmium, commonly found in indoor dust, rank among the most toxic micropollutants and are associated with neurotoxicity and nephrotoxicity[29]. Other VOCs/SVOCs, such as phenothrin, chlorfenapyr, and benzo(e)pyrene, originate from building materials, insecticides, and combustion emissions[30]. Furthermore, a broad spectrum of emerging contaminants, including phthalates[31], organophosphate esters[32], flame retardants, and PFASs[33], has been increasingly reported. These compounds were characterized by high stability, hydrophobicity, and bioaccumulative potential, underscoring a historical shift in indoor-pollution research from short-lived volatile organics toward semi-volatile and persistent pollutants those are expected to remain entrapped and enriched within indoor matrices such as dust and pet hair.

https://doi.org/10.66178/aie-0026-0006

Scenario-specific mass spectral libraries screening

Recently, HRMS features have been defined as ions characterized by a specific *m/z* value at a given retention time (RT)[34,35] and have been widely used to characterize the complexity of chemical mixtures[36]. After quality control and field-blank correction, the combined analyses of dust and pet hair samples revealed 143,736 molecular features, highlighting the enormous data volume and the challenge of direct interpretation. The total ion chromatograms showed dense and widely distributed signals, consistent with broad chemical coverage (Fig. 3a). CH_2-based Kendrick mass defect (KMD) analysis (Fig. 3b) further revealed several homologous series, indicating the presence of structurally related aliphatic compounds[37]. However, the wide dispersion of data points beyond these series demonstrates considerable molecular diversity, suggesting contributions from multiple chemical sources and transformation processes. This coexistence of homologous patterns and highly heterogeneous signals[38] highlights the limited utility of conventional libraries and underscores the need for scenario-specific spectral databases to enable more meaningful feature annotation.

The robustness of the developed libraries was evaluated by benchmarking reference standard identification, compound annotation coverage, and data-processing efficiency against mzCloud. In dust samples, 11 reference standards were confirmed, of which only 27% (3/11) were matched by mzCloud, 64% (7/11) were identified using the molecular ion database, 18% (2/11) were confirmed by both the molecular and diagnostic ion libraries, and 9% (1/11) were uniquely identified by the diagnostic ion library (Fig. 3c and Supplementary Sheet 03–04). In pet hair samples, 8 reference standards were identified, with 25% (2/8) matched by mzCloud, 62.5% (5/8) additionally identified via the molecular ion library, 12.5% (1/8) by the diagnostic ion library, and 25% (2/8) confirmed by both molecular and diagnostic ion libraries (Fig. 3c and Supplementary Sheet 03–04). Collectively, these results demonstrated that the scenario-specific library substantially improved the coverage of reference standards identification compared to conventional library matching[39].

Incorporation of the scenario-specific library increased compound annotations by approximately 2.2–28-fold (+116% to +2,700%) across dust and hair samples, substantially enhancing chemical detection coverage (Fig. 3d). In dust samples, the number of identified compounds increased from 85 to 184 (+116%) for commercialized industrial chemicals, from 30 to 100 (+233%) for PPCPs, from 11 to 38 (+245%) for industrial intermediates and transformation products, from 13 to 36 (+177%) for natural toxins, and from 2 to 23 (+1050%) for pesticides. Hair samples showed similar trends, with commercialized industrial chemicals increasing from 64 to 180 (2.8-fold), pharmaceuticals and personal care products from 29 to 108 (3.7-fold), industrial intermediates and transformation products from 12 to 37 (3.1-fold), natural toxins from 11 to 30 (2.7-fold), and pesticides from 1 to 28 (28-fold). By contrast, minimal changes were observed for categories such as disinfection by-products. These results highlight that the scenario-specific library preferentially enhanced annotation of compound classes that are underrepresented in conventional databases, particularly those of regulatory or exposure relevance, highlighting the value of matrix- and scenario-tailored spectral resources.

Processing efficiency was also substantially improved by applying the scenario-specific library. Using the same raw NTS dataset, a common Compound Discoverer process required 12 h 30 min to complete the mzCloud-based workflow, largely due to network-dependent spectral matching and the high computational load associated with full-fragmentation searching. In contrast, analysis using the in-house molecular ion library was completed within 16 min under identical conditions, as the search was restricted to precursor ions and executed locally without network latency. This ~47-fold reduction in processing time demonstrates that a scenario-specific molecular ion library not only enhances annotation coverage but also enables rapid and scalable data interpretation in large NTS datasets. Collectively, these results highlight that scenario-specific databases can provide both robust and highly efficient support for NTS in complex indoor matrices.

Contaminants accumulated in pet hair

Non-target screening showed that pet hair and indoor dust contained highly comparable numbers of annotated compounds, with more than 50% of compounds shared between the two matrices. This is similar to our previous finding using target analysis in GC-MS[21]. Functional use analysis (Fig. 4a) showed similar profiles for both matrices, dominated by commercial and industrial chemicals[21], and PPCPs[40]. Among the detected chemicals, organophosphate flame retardants (OPFRs) warranted particular attention. OPFRs[41], including triphenyl phosphate, tris (2-butoxyethyl) phosphate, and dibutyl phosphate, were detected in dust and were also reported in pet hair[42]. These OPFRs are of particular concern due to documented toxicological effects, including carcinogenicity[43], endocrine disruption[44], neurotoxicity[45], and reproductive toxicity[46]. Meanwhile, PPCPs such as ibuprofen, propacetamol, gabapentin, propylparaben, DEET, and octinoxate were frequently detected[47,48]. Several of these compounds, notably propylparaben and DEET, have been linked to endocrine[49,50] and neurological effects[51,52], highlighting potential health risks associated with chronic indoor exposure. Additional detected groups included industrial intermediates, transformation products, natural toxins, food additives, and pesticides[22]. The frequent presence of industrial and biological transformation products in both matrices, which are not typically anticipated in conventional indoor exposure studies, suggests overlooked exposure pathways and complex indoor chemical dynamics. Structural category analysis (Fig. 4b) revealed high similarity between the two matrices, with aromatic rings, ketones, carboxylic acids, alcohols, and amines as the predominant functional groups. These structures are commonly associated with surfactants, fragrances, fatty acid derivatives, plastic additives, and personal care products.

To identify molecular determinants of contaminant accumulation, ANN modeling combined with SHAP values indicated that no single descriptor dominated model performance (Supplemental Fig. S3); instead, accumulation was governed by a combined contribution of multiple molecular properties. Key contributors included hydrophobicity (log P), ionization-related parameters (pK_a), aromaticity, molecular flexibility, polarizability, and bioactivity-related indices[53]. Collectively, these properties favor chemical partitioning on organic-rich and biomolecular surfaces, providing a mechanistic explanation for the pronounced compositional similarity observed between indoor dust and pet hair (Supplemental Fig. S4).

According to the ANN and SHAP results, six representative physicochemical descriptors were selected to further characterize NTS-annotated compounds in dust and pet hair, representing complementary dimensions of solubility, hydrophobicity, polarity, and molecular polarizability. These included molar refraction[54], log S, refractive index, log K_{ow}, log P, and topological polar surface area (TPSA). As shown in Fig. 4c, molar refraction, a proxy for molecular polarizability and volume, exhibited a strong negative correlation with aqueous solubility (log S) across the indoor chemical library, dust, and hair datasets. The near-overlapping distributions of these

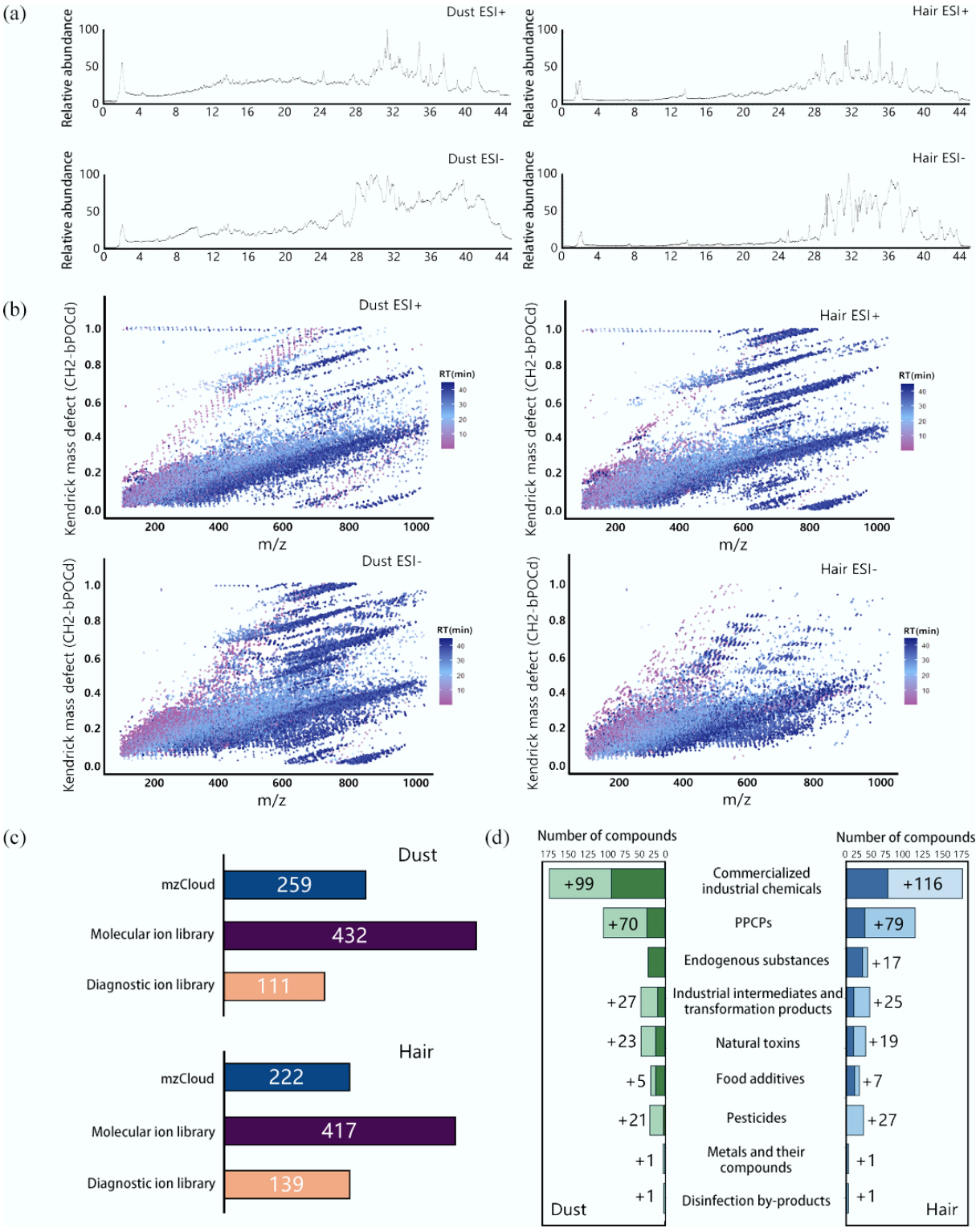


Fig. 3 (**a**) Total ion chromatograms (TIC) of dust and pet hair samples analyzed by LC-HRMS. (**b**) Scatter plot (Kendrick plot) of non-target screening of dust and pet hair samples using LC-HRMS in both positive (ESI^+) and negative (ESI^-) electrospray ionization modes (DIA). The x-axis represents *m/z*, the y-axis represents Kendrick mass defect, and color shading indicates retention time. (**c**) Bar chart showing the number of compounds identified only via mzCloud or only via the scenario-based database. (**d**) Bar chart showing the incremental number of compounds in different chemical usage categories obtained by integrating the scenario-based database with mzCloud.

parameters across all datasets, including text-mined literature-available indoor contaminants, indicate consistent physicochemical selection pressures across indoor compartments. Compounds with high molar refraction and low solubility, typically hydrophobic or aromatic species with larger molecular volumes, were preferentially enriched in both dust and hair. Hydrophobicity-related descriptors

https://doi.org/10.66178/aie-0026-0006

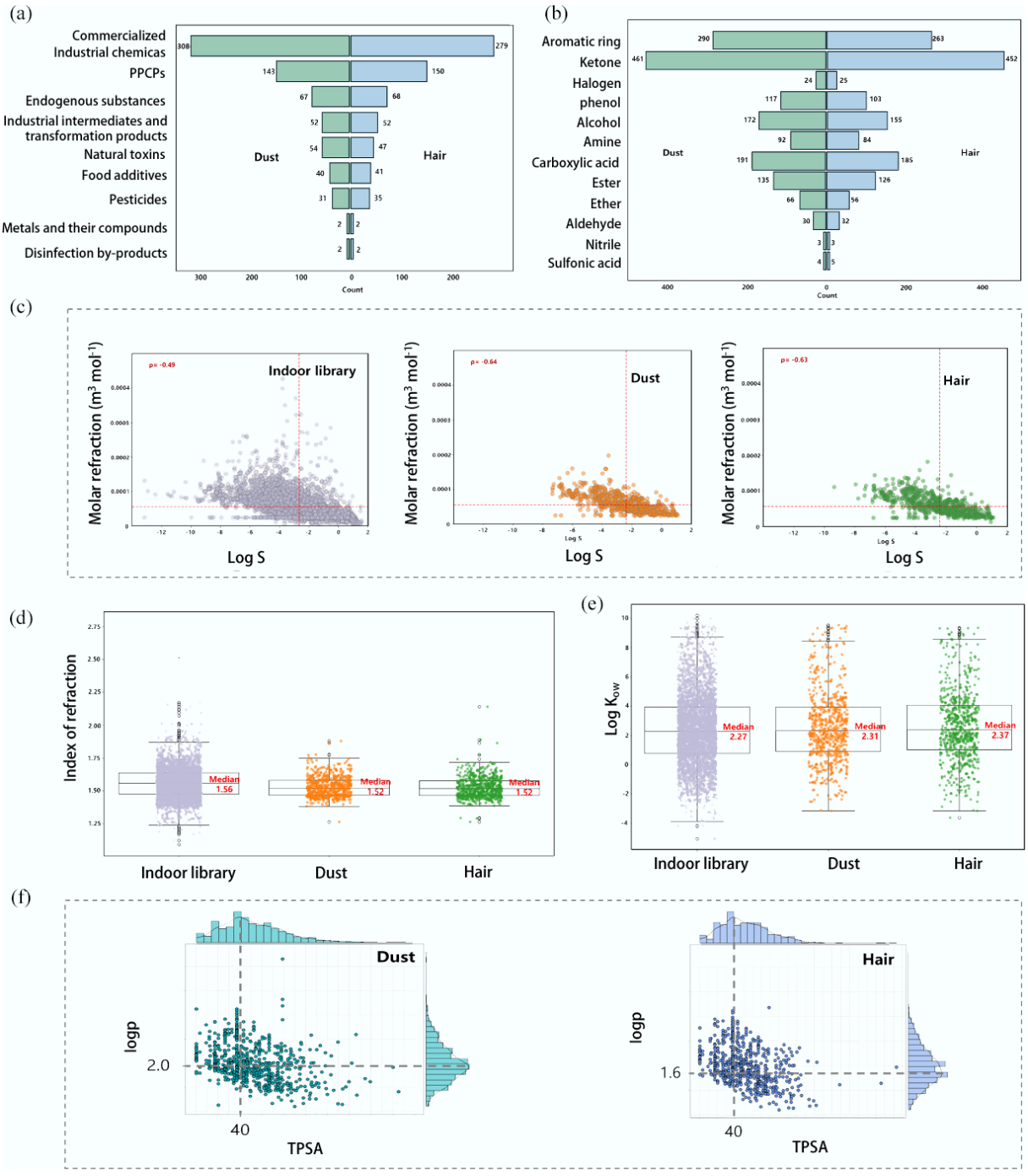


Fig. 4 Physicochemical property analysis of NTS-annotated pollutants in dust and pet hair. (**a**) Usage-based classification of detected pollutants. (**b**) Functional group-based classification. (**c**) Correlation between molar refraction and aqueous solubility (log S) for the indoor library, dust, and hair datasets. (**d**) Distribution of the index of refraction in the indoor library, dust, and hair samples. (**e**) Distribution of log K_{ow} values. (**f**) Relationship between log P and topological polar surface area (TPSA) in dust and hair samples.

further supported this trend. The refractive index of identified compounds clustered between 1.52 and 1.56 (Fig. 4d), a characteristic of moderately polarizable SVOCs, while the log K_{ow} distribution centered at approximately 2.3, indicating moderate hydrophobicity. These properties favor sorption to organic-rich dust and keratinous hair surfaces while limiting aqueous dissolution. Analysis of polarity constraints revealed that most compounds occupied a narrow window defined by a log P of approximately 2 and a TPSA below 60 Å^2 (Fig. 4f), consistent with neutral, weakly polar molecules with limited hydrogen-bonding capacity. Such compounds possess sufficient lipophilicity to adsorb onto dust-associated organic carbon and hair keratin, while retaining enough polarity to avoid rapid volatilization. Collectively, these results demonstrate that both dust and pet hair selectively retain compounds that fall within a shared physicochemical window: moderately hydrophobic, weakly polar, low-solubility molecules which are often characterized by aromatic or aliphatic structures. This close agreement between ANN-identified key descriptors and empirical distributions confirms that

contaminant accumulation in both matrices is governed by physiochemically driven partitioning and sorption processes.

Future Perspectives and Implications

In the present study, we developed an integrated data-to-laboratory framework combining large-scale text mining, machine learning, and HRMS-based NTS to investigate indoor air contaminants in pet hair and house dust. By constructing scenario-specific molecular and diagnostic ion libraries, our approach enabled comprehensive NTS and substantially enhanced chemical annotation coverage compared to conventional spectral databases. The results revealed a significant overlap in the chemical space of contaminants between dust and pet hair, highlighting that pet hair functions as a complementary exposure matrix reflecting indoor chemical exposures rather than a distinct contaminant reservoir. Functional and structural analyses showed that both matrices predominantly retain PPCPs and industrial chemicals characterized by moderate hydrophobicity, low aqueous solubility, and aromatic or aliphatic frameworks. ANN modeling identified key physicochemical descriptors such as log P, TPSA, and molecular flexibility as primary determinants governing contaminant accumulation, providing mechanistic insights into the selective enrichment processes in indoor matrices.

Despite these advances, several limitations warrant consideration. Owing to the limited sample size, data from 14 pet hair samples and 10 house dust samples were pooled for analysis. However, this approach may have masked inter-individual variability related to household characteristics, pet behavior, and temporal exposure dynamics. Consequently, the present findings primarily reflect scenario-directed exposure patterns rather than individual-specific profiles. Future studies should expand sample sizes and incorporate time-resolved sampling designs to better resolve individual and temporal variability. Beyond dust and pet hair, the developed framework is readily extendable to chemically complex matrices characterized by partial prior knowledge yet persistent unknown contaminants, including silicone wristband-based exposure monitoring scenarios for personal exposure assessment and environmental systems such as wastewater treatment plants, where numerous contaminants remain poorly characterized.

Ethical statements

Pet hair and indoor dust samples were collected from ten households after informed consent was obtained from all pet owners (residents) prior to sample collection. Pet hair was trimmed non-invasively near the abdominal skin by the pet owners themselves, without causing pain or distress to the animals; no experimental interventions were performed on the animals. This non-invasive sampling procedure falls below the threshold specified in Directive 2010/63/EU of the European Parliament and of the Council and is therefore outside the scope of this Directive. Therefore, formal ethical approval was not required.

Author contributions

The authors confirm their contributions to the paper as follows: Liwei He: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. Fei Cheng: Writing – review & editing, Visualization, Validation, Conceptualization, Funding acquisition, Formal analysis. Jingwen Shi: Methodology, Investigation. Huizhen Li: Writing – review & editing, Conceptualization, Investigation. Zhiqiang Yu: Writing – review & editing, Supervision, Project administration, Funding acquisition, Validation. Jing You: Writing – review & editing, Conceptualization, Supervision, Project administration, Funding acquisition.

All authors disclosed no relevant relationships.

Data availability

The data and code supporting the findings of this study are available from the corresponding author on reasonable request.

Funding

This work was financially supported by the National Natural Science Foundation of China (42577301, 42377270, 42321003), Guangdong Major Project of Basic and Applied Basic Research (2023B0303000007), Guangdong Basic and Applied Basic Research Foundation (2025A1515010960), China Postdoctoral Science Foundation (2020M68392), and the Innovative Research Team of Department of Education of Guangdong Province (2020KCXTD005). The authors thank Thermo Fisher Scientific for Instrumental support.

Declarations

Competing interests

The authors declare that they have no conflict of interest.

Supplementary information

The Supporting Information is available at: https://doi.org/10.66178/aie-0026-0006.

Text detailing chemicals and reagents, data collection and curation, diagnostic ion prediction by artificial neural network, instrumental analysis, and code for calculating chemical functional groups. Figures showing HRMS screening workflow and ANN parameter selection results;

Supplementary excel sheets with text-mined chemicals, a molecular ion library, nontarget compounds identified in dust and hair samples.

Author details

[1]State Key Laboratory of Advanced Environmental Technology, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China; [2]Guangdong Provincial Key Laboratory of Environmental Pollution and Health, College of Environment and Climate, Jinan University, Guangzhou 511443, China; [3]University of the Chinese Academy of Sciences, Beijing 100049, China

References

[1] AVMA. 2013. *U. S. pet ownership & demographics sourcebook*. Schaumburg, IL: American Veterinary Medical Association. 42 pp. https://www.avma.org/javma-news/2013-02-01/vital-statistics

[2] Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, et al. 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science & Environmental Epidemiology* 11:231–252

[3] Sheu R, Stönner C, Ditto JC, Klüpfel T, Williams J, Gentner DR. 2020. Human transport of thirdhand tobacco smoke: A prominent source of hazardous air pollutants into indoor nonsmoking environments. *Science Advances* 6:eaay4109

[4] Weschler CJ, Hodgson AT, Wooley JD. 1992. Indoor chemistry: ozone, volatile organic compounds, and carpets. *Environmental Science & Technology* 26:2371–2377

[5] Klein F, Baltensperger U, Prévôt ASH, El Haddad I. 2019. Quantification of the impact of cooking processes on indoor concentrations of

volatile organic species and primary and secondary organic aerosols. *Indoor Air* 29:926–942
[6] Mizukawa H, Nomiyama K, Nakatsu S, Yamamoto M, Ishizuka M, et al. 2017. Anthropogenic and naturally produced brominated phenols in pet blood and pet food in Japan. *Environmental Science & Technology* 51:11354–11362
[7] Zhang Y, Tong Y, Cheng F, Shi J, Huang J, et al. 2023. Occurrence of emerging contaminants in pet hair and indoor air: integrative health risk assessment using multiple ToxCast endpoints. *Environmental Science. Processes & Impacts* 25:1839–1849
[8] Song Z, Shi M, Ren X, Wang L, Wu Y, et al. 2023. An integrated non-targeted and targeted analysis approach for identification of semi-volatile organic compounds in indoor dust. *Journal of Hazardous Materials* 459:132202
[9] Yang Y, Wang Y, Tan F, Zhang Z, Rodgers TFM, Chen J. 2020. Pet hair as a potential sentinel of human exposure: Investigating partitioning and exposures from OPEs and PAHs in indoor dust, air, and pet hair from China. *Science of The Total Environment* 745:140934
[10] Bukowski JA, Wartenberg D. 1997. An alternative approach for investigating the carcinogenicity of indoor air pollution: pets as sentinels of environmental cancer risk. *Environmental Health Perspectives* 105:1312–1319
[11] Hegedus C, Andronie L, Uiuiu P, Jurco E, Lazar EA, Popescu S. 2023. Pets, genuine tools of environmental pollutant detection. *Animals* 13:2923
[12] Ruan T, Li P, Wang H, Li T, Jiang G. 2023. Identification and prioritization of environmental organic pollutants: from an analytical and toxicological perspective. *Chemical Reviews* 123:10584–10640
[13] Hollender J, Schymanski EL, Ahrens L, Alygizakis N, Béen F, et al. 2023. NORMAN guidance on suspect and non-target screening in environmental monitoring. *Environmental Sciences Europe* 35:75
[14] Zhu L, Bossi R, Carvalho PN, Rigét FF, Christensen JH, et al. 2024. Suspect and non-target screening of chemicals of emerging Arctic concern in biota, air and human serum. *Environmental Pollution* 360:124605
[15] Cheng F, Li H, Lou X, He L, Wu X, et al. 2025. Event-driven taxonomy (EDT) screening: Leveraging effect-based spectral libraries to accelerate semiquantitative nontarget analysis of AhR agonists in sediment in the era of big data. *Environmental Science & Technology* 59:14359–14371
[16] Liu P, Pan D, Jiao X-Y, Liu J-N, Du P-H, et al. 2025. Integrating non-target analysis and machine learning: a framework for contaminant source identification. *npj Clean Water* 8:78
[17] Huang J, Cheng F, He L, Lou X, Li H, You J. 2024. Effect driven prioritization of contaminants in wastewater treatment plants across China: A data mining-based toxicity screening approach. *Water Research* 264:122223
[18] Cheng F, Escher BI, Li H, König M, Tong Y, et al. 2024. Deep learning bridged bioactivity, structure, and GC-HRMS-readable evidence to decipher nontarget toxicants in sediments. *Environmental Science & Technology* 58:15415–15427
[19] Zhang S, Yan X, Tang B, Luo W, Chen S, et al. 2023. Human hair as a noninvasive matrix to assess exposure to micro-organic contaminants: State of the art review. *Science of The Total Environment* 892:164341
[20] Qiao L, Zheng XB, Zheng J, Chen SJ, Zhong CQ, et al. 2019. Legacy and currently used organic contaminants in human hair and hand wipes of female e-waste dismantling workers and workplace dust in south china. *Environmental Science & Technology* 53:2820–2829
[21] Mitro SD, Dodson RE, Singla V, Adamkiewicz G, Elmi AF, et al. 2016. Consumer product chemicals in indoor dust: A quantitative meta-analysis of U. S. Studies. *Environmental Science & Technology* 50:10661–10672
[22] Tran VV, Park D, Lee YC. 2020. Indoor air pollution, related human diseases, and recent trends in the control and improvement of Indoor air quality. *International Journal of Environmental Research and Public Health* 17:2927
[23] Mai J-L, Yang W-W, Zeng Y, Guan Y-F, Chen S-J. 2024. Volatile organic compounds (VOCs) in residential indoor air during interior finish period: Sources, variations, and health risks. *Hygiene and Environmental Health Advances* 9:100087
[24] Kang K, Kim T, Kim DD. 2023. An Investigation of concentration and health impacts of aldehydes associated with cooking in 29 residential buildings. *Indoor Air* 2023:2463386
[25] Hisham MWM, Grosjean D. 1991. Air pollution in southern California museums: Indoor and outdoor levels of nitrogen dioxide, peroxyacetyl nitrate, nitric acid, and chlorinated hydrocarbons. *Environmental Science & Technology* 25:857–862
[26] Chen H-M, Wu M-T. 2017. Residential exposure to chlorinated hydrocarbons from groundwater contamination and the impairment of renal function-An ecological study. *Scientific Reports* 7:40283
[27] Loh MM, Levy JI, Spengler JD, Houseman EA, Bennett DH. 2007. Ranking cancer risks of organic hazardous air pollutants in the United States. *Environmental Health Perspectives* 115:1160–1168
[28] Levasseur JL, Hammel SC, Hoffman K, Phillips AL, Zhang S, et al. 2021. Young children's exposure to phenols in the home: Associations between house dust, hand wipes, silicone wristbands, and urinary biomarkers. *Environmental International* 147:106317
[29] Roy A, Jha AK, Kumar A, Bhattacharya T, Chakraborty S, et al. 2024. Heavy metal pollution in indoor dust of residential, commercial, and industrial areas: a review of evolutionary trends. *Air Quality, Atmosphere & Health* 17:891–918
[30] Kumar P, Singh AB, Arora T, Singh S, Singh R. 2023. Critical review on emerging health effects associated with the indoor air quality and its sustainable management. *Science of The Total Environment* 872:162163
[31] Yang C, Harris SA, Jantunen LM, Kvasnicka J, Nguyen LV, Diamond ML. 2020. Phthalates: Relationships between Air, Dust, Electronic Devices, and Hands with Implications for Exposure. *Environmental Science & Technology* 54:8186–8197
[32] Balasch A, Moreno T, Eljarrat E. 2023. Assessment of Daily Exposure to Organophosphate Esters through PM2.5 Inhalation, Dust Ingestion, and Dermal Contact. *Environmental Science & Technology* 57:20669–20677
[33] Morales-McDevitt ME, Becanova J, Blum A, Bruton TA, Vojta S, et al. 2021. The Air That We Breathe: Neutral and Volatile PFAS in Indoor Air. *Environmental Science & Technology Letters* 8:897–902
[34] Black GP, He G, Denison MS, Young TM. 2021. Using estrogenic activity and nontargeted chemical analysis to identify contaminants in sewage sludge. *Environmental Science & Technology* 55:6729–6739
[35] Papazian S, D'Agostino LA, Sadiktsis I, Froment J, Bonnefille B, et al. 2022. Nontarget mass spectrometry and in silico molecular characterization of air pollution from the Indian subcontinent. *Communications Earth & Environment* 3:35
[36] Bonnefille B, Karlsson O, Rian MB, Raqib R, Parvez F, et al. 2023. Nontarget analysis of polluted surface waters in Bangladesh using open science workflows. *Environmental Science & Technology* 57:6808–6824
[37] Kune C, McCann A, Raphaël LR, Arias AA, Tiquet M, et al. 2019. Rapid visualization of chemically related compounds using kendrick mass defect as a filter in mass spectrometry imaging. *Analytical Chemistry* 91:13112–13118
[38] Kang G, Li J, Xue W, Zhou K, Jin Z, Lv H. 2025. Kendrick Mass Defect Analysis for Structural Characterization of Additives in Low-Viscosity Engine Oils: A Concise Review. *ACS Omega* 10:29950–29968
[39] Wong JW, Wang J, Chow W, Carlson R, Williams AJ, et al. 2025. Multilaboratory study of a nontarget data acquisition for target analysis (nDATA) workflow using ultrahigh-performance liquid chromatography-high-resolution mass spectrometry for the screening of 1087 pesticides in fresh fruits and vegetables. *Journal of Agricultural and Food Chemistry* 73:8632–8650
[40] Liu J-L, Wong M-H. 2013. Pharmaceuticals and personal care products (PPCPs): A review on environmental contamination in China. *Environment International* 59:208–224
[41] Yang J, Yao Y, Li X, He A, Chen S, et al. 2024. Nontarget identification of novel organophosphorus flame retardants and plasticizers in indoor air and dust from multiple microenvironments in China. *Environmental Science & Technology* 58:7986–7997
[42] Brits M, Brandsma SH, Rohwer ER, De Vos J, Weiss JM, de Boer J. 2019. Brominated and organophosphorus flame retardants in South African indoor dust and cat hair. *Environmental Pollution* 253:120–129

[43] Hoffman K, Sosa JA, Stapleton HM. 2017. Do flame retardant chemicals increase the risk for thyroid dysregulation and cancer? *Current Opinion in Oncology* 29:7–13

[44] Zhang G, Meng L, Guo J, Guan X, Liu M, et al. 2023. Exposure to novel brominated and organophosphate flame retardants and associations with type 2 diabetes in East China: A case-control study. *Science of The Total Environment* 871:162107

[45] Liu W, Luo D, Xia W, Tao Y, Wang L, et al. 2021. Prenatal exposure to halogenated, aryl, and alkyl organophosphate esters and child neurodevelopment at two years of age. *Journal of Hazardous Materials* 408:124856

[46] Luo D, Liu W, Wu W, Tao Y, Hu L, et al. 2021. Trimester-specific effects of maternal exposure to organophosphate flame retardants on offspring size at birth: A prospective cohort study in China. *Journal of Hazardous Materials* 406:124754

[47] Friedman A, Falakdin P, Govande A, Gommers G, Ryckebusch Y, et al. 2025. Pesticide residues in bedroom dust: occurrence, determinants, and health risk assessment. *Environmental Pollution* 387:127306

[48] Scott LN, Fiume M, Bergfeld WF, Belsito DV, Hill RA, et al. 2022. Safety assessment of panthenol, pantothenic acid, and derivatives as used in cosmetics. *International Journal of Toxicology* 41:77–128

[49] Nowak K, Ratajczak–Wrona W, Górska M, Jabłońska E. 2018. Parabens and their effects on the endocrine system. *Molecular and Cellular Endocrinology* 474:238–251

[50] Oates RP, Longley G, Hamlett P, Klein D. 2017. Pharmaceutical and Endocrine Disruptor Compounds in Surface and Wastewater in San Marcos, Texas. *Water Environment Research* 89:2021–2030

[51] Merola C, Caioni G, Bertolucci C, Lucon-Xiccato T, Savaşçı BB, et al. 2024. Embryonic and larval exposure to propylparaben induces developmental and long-term neurotoxicity in zebrafish model. *Science of The Total Environment* 912:168925

[52] Schoenig GP, Hartnagel RE Jr, Schardein JL, Vorhees CV. 1993. Neurotoxicity evaluation of N, N-Diethyl-m-toluamide (DEET) in rats1. *Fundamental and Applied Toxicology* 21:355–365

[53] Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. 2024. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science* 17:e70056

[54] Liu Y, Bé AG, Or VW, Alves MR, Grassian VH, Geiger FM. 2020. Challenges and opportunities in molecular-level indoor surface chemistry and physics. *Cell Reports Physical Science* 1:100256

Copyright: © 2026 by the author(s). Published by NEW PRESS LIMITED, TSIM SHA TSUI, HK, China. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0.