RESEARCH ARTICLE

Development and Internal Evaluation of an Indoor Building Exposure Index for Healthy Housing in a Monsoon-Affected Indian Tier-II City

Vrinda S Bhalerao[1*], Vishvanath N Kanthe[2], Pramod S Patil[3], Lalit B Pawar[4], Kamlesh P Bhagat[5]

[1]PhD, Assistant Professor, Department of Civil Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India*
[2]PhD, Associate Professor, Department of Civil Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India
[3]PhD, Assistant Professor, Department of Civil Engineering, R. C. Patel Institute of Technology, Shirpur, Dhule, Maharashtra, India
[4]PhD, Assistant Professor, Department of Civil Engineering, SNJB's LSKBJ College of Engineering, Chandwad, Maharashtra, India
[5]Assistant Professor, Department of Civil Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India

Corresponding Author: Vrinda S Bhalerao*

ABSTRACT
In indoor exposure screening, individual variables related to air quality or sources are commonly the primary focus, while building conditions are traditionally less often included in a comprehensive, integrated decision-making process. In the present framework-development study using 442 dwelling records in Nashik (India), the Indoor Building Exposure Index for Healthy Housing (IBEI-HH) was developed and internally evaluated. The analytic data set contains the original survey fields and new or improved building and sensor fields developed to test the workflow. Since these improved fields were not observed in the field, they are a proof of concept. Six domains of the psychological scale, namely moisture/dampness, ventilation/daylight deficiency, combustion, chemical use, dust/material reservoirs, and sensor-based exposure, were translated into a screening score ranging from 0 to 100. This development sample contained 35 low-risk, 366 moderate-risk, 41 high-risk, and no very high-risk Records. The range of agreement for the other weighting schemes was spread from 91.63% to 95.70%. The binary random forest had an ROC-AUC of 0.969 and a PR-AUC of 0.825, with a high recall (1 of 8) for the very high and high classes, highlighting that when discrimination between the classes is high, case detection is inadequate. The use of IBEI-HH is therefore an interpretable framework. It must be validated in the field, in addition to external validation, to be used as a health risk or operational screening tool.
Keywords: IBEI-HH; healthy housing; indoor exposure risk; dampness; ventilation deficiency; GIS hotspot mapping.
How to cite this article: Vrinda S Bhalerao, Vishvanath N Kanthe, Pramod S Patil, Lalit B Pawar, Kamlesh P Bhagat. Development and Internal Evaluation of an Indoor Building Exposure Index for Healthy Housing in a Monsoon-Affected Indian Tier-II City. International Journal of Drug Delivery Technology. 2026;16(80s):420-433. DOI: 10.25258/ijddt.16.80s.49
Source of support: Nil.
Conflict of interest: None

INTRODUCTION

Residential indoor environmental quality is important because it is where most people spend the majority of their time. Factors that can lead to poor indoor conditions include lack of ventilation, combustion processes, household chemicals (e.g., cigarette smoke), dust-retaining materials, high occupancy density, particulate matter (e.g., PM2.5 and PM10), CO2 and CO emissions, humidity, environmental temperature, and lack of daylight. However, the impact of physical building conditions tends to be understudied, and exposure is more often assessed using a single air-quality indicator or a single indoor activity performed by occupants. Exposure is, in reality, the consequence of interactions among building conditions, household practices, and environmental context. Dampness, wall leakage, external water accumulation, cracks, visible mould, lack of waterproofing, old pipes, and waterlogging can lead to increased moisture retention and may significantly contribute to the creation of unhealthy environments. Pollutant dilution and moisture flushing will also be reduced when ventilation in the kitchen or bathroom is poor, when there is no window opening, when there is no bathroom or kitchen exhaust system, or when the window opening is ineffective. These pathways are particularly pertinent in monsoon-prone cities, where precipitation, high humidity, leaks, and waterlogging are likely to increase indoor exposure above normal [1-3].

Housing in Tier II cities like Nashik is of varied types, ranging from flats to bungalows, chawls, old masonry-built houses, and mixed-construction houses. The ages of these buildings, the materials they are made of, their ventilation equipment, maintenance, sunlight exposure, dampness exposure, household practices, etc., vary. For this reason, an assessment focusing only on pollutants or only on behaviours could not capture the overall level of exposure in a dwelling. A building-centric framework is hence required that combines building defects, performance related to ventilation and daylighting, indoor/sensorial

**Author for Correspondence: Vrinda S Bhalerao*

sources, occupant density, and sensor measurements. This study develops IBEI-HH for the residences of Nashik by considering six domains that are normalized and serve as the basis for classifying risk, identifying the dominant pathways of exposure, and mapping provisional engineering interventions. The intervention logic, the dwelling characteristics, and the exposure typology are all connected to their contributions through a transparent workflow. There are currently no tools to challenge the findings of internal methodological evaluation, and further external field and health-outcome validation are needed [4-6].

LITERATURE REVIEW

Indoor exposure and building-performance pathways

There is growing focus in healthy housing research on the importance of indoor exposure, which is influenced by air quality, ventilation, moisture, materials, household practices, and livability. So, measuring only pollutants can give only a partial view of what it is like to live in a dwelling. Leaks, condensation, water infiltration, inadequate water removal, and ventilation are causes of moisture and mould. They are cases of building failures in which the persistence of moisture allows micro-organisms to flourish, producing odours and degrading building finishes, particularly in monsoon areas. Moist areas, cleaning agents, and cooking will help create moisture, and ventilation will help dilute pollutants. Openness, orientation, and passive control of the environment are also successfully demonstrated by daylight and sunlight.

Indoor sources, measurements, and integrated assessment

Indoor contaminants may be generated or trapped in buildings due to combustion, consumer chemicals, smoking, carpets and upholstery, pets, and high-occupancy populations. These effects can be very reliant on ventilation and building conditions and need to be assessed accordingly. Measurements of PM2.5, PM10, CO2, CO, TVOCs, relative humidity, temperature, and illuminance can provide direct evidence of the environment. Composite indices can then be used to convert multidimensional evidence into interpretable risk classes. Yet, there are many approaches that do not integrate building defects, ventilation, indoor sources, material reservoirs, measurements, and engineering responses into a single decision pathway [7-10].

Spatial interpretation and research gap

To maintain and identify neighbourhood clusters with respect to drainage, waterlogging, age of housing construction, etc., GIS analysis can be used. It can inform ward-specific targeting and ward-level healthy housing programs. Major gaps include the failure to integrate building-defect data with behavior- and sensor-based evidence, and the lack of data linking risk scores/dominant pathways to interventions for monsoon-impacted Tier-II

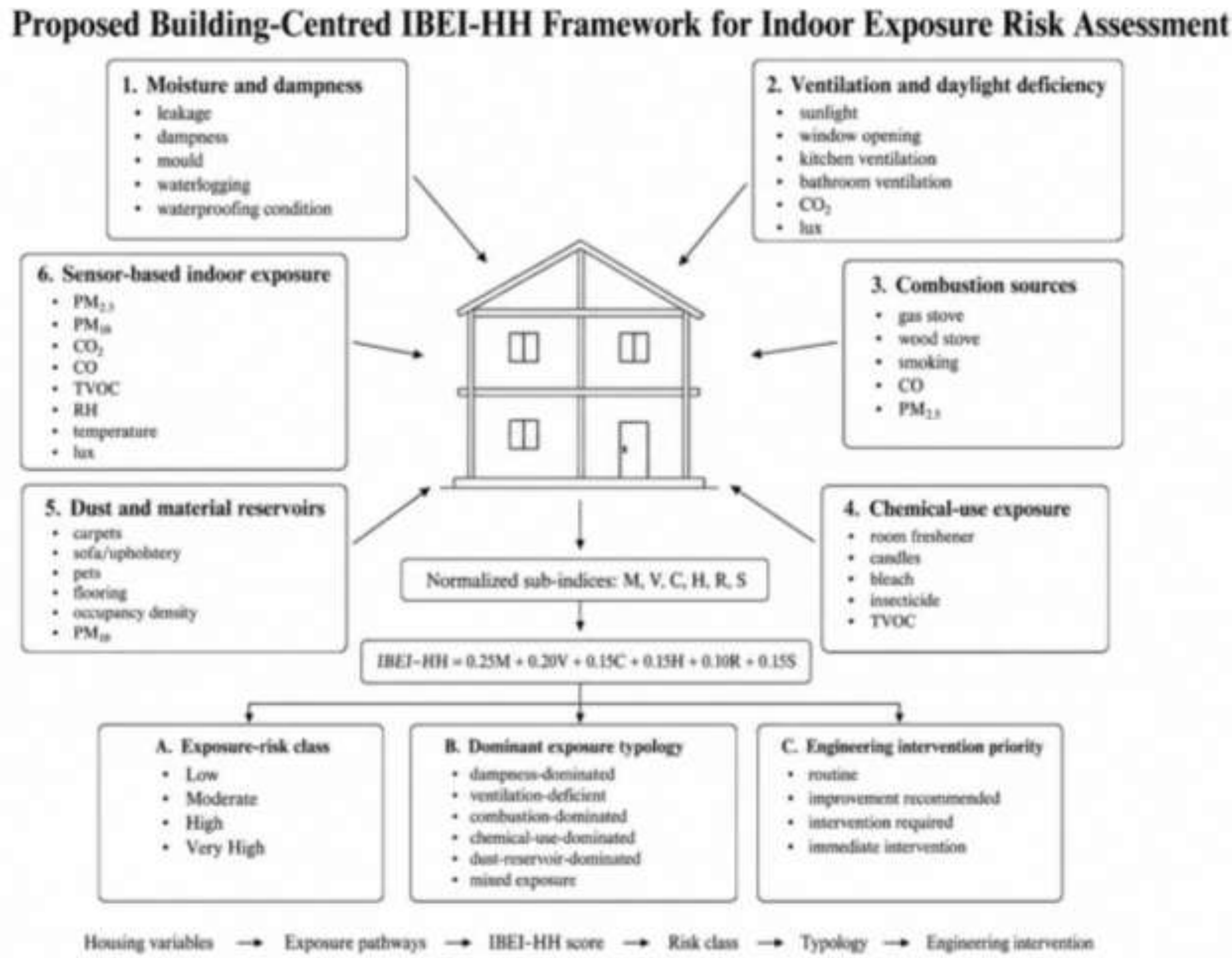

Figure 1. Conceptual workflow of the IBEI-HH framework

Table 1. Interpretation of the IBEI-HH risk classes

Risk class	Interpretation
Low	Lowest relative screening concern
Moderate	Review maintenance and ventilation conditions.
High	Prioritize field inspection and source review.
Very High	Immediate confirmatory inspection under the screening rule

Table 2. Dominant exposure components and corresponding typology labels

Dominant component	Typology
CM_i	Dampness-dominated
CV_i	Ventilation-deficient
CC_i	Combustion-dominated
CH_i	Chemical-use-dominated
CR_i	Dust-reservoir-dominated
CS_i	Sensor-elevated mixed exposure

cities in India. IBEI-HH aims to address these gaps by creating a holistic building–behavior–environment model. [11-14]

METHODOLOGY

Study design and evidentiary scope

This is a computational cross-sectional framework-development study based on 442 dwelling records. A source file with 119 raw variables and an analytical file with 181 variables, based on the source file (normalized indicators, sub-indices, classes, typologies, and intervention outputs), were constructed. The dates shown on all records are for the survey conducted on 20 May 2026. An estimated building and sensor field was added to the enhanced building and sensor fields that were not observed in the field; engineer verification has yet to be completed for all 442 building records. These fields enable evaluation of behaviors and the workflow's internal coherence, but their evaluation with respect to exposure prevalence, causation, medical risks, and valid intervention rules remains undecided. Anonymization of GPS coordinates. Available material doesn't record participant consent, how participants were recruited, the hardware or calibration of the sensors, where they were placed, and how long they were monitored before confirmatory field use [15-18].

Conceptual framework

The approach is to consider each dwelling as a coupled building–behavior–environment system. Leakage, dampness, cracks, waterproofing failures, and mould are classified as building conditions exposure-relevant conditions because they can lead to higher moisture retention, microbial growth, persistence of pollutants, odours, and wear of the surface – all of which are challenging for maintenance. Ventilation and daylight are also seen as means of pollutant dilution, moisture control, and passive IECD systems. As shown in Figure 1, indicators for dwellings are used to create component scores, risk classes, dwelling-exposure typologies, and intervention pathways as part of the workflow. This series helps to explain how raw input data are translated into accessible screening results [19-21].

Table 3. Exposure typologies and associated intervention logic

Typology	Primary intervention logic
Dampness-dominated	waterproofing, leakage repair, plumbing correction, mould control
Ventilation-deficient	cross-ventilation improvement, exhaust fan, kitchen/bathroom ventilation upgrade, daylight improvement
Combustion-dominated	kitchen exhaust, source isolation, cooking-area ventilation improvement
Chemical-use-dominated	VOC source reduction, product substitution, and ventilation enhancement
Dust-reservoir-dominated	carpet reduction, cleanable surfaces, dust-control maintenance
Mixed high exposure	combined building repair, source control, and indoor monitoring

Table 4. Variable-domain configurations used in the ablation analysis

Model	Variables used
A	Building condition only
B	Building condition + ventilation/daylight
C	Building condition + ventilation/daylight + indoor sources
D	Building condition + ventilation/daylight + indoor sources + dust reservoirs
E	Full model including sensor indicators

The workflow "groups" the variables into six domains, transforms each variable (indicator) into a 0-1 risk score, aggregates the six sub-indices into an IBEI-HH, defines a risk class and predominant typology of exposure, and maps these to a provisional intervention priority.

Analytical structure of the IBEI-HH model

For each dwelling, the input data are arranged into six exposure domains:

The analytical representation begins by organizing every dwelling into six conceptually distinct exposure domains, thereby preventing unlike mechanisms from being prematurely collapsed. Equation (1) defines the six-domain vector used throughout the subsequent stages of scoring, classification, and modeling.

$$X_i = [X_i^M, X_i^{In}, X_i^C, X_i^H, X_i^R, X_i^S] \tag{1}$$

Separating the domains preserves causal interpretation: dwellings with similar total scores may require different responses when their risk is driven by dampness, ventilation, combustion, chemicals, reservoirs, or measured exposure.

Variable coding and normalization

Variables were converted to a common 0–1 screening scale, where 0 denotes the lowest encoded concern and 1 the highest. These values are relative analytical scores, not health-based exposure limits. Categorical survey responses require transparent numerical coding before they can be combined with continuous sensor measurements. Equation (2) defines condition presence, Equation (3) defines severity, Equation (4) normalizes positively oriented variables, and Equation (5) reverse-normalizes indicators for which higher values imply lower concern.

$$s(x) = \begin{cases} 0, & \text{condition absent} \\ 0.5, & \text{condition uncertain, occasional, or unknown} \\ 1, & \text{condition present} \end{cases} \tag{2}$$

$$s(x) = \begin{cases} 0, & \text{none} \\ 0.33, & \text{minor} \\ 0.67, & \text{moderate} \\ 1, & \text{severe} \end{cases} \tag{3}$$

$$s_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{4}$$

$$s_{ij}^{rev} = 1 - \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{5}$$

Construction of domain sub-indices

Moisture and dampness sub-index

The moisture domain brings together visible faults and contextual issues that may contribute to dampness or mould in a home. The moisture and dampness sub-index is obtained from Equation (6) as the average of the 10 indicators normalized according to the Equation.

$$M_i = \frac{L_i + ID_i + ED_i + DA_i + DH_i + VM_i + WL_i + LS_i + WP_i + PA_i}{10} \tag{6}$$

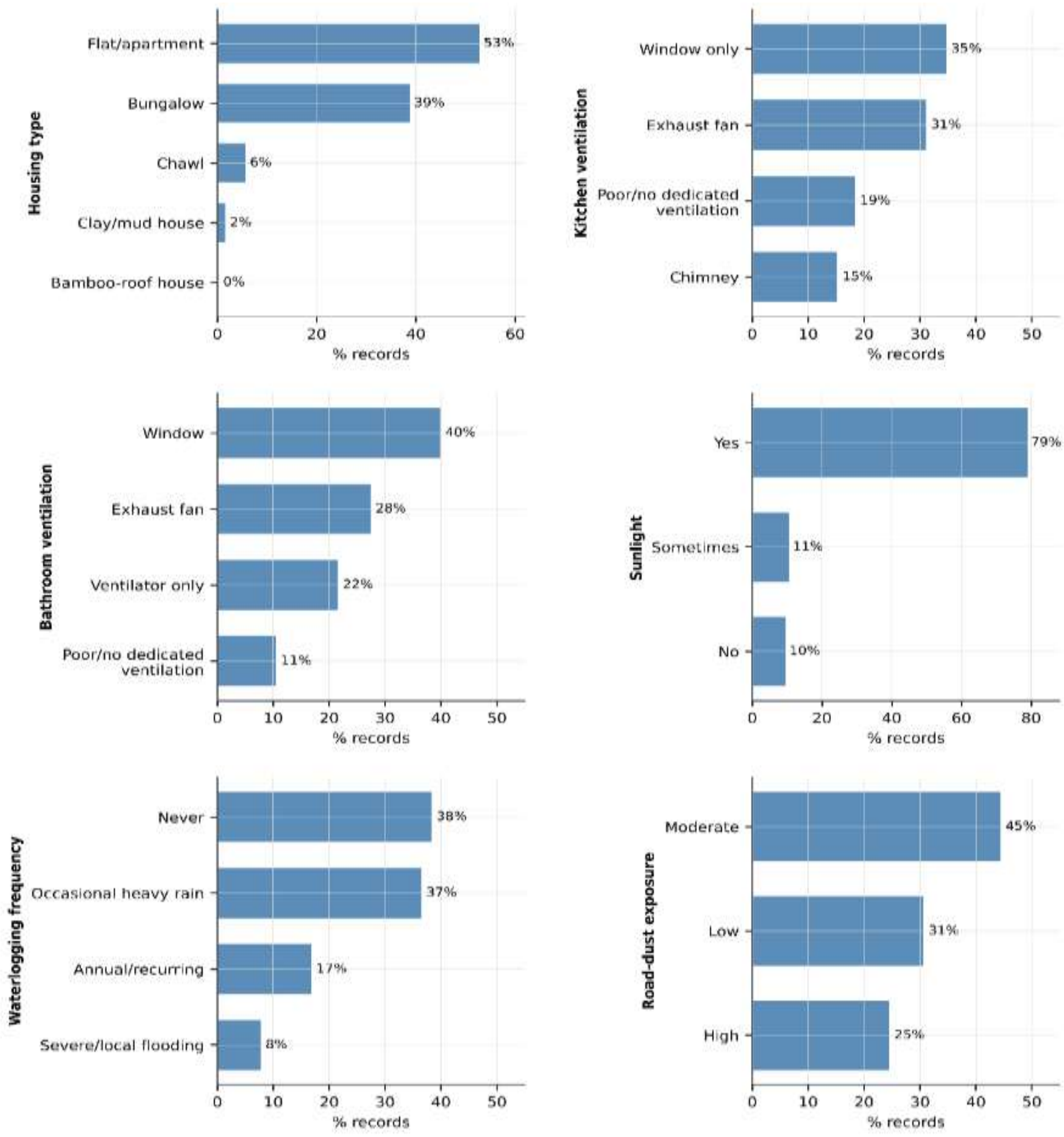

Figure 2. Sample and exposure characteristics of the 442 development-sample dwelling records

Higher moisture scores indicate greater exposure to leakage, dampness, mould, waterlogging, waterproofing deficiencies, and aging pipes.

Ventilation and daylight deficiency sub-index

The ventilation performance data are shown along with daylight and measured carbon dioxide levels, as these factors characterize other aspects of dilution and passive environmental control. Equations (7-9) are used to examine the domain score, reverse-engineer daylight (deficiency indicates greater risk), and normalize the observed CO2.

$$In_i = \frac{SL_i+WFR_i+CV_i+KV_i+BV_i+EF_i+CH_i+CO2_i+LX_i}{9} \quad (7)$$

$$LX_i = 1 - \frac{Lux_i - Lux^{\min}}{Lux^{\max} - Lux^{\min}} \quad (8)$$

$$CO2_i = \frac{CO2_i^{obs} - CO2^{\min}}{CO2^{\max} - CO2^{\min}} \quad (9)$$

Higher ventilation/daylight scores indicate weaker pollutant dilution, moisture removal, sunlight access, and passive environmental control.

Combustion source sub-index

The combustion pathway encompasses the presence of a source and the conditions that can affect the accumulation of pollutants. The combustion-source sub-index is calculated from Equation (10), while Equation (11) converts the observed CO and PM2.5 values to a comparable normalized value.

$$C_i = \frac{GS_i+WS_i+SM_i+TB_i+PKV_i+CO_i+PM25_i}{7} \quad (10)$$

$$CO_i = \frac{CO_i^{obs} - CO^{\min}}{CO^{\max} - CO^{\min}}$$
$$PM25_i = \frac{PM2.5_i^{obs} - PM2.5^{\min}}{PM2.5^{\max} - PM2.5^{\min}} \quad (11)$$

The combustion component captures the joint effect of indoor combustion sources, ventilation, PM2.5, and CO.

Chemical-use sub-index

Household product use, building-related chemical indicators, and measured VOCs are indicators of chemical exposure. The chemical-use sub-index is given by Equation (12), and the measured TVOC concentration is normalized (Equation (13)).

$$H_i = \frac{RF_i+CD_i+BL_i+IN_i+NC_i+TVOC_i}{6} \quad (12)$$

$$TVOC_i = \frac{TVOC_i^{obs} - TVOC^{\min}}{TVOC^{\max} - TVOC^{\min}} \quad (13)$$

Chemical exposure becomes more consequential where product use coincides with limited ventilation.

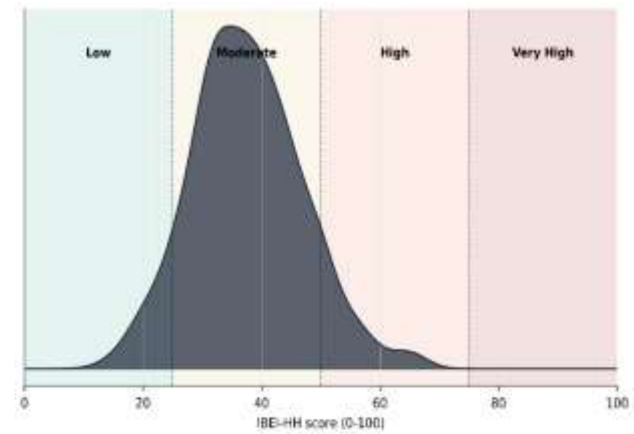

Figure 3. Distribution of IBEI-HH scores and classification thresholds

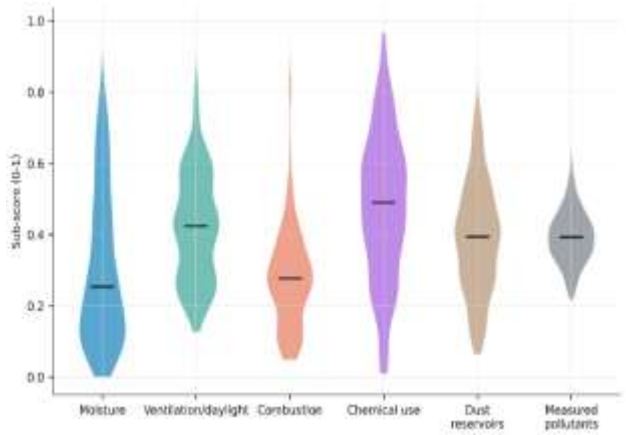

Figure 4. Distribution of the six IBEI-HH sub-indices

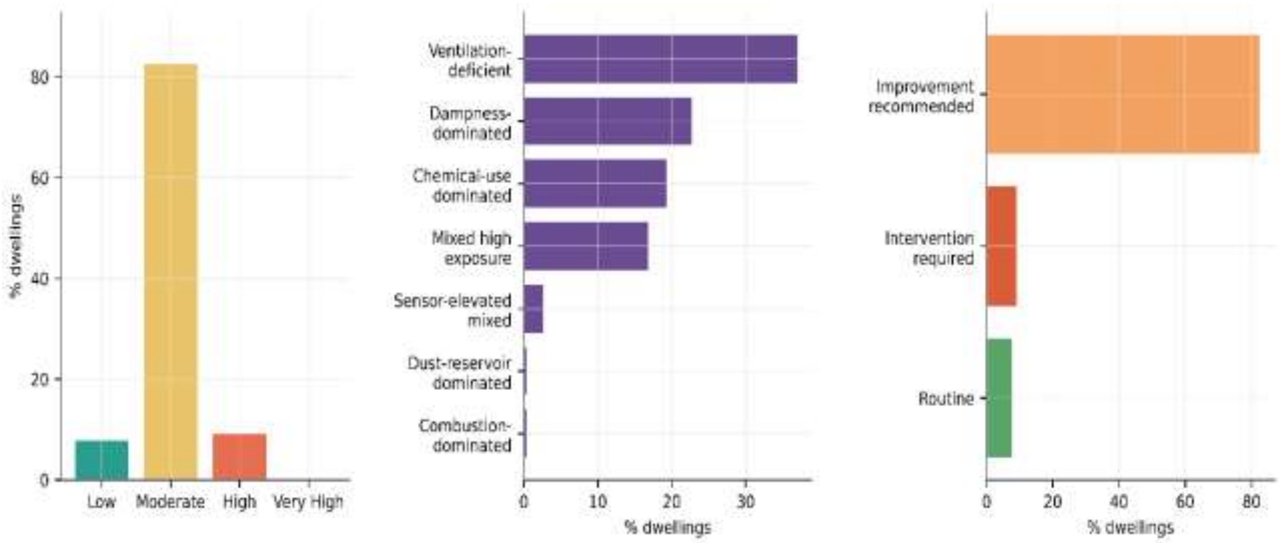

Figure 5. IBEI-HH risk classes, dominant exposure typologies, and intervention priorities

Table 5. Analytical dataset domains, variables, purposes, and derived outputs

Dataset component	Variables included	Purpose in Paper 2	Derived output
Identification and location	response_id, ward_no, locality, gps_lat_anonymized, gps_lon_anonymized, urban_context	Identifies each dwelling and supports spatial or ward-level indoor exposure analysis	Ward/locality exposure summaries and map-ready outputs
Moisture and dampness	raw_wall_leakage, raw_internal_dampness, raw_external_dampness, dampness_area_pct, dampness_height_cm, visible_mould, mould_smell, waterlogging_frequency, leakage_source_primary, raw_waterproofing, raw_pipe_age	Captures moisture ingress, dampness, mould potential, leakage source, and waterlogging-related indoor exposure	Moisture and dampness score (M i)
Ventilation and daylight	raw_sunlight, window_to_floor_ratio_pct, cross_ventilation, kitchen_ventilation, bathroom_ventilation, exhaust_fan_present, chimney_present, indoor_co2_ppm, lux_living_room	Measures pollutant dilution, moisture removal potential, daylight access, and ventilation adequacy	Ventilation and daylight deficiency score (Vi)
Combustion sources	raw_gas_stove, raw_wood_stove, raw_smoking, raw_tobacco, kitchen_ventilation, indoor_co_ppm, indoor_pm25_ugm3	Represents combustion-related indoor sources and measured combustion indicators	Combustion source score, (C i)
Chemical-use exposure	raw_room_freshener, raw_candles, raw_chlorine_bleach, raw_insecticide, raw_nonstick, indoor_tvoc_ppb	Captures household chemical-use practices and VOC-related exposure	Chemical-use score, (H i)
Dust and material reservoirs	raw_carpet, raw_sofa, raw_pets, raw_indoor_plants, raw_flooring, occupancy_density_person_per_room, indoor_pm10_ugm3	Represents indoor surfaces, reservoirs, occupancy, and dust-retention or resuspension conditions	Dust and reservoir score, (R i)
Sensor-based indoor environment	indoor_pm25_ugm3, indoor_pm10_ugm3, indoor_co2_ppm, indoor_co_ppm, indoor_tvoc_ppb, indoor_rh_pct, indoor_temp_c, lux_living_room	Estimated sensor variables used for internal workflow testing; values require field replacement	Sensor-based exposure score, (S i)
IBEI-HH derived variables	moisture_score_M, ventilation_daylight_score_V, combustion_score_C, chemical_score_H, dust_reservoir_score_R, sensor_exposure_score_S, ibei_score, ibei_score_0_100	Quantifies indoor exposure risk using six weighted sub-indices	Composite IBEI-HH score
Typology and intervention outputs	ibei_risk_class, dominant_exposure_component, dominant_exposure_typology, high_component_count, intervention_priority, intervention_recommendation, indoor_intervention_priority_score_IIPS	Converts exposure score into risk class, exposure pathway, and engineering intervention priority	Indoor exposure typology and intervention ranking

Dust and material reservoir sub-index

Materials that retain dust, occupancy, pets, plants, flooring, and coarse particles are combined to reflect potential conditions for reservoir and resuspension. The reservoir sub-index is computed in accordance with Equation (14), the density of occupancy in accordance with Equation (15), and the observed PM10 in accordance with Equation (16).

$$R_i = \frac{CP_i + SF_i + PT_i + PL_i + FL_i + Oh,D._i + PM10_i}{7} \quad (14)$$

$$Oh,D._i = \frac{Oh,D_i^{obs} - Oh,D^{\min}}{Oh,D^{\max} - Oh,D^{\min}} \quad (15)$$

$$PM10_i = \frac{PM10_i^{obs} - PM10^{\min}}{PM10^{\max} - PM10^{\min}} \quad (16)$$

The reservoir component represents dust retention or resuspension associated with furnishings, pets, plants, flooring, occupancy, and PM10.

Sensor-based exposure sub-index

The sensor's domain is a measurement-based alternative to the domains of the questionnaire and the building-condition variables. The relative humidity is transformed using Equation (18), and the temperature deviation is expressed as a consistently oriented risk term using Equation (19); Equation (17) is then used to calculate the combined sensor score.

$$S_i = \frac{PM25_i + PM10_i + CO2_i + CO_i + TVOC_i + RH_i + TEMP_i + LX_i}{8} \quad (17)$$

$$RH_i = \frac{RH_i^{obs} - RH^{min}}{RH^{max} - RH^{min}} \quad (18)$$

$$TEMP_i = \frac{|T_i - T_{ref}|}{\max|T_i - T_{ref}|} \quad (19)$$

Sensor indicators provide measurement-supported evidence and reduce reliance on self-reported housing conditions.

Indoor Building Exposure Index calculation

The composite index maintains the relative weight of each exposure pathway and yields a boundable, easy-to-interpret index. Equations (20) and (22) show all six weighted domains; Equation (21) specifies that the weights sum to 1;

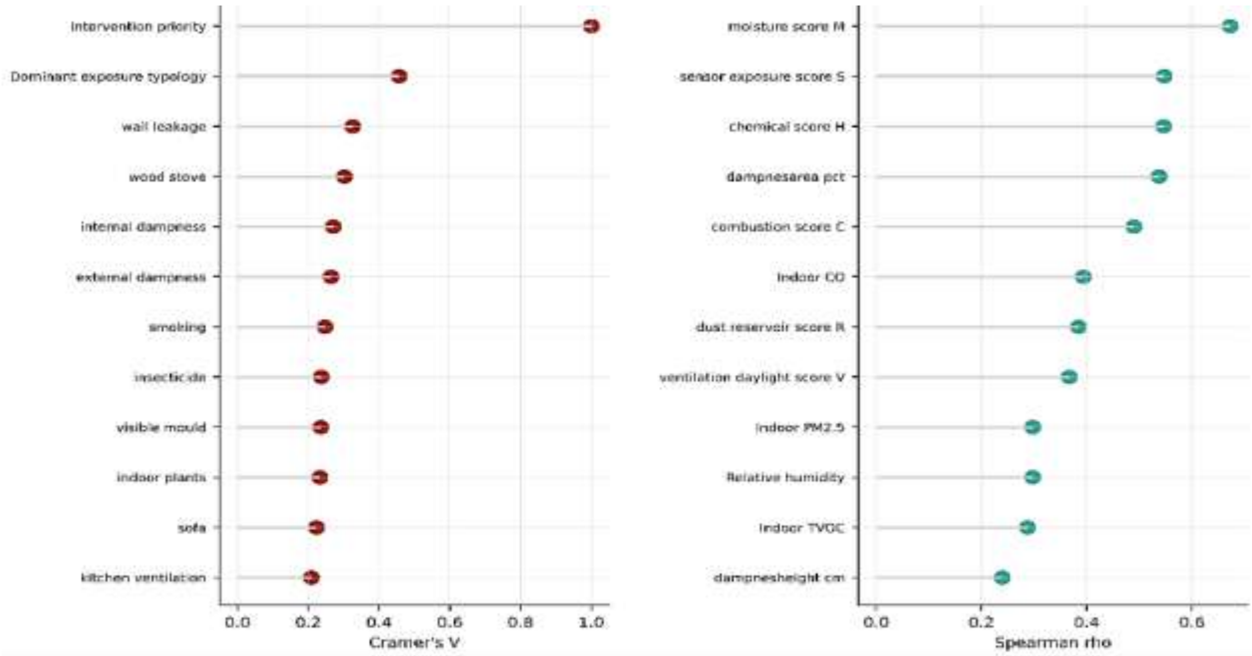

Figure 6. Internal association evidence for IBEI-HH risk classification

Figure 7. Encoded ventilation, mould, and waterlogging profiles across risk classe

Table 6. Core setup of the IBEI-HH internal evaluation

Item	Specification
Framework	Indoor Building Exposure Index for Healthy Housing, IBEI-HH
Study unit	One of 442 development-sample dwelling records
Main objective	Indoor exposure-risk classification and intervention prioritization
Primary output	IBEI-HH score and indoor exposure-risk class
Risk classes	Low, Moderate, High, Very High
Main decision group	High + Very High dwellings
Validation reference	Existing development-sample exposure score/class; internal comparison only
Prediction target 1	Derived four-class IBEI-HH label
Prediction target 2	Derived binary high-risk label
Final ranking score	$IIPS_i = 0.60IBEI_i + 0.40P_i^{HVH}$

Table 7. IBEI-HH components, principal indicators, and assigned weights

Component	Symbol	Main indicators	Weight
Moisture and dampness	M_i	Leakage, dampness, mould, dampness area/height, waterlogging, waterproofing, pipe age	0.25
Ventilation and daylight deficiency	V_i	Sunlight, window-to-floor ratio, kitchen/bathroom ventilation, CO2, lux	0.20
Combustion source	C_i	Gas stove, wood stove, smoking, tobacco, kitchen ventilation, CO, PM2.5	0.15
Chemical-use exposure	H_i	Room freshener, candles, bleach, insecticide, non-stick cookware, TVOC	0.15
Dust and reservoir	R_i	Carpets, sofa, pets, indoor plants, flooring, occupancy density, PM10	0.10
Sensor-based exposure	S_i	PM2.5, PM10, CO2, CO, TVOC, RH, temperature, lux	0.15

Table 8. IBEI-HH equation, classification thresholds, and priority metric

Parameter	Specification
Index equation	$IBEI_i = 0.25M_i + 0.20V_i + 0.15C_i + 0.15H_i + 0.10R_i + 0.15S_i$
Low	$0 \leq IBEI_i < 0.25$
Moderate	$0.25 \leq IBEI_i < 0.50$
High	$0.50 \leq IBEI_i < 0.75$
Very High	$0.75 \leq IBEI_i \leq 1.00$
Intervention-priority metric	Recall $_{HVH} = TP_{HVH}/(TP_{HVH} + FN_{HVH})$

Table 9. Random forest modeling and internal evaluation settings

Item	Specification
Main models	Four-class and binary random forest classifiers
Train-test split	80:20
Random seed	42
Class imbalance	Balanced class weights
Random forest trees	500
Minimum leaf size	2
Categorical variables	One-hot encoded
Numerical variables	Standard scaled
Model metrics	Accuracy, macro-F1, weighted-F1, High/Very-High recall, ROC-AUC, PR-AUC
Explainability	SHAP; permutation importance if SHAP is unavailable
Robustness checks	Descriptive Cronbach's alpha; weighting sensitivity; ablation study
Typology rule	Largest weighted component contribution; mixed exposure if ≥ 2 component scores exceed 0.60
Intervention output	Rule-based pathway recommendation requiring field confirmation

and Equation (22) shows the normalization that yields a reported 0–100 score.

$$IBEI_i = 0.25M_i + 0.20V_i + 0.15C_i + 0.15H_i + 0.10R_i + 0.15S_i \quad (20)$$

$$0.25 + 0.20 + 0.15 + 0.15 + 0.10 + 0.15 = 1 \quad (21)$$

Moisture/dampness was given a weight of 0.25, and ventilation/daylight a weight of 0.20 a priori; combustion, chemical use, and sensors each a weight of 0.15, and reservoir a weight of 0.10. These weights are expert-based and not health-outcome-based, but rather reflect the building-performance focus of this study. Agreement on alternative weighting schemes is thus not necessarily interpreted as evidence that the baseline is the best weighting scheme. Table 1 describes 4 provisional screening classes and the level of review for each; these screening bands are not validated health reference limits. In Table 2, the largest weighted contribution to a component is associated with an interpretable typology of exposures, thereby retaining the most influential pathway despite total scores that appear similar across the various combinations.

$$IBEI_{100,i} = 100 \times IBEI_i \tag{22}$$

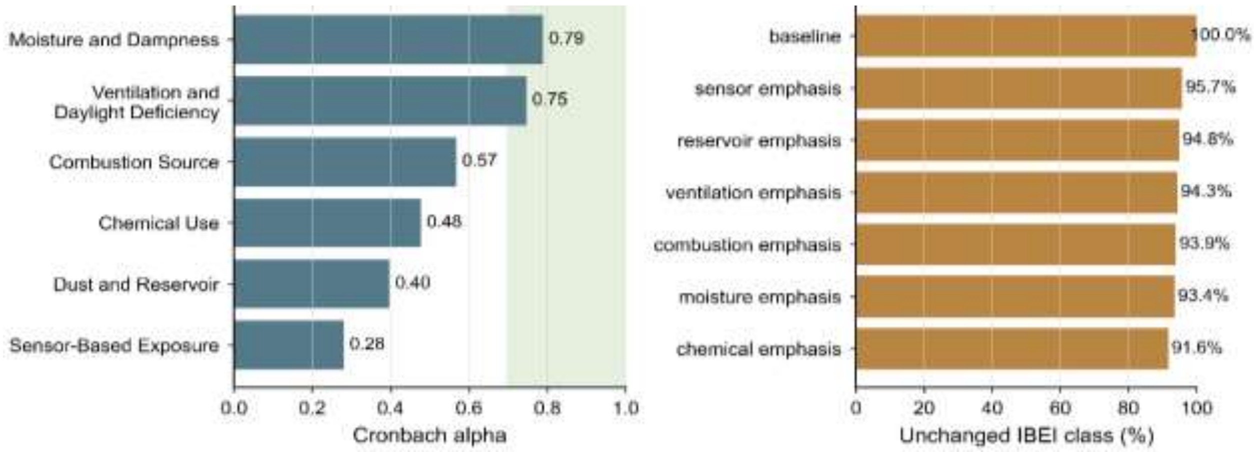

Figure 8. Internal consistency and weighting sensitivity of IBEI-HH components

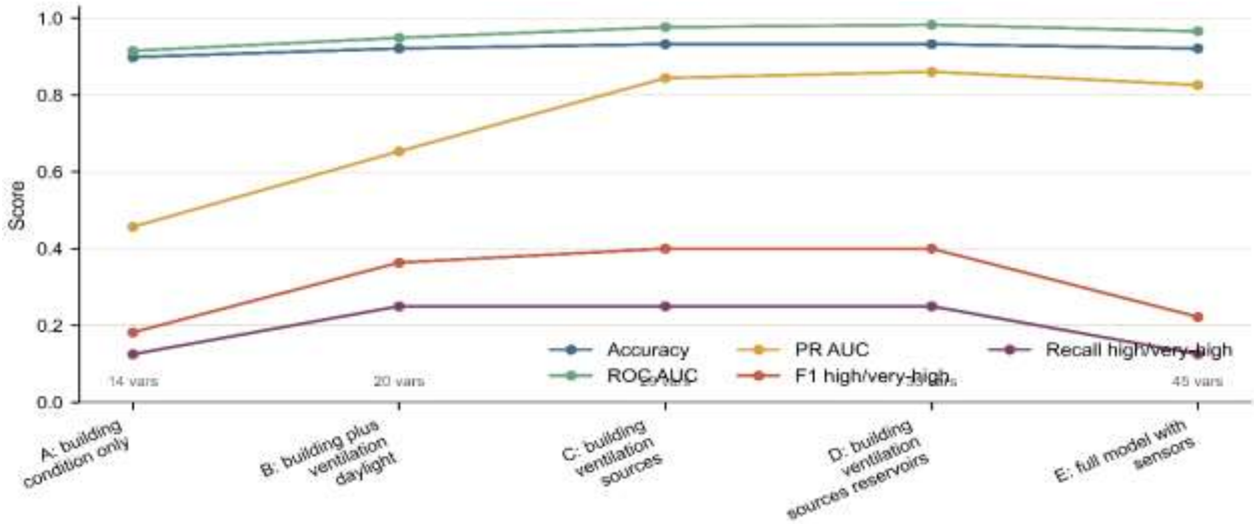

Figure 9. Ablation performance of the random forest models

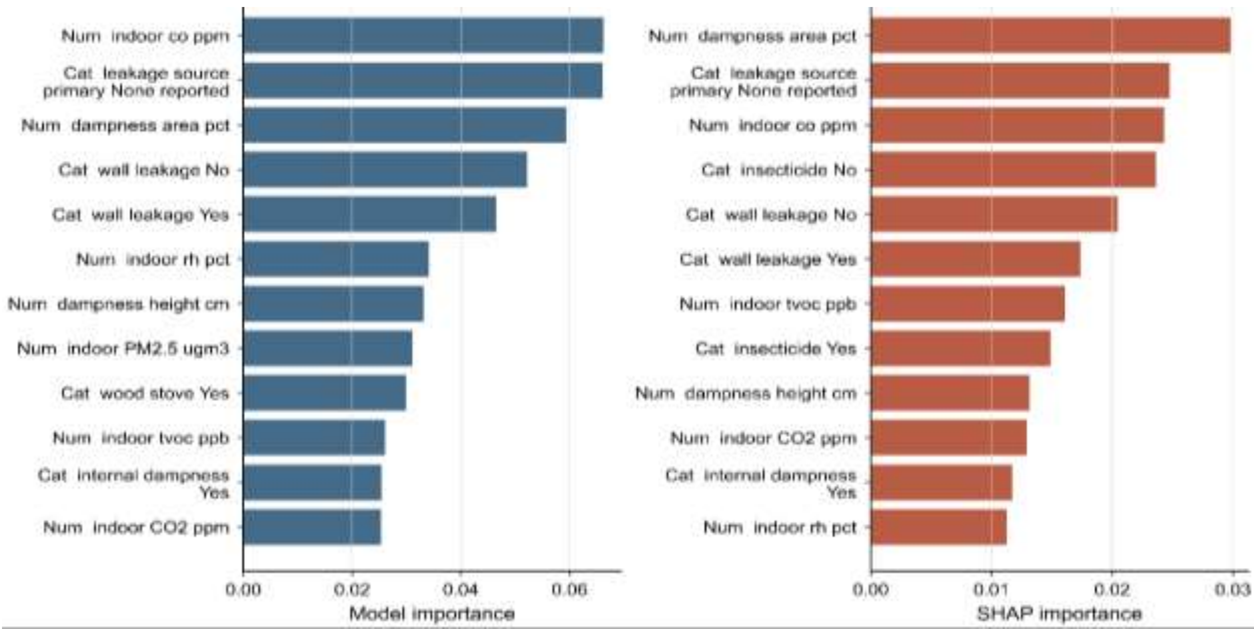

Figure 10. Random-forest feature importance and SHAP explainability

Table 10. Development-sample risk classes and provisional intervention priorities

Risk class	Records	Share	Provisional priority	Interpretation
Low	35	7.92%	Routine review	Maintain conditions and verify during routine inspection
Moderate	366	82.81%	Improvement recommended	Review the dominant pathway and plan targeted improvements
High	41	9.28%	Intervention required	Prioritize field inspection and confirm the encoded drivers
Very high	0	0.00%	Immediate intervention	No development-sample records; retain category for prospective use

Risk classification, typology, and intervention mapping

Exposure-risk classification

Risk categories are used as screening bands rather than clinical or regulatory thresholds. Equation (23) assigns low, moderate, high, and very-high classes from the normalized IBEI-HH score.

$$\text{Risk}_i = \begin{cases} \text{Low}, & 0 \leq IBEI_i < 0.25 \\ \text{Moderate}, & 0.25 \leq IBEI_i < 0.50 \\ \text{High}, & 0.50 \leq IBEI_i < 0.75 \\ \text{Very High}, & 0.75 \leq IBEI_i \leq 1.00 \end{cases} \tag{23}$$

Exposure typology identification

The typology assignment preserves the dominant mechanism underlying a dwelling's total score and also allows multiple elevated pathways to be recognized. Equation (24) defines weighted component contributions, Equation (25) selects the largest contribution, and Equation (26) identifies mixed high-exposure cases.

$$\begin{aligned} CM_i &= 0.25M_i \\ CV_i &= 0.20V_i \\ CC_i &= 0.15C_i \\ CH_i &= 0.15H_i \\ CR_i &= 0.10R_i \\ CS_i &= 0.15S_i \end{aligned} \tag{24}$$

$$T_{ypology} = \arg\max[CM_i, CV_i, CC_i, CH_i, CR_i, CS_i] \tag{25}$$

$$\sum_{k=1}^{6} I(Z_I \geq 0.60) \geq 2 \tag{26}$$

This typology layer distinguishes dwellings with comparable total scores but different dominant causes; cases with two or more component scores above 0.60 are treated as mixed high exposure.

Table 3 translates each dominant exposure pathway into a provisional engineering review pathway. The listed actions indicate what should be examined during subsequent field verification.

Intervention mapping

The decision layer translates analytical outputs into provisional engineering priorities that still require field confirmation. Equation (27) maps risk and typology to an intervention pathway, and Equation (28) assigns the corresponding action-priority category.

$$\text{Intervention}_i = f(R\text{ i s }k_i, \text{ Typology}_i) \tag{27}$$

$$\text{Priority}_i = \begin{cases} \text{Routine}, & \text{Risk}_i = \text{Low} \\ \text{Improvement recommended}, & \text{Risk}_i = \text{Moderate} \\ \text{Intervention required}, & \text{Risk}_i = \text{High} \\ \text{Immediate intervention}, & \text{Risk}_i = \text{Very High} \end{cases} \tag{28}$$

Risk and typology are translated into actionable priorities for building repair, ventilation improvement, source control, or combined intervention.

Internal association analysis

Internal association tests were used to examine relationships among encoded dwelling variables, component scores, and derived classes without implying causation. Equation (29) defines the chi-square statistic, Equation (30) defines Cramér's V, and Equation (31) defines Spearman's rank-correlation coefficient.

$$x^2 = \sum \frac{(The - E)^2}{AND} \tag{29}$$

$$In = \sqrt{\frac{x^2}{N(k-1)}} \tag{30}$$

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{31}$$

Chi-square tests with Cramér's V and Spearman's rank correlations characterized the internal associations between the encoded variables and IBEI-HH classes or scores. Associations with component variables were expected because those variables contribute mathematically to the index; they were therefore treated as evidence of internal coherence rather than as independent validation. The derived intervention category was excluded from validity claims because it is deterministically assigned from the risk class.

Predictive modelling

Predictive evaluation considered both the original four-class outcome and a priority-oriented binary outcome, as class imbalance can obscure the detection of high-risk cases. Equation (32) defines the four-class target, Equation (33) defines the binary target, Equation (34) defines the predicted high/very-high probability, Equation (35) constructs the intervention-priority score, and Equation (36) defines recall for the priority class.

$$AND_i \in \{\text{ Low, Moderate, High, Very High }\} \tag{32}$$

$$AND_i^{HVH} = \begin{cases} 1, & AND_i \in \{\text{ High, Very High }\} \\ 0, & AND_i \in \{\text{ Low, Moderate }\} \end{cases} \tag{33}$$

The binary task directly addresses the practical decision of whether a dwelling requires high-priority exposure reduction.

$$P_i^{HVH} = P(\text{ And }_i^{HVH} = 1 \mid X_i) \tag{34}$$

$$IIPS_i = 0.60IBEI_i + 0.40P_i^{HVH} \tag{35}$$

Two random-forest models: a four-class model and a binary (high/very-high) model, were both successfully evaluated.

The records were randomly divided into training and test sets, comprising 80% (353) and 20% (89) of the records, respectively, with Random Seed 42. The class weights were equalized, 500 trees were used, and the minimum leaf size was set to 2. It is the internal recoverability of the rule-based class; it is not an external validity (predictive validity) measure; its index input is the same for the performance measure and the prediction target; Performance is measured using the same index input used to derive the prediction target. To account for nonlinear relationships between building defect variables, indoor sources and activities, and measured environmental variables, the random forest model is used.

$$\text{Recall}_{HVH} = \frac{TP_{HVH}}{TP_{HVH}+FN_{HVH}} \tag{36}$$

Evaluation used accuracy, macro- and weighted F1, high/very-high recall, ROC-AUC, and PR-AUC. Because only 41 of 442 records were high risk and none were very high risk, threshold recall and the confusion matrix were prioritized over accuracy.

Explainability and ablation analysis

Model explanations were summarized using additive SHAP contributions, enabling interpretation of influential inputs at both the dwelling and dataset levels. Equation (37) decomposes a prediction into its baseline and feature contributions, while Equation (38) defines global mean absolute SHAP importance.

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j \tag{37}$$

$$I_j = \frac{1}{N}\sum_{i=1}^{N} |\phi_{ij}| \tag{38}$$

SHAP values identify whether dampness, ventilation, pollutants, chemical use, combustion, or reservoirs drive predictions. Table 4 shows how successive exposure domains were added during the ablation analysis. Comparing these configurations reveals the incremental contributions of ventilation, sources, reservoirs, and sensors to the modeling effort.

Five ablation models progressively add ventilation, sources, reservoirs, and sensors to building-condition variables and compare accuracy, precision, recall, ROC-AUC, and PR-AUC.

Sensitivity and internal consistency

Robustness checks examine whether conclusions remain stable when domains are removed, or their weights are varied, while internal consistency is reported only as a descriptive diagnostic. Equation (39) defines the ablation change, Equation (40) calculates classification stability, and Equation (41) defines Cronbach's alpha.

$$\Delta IBEI_i = \left| IBEI_i^{\text{all}} - IBEI_i^{\text{base}} \right| \tag{39}$$

Figure 11. Exploratory spatial surface of mean IBEI-HH scores

$$\text{Stability } = \frac{\sum_{i=1}^{N} I\left(Risk_i^{\text{all}} = Risk_i^{\text{base}}\right)}{N} \quad (40)$$

$$a = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} s_j^2}{s_T^2}\right) \quad (41)$$

Sensitivity was evaluated using classification agreement across six different weighting schemes. Cronbach's alpha was reported descriptively only: These are formative, not interchangeable, indicators of IBEI-HH, and therefore, a low alpha does not automatically indicate that they were invalid. The analytical domains, constituent variables, purpose, and derived outputs are reported in Table 5. A summary of the study unit, analytical objectives, outputs, and internally derived prediction targets is provided in Table 6. The six IBEI-HH components, their definitions, and the a priori weights are presented in Table 7. Table 8 brings together these three components of the index: index specification, screening thresholds, and priority metric. The internal evaluation procedures and settings for the random forest in the modeling workflow that can be used for replication are provided in Table 9.

A transparent methodology was developed to convert the different types of dwelling indicators and to compute comparable exposure scores and typologies, and to internally evaluate priority classes. This framework will help with future field validation and outside validation efforts, but will not lead to health or engineering conclusions.

RESULTS

Results include descriptions of the 442-record development dataset, use of indices as behavior, internal association, sensitivity, model diagnostics, and spatial patterns. Does not attempt to estimate prevalence in the field because the above data on the enhanced building and sensor variables were not field-observed.

Sample characteristics

The flats/"apartments" accounted for 53%, the bungalows accounted for 39%, the chawls accounted for 6%, and the clay/mud houses accounted for 2% of the sample. 19% of households lacked dedicated kitchen ventilation or had poor dedicated kitchen ventilation, 21% had no or only occasional sunlight, and 54% experienced waterlogging at least sometimes when it rained heavily. The 442-record development set shown in Figure 2 briefly characterizes the basic housing and exposure features encoded in those records, offering a glimpse of the housing and exposure characteristics that underlie the subsequent score distributions.

Score distribution and exposure pathways

IBEI-HH scores were concentrated in the moderate band (25–50), with a limited upper tail in the high-risk band and no very-high-risk cases. Figure 3 shows this distribution and the resulting class imbalance considered during predictive evaluation.

The assigned weights focused on moisture/dampness (0.25) and ventilation/daylight (0.20). Ventilation/daylight (0.0851), moisture (0.0774), chemical use (0.0718), sensors (0.0593), combustion (0.0435), and dust reservoirs (0.0393) had the highest weighted contribution. The median for the sub-scores ranged from 0.4898 (chemical exposure) to 0.2537 (Moisture). However, there was a wide upper tail for moisture. Figure 4 shows the distribution of all six component scores and also reveals the different central tendencies and spreads of the scores.

Of the 442 development-sample records, 35 (7.92%) were low risk, 366 (82.81%) moderate risk, 41 (9.28%) high risk, and none were very high risk. The most common typologies were those with the highest numbers of ventilation-deficient (164 records, 37.10%), dampness-dominated (101 records, 22.85%), and chemical-use-dominated (86 records, 19.46%) typologies. These percentages refer to the output of the development sample, not population prevalence. A numerical score is transformed into an interpretable decision pathway in figure 5 by combining the composite score and the major exposure pathway.

Internal association and robustness results

Among non-derived categorical variables, wall leakage showed Cramér's V=0.325 ($p<0.001$), wood-stove use V=0.302 ($p<0.001$), internal dampness V=0.270 ($p<0.001$), external dampness V=0.264 ($p<0.001$), smoking V=0.247 ($p<0.001$), insecticide use V=0.236 ($p<0.001$), and visible mould V=0.235 ($p<0.001$). The ρ values between the components and IBEI-HH varied between 0.367 for ventilation/daylight to 0.673 for moisture ($p<0.001$ for all components). These inputs are mathematically part of the computation of IBEI-HH, so the associations aren't necessarily valid; they just reflect how the scores behave internally. There was a higher proportion of high-risk records than low-risk ones with poor or missing kitchen ventilation (32%) and visible mould (39%); none of the low-risk records had visible mould; high-risk records also had a higher proportion of those reported to have waterlogging at least once or annually (24%). The contrasts presented here are descriptive only and do not provide independent field validation. The highest categorical and rank relationships are summarized in Figure 6, whereas comparisons across the derived risk classes of the encoded ventilation, mould, and waterlogging/inundation are presented in Figure 7.

The person value of moisture/dampness was 0.789, ventilation/daylight was 0.746, combustion was 0.567, chemical use was 0.477, reservoirs was 0.396, and sensors was 0.280. The lower values are suitable for heterogeneous formative indicators and help pinpoint areas in need of field reassessment. Class agreement for the alternative weighting schemes ranged from 91.63% to 95.70%. Figure 8 compares descriptive internal consistency and classification stability when using alternative component weights.

Predictive performance and explainability

The random forest with four classes achieved an accuracy of 0.888, a macro-F1 of 0.642, and a weighted-F1 of 0.863. Recall was 0.286 for low-risk records, 1.000 for moderate-risk records, and 0.375 for high-risk records; there were no very high-risk records. After adding the ventilation, source, and reservoir variables, the rank discrimination in the ablation results improved, but it was still an internal rule-catch. The model's accuracy, ROC-AUC, and PR-AUC for binary high/very-high classification were 0.921, 0.969, and 0.825, respectively. The confusion matrix had 81 true negatives, 0 false positives, 7 false negatives, and 1 true positive, with high/very-high precision 1.000, recall 0.125, and F1 0.222. The model achieved good rankings and low sensitivity insufficiency at the tested threshold. Relative humidity, leakage source, dampness area, wall leakage, indoor CO, PM2.5, TVOCs, internal dampness, and CO2 were the factors that dominate the feature importance and SHAP scores. The discrimination changes when exposure domains are added, as seen in Figure 9, as well as in which variables most strongly influenced the internally derived classification, as seen in Figure 10. These are the importance estimates; they do not suggest causation.

Descriptive spatial patterns

Ward 17 had the highest development-sample mean IBEI-HH score (41.05), and Ward 22 had the highest high-risk share (20.00%). Adgaon locality had the highest mean score (41.05), and Rane Nagar locality had the highest high-risk score (20.00%). These values are not inferential hotspot identification values because no spatial significance test was conducted. The descriptive surface revealed the presence of localized concentrations (1) in the western, south-western, central-southern, Tapovan, Adgaon, Nashik Road, and south-eastern areas. The map is exploratory, as the coordinates are anonymized and no spatial autocorrelation testing was conducted. The rule-based intervention map shared higher-priority records across clusters, providing an example of the rule-based framework that can be used to organize field inspections without identifying that an individual dwelling needs to be addressed. Waterlogging of higher scores was also descriptively associated with annual/recurring waterlogging and severe/waterlogging in the locality of the waterlogging. The locations of concentrations in Figure 11 are descriptive, not causal, of mean development-sample scores.

Intervention-priority summary

The evidence provided was to offer feasibility only for classes and typologies created through the analytical workflow that were deemed interpretable. Table 10 shows the number and/or percentage of development-sample records in each risk class and their respective provisional intervention priority. These priorities are to establish future inspection exposure and not to be a substitute for field exposure measurements, independent outcome, engineer verification, or external validation.

DISCUSSION

Principal findings

The geo-referenced development variables, which are typically heterodox in nature, were transformed by IBEI-HH into a usable screening score, a dominant-pathway typology, and an intervention logic. Its main impact is on interpretability (similar total scores may have been attributed to different moisture, ventilation, combustion, chemical, reservoir, or sensor causes). Note that the preponderance of the 'moderate' class cannot be inferred to be the same for Nashik housing since improved variables were estimated and the sampling design was not specified as population-representative. The class distribution is representative of the flow of classes, not of the presence of classes.

Building-condition, ventilation, and source pathways

The index acted as intended: dampness, leakage, mould and aging pipes, waterlogging, all of this added to the moisture pathway. This is physically plausible in the context of a monsoon climate, but for real-world impacts, the estimated fields should be replaced with site observations and repeated moisture measurements from the present data set. Despite moisture being assigned the greatest weighting, maximum ventilation/daylight contributed to the greatest mean weighted contribution. The prevalence of this finding would indicate either a general failure in ventilation encoding or more frequent and elevated encoding. To determine whether this pattern continues, field measurements of air change, window operation, daylight, and occupancy are needed.

Exposure typologies and practical interpretation

The typology layer is a practical benefit over an undifferentiated composite score by reporting the predominant pathway responsible for each score. However, the recommended actions are called "decision rules," not actual prescriptions. Before intervention, the defects should be identified, including the extent of the defect, its source, and feasibility and cost, through a qualified inspection. In this phase, IBEI-HH is more of an organizer for data collection and inspection, not a tool for decision-making automation. This is useful for documenting the rationale for flagging a dwelling and for correlating each pathway with a response.

Internal evaluation, model performance, and robustness

Correlations for the components, as well as high random-forest AUC values, must be interpreted carefully when the outcome class is drawn from the same inputs. The low binary recall of 0.125 and the number of high-risk records not captured (7) are more important for screening than the AUC value. For truly predictive validation, independent field outcomes must be used. Most records showed weight sensitivity within the normal bounds of reweighting. Cronbach's alpha does not have such an impact in this case, as the domains are formative (i.e., leakage, waterlogging, mould, and pipe age do not necessarily co-occur). Future studies need to focus more on assessing experts' content validity, test–retest reliability, inter-rater reliability, and criterion validity.

Spatial planning implications

Despite the ability to show a potential planning interface through the maps, anonymized coordinates, unknown sampling density, and the lack of spatial significance testing prevent neighborhood-level inferences. In future research, structure spatial units, minimal counts, autocorrelation, uncertainty, and privacy should be reported.

Limitations and future work

The problem is that the building and sensor parameters were again modeled in a workflow test, but not actually measured in the field – verification of the engineers is still pending for all 442 records. These include an undocumented recruitment strategy, a single survey date, few data on the history of the birds or data relating to the sensor hardware or the calibration of the sensors, weights and thresholds by experts, mathematical linkage of predictor and data outcomes, a single train–test split, severe class imbalance, no very-high-risk cases, and descriptive rather than inferential spatial analyses. Confirmatory works entail preregistered coding, parallel measurements during monsoon and non-monsoon conditions with the same wave or dwellings from the same wave, additional measurements on dwellings in the same wave (nested cross-validation), additional measurements on dwellings from a different wave (cross-validation), remeasurement of dwellings with independent health outcomes, and measurement of dwellings from an external city (city external validation). Within these constraints, IBEI-HH offers a tested architecture that connects building condition, behavior, environmental measurements, environmental space, and engineering review [22] [23].

CONCLUSION

This research created IBEI-HH, a robust six-domain framework that organizes information about exposures to healthy housing. The classifications generated by the workflow across 442 records in the POC were 35 low-risk and 366 moderate-risk, with no very high-risk and 41 high-risk, yielding agreement of 91.63%–95.70% when using other weights. The 0.969 ROC-AUC of the binary classifier resulted in a low recall (0.125) in the high-risk group; therefore, the tested model is not suitable for screening deployment. More importantly, an estimate of the enhanced building and sensor fields was done, but not a field observation. The findings can thus provide information on "computational feasibility" and "interpretability," but not on exposure prevalence, health risk, or external validity. For IBEI-HH to support operational healthy-housing decisions, it must be both field-measured and engineer-verified; independently, outcomes must occur; robust validation is necessary; and external replication must be achieved.

REFERENCES

1. N. E. Klepeis et al., "The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants," J. Expo. Anal. Environ. Epidemiol., vol. 11, no. 3, pp. 231–252, 2001, doi: 10.1038/sj.jea.7500165.
2. A. P. Jones, "Indoor air quality and health," Atmos. Environ., vol. 33, no. 28, pp. 4535–4564, 1999, doi: 10.1016/S1352-2310(99)00272-1.
3. J. D. Spengler and Q. Chen, "Indoor air quality factors in designing a healthy building," Annu. Rev. Energy Environ., vol. 25, pp. 567–600, 2000, doi: 10.1146/annurev.energy.25.1.567.
4. M. J. Mendell, A. G. Mirer, K. Cheung, M. Tong, and J. Douwes, "Respiratory and allergic health effects of dampness, mold, and dampness-related agents: A review of

the epidemiologic evidence," Environ. Health Perspect., vol. 119, no. 6, pp. 748–756, 2011, doi: 10.1289/ehp.1002410.

5. R. Quansah, M. S. Jaakkola, T. T. Hugg, S. A. M. Heikkinen, and J. J. K. Jaakkola, "Residential dampness and molds and the risk of developing asthma: A systematic review and meta-analysis," PLoS One, vol. 7, no. 11, Art. no. e47526, 2012, doi: 10.1371/journal.pone.0047526.
6. C.-G. Bornehag et al., "Dampness in buildings and health: Nordic interdisciplinary review of the scientific evidence on associations between exposure to 'dampness' in buildings and health effects (NORDDAMP)," Indoor Air, vol. 11, no. 2, pp. 72–86, 2001, doi: 10.1034/j.1600-0668.2001.110202..
7. J. Sundell et al., "Ventilation rates and health: Multidisciplinary review of the scientific literature," Indoor Air, vol. 21, no. 3, pp. 191–204, 2011, doi: 10.1111/j.1600-0668.2010.00703.x.
8. W. J. Fisk, Q. Lei-Gomez, and M. J. Mendell, "Meta-analyses of the associations of respiratory health effects with dampness and mold in homes," Indoor Air, vol. 17, no. 4, pp. 284–296, 2007, doi: 10.1111/j.1600-0668.2007.00475.x.
9. M. Frontczak and P. Wargocki, "Literature survey on how different factors influence human comfort in indoor environments," Build. Environ., vol. 46, no. 4, pp. 922–937, 2011, doi: 10.1016/j.buildenv.2010.10.021.
10. P. Wolkoff, "Indoor air humidity, air quality, and health—An overview," Int. J. Hyg. Environ. Health, vol. 221, no. 3, pp. 376–390, 2018, doi: 10.1016/j.ijheh.2018.01.015.
11. T. Salthammer, S. Mentese, and R. Marutzky, "Formaldehyde in the indoor environment," Chem. Rev., vol. 110, no. 4, pp. 2536–2572, 2010, doi: 10.1021/cr800399g.
12. C. J. Weschler, "Changes in indoor pollutants since the 1950s," Atmos. Environ., vol. 43, no. 1, pp. 153–169, 2009, doi: 10.1016/j.atmosenv.2008.09.044.
13. W. W. Nazaroff, "Indoor particle dynamics," Indoor Air, vol. 14, suppl. 7, pp. 175–183, 2004, doi: 10.1111/j.1600-0668.2004.00286.x.
14. S. B. Gordon et al., "Respiratory risks from household air pollution in low and middle income countries," Lancet Respir. Med., vol. 2, no. 10, pp. 823–860, 2014, doi: 10.1016/S2213-2600(14)70168-7.
15. K. R. Smith et al., "Millions dead: How do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution," Annu. Rev. Public Health, vol. 35, pp. 185–206, 2014, doi: 10.1146/annurev-publhealth-032013-182356.
16. U. Satish et al., "Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance," Environ. Health Perspect., vol. 120, no. 12, pp. 1671–1677, 2012, doi: 10.1289/ehp.1104789.
17. J. G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler, "Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers," Environ. Health Perspect., vol. 124, no. 6, pp. 805–812, 2016, doi: 10.1289/ehp.1510037.
18. S. Vardoulakis et al., "Impact of climate change on the domestic indoor environment and associated health risks in the UK," Environ. Int., vol. 85, pp. 299–313, 2015, doi: 10.1016/j.envint.2015.09.010.
19. B. Maag, Z. Zhou, and L. Thiele, "A survey on sensor calibration in air pollution monitoring deployments," Internet Things, vol. 4, pp. 35–54, 2018, doi: 10.1016/j.iot.2018.09.001.
20. J. Wang, S. Viciano-Tudela, L. Parra, R. Lacuesta, and J. Lloret, "Evaluation of suitability of low-cost gas sensors for monitoring indoor and outdoor urban areas," IEEE Sensors J., vol. 23, no. 18, pp. 20968–20975, 2023, doi: 10.1109/JSEN.2023.3301651.
21. J. Lloret, E. Macías, A. Suárez, and R. Lacuesta, "Ubiquitous monitoring of electrical household appliances," Sensors, vol. 12, no. 11, pp. 15159–15191, 2012, doi: 10.3390/s121115159.
22. P. O. S. Vaz de Melo, A. Carneiro Viana, M. Fiore, K. Jaffrès-Runser, F. Le Mouël, A. A. F. Loureiro, L. Addepalli, and G. Chen, "RECAST: Telling apart social and random relationships in dynamic networks," Perform. Eval., vol. 87, pp. 19–36, 2015, doi: 10.1016/j.peva.2015.01.005.
23. H. Zhang and R. Srinivasan, "A systematic review of air quality sensors, guidelines, and measurement studies for indoor air quality management," Sustainability, vol. 12, no. 21, Art. no. 9045, 2020, doi: 10.3390/su12219045.