Advanced Sensor Research

ADVANCED SENSOR RESEARCH Open Access

www.advsensorres.com

RESEARCH ARTICLE OPEN ACCESS

A Sensor-to-Algorithm Framework for Respiration Monitoring: Stretchable Strain Sensors Coupled With Machine Learning Feature Optimization

Youngjun Lee[1] | Seokkyoon Hong[1,2] | Jinheon Jeong[1] | Dong Rip Kim[3] | Chi Hwan Lee[1,4,5,6,7] | Rachel K. Surowiec[1]

[1]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, USA | [2]School of Polymer Science and Engineering, Chonnam National University, Gwangju, Republic of Korea | [3]School of Mechanical Engineering, Hanyang University, Seoul, Republic of Korea | [4]School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, USA | [5]School of Materials Engineering, Purdue University, West Lafayette, Indiana, USA | [6]Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA | [7]Center for Implantable Devices, Purdue University, West Lafayette, Indiana 47907, USA

Correspondence: Chi Hwan Lee (lee2270@purdue.edu) | Rachel K. Surowiec (rsurowie@purdue.edu)

Received: 8 January 2026 | **Revised:** 19 June 2026 | **Accepted:** 15 July 2026

Keywords: classification | feature extraction | machine learning | respiration | strain sensor

ABSTRACT

Reliable classification of respiration states is essential for wearable technologies and clinical monitoring, yet robust methods for distinguishing subtle waveform differences across breathing conditions remain limited. This study aims to develop a machine learning (ML)–optimized framework for accurately classifying shallow, normal, and deep respiration patterns from a stretchable strain sensor. Respiration waveforms were transformed into feature representations derived from both handcrafted ML features and convolutional neural network (CNN)-based latent embeddings. Multiple classifier families were evaluated using ten-fold stratified cross-validation. Random forest achieved the strongest overall performance, yielding 95.0% accuracy for shallow versus normal breathing (59/60 and 55/60 correctly classified), 99.2% accuracy for shallow versus deep (60/60 and 59/60 correct), and 91.7% accuracy for normal versus deep (56/60 and 54/60 correct). Shapley Additive exPlanations (SHAP) analysis and hierarchical clustering confirmed that high-ranking features consistently reflected amplitude variability and signal-energy differences across breathing states. These results demonstrate that combining strain sensor waveforms with ML models enables preliminary and interpretable classification of respiratory patterns. Unlike prior respiration-classification approaches, our framework explicitly links breathing-induced strain-sensor deformation to interpretable waveform features through SHAP-guided model optimization, enabling a physically grounded and explainable pathway from material response to classifier decision.

1 | Introduction

Continuous monitoring of human respiration has become increasingly important in modern health assessment, offering key insights into pulmonary function, autonomic regulation, stress response, and early signs of cardiorespiratory instability [1–3]. As respiratory abnormalities often appear intermittently and vary across daily activities, real-time monitoring outside clinical environments is essential [4, 5]. Recent advances in wearable sensing technologies, particularly stretchable strain sensors based

Youngjun Lee and Seokkyoon Hong contributed equally to this work.

on carbon nanotube–polydimethylsiloxane (CNT–PDMS) composites, have improved access to high-resolution respiratory time-series data [6–10]. In this composite, CNTs provide a conductive network with exceptional mechanical and electronic properties, while the PDMS elastomer serves as a stretchable support matrix [11, 12]. Owing to their high stretchability, stable piezoresistive response, and dynamic durability, CNT–PDMS strain sensors can reliably capture thoracic and abdominal expansions during different breathing states [13, 14]. However, respiration signals acquired in real-world settings remain challenging to interpret due to waveform variability and feature overlap, underscoring the need for systematic ML–based feature optimization and model selection [15–17].

Despite the rapid adoption of ML in wearable-based respiratory research [18, 19], most existing studies focus primarily on overall classification accuracy without systematically examining the role of feature engineering, classifier selection, and model calibration [20, 21]. Respiration waveforms contain rich temporal, statistical, amplitude-based, and frequency-domain characteristics, yet prior work rarely evaluates which feature types contribute most to distinguishing different breathing states [22]. Similarly, comprehensive comparisons across classifier families, including tree-based ensemble models, kernel-based methods, and probabilistic classifiers, remain limited [23]. Furthermore, the impact of calibration strategies on predicted probability outputs is often overlooked, despite its importance for applications involving real-time alerts or risk stratification [24, 25].

This study advances the current state of wearable respiratory monitoring by establishing a proof-of-concept, fully integrated strain sensor–to–machine learning pipeline that links deformation-induced resistance modulation in a stretchable strain sensor to discriminative waveform features. By integrating multimodal, engineered handcrafted descriptors with CNN-derived latent embeddings, we systematically determine which signal domains contribute most to class separation, which is an analysis largely missing in existing respiration-classification research. The incorporation of SHAP-based interpretability further enables a mechanistic connection between breathing-induced CNT network deformation and key waveform characteristics such as amplitude variability and signal-energy differences. Together, these results show that combining a stretchable CNT–PDMS strain sensor with a cross-validated, model-agnostic ML framework enables interpretable and proof-of-concept respiration-state classification for future wearable monitoring systems.

2 | Results and Discussion

2.1 | Machine Learning Optimization for Respiration Using Strain Sensor

Figure 1 illustrates the strain sensor-based respiration monitoring system, integrating thoracic breathing dynamics, strain sensing, and ML classification. Figure 1a presents a schematic illustration of respiration monitoring with the stretchable strain sensor and its photograph. Figure 1b shows schematics and SEM images of the stretchable strain sensor under mechanical stretching, showing the disruption of the CNT network. During respiration, abdominal expansion and contraction induce deformation of the strain sensor, which alters the conductive pathways and results in resistance changes, thereby enabling respiration monitoring. Figure 1c illustrates the feature-extraction strategies applied to the acquired respiration signals. Two complementary approaches were used: (i) a handcrafted feature pipeline, where each respiration waveform is transformed into four categories of quantitative descriptors–-statistical features (e.g., mean, and variance), energy-related metrics (e.g., root mean square (RMS), and sum of squares), frequency-domain features (e.g., Fast Fourier Transform (FFT) power, and spectral entropy), and peak- and temporal-based characteristics (e.g., peak intervals, and zero crossings); and (ii) a deep learning pipeline, where raw 1D respiration segments are processed through a CNN to learn latent representations directly from the waveform without manual engineering. Together, these feature sets capture amplitude variability, signal energy, spectral composition, temporal rhythm, and higher-level nonlinear patterns embedded in the strain sensor output. The CNN-based latent embeddings complement the handcrafted feature pipeline by capturing higher-order nonlinear temporal patterns and subtle waveform morphology that are difficult to encode through manually defined statistical descriptors. While handcrafted features provide physically interpretable representations grounded in signal characteristics such as amplitude, energy, and spectral content, CNN-derived embeddings offer a data-driven perspective that can identify abstract signal structures beyond the reach of predefined feature engineering, thereby broadening the representational capacity of the overall classification framework. Figure 1d summarizes the ML classifiers used to distinguish among breathing states based on these extracted features. Models include tree-based classifiers (e.g., decision tree, random forest, eXtreme Gradient Boosting (XGBoost), non-tree models (e.g., support vector machine (SVM), logistic regression, Naïve Bayes), and calibrated variants (e.g., sigmoid- and isotonic-regression models) for improved probability estimation. Collectively, Figure 1c,d describes the complete computational workflow–-from transforming raw physiological signals into meaningful features to applying ML algorithms for optimized respiration state classification.

2.2 | Strain Sensing Performance and Respiration Monitoring

Figure 2a presents the stress-strain curve of the stretchable strain sensor that features a network of CNTs and polydimethylsiloxane (PDMS), as shown in Figure S1. The stretchable strain sensor exhibited a mechanical modulus of 1.4 MPa and a tensile breaking strain of ~160% (Figure 2a). The relative resistance change (ΔR/R0) of the stretchable strain sensor increased under stretching due to the disconnection of CNT network, exhibiting high linearity ($R^2 > 0.99$) and a gauge factor of 0.94 (Figure 2b). Figure 2c shows the SEM image of the stretchable strain sensor, showing the CNT network after stretching. The disconnection of CNT pathways has increased electrical resistance, which constitutes the strain sensing mechanism (Figure 2c and Figure S2).

For practical wearable sensing, the stretchable strain sensor must demonstrate stable performance under repeated loading–unloading cycles [26–28]. Figure 2d shows the stretchable strain

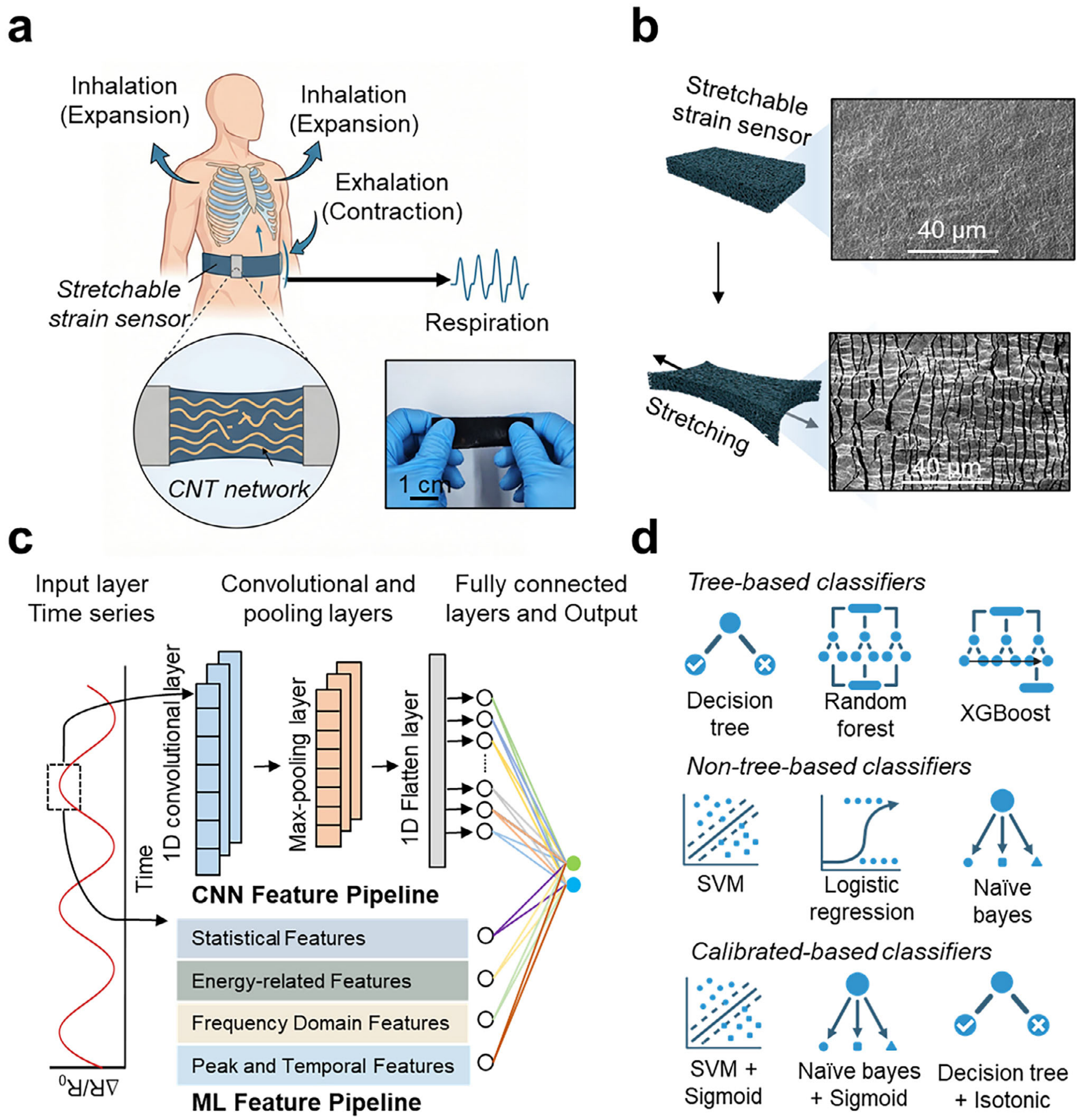

FIGURE 1 | Machine learning optimization for respiration using strain sensors. (a) Conceptual illustration and photograph of the stretchable strain sensor for respiration monitoring. (b) Schematics and SEM images of the stretchable strain sensor under stretching. (c) Latent features extracted from a 1D-CNN and handcrafted-derived ML feature-extraction pipeline, where respiration waveforms are transformed into four categories of quantitative descriptors: statistical features, energy-related features, frequency-domain features, and peak/temporal features. (d) ML classifiers evaluated in this study, including tree-based models (decision tree, random forest, XGBoost), non-tree models (SVM, logistic regression, Naïve Bayes), and calibrated variants (sigmoid- and isotonic-calibrated classifiers), used to classify breathing states based on extracted features.

sensor showed no notable alteration in $\Delta R/R_0$ against > 1000 cycles of stretching at an applied strain of up to 10%. Consistent strain sensing performance was further maintained at higher strain levels of up to 50% over 1000 cycles, confirming the mechanical durability and sensing reliability of the stretchable strain sensor under repeated loading-unloading cycles. (Figure S3). This durability is attributed to the high fracture toughness of CNTs [29]. The stretchable strain sensor can be stretchable, bendable, and twistable without compromising its material properties (Figure 2e).

The stretchable strain sensor was utilized to monitor the respiratory activity of a subject during inhalation and exhalation, enabling real-time detection of magnitude, intensity, and consistency of abdominal strain [30]. Respiration was recorded under three breathing states (e.g., shallow, normal, and deep breathing)

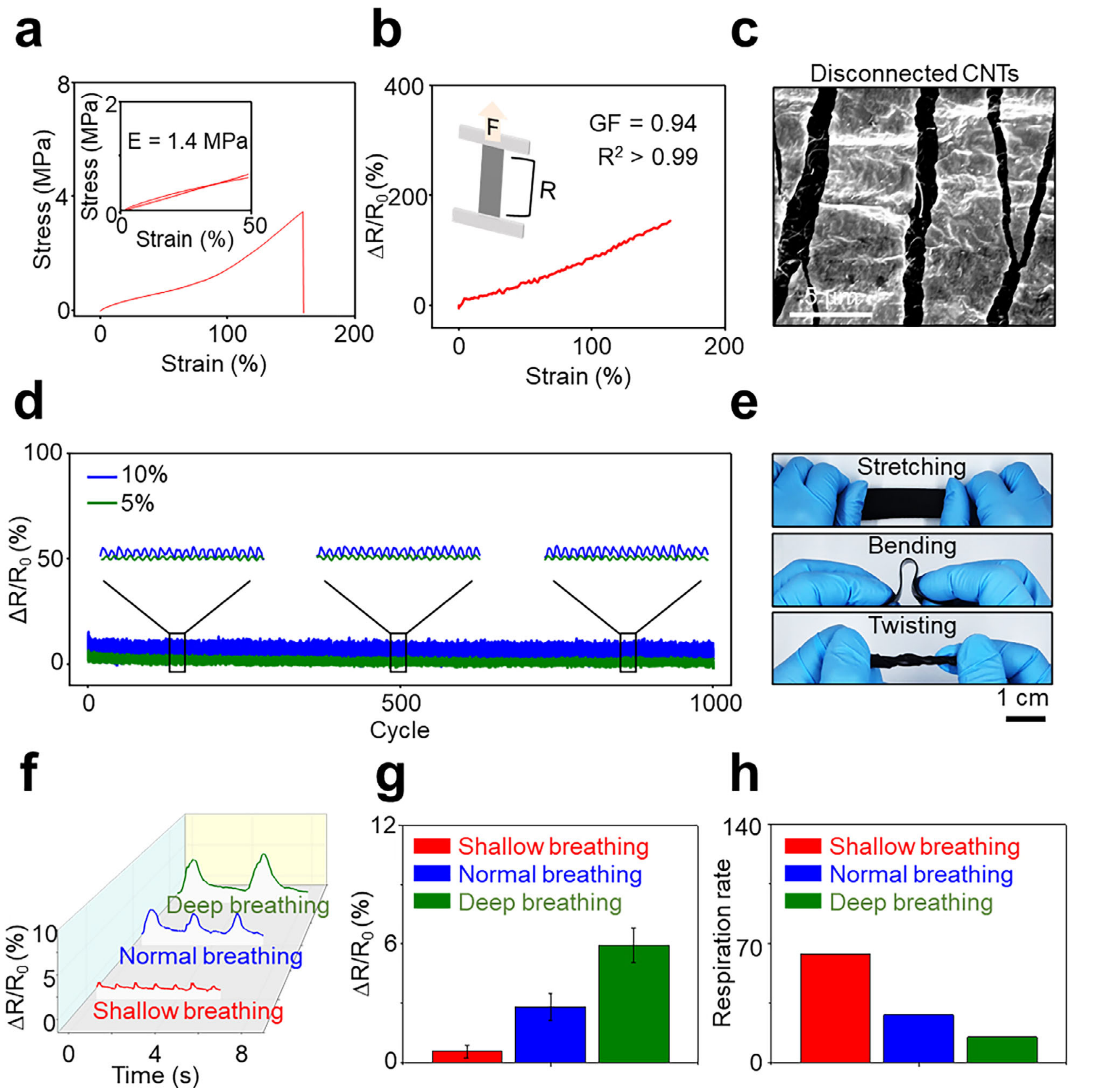

FIGURE 2 | Strain sensing performance and respiration monitoring. (a) Stress–strain curves of the stretchable strain sensor. (b) $\Delta R/R_0$ of the stretchable strain sensor under stretching. (c) SEM image of the stretchable strain sensor under stretching. (d) $\Delta R/R_0$ of the stretchable strain sensor throughout 1,000 stretching–releasing cycles with strains of 5% and 10%. (e) Photographs of the stretchable strain sensor under various mechanical deformations, including stretching, bending, and twisting. (f) Respiratory signals collected by the stretchable strain sensor under different breathing states (e.g., shallow, normal, and deep breathing). Bar graphs of (g) the relative change in respiratory signal amplitude ($\Delta R/R_0$) and (h) respiration rate across breathing status.

using the stretchable strain sensor from a single healthy adult participant, which showed respiration patterns associated with each breathing state (Figure 2f), as a proof-of-concept evaluation of sensor–algorithm coupling rather than population-level generalization. The respiration intensity and rate across each breathing state can be seen in Figure 2g,h. Specifically, shallow breathing exhibited an average respiration rate of 64 breaths per minute (bpm) with a signal amplitude of 0.6% ($\Delta R/R_0$). Under normal breathing, the respiration rate decreased to 28 bpm with an increased amplitude of 2.8%, while deep breathing further decreased the rate to approximately 15 bpm with an increase in amplitude of 5.9%.

2.3 | Feature Selection Reveals Deformation-Driven Signal Characteristics

Figure 3 summarizes the feature selection results across three breathing comparisons using the random forest classifier, which demonstrated the highest performance among all tested clas-

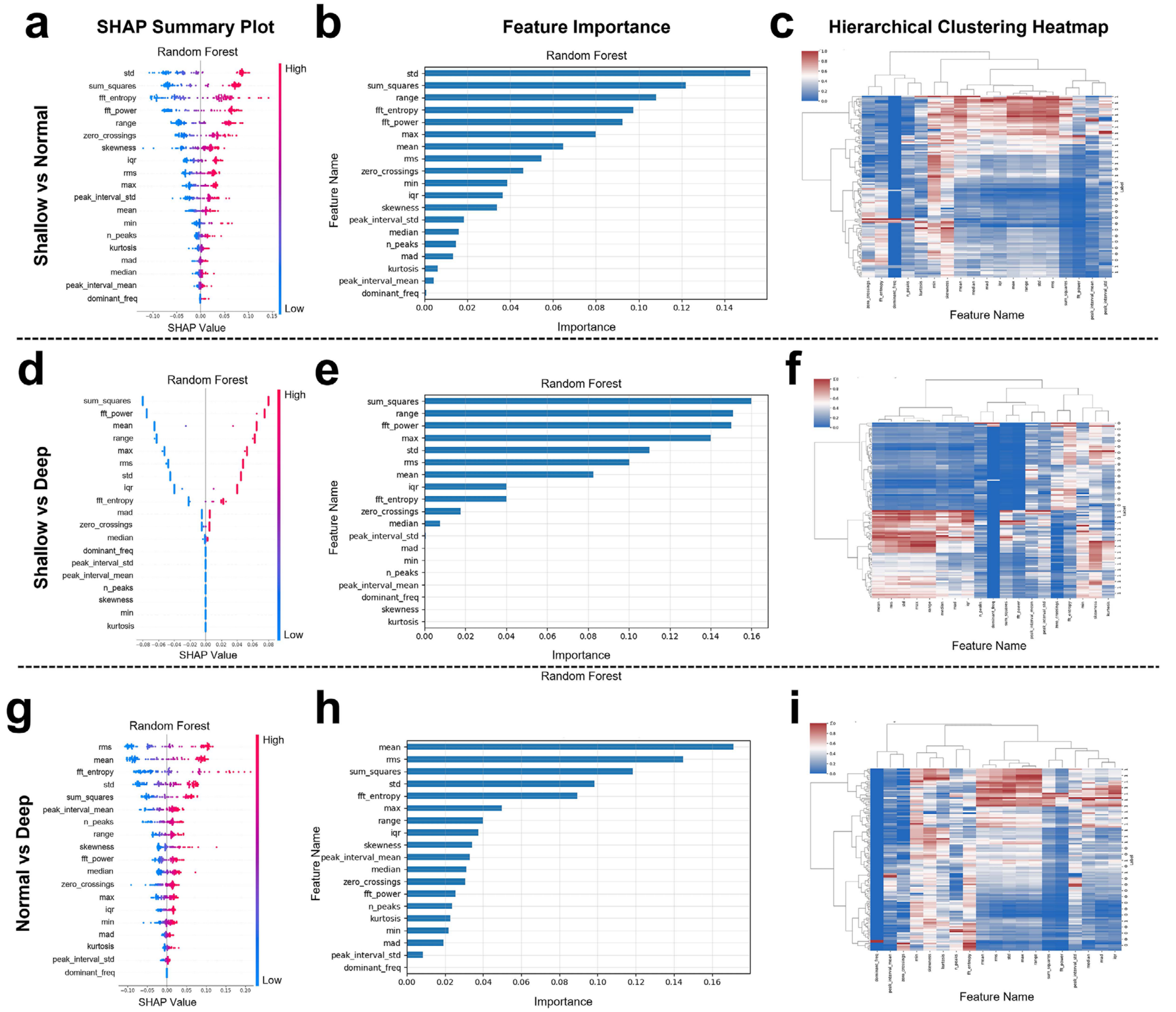

FIGURE 3 | Feature-selection results for the three binary breathing comparisons using the random forest classifier. (a–c) Shallow vs. Normal: SHAP summary plot (a), feature-importance ranking (b), and hierarchical clustering heatmap (c). Key discriminative features included standard deviation, sum of squares, range, FFT power, and FFT entropy, reflecting lower variability and signal energy in shallow breathing. (d–f) Shallow vs. Deep: SHAP plot (d), feature importance (e), and clustering heatmap (f). Energy- and amplitude-related metrics–-sum of squares, FFT power, RMS, maximum amplitude, and range–-dominated, capturing the markedly larger respiratory excursion of deep breathing. (g–i) Normal vs. Deep: SHAP plot (g), feature importance (h), and clustering heatmap (i). Mean amplitude, RMS, sum of squares, standard deviation, and FFT entropy were most influential, with deep breathing showing consistently higher-amplitude and higher-energy waveform patterns.

sifiers. For shallow versus normal breathing (Figure 3a–c and Figure S4), SHAP analysis and random forest feature importance both identified standard deviation, sum of squares, range, FFT power, and FFT entropy as the most influential features. These metrics reflect the reduced amplitude variability, lower signal energy, and diminished spectral complexity characteristic of shallow breathing. Hierarchical clustering confirmed clear separation between groups, indicating that shallow respiration exhibits consistently lower variability and energy patterns. In the shallow versus deep comparison (Figure 3d–f and Figure S5), signal energy and amplitude-related metrics dominated. Sum of squares, FFT power, range, maximum amplitude, and RMS ranked highest, reflecting the substantially greater respiratory excursion during deep breathing. SHAP values showed positive contributions toward deep breathing and negative contributions toward shallow breathing for these features. Hierarchical clustering revealed distinct separation, with deep breathing characterized by uniformly higher energy and amplitude values. For normal versus deep breathing (Figure 3g–i and Figure S6), amplitude and energy-related features remained the primary determinants. Mean, RMS, sum of squares, standard deviation, and FFT entropy ranked highest, reflecting the greater respiratory effort and signal magnitude associated with deep breathing. SHAP contributions were strongly positive toward deep breathing, while normal breathing showed lower values across these metrics. Hierarchical clustering confirmed distinct grouping, with deep respiration producing consistently higher-amplitude and higher-energy waveform patterns.

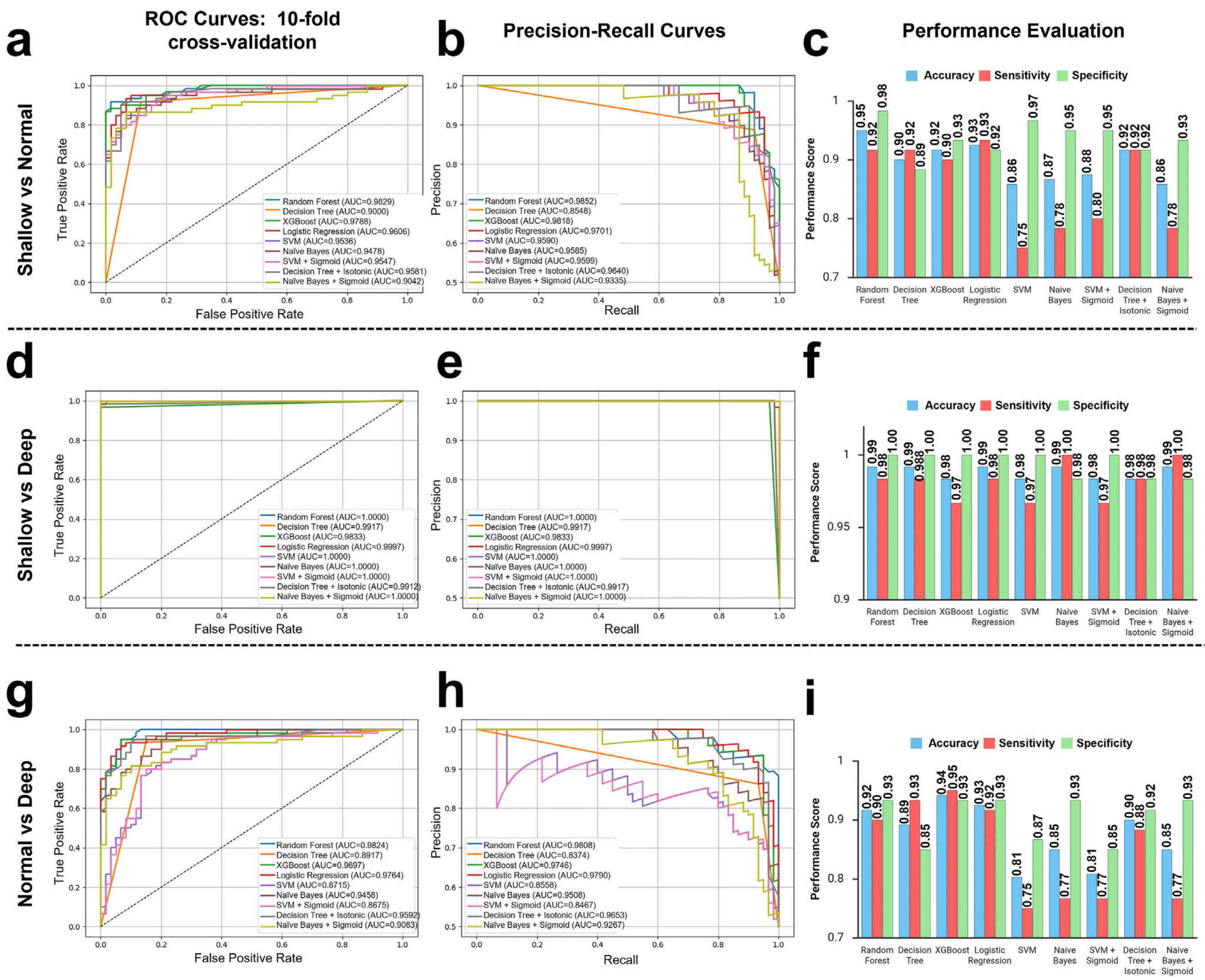

FIGURE 4 | Classification performance across the three respiration comparisons using multiple ML classifiers. (a–c) Shallow vs. Normal breathing: (a) ROC curves showing classifier discrimination ability with AUC values. Random forest achieved the highest AUC (0.9829), followed by XGBoost (0.9788) and logistic regression (0.9606). (b) PR curves demonstrating trade-offs between precision and recall. Random forest maintained the highest AP (0.9852). (c) Performance metrics comparing accuracy, sensitivity, and specificity across all classifiers. Random forest showed the most balanced performance (accuracy: 0.9500, sensitivity: 0.9167, specificity: 0.9833). (d–f) Shallow versus Deep breathing: (d) ROC curves showing near-perfect separation across all classifiers, with most models achieving AUC ≥ 0.9800. (e) PR curves demonstrating consistently high precision across the recall range for all models. (f) Performance metrics showing uniformly excellent classification with accuracy, sensitivity, and specificity all exceeding 0.97 for top-performing models. (g–i) Normal versus Deep breathing: (g) ROC curves revealing reduced discrimination compared to other comparisons. Random forest achieved the highest AUC (0.9824), followed by logistic regression (0.9764) and XGBoost (0.9697). (h) PR curves showing greater variability in precision-recall trade-offs. (i) Performance metrics indicating lower overall accuracy compared to previous comparisons, with random forest maintaining the highest balanced performance (accuracy: 0.92, sensitivity: 0.90, specificity: 0.93). All evaluations were conducted using ten-fold cross-validation. Calibrated models (sigmoid and isotonic) are indicated with "+" symbols.

2.4 | Classification Performance and Evaluation Across Breathing Conditions

For the shallow versus normal task (Figure 4a–c), the random forest model was the strongest performer within the tree-based family, achieving an accuracy of 0.9500 and the highest ROC AUC of 0.9829 (95% CI: 0.9674–0.9984), and an average precision (AP) of 0.9858. Among the non-tree models, logistic regression yielded the best performance, with an accuracy of 0.9250, a ROC AUC of 0.9606 (95% CI: 0.9320–0.9892), and an AP of 0.9701. Within the calibrated family, the decision tree with isotonic calibration performed best, reaching an accuracy of 0.9167, a ROC-AUC of 0.9581 (95% CI: 0.9339–0.9822), and an AP of 0.9640. When comparing across families, random forest exceeded the top non-tree and calibrated models in both accuracy and AUC—with a distinctly higher upper-bound AUC confidence limit—indicating that ensemble tree-based classifiers provided the most effective discrimination between shallow and normal breathing. Confusion matrices (Figure 5a–c and Figure S7) confirmed these performance differences. Random forest achieved

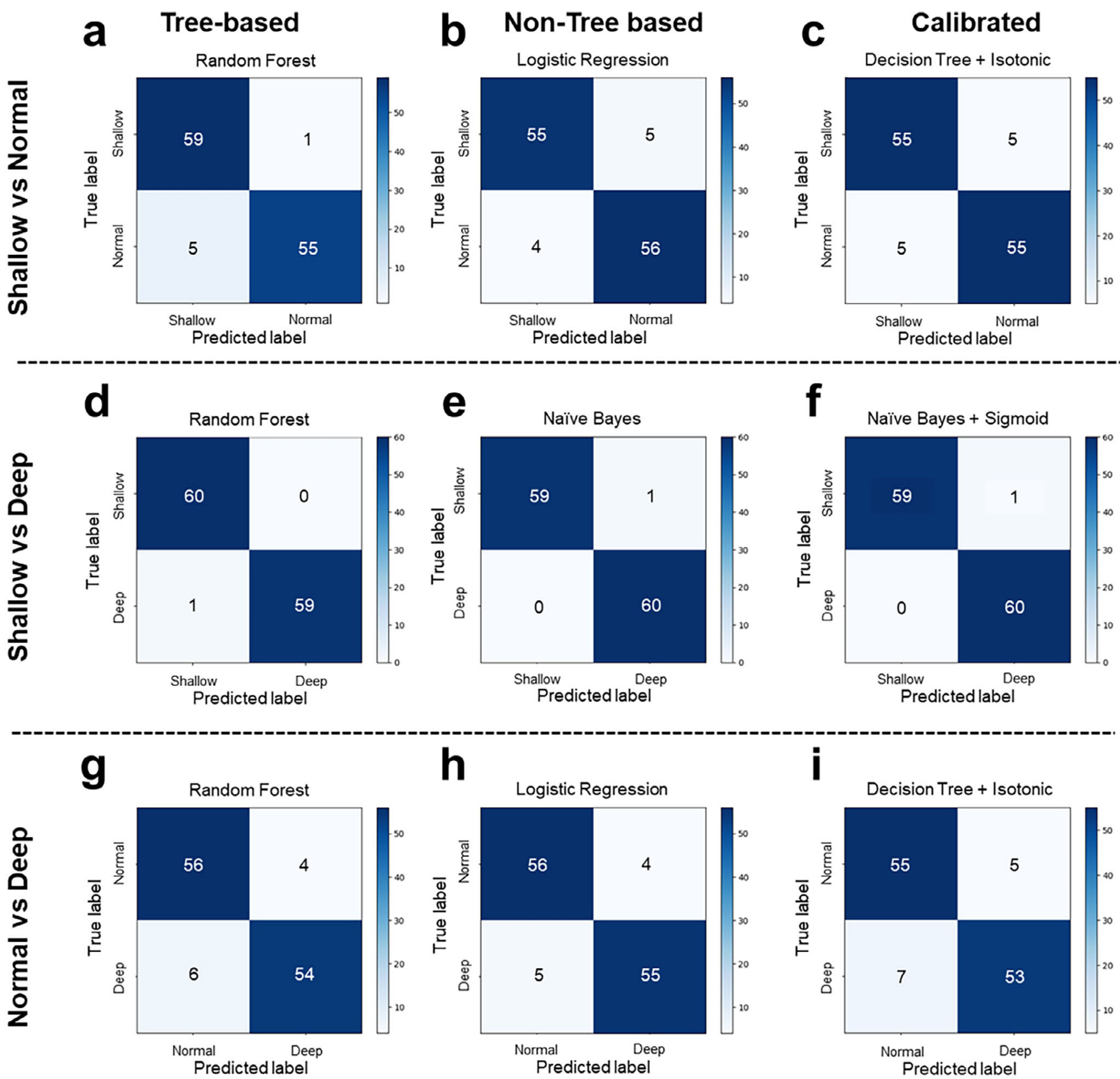

FIGURE 5 | Confusion matrices for the three respiration comparisons using representative classifiers. (a–c) Shallow versus Normal: Random forest showed the most balanced predictions (59–55 correct), while logistic regression and the isotonic-calibrated decision tree exhibited slightly higher misclassification rates. (d–f) Shallow versus Deep: All models demonstrated near-perfect classification. Random forest achieved 60/60 and 59/60 correct predictions, and Naïve Bayes (with or without sigmoid calibration) showed similarly high accuracy. (g–i) Normal vs. Deep: Error rates increased compared to shallow–deep, with random forest correctly classifying 56/60 normal and 54/60 deep breaths. Logistic regression and the isotonic-calibrated decision tree showed comparable but slightly lower performance.

the most balanced classification, correctly identifying 98.3% of shallow breaths (59/60) and 91.7% of normal breaths (55/60) with minimal cross-class confusion. Logistic regression showed slightly higher misclassification rates for shallow breaths (55/60, 91.7%) and normal breaths (56/60, 93.3%), consistent with its lower AUC. The isotonic-calibrated decision tree exhibited the largest errors among top-performing models, correctly classifying 91.7% for both shallow and normal breathing (55/60 each).

For the shallow versus deep comparison (Figure 4d–f), all model families achieved uniformly high discrimination, with only small differences in performance. Within the tree-based models, random forest performed best, with an accuracy of 0.9917, a ROC-AUC of 0.9999 (95% CI: 0.9993–0.9999), and an AP of 0.9999. Among the non-tree classifiers, logistic regression showed the strongest performance, matching the top accuracy and yielding a ROC-AUC of 0.9997 (95% CI: 0.9984–0.9998) and an AP of 0.9997. For the calibrated models, Naïve Bayes with sigmoid calibration performed highest, also reaching an accuracy of 0.9917, a ROC AUC of 0.9999 (95% CI: 0.9994–0.9999), and an AP of 0.9999. These uniformly high results are expected, as shallow and deep breathing exhibit markedly different waveform amplitudes and signal energy, making the two conditions inherently easier to discriminate. Confusion matrices (Figure 5d–f and Figure S8) showed uniformly strong performance across all classifiers, reflecting the large amplitude and energy differences between shallow and deep breathing. Random forest demonstrated near-perfect discrimination, correctly classifying 98.3% of both shallow and deep breaths (59/60 each) with only a single error per class. Naïve Bayes achieved 98.3% accuracy for shallow breaths (59/60) and 100% for deep breaths (60/60), indicating that probabilistic modeling effectively captured the clear separability between conditions. The sigmoid-calibrated Naïve Bayes model produced identical results.

For the normal versus deep comparison (Figure 4g–i), within the tree-based models, random forest achieved the highest performance, with an accuracy of 0.9167, a ROC-AUC of 0.9824 (95% CI: 0.9552–0.9987), and an AP of 0.9808. Among the non-tree models, logistic regression performed best, reaching an accuracy of 0.9250, a ROC-AUC of 0.9764 (95% CI: 0.9591–0.9937), and an AP of 0.9790. For calibrated classifiers, the decision tree with isotonic calibration showed the strongest results, achieving an accuracy of 0.90, a ROC AUC of 0.9592 (95% CI: 0.9163–0.9754), and an AP of 0.9653. Compared across families, random forest and logistic regression showed slightly lower performance than in the shallow-deep task, with random forest accuracy decreasing by 7.5% in accuracy and 1.8% in ROC-AUC, and logistic regression decreasing by 6.7% in accuracy and 2.3% in ROC-AUC, reflecting the smaller waveform differences between normal and deep breathing. Confusion matrices (Figure 5g–i and Figure S9) revealed higher misclassification rates than the shallow–deep task, reflecting more moderate waveform differences between normal and deep breathing. Random forest correctly identified 93.3% of normal breaths (56/60) and 90.0% of deep breaths (54/60). Logistic regression showed similar performance at 93.3% for normal breaths (56/60) and 91.7% for deep breaths (55/60), consistent with its slightly higher ROC-AUC. The isotonic-calibrated decision tree exhibited the largest errors, correctly identifying 91.7% of normal breaths (55/60) and 88.3% of deep breaths (53/60).

Calibration analysis further evaluated the reliability of predicted class probabilities across classifier families. For the shallow versus normal comparison, the decision tree with isotonic calibration performed best among calibrated models, achieving a ROC-AUC of 0.9581 and an accuracy of 0.9167. In the shallow versus deep task, Naïve Bayes with sigmoid calibration achieved the highest calibrated performance with an accuracy of 0.9917 and a ROC-AUC of 0.9999, consistent with the inherently high separability between these two conditions. For the normal versus deep comparison, the decision tree with isotonic calibration again led among calibrated classifiers, with a ROC-AUC of 0.9592 and an accuracy of 0.90. These results suggest that isotonic calibration more effectively corrects for overconfidence in tasks with moderate waveform overlap, whereas sigmoid calibration provides sufficient correction when class separation is more pronounced. Overall, the inclusion of calibrated classifiers confirms that post-hoc probability calibration can meaningfully improve the reliability of classifier outputs, which is particularly relevant for threshold-based alert systems in real-time wearable respiratory monitoring.

Across all tasks, random forest consistently outperformed other classifier families, which aligns with the underlying signal physics of strain-sensor respiration waveforms. Because breathing-induced CNT network deformation produces nonlinear changes in amplitude, energy, and temporal variability, the resulting feature space contains hierarchical and nonlinear interactions that tree-ensemble models naturally capture. Random forest leverages multiple decorrelated decision boundaries, enabling it to model amplitude–energy coupling and subtle distribution shifts more effectively than linear classifiers or single-tree models. This makes it particularly well suited for soft-material–based sensing systems, where physiological deformation propagates through inherently nonlinear electrical pathways.

2.5 | External Validation With Additional Subjects

To evaluate the generalizability of the proposed framework and to extend the binary classification approach to a unified multi-class setting, external validation was performed on three additional subjects (healthy young males; Subject 1, Subject 2, and Subject 3) under the same experimental protocol. A single Random Forest classifier trained on the original dataset was applied to each new participant independently, classifying all three respiration states (Shallow, Normal, and Deep) simultaneously. Subject-level separation was strictly enforced, with feature scaling fitted exclusively on the training data and applied without modification to each test subject. Each subject contributed 60 windows per breathing state (180 windows total per subject). As shown in Figure 6, the Random Forest classifier achieved classification accuracies of 91.1%, 90.6%, and 93.9% for Subject 1, Subject 2, and Subject 3, respectively, with a mean accuracy of 91.9% across all three external subjects. Classification accuracy for all nine classifiers across the three subjects is additionally summarized in Table 1.

3 | Conclusion

This work establishes a fully optimized ML framework for classifying respiration states using signals acquired from a stretchable strain sensor. By integrating handcrafted ML features with latent representations extracted from a 1D CNN, the study systematically evaluated classifier families and identified the most effective algorithms for distinguishing shallow, normal, and deep breathing. Random forest consistently demonstrated top performance across tasks, achieving accuracies of 95.0% (shallow vs. normal), 99.2% (shallow vs. deep), and 91.7% (normal vs. deep), along with correspondingly high AUC values. SHAP analysis and hierarchical clustering further revealed that the most discriminative features were those capturing amplitude variability and signal-energy differences. To further assess generalizability, external validation was conducted on three additional healthy young male subjects, with the Random Forest classifier achieving a mean accuracy of 91.9% across subjects, providing preliminary evidence that the framework may extend beyond a single-subject setting.

While these preliminary results support the feasibility of the proposed framework beyond a single-subject setting, several limi-

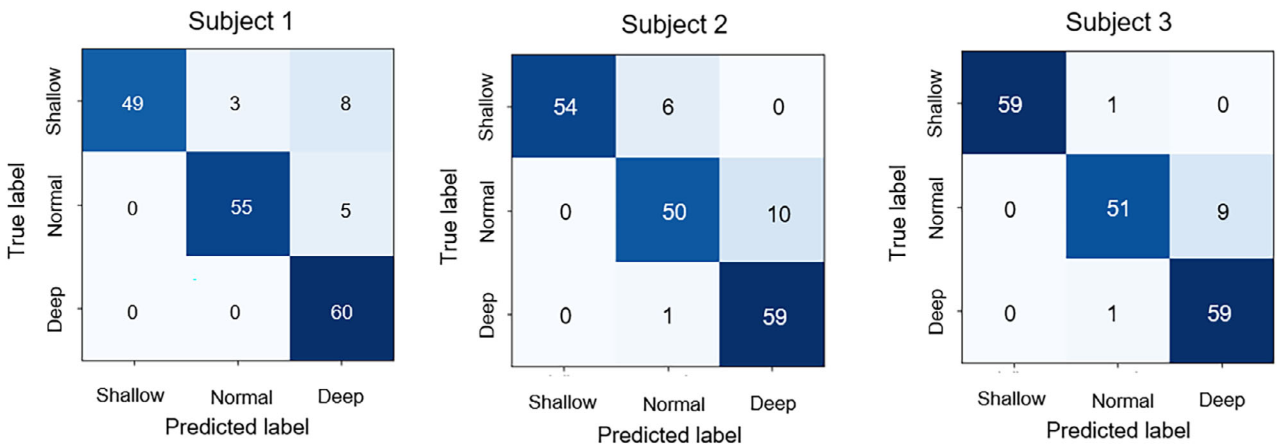

FIGURE 6 | External validation of the three-class respiration classification model across three additional subjects. Confusion matrices show the predicted versus true breathing states (Shallow, Normal, and Deep) for Subject 1, Subject 2, and Subject 3 using the Random Forest classifier trained on the original dataset.

TABLE 1 | Classification accuracy of nine machine learning models across three external validation subjects.

Model	Accuracy		
	Subject 1	**Subject 2**	**Subject 3**
Random Forest	0.9111	0.9056	0.9389
Decision Tree	0.9056	0.8222	0.8722
XGBoost	0.9078	0.8989	0.9178
Logistic Regression	0.7889	0.8778	0.8000
SVM	0.8556	0.9667	0.8944
Naive Bayes	0.7389	0.7000	0.7278
SVM + Sigmoid	0.8611	0.9333	0.8500
Decision Tree + Isotonic	0.9111	0.8722	0.8778
Naive Bayes + Sigmoid	0.7722	0.6944	0.7500

Note. All models were trained on the original single-subject dataset and tested independently on each external validation subject. Feature scaling was fitted exclusively on the training data and applied without modification to each test subject. Accuracy values represent the proportion of correctly classified windows across all three breathing states (Shallow, Normal, and Deep) out of 180 windows per subject (60 windows per state). SVM = Support Vector Machine.

tations should be acknowledged. The breathing protocol required subjects to perform deliberate, sustained breathing states, which does not reflect the continuous and unsegmented nature of real respiratory streams, where state transitions occur gradually and window boundaries cannot be defined a priori. Additionally, the study population was limited to healthy young males performing controlled breathing, and classification performance may not transfer to clinical populations where respiratory patterns differ substantially from healthy baselines, such as patients with obstructive or restrictive pulmonary conditions. Sensor placement reproducibility across sessions was not evaluated and represent additional sources of real-world variability that should be addressed before clinical deployment. Furthermore, repeated-session validation and direct comparison against a reference respiration-monitoring device were not performed in this study, representing further limitations that should be addressed in future work to establish the physiological validity and inter-session reproducibility of the proposed framework.

This study highlights how combining stretchable strain sensors with systematic ML optimization can advance real-time, non-invasive respiration monitoring for wearable and clinical applications. By revealing which signal features and classifier architectures most effectively encode breathing-induced deformation, the proposed framework offers a proof-of-concept approach for analyzing soft, stretchable sensing platforms. CNN-derived latent embeddings complement physics-informed features by capturing complex nonlinear waveform patterns, motivating future integration of explainable AI and robustness-aware modeling to support deployment under diverse environmental conditions. Together, these advances position the sensor-to-algorithm pipeline as a foundation for multimodal physiological monitoring. Future work should expand validation to larger and more diverse cohorts under real-world conditions to establish the robustness and clinical applicability of the proposed framework.

4 | Experimental Section

4.1 | Materials

The average diameter and length of the multi-walled carbon nanotubes (MWCNTs) (Cheaptubes. Inc.; purity > 95 wt.%) were measured as 10–20 nm and 10–30 μm, respectively. The MWCNTs were blended with isopropanol at a weight ratio of 0.5 wt.% and sonicated to produce a uniform ink. All human studies were conducted in accordance with university regulations and were approved by the Institutional Review Board (IRB, protocol #: 202212-009-03).

4.2 | Fabrication of the Stretchable Strain Sensor

Stretchable strain sensors were fabricated as illustrated in Figure S10. The MWCNTs were first pasted onto a glass substrate to form a conductive layer. An uncured PDMS precursor, prepared by mixing the base and curing agent at a weight ratio of 10:1, was then poured onto the MWCNT-coated glass substrate. Then, the PDMS was cured at 50°C overnight to obtain a CNT–PDMS composite. The resulting stretchable strain sensor was subsequently peeled off from the glass substrate.

4.3 | Mechanical and Electrical Characterizations

A mechanical testing system (Mark-10) was used to generate a stress-strain curve of the stretchable strain sensor. The resistance of the stretchable strain sensor was measured using a source meter (Keithley 2400; Tektronix, Inc.) for electrical testing. The structural details of the stretchable strain sensor were analyzed using a high-resolution SEM (Teneo Volumescope, Thermo Fisher Scientific). The respiration signals were sampled at 20 Hz.

4.4 | Gauge Factor

Gauge factor is defined by the following equation:

$$GF = (\Delta R/R_0)/\varepsilon$$

where R_0, ΔR, and ε are the initial resistance before stretching, difference of resistance under stretching, and applied strain, respectively.

4.5 | Structure of ML/CNN Algorithms

Prior to segmentation, raw respiration signals were preprocessed using a second-order Butterworth low-pass filter (cutoff = 3 Hz) to remove high-frequency noise, and motion-related outliers were suppressed using median absolute deviation (MAD)-based thresholding [6]. Respiration signals from the strain sensor were then segmented into 60 non-overlapping 6-second windows for each respiratory status and processed through two complementary feature extraction pipelines [31, 32]: a handcrafted ML-based approach and a CNN-based deep learning approach. In the ML-based statistical approach, each 6-second window was transformed into a structured feature vector consisting of four domains: (1) Statistical features, including mean, median, standard deviation, interquartile range (IQR), MAD, minimum, maximum, skewness, and kurtosis; (2) Energy-related features, such as RMS, sum of squares, and total FFT power; (3) Frequency-domain features, including dominant frequency and spectral entropy; (4) Peak and temporal features, including peak-to-peak amplitude, signal slope, number of peaks, inter-peak intervals, signal range, and zero-crossings. These handcrafted descriptors captured amplitude, variability, and frequency characteristics essential for differentiating between respiration states [33]. These feature domains were selected based on the underlying sensor physics: breathing-induced deformation of the CNT network produces nonlinear changes in resistance that manifest as variations in amplitude, signal energy, and spectral content. Accordingly, the chosen statistical, energy-, frequency-, and temporal-domain features directly reflect measurable physical responses of the strain sensor to respiratory motion.

In the CNN pipeline, raw respiration segments were passed through a 1D convolutional network. The network consisted of three Conv1D layers with 32, 64, and 128 filters (kernel size = 3, ReLU activation), each followed by max-pooling (pool size = 2). The extracted feature maps were flattened and fed into two fully connected layers (128 and 64 units) before entering a softmax classification head. Unlike the engineered descriptors, the CNN-derived latent embeddings capture higher-order and nonlinear waveform patterns, arising from CNT network rearrangements during breathing, that are not easily represented by handcrafted features. This architecture ensured consistent learning of hierarchical temporal features relevant to respiration-state discrimination [34, 35]. All handcrafted and CNN-derived feature inputs were finally normalized on a per-window basis using min–max scaling to the [0,1] range, ensuring consistent feature magnitudes across respiration states and recordings.

4.6 | ML Performance Evaluation

To ensure rigorous and leakage-free evaluation, all models were trained and tested using ten-fold stratified cross-validation, where each fold used 90% of the data for training and 10% for testing [36]. The model was retrained from scratch within each fold using only the training portion, and evaluated exclusively on the held-out test data. Feature extraction, including both handcrafted statistical descriptors and CNN-derived embeddings, was performed exclusively within each training split to avoid information leakage. Respiration signals were segmented into 60 non-overlapping 6-second windows per breathing state at a sampling rate of 20 Hz, yielding 120 data points per window. All preprocessing steps, including min-max feature scaling, handcrafted feature extraction, and CNN embedding generation, were performed strictly within each cross-validation fold using only the training partition, and applied without refitting to the held-out test fold to prevent data leakage. Classifier hyperparameters were configured as follows: Random Forest used 300 estimators with square-root feature sampling and a minimum of two samples per split; Decision Tree used Gini impurity as the splitting criterion with no maximum depth constraint; Logistic Regression used a regularization strength of C = 0.5 with a maximum of 1000 iterations; SVM used an RBF kernel with C = 1.0 and gamma = 'scale' with probability estimation enabled; XGBoost used 300 estimators, a maximum tree depth of 6, and a learning rate of 0.3 with log loss as the evaluation metric; Naïve Bayes used a variance smoothing factor of 1e-9. Calibrated variants were fitted using 5-fold internal cross-validation with ensemble averaging. All models used random_state = 42 for reproducibility.

A diverse suite of ML classifiers was evaluated to determine which modeling approach best captures the characteristics of strain sensor–derived respiration waveforms. Tree-based models (e.g., random forest, decision tree, and XGBoost) were included due to their ability to capture non-linear feature interactions and handle multi-scale amplitude variations inherent in respiratory signals without requiring explicit feature scaling or normalization [23, 24]. Random forest and XGBoost additionally provide ensemble robustness through aggregation of multiple decision boundaries, reducing overfitting risk in high-dimensional feature spaces.

Non-tree models (e.g., SVM, logistic regression, and Naïve Bayes) were selected for their complementary strengths [37, 38]: SVM establishes optimal hyperplane separation in transformed feature spaces, making it effective for margin-based discrimination of overlapping breathing states; Logistic regression offers interpretable probabilistic outputs and computational efficiency for real-time deployment; Naïve Bayes provides a probabilistic baseline that assumes feature independence, serving as a benchmark to assess whether feature correlations significantly impact performance.

Calibrated variants (e.g., sigmoid- and isotonic-calibrated models) were included to improve the reliability of predicted class probabilities, which is essential for threshold-based alert systems used in real-time respiration monitoring. Without proper calibration, classifiers may output overconfident or poorly calibrated probability estimates, potentially triggering false alarms or missed detections in wearable applications [39, 40].

Model performance was quantified using accuracy, sensitivity, specificity, and ROC-AUC, with additional evaluation using precision-recall (PR) curves and confusion matrices to characterize class-wise behavior [41, 42]. To interpret model behavior and identify discriminative features, SHAP value analyses and feature importance were performed for tree-based models, and hierarchical clustering was used to visualize feature-level separability across breathing conditions [43, 44]. All preprocessing, feature extraction, and classification procedures were implemented in Python using NumPy, SciPy, and Scikit-learn [45, 46].

Author Contributions

Y.L., and S.H. contributed equally to this work. Y.L. and S.H. performed conceptualization; data curation, methodology; validation; formal analysis; visualization; Writing – review and editing; writing – original draft; project administration; software. J.J. and D.R.K. performed methodology and visualization. C.H.L. and R.S. performed conceptualization; methodology; formal analysis; formal investigation; writing – review and editing; funding acquisition; project administration; resources; supervision. All authors made substantial contributions to the manuscript and approved the final version.

Acknowledgments

C.H.L. acknowledges support from the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea, through the Global Industrial Technology Cooperation Center (GITCC) Program, supervised by the Korea Institute for Advancement of Technology (KIAT) (Task No. P0028319), as well as from the University Faculty Scholar Endowment.

Conflicts of Interest

The authors state that they have no conflicts of interest.

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. C. Massaroni, A. Nicolò, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, "Contact-based Methods for Measuring Respiratory Rate," *Sensors* 19, no. 4 (2019): 908, https://doi.org/10.3390/s19040908.

2. A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, "The Importance of respiratory Rate Monitoring: From Healthcare to Sport and Exercise," *Sensors* 20, no. 21 (2020): 6396.

3. H. C. Ates, A. Brunauer, F. von Stetten, et al., "Integrated Devices for Non-invasive Diagnostics," *Advanced Functional Materials* 31, no. 15 (2021): 2010388.

4. F. Salaorni, G. Bonardi, F. Schena, M. Tinazzi, and M. Gandolfi, "Wearable Devices for Gait and Posture Monitoring via Telemedicine in People With Movement Disorders and Multiple Sclerosis: A Systematic Review," *Expert Review of Medical Devices* 21, no. 1–2 (2024): 121–140, https://doi.org/10.1080/17434440.2023.2298342.

5. J. Kim, A. S. Campbell, B. E.-F. de Ávila, and J. Wang, "Wearable Biosensors for Healthcare Monitoring," *Nature Biotechnology* 37, no. 4 (2019): 389–406, https://doi.org/10.1038/s41587-019-0045-y.

6. M. Chu, T. Nguyen, V. Pandey, et al., "Respiration Rate and Volume Measurements Using Wearable Strain Sensors," *NPJ Digital Medicine* 2, no. 1 (2019): 8, https://doi.org/10.1038/s41746-019-0083-3.

7. O. Atalay, W. R. Kennon, and E. Demirok, "Weft-knitted Strain Sensor for Monitoring Respiratory Rate and Its Electro-mechanical Modeling," *IEEE Sensors Journal* 15, no. 1 (2014): 110–122.

8. R. Wang, L. Sun, X. Zhu, et al., "Carbon Nanotube-Based Strain Sensors: Structures, Fabrication, and Applications," *Advanced Materials Technologies* 8, no. 1 (2023): 2200855.

9. J. H. Lee, K. Cho, and J. K. Kim, "Age of Flexible Electronics: Emerging Trends in Soft Multifunctional Sensors," *Advanced Materials* 36, no. 16 (2024): 2310505.

10. S. Hong, T. Park, Y. Lee, et al., "Smart Garment for Continuous Respiration Monitoring in Canines," *ACS sensors* 11, no. 3 (2026): 2111–2120, https://doi.org/10.1021/acssensors.5c03783.

11. X. Wang, J. Li, H. Song, H. Huang, and J. Gou, "Highly Stretchable and Wearable Strain Sensor Based on Printable Carbon Nanotube Layers/Polydimethylsiloxane Composites With Adjustable Sensitivity," *ACS Applied Materials & Interfaces* 10, no. 8 (2018): 7371–7380.

12. M. O. Tas, M. A. Baker, M. G. Masteghin, J. Bentz, K. Boxshall, and V. Stolojan, "Highly Stretchable, Directionally Oriented Carbon Nanotube/PDMS Conductive Films With Enhanced Sensitivity as Wearable Strain Sensors," *ACS Applied Materials & Interfaces* 11 (2019): 39560–39573, https://doi.org/10.1021/acsami.9b13684.

13. M. Liu, Y. Wang, J. Chen, et al., "Non-Invasive Flexible Electro-Mechanical Sensors for Human Respiratory Monitoring and Chronic Disease Management," *Advanced Materials Technologies* 9, no. 10 (2024): 2302010, https://doi.org/10.1002/admt.202302010.

14. M. Amjadi, K.-U. Kyung, I. Park, and M. Sitti, "Stretchable, Skin-Mountable, and Wearable Strain Sensors and Their Potential Applications: A Review," *Advanced Functional Materials* 26, no. 11 (2016): 1678–1698, https://doi.org/10.1002/adfm.201504755.

15. P. H. Charlton, D. A. Birrenkott, T. Bonnici, et al., "Breathing Rate Estimation From the Electrocardiogram and Photoplethysmogram: A Review," *IEEE Reviews in Biomedical Engineering* 11 (2018): 2–20.

16. T. Pereira, N. Tran, K. Gadhoumi, et al., "Photoplethysmography Based Atrial Fibrillation Detection: A Review," *NPJ Digital Medicine* 3, no. 1 (2020), https://doi.org/10.1038/s41746-019-0207-9.

17. S. Hassani and U. Dackermann, "A Systematic Review of Advanced Sensor Technologies for Non-destructive Testing and Structural Health Monitoring," *Sensors* 23, no. 4 (2023): 2204.

18. V. Ravichandran, B. Murugesan, and V. Balakarthikeyan, "RespNet: A Deep Learning Model for Extraction of Respiration From Photoplethys-

mogram," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering In Medicine And Biology Society (EMBC)* (IEEE, 2019).

19. K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Estimation of Respiration Rate From ECG, BP and PPG Signals Using Empirical Mode Decomposition," in *Proceedings of the 2011 IEEE International instrumentation and measurement technology conference* (IEEE, 2011).

20. K. S. Moon and S. Q. Lee, "A Wearable Multimodal Wireless Sensing System for respiratory Monitoring and Analysis," *Sensors* 23, no. 15 (2023): 6790, https://doi.org/10.3390/s23156790.

21. A. Chen, J. Zhang, L. Zhao, et al., "Machine-learning Enabled Wireless Wearable Sensors to Study Individuality of respiratory Behaviors," *Biosensors and Bioelectronics* 173 (2021): 112799, https://doi.org/10.1016/j.bios.2020.112799.

22. M. Folke, L. Cernerud, M. Ekström, and B. Hök, "Critical Review of Non-invasive Respiratory Monitoring in Medical Care," *Medical & Biological Engineering & Computing* 41, no. 4 (2003): 377–383, https://doi.org/10.1007/BF02348078.

23. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?," *Journal of Machine Learning Research* 15, no. 1 (2014): 3133–3181.

24. A. Niculescu-Mizil and R. Caruana, "Predicting Good Probabilities With Supervised Learning," in *Proceedings of the Proceedings of the 22nd International Conference on Machine Learning* (2005), 625–832.

25. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in *Proceedings of the International Conference On Machine Learning* (PMLR, 2017), 1321–1330.

26. S. Hong, T. Park, J. Lee, et al., "Tough Conductive Organohydrogel for Wearable Sensing in Extreme Environmental Conditions," *Advanced Materials Technologies* 9, no. 2 (2024): 2301398, https://doi.org/10.1002/admt.202301398.

27. S. Hong, T. Yu, Z. Wang, and C. H. Lee, "Biomaterials for Reliable Wearable Health Monitoring: Applications in Skin and Eye Integration," *Biomaterials* 314 (2025): 122862.

28. S. Hong, J. Lee, T. Park, et al., "Spider Silk-Inspired Conductive Hydrogels for Enhanced Toughness and Environmental Resilience via Dense Hierarchical Structuring," *Advanced Science* 12, no. 12 (2025): 2500397, https://doi.org/10.1002/advs.202500397.

29. A. S. Dahiya, T. Gil, J. Thireau, et al., "1D Nanomaterial-Based Highly Stretchable Strain Sensors for Human Movement Monitoring and Human–Robotic Interactive Systems," *Advanced Electronic Materials* 6, no. 10 (2020): 2000547, https://doi.org/10.1002/aelm.202000547.

30. S. Hong, H. Zhang, J. Lee, et al., "Spongy Ag Foam for Soft and Stretchable Strain Gauges," *ACS Applied Materials & Interfaces* 16, no. 20 (2024): 26613–26623, https://doi.org/10.1021/acsami.4c04719.

31. A. K. Kumar, M. Ritam, L. Han, S. Guo, and R. Chandra, "Deep Learning for Predicting Respiratory Rate From Biosignals," *Computers in Biology and Medicine* 144 (2022): 105338, https://doi.org/10.1016/j.compbiomed.2022.105338.

32. K. McClure, B. Erdreich, J. H. Bates, R. S. McGinnis, A. Masquelin, and S. Wshah, "Classification and Detection of Breathing Patterns With Wearable Sensors and Deep Learning," *Sensors* 20, no. 22 (2020): 6481, https://doi.org/10.3390/s20226481.

33. O. Karaarslan, K. D. Belcastro, and O. Ergen, "Respiratory Sound-base Disease Classification and Characterization With Deep/Machine Learning Techniques," *Biomedical Signal Processing and Control* 87 (2024): 105570, https://doi.org/10.1016/j.bspc.2023.105570.

34. S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, "1-D Convolutional Neural Networks for Signal Processing Applications," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2019), 8360–8364.

35. S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time Patient-specific ECG Classification by 1-D Convolutional Neural Networks," *IEEE Transactions on Biomedical Engineering* 63, no. 3 (2015): 664–675.

36. T. Fushiki, "Estimation of Prediction Error by Using K-fold Cross-validation," *Statistics and Computing* 21, no. 2 (2011): 137–146, https://doi.org/10.1007/s11222-009-9153-8.

37. P. Golpour, M. Ghayour-Mobarhan, A. Saki, et al., "Comparison of Support Vector Machine, Naïve Bayes and Logistic Regression for Assessing the Necessity for Coronary Angiography," *International Journal of Environmental Research and Public Health* 17, no. 18 (2020): 6449.

38. A. B. Musa, "Comparative Study on Classification Performance Between Support Vector Machine and Logistic Regression," *International Journal of Machine Learning and Cybernetics* 4, no. 1 (2013): 13–24, https://doi.org/10.1007/s13042-012-0068-x.

39. M. Kull, T. M. Silva Filho, and P. Flach, "Beyond Sigmoids: How to Obtain Well-calibrated Probabilities From Binary Classifiers With Beta Calibration," *Electronic Journal of Statistics* 11 (2017): 5052–5080, https://doi.org/10.1214/17-EJS1338SI.

40. X. Jiang, M. Osl, J. Kim, and L. Ohno-Machado, "Smooth Isotonic Regression: A New Method to Calibrate Predictive Models," *AMIA Summits on Translational Science Proceedings* 2011 (2011): 16.

41. M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *International Journal of Data Mining & Knowledge Management Process* 5, no. 2 (2015): 01–11.

42. S. Narkhede, "Understanding AUC-ROC curve," *Data Science* 26 (2018): 220.

43. F. Murtagh and P. Contreras, "Algorithms for Hierarchical Clustering: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2 (2012): 86–97.

44. Z. Li, "Extracting Spatial Effects From Machine Learning Model Using Local Interpretation Method: An Example of SHAP and XGBoost," *Computers, Environment and Urban Systems* 96 (2022): 101845, https://doi.org/10.1016/j.compenvurbsys.2022.101845.

45. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

46. W. McKinney, "pandas: A Foundational Python Library for Data Analysis and Statistics," *Python for High Performance and Scientific Computing* 14, no. 9 (2011): 1–9.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: adsr70188-sup-0001-SuppMat.docx.