

Supervised Classification of Respiratory Ciliary Motion from Video Microscopy

Camilly Costa, Victor Rodrigues, Joao C W A Costa and Moisés Silva

Abstract—Reliable assessment of respiratory ciliary motion is essential for evaluating mucociliary clearance and supporting the diagnosis of respiratory dysfunctions such as Primary Ciliary Dyskinesia (PCD). Current analysis relies on expert visual inspection of high-speed videos, which is subjective and difficult to scale. This study proposes a supervised, phenotype-driven framework for classifying ciliary motion from video microscopy, comparing two spatial preprocessing strategies: restricting feature extraction to the masked ciliated region and using the entire image, as well as different motion descriptors. Basal motility versus hypomotility was best classified with masked frequency descriptors and Support Vector Machine with Radial Basis Function (SVM-RBF), reaching 78.06% balanced accuracy. Hypomotility versus secondary ciliary alteration required combined frequency and autoregressive descriptors, reaching 78.81% balanced accuracy with logistic regression, while Random Forest achieved the highest macro F1-score of 79.24%. Thus, the optimal pipeline is phenotype-specific: masked frequency descriptors best separate basal motility from hypomotility, whereas combined spectral and temporal descriptors better distinguish abnormal phenotypes.

Keywords—Respiratory ciliary motion, high-speed video microscopy, supervised classification, signal processing.

I. INTRODUCTION

Reliable assessment of respiratory ciliary motion is clinically crucial because impaired mucociliary clearance is the hallmark of severe respiratory conditions, most notably Primary Ciliary Dyskinesia (PCD) [1]. The lack of timely and accurate diagnosis of PCD can lead to progressive and irreversible lung damage [1]. Therefore, quantitative and reproducible assessment of ciliary motion is important for supporting clinical decision-making and improving the characterization of respiratory dysfunctions.

Currently, assessment of ciliary motion by high-speed video microscopy relies heavily on expert visual inspection [1], [2]. However, this manual approach is subjective, time-consuming, and difficult to scale, highlighting the need for automated diagnostic methods [2], [3]. Automating this analysis remains challenging because effective mucociliary clearance depends not only on the presence of cilia, but also on the spatial and temporal organization of their beating. Functionally relevant abnormalities may therefore arise from changes in coordination and motion pattern, rather than from frequency alone [3].

Camilly Costa, Victor Rodrigues, and João C W A Costa, Universidade Federal do Pará (UFPA), Belém-PA, Brazil, e-mails: camilly.costa@itec.ufpa.br, victorhrc20@gmail.com, joaoweyl@gmail.com; Moisés Silva, Los Alamos National Laboratory, Los Alamos-NM, USA, e-mail: mfelipe@lanl.gov.

For this reason, high-speed video microscopy provides a valuable basis for computational analysis aimed at identifying physiologically meaningful alterations in ciliary motion. Figure 1 summarizes the respiratory epithelial structures involved in mucociliary clearance and provides the biological context for the motion analysis performed in this study.

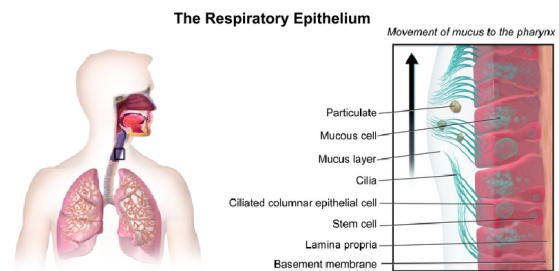


Fig. 1. Schematic representation of the respiratory epithelium, illustrating key structural components involved in mucociliary clearance. Adapted from BruceBlasus/Blasus.com staff, Blasus Medical 2014 [4].

This study investigates supervised classification of respiratory ciliary motion using the open Video Dataset of Respiratory Cilia Motion Phenotypes, with a focus on evaluating different feature extraction methods for improving classification performance [5]. Two biologically relevant pairwise comparisons are considered: basal motility versus hypomotility, and hypomotility versus secondary ciliary alteration. These comparisons assess whether computational descriptors can distinguish preserved motion from reduced motility and, more challengingly, separate distinct abnormal motion phenotypes.

The analysis compares masked and unmasked spatial preprocessing, frequency and autoregressive descriptors, and three supervised classifiers. This design allows the influence of preprocessing, feature representation, and model choice to be evaluated under a consistent phenotype-driven framework.

Related studies have used dynamic-texture descriptors, 3D convolutional neural networks (CNNs), and unsupervised spatiotemporal representations [2], [6], [7]; Li documented the present mouse-trachea dataset and a Gabor-enhanced U-Net segmentation approach [5], [8]. Here, masking, feature representation, and classifier choice are compared under a common phenotype-driven protocol.

II. MATERIALS AND METHODS

A. Study design and dataset

The open *Video Dataset of Respiratory Cilia Motion Phenotypes* was used in this study [5], [8]. It contains 872 high-speed video microscopy recordings from

mouse tracheal samples under different stimuli, including temperature-treatment recordings at 18, 24, 26, and 29 °C and room-temperature controls at 22 °C. Ground-truth masks label the ciliated area. Each video was recorded at 250 frames per second for 1 s, resulting in 250 frames per sample. The spatial resolution is 800×800 pixels, with a pixel size of $0.07987 \mu\text{m}$. A CSV file provides the ciliary beat frequency (CBF) values.

The masks identify the ciliated region of interest and were used as binary spatial supports. The ten experimental groups are control, temperature treatment, trifluoperazine (TFP) treatment, terbutaline treatment, N-acetylcysteine (NAC) treatment, Ca treatment, AZL treatment, adenosine triphosphate (ATP) treatment, amphotericin B treatment, and angiotensin-converting enzyme 2 (ACE2) manipulation [5]. The supervised class labels were derived from these original groups.

The analysis was structured around experimental groups derived from the original treatment categories. The core study considered three central groups: (1) basal motility, represented by control samples; (2) experimental hypomotility, represented by trifluoperazine-treated samples, consistent with the reported reduction of baseline ciliary beat frequency after trifluoperazine exposure [9]; and (3) secondary ciliary alteration, represented by temperature-treated samples, since the dataset documents temperature-dependent ciliary beat frequency [8].

From these groups, two core biological comparisons were defined to address the most relevant diagnostic challenges: basal motility versus hypomotility, which evaluates the detection of direct functional loss, and hypomotility versus secondary ciliary alteration, which evaluates the distinction between different abnormal motion states.

Additional scenarios were examined to provide broader context, including basal motility versus secondary alteration, extended hypomotility, grouped abnormal classes, exploratory hyperactivity, exploratory secondary alteration, and multiclass settings. However, the main interpretation of the study remained centered on the two core biological comparisons because they yielded the clearest biological interpretation and the most informative methodological comparison.

B. Overview of the computational pipeline

The complete analytical workflow is summarized in Figure 2. For each phenotypic comparison, the videos were processed through the same sequence of steps, allowing preprocessing, feature representation, and classifier choice to be compared under a consistent evaluation protocol.

First, the spatial support of the signal was defined using two alternatives: analysis restricted to the ciliary-region mask and analysis without spatial masking. Next, the videos were transformed into temporal signals and processed using the Hilbert transform, principal component analysis (PCA), and blind source separation (BSS) to obtain representative motion components.

This signal-processing sequence was motivated by established frameworks for structural motion estimation,

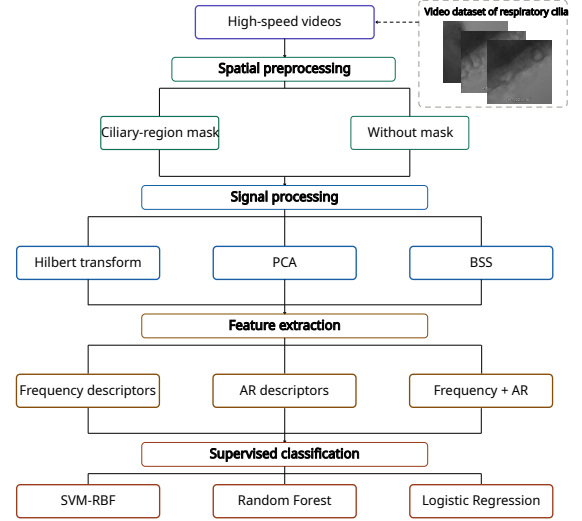


Fig. 2. Overview of the computational pipeline, detailing the workflow from spatial preprocessing and signal transformation to feature extraction and supervised classification.

which show that phase-based video transformations combined with modal identification can isolate subtle and independent kinematic patterns from complex visual scenes [10], [11]. Alternative feature representations were then extracted from these components, including frequency descriptors, autoregressive descriptors, and their combination. Finally, these feature sets were supplied to three supervised classifiers—SVM-RBF, Random Forest, and Logistic Regression—and evaluated using stratified cross-validation.

This workflow made it possible to examine how each design choice contributed to classification performance while preserving a consistent interpretation of the phenotypic comparisons under study.

C. Preprocessing

Each video was read frame by frame and reshaped into a space-time matrix, with one row per selected pixel and one column per frame; masked analyses retained only pixels within the segmented ciliated region. The Hilbert transform was applied along time, followed by PCA on the real and imaginary representations and BSS with the Complexity Pursuit algorithm [10], [11], [12], [13]. The resulting temporal components (modal coordinates) were used for frequency and autoregressive feature extraction.

The pipeline was implemented in Python. NumPy and SciPy were used for numerical and signal-processing operations, OpenCV for video reading and mask application, and scikit-learn for PCA, feature standardization, supervised classification, and stratified cross-validation.

D. Feature extraction

The feature-extraction stage compared different computational representations of ciliary motion derived from the same processed signals. The first feature class comprised frequency descriptors, defined as the dominant frequencies extracted from the representative temporal

components of each video. These descriptors were selected to capture the basic oscillatory speed of the cilia and to provide a direct measure for detecting reductions in movement, such as hypomotility.

The second feature class comprised autoregressive (AR) descriptors, obtained from autoregressive coefficients estimated from modal coordinates. These features were used to model broader temporal organization and coordination patterns that frequency descriptors alone may not capture.

A third feature set was built by combining frequency and AR descriptors into a single feature vector. This combined representation was intended to integrate complementary spectral and temporal-dynamic information, especially for distinguishing between more complex abnormal phenotypes.

This process produced six representation settings for each experiment: (1) frequency descriptors with mask, (2) frequency descriptors without mask, (3) AR descriptors with mask, (4) AR descriptors without mask, (5) combined descriptors with mask, and (6) combined descriptors without mask.

E. Supervised classification

Three classifiers were evaluated: SVM-RBF, Random Forest, and logistic regression. Feature vectors were standardized with `StandardScaler` (Z-score normalization) within each training fold, and all classifiers used `class_weight=balanced` to account for unequal class distributions.

Hyperparameters were fixed a priori using the values reported below. SVM-RBF used $C = 1.0$, $\gamma = \text{scale}$, and the RBF kernel. Random Forest used `n_estimators=100`, `max_depth=None`, and `random_state=42`. Logistic regression used `solver=liblinear`, L2 penalty, `max_iter=1000`, and `random_state=42`. Other settings retained scikit-learn defaults, and no probability calibration was applied.

For each phenotypic comparison, each classifier was evaluated separately with the alternative feature representations, so performance reflects the combined effects of preprocessing, feature representation, and model choice.

F. Evaluation protocol

Model performance was assessed with stratified five-fold cross-validation using shuffled folds and a fixed random seed (`random_state=42`). Out-of-fold predictions were aggregated, and performance was measured using accuracy (Acc), balanced accuracy (BA), and macro F1-score ($F1_{\text{macro}}$), defined in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

$$\begin{aligned} \text{Acc} &= \frac{TP+TN}{TP+TN+FP+FN}, \\ \text{BA} &= \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right), \\ F1_{\text{macro}} &= \frac{1}{C} \sum_{i=1}^C \frac{2TP_i}{2TP_i+FP_i+FN_i}. \end{aligned} \quad (1)$$

III. RESULTS

The results are presented according to the main classification tasks that define the central narrative of

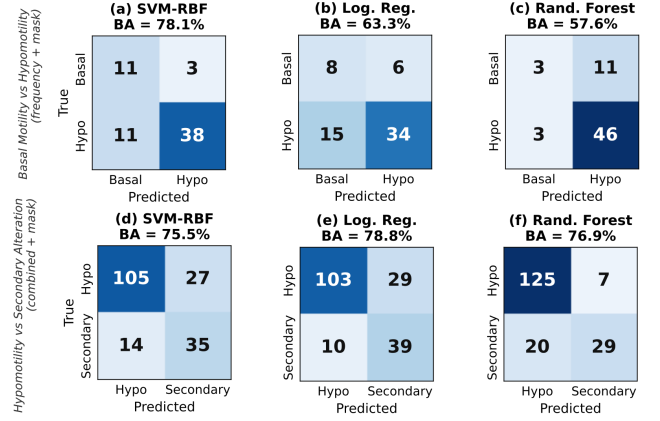


Fig. 3. Confusion matrices for the two core phenotypic comparisons. Panels (a)–(c) show basal motility versus hypomotility using frequency descriptors with spatial masking. Panels (d)–(f) show hypomotility versus secondary ciliary alteration using combined frequency and autoregressive descriptors with spatial masking.

this study: the distinction between basal motility and hypomotility, and the distinction between hypomotility and secondary ciliary alteration. Across both comparisons, predictive performance depended on the interaction among feature representation, masking strategy, and classifier choice. The results are therefore reported by emphasizing the best-performing configurations within each biologically relevant comparison and by summarizing the remaining scenarios separately.

A. Basal motility versus hypomotility

The comparison between basal motility, represented by control samples, and hypomotility, represented by trifluoperazine-treated samples, yielded one of the strongest results in the study. As shown in Table I, the best performance was obtained with SVM-RBF using frequency descriptors with masking of the ciliated region, reaching 78.06% balanced accuracy and 72.78% macro F1-score. Under this configuration, the model produced 3 false positives and 11 false negatives.

From the perspective of feature representation, this comparison clearly favored spectral information. As shown in Table II, under the same classifier and masking condition, the combined representation achieved 58.16% balanced accuracy, whereas AR descriptors achieved 47.45%. This pattern suggests that the distinction between preserved motility and experimentally induced hypomotility is better characterized by dominant spectral cues than by autoregressive temporal descriptors.

Figure 3(a)–(c) shows the confusion matrices for this comparison, illustrating the error distribution among classifiers.

The effect of preprocessing was especially pronounced in this comparison. As shown in Table III, for the same classifier and descriptor type, the use of the ciliary mask increased balanced accuracy from 39.29% without masking to 78.06% with masking. This marked difference indicates that restricting the analysis to the ciliated region was critical for isolating the signal components most relevant to the transition from

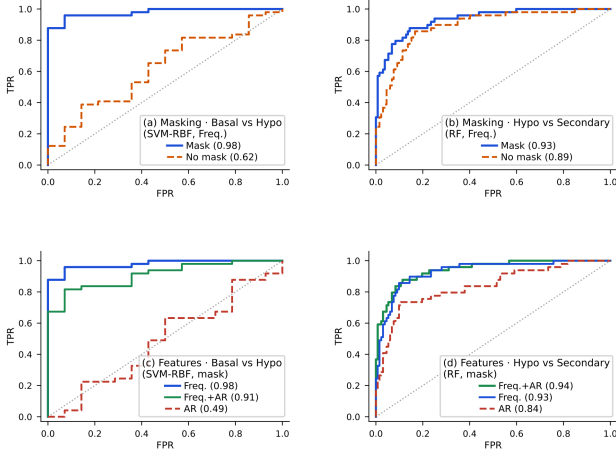


Fig. 4. Receiver operating characteristic (ROC) curves for the two core phenotypic comparisons. Panels (a) and (b) show the effect of spatial masking, while panels (c) and (d) compare feature representations. TPR = true positive rate; FPR = false positive rate.

basal motility to hypomotility. The effect of spatial masking is further illustrated in panels (a) and (b) of Figure 4.

B. Hypomotility versus secondary ciliary alteration

The comparison between hypomotility, represented by trifluoperazine-treated samples, and secondary ciliary alteration, represented by temperature-treated samples, produced the best overall pairwise performance observed in the study, although the optimal metric depended on the classifier considered. As shown in Table I, using combined descriptors with masking, logistic regression achieved the highest balanced accuracy, reaching 78.81%, whereas Random Forest achieved the highest macro F1-score, reaching 79.24%. Under the same representation setting, SVM-RBF yielded slightly lower values.

TABLE I
CLASSIFIER COMPARISON

Model	Bal. Acc. (%)	Macro F1 (%)	FP	FN
Basal motility vs. hypomotility (frequency descriptors with mask)				
SVM (RBF)	78.06	72.78	3	11
Logistic Regression	63.27	59.82	6	15
Random Forest	57.65	58.40	11	3
Hypomotility vs. secondary ciliary alteration (combined descriptors with mask)				
SVM (RBF)	75.49	73.36	27	14
Logistic Regression	78.81	75.37	29	10
Random Forest	76.94	79.24	7	20

The error profiles of the two best-performing models were also distinct. Logistic regression produced fewer false negatives (10) but a larger number of false positives (29), whereas Random Forest showed a more balanced error distribution, with 7 false positives and 20 false negatives. This indicates that logistic regression offered the highest class-balanced separation, while Random Forest delivered the strongest balance between precision and recall.

The corresponding confusion matrices for the hypomotility versus secondary ciliary alteration comparison are shown in Figure 3(d)–(f).

With respect to feature representation, this comparison differed from the basal motility versus hypomotility comparison. As shown in Table II, here, the combined representation outperformed the isolated feature sets. With Random Forest and masking, combined descriptors reached 76.94% balanced accuracy and 79.24% macro F1-score, compared with 75.80% and 77.48% for frequency descriptors, and 70.44% and 72.72% for AR descriptors. These results indicate that distinguishing hypomotility from secondary alteration benefits from integrating complementary information about spectral structure and temporal dynamics.

TABLE II
FEATURE REPRESENTATION COMPARISON

Feature set	Bal. Acc. (%)	Macro F1 (%)	FP	FN
Basal motility vs. hypomotility (SVM-RBF with mask)				
Frequency	78.06	72.78	3	11
Frequency + AR	58.16	56.89	8	13
AR	47.45	46.15	9	20
Hypomotility vs. secondary ciliary alteration (Random Forest with mask)				
Frequency	75.80	77.48	10	20
Frequency + AR	76.94	79.24	7	20
AR	70.44	72.72	8	26

Panels (c) and (d) of Figure 4 compare the ROC curves obtained with different feature representations for both core comparisons.

Masking also improved performance in this comparison, although less dramatically than in the previous comparison. As shown in Table III, for example, with Random Forest and frequency descriptors, balanced accuracy increased from 70.70% without masking to 75.80% with masking. This suggests that the ciliated-region mask improved discrimination, but the comparison remained reasonably solvable even when the full field of view was retained.

TABLE III
EFFECT OF MASKING ON THE MAIN CLASSIFICATION TASKS

Preprocessing	Bal. Acc. (%)	Macro F1 (%)	FP	FN
Basal motility vs. hypomotility (SVM-RBF with frequency descriptors)				
With mask	78.06	72.78	3	11
Without mask	39.29	39.71	11	21
Hypomotility vs. secondary ciliary alteration (Random Forest with frequency descriptors)				
With mask	75.80	77.48	10	20
Without mask	70.70	72.64	10	25

IV. DISCUSSION

The results show that supervised classification of respiratory ciliary motion does not depend solely on the learning algorithm, but on the interaction among the biological problem, data representation, and preprocessing. Rather than revealing a single universally optimal pipeline, the experiments demonstrate that the most informative computational strategy must be tailored to the specific phenotypic comparison being modeled.

In the distinction between basal motility and hypomotility, the best performance was achieved by SVM-RBF using masked frequency descriptors, reaching 78.06% balanced accuracy. This result is biologically consistent because

hypomotility is characterized primarily by a direct reduction in movement, making dominant spectral information highly informative. In this setting, masking acted as an important noise-suppression mechanism by restricting the analysis to the ciliated region and reducing interference from background variation.

Conversely, the comparison between hypomotility and secondary ciliary alteration favored the integration of frequency information and autoregressive modeling. Logistic regression reached 78.81% balanced accuracy using masked combined descriptors, indicating that distinguishing between two abnormal phenotypes requires more than oscillatory speed alone. Secondary alterations may affect the broader temporal organization of the motion, making descriptors of dynamic structure useful for classification. The confusion matrices are shown in Figure 3, with panel (a) for the SVM-RBF result in the first comparison and panels (e) and (f) for the logistic-regression and Random-Forest results in the second comparison.

From a practical and clinical perspective, these findings support the development of automated tools that reduce the subjectivity and operational burden of expert visual inspection. Spatial masking improved performance in the main tasks, especially for basal motility versus hypomotility, but the unmasked condition still retained informative signal in the hypomotility versus secondary alteration comparison. This suggests that future diagnostic workflows may benefit from flexible pipelines: mask-based analysis when maximum discrimination is required, and full-field analysis when manual or semi-automatic region annotation is unavailable or impractical. Such phenotype-aware pipelines could support faster, more reproducible, and less annotation-dependent assessment of respiratory ciliary motion.

The main contribution of this study is the systematic evaluation of spatial preprocessing, frequency and autoregressive descriptors, and supervised classifiers within a common phenotype-driven framework, combined with a physiological interpretation of the resulting classification patterns. Previous studies have investigated ciliary-motion classification using conventional descriptors, 3D convolutional neural networks, and unsupervised spatiotemporal representations [2], [6], [7]. Although direct performance comparisons are limited by differences in datasets, experimental conditions, and classification tasks, these studies reinforce the relevance of computational approaches for ciliary-motion analysis. In this context, the proposed classical pipeline offers explicit and interpretable motion descriptors while requiring substantially less model complexity than end-to-end deep-learning approaches. The different false-positive and false-negative profiles observed among classifiers further indicate that model selection may be adapted to the specific phenotype being investigated. As the present evaluation was performed on a single open dataset using stratified cross-validation, future validation on independent datasets will be important to assess the generalizability of these findings.

V. CONCLUSIONS

This study demonstrated that automated classification of respiratory ciliary motion is fundamentally a phenotype-driven task. The results showed that distinct biological comparisons require tailored combinations of spatial preprocessing, feature representation, and learning algorithms. Specifically, distinguishing basal motility from hypomotility was best achieved using masked frequency descriptors with SVM-RBF, whereas identifying secondary ciliary alterations required combining spectral and temporal-dynamic representations, best modeled by logistic regression and Random Forest.

These findings indicate that dominant spectral information is particularly effective for identifying direct reductions in ciliary motility, whereas more complex abnormal motion states require descriptors representing both oscillatory behavior and temporal-dynamic organization. Ciliary-region masking was especially important for improving discrimination between basal motility and hypomotility, reinforcing spatially targeted analysis when reliable region-of-interest annotations are available. At the same time, unmasked results for hypomotility versus secondary alteration suggest that full-field representations may still be useful when manual or semi-automatic annotation is impractical.

Finally, these results confirm the viability of conventional supervised machine learning for high-speed video microscopy, while also showing that further optimization remains possible. Future work should explore advanced feature fusion, deep-learning models, formal statistical comparisons, and validation using independent datasets from multiple centers to assess the generalizability of the proposed framework.

REFERENCES

- [1] A. Shoemark *et al.*, "European Respiratory Society and American Thoracic Society guidelines for the diagnosis of Primary Ciliary Dyskinesia," *Eur. Respir. J.*, 2025.
- [2] S. P. Quinn *et al.*, "Automated identification of abnormal respiratory ciliary motion in nasal biopsies," *Sci. Transl. Med.*, vol. 7, no. 299, p. 299ra124, 2015.
- [3] S. Blanchon *et al.*, "Deep phenotyping, including quantitative ciliary beating parameters, and extensive genotyping in primary ciliary dyskinesia," *J. Med. Genet.*, vol. 57, no. 4, pp. 237–244, 2020.
- [4] Blausen.com staff, "Medical gallery of Blausen Medical 2014," *WikiJ. Med.*, vol. 1, no. 2, 2014, doi: 10.15347/wjm/2014.010.
- [5] C. Li and S. Quinn, "Video Dataset of Respiratory Cilia Motion Phenotypes," *Zenodo*, 2023. DOI: 10.5281/zenodo.8135892.
- [6] C. Lu and S. Quinn, "Classification of ciliary motion with 3D convolutional neural networks," in *Proc. ACM Southeast Regional Conf.*, pp. 235–238, 2017, doi: 10.1145/3077286.3077303.
- [7] M. Zain *et al.*, "Towards an unsupervised spatiotemporal representation of cilia video using a modular generative pipeline," in *Proc. 19th Python Sci. Conf.*, pp. 166–175, 2020, doi: 10.25080/Majora-342d178e-017.
- [8] C. Li, "Investigating ACE2 localization in respiratory cilia and characterizing cilia motion phenotypes," Ph.D. dissertation, Univ. Georgia, 2023.
- [9] G. Di Benedetto, C. J. Magnus, P. T. Gray, and A. Mehta, "Calcium regulation of ciliary beat frequency in human respiratory epithelium in vitro," *J. Physiol.*, vol. 439, pp. 103–113, 1991, doi: 10.1113/jphysiol.1991.sp018659.
- [10] G. Liu, M. Li, Z. Mao, and Q. Yang, "Structural motion estimation via hilbert transform enhanced phase-based video processing," *Mech. Syst. Signal Process.*, vol. 166, p. 108418, 2022.
- [11] M. Silva, A. Green, and D. Mascareñas, "A generalized technique for full-field blind identification of travelling waves and complex modes from video measurements with hilbert transform," in *Proc. 39th IMAC, Data Science in Engineering*, vol. 9, pp. 233–236, 2022.
- [12] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. A*, vol. 374, no. 2065, p. 20150202, 2016.
- [13] A. Hyvärinen, "Complexity pursuit: separating interesting components from time series," *Neural Comput.*, vol. 13, no. 4, pp. 883–898, 2001.