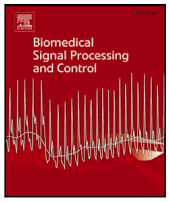
Biomedical Signal Processing and Control 90 (2024) 105808

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



DeepCilia: Automated, deep-learning based engine for precise ciliary beat frequency estimation

Giovanni Dimauro [a,1,*], Nicola Barbaro [a,1], Mauro Giuseppe Camporeale [a,1], Valeria Fiore [b,1], Matteo Gelardi [b,1], Michele Scalera [a,1]

[a] Department Of Computer Science, University of Bari "Aldo Moro", Bari, Puglia, Italy
[b] Unit of Otolaryngology, Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Puglia, Italy

ARTICLE INFO

Keywords:
Signal processing
Nasal cytology
CBF estimation
Object detection

ABSTRACT

Technology advancements over recent decades have greatly impacted medical treatments and diagnostic methods. These advancements could be useful in nasal cytology, which is becoming increasingly critical in diagnosing nasal conditions. In this study we propose DeepCilia, a Ciliary beating frequency (CBF) decay estimation engine that autonomously detects ciliated strias using YOLOv8 to perform the detection task and is able to produce accurate estimations even for low FPS videos, by using the Short-Time Fourier Transform (STFT). DeepCilia achieves very good results with an average computation time below 2 s for each video in the dataset, and an RMSE of less than 1.5 Hz. These achievements make DeepCilia an ideal tool in the field of ciliary motility analysis to estimate the time of death in the context of forensics medicine.

1. Introduction

Technology advancements over recent decades have greatly impacted medical treatments and diagnostic methods. The integration of computer vision in the medical and biomedical fields has led to the availability of additional tools for specialists [1–5]. These tools have enhanced their ability of specialists to acquire, transmit, and analyze digital images and signals, thereby improving their overall performance [6,7]. Much research has focused on the segmentation and classification of cells from digital images, as in the case of nasal cytology, which is becoming increasingly critical in diagnosing nasal conditions [8,9]. This field focuses on the examination of nasal mucosa cells with the objective of recognizing changes in the epithelium, which is frequently subjected to acute or chronic irritation and inflammation caused by viruses, bacteria, or fungi.

Among the cell population of the nasal mucosa, ciliated cells make up 80% of the epithelium in the upper airways [10]. These cells are characterized by the presence of cilia that normally beat in a consistent, fluid pattern, easing effective mucociliary clearance. Ciliated cells in humans are usually 15–20 nm in length and beat at a rate of 16 Hz, but this can vary depending on factors such as infections, temperature, age, or inflammation in the upper airway [11]. This cytotype results quite important since its study allows for the diagnosis of primary ciliary dyskinesia (PCD), a rare pathology that refers to the irregular and disorganized movement of cilia and is usually associated with other severe pathologies such as situs inversus, cardiac diseases and male infertility. The diagnosis is typically made through a biopsy or brushing of the ciliary mucosa and the assessment of ciliary function through the measurement of the ciliary beating frequency (CBF).

In addition, [9] demonstrated that the CBF decay of the ciliated cells in a dead body may help to trace the time of death, since a statistically significant relationship between decreasing ciliary movements and increasing postmortem interval was detected even in the presence of putrefactive changes in nasal ultrastructure integrity.

These motivations have fueled the research seeking an architecture that supports such long, complex estimations, and feels necessary in medical and forensics contexts.

As per the available literature, the tools used to measure CBF exhibit practical limitations and often require expensive technology or manual, constant detection of ciliated cells by physicians to obtain reliable results. Consequently, domain experts without specialized laboratories must rely on qualitative evaluation of ciliated epithelium. Thus, is important to invest in research for the development of an efficient and cost-effective CBF measurement system.

Assessing CBF is a challenging task because high oscillatory motion cannot be observed directly under the naked eye. Over the years, various techniques have been developed to tackle this challenge. Initially, methods relying on photoelectric effects [12], photodiodes [13],

* Corresponding author.
E-mail address: giovanni.dimauro@uniba.it (G. Dimauro).
[1] The authors' contribution is shared and equal on all activities.

https://doi.org/10.1016/j.bspc.2023.105808
Received 20 July 2023; Received in revised form 10 October 2023; Accepted 27 November 2023
Available online 7 December 2023

and stroboscopes [14,15] were proposed. Later, high-speed video cameras allowed for more accurate CBF analysis [16,17], but the process remained semiautomated with CBF determined manually. In 2003, Sisson et al. introduced the first automated technique for measuring CBF, named the Sisson-Ammons Video Analysis (SAVA) [18]. This method is based on the comparison of both analog and digital CBF measurements of the same region of interest (RoI) while varying the temperature, yielding fairly equal results. They also effectuated another measurement with an innovative analysis algorithm called whole-field analysis (WFA), where the entire image is analyzed for each frame. Both the methods have pros and cons: while the RoI method does not need a great deal of computational power to achieve its results, it can usually introduce a selection bias because "the human eye is naturally drawn toward faster beating cells". On the other hand, WFA does not require any human interaction, but it is computationally expensive if algorithms for motion detection are not implemented to ease the task of detecting motile cells. The study presented in [19] details a CBF measurement technique using the Fast Fourier Transform (FFT). The authors use an iPhone camera to capture images, which are then analyzed on a desktop system while emphasizing the importance of accurate RoI selection, stating that it is "challenging and crucial". Consequently, the authors suggested the need for improvement in RoI selection strategies.

In earlier studies [20,21], CBF was determined by analyzing the variations in pixel intensity. However, the results thus obtained were limited by the specific experimental conditions for video acquisition. In [22,23], the authors proved the ability to determine the motile areas and CBF using optical flow (OF). OF is described as the representation of the movement of the imaging surface in 2-D images as a result of the actual movement of objects in 3-D space [24]. Nevertheless, it has been emphasized that the accuracy of CBF estimation is highly dependent on the quality of the image and the frequency resolution, which can be affected by factors such as the sampling frequency and the number of frames captured [21,23–25].

In this study, we propose Deep Cilia, a completely autonomous CBF decay estimation system, whose novelty lies in the following:

- a hybrid WFA/RoI analysis of the video which is completely handled by deep learning models (YOLOv8) that are trained to detect the upper part of a ciliated cell called the ciliated stria. The Deep Cilia architecture is a hybrid between the WFA and the RoI analysis techniques since it analyzes the whole video frame by frame but only returns the frame sections where the ciliated stria is present.
- the ability to correctly estimate the CBF even for low FPS videos through the variation of the light intensity in the frequency domain, calculated using the Short-Time Fourier Transform (STFT).

2. Materials and methods

2.1. Dataset

All the experiments performed in this study were carried out on a dataset specifically collected for this study available in the IEEE Dataport repository (Beating Ciliated Cells, DOI 10.21227/qkm1-k152). It consists of 119 videos of ciliated cells acquired by an expert at the Unit of Otolaryngology, Department of Clinical and Experimental Medicine, University of Foggia, Italy. The videos were acquired with the use of a Nikon Eclipse 600 (with 1000x magnification) microscope equipped with a camera model MD6iS (Sony IMX236 Sensor) (see Fig. 1).

Each video exhibits a variable number of ciliated cells varying from 1 to more than 5 (in contexts where many cells are amassed). The average video duration in the dataset is 93 s, with an upper bound of 352 s and the shortest being only 3 s long.

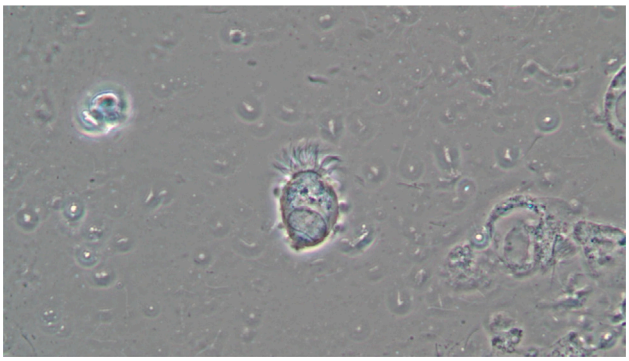


Fig. 1. Example of frame from one of the dataset's videos.

As previously underlined, the cardinality of ciliated cells recorded in each video can vary. From the 119 videos a total of 246 clearly visible strias (Fig. 2) were reported, labeled and manually extracted by the authors on each frame.

In addition to the number of cells and video duration, the videos differ from each other in terms of the following:

- Frames per Second (FPS) [$\mu = 23.63$ FPS, $\delta = \pm 2.54$]
- Steadiness of the camera

Such characteristics can generally influence the results of the experiments: video recordings with a framerate lower than double the cell CBF (folding frequency for that particular cell) can alter the estimation of the signals' frequency [26]. Similarly, a video reporting a few seconds of trembling can irrevocably damage the frequency estimation calculations in that time window (camera shakes will be detected and overlap the metachronal movement of ciliated cells).

To train and evaluate the Ciliated Stria Detection component, all the videos were viewed to allow an expert to manually annotate the ciliated strias of each cell. Then the dataset was partitioned into training, validation and test set containing 89 (185), 18 (36), and 12 (25) videos (cells) respectively.

2.2. Preprocessing and augmentation

Camera lens-equipped microscopes can differ in color calibration, light, resolution, and framerate. For this reason, a standardizing preprocessing phase was defined. Every video in the dataset was first converted to grayscale, and then resized to a resolution of 640 × 640. These operations make the training phase both consistent and RAM wise light in every part of the analysis thanks to the reduction in both image size and number of color channels (from 3 RGB color channels to one grayscale channel).

This same preprocessing will also be applied to every video that will be submitted to DeepCilia for analysis.

In preparation for the CNN training, the number of images was increased through a data augmentation step. In particular, the images underwent minor modifications between shearing (both vertical and horizontal) and random rotations in the range of ±180 degrees. This technique has been proven to be an effective method to also prevent overfitting [27,28], consequently giving the model the chance to better generalize the entities even in different conditions and backgrounds from the ones shown in the training set.

2.3. Stria identification

In the current architectural design, DeepCilia leverages a state-of-the-art CNN such as YOLOv8 [29], which is widely regarded for its good performance in object detection.

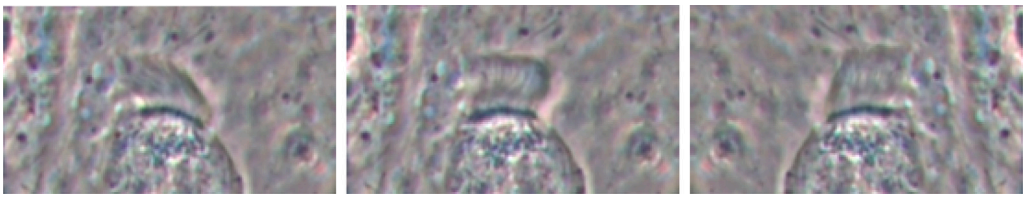


Fig. 2. Manually segmented ciliated strias, used as ground truths when training YOLO.

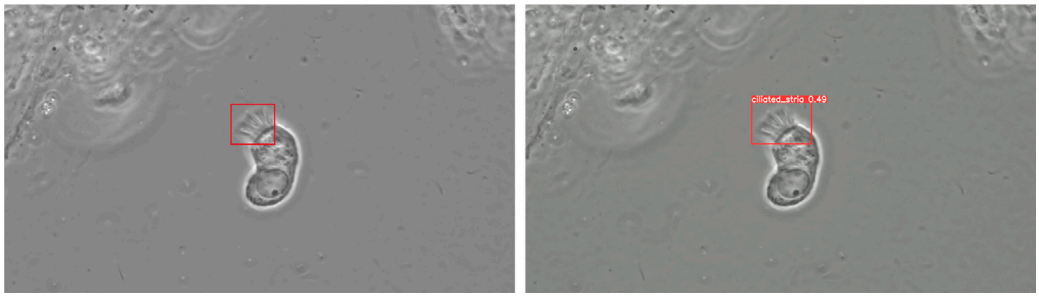


Fig. 3. Comparison between the manually annotated ground truths (a) and the predicted bounding boxes (b).

After training for 200 epochs on the training set, described in Section 2.1, YOLOv8 could reliably detect instances of ciliated strias from each frame and return, for each instance, a bounding box of coordinates to track the ciliated stria (Fig. 3). This tracking allows the engine to not lose track of where the ciliated stria is thus making it robust to small camera movements and translation of the cells in the video.

In order to test YOLOv8 on the task of detection of Ciliated Striae, we submitted to the CNN model a testing set consisting of 5350 images containing 7785 manually-labeled Striae, extracted from 12 videos of the dataset discussed in Section 2.1. For each and every image in the testing set, YOLOv8 yields a bounding box for any plausible Stria detection, and the relative coordinates will be compared to the ground truth in order to obtain the metrics of mAP, Precision and Recall. Different Intersection-Over-Union (IoU) thresholds have been tested, with the best results landing on a precision score of 81% when the IoU is set to 0.5. YOLOv8 also realized a mAP50% of 52% in this testing setting.

Since ciliated cells do not move particularly fast, DeepCilia does not submit to YOLO all the frames of the video but only one per second. This greatly reduces the inference time while still obtaining a precise identification of the ciliated strias.

2.4. CBF estimation

Digital signals (which in this study are frame-by-frame light intensity variations of a video pixel) are characterized by a discrete set of samples that represent the amplitude and time of the signal. These signals may contain noise or unwanted components that need to be attenuated or completely removed in order to extract the desired information. For example, videos recorded with generic cameras can suffer from a variety of issues, such as noise, blur, artifacts, and distortion, which can degrade the analysis. To extract meaningful components of the digital signal, a complex filtering pipeline that elaborates the signal through the selective pass/block of certain band frequencies is needed. Such a pipeline is described in the following from Sections 2.4.1 to 2.4.4.

2.4.1. Pixel intensity through-time filtering

The DeepCilia CBF estimation component is based on the analysis of the light variation of each pixel throughout the video. The bounding boxes returned by the Stria identification component that encompass each detected ciliated stria are used to cut windows from the original video V and from it create many "smaller" videos G_k (one for each detection) that will have as resolution the shape of the bounding box.

For each of these new videos we calculate a function called pixel intensity through-time (PITT) which is defined as a function of a single pixel $p(x, y)$ over the whole video such that $PITT(x, y)$ returns a 1-D array of length N where N is the cardinality of the frames of the videos. Each entry $PITT_{i,x,y}$ of the 1-D array is a value in the $[0-255]$ interval (videos are in grayscale, see 2.2) representing the light intensity of pixel $p(x, y)$ at the ith moment (Fig. 4).

2.4.2. Pixel wise filtering

Each of the k grayscale subvideos G_k (and consequently every pixel $G_{k(x,y)}$ composing them) will be processed using a sequence of filters, namely:

- *Frames resizing through interpolation*: Every frame $G_{k(i)}$ of video G_k is resized to become a square with side length equal to min(100, smaller side of G_k), thus returning a new square video called SG_k. This has been proven effective for faster computation and allows videos of the same size to be stacked to be further processed in parallel.
- *Removal of the DC-Offset*: In the context of camera recordings, DC-Offset can be referred to as a baseline signal that is present in the image sensor's output caused by imperfections in the sensor itself or readout circuitry. DC-Offset can be found as a uniform gray or black level in the captured image which reduces the available dynamic range of the sensor and leads to artifacts. A simple filter is applied to mitigate the effects of DC-Offset. This filter, given an SG_k, calculates the average value of each pixel of the video $P(SG_{k,x,y})$ (1) and then subtracts it from the values of the pixel itself in every frame (2):

$$A_{(SG_{k,x,y})} = \frac{1}{N}\sum_{i=1}^{N} P_i(SG_{k,x,y}) \tag{1}$$

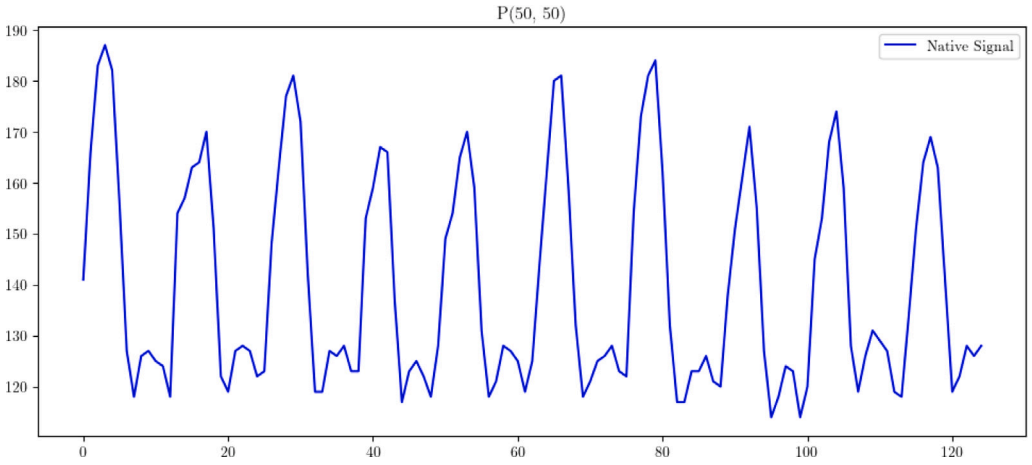


Fig. 4. 5 seconds long PITT array of one video pixel.

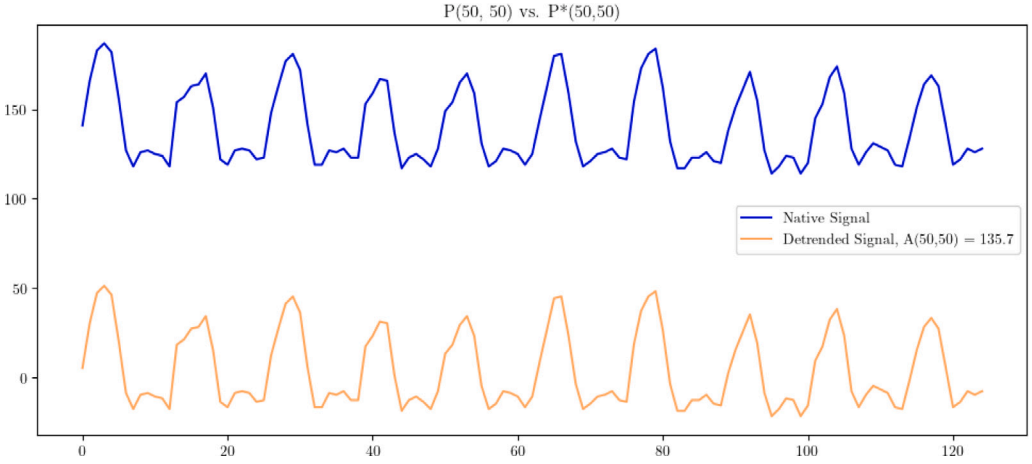


Fig. 5. Signal from Fig. 4 compared to the detrended signal. The plot for P*(50,50) is lower because the average value of the pixel throughout the video has been subtracted.

$$P_i^*(SG_{k,x,y}) = P_i(SG_{k,x,y}) - A_{(SG_{k,x,y})} \qquad 1 \le i \le N \tag{2}$$

The above high-pass filter effectively removes the 0 Hz frequency from the frequency spectrum of a signal (Fig. 5). Fig. 7(b) shows that applying this filter returns an image with less contour definition (e.g. the black stria clearly visible in the lower part of Fig. 7(a) is no longer present). The intuition behind this filtering technique is that every portion of the image that is constant in position and luminosity shall be removed.

- *First-Order Difference Filter*: a type of digital filter used to compute adjacent samples of a signal. It is a high-pass filter, and is used to detect edges or peaks, to analyze transient behaviors such as sudden changes or perturbations. Consecutive positive (resp. negative) elicitations in the digital signal will be amplified. These results are useful to detect minor camera shakes or image noise, with the downside of injecting a small quantity of noise in the signal. This filter also operates pixelwise, and it is implemented as a finite impulse response (FIR) digital filter:

$$EQ_i(SG_{k,x,y}) = P_i^*(SG_{k,x,y}) - P_{i-1}^*(SG_{k,x,y}) \qquad 2 \le i \le N \tag{3}$$

It can be shown that the variance of the noise for each sample obtained by this process is equal to $2\sigma^2$. Suppose that for the ith frame:

$$P_i^*(SG_{k,x,y}) = C_i^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y}) \tag{4}$$

where $C_i^*(SG_{k,x,y})$ is the light intensity value given by the beating ciliated cell in the ith frame for pixel $p(x, y)$ and $N_i^*(SG_{k,x,y})$ is an additive white Gaussian noise of variance σ^2 in the ith frame for $p(x, y)$, in the bandwidth which we are currently focusing. Performing a first-order difference filter as in (3) will give:

$$\begin{aligned} &P_i^*(SG_{k,x,y}) - P_{i-1}^*(SG_{k,x,y}) = \\ &(C_i^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y})) - (C_{i-1}^*(SG_{k,x,y}) + N_{i-1}^*(SG_{k,x,y})) = \\ &C_i^*(SG_{k,x,y}) - C_{i-1}^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y}) - N_{i-1}^*(SG_{k,x,y}) \end{aligned} \tag{5}$$

Focusing on $N_i^*(SG_{k,x,y}) - N_{i-1}^*(SG_{k,x,y})$, one can note that after the filtering phase the variance has doubled, and now has a value of $2\sigma^2$.

Fig. 6 shows that higher peaks in the detrended signal (in orange) exhibit a smaller amplitude in the first-difference filtered signal (light blue), and conversely, small peaks are now much higher. This process, at the cost of the aforementioned noise injection, allows peaks with different amplitudes and the same bandwidth to reduce the amplitude gap itself.

Fig. 7 shows the effect of this filtering technique on a detrended image. In this case, the most agitated parts of the frame (the cilia) remain visible while static parts are canceled thus making the process of frequency estimation more robust and focused only on the pixels that truly count.

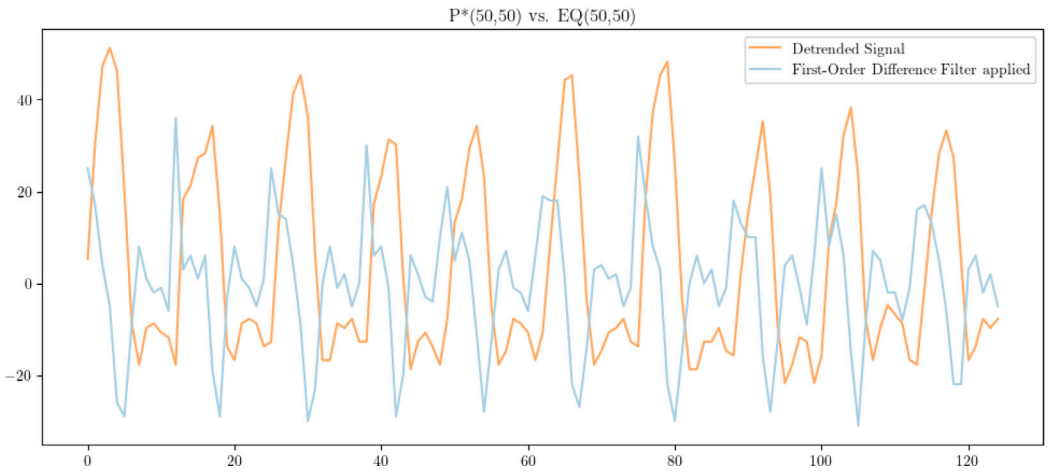


Fig. 6. First-order difference filter (in light blue) shows smaller peaks than the Detrended signal (in orange), bringing the whole light wave closer to the smaller peaks always preceding the higher ones.

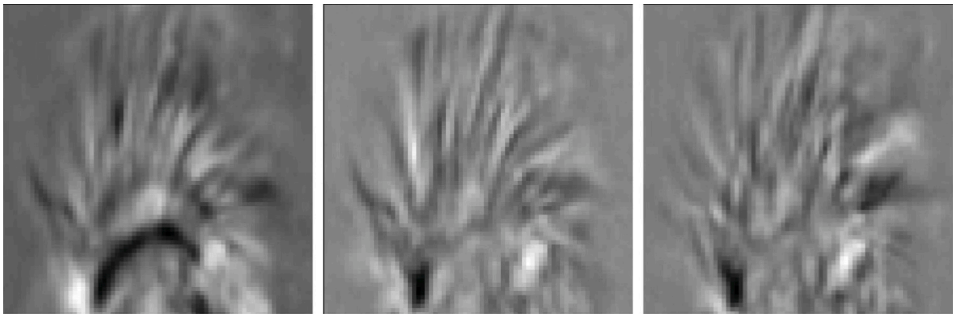


Fig. 7. Detection of a ciliated stria made by YOLOv8 (a); same frame but without the camera DC-offset (b); same frame but with first-difference filtering applied: static parts are darker, thus making the process of CBF estimation simpler and more focused on pixels drawing a ciliated stria (c).

2.4.3. Simple exponential smoothing

Once every $EQ(SG_{k,x,y})$ PITT for each of the k ciliated cells detected by YOLO has been processed, the next step is to generate two distinct digital signal sequences for each PITT.

Given the values $0 < \alpha < 1$ and $\beta = 1 - \alpha$ (with $\alpha = 0.3$), we define:

- FORWARD PASS:

$$F_i(SG_{k,x,y}) = \begin{cases} EQ_1(SG_{k,x,y}), & \text{if } i = 1 \\ \beta * EQ_i(SG_{k,x,y}) + \alpha * F_{i-1}(SG_{k,x,y}), & \text{if } 1 < i \leq N \end{cases} \quad (6)$$

- BACKWARD PASS:

$$B_i(SG_{k,x,y}) = \begin{cases} EQ_N(SG_{k,x,y}), & \text{if } i = N \\ \beta * EQ_i(SG_{k,x,y}) + \alpha * B_{i+1}(SG_{k,x,y}), & \text{if } 1 \leq i < N \end{cases} \quad (7)$$

The F (respectively B) sequence is called *forward* (*backward*) *exponential smoothing* (FES/BES). The properties of FES can be summarized in its ability to stand for values being influenced by what has happened in the immediate past, thus giving the impression that a lag within the signal has been injected. Conversely, BES results in a sequence of values that foresees the real trend of the signal (Fig. 8).

Next, the arithmetical average between the FES and BES is calculated:

$$R_i(SG_{k,x,y}) = \frac{F_i(SG_{k,x,y}) + B_i(SG_{k,x,y})}{2}, \text{ for } 1 \leq i \leq N \quad (8)$$

Over $R(SG_{k,x,y})$, the two injected lags from F and B cancel themselves out when the two overlapped signal cross their paths over similar features of the original signals (Fig. 9).

This exponential averaging function acts as a smoothing factor of small disturbances in the signal. It serves as a low-pass filter where the number of high frequencies flattened is proportional to the value of α. The significant, yet precise dampening factor of the *average exponential smoothing* counterbalances the noise injected by the *first-order difference*. This makes the signal less prone to disturbances while still replicating the sinusoidal behavior of the original chirped signal. This in turn, allows for a more robust signal analysis without altering the phase (Fig. 9).

2.4.4. Short time fourier transform

The DeepCilia pipeline ends with the application of the *Short-Time Fourier Transform* (STFT) on every signal $R(SG_{k,x,y})$. This generates an array of results with cardinality $m = \frac{L - windowsize}{hopsize} + 1$ for each of the $R(SG_{k,x,y})$ where L is the length of the signal (windowsize = 128, hopsize = 0).

From each of the m windows, the principal frequency is extracted by calculating the mode between the frequency values of all pixels. This allows DeepCilia to plot the CBF decay through the entire length of the video of the ciliated cell.

3. Results

The pipeline proposed and explained in Section 2.4 was validated by submitting 20 videos of ciliated strias manually cut-out from 20 different videos of the dataset.

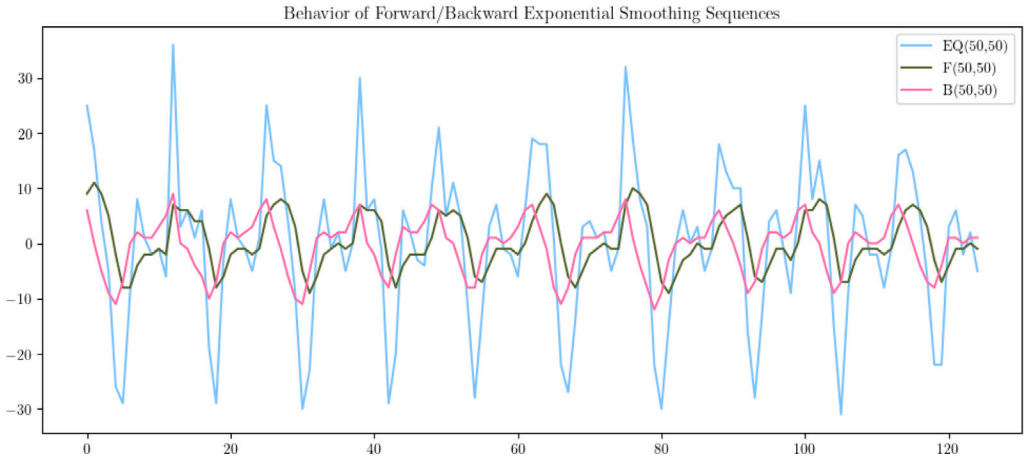


Fig. 8. FES and BES w.r.t. the detrended signal.

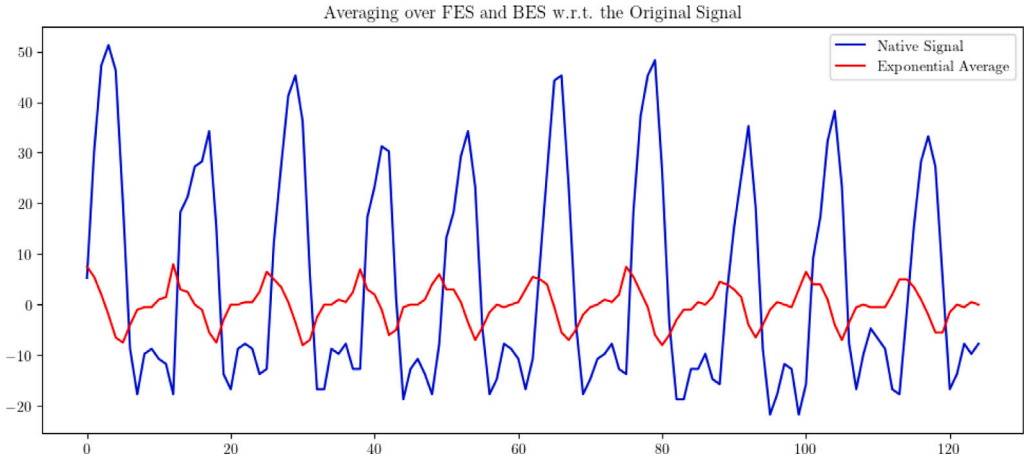


Fig. 9. Averaged exponential smoothed signal (red) and the detrended signal (blue).

To evaluate the DeepCilia pipeline in different working contexts, between these 20 cutout videos, some were purposely extracted from video with very different lengths (15 s to 5 min) and different quality (cells very close to each other, minor camera shakes, slight changes of focus).

The prediction of the CBF decay obtained by the pipeline was compared with a ground truth CBF that was estimated by the authors by playing in slow motion the 20 videos, keeping track of the full ciliary motion to count a single beat. The final CBF ground truth estimate of each part was selected by majority vote between the manual counts made by the authors.

The results were evaluated and compared in terms of root mean square error (RMSE) and of mean bias error (MBE). This is because the former is useful to detect where and when the algorithm's estimated CBFs are showing large gaps from the ground truth, while the latter is useful to understand if the pipeline tends to over/underestimate the true ciliary beat.

Table 1 reports the results achieved, while Fig. 10 reports two example plots of ground truth CBF (blue) against the CBF estimated by DeepCilia (red) of two different ciliated cells in the two longest videos of the dataset.

After a thorough analysis of these 20 videos in the test set, the videos on which DeepCilia reports a brief gap ($RMSE > \tilde{0}.15$) with respect to the ground truth are only the ones where huge camera jolts are recorded either repeatedly or for an extended period of time (more than 1 s in any video window). This shows the overall resilience of the system to small camera shakes and video impurities.

Table 1
CBF estimation algorithm performance scores.

Video Id	Length (s)	Fps	RMSE (Hz)	MBE (Hz)
1	21.48	25	0.099	0.060
2	38.60	25	0.174	0.112
3	28.80	25	0.008	0.008
4	61.64	31	0.109	0.052
5	69.78	31	0.154	0.101
6	29.76	25	0.147	0.118
7	60.76	24	0.259	0.225
8	121.38	24	0.172	0.113
9	71.78	24	0.111	0.067
10	119.53	24	0.079	0.035
11	82.74	24	0.077	0.033
12	328.43	19	0.252	0.134
13	67.92	24	0.175	0.113
14	38.44	25	0.127	0.088
15	28.99	24	0.127	0.208
16	17.32	25	0.075	0.038
17	24.36	25	0.089	0.049
18	31.04	25	0.113	0.065
19	64.38	19	0.238	0.168
20	46.05	19	0.117	0.098
Avg	–	–	0.135	0.094

The focus of this study is also to achieve good performances in terms of time efficiency. Keeping in mind this objective, the architecture of DeepCilia has been designed to be GPU-based to allow parallelization of the different filtering and STFT operations that have to be performed

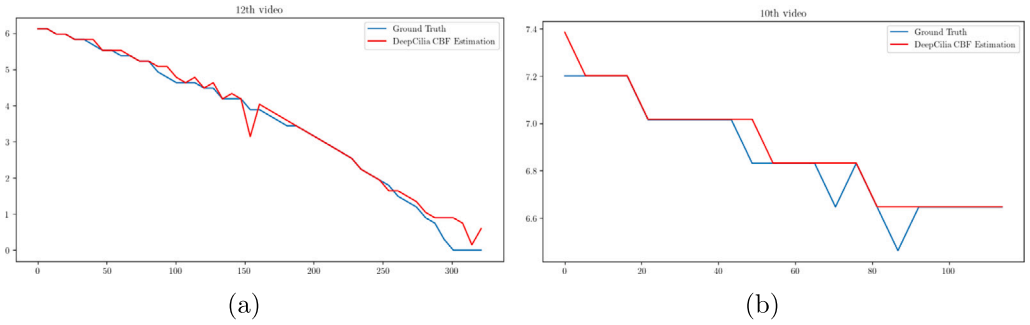


Fig. 10. GT CBF (in blue) against estimated CBF (in red) of video 12 (a) and video 10 (b).

over every single pixel of each video. This optimization guarantees precise results in a matter of seconds. Specifically, the CBF estimation on each of the 20 videos of the test dataset was performed in less than two seconds.

DeepCilia has been implemented with Python (ver. 3.8.12), using the OpenCV library for image manipulations and tensor-computing enabled Pytorch framework for both YOLOv8 deployment, frames parallel preprocessing and STFT computation.

All the experiments were carried out on a machine configured with an Intel-7700HQ CPU, NVIDIA 1050 GPU and 16 GB RAM.

4. Discussion

The evaluation performed on the 20 videos experimental sample revealed its ability to accurately estimate the ciliary beat frequency decay over time, in fact the average RMSE of only 0.135 Hz between the estimation and the ground truth CBF across all 20 videos indicates how the system can precisely track changes in the ciliary beat frequency over time. The heterogeneity of the test dataset and error independence to the video length and Fps, allow to demonstrate the robustness of DeepCilia to real-word variation in video quality; moreover as reported in the previous section, the system resilience to video impurities like small camera shake can be shown by noting that larger error occurred only in videos where huge camera jolts are recorded repeatedly or for an extended period of time. Also, by looking at the Mean Bias Error values, it can be asserted that estimated values are indeed centered around the true value and not systematically over or underestimated.

These promising results can be attributed to the proposed image filtering a STFT pipeline where the combination of filters enhanced relevant signals while suppressing noise and artifacts allowing the STFT to faithfully extract the principal frequency of the ciliary beat even from lower quality videos, thus proving the pipeline to be capable of accurately estimate the CBF. Moreover, the good performances of YOLOv8 in handling the stria detection confirmed the intuition of proposing an hybrid WFA/RoI analysis approach that was able to focus computations only on relevant regions while still analyzing the whole frame, avoiding the selection bias of traditional RoI methods.

5. Conclusions

In this study we propose DeepCilia, a software system that positions itself as an AI-enabled, fully automated end-to-end architecture for fast and reliable CBF estimation of ciliated cells.

DeepCilia can autonomously detect the RoI of the cells in the submitted videos by employing a novel approach to reliably estimate the CBF and its decay through the use of image filtering and STFT analysis. Thus, it can offer a real-time and affordable service without the need for expensive equipment.

The proposed system provides the physician with a tool to automatically estimate the CBF decay of a cell. In addition, thanks to its autonomous stria identification component, it can also relieve the doctor from the tedious task of searching and identifying where the cell lies before starting to count its beats.

The subsequent state-of-the-art computer vision system can handle videos of any length to estimate CBF both quickly, with an average computation time always below 2 s for each video in the dataset, and precisely with an average RMSE of approximately0 0.15 Hz estimated on the 20 ground truths.

These achievements make DeepCilia a step forward in the task of automating the analysis of cell behavior in the context of cytology and make it an ideal tool in the field of ciliary motility analysis to estimate the time of death in the context of forensics medicine.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M. Erten, I. Tuncer, P.D. Barua, K. Yildirim, S. Dogan, T. Tuncer, R.-S. Tan, H. Fujita, U.R. Acharya, Automated urine cell image classification model using chaotic mixer deep feature extraction, J. Digit. Imaging (2023) 1–12.

[2] S.Y. Rubaiat, M.M. Rahman, M.K. Hasan, Important feature selection & accuracy comparisons of different machine learning models for early diabetes detection, in: 2018 International Conference on Innovation in Engineering and Technology (ICIET), IEEE, 2018, pp. 1–6.

[3] G. Casalino, G. Castellano, G. Zaza, Evaluating the robustness of a contact-less mHealth solution for personal and remote monitoring of blood oxygen saturation, J. Ambient Intell. Humaniz. Comput. (2022) 1–10.

[4] G. Dimauro, M.E. Griseta, M.G. Camporeale, F. Clemente, A. Guarini, R. Maglietta, An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset, Artif. Intell. Med. 136 (2023) 102477.

[5] M.K. Hasan, M.H. Aziz, M.I.I. Zarif, M. Hasan, M. Hashem, S. Guha, R. Love, S. Ahamed, HeLP ME: Recommendations for non-invasive hemoglobin level prediction in mobile-phone environment, JMIR Mhealth Uhealth (2019).

[6] G. Dimauro, M.G. Camporeale, A. Dipalma, A. Guarini, R. Maglietta, Anaemia detection based on sclera and blood vessel colour estimation, Biomed. Signal Process. Control 81 (2023) 104489.

[7] M. Valueva, G. Valuev, N. Semyonova, P. Lyakhov, N. Chervyakov, D. Kaplun, D. Bogaevskiy, Construction of residue number system using hardware efficient diagonal function, Electronics 8 (6) (2019) 694.

[8] M. Gelardi, et al., Atlas of Nasal Cytology Stampa for the Differential Diagnosis of Nasal Diseases, second ed., ITA, 2012.

[9] M. Gelardi, L. Iannuzzi, N. Quaranta, M. Landi, G. Passalacqua, NASAL cytology: practical aspects and clinical relevance, Clin. Exp. Allergy 46 (6) (2016) 785–792.

[10] G. Caruso, M. Gelardi, G.C. Passali, M.M. De Santi, Nasal scraping in diagnosing ciliary dyskinesia, Am. J. Rhinol. 21 (6) (2007) 702–705.

[11] B.T. Lemieux, J.J. Chen, J. Jing, Z. Chen, B.J. Wong, Measurement of ciliary beat frequency using Doppler optical coherence tomography, in: International Forum of Allergy & Rhinology, Vol. 5, Wiley Online Library, 2015, pp. 1048–1054.

[12] J. Yager, T.-M. Chen, M.J. Dulfano, Measurement of frequency of ciliary beats of human respiratory epithelium, Chest 73 (5) (1978) 627–633.

[13] H. Teichtahl, P. Wright, R. Kirsner, Measurement of in vitro ciliary beat frequency: a television-video modification of the transmitted light technique, Med. Biol. Eng. Comput. 24 (1986) 193–196.
[14] J. Gray, The mechanism of ciliary movement.—VI. Photographic and stroboscopic analysis of ciliary movement, Proc. R. Soc. B 107 (751) (1930) 313–332.
[15] A. Coste, M.-C. Millepied, C. Chapelin, P. Reinert, F. Poron, M. Boucherat, E. Escudier, Incidence of primary ciliary dyskinesia in children with recurrent respiratory diseases, Ann. Otol. Rhinol. Laryngol. 106 (10) (1997) 854–858.
[16] I. Schipor, J.N. Palmer, A.S. Cohen, N.A. Cohen, Quantification of ciliary beat frequency in sinonasal epithelial cells using differential interference contrast microscopy and high-speed digital video imaging, Am. J. Rhinol. 20 (1) (2006) 124–127.
[17] C. O'Callaghan, K. Sikand, M.A. Chilvers, Analysis of ependymal ciliary beat pattern and beat frequency using high speed imaging: comparison with the photomultiplier and photodiode methods, Cilia 1 (2012) 1–7.
[18] J.H. Sisson, J. Stoner, B. Ammons, T.A. Wyatt, All-digital image capture and whole-field analysis of ciliary beat frequency, J. Microsc. 211 (2) (2003) 103–111.
[19] J.J. Chen, B.T. Lemieux, B.J. Wong, A low-cost method of ciliary beat frequency measurement using iPhone and MATLAB: rabbit study, Otolaryngol. Head Neck Surg. 155 (2) (2016) 252–256.
[20] G. Mantovani, M. Pifferi, G. Vozzi, Automated software for analysis of ciliary beat frequency and metachronal wave orientation in primary ciliary dyskinesia, Eur. Arch. Oto-Rhino-Laryngol. 267 (2010) 897–902.
[21] C.M. Smith, J. Djakow, R.C. Free, P. Djakow, R. Lonnen, G. Williams, P. Pohunek, R.A. Hirst, A.J. Easton, P.W. Andrew, et al., ciliaFA: a research tool for automated, high-throughput measurement of ciliary beat frequency using freely available software, Cilia 1 (1) (2012) 1–7.
[22] W. Kim, T.H. Han, H.J. Kim, M.Y. Park, K.S. Kim, R.W. Park, An automated measurement of ciliary beating frequency using a combined optical flow and peak detection, Healthc. Inform. Res. 17 (2) (2011) 111–119.
[23] E. Puybareau, H. Talbot, G. Pelle, B. Louis, J.-F. Papon, A. Coste, L. Najman, A regionalized automated measurement of ciliary beating frequency, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 528–531.
[24] O. Meste, F. Brau, A. Guyon, Robust estimation of the motile cilia beating frequency, Med. Biol. Eng. Comput. 53 (2015) 1025–1035.
[25] G. Dimauro, A new image quality metric based on human visual system, in: 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS) Proceedings, IEEE, 2012, pp. 69–73.
[26] L. Lévesque, Nyquist sampling theorem: understanding the illusion of a spinning wheel captured with a video camera, Phys. Educ. 49 (6) (2014) 697.
[27] J. Wang, L. Perez, et al., The effectiveness of data augmentation in image classification using deep learning, Convolutional Neural Netw. Vis. Recognit. 11 (2017) (2017) 1–8.
[28] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[29] G. Jocher, A. Chaurasia, J. Qiu, YOLO by ultralytics, 2023, URL https://github.com/ultralytics/ultralytics.