wes-2025-26
Response to Anonymous Referee #1

Thank you for taking the time to review our manuscript. We sincerely appreciate your constructive feedback and careful evaluation. In response, we have revised the manuscript accordingly, with all changes clearly marked in blue for your convenience. Corresponding updates are also indicated in blue font within this response letter. We trust that these revisions address your comments and enhance the clarity and quality of our work, and we hope the updated version meets your expectations.

Key concerns:

1. Material and Damage Representation: The structural damage was introduced in metal rather than in composite materials, which are more representative of real-world turbine blades. Furthermore, a saw cut does not replicate the characteristics of a crack as it would naturally occur.

   Reply: We thank the reviewer for this valuable observation. We fully acknowledge that our experimental setup, specifically, the use of an aluminum cantilever and the introduction of damage via saw cuts, does not replicate the material composition or crack morphology typical of modern composite wind turbine blades.

   Our decision to use a metallic cantilever was driven by the need for a controlled and repeatable environment to establish a proof of concept. The use of a saw cut enables us to systematically vary the damage severity and assess its measurable impact on the aerodynamic pressure field under well-defined and reproducible conditions. While the induced damage may not mimic natural crack propagation mechanisms in composites, it introduces a local stiffness reduction that suffices to validate the hypothesis that structural changes manifest as measurable perturbations in the aerodynamic pressure distribution.

   This study is intended as a first step to demonstrate the feasibility of using aerodynamic pressure measurements for structural condition assessment. Future work will focus on extending the methodology to composite specimens and more realistic damage types, such as delaminations or matrix cracking.

   We have modified the manuscript in several locations and added a dedicated paragraph to the revised manuscript explicitly acknowledging these limitations and clarifying the motivation behind our experimental choices:

   - line 160: We approximate a fixed WTB by mounting an airfoil on an aluminum cantilever beam with a rectangular cross-section and placing it in a wind tunnel test section.
   - lines 173-178: An important limitation of our setup lies in the choice of material and the manner in which damage is introduced. While real-world wind turbine blades are composed of layered composite materials exhibiting complex failure modes such as delamination and fiber breakage, our experiments employ an aluminum cantilever with damage emulated via saw cuts. This simplification allows for controlled, repeatable tests and the ability to systematically vary damage severity. Although the artificial crack does not fully replicate the morphology or fracture mechanics of a naturally occurring defect in composites, it produces a measurable stiffness reduction, which is central to our proof-of-concept study.

- lines 589-592: Additionally, more advanced material models and realistic damage representations will be necessary to accurately account for variations in material properties and structural integrity. Future research efforts will aim to translate the proposed methodology to composite specimens to more closely align with practical applications in wind turbine blade monitoring.

Additionally, we sketch in the outlook of the revised manuscript a strategy, how the proposed measurement and detection concept could be extended towards real-world applicability (see lines 583-607):

Beside these immediate next steps, scaling the damage detection approach proposed in this study to full-scale wind turbines and real-world environmental and operating conditions (EOCs) requires substantial further research and development. We propose the following multi-stage strategy to facilitate this scaling process:

- **Simulation and experimental validation:** Further numerical simulations and experimental validation—such as wind tunnel testing using a miniature wind turbine under varying wind speed conditions and turbulence intensities—are essential. These efforts aim to deepen our understanding of how realistic inflow conditions, rotational aerodynamic effects, and real-world damage scenarios influence the pressure distribution along turbine blades. Additionally, more advanced material models and realistic damage representations will be necessary to accurately account for variations in material properties and structural integrity. Future research efforts will aim to translate the proposed methodology to composite specimens to more closely align with practical applications in wind turbine blade monitoring.
- **Spatial distribution of pressure sensors:** Scaling damage detection to cover the full blade span can be realized by deploying multiple Aerosense sensor nodes along each blade, as illustrated in Figure 1(a) of the manuscript. This setup enables the simultaneous acquisition and processing of aerodynamic pressure data at several chord- and span-wise locations, thereby enhancing spatial resolution and detection capability.
- **Unsupervised or self-supervised damage detection:** In real-world applications, labeled data are typically unavailable. Therefore, an unsupervised or self-supervised approach to anomaly or damage detection is required. In ongoing work, we are developing such a method tailored to the dataset presented in this study, to be reported in a forthcoming publication. To adapt this to operational and environmental variability, we propose leveraging local inflow information estimated via other Aerosense methods (see Section 4 of [2] and Section 4.5 of [8]). Moreover, fusing aerodynamic pressure data with measurements from the 6-DOF inertial measurement unit embedded in each Aerosense node may further enhance robustness and sensitivity.
- **Field deployment and scaling:** The final step involves implementing the proposed sensor layout and detection methods on a small-scale operational wind turbine. Field testing will serve to validate the performance of the unsupervised detection framework. The knowledge and insights gained through this process will inform the subsequent upscaling to full-scale wind turbines, enabling robust aerodynamic pressure-based damage detection under realistic conditions.

According to the third bullet point from above, we add reference [8] to the manuscript.

2. Experimental Conditions: Both the wind excitation and imbalance excitation were kept constant throughout the experiments, a condition not reflective of the variable nature of real-world wind turbine environments.

   Reply: We appreciate the reviewer's thoughtful comment. We fully agree that the experimental conditions employed, namely constant wind inflow and harmonic excitation, do not capture the full complexity of real-world wind turbine operation.

   However, this study was designed as a proof of concept to assess whether structural damage can be reliably detected from aerodynamic pressure measurements in a simplified and well-controlled setting. Given the indirect nature of aerodynamic pressure as a proxy for structural condition, a controlled environment was necessary to observe and isolate the underlying mechanisms governing damage detectability.

   Moreover, we acknowledge that our findings are not directly transferable to full-scale wind turbines. Nonetheless, they lay the groundwork for future research and scaled-up experimental campaigns under more realistic and variable operating conditions. We have revised manuscript at several locations to more clearly articulate this scope:

   - line 6: This proof of concept study is based on a series of wind tunnel experiments on a NACA 633418 airfoil.
   - lines 180-188: Our investigation is conducted in a wind tunnel facility under controlled environmental and operational conditions (EOCs) which do not reflect the complexity of the EOCs real world wind turbines. However, given the indirect nature of aerodynamic pressure as a proxy for structural condition, a controlled environment is necessary to observe and isolate the underlying mechanisms governing damage detectability. Thus, this paper does not aim to answer whether it is possible to detect and rank the severity of structural damage under real operational (rotating wing aerodynamics, pitching, tension stiffening, etc.) and environmental conditions (high turbulence, varying temperature and weather) of a wind turbine and its findings are not directly transferable to full-scale wind turbines. Instead, this paper rather aims to offer a proof of concept as to whether such highly indirect pressure measurements can be conceived for use within an SHM setting and to lay the groundwork for further research and the justification for scaled-up experimental campaigns under more realistic and variable EOCs.
   - lines 586-589: These line point to the first bullet point of our scaling strategy in response to key concern 1. Therefore, please find this above.

3. Evaluation Methodology: A supervised classification method based on CNNs was employed. For real applications, datasets typically do not include labeled damage states, necessitating unsupervised methods that do not rely on such data.

   Reply: Thank you for this insightful observation. We fully agree that in real-world applications, labeled data corresponding to specific damage states are generally not available. As a result, supervised classification methods are not directly applicable, and unsupervised or self-supervised anomaly detection methods must be pursued for practical deployment.

   In this study, we deliberately begin with the simpler, supervised classification setting to assess the fundamental viability of using indirect aerodynamic pressure measurements for structural damage detection. This controlled setup allows us to verify whether the pressure signals—despite the complexity of the underlying aeroelastic dynamics—indeed carry identifiable signatures of damage severity. Demonstrating success in this supervised task thus provides a crucial first step in validating the information content and relevance of the measurement modality itself.

Moreover, the findings from this supervised task offer valuable architectural and signal-processing insights that will inform the development of unsupervised approaches. For instance, the convolutional neural network (CNN) architecture used here may serve as a robust encoder in a future autoencoder-based anomaly detection framework. We are actively working on this transition and will present the results in a forthcoming follow-up publication.

To highlight this rationale and future direction in the manuscript, we have added the following clarifying statements:

- lines 392-395: Although the supervised learning approach employed in this study is not directly applicable to real-world scenarios due to the absence of labeled data, it serves as a first step to assess the viability of indirect aerodynamic pressure measurements for damage detection. The CNN architecture developed here may also provide a suitable encoder for future unsupervised anomaly detection approaches.
- lines 597-603: These lines point to the third point of our scaling strategy in response to key concern 1. As outlined there, future work will focus on the development of unsupervised and self-supervised anomaly detection methods that are suitable for real-world deployment.

4. Given these points, the manuscript's findings have limited transferability to real-world settings. Furthermore, the impact and scope of the work may be misaligned with the target journal's focus.

   Reply: We appreciate your comment and fully acknowledge that our experimental setup does not capture the full complexity of real-world operational conditions. However, we respectfully argue that the value of our study lies in its role as a foundational proof-of-concept. Conducting the investigation under controlled and simplified conditions was an intentional and necessary first step to rigorously evaluate our central hypothesis: that structural damage induces measurable perturbations in the aerodynamic pressure distribution.

   This controlled framework allows us to isolate key mechanisms governing damage detectability, validate the efficacy of our indirect sensing approach, and assess the potential of data-driven detection models before transitioning to more complex and variable real-world environments. Such early-stage, hypothesis-driven studies are critical in establishing scientific feasibility and reducing risks in the subsequent development of scalable monitoring systems.

   In this sense, while the study may be limited in direct transferability, it contributes meaningful insight to the wind energy and structural health monitoring communities by advancing the understanding of aerodynamic sensing as a viable pathway for damage detection. We believe this aligns with the journal's scope, which includes the development of novel methods with potential for future real-world application.

Specific points for improvement include:

1. Line 281 mentions that the parameters of the CNNs are fewer compared to alternative methods. It would be helpful to specify the number of parameters used in the CNNs.

Reply Thank you for this helpful suggestion. We have now included the exact number of trainable parameters used in our CNN model in the revised manuscript (lines 294-296) to provide greater clarity and transparency.

Our decision is motivated by the fact that the number of trainable parameters in the proposed CNN architecture—302,342 parameters—is substantially lower than in the above mentioned alternative architectures, which typically involve significantly more complex networks.

2. Line 284 notes the use of the Adam algorithm. Given that AdamW is now the standard, why was the Adam algorithm chosen?

   Reply: Thank you for this thoughtful comment. We were not aware of the advantages of AdamW over Adam and compared our proposed CNN against a CNN with the same architecture and training routine using AdamW and a small weight decay of 0.0001. However, the overall classification accuracy yielded via the AdamW model was lower compared to the classification accuracy reported in the preprint, as can be seen in Table 1. The table reports slightly lower classification accuracies for the different splits and a lower average classification accuracy for both datasets.

Table 1: Classification results using the proposed CNN architecture and training procedure together with the AdamW optimizer and a weight decay of 0.0001.

AoA	split 1	split 2	split 3	average
0°	80.00%	97.80%	94.66%	90.82%
8°	79.40%	94.43%	87.12%	86.98%

To increase the classification accuracy, we introduced - further to the weight decay implemented by AdamW - additional regularization in the form of standard dropout layers after the activation of every convolutional layer.[1] For a dropout rate of 0.2 and a weight decay of 0.0001, we obtain the accuracies on the different splits shown in Table 2:

Table 2: Classification results using the proposed CNN architecture enhanced by dropout layers (dropout rate of 20%) after every convolutional layer and proposed training procedure together with the AdamW optimizer and a weight decay of 0.0001.

AoA	split 1	split 2	split 3	average
0°	82.77%	98.70%	95.10%	92.19%
8°	82.63%	95.70%	88.20%	88.20%

The accuracy on the different splits shown in Table 2 is similar to the accuracy of the model proposed in the preprint, for each single split and thus also for the average values. Training either the AdamW or the AdamW + dropout model for more epochs, using a different learning rate scheme, for example cosine annealing or exponential decay, or varying the weight decay didn't increase the accuracy of the model.

Furthermore, in response to your remark 5, we compare the loss curves of our proposed model from the preprint with those of the AdamW + dropout variant, as both achieve

[1] This model is subsequently referred to as 'AdamW + dropout' model

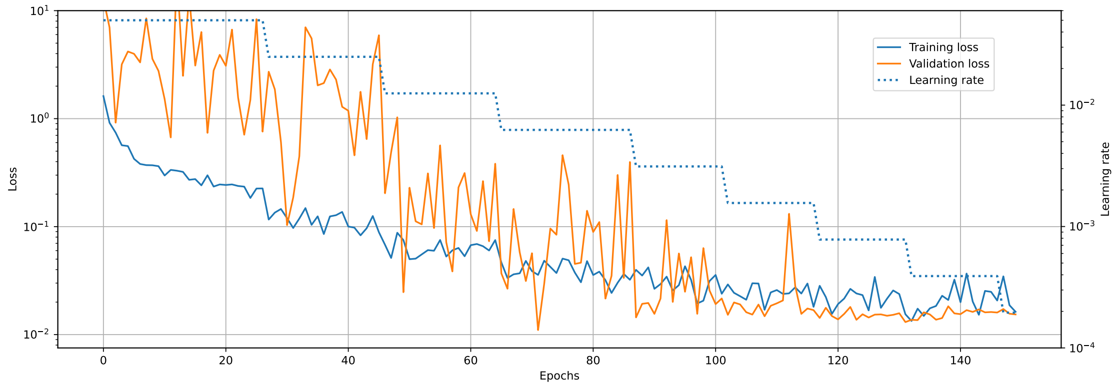

Figure 1: Training loss curve, validation loss curve and the curve of the stepwise decreasing learning rate of the model suggested in the preprint, plotted over the training epochs. These curves correspond to training on split 2 of the 0° AoA dataset.

similar accuracy across all data splits. Figure 1 presents the training loss, validation loss, and learning rate curves for the model proposed in the preprint. While the training loss decreases steadily with only minor fluctuations, the validation loss exhibits pronounced oscillations during the early training phase—particularly when the learning rate is high. As the learning rate progressively decreases, the magnitude of these fluctuations in the validation loss also diminishes. In the final ~ 40 iterations, the validation loss stabilizes and fluctuates only mildly. Overall, the model demonstrates good convergence behavior, despite the initial variability in the validation loss.

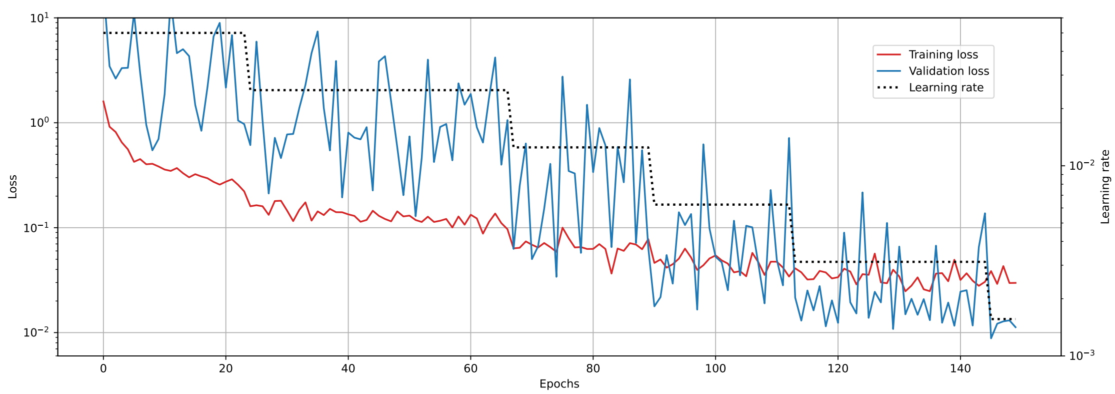

Figure 2: Training loss curve, validation loss curve and the curve of the stepwise decreasing learning rate of the AdamW and dropout model described above, plotted over the training epochs. These curves correspond to training on split 2 of the 0° AoA dataset.

Figure 2 shows the training loss, validation loss, and learning rate curves for the AdamW + dropout model. Similar to the model proposed in the preprint, the training loss decreases consistently over time (see panel a) of Figure 3). However, the validation loss exhibits more pronounced fluctuations during the final $\sim$50 training epochs when compared to the validation loss of the preprint model (see panel b) of Figure 3). Additionally, the learning rate does not decay as substantially as it does in the preprint model (again, see panel a) of Figure 3).

When examining the smoothed validation loss curves in panel b)—of Figure 3 obtained using debiased exponential smoothing with a smoothing factor of 0.7 (as implemented in TensorBoard)—it becomes evident that the preprint model converges toward lower validation loss values. Given this slightly more stable training behavior, the reduced architectural complexity, and the absence of any significant difference in classification accuracy between the two models, we have opted to retain the preprint model in this study.

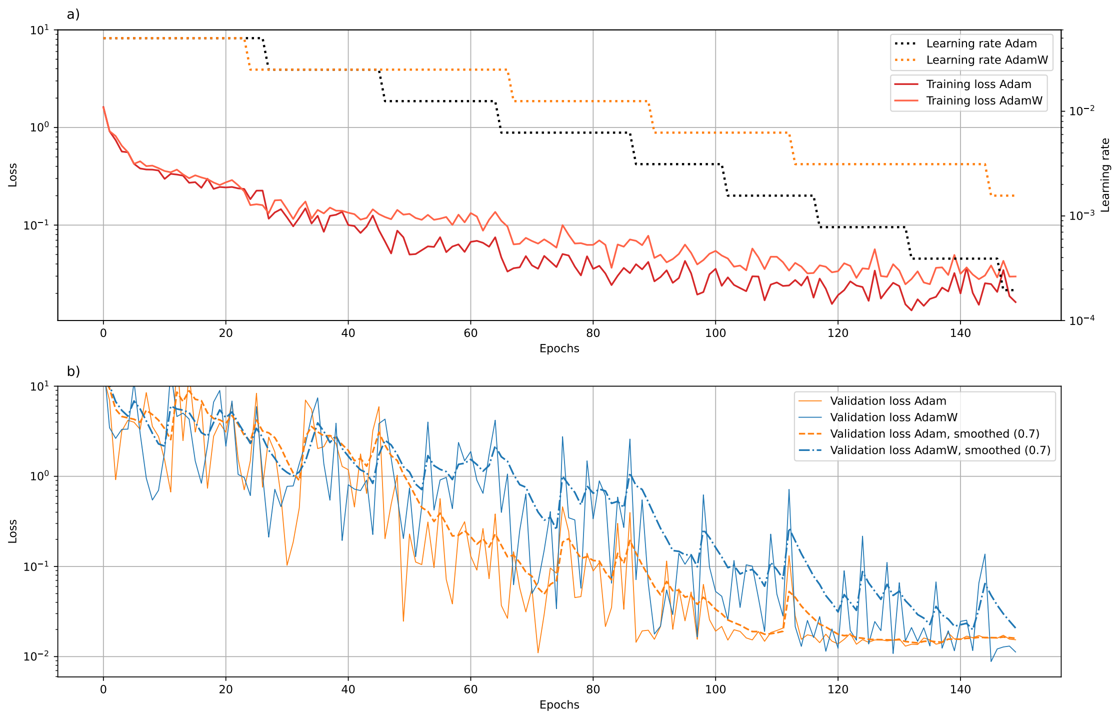

Figure 3: Comparison of loss and learning rate curves for the model proposed in the preprint (referred to by *Learning rate Adam*, *Training Loss Adam* and *Validation Loss Adam* in the legend) and the AdamW + dropout variant. Panel a): Training loss and learning rate over training epochs for data split 2 at an angle of attack (AoA) of 0°. Panel b): Smoothed training and validation loss curves over training epochs for the same split.

3. Figure 8 suggests that labeled data is required. In real-world scenarios, where would labeled data come from? What can be done if no labeled data is available?

   Reply: Thank you for this important and insightful comment. As we also discussed in our response to key concern 3, we fully acknowledge that in real-world scenarios, labeled data indicating specific damage states are typically not available. This makes supervised learning approaches impractical for real deployment and highlights the necessity of developing unsupervised or self-supervised anomaly detection methods.

   In the present study, our primary goal is to demonstrate the feasibility of using aerodynamic pressure measurements for structural damage detection through a simplified and controlled setting. To this end, we rely on labeled experimental data as a first step to validate that damage-related information is indeed encoded in the pressure signals and can be learned by even relatively simple machine learning models.

   We are currently working on a follow-up publication that addresses this exact challenge by developing and evaluating unsupervised approaches tailored to the characteristics of the pressure data. We have also revised the manuscript to clarify this research trajectory and the rationale behind our current focus, as described in the response to key concern 3.

4. Figure 10 is confusing due to the inconsistent decimal places, e.g. 0.99 and 0.011.

   Reply: Thank you for this remark. We adapt this in the revised manuscript (see line 359) as follows:

justification=centering

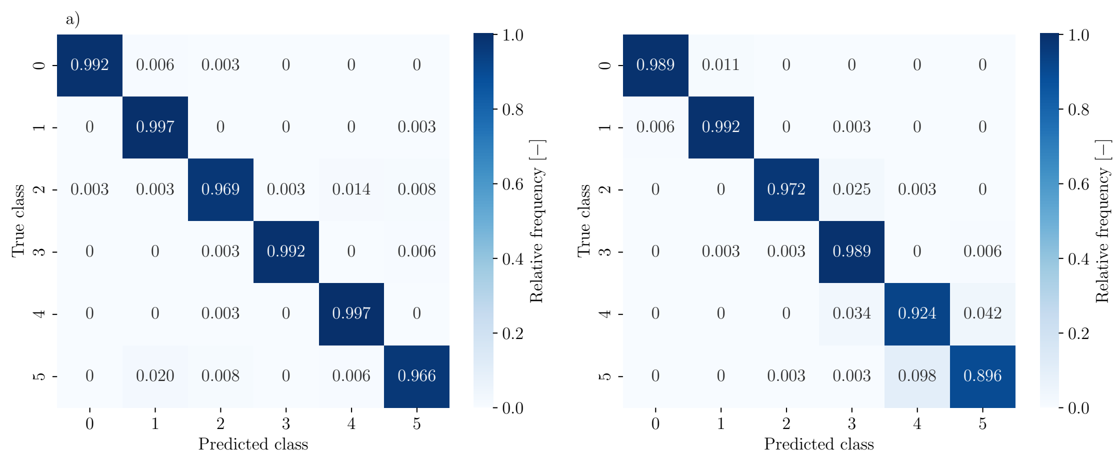

Figure 4: Classification results for split 2 of both datasets depicted by confusion matrices and rounded to three decimal places for non-zero values. The rows correspond to the true class of a sample; the columns to the class predicted by the proposed method. The relative frequencies are rounded to three decimals places. a) shows the results for 0° AoA and b) for 8° AoA.

5. While line 365 mentions fast classification times, how long does CNN training take? Showing a training loss curve would be beneficial.

   Reply: Thank you for this remark. The training and validation loss curve our CNN model is shown above in Figure 1. We include the training time to the revised manuscript as subsequently shown in lines 389-391:

   From a computational perspective our algorithm is efficient: while the model is trained in approximately 445s on a cluster node using an Intel Xeon Gold 6336Y CPU and a NVIDIA HGX A100 80GB GPU, the test set is classified in approximately 1.8s.

6. Section 5 would benefit from examining the model's performance when presented with data not included within the training data (e.g. other windspeed, bigger crack, etc.)

   Reply: Thank you for this valuable suggestion. We agree that assessing the generalization capability of the model to unseen conditions—such as different inflow velocities or damage severities—is critical for real-world applicability. However, at this stage, our purely data-driven models, including the relatively small CNNs used in this study, do not generalize well to conditions outside the training domain. This is a well-known limitation of data-driven approaches, which tend to struggle with extrapolation beyond the distribution of the training data.

   We expect improved generalization by incorporating additional inputs that explicitly describe the operational state—such as local angle of attack and relative inflow velocity—which can be estimated as outlined in Section 4 of[2] and Section 4.5 of[8]. Additionally, the use of more expressive architectures and expanded, more diverse training datasets will be necessary to robustly address this limitation.

As a concrete next step, we plan to explore transfer learning strategies to enhance generalization across varying inflow conditions. For instance, we aim to develop unsupervised anomaly detection models trained on data at 0° angle of attack and evaluate their applicability to data collected at 8°, thereby reducing the dependence on exhaustive retraining.

These considerations have been incorporated into the revised manuscript as part of the third bullet point in the forward-looking strategy outlined in response to key concern 1.

7. Section 6.2 conducted studies under ambient excitation only, which depends on laboratory conditions and may not be comparable to the controlled excitation in wind and imbalance conditions in Section 5. This could also explain why certain frequencies were not identified for some damage states.

   Reply: Thank you for this pertinent observation. We fully acknowledge that the studies presented in this work were conducted on a laboratory-scale mock-up under ambient excitation conditions, and not on a rotating wind turbine in the field. The aim of this setup is to provide a controlled, repeatable environment suitable for proof-of-concept investigations.

   While this setup does not fully replicate the complexities of real operational conditions, including rotation of the blade, environmental variability, and less controlled excitation, it allows us to systematically introduce and study damage scenarios, which would be significantly more challenging to implement and evaluate on an operational turbine.

   We agree that field deployment under realistic operational conditions, in the field, represents a crucial next step and a significantly more complex challenge. We hope to address this in future work as we move from controlled laboratory validation toward real-world implementation.

8. The mode shapes in Figure 14 vary significantly between model orders. How were these complex mode shapes transformed into real space? The first mode shape seems improperly identified — what caused this?

   Reply: Thank you for this insightful observation. In the original analysis, we employed stochastic subspace identification (SSI) to extract modal parameters from acceleration measurements collected under ambient excitation. However, these measurements were limited to a total duration of approximately 60s, of which only 40s were actually used for identification (the first 20s were discarded due to transient effects). We note here that the manuscript mistakenly stated that 60s windows were used; this has now been corrected in the revised version.

   Given the relatively short signal length, SSI—being a time-domain method—struggled to produce stable results, particularly for lower-energy modes. We believe this is the primary reason for the high variability in mode shapes observed in Figure 14 and the incorrect identification of the first mode. To address this issue, we re-analyzed the acceleration data using a frequency-domain approach, specifically the Automated Frequency Domain Decomposition (AFDD) method [6], which is better suited to shorter ambient response signals.

This revised analysis yielded more consistent results. Many of the natural frequencies and several mode shapes previously identified via SSI were confirmed. However, the first mode was indeed found to be incorrectly identified in the original analysis. Additionally, the mode previously labeled as ϕ_6 at approximately 53 Hz appears to be a spurious mode and does not appear in the AFDD results. As also relevant to your remark 10, we can confirm that mode 4 is now detected in the undamaged state, not only from damage class 3 onward as previously reported.

We have revised Subsection 6.2 ("Evolution of the eigenfrequencies with increasing damage") and Subsection 6.3 ("Influence of the crack characteristics") to reflect these updated findings and corrected methodology:

- lines 450-462: We use automated frequency domain decomposition (AFDD), introduced by [6], and its implementation in MATLAB. AFDD is based on frequency domain decomposition that was originally introduced by [4, 3, 5]. The acceleration measurements collected under ambient excitation have an approximate length of 60s. We discard the first 20s of each acceleration measurement and extract signal windows of 10s duration from the remaining signal, using a sliding increment of 0.5s. Subsequently, the signal windows are preprocessed with a low-pass filter with a cut-off frequency of 100s. In an additional step, we employ k-means clustering, introduced by [11], to group the previously identified natural frequencies and mode shapes into distinct clusters that represent different vibration modes of the analyzed structure. By computing the mean value $\mu_{f,i}$ and the standard deviation $\sigma_{f,i}$ of each cluster i of natural frequencies, we obtain the uncertainty associated with the respective natural frequency. We use the same approach to compute the mean values and standard deviations of the relative displacements of each mode shape ϕ_i at the location of the sensors, thereby assessing the uncertainty related to the identified mode shapes. Finally, we refine the determined frequency clusters to identify reliable modes by eliminating natural frequencies that exceed a 5% deviation from the respective cluster mean or are only detected in a small number of signal windows. The associated mode shapes are removed accordingly.
- lines 466-478: The results of $\mu_{f,i}$ and $\sigma_{f,i}$ for the eigenmodes up to 60Hz in all structural states are given in Table A4. The evolution of these natural frequencies f_i for increasing damage is presented graphically in Figure 5. Additionally, the eigenmodes ϕ_i that have been identified for experiment 15 are exemplarily given in Figure 6. While the mode shapes ϕ_1, ϕ_2 and ϕ_5 are consistently detected in the data of experiment 15, the mode shapes related to f_3 and f_4 exhibit higher variance. In Figures 5 a) and b) the eigenfrequencies f_1 and f_2 of the first two vertically oscillating eigenmodes ϕ_1 and ϕ_2 exhibit a an overall decreasing trend with increasing crack length (not considering damage class 1 with the added mass here). Considering the associated standard deviations of f_1 and f_2, this trend is more pronounced for f_1. In contrast, the evolution of the eigenfrequencies f_3, f_4 and f_5 is non-monotonic. Possibly, the eigenmodes corresponding to f_3, f_4 and f_5 are sensitive towards perturbations in the boundary conditions of the system. Moreover, due to their higher variability, f_3 and f_4 might not represent merely vertically oscillating eigenmodes, but coupled ones. To obtain further insight in the these eigenfrequencies and eigenmodes, we offer the results of a simulation in the next section, which aims to replicate the experiment. For the monotonically and approximately monotonically decreasing eigenfrequencies f_1 and f_2 holds that the absolute change between the least and the most damaged state is approximately 0.06Hz and 0.14Hz respectively.

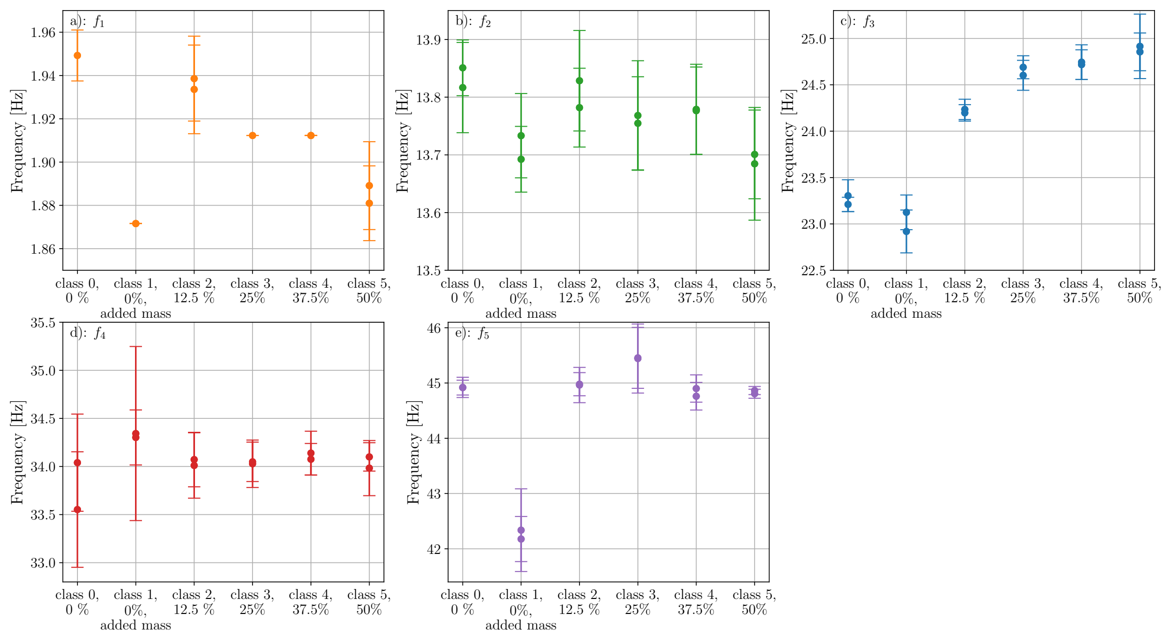

Figure 5: Evolution of the mean values (circular markers) and standard deviations (error bars) of the first five eigenfrequencies f_i, determined with AFDD and clustering, plotted over the damage classes. The exact values of f_i for each damage class are given in Table A4. Eigenmode ϕ_1 could not be detected in experiment 6 and 63 which correspond to damage class 0 and 3. Thus for f_1, there is only one value presented in damage classes 0 and 3.

- lines 510-514: The mode shapes $\phi_{2,m}$ and $\phi_{4,m}$ of the FE-model oscillate only horizontally (see Figure 15). As a consequence, these cannot be consistently identified with the experimental acceleration data and in the OMA, as the accelerometers along the cantilever beam only measure vertical accelerations. Comparing the experimentally determined mode shapes with the vertically oscillating mode shapes of the FE-model, it is noticeable that the eigenvalue analysis of the FE-model does not predict modes the ϕ_3 and ϕ_4 at 23.304Hz and 34.040Hz. Furthermore, looking at the mode shape ϕ_3 and ϕ_4 (see Figure 13, c) and d)), it is observed that the cantilever beam remains approximately straight between 200mm and 800mm.

Additionally, we update the Table A4 in the appendix, which includes all the eigenfrequencies for all detected modes up to 60Hz and all damage states.

Furthermore, we added the references [6, 4, 3, 5, 11] to the revised manuscript.

9. Line 482 suggests comparing identified mode shapes with those from the FE model using the MAC for clarity.

Reply: Thank you for this remark. We added a comparison between the modes shapes from the FE model and the operational modal analysis in terms of the Modal Assurance Criterion to the manuscript in lines 517-527:

The modes shapes ϕ_1, ϕ_2 and ϕ_5 inferred via OMA (see Figure 13 a), b) and f)) seem to be the mode shapes most similar to the FE-estimated mode shapes $\phi_{1,m}$, $\phi_{3,m}$ and $\phi_{5,m}$ (see Figure 15). To evaluate this, we compute the modal assurance criterion (MAC), introduced by [1], between the model-based and experimental (from experiment 15,

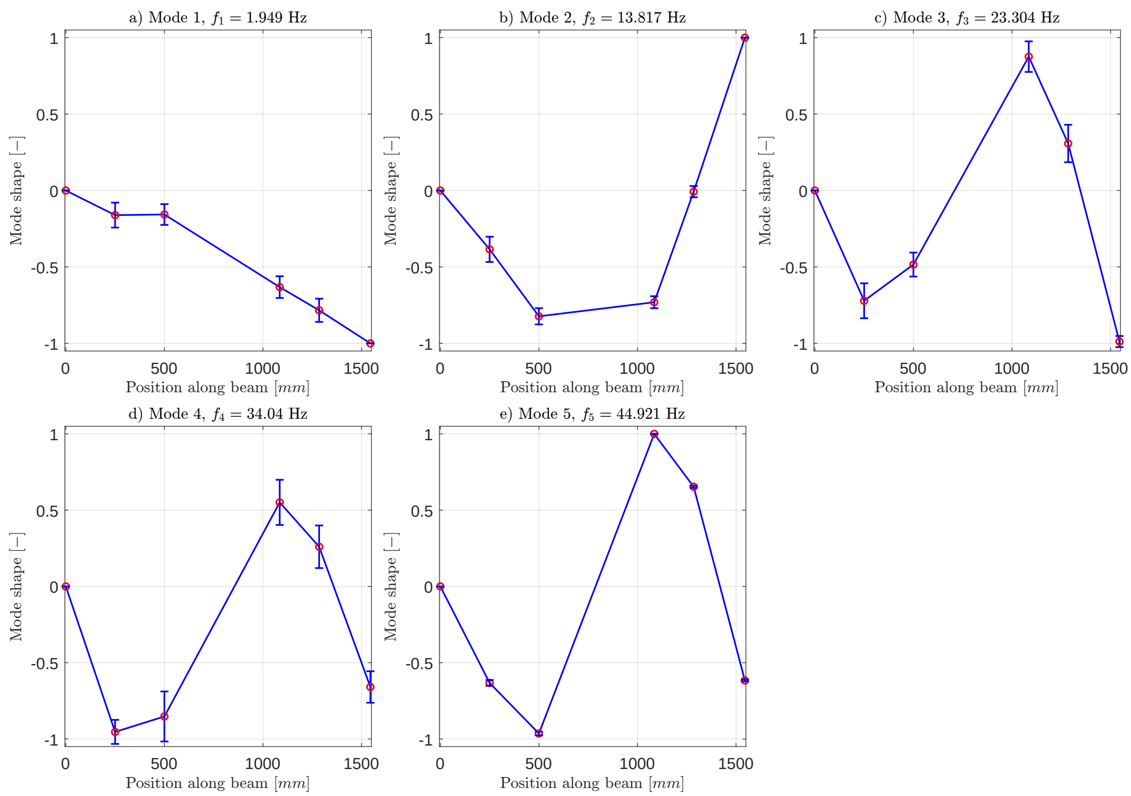

Figure 6: Mode shapes of the first five vertically oscillating eigenmodes ϕ_i of experiment 15, determined with AFDD and k-means clustering. The errorbars indicate the standard deviation associated with the relative vertical displacements of the degrees of freedom of each mode shape. The mode shapes are scaled to comprise a maximum value of 1.0.

Table 3: Values of the modal assurance criterion (MAC) between the modes ϕ_i, determined by OMA, and the modes $\phi_{i,m}$, computed from the FE-model, rounded to two decimal digits. The most similar mode shapes are highlighted by a bold MAC value.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
$\phi_{1,m}$	**0.99**	0.04	0.02	0.02	0.01
$\phi_{3,m}$	0	**0.98**	0.16	0	0.02
$\phi_{5,m}$	0.08	0.02	0.74	0.70	**0.96**

see Figure 13) mode shapes from the undamaged system. For that purpose, we use the data of the five accelerometers and of the FE-model at the same locations. The results shown in Table 3 confirm that the mode shapes ϕ_1 and $\phi_{1,m}$, ϕ_2 and $\phi_{3,m}$, and ϕ_5 and $\phi_{5,m}$ are the most similar mode shapes, as these exhibit the highest MAC, given the five vertical degrees of freedom considered in the computation. Also, the corresponding eigenfrequencies of $f_1 \approx 1.95$Hz and $f_2 \approx 13.88$Hz, and $f_{1,m} = 1.93$Hz and $f_{3,m} = 11.91$Hz are close to each other. However, $f_5 \approx 44.921$Hz and $f_{5,m} = 52.070$Hz, differ more. Nevertheless, due to the good match for the first two natural frequencies and eigenmodes, we conclude that the simulation model offers a sufficient approximation of the experimental setup and may be used for further analysis.

10. Line 490 raises doubts regarding the assumption that Mode 4 results from the cut, as it appears in the stabilization diagram in Figure 12 for the healthy state and could be probably identified at higher model orders. The saw cut does not have the same dynamic characteristics as a real crack, which would only become apparent at higher vibration amplitudes.

    Reply: Thank you for this insightful remark. We conducted another OMA based on frequency domain decomposition and there the mode at approximately 34Hz is identified also in the healthy state. Thus we can confirm your remark and have updated the revised manuscript as shown in the answer to remark 8. Thank you very much for this remark.

Response to Anonymous Referee #2

Thank you for taking the time to review our manuscript. We sincerely appreciate your constructive feedback and careful evaluation. In response, we have revised the manuscript accordingly, with all changes clearly marked in blue for your convenience. Corresponding updates are also indicated in blue font within this response letter. We trust that these revisions address your comments and enhance the clarity and quality of our work, and we hope the updated version meets your expectations.

The paper "On the Potential of Aerodynamic Pressure Measurements for Structural Damage Detection" by Franz et al. explores a proposed novel method for detecting structural damage in wind turbine blades (or similar elastic, aerodynamically loaded structures) using aerodynamic pressure measurements instead of traditional vibration-based sensing. The authors conduct a series of wind tunnel experiments using a heaving airfoil mounted on a cantilever beam, where structural damage is simulated by incrementally sawing the beam near its support. The aerodynamic pressure distribution across the airfoil is captured using a wireless, MEMS-based sensing system. A convolutional neural network (CNN) is then trained to classify damage severity based solely on these pressure time series, achieving high accuracy across various inflow conditions and damage states.

In summary, the proposed method is interesting and could be very useful to researchers and practitioners alike. However, before being reconsidered for full acceptance, the following remarks should all be addressed by the authors

Specific Remarks:

1. Remark 1 & Remark 2: The paper emphasises a controlled wind tunnel proof-of-concept, but it's important to better discuss how the proposed approach could scale to actual, real-life, real-size, field operational conditions. Related to the first remark, further discussion of challenges or plans to generalize from airfoil-level to full blade implementation would add value to the scientific paper.

   Reply: Thank you for these closely related and highly relevant remarks. We fully agree that providing a clear roadmap for scaling the proposed approach to real-world, full-scale applications is essential and adds significant value for the readership of Wind Energy Science. In response, we have included a new paragraph in the outlook section of the revised manuscript (see lines 583-607), which outlines a multi-stage strategy for transitioning from this proof-of-concept study to field deployment under realistic environmental and operational conditions (EOCs). The new section reads:

   Beside these immediate next steps, scaling the damage detection approach proposed in this study to full-scale wind turbines and real-world environmental and operating conditions (EOCs) requires substantial further research and development. We propose the following multi-stage strategy to facilitate this scaling process:

   - **Simulation and experimental validation:** Further numerical simulations and experimental validation—such as wind tunnel testing using a miniature wind turbine under varying wind speed conditions and turbulence intensities—are essential. These efforts aim to deepen our understanding of how realistic inflow conditions, rotational aerodynamic effects, and real-world damage scenarios influence the pressure distribution along turbine blades. Additionally, more advanced material models and realistic damage representations will be necessary to accurately account for variations in material properties and structural integrity. Future research efforts will aim to translate the proposed methodology to composite specimens to more closely align with practical applications in wind turbine blade monitoring.

- **Spatial distribution of pressure sensors:** Scaling damage detection to cover the full blade span can be realized by deploying multiple Aerosense sensor nodes along each blade, as illustrated in Figure 1(a) of the manuscript. This setup enables the simultaneous acquisition and processing of aerodynamic pressure data at several chord- and span-wise locations, thereby enhancing spatial resolution and detection capability.
- **Unsupervised or self-supervised damage detection:** In real-world applications, labeled data are typically unavailable. Therefore, an unsupervised or self-supervised approach to anomaly or damage detection is required. In ongoing work, we are developing such a method tailored to the dataset presented in this study, to be reported in a forthcoming publication. To adapt this to operational and environmental variability, we propose leveraging local inflow information estimated via other Aerosense methods (see Section 4 of [2] and Section 4.5 of [8]). Moreover, fusing aerodynamic pressure data with measurements from the 6-DOF inertial measurement unit embedded in each Aerosense node may further enhance robustness and sensitivity.
- **Field deployment and scaling:** The final step involves implementing the proposed sensor layout and detection methods on a small-scale operational wind turbine. Field testing will serve to validate the performance of the unsupervised detection framework. The knowledge and insights gained through this process will inform the subsequent upscaling to full-scale wind turbines, enabling robust aerodynamic pressure-based damage detection under realistic conditions.

This proposed roadmap emphasizes that while our current work is a controlled proof-of-concept, it provides a critical foundation for the long-term goal of achieving robust structural health monitoring on operational wind turbines.

According to the third bullet point from above, we add reference [8] to the manuscript.

2. The paper acknowledges that sensor placement and distribution on full blades have not been explored; this aspect may be investigated a little more, or discussed in further detail with the current information.

   Reply: Thank you for this important comment. We have conducted a preliminary analysis of sensor placement at the airfoil level, specifically investigating spanwise sensor distribution. Initial results suggest that the relevance of individual sensor positions for damage detection varies depending on the inflow conditions. These findings indicate that sensor importance is not uniform and could be influenced by local aerodynamic behavior. Given the practical implications of this observation, we are currently working on a follow-up publication that will investigate this aspect more systematically.

   Regarding the placement of sensors along the full length of a wind turbine blade, our conceptual approach envisions multiple sensor nodes distributed along the span. However, this aspect cannot be addressed using the currently available experimental dataset.

A thorough investigation of full-blade sensor distribution would require high-fidelity simulation models capable of capturing the coupled interaction between blade vibrations and surrounding aerodynamic flow fields.

While we fully agree that this is a highly relevant and practically significant question, it is also a complex one that falls outside the scope of the present proof-of-concept study. Nevertheless, we believe this constitutes an important direction for future research, and we now briefly mention this in the outlook section of the revised manuscript (see lines 593-596 - these lines point to bullet point two of the scaling strategy described above).

3. The CNN is selected for its small parameter count, but its generalization to unseen data (especially with small training sets) may need stronger justification or comparison with simpler ANN models.

   Reply: Thank you for this thoughtful remark. We acknowledge the importance of justifying the choice of model architecture, especially in light of the limited training data and the need for generalization. While we understand your suggestion to compare our CNN to simpler ANN architectures, our selection of the CNN was based on its favorable trade-off between model complexity, classification performance, and suitability for small datasets.

   Prior to selecting the CNN architecture, we conducted preliminary evaluations using alternative classification approaches, including feature extraction methods such as the `catch22` time-series feature library (introduced by [12]), as well as time-series transformation techniques like ROCKET and MiniROCKET, followed by standard classifiers (e.g., Random Forests, Gradient Boosting, and Multi-Layer Perceptrons). However, all of these combinations yielded lower classification accuracies compared to the proposed CNN model.

   At present, we have no evidence that simpler ANN architectures—or hybrid models combining feature transforms with ANNs—can achieve comparable performance with the available data. That said, we agree that including a broader architectural comparison would strengthen the evaluation and is a valuable direction for future work. Due to space and scope constraints, we have chosen to focus this study on demonstrating the feasibility and performance of our proposed CNN architecture. A more comprehensive investigation, including comparisons with alternative and potentially simpler models, will be addressed in our planned follow-up work on unsupervised damage detection.

4. The consistent confusion between damage classes 4 and 5 at higher AoA suggests model limitations or overlapping feature distributions. This should be discussed more explicitly and in more detail.

   Reply: Thank you for this valuable observation. We agree that the consistent misclassification between damage classes 4 and 5 at an 8° angle of attack (AoA) merits further discussion. As shown in Figure 11 of the manuscript, particularly the heaving amplitudes for experimental series 5 and 6 (corresponding to damage classes 4 and 5) are highly similar at 8° AoA and thus indicate similar pressure distributions. This, in combination with only minor shifts in the structural eigenfrequencies between these two damage states, likely contributes to significant overlap in the learned feature representations.

   Moreover, the increased aerodynamic unsteadiness and turbulence at higher AoA further reduces the signal-to-noise ratio, making it more difficult for the model to distinguish subtle structural differences. Given these factors, and considering the limited capacity of the small CNN architecture used in this study, the observed confusion is not unexpected. While a more in-depth interpretation (e.g., using explainability tools such as SHAP) could provide further insight, it is beyond the current scope. We plan to include such analyses in future work to better understand model decisions and improve class separability under more complex conditions. We now acknowledge and discuss this limitation more explicitly in the revised manuscript (see lines 367-376).

Furthermore, we observe a consistent confusion between damage classes 4 and 5 at 8° AoA in Figure 10. In the set of considered damage states, the variation from class 4 to 5 represents the smallest relative change between two damage states and, following Figure 11, particularly the acceleration amplitudes and thus also the velocity amplitudes for experimental series 5 and 6 (corresponding to damage classes 4 and 5) are highly similar at 8° AoA. Based on the simplified relationship of Equation (7), this indicates similar pressure distributions. This, in combination with only minor shifts in the structural eigenfrequencies between these two damage states, likely contributes to significant overlap in the learned feature representations. Moreover, the increased aerodynamic unsteadiness and turbulence at higher AoA further reduces the signal-to-noise ratio, making it more difficult for the model to distinguish subtle structural differences. Given these factors, and considering the limited capacity of the small CNN architecture used in this study, the observed confusion is not unexpected.

5. The concept of pressure-based damage indicators vs conventional vibration-based ones is very interesting from a conceptual perspective, however, a more quantitative comparison would be useful, even only in a small dedicated subset, to better compare the novel proposed and conventional methodologies. For instance, RMS acceleration (Section 6.1) and eigenfrequency evolution (Section 6.2), which are widely used damage-sensitive features, can be used similarly to train a CNN - or other kinds of ANNs.

   Reply: Thank you for this thoughtful comment. We agree that a quantitative comparison between pressure-based and conventional vibration-based damage indicators would offer valuable insight and context for the reader. However, such a comparison lies beyond the scope of the current study, which is focused on exploring the feasibility of using aerodynamic pressure measurements as a basis for damage detection. We fully acknowledge that vibration-based features—such as RMS acceleration and eigenfrequency shifts—are well-established and often more directly sensitive to structural changes. From a purely diagnostic standpoint, they may indeed offer superior performance in many scenarios.

   That said, one of the core motivations of our study is to investigate alternative sensing modalities that may be more readily available or easier to integrate in certain applications. For example, in the context of wind turbines, aerodynamic pressure measurements are often already acquired for control or performance monitoring purposes, while the installation and maintenance of vibration sensors (especially on rotating components) can be logistically complex and costly. Thus, while not necessarily superior in terms of raw sensitivity, pressure-based indicators can offer a pragmatic and scalable option for structural health monitoring in environments where access to direct structural measurements is limited. We agree that future work should include a dedicated quantitative comparison to assess the trade-offs between these sensing approaches, and we now mention this explicitly in the revised manuscript (see lines 608-612):

   Moreover, another future research direction consists in quantitative comparison between pressure-based and conventional vibration-based damage indicators. Although from a purely diagnostic standpoint, vibration based features may offer superior performance in many scenarios, aerodynamic pressure-based indicators might offer a pragmatic and scalable option for structural health monitoring in environments where access to direct structural measurements is limited. Hence, a quantitative comparison to assess the trade-offs between these sensing approaches, would offer valuable insights.

6. Since the approach is fully data-driven, more insight into the physical interpretability of the CNN outputs and feature relevance (even only in a qualitative, descriptive manner) would improve trust in the method.

   Reply: Thank you for this important remark. We fully agree that interpretability is a critical aspect of data-driven methods, particularly in safety-critical domains such as structural health monitoring, where trust and transparency are essential. In the present study, our focus was on demonstrating the feasibility of using aerodynamic pressure measurements in combination with a CNN to detect structural damage. We acknowledge that the CNN is used here more as a black-box classifier, and we have not yet conducted a detailed post-hoc interpretability analysis.

   Initially, we explored more physically interpretable approaches using feature libraries such as `catch22` ([12]), as well as hand-crafted features based on aerodynamic principles. However, these methods did not yield satisfactory performance on our dataset, motivating the shift to a purely data-driven approach. Although we do not currently provide formal interpretability results, we recognize the need to enhance understanding of what the CNN is learning—particularly which pressure patterns or dynamic signatures contribute most to damage classification. Techniques such as saliency maps, feature attribution, or layer-wise relevance propagation could offer valuable insights and are excellent candidates for future investigation. We have revised the manuscript accordingly and added the following sentence to the outlook to emphasize this point (see lines 578-582):

   Further future work should also focus on applying interpretable machine learning techniques—such as saliency maps, feature attribution methods, or layer-wise relevance propagation—to better understand the features learned by the model. Enhancing interpretability is essential for building trust in data-driven damage indicators and supporting informed decision-making in structural health monitoring, especially if these are based on such indirect proxies for structural damage as aerodynamic pressure measurements.

7. Including added mass as a damage class, to simulate damage occurrence without actually damaging the specimens, has been done several times by well-known Authors. However, this Reviewer has never been fully convinced about the reliability of this approach. Apart from similar shifts in natural frequencies (of course, can be manipulated by decreasing the n-th modal stiffness or increasing the corresponding n-th modal mass), the structural implications of adding masses differ significantly from reducing stiffnesses (which itself is only an approximation of actual crack initiation, growth, and coalescence). Hence, further justification to use this approach should be provided, and its limitations should be fully and explicitly discussed.

   Reply: Thank you for your thoughtful remark. We understand your concern and would like to clarify that the added mass in the experimental setup is not intended to represent a different type of crack or structural damage. Instead, it is meant to simulate the effect of ice accretion ("icing") on a wind turbine blade, which also results in an increase in mass. We realize that this was not clearly explained in the original manuscript and apologize for any confusion this may have caused. We added an additional sentence in Section 3.3 "Design of Experiments" (see lines 250-251) to clarify this point in the revised manuscript:

While the cuts with different lengths should represent the stiffness reduction of cracks of different length, the added mass is meant to represent ice accretion, also known as "icing", on the wind turbine blade.

8. Apart from wind turbine blades, aerodynamic pressure plays a key, pivotal role in long-span suspended bridges – see e.g. https://doi.org/10.1016/j.proeng.2017.09.576, https://doi.org/10.1007/978-3-031-61425-5_46, https://doi.org/10.1061/JSENDH.STENG-12095. These works may be cited in the paper to provide a broader context.

   Reply: Thank you for this valuable suggestion. We were not aware of these publications and agree that long-span bridges may offer further interesting application scenarios for the Aerosense measurement system. We cite the suggested publications in the outlook (see lines 612-617) of the revised manuscript and address the application of the Aerosense system to long-span suspended bridges as follows:

   Finally, another interesting application scenario, beyond wind turbines, for the Aerosense measurement system and for aerodynamic pressure based damage detection, consists in slender, long-span suspension bridges. For such bridges, excitation by wind loading and thus aerodynamic pressure plays an essential role [9, 7, 10]. Here, the Aerosense system could not only contribute by monitoring the wind loading at selected points, but also by enhancing existing acceleration based damage detection methods by information derived from aerodynamic pressure data.

9. Similarly to comment 8, referring to Fig 15, saw cutting a specimen to simulate damage is extremely common and acceptable, but the difference between this unrealistic damage scenario, which is straight cut and thick, and actual real-life cracks, which are irregular in shape and hair-like thin, should be explicitly discussed.

   Reply: Thank you for your insightful remark. We agree with you and modify the manuscript at several locations to explain the difference between the introduced unrealistic damage scenario and actual, real-world cracks as follows:

   - lines 173-178: An important limitation of our setup lies in the choice of material and the manner in which damage is introduced. While real-world wind turbine blades are composed of layered composite materials exhibiting complex failure modes such as delamination and fiber breakage, our experiments employ an aluminum cantilever with damage emulated via saw cuts. This simplification allows for controlled, repeatable tests and the ability to systematically vary damage severity. Although the artificial crack does not fully replicate the morphology or fracture mechanics of a naturally occurring defect in composites, it produces a measurable stiffness reduction, which is central to our proof-of-concept study.
   - lines 589-592: Additionally, more advanced material models and realistic damage representations will be necessary to accurately account for variations in material properties and structural integrity. Future research efforts will aim to translate the proposed methodology to composite specimens to more closely align with practical applications in wind turbine blade monitoring.

10. The reference https://doi.org/10.5281/ZENODO.8018677 is linked to the powerpoint of a presentation; it would be better to use a peer-reviewed journal or conference article

instead.
Reply: Thank you for your comment. We only used one of the pictures for Figure 1 b) from the conference presentation of our co-author and thus cited it here initially. However, as we have not used any other contents from this presentation, we removed the citation in the revised version.

References

[1] R. J. Allemang and D. L. Brown. A correlation coefficient for modal vector analysis. In *1st International Modal Analysis Conference, Society for Experimental Mechanics, Orlando, FLorida, USA, pp. 110-116*, 1982.

[2] S. Barber, J. Deparday, Y. Marykovskiy, E. Chatzi, I. Abdallah, G. Duthé, M. Magno, T. Polonelli, R. Fischer, and H. Müller. Development of a wireless, non-intrusive, mems-based pressure and acoustic measurement system for large-scale operating wind turbine blades. *Wind Energy Science*, 7(4):1383–1398, 2022.

[3] R. Brincker, C. Ventura, and P. Andersen. estimation by frequency domain decomposition. In *Proceedings of IMAC 19*, pages 698–703, United States, 2001. Society for Experimental Mechanics. The International Modal Analysis Conference ; Conference date: 05-02-2001 Through 08-02-2001.

[4] R. Brincker, L. Zhang, and P. Andersen. Modal identification from ambient responses using frequency domain decomposition. In *IMAC 18 : Proceedings of the International Modal Analysis Conference (IMAC), San Antonio, Texas, USA, February 7-10, 2000*, pages 625–630, 2000. The International Modal Analysis Conference ; Conference date: 07-02-2000 Through 10-02-2000.

[5] R. Brincker, L. Zhang, and P. Andersen. Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures*, 10(3):441, jun 2001.

[6] E. Cheynet, J. B. Jakobsen, and J. Snæbjörnsson. Damping estimation of large wind-sensitive structures. *Procedia Engineering*, 199:2047–2053, 2017. X International Conference on Structural Dynamics, EURODYN 2017.

[7] M. Civera, M. M. Rosso, G. C. Marano, and e. Chiaia. Validation and comparison of two aoma approaches for the ambient vibration testing of long suspension bridges under strong wind loads. In C. Rainieri, C. Gentile, and M. Aenlle López, editors, *Proceedings of the 10th International Operational Modal Analysis Conference (IOMAC 2024)*, pages 475–484, Cham, 2024. Springer Nature Switzerland.

[8] J. Deparday, Y. Marikovskiy, I. Abdallah, and S. Barber. An aerodynamic measurement system to improve the efficiency of wind turbine rotor blades [preprint], 2025.

[9] A. Fenerci and O. Øiseth. The hardanger bridge monitoring project: Long-term monitoring results and implications on bridge design. *Procedia Engineering*, 199:3115–3120, 2017. X International Conference on Structural Dynamics, EURODYN 2017.

[10] K. A. Kvåle, A. Fenerci, Øyvind Wiig Petersen, A. Rønnquist, and O. Øiseth. Data set from long-term wave, wind, and response monitoring of the bergsøysund bridge. *Journal of Structural Engineering*, 149(9):04723002, 2023.

[11] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[12] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33(6):1821–1852, 2019.