





RESEARCH ARTICLE | APRIL 09 2024


A deep learning approach to classifying flow-induced vibration response regimes of an elliptical cylinder

Jonathan C. C. Lo ; Mark C. Thompson ; Kerry Hourigan ; Jisheng Zhao 




Physics of Fluids 36, 044110 (2024)

<https://doi.org/10.1063/5.0202497>



Physics of Fluids
Special Topic:
Flow and Climate
Guest Editors: Khaled Ghannam and Mostafa Momen
[Submit Today!](#)



A deep learning approach to classifying flow-induced vibration response regimes of an elliptical cylinder

Cite as: Phys. Fluids **36**, 044110 (2024); doi: 10.1063/5.0202497

Submitted: 5 February 2024 · Accepted: 23 March 2024 ·

Published Online: 9 April 2024



View Online



Export Citation



CrossMark

Jonathan C. C. Lo,^{1,a)}  Mark C. Thompson,¹  Kerry Hourigan,¹  and Jisheng Zhao^{1,2} 

AFFILIATIONS

¹Fluids Laboratory for Aeronautical and Industrial Research (FLAIR), Department of Mechanical and Aerospace Engineering, Monash University, Victoria 3800, Australia

²School of Engineering and Technology, University of New South Wales, Canberra ACT 2600, Australia

^{a)}Author to whom correspondence should be addressed: jonathan.lo1@monash.edu

ABSTRACT

This study proposes a new approach that leverages deep learning to the study of flow-induced vibration (FIV), specifically to automate flow regime classification and to visualize the transitions between these regimes. Using previously obtained data on the amplitude response of an elastically mounted cylinder as a function of reduced velocity for a range of structural damping ratios, the time trace of the body displacement and fluid driving forces are first converted into a frequency-time representation using continuous wavelet transforms before being input to several pre-trained convolutional neural networks for feature extraction. When utilizing the outputs of each convolutional neural network for regime classification, we found that almost all the machine learning approaches had high cross-validation accuracy that was statistically insignificant from each other. The five best-performing classifiers were then used as an ensemble method, yielding a weighted accuracy of 99.1% on the test data. The FIV response regimes were further investigated by projecting the outputs of the pre-trained convolutional neural networks onto the first three modes identified with principal component analysis (PCA). The PCA plots indicated that, among all the models considered, Xception showed superior capability in delineating distinct locations for different FIV response regimes, based on the lock-in frequency and the presence of harmonics in the driving fluid forces. Moreover, the PCA plots also showed that increasing the structural damping ratio resulted in a diminished disparity in the dynamics of the identified FIV response regimes, leading to a less discernible separation between the regimes in the plots.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0202497>

I. INTRODUCTION

Flow-induced vibration (FIV), arising from the complex coupled interaction between a fluid and a structure [often termed fluid–structure interaction (FSI)], is an aerodynamically resonant phenomenon important to many important areas of engineering. Often observed as the swaying of large structures, such as bridges and high-rise buildings, in strong winds, FIV is both detrimental in applications where structural failure or long-term fatigue is undesirable and advantageous as a potential source of renewable energy (Soti *et al.*, 2018; Lv *et al.*, 2021; and Wang *et al.*, 2017). As such, this has motivated ongoing research to characterize, predict, and control FIV (Govardhan and Williamson, 2000; Khalak and Williamson, 1996; Morse and Williamson, 2009; and Wong *et al.*, 2017).

The FIV response of an aeroelastic or elastically mounted bluff body can typically be characterized in terms of two distinct

phenomena: “vortex-induced vibration” (VIV) and “galloping.” VIV occurs as a result of the periodic shedding of vortices or at least shear layer vorticity variation (Menon and Mittal, 2020), from an elastic or elastically mounted body leading to a periodic cross-flow forcing. VIV is self-limited due to the breakup of the symmetrically alternating vortex pattern at large vibration amplitudes (Blevins, 1990). Galloping is instead driven by the asymmetric pressure distribution across the body caused by the changes in the instantaneous flow incidence angle as the body oscillates in the fluid. As both manifestations of FIV are dependent on the properties of the flow and the cylinder (e.g., flow velocity, Reynolds number, geometry, mass ratio, applied damping, and structural stiffness), many past works have chosen parameters such that VIV and galloping occur separately and can be individually investigated (Brooks, 1960). However, recent studies (see Nemes *et al.*, 2012;

Zhao *et al.*, 2018) have shown that rich and complex fluid–structure interactions can also be observed when both FIV phenomena concurrently manifest in the dynamic response of the elastic or elastically mounted cylinder system.

Such a system can be modeled as a linear second-order damped harmonic oscillator of the following form:

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = F_y(t), \quad (1)$$

which relates the total fluid force acting in the cross-flow direction [$F_y(t)$] to the system mass (m), spring constant (k), structural damping constant (c), and cylinder displacement [$y(t)$]. The oscillation amplitude (A), flow speed (U), and structural damping (c) are henceforth represented by their non-dimensional forms: the normalized amplitude parameter $A^* = A/b$, reduced velocity $U^* = U/(f_{nw}b)$, and the structural damping ratio accounting for added mass effects $\zeta = c/(2\sqrt{k(m+m_A)})$. Here, f_{nw} and m_A denote the natural frequency of the elliptical cylinder when submerged in quiescent water and potential added mass, respectively (Blevins, 1990; Sumner *et al.*, 2006). Additionally, b is the cross-stream (major) axis length of the cylinder. It is also noted that both the transverse lift and stream-wise drag forces are described in their non-dimensionalized form and are defined as $C_y = F_y(t)/(\rho U^2 bL/2)$ and $C_x = F_x(t)/(\rho U^2 bL/2)$, respectively, where ρ is the fluid density and L denotes the immersed length of the cylinder. The vortex force coefficient [$C_v = F_v(t)/(\rho U^2 bL/2)$], which describes the contribution of vortex shedding to the overall lift force, was computed through a decomposition of the transverse force into vortex and potential [$F_p(t) = -m_A\ddot{y}(t)$] force components based on $F_v(t) = F_y(t) - F_p(t)$ [see Govardhan and Williamson (2000); Morse and Williamson (2009); and Zhao *et al.* (2014a); (2014b)].

One example of the co-manifestation of VIV and galloping relates to cylinders with elliptical geometries as parameterized by the elliptical ratio $\varepsilon = b/a$, where a and b are the stream-wise and cross-flow (transverse) dimensions, respectively. Since the circular cylinder is considered a special case of the elliptical geometry (with $\varepsilon = 1$) whose FIV response is VIV due to the inherent symmetry of the system (see the reviews of Williamson and Govardhan, 2004; Bearman, 1984), numerous studies have shown that the introduction of deformation, where $\varepsilon > 1$, causes the flow induced by a transversely oscillating elliptical cylinder to become more unstable (Hall, 1984; Navrose *et al.*, 2014; Zhao *et al.*, 2019; and Vijay *et al.*, 2020). However, it was only recently when Lo *et al.* (2023) investigated the effect of structural damping on the FIV dynamics of a $\varepsilon = 5$ cylinder (see Fig. 1) that large-scale oscillation amplitudes, previously unseen for other geometries at similar U^* values, was discovered. Coined the “hyper-branch,” the maximum amplitude of $A \approx 8b$ observed for this region is only limited by the width of the water channel facility in which the experiments were conducted. Additionally, the presence of a secondary vortex street implies that flow attachment occurs as the cylinder translates across the oncoming flow and further indicates that the approximately factor-of-five increase in maximum amplitude over the studies by Zhao *et al.* (2019) and Vijay *et al.* (2020) can be attributed to the combined effect of VIV and a galloping-like movement-induced instability in the hyper-branch. Lo *et al.* (2023) also showed that the oscillation amplitude is reduced with increasing structural damping, with the hyper-branch response suppressed for $m^*\zeta \geq 0.334$.

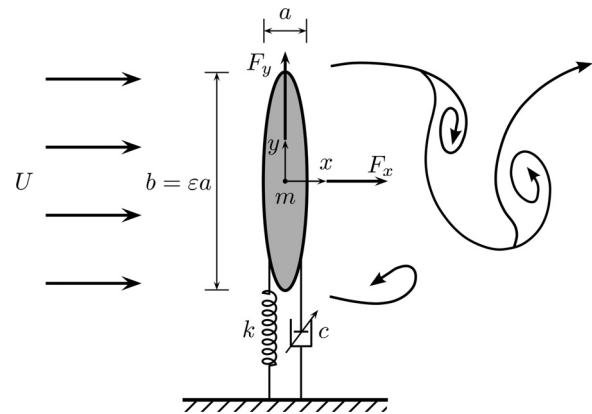


FIG. 1. A schematic defining the problem of interest: an elastically mounted elliptical cylinder model constrained to oscillate transverse (y) to the free stream flow of velocity U and in the positive x direction. Here, the geometry is characterized by the elliptical ratio $\varepsilon = b/a$, where a and b are the stream-wise and cross-flow dimensions, respectively. Additionally, m is the oscillating mass, k denotes the spring constant, c is the adjustable structural damping, and F_x and F_y represent the respective drag and the transverse (lift) fluid forces acting on the body.

While the above-mentioned parametric study has demonstrated the varied FIV responses that arise from different fluid and structural variables, identifying and characterizing the oscillatory responses and their corresponding wake modes into different flow regimes (e.g., VIV response, the hyper-branch response) were conducted based on a manual classification by the researcher. Although each identified flow regime is well-defined and based on the time traces of body vibration and fluid forces acting on the cylinder, their relationships with other regimes remain unclear. As such, with the literature spanning different ranges of structural and fluid variables, it becomes a challenge to determine if regimes found in previous studies under different names arise from similar FIV dynamics. Additionally, although such a manual process is reasonable for investigations that focus on the effect of a single parameter on the amplitude response (as a function of U^*), this becomes, at best, inconvenient when multiple system parameters affect the FIV dynamics.

To address these issues, a deeper understanding of the fluid–structure coupling may be gained by using a data-driven approach to analyze the relationships between the dynamic response (e.g., cylinder displacement and force time traces) and the corresponding wake structure. An advantage to this approach is that prior physics knowledge or mathematical models of the system are not required for machine learning methods to detect patterns or relationships from the data, although, of course, it is important to relate predictions back to the flow physics after the machine learning model has done its work. Furthermore, since each type of wake structure can be treated as a label, similarities can be drawn between the classification problem in other engineering fields where signals are categorized into one of multiple possible unique classes. As emphasized in the review by Brunton *et al.* (2020), machine learning has succeeded in handling the non-linear nature of fluid dynamics in applications like flow feature extraction, modeling flow dynamics, as well as flow optimization and control. For the study of FIV, in particular, examples of problems where data-driven approaches have been applied include predicting

the crosswind vibrations of rectangular cylinders (Lin *et al.*, 2021) as well as improving the FIV prediction of existing semi-empirical models (Shi *et al.*, 2023).

However, to the best of our knowledge, leveraging machine learning methods to study the different flow regimes and wake structures of elastically mounted cylinders undergoing FIV has not been attempted, thus far. Fortunately, signal classification is an area of great interest due to its applicability in many research fields where one-dimensional time-series sensor data are important. Notable examples include the medical field, where diagnosing different types of cardiac arrhythmias and the identification of epilepsy requires interpretation of biosignals, such as electrocardiogram (ECG) and electroencephalogram (EEG) data (Hong *et al.*, 2020; Abbasi and Goldenholz, 2019); audio processing, where signals from a microphone are used for voice recognition; and anomaly detection, which is used to identify fraudulent transactions (Priscilla and Prabha, 2019) and to monitor faults in real-time during machining operation (Teti *et al.*, 2010). Although common machine learning algorithms to tackle classification problems are k -nearest neighbors, support vector machines, and tree-based algorithms (i.e., decision trees, random forest, boosted decision trees, etc.), the aforementioned approaches require human experts to represent the raw signal with insightful features (e.g., statistical parameters, such as mean, standard deviation, and root mean square) (Celin and Vasanth, 2018). While this technique of “feature engineering” is useful in simplifying the classification problem, the model accuracy is dependent on the quality of the chosen features, with the latter often only applicable to a specific problem (Chollet, 2022).

To address this challenge, there has been significant interest in deep learning algorithms that forego the need for feature engineering and can instead “learn” the features directly from the raw input. Due to the temporal aspect of signal data, a type of deep learning model known as the recurrent neural network (RNN) is often employed. Recurrent neural networks iterate through the entire signal sequentially while also taking into account the values that the signal had taken before, thus enabling the long-range temporal context of the signal to be captured. Salloum and Kuo (2017) showed that recurrent neural networks were able to interpret ECG signals for biometric and authentication identification, while Pham (2021) showed that both time-space and time-frequency features of the signal can be used as inputs to a long short-term memory (LSTM) recurrent neural network to detect irregular heartbeat from ECG data and Parkinson’s disease from gait data. However, as noted by Vaswani *et al.* (2017), the sequential nature of recurrent neural networks precludes their parallelization during training and, as such, can be computationally inefficient.

Another deep learning approach utilizes continuous wavelet transform (CWT) to first represent the time-frequency information of each signal in a two-dimensional matrix (Janke and Diener, 2017; Tran *et al.*, 2020). This array is then used as an input for a deep learning architecture known as a convolutional neural network (CNN) to perform the classification. By interpreting the input as a collection of different spatial features and patterns, convolutional neural networks have greater parallelism (Bradbury *et al.*, 2016) and a faster training time (Weytjens and Weerd, 2020) than recurrent neural networks. Additionally, when comparing the performance of recurrent neural networks and convolutional neural network architectures for non-audible speech classification and crop identification, Fernandes *et al.* (2019) and Zhao *et al.* (2019), respectively, concluded that the

convolutional neural network approach performed with greater accuracy. Another advantage, as demonstrated by Narin (2020) for the detection of focal and non-focal epileptic seizure, is that the resemblance of the continuous wavelet transform outputs to an image means that a pre-trained model (a saved neural network previously trained on a large image dataset) can be utilized instead of training a model from scratch. As such, less data and shorter training time are required to achieve the same performance since the knowledge already learnt by the pre-trained model is “transferable.” Finally, as flow regime classification in FIV research is predominantly driven by the analysis of frequency components in the time trace signals, the use of continuous wavelet transforms to represent the frequency-time information is a natural framework in which to elucidate the relationships between the dynamic response and wake modes of an oscillating cylinder.

In summary, the current literature has shown that the FIV response of elliptical cylinders is dependent on the parameters of the fluid–structure interaction, such as elliptical ratios, reduced velocities, and Reynolds number. Although past studies have manually identified the different flow regimes and their associated wake modes, there is a lack of knowledge on the relationship with other regimes, as well as with those identified in previous studies (Zhao *et al.*, 2019; Lo *et al.*, 2023). This, along with the intractability of performing regime classification with the constant supervision of the researcher for studies spanning multiple structural and fluid variables, restricts the ability to scale parametric studies to incorporate a more holistic understanding of the underlying fluid–structure coupling that drives the FIV dynamics. With machine learning becoming more prominent in fluid dynamics research, this issue may be addressed with a data-driven approach. Specifically, the deployment of continuous wavelet transforms and convolutional neural networks in combination shows potential as a classification strategy, owing to their success in other research fields.

The article proceeds by outlining the methodology in Sec. II, which includes the dataset (Sec. II A) and machine learning approaches used (Sec. II C). Next, the application of the machine learning model as a classification tool and as a method to compare and distinguish between FIV regimes is discussed in Secs. III A and III B, respectively. Finally, the conclusions are drawn in Sec. IV, highlighting the important findings and the significance of the current study.

II. METHODOLOGY

This study will primarily be conducted on a dataset obtained by Lo *et al.* (2023), with the machine learning (ML) methodology implemented using Python and its additional open-source ML-related packages. Continuous wavelet transforms are implemented using the PyWavelets (Lee *et al.*, 2019) package, while Keras (Chollet *et al.*, 2015) and Tensorflow (Abadi *et al.*, 2015) packages are used to build and train the neural networks. To gain greater insight into the outputs from the deep learning model, as later discussed in Sec. III B, principal component analysis (PCA) is implemented using the scikit-learn library.

Figure 2 outlines the general machine learning workflow used to classify the FIV response based on the body displacement and transverse fluid force components (i.e., the total lift force, $C_{L,t}$, and the vortex force, $C_{L,v}$) acting on the elliptical cylinder. Starting with the time traces of these signals [Fig. 2(a)] obtained by Lo *et al.* (2023), a continuous wavelet transform is used to represent the time-frequency and amplitude information of the measurements as a two-dimensional array [Fig. 2(b)]. Since the result (known as “scalograms”) resembles an

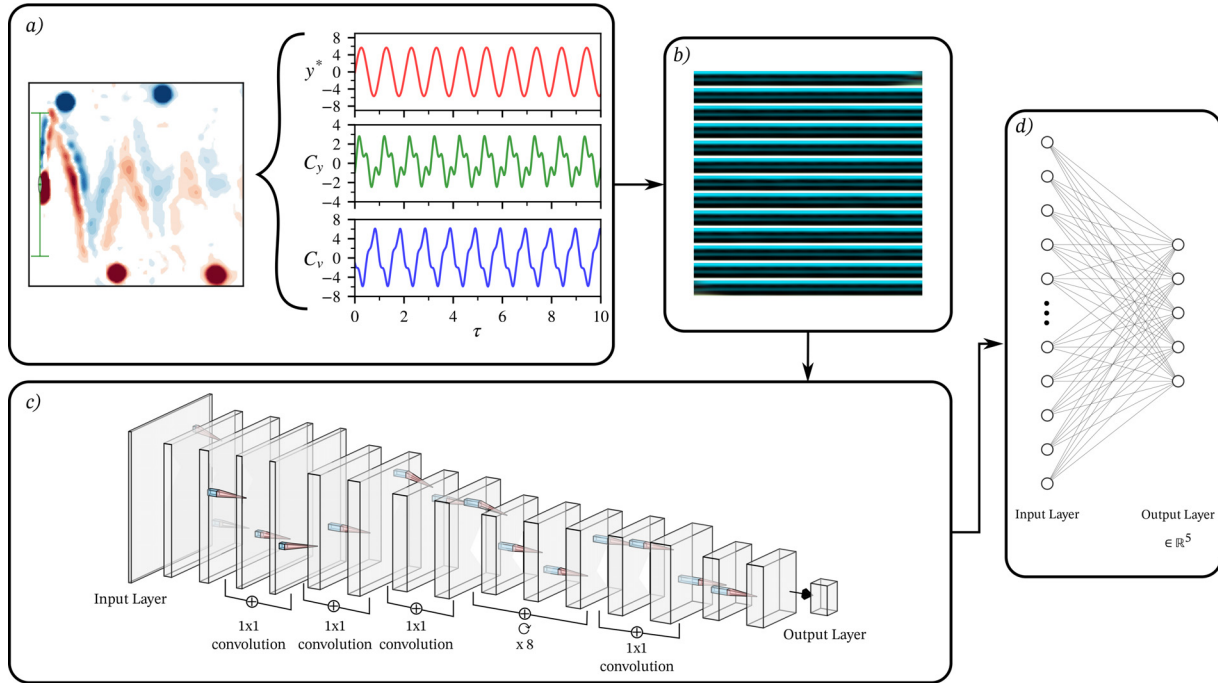


FIG. 2. Overview of the machine learning workflow of this study. (a) The normalized amplitude y^* , lift coefficient C_y , and vortex force coefficient C_v time traces resulting from the fluid–structure interaction between the flow and the cylinder, as a function of oscillation period τ , is converted into (b) scalograms using a continuous wavelet transform with complex Morlet wavelets. (c) The convolutional neural network (the graphical representation of the Xception model is shown here as an example, where $\odot \times 8$ symbol denotes the number of times the section has been repeated) is then utilized to further transform the scalogram, with the resultant outputs used as inputs for (d) a fully connected neural network classifier.

image, a convolution neural network [Fig. 2(c)] is then used to further extract the key features of the signals required by the feed-forward neural network classifier [Fig. 2(d)] to perform regime identification. Further details about the dataset can be found in Sec. II A, while the continuous wavelet transform and deep learning implementations (i.e., the convolution neural network and feed-forward neural network) will be discussed in Secs. II B and II C, respectively.

A. Dataset for machine learning training

Using the setup as defined by the schematic in Fig. 1, Lo *et al.* (2023) conducted a parametric investigation to obtain the effect of damping ratios on the FIV response as a function of reduced velocity and whose data (i.e., body vibration and fluid force time traces) will be used for the training and testing of models in this study. The experiments were conducted within the free-surface recirculating water channel of the Fluids Laboratory for Aeronautical and Industrial Research (FLAIR) at Monash University. To implement free transverse vibration, the $\varepsilon = 5$ elliptical cylinder was mounted on a low-friction air-bearing rig and was elastically constrained by high-precision extension springs. The mass ratio of the total oscillating system was $m^* = m/m_d = 17.4$, where m_d is the mass of the displaced water. The structural damping of the air-bearing rig was controlled using an electromagnetic damper system developed by Soti *et al.* (2018). Further details on the experimental methodology, including the elliptical cylinder, air bearing rig, and damper system, can be found in Lo *et al.* (2023).

The fluid–structure interactions between the fluid flow and elliptical cylinder were investigated for structural damping ratios over the range $3.62 \times 10^{-3} \leq \zeta \leq 1.87 \times 10^{-1}$ for reduced velocities $2.3 \leq U^* = U/(f_{nw}b) \leq 10.0$. Measurements at each U^* value were conducted for 300 s with a sampling rate of 100 Hz. The corresponding free-stream velocity and Reynolds number in the water channel were $40 \leq U \leq 180$ mm/s and $980 \leq Re = UD/\nu \leq 4410$, respectively, where ν is the kinematic viscosity of the fluid. The free-stream turbulence level was less than 1% over the flow velocities of interest. To further explore the movement-induced nature of the different FIV regimes and the effect of hysteresis, experiments with increasing and decreasing reduced velocities were conducted.

After characterizing the FIV response based on the body displacement and fluid forces acting on the elliptical cylinder, Lo *et al.* (2023) identified four main flow regimes corresponding to non-negligible amplitudes as shown by the time traces [i.e., $y^*(t)$, $C_y(t)$, and $C_v(t)$] in Fig. 3 as well as the frequency decomposition of both the transverse lift and the vortex forces in Fig. 4. These regions arise from the vortex shedding frequency “locking” onto the body oscillation frequency at close to the natural frequency of the system when submerged in quiescent water, $f^* \approx f_{C_y}^* \approx f_{nw}$. In line with the naming convention by Lo *et al.* (2023), the subplots of Figs. 3(a)–3(d) and 4(a)–4(d) both correspond to regime I, regime II, hyper-branch regime, and regime III, respectively.

While regime I is mainly characterized by C_y oscillating at $f^* = 1$, the appearance of regime II is marked by the development of

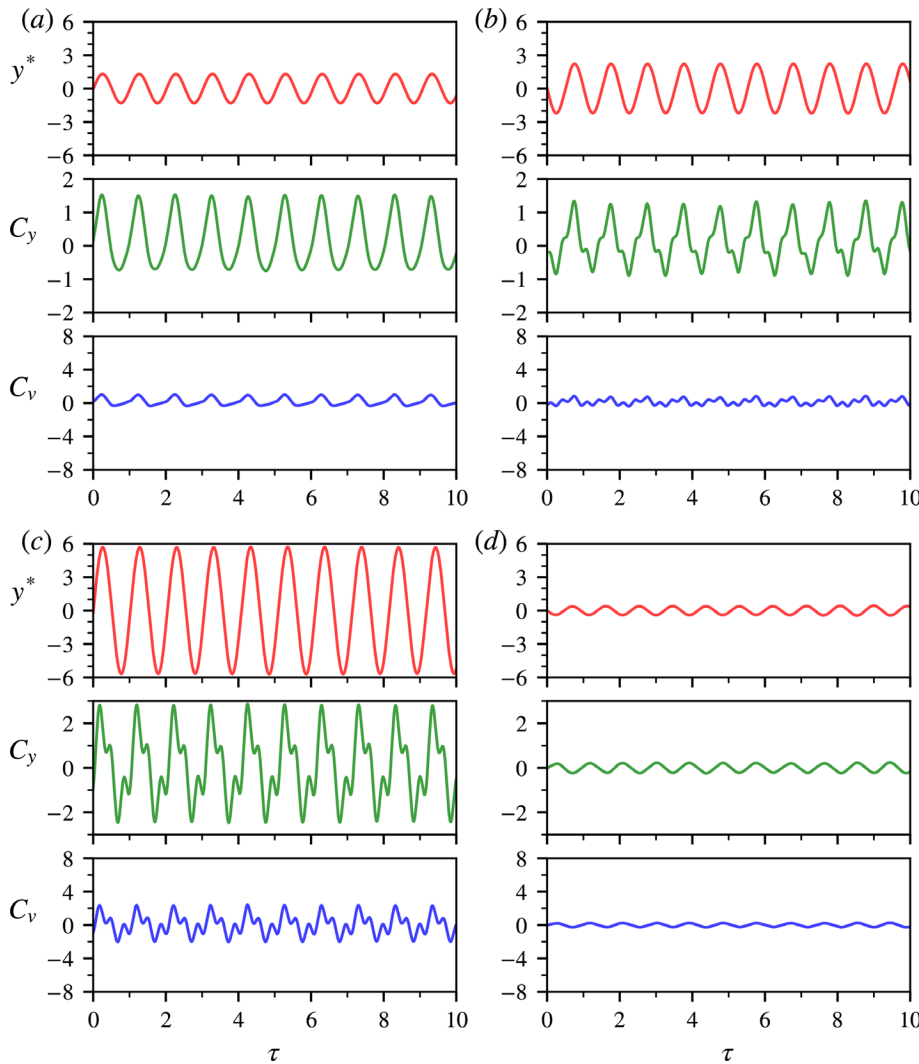


FIG. 3. Sample time traces of the cylinder vibration for damping ratio $\zeta = 3.62 \times 10^{-3}$ at different reduced velocities corresponding to the four identified lock-in regimes: (a) $U^* = 3.0$ (I), (b) $U^* = 4.0$ (II), (c) $U^* = 6.0$ (hyper-branch regime), and (d) $U^* = 8.0$ (III). Note that the time as indicated by the horizontal axis is scaled by the natural system frequency such that $\tau = t f_{nw}$.

the second and third harmonics in the transverse lift and vortex forces. In the hyper-branch regime, large vibrational amplitudes are observed along with a slight drop in the third harmonic of C_y and C_v to below $f^* = 3$. Finally, regime III is identified by the transverse fluid and vortex forces oscillating at a frequency marginally above the natural frequency f_{nw} , which increases slowly with U^* . In addition to the aforementioned “lock-in” regimes, a fifth regime known as the desynchronized region was observed and corresponds to chaotic amplitudes with frequency contributions mainly by vortex shedding at the Strouhal frequency. The contour map in Fig. 5 shows how the FIV amplitude of the cylinder varies with reduced velocity and damping ratio as well as the location of each flow regime in the $U^* - \zeta$ parameter space.

As shown in Fig. 6, particle image velocimetry (PIV) was then conducted to elucidate the wakes associated with the vibrational response of the elliptical cylinder in each regime. While no discernible flow patterns were observed in the desynchronized regime, the alternating shedding of two singular vortices from either side of the

cylinder every oscillation cycle was the main wake structure present in the four lock-in regimes. Labeled the 2S wake mode, the vortices are shed approximately in-phase relative to the body vibration, except for regime III [Fig. 6(d)], where vortex shedding occurs in an anti-phase with a delay of 180° . In addition to the main 2S mode that was the sole flow feature in regime I [Fig. 6(a)], regime II [Fig. 6(b)], and the hyper-branch regime [Fig. 6(c)] are also characterized by the presence of a continuum of starting vortices (i.e., the secondary vortex street) that form a zigzag pattern aft of the cylinder. For regime III [Fig. 6(d)], the ancillary structure is a thin shear layer that forms a tail connecting the major vortices to the elliptical cylinder. These additional wake structures are the source of the second and/or third harmonics in the transverse fluid force (i.e., C_y , C_v) frequency responses.

The study by Lo et al. (2023) contains FIV responses for 18 structural damping ratios, with a total of 3255 points in the $U^* - \zeta$ parameter space. The distribution of samples across the five identified wake-mode classes is skewed, with approximately 12% of the data belonging to regime I, 4% in regime II, 10% in the hyper-branch regime, 34% in

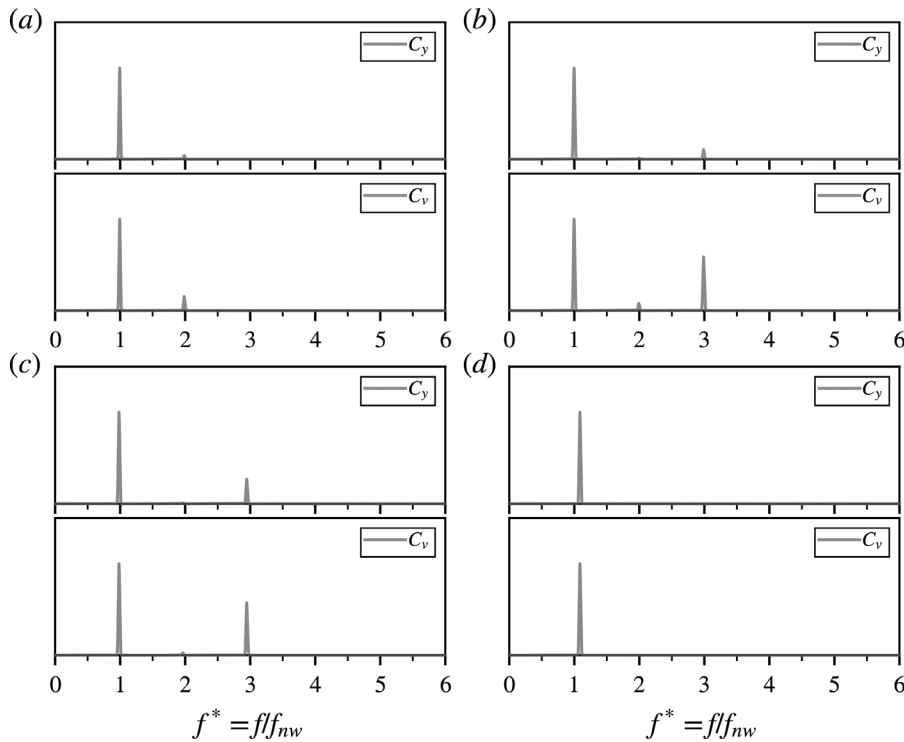


FIG. 4. Frequency decomposition of the transverse lift C_y and vortex C_v coefficient time traces for the four identified lock-in regimes when $\zeta = 3.62 \times 10^{-3}$: (a) regime I ($U^* = 3.0$), (b) regime II ($U^* = 4.0$), (c) hyper-branch regime ($U^* = 6.0$), and (d) regime III ($U^* = 8.0$). Note that the spectral power as denoted by the y axes is normalized relative to the largest frequency peak.

regime III, and 40% in the desynchronized region. As such, the wake-mode identification part of this study as discussed in Sec. III A becomes an imbalanced classification problem. To have a true measurement of model accuracy, training data are used to train and optimize the machine learning model, while the test set is withheld and exclusively used to provide an unbiased out-of-sample evaluation of the final model accuracy at the end of the study. In this study, 60% of the total data points was used for model training and validation while the remaining 40% of the data was taken as the 1302-point test set. To ensure that both sets reflect the total database, stratified splitting was performed such that the percentage of samples belonging to each class is identical across the partitions.

B. Continuous wavelet transform (CWT)

Utilizing the dataset and the knowledge gained from the study by Lo *et al.* (2023), the relationship between the wake and the structural response of the elliptical cylinder is ascertained by first transforming the time trace signals into a form that is amenable to the deep learning architecture shown in Figs. 2(b) and 2(c). While the frequency information is important in determining the physics of the FIV system for a given U^* and ζ as discussed in Sec. II A, the temporal nature of the signal is also important in determining if any time-dependent phenomena are occurring. Examples include the branch-mode competition observed by Zhao *et al.* (2018) for the square cylinder and the intermittent mode-switching between the upper and lower branches of the circular cylinder FIV response at low mass ratios. An approach that allows the characteristics of the structural response in both the frequency and time domain to be captured is the continuous wavelet transform (CWT). Mathematically, this process is a cross correlation

between a signal $s(t)$ (standardized with the z-score) and a reference wavelet $\Psi(t)$ such that

$$W_s(n, m) = \left\langle s(t), \Psi_{n,m} \left(\frac{t-m}{n} \right) \right\rangle = \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} s(t) \cdot \Psi^* \left(\frac{t-m}{n} \right) dt, \quad (2)$$

where the time-frequency representation of the signal W_s is a function of the reference wavelet's scale n and time shift m (Tary *et al.*, 2018). Wavelets are a family of functions localized in time and have zero mean and unit energy. The complex Morlet wavelet is used, taking the following form:

$$\Psi(t; B, C) = \frac{1}{\sqrt{\pi B}} \exp^{-\frac{t^2}{C}} \exp^{j2\pi Ct}, \quad (3)$$

where B and C are the bandwidth and center frequency, respectively, and $\Psi^*(t)$ is its complex conjugate. Since the inner product $\langle \cdot \rangle$ at time t is maximum when both functions are identical in frequency space, continuous wavelet transforms capture both the frequency and time information of the signal. The wavelet used to interrogate the $s(t)$ signal can be considered as a complex sine wave localized in time (due to the minimal amplitude away from the signal center), so the role of m is to translate the center of the wavelet to $t = m$ while n scales the frequency C . The resultant representation after taking the absolute value $|W_s(n, m)|$ is known as a scalogram and is plotted as a two-dimensional contour with time (m) and frequency (n) as the axes. Their resemblance to an image allows the product of the continuous wavelet transform to be used as inputs for the convolutional neural network discussed in Sec. II C 1.

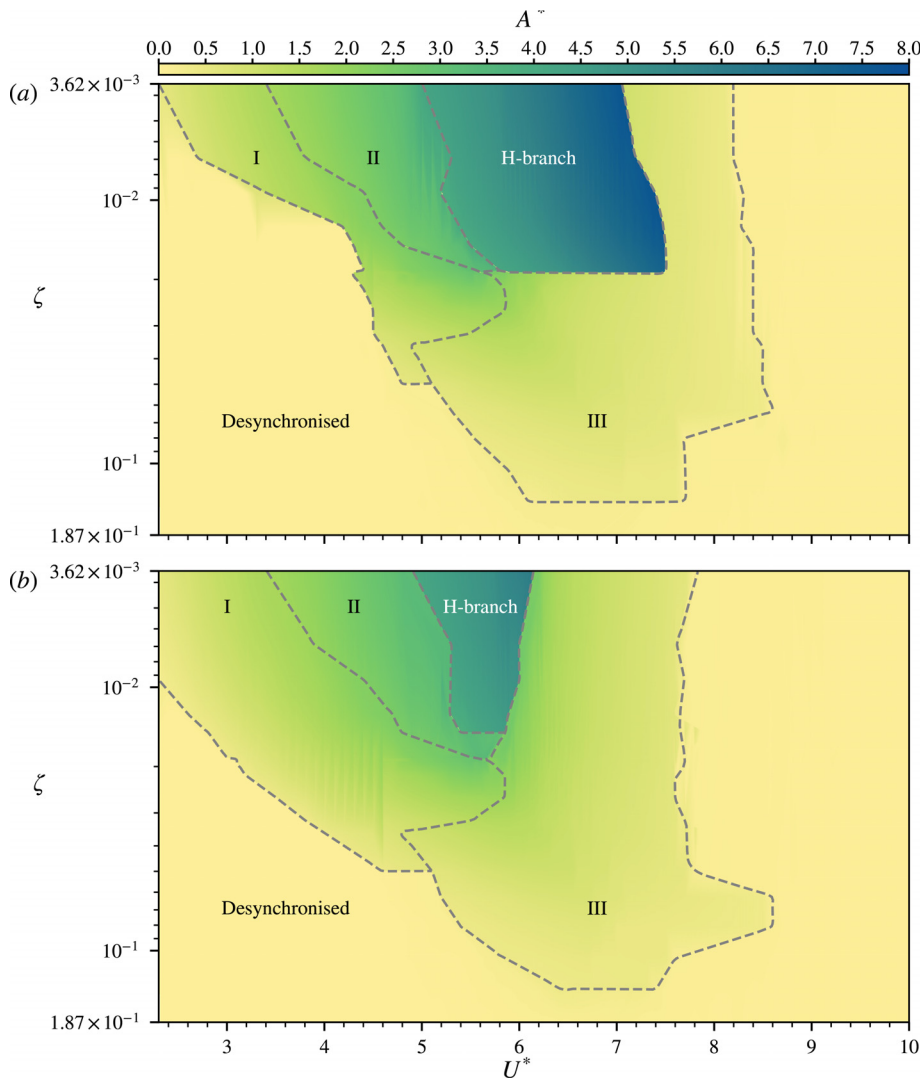


FIG. 5. Two-dimensional contours of the amplitude response (data taken from Lo et al., 2023) in $U^* - \zeta$ space. The parameter space is divided into five unique vibration regimes based on amplitude, phase, and frequency responses: regime I, regime II, hyper-branch regime (H-branch), regime III, and the desynchronized region. The dashed line marks the approximate boundaries of each region. (a) corresponds to U^* increasing and (b) to U^* decreasing.

Due to the wake-body synchronization that occurs during the lock-in regimes in the FIV response of the elliptical cylinder, the complex Morlet wavelet used in this study was chosen to have a center frequency of $C = f_{nw}$ and a bandwidth B corresponding to three oscillation cycles. The frequency scales of interest include 23 points uniformly distributed over a range of $a \in [0.9, 3.1]f_{nw}$. As will be discussed in Sec. II C 1, the convolutional neural network accepts three-dimensional arrays as inputs, where the first two dimensions denote the width and height of the image while the last dimension represents the intensities in the red, green, and blue channels.

To maximize the information available to the deep learning model, three different time traces can be simultaneously incorporated into the scalogram image corresponding to each of the three color channels. The cylinder displacement (y^*), cross-flow lift force (C_y), and vortex fluid force (C_v) were specifically chosen as the parameters that best describe the FIV response due to the system being constrained to only transversely oscillate, with the inclusion of C_v to account for the vortex shedding dynamics. Furthermore, additional

knowledge about the amplitudes of each sample can be encoded into the scalograms by normalizing the color intensities of each channel in the image array relative to some fixed reference. This normalization method is mathematically defined as

$$I(i, j, c) = \frac{M(s(t; c))}{R(c)} \left(\frac{W_s(i, j; c)}{\max W_s(i, j; c)} \right), \quad c = \{0, 1, 2\}, \quad (4)$$

where i and j refer to the position along the width and height, respectively, for a scalogram image. In order, c refers to the red, green, and blue color channels, with each channel corresponding to a time trace such that $s(t; c) = [y^*(t), C_y(t), C_v(t)]$. $M(s)$ is the maximum value for a given $s(t; c)$, which is the mean of the top 10% of amplitude peaks for y^* , and is the maximum root mean square value for the lift and vortex forces. Finally, $R(c) = [8, 1.8, 1.8]$ is the fixed reference for each parameter and was determined *a priori* by first identifying the maximum value of the time traces across the entire dataset. Figure 7 shows an example of the resultant images for each lock-in regime after

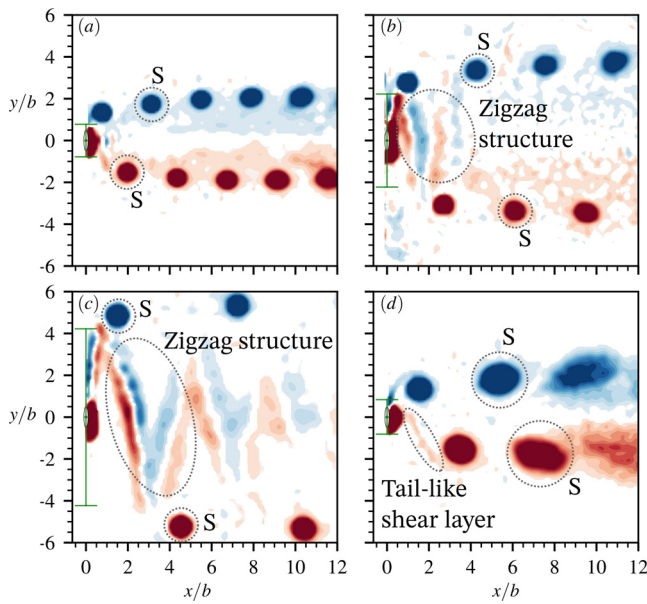


FIG. 6. Phase-averaged normalized vorticity contours of the wake structures downstream of the elliptical cylinder when the cylinder is in the center position and moving downward. Here, the flow evolves from left to right with the blue and red contours representing clockwise and anti-clockwise vorticity, respectively. In each plot, the green dot and vertical line between two horizontal bars denote the body center position and peak-to-peak vibration amplitude of the cylinder, respectively, while a single shed vortex every half-cycle is denoted by S. The normalized vorticity field $\omega_z^* = \omega_z b / U$, where ω_z is the vorticity out of the x - y plane. The FIV flow regime and (U^*, ζ) value corresponding to each frame is (a) regime I— $(3.5, 1.88 \times 10^{-2})$, (b) regime II— $(4.0, 3.58 \times 10^{-3})$, (c) hyper-branch regime— $(6.0, 1.88 \times 10^{-2})$, and (d) regime III— $(7.2, 1.88 \times 10^{-2})$.

performing continuous wavelet transform on the time trace signals and subsequently applying normalization. Scalograms of the desynchronized regime are not presented here due to the negligible vibrational dynamics causing the resultant images to be almost completely black.

The appearance of each sample, when expressed using the CWT method as described above, will be unique and dependent on the fluid forcing, vortex forcing, and vibrational dynamics of the elliptical cylinder undergoing cross-flow FIV for a given U^* and ζ value. Thus, the role of the machine learning models will be, using the characterization of the different FIV responses by [Lo et al. \(2023\)](#) as the “ground-truth,” to identify the common visual features in the scalogram that allows one regime to be distinguished from another.

C. Deep learning architectures

Once the time trace signals have been converted into a scalogram using the continuous wavelet transform, the relationship between the resultant contour (in the form of an image with red, green, and blue color channels) and the wake structure will be identified using both the convolutional as well as a feed-forward neural network. Acting as universal function approximators, an advantage of this approach is that the complex connection between the inputs and outputs does not need to be derived analytically but instead can be mimicked using pre-obtained data. In implementations of deep learning, the “depth” of the aforementioned neural networks refers to the numerous successive layers of representation utilized to transform the data into a high-dimensional space such that patterns can be clearly extracted. A typical architecture consists of an input layer, hidden layers that provide the main computation power of the model, and an output layer that generates the prediction. As each layer consists of a basic tensor operation, a chain of these layers can be likened to a multistage process where the

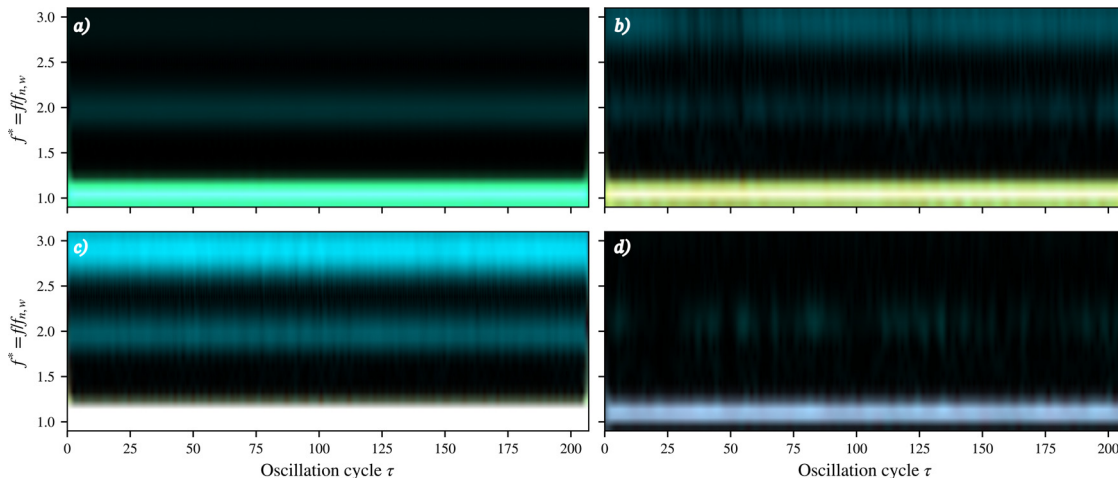


FIG. 7. Contour plot of the scalograms after the continuous wavelet transform was applied to the 300-second long time trace obtained at $\zeta = 3.62 \times 10^{-3}$ in decreasing U^* increment when $U^* =$ (a) 2.5, (b) 4.0, (c) 5.5, and (d) 7.0. The frequency resolution on the vertical axis is $0.1 f_{nw}$ over a range from $f^* = 0.9$ to 3.1. The resultant RGB colors are obtained by the magnitude of each scalogram value (normalized by their maximum). RGB channels: red—body displacement, blue—transverse fluid force, and green—vortex force. N.B.: for better contrast, the intensity of the images has been increased by a factor of five.

data progressively become more “meaningful” with respect to some desired task. Since both the convolutional and feed-forward neural networks consist of parameters that dictate how the inputs are related to the outputs, a process known as training is utilized to determine, from the existing experimental measurements, the parameters that would correctly identify the relationships between the inputs and the outputs. Mathematically, this process can be expressed as determining the model parameter values that minimize the loss function, which defines the error between the predicted regime and the true value observed in the data (i.e., the “label”). As the problem of interest involves multi-class, single-label classification, the categorical cross-entropy was chosen to be the loss function and is defined as

$$\ell = - \sum_{i=1}^n t_i \log(p_i), \quad (5)$$

which sums over all n classes, where t_i is the truth label (i.e., the true flow regime) and p_i is the probability determined by the model that the given input belongs to the i th class. As there are a total of five possible flow regimes in the FIV response, the labels are one-hot encoded such that they are represented as an array of length n where each position in the array corresponds to a given class. As such, $t_i = 1$ only occurs for the location of the flow regime we wish to represent while t_i is 0 for the other positions.

Although deep learning models are trained by minimizing the loss, the model performance is measured using a metric instead. As such, the former acts as a proxy for the latter since some desired metric functions are not differentiable and hence cannot be optimized via gradient descent methods used by neural networks during model training. In this study, the loss minimization using gradient descent is conducted using the “RMSprop” optimizer (Hinton *et al.*, 2012) in batches of 32 samples with a learning rate of 2×10^{-5} . To account for the class imbalance in the dataset, the performance of the models will be evaluated using the “weighted accuracy” (i.e., the fraction of the total number of correct predictions) metric where the individual contributions by a sample to the loss is dependent on the class that the data point belongs to. The weighting for samples in the n th class w_n is taken as

$$w_n = \frac{P_{\text{tot}}}{N_{\text{class}} P_n}, \quad \{n = 1, 2, \dots, N_{\text{class}}\}, \quad (6)$$

where a ratio of the total number of points in the total dataset P_{tot} to the number of classes N_{class} is used to scale the P_n points in the n th class. As such, this allows under-represented wake modes with few samples (such as regime II) to be as equally important to the model as other categories with many examples.

Beyond the aforementioned parameters found during the training of deep learning models, hyper-parameters also affect model performance. Defined as a parameter whose value controls the learning process, hyper-parameters must be optimized outside of model training by training the model multiple times to find the combination that yielded the best performance as measured by the metric. The only hyper-parameter of interest in this study, which was deemed to have the strongest influence on model performance, is the number of epochs (i.e., the number of times the entire training dataset was used during training).

However, this hyper-parameter tuning process is susceptible to a detrimental phenomenon known as “overfitting.” A model is

considered to be overfitted when a model starts to learn the noise (i.e., irrelevant information) specific only to the training data, at the expense of its predictive ability to generalize to unseen data based on patterns learnt during training. As such, to accurately ascertain the effect of hyper-parameters (i.e., the number of training epochs) on the loss, k -fold cross-validation is often used to minimize the effect of overfitting on the final estimate. For this approach, the training data are split into k sets, where one set is selected for validation and is used to evaluate the performance of a model trained on the remaining $k - 1$ sets. This process is repeated k times such that each set is the validation set once, with the resultant k performance scores averaged to provide the final estimate. Fivefold cross-validation was utilized in this study with stratified sampling conducted to ensure the same ratio of flow regimes in all the sets. To account for the class imbalance in the dataset during training, an oversampling approach is utilized, whereby the minority classes are randomly sampled with replacement such that the number of instances in each class is balanced. As the models are trained in a batch-wise manner, this method allows for a smoother loss gradient signal over the weighting method as a sample belonging to a minority class will appear in many different batches with equal weighting instead of just one batch with a large weight.

The specifics of the layers that make up the convolutional neural network and the feed-forward neural network classifier used in this study will now be discussed in Secs. II C 1 and II C 2, respectively.

1. Feature extraction using convolutional neural networks

As their name suggests, the interpretation power of the convolutional neural network arises from the convolutional operation conducted by each layer of the model to learn local patterns in the input images. Conceptually, the convolutional layer can be thought of as passing the inputs through different filters to spatially locate the positions corresponding to the specific features (i.e., corners, curves, etc.) that are of interest for each filter (Chollet, 2022). This is implemented by sliding a n -by- n window over the input array, and at each window location, taking a tensor product of the elements bounded by the window with the convolutional kernel k . After adding a constant offset called *bias* \bar{b} , the output of the convolution layer is calculated by applying an activation function on the resultant tensor to introduce non-linearity. Mathematically, this is described as

$$\vec{O}^{(i)} = A \left(\vec{O}^{(i-1)} * \vec{k}^{(i)} + \bar{b}^{(i)} \right), \quad (7)$$

where $O^{(i-1)}$ and $O^{(i)}$ are the output of the previous and current layers, respectively. Although the initial image input has three channels corresponding to the color intensity of red, green, and blue, each channel of the outputs refers to a particular filter in the kernel. An advantage of such an approach is that the patterns learnt by the convolution layers are generalizable and can be recognized anywhere in the image. Additionally, the use of successive layers allows the model to represent increasingly larger patterns in the later layers as a conglomeration of smaller features captured in the initial layers. As such, the translation invariance and ability to identify the spatial hierarchy of patterns allow complex and abstract visual concepts, particularly of images, to be learnt in a data-efficient manner.

In the case of this study, scalograms are used as the image inputs instead. Since a scalogram represents the frequency and time

information of a signal across its vertical height axes and horizontal width axes, respectively, the outputs of the convolutional neural network can hence be thought of as an interpretation of the FIV-related signals based on the latter's frequency and time characteristics. As such, the patterns and bands that appear (Fig. 7) in, as well as their location within, the scalogram are the main features used by the machine learning model to categorize the different samples of the dataset into the five different FIV regimes identified by Lo *et al.* (2023).

Although convolutional neural networks are well-suited to classifying image-like scalograms, many samples are usually required to train an accurate model. To overcome this limitation, a “transfer learning” approach is undertaken by re-purposing the hidden layers of pre-trained convolutional neural networks for image classification instead of identifying image patterns from scratch. Although the original model may have been trained on completely different images to the current application, the main assumption is that, provided a sufficiently vast and diverse dataset, the learnt features are general enough to interpret the visual world. Of the many pre-trained convolutional neural network models available in the Keras library, six commonly used architectures were selected in this study to evaluate their performance in classifying the scalograms based on the type of synchronization regime (i.e., regime I, regime II, hyper-branch regime, regime III, or desynchronized regime). The six chosen architectures are the Xception (Chollet, 2017), InceptionV3 (Szegedy *et al.*, 2015), ResNet50 (He *et al.*, 2016a), ResNet50V2 (He *et al.*, 2016b), InceptionResNetV2 (Szegedy *et al.*, 2016), and VGG16 (Simonyan and Zisserman, 2014) models.

While all the models of interest in this study utilize convolutions as the main mathematical approach to interpret the multi-dimensional inputs, the way these low-level operations are implemented differs amongst the different architectures. The dissimilarity often arises in the number of layers (i.e., model depth); the number, size, and stride length of the convolutional kernels; the use of pointwise convolutions and separable convolutions; the inclusion of shortcut connections; as well as the frequency at which the pooling layer is applied. The unique combinations of the aforementioned factors mean that each architecture will approach the task of analyzing the input data differently, with varying levels of accuracy and computational efficiency. Therefore, this study aims to elucidate the model architecture that is best suited for classifying scalograms of FIV-related signals and also to identify the models that can represent the data such that it is “interpretable” by a human researcher (see Sec. III B).

The pre-trained models use parameters learnt through training on the ImageNet database, which was developed by Deng *et al.* (2009) and contains over 1.4×10^6 images spanning 1000 classes. Although each model had different input image sizes when trained on ImageNet, the size of the images in this study for all models was chosen to be 299×299 pixels², in line with the largest inputs for the models of interest (i.e., Inception, Xception, and InceptionResNetV2). As the images are in color with three channels corresponding to red, green, and blue intensity, each image has a total of $299 \times 299 \times 3$ inputs. Since the scalograms generated in Sec. II B have a size of $23 \times 30\,000$ pixels² due to the 100 Hz acquisition rate, they were first reconfigured by cutting the array into 13 sections and stacking them together to form a 299×2307 pixels² image. The result was then resized in the temporal direction to obtain the required size [Fig. 2(b)]. Although this operation results in an approximately $7.7 \times$ compression loss with

each pixel spanning 77 milliseconds in the horizontal direction, the resultant Nyquist rate is approximately $9f_{\text{inw}}$ with an overall resolution of approximately 18 points per cycle corresponding to an oscillation period of ~ 1.4 s. As such, any intermittent behavior will still be well captured since the highest harmonic of interest is $f^* = 3$. After feeding the scalogram images as inputs to the pre-trained convolutional neural network, the outputs of this feature extraction process are then stored and later used as inputs to the fully connected output layer for wake-mode classification (Sec. II C 2).

2. Fully connected output layer for classification

Using the outputs of the pre-trained convolutional neural network as discussed above, the five-node fully connected output layer shown in Fig. 2(d) is then used to perform the wake-mode classification. By assuming that the convolutional neural network has extracted all the salient frequency-time information from the scalogram, the samples corresponding to each FIV regime will occupy a unique subspace within the high-dimensional output space of this network. Therefore, the role of the final fully connected layer in the overall machine learning workflow is to identify the boundaries in the high-dimensional space that can separate the five response regimes with the “minimum” error. As such, the output of the convolutional neural network becomes the input for the fully connected layer, with the final output being the prediction of the FIV regime for a given input sample.

While similar in architecture to their convolutional counterpart, the fully connected layer is characterized by each node being “fully connected” to every node in the layers immediately preceding and succeeding it. Mathematically, this is represented by the dot product operation and defined as

$$\vec{O}^{(i+1)} = A \left(\vec{O}^{(i)} \cdot \vec{w} + \vec{b} \right), \quad (8)$$

where, instead of convolution with a kernel, a matrix of weights \vec{w} is used to define the relationships between nodes from layer to layer. Although the number of inputs corresponds to the number of outputs of the convolutional neural network, the fully connected layer has five outputs corresponding to each of the five regions of interest in this study. The activation for this layer is achieved using the softmax function, which is often used for multi-class and single-label classification tasks where the value of each output node corresponds to a possible response regime. This is mathematically defined as

$$A(x_i) = p_i = \frac{\exp x_i}{\sum_{j=1}^n \exp x_j}, \quad (9)$$

where the resultant values [equivalent to p_i in Eq. (5)] correspond to the probability that a given input belongs to the i th regime as predicted by the model. Here, \vec{x} denotes an array where each element represents one of the total $n = 5$ regimes. Therefore, the prediction of this feed-forward fully connected neural network classifier is the regime with the highest probability as measured by the fully connected softmax-activated final layer.

III. RESULTS

A. Wake mode and response regime classification

Before any discussion can begin, the effect that the skewed nature of the dataset has on the measures of model performance must first be

accounted for. As such, care must be taken to establish a baseline such that the true “statistical power” of any predictive model in this study can be ascertained. This reference is provided by a “naïve” classifier, where performance below the minimum baseline indicates that the model is not useful as it does not capture any relationships in the data. Since the metric of interest is the weighted accuracy, the naive classifier with no domain knowledge about the problem will randomly predict one of the five classes. As the weighting approach negates the effect of class imbalance on the measure of overall model performance, the baseline weighted accuracy used for comparison in this study is 20%.

For hyper-parameter optimization, Figs. 8(a)–8(f) show the weighted accuracy (with error bars denoting standard errors) for each architecture during training, estimated using fivefold cross-validation, as a function of epochs (the number of times the entire training dataset has passed through the ML algorithm). Except for ResNet50V2 and InceptionResNetV2, all the models tested in this study demonstrate a smooth convergence to a maximum weighted accuracy value within approximately 25 epochs. To ascertain the optimal number of epochs used for training, a variation of the one-standard-error rule suggested by Hastie *et al.* (2001) is utilized to determine the ideal epoch for each model. Here, the most parsimonious model whose accuracy is no more than one standard error below the accuracy of the best model (denoted by the open red circle) is chosen. Since parsimony is considered to be negatively correlated with the number of epochs, the epoch reported in Table I was chosen as the lowest number whose accuracy was within one standard error of the maximum accuracy for each given architecture.

After optimizing the weighted accuracy of each model, their performance is then compared to select the best-performing architecture for the problem presented in this study. To ensure that the differences in model performance are statistically significant, a “corrected” resampled student *t*-test, first proposed by Nadeau and Bengio (1999),

is conducted using 15 folds of the training data (obtained by repeating the fivefold cross-validation three times) with a significance level of 0.05. The adjustment of the *t*-statistic is necessary as the training and validation data overlap across each iteration of a single cross-validation experiment, thereby violating the assumption of independence needed for an unbiased estimate of variance required in a true test of statistical significance (Witten *et al.*, 2011). The corrected *t*-test used in this study is

$$t_c = \frac{\bar{d}}{\sqrt{\left(\frac{1}{k} + \frac{n_2}{n_1}\right)\sigma_d^2}}, \quad (10)$$

where the random variable *d*, the difference in weighted accuracy of the two models, has a mean of \bar{d} and a variance of σ_d^2 . Unlike the uncorrected *t*-statistic, *t_c* cannot increase boundlessly with the number of folds (*k*) and is instead also dependent on the number of samples used for training (*n*₁) and validation (*n*₂). The choice of 15 folds is in line with the suggestion made by Nadeau and Bengio (1999), who argued that the increase in the statistical power of the test by choosing beyond *k* = 15 folds is likely not worth the additional computational effort. As such, Table I shows the weighted accuracy and standard error for each model averaged over 15 folds.

Interestingly, the pre-trained Xception model had the highest weighted accuracy of 98.9 ± 0.2 , agreeing with the results presented by Copiaco *et al.* (2019) for audio classification with scalogram representations. However, when comparing the statistical significance of this result using the *t_c* statistic, only the difference in the weighted accuracy of the InceptionResNetV2 model was deemed statistically significant with a *p*-value of 0.06%. Relative to the best-performing Xception model, the null hypothesis could not be rejected for the other

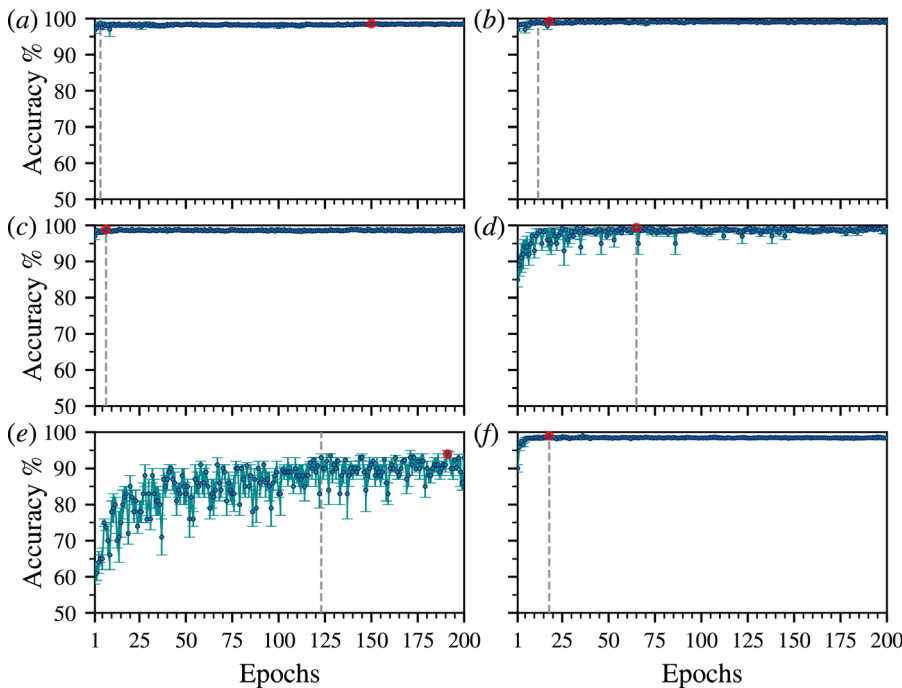


FIG. 8. Evolution of weighted accuracy as a function of epochs for the training data using (a) InceptionV3, (b) Xception, (c) ResNet50, (d) ResNet50V2, (e) InceptionResNetV2, and (f) VGG16 pre-trained convolutional neural networks. The plots are taken as the average of the accuracy curves obtained during each iteration of the fivefold cross-validation, with the error bars denoting the standard error. The red point shows the maximum accuracy achieved by the model, with the gray dotted lines and annotations denoting the epoch and accuracy of the idea model as determined using the “one-standard-error” rule.

TABLE I. Weighted accuracy and standard error (SE), evaluated using fivefold cross-validation repeated three times, of the pre-trained models after hyper-parameter optimization of the epoch number. Models were compared to the highest mean accuracy (Xception), with modified *t*-tests conducted to measure the statistical significance of this difference (Δ Accuracy) as indicated by the *p*-value. Training time indicates the duration of feature extraction and a single fivefold cross-validation used in hyper-parameter optimization.

	Epochs	Weighted accuracy (%)—pre-trained model	SE (%)	Δ Accuracy (%)	<i>p</i> -value (%)	Training time (h)
Xception	12	98.9	0.2	0.73
InceptionV3	4	98.4	0.2	0.5	41.85	0.58
ResNet50	7	98.6	0.2	0.3	41.11	0.80
ResNet50V2	65	98.8	0.5	0.1	91.02	0.76
InceptionResNetV2	123	90.9	0.8	8.0	0.06	0.46
VGG16	18	98.8	0.2	0.1	98.37	0.27

architectures. As this lack of definitive difference in the weighted accuracy of the top five machine learning models may be attributed to the data in the present study only containing samples across U^* and ζ , employing machine learning models to classify samples collected over a wider parameter space may yield a more significant performance difference. Nevertheless, the average weighted accuracy across the five best pre-trained models was 98.7%, approximately 4.9 times the naive classifier baseline accuracy of 20%. This indicates that the machine learning workflows proposed in this study can correctly identify the relationship between the wake mode and the structural response in FIV of an elliptical cylinder.

To demonstrate the validity of using the pre-trained convolutional neural networks over training one entirely from scratch (i.e., training both the convolutional neural network base and the classifier together), Figs. 9(a)–9(f) show the fivefold cross-validated accuracy for each randomly initialized architecture as a function of epochs using the same training data as before. Since both the convolutional base and

the classifier were initially untrained, they were first combined and then trained as a singular machine learning model. Except for ResNet50V2 and InceptionResNetV2, the models performed approximately the same as a naive classifier with an accuracy of $\sim 20\%$ at first, before an almost instantaneous jump in performance and a rapid convergence to the optimal weighted accuracy. Again, the one-standard-error rule was used to ascertain the optimal epoch number hyper-parameter, with the ideal epochs for each tested architecture reported in Table II. The corresponding accuracy and standard error for each optimized model were obtained using fivefold cross-validation repeated three times and are also presented in the table.

Although the convolutional bases have now learnt spatial features and patterns that are specific to the scalogram images, the differences in weighted accuracy are not statistically significant (hypothesis testing using the t_c statistic with a significance level of 0.05) when compared to the pre-trained convolutional neural networks. This suggests that the ImageNet database, from which the Xception model was trained, is

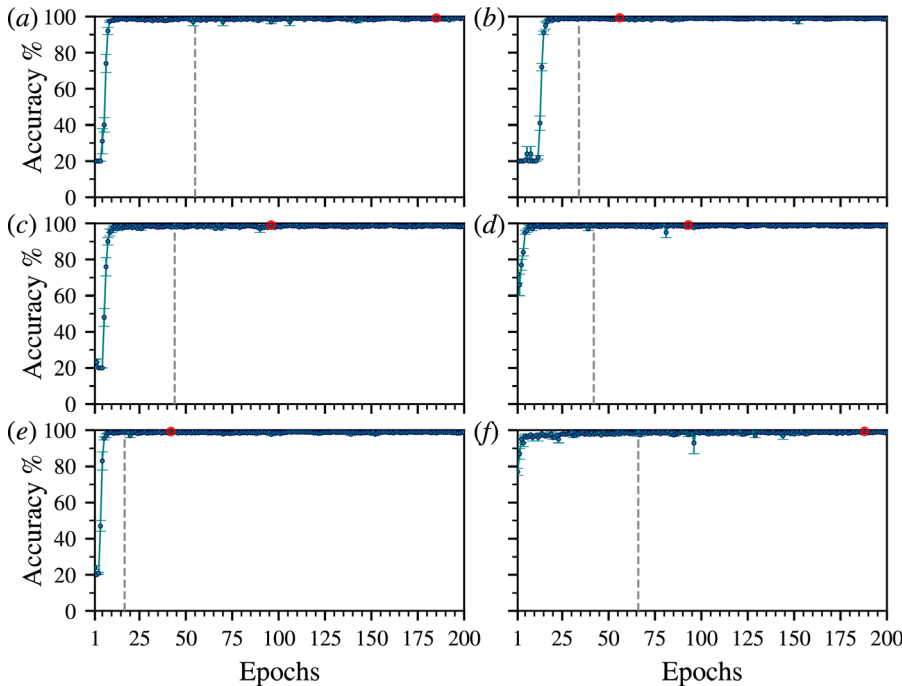


FIG. 9. Evolution of weighted accuracy as a function of epochs when training both the convolutional neural network bases and the fully connected classifiers from scratch. The bases are (a) InceptionV3, (b) Xception, (c) ResNet50, (d) ResNet50V2, (e) InceptionResNetV2, and (f) VGG16. More details can be found in the caption of Fig. 8.

TABLE II. Weighted accuracy and standard error (SE), evaluated using fivefold cross-validation repeated three times, of models trained from scratch after hyper-parameter optimization of the epoch number. Each architecture was compared to its pre-trained counterpart, with modified *t*-tests conducted to measure the statistical significance of this difference (Δ Accuracy) as indicated by the *p*-value. Training time indicates the duration of a single fivefold cross-validation used in hyper-parameter optimization.

	Epochs	Weighted accuracy (%) —trained from scratch	SE (%)	Δ Accuracy with respect to pre-trained model (%)	<i>p</i> -value (%)	Training time (h)
Xception	34	98.6	0.3	−0.3	74.21	9.95
InceptionV3	55	98.5	0.2	0.1	82.63	6.75
ResNet50	44	98.5	0.1	−0.1	83.06	8.61
ResNet50V2	42	98.5	0.3	−0.3	83.85	7.68
InceptionResNetV2	17	98.5	0.2	7.6	0.10	11.97
VGG16	66	98.3	0.3	−0.5	52.87	11.78

general enough to accurately extract the features of the scalogram images in this study. The only exception is the InceptionResNet50V2 model, where the increase in weighted accuracy of 7.6% was deemed statistically significant with a *p*-value of 0.1%. This result indicates that although the architecture can learn the features to interpret the scalogram images accurately, the weights the model learnt from ImageNet are not as transferable as those for the other models investigated in this study.

Furthermore, an advantage of just training the fully connected classifier over training a model from scratch is the reduction in computational complexity. Along with fewer parameters required to be optimized, the dot product operations used in the fully connected neural classifier also require fewer operations to compute than the operations in the convolutional layers of the convolutional neural networks. Hence, training is much less resource- and time-consuming to complete. On our computational setup (Intel Core i7 5930K CPU and 64 GB of RAM, with machine learning operations conducted on a 12 GB VRAM ASUS GeForce RTX 3060 GPU), training both the convolutional neural network base and the classifier (Table II) was on average 19.3 times as long as the time required to perform feature extraction on the scalogram images using the pre-trained model and to complete a single fivefold cross-validation of the classifier alone (Table I).

After ascertaining the robustness of the proposed machine learning workflow (leveraging pre-trained convolutional networks for wake-mode classification) for the classification of wake modes associated with each identified FIV regime, we found that the performance of InceptionV3, ResNet50, ResNet50V2, and VGG16 models was not statistically significant from the highest mean weighted accuracy achieved by the Xception model. Since the performance of the five models could not be differentiated, we propose a new approach where the models are combined to form an *ensemble*. By obtaining the outputs of each model, an ensemble method will make a prediction that gets the most votes. Assuming that each model makes different types of errors, aggregating the predictions of all the individual models can generally reduce the bias and variance of the predictor (Geron, 2019). As such, the optimal epoch hyper-parameter identified in Fig. 8 for each selected model was then utilized to train the feed-forward neural network classifiers on the entire training set. After aggregating all five models, the final unbiased performance of this ensemble approach was ascertained by evaluating the test dataset. Both the overall unweighted and weighted accuracy were 99.1%, with a breakdown of the number

of samples correctly and incorrectly predicted for the five classes shown by the confusion matrix in Fig. 10.

Out of the 1302-point test dataset, only 12 samples were misclassified by the machine learning workflow outlined in Fig. 2. Furthermore, when overlaying these points on top of the amplitude contour map (Fig. 5) as shown by Fig. 11, the regime predicted by the ensemble machine learning method for each test sample is almost identical to that identified by Lo *et al.* (2023). This indicates that, despite never encountering the test dataset, our proposed deep-learning approach can reproduce the boundaries of the FIV flow regimes in the $U^* - \zeta$ parameter space with reasonable fidelity. Therefore, it can be concluded that the ensemble can learn the key features in the scalogram images that allow the different wake-mode regimes to be identified for previously unseen data.

From Fig. 11, the observed misclassifications in Fig. 10 predominantly arise at or near the boundaries between the FIV regimes.

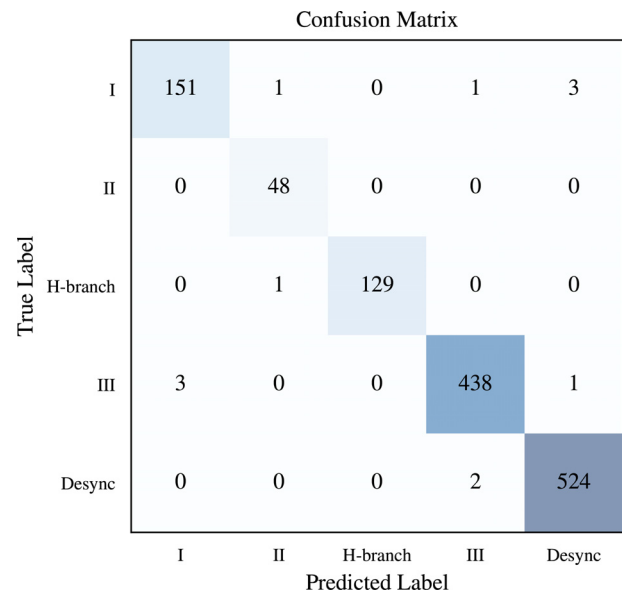


FIG. 10. Breakdown of the number of samples in the test dataset that belong to each wake-mode category. The vertical axis denotes the “true” label of the data point (as identified by the author in Fig. 5), while the horizontal axis is the category predicted by the machine learning workflow.

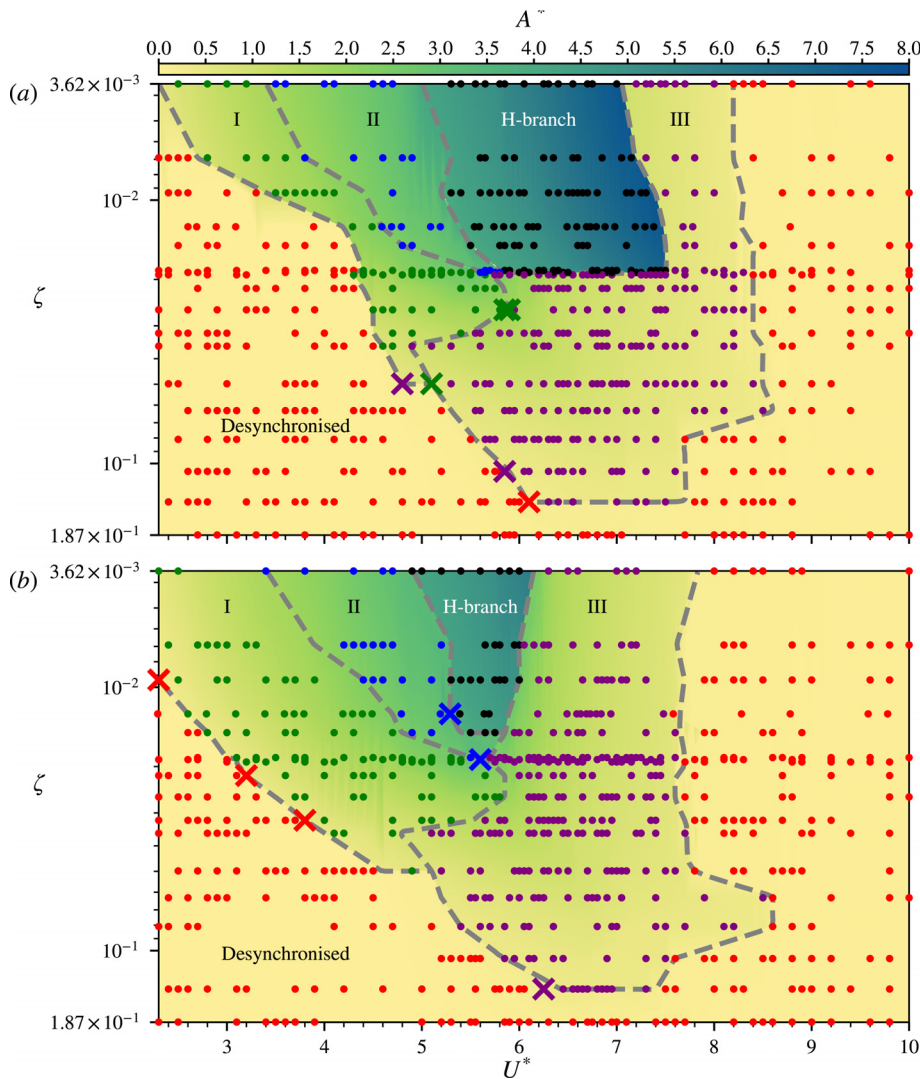


FIG. 11. The samples of the test dataset, represented with markers, overlaid on the two-dimensional contours presented in Fig. 5. (a) corresponds to U^* increasing, and (b) to U^* decreasing. Circular and crossed markers denoted correct and incorrectly classified points, respectively, with the color of each marker representing the regime predicted by the ML workflow: green—regime I, blue—regime II, black—hyper-branch regime, purple—regime III, and red—desynchronized. The gray dashed lines are the boundaries of the FIV regimes as identified by Lo *et al.* (2023).

This source of error is likely due to the body vibration and fluid forcing dynamics being very similar at the transitions between the FIV regimes, thereby making the classification difficult. A possible extension to this study is to ascertain if including additional dynamical information (such as phase and drag) as inputs to the machine learning model would improve the accuracy at the regime boundaries. It may also be interesting to investigate the combinations of time traces that would yield the highest accuracy.

In addition, the robustness of the proposed ensemble method to the dataset size can be estimated by plotting the weighted accuracy (when evaluated on the test set) as a function of the percentage of the total data used for training. From Fig. 12, the general trend observed is that increases in training samples result in a rise in the accuracy, with a maximum improvement of approximately 10% in accuracy corresponding to an increase from 5% to 15% of the total data used for training. This is likely due to the unbalanced nature of the dataset, with 5% of the total data approximately equivalent to six samples in regime II. Therefore, the rapid increase in accuracy in the low data

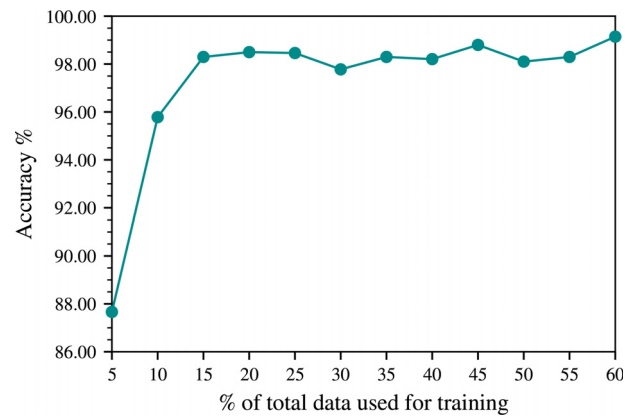


FIG. 12. The effect of the quantity of training data on the weighted accuracy of the proposed ensemble machine learning workflow.

limit could be caused by a lack of training data for under-represented regimes, which indicates that the difference in performance may be less for more balanced datasets. The accuracy then plateaus when more than 15% of the total dataset is used for training. This observed behavior arises as the data are only being used to train the fully connected classifier that relates the outputs of the pre-trained convolutional neural network to the wake-mode classes established by [Lo et al. \(2023\)](#). Since the spatial patterns used to interpret the scalogram images are already known to the convolutional neural network, fewer samples are required during training as fewer model parameters must be learnt. Therefore, this indicates that our proposed machine learning workflow is generally robust to the size of the dataset and produces accurate results even in the low sample limit.

To summarize, the proposed machine learning method of using a pre-trained convolutional neural network to extract features in the scalogram images and using the result as inputs to a classifier was shown to be effective in classifying the FIV regimes identified by [Lo et al. \(2023\)](#). While training the models from scratch takes an average of 19.3 times that of utilizing a pre-trained convolutional neural network base and only training the classifier, it was also shown that, except for InceptionResNetV2, there was no statistically significant improvement in performance. It can, hence, be concluded that pre-trained networks are an effective and time-efficient approach (even in the low-data limit) for wake-mode classification without sacrificing accuracy.

B. Interpretability of the proposed data-driven approach for wake-mode analysis

The approximately 99% accuracy achieved by leveraging the pre-trained convolutional neural networks for feature extraction indicates that the proposed machine learning method can differentiate between the lock-in regimes occurring in the FIV response of the elliptical cylinder. Hence, the following question arises: in addition to applications in the wake-mode classification, can the outputs of the convolutional neural network also be utilized to visualize the dynamics behind the FIV of the elliptical cylinder? More simply, are the transformations of the data by the proposed data-driven approach interpretable?

To answer this question, the processes outlined in [Figs. 2\(a\)–2\(c\)](#) were again employed to project the entire dataset into high-dimensional space. However, instead of feeding the result of the transformations by each convolutional neural network into a classifier as depicted in [Fig. 2\(d\)](#), the high dimensional outputs (of length 41 472–204 800, depending on the model) were instead mapped into a three-dimensional representation using principal component analysis. PCA, which is often referred to as proper orthogonal decomposition (POD) in many fluid dynamics studies, is a type of linear dimensionality reduction technique. Transforming the data onto a new coordinate system, PCA identifies the important features by expressing the high-dimensional samples using only a few parameters that maximally explain the variations in the data.

[Figures 13–18](#) plot the resultant outputs as a function of the first three PCA modes along with the percentage of the total explained variance in the data. A plot of the second PCA mode against the third mode was omitted from the figure as combined components did not capture enough variance to yield any clear and meaningful separation of the data.

Except for the ResNet50 model, all of the CNNs tested in this study show the desynchronized region and the hyper-branch regime on the left- and right-hand sides of the PCA plots, respectively. The desynchronized response is characterized by negligible amplitude and no lock-in or harmonics, while the hyper-branch regime is identified by the large amplitudes and lock-in with significant harmonic components ($f_C^* \approx 2$ and 3). Interestingly, the visualization of the ResNet50 outputs has a completely different pattern from the plots of the other models, with each regime continuously connected to the other. Although this structure does not appear to be very interpretable, the lack of overlap between the different regimes explains the high predictive accuracy of the ResNet50 model.

Of the remaining CNNs, the InceptionResNetV2 PCA plot in [Fig. 13](#) appears to lack an overall structure and has the least separation of the regimes as shown by the significant overlap between regimes I and II and between regimes I and III. This, along with the low fivefold cross-validated weighted accuracy estimate when compared to the other tested models, indicates that the features

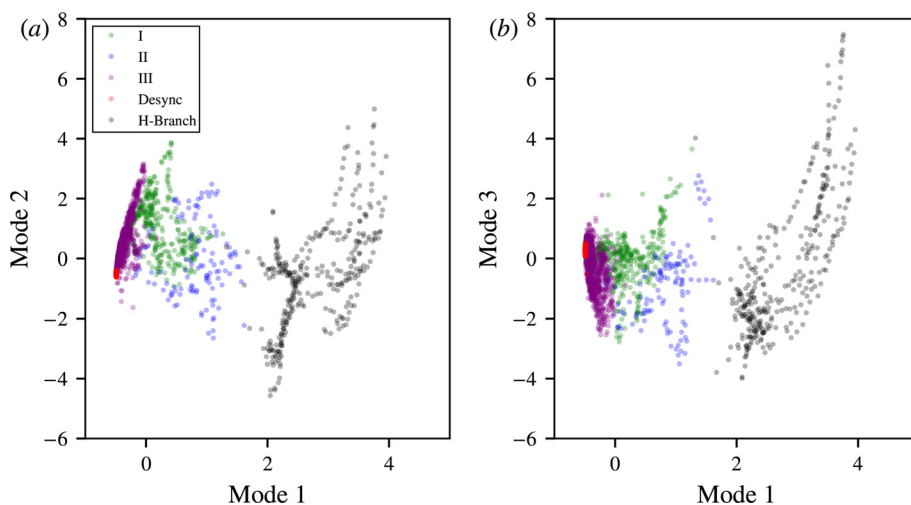


FIG. 13. Plots of (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained InceptionResNetV2 convolutional neural network outputs. The three modes explain 99.9% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

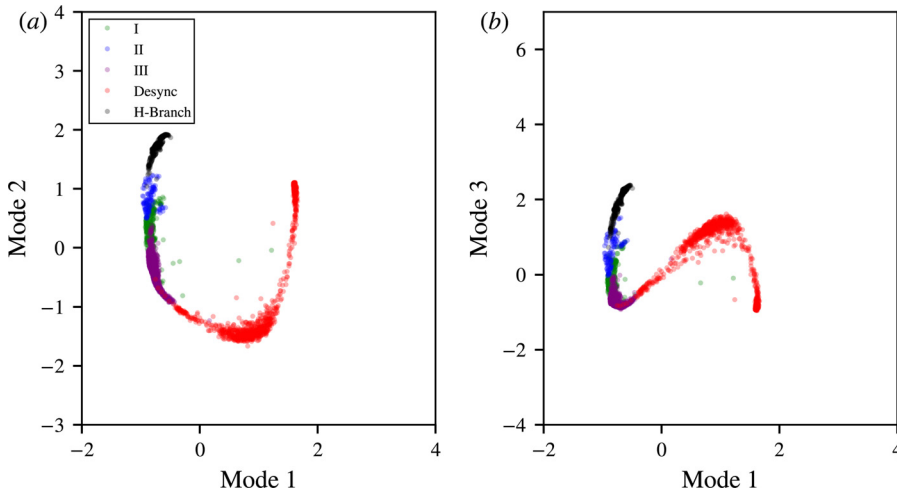


FIG. 14. Plot of the (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained ResNet50 convolutional neural network outputs. The three modes explain 78.1% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

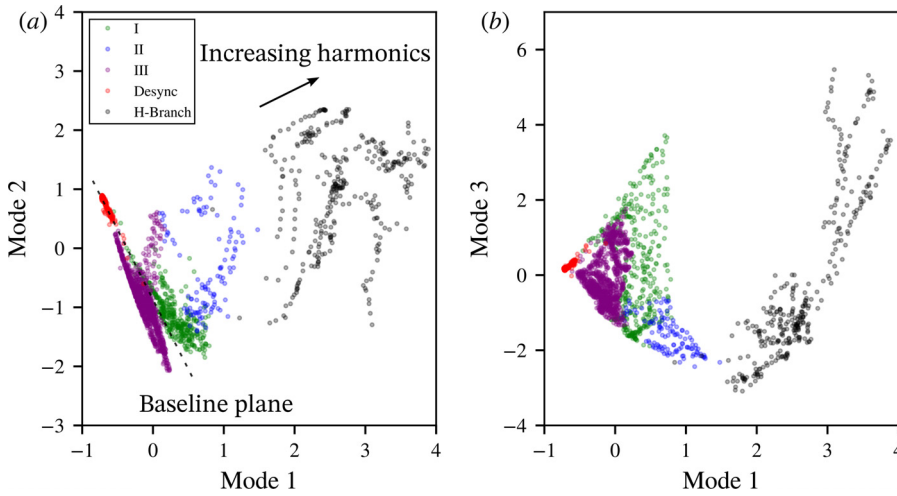


FIG. 15. Plot of the (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained Xception convolutional neural network outputs. The three modes explain 85.9% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

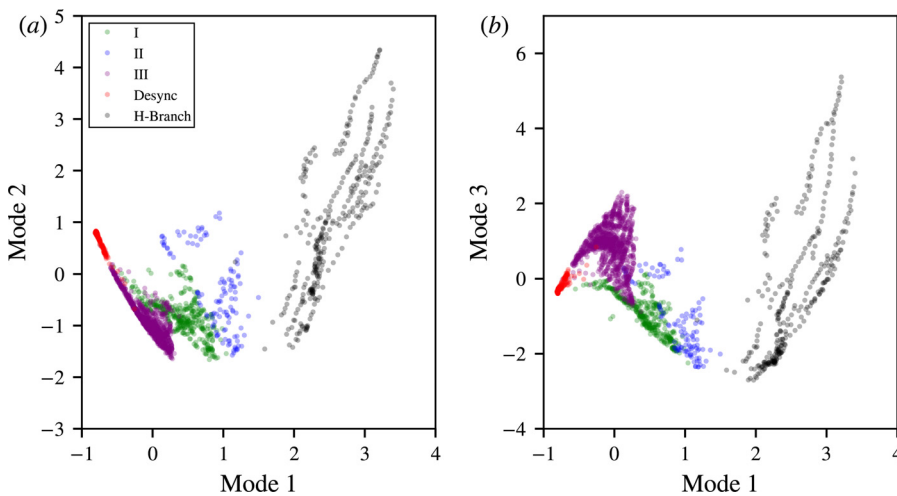


FIG. 16. Plot of the (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained ResNet50V2 convolutional neural network outputs. The three modes explain 97.9% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

09 July 2024 06:31:45

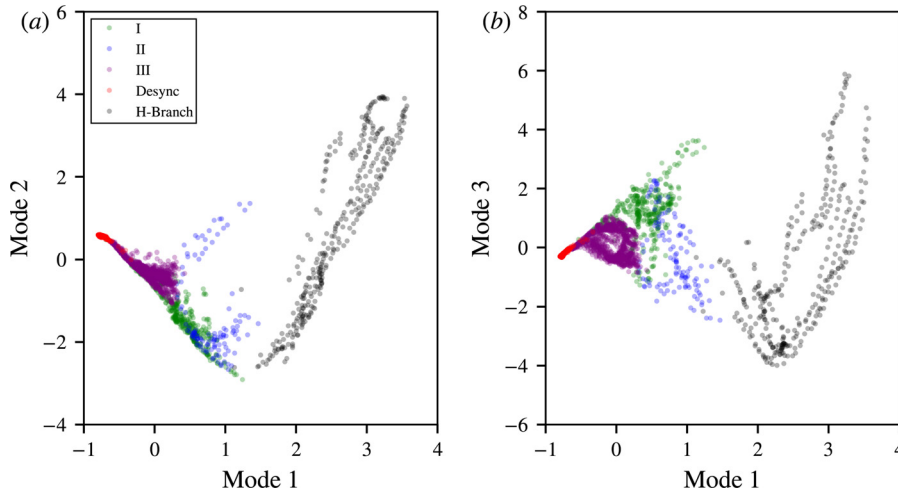


FIG. 17. Plot of the (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained InceptionV3 convolutional neural network outputs. The three modes explain 84.4% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

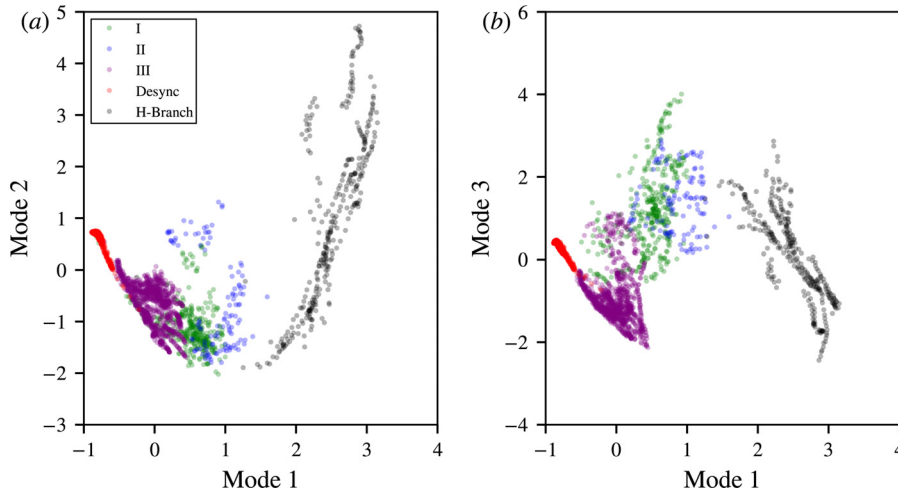


FIG. 18. Plot of the (a) the first and second modes and (b) the first and third modes obtained after conducting principal component analysis (PCA) on the pre-trained VGG16 convolutional neural network outputs. The three modes explain 94.4% of the total variance, with the axes showing the standard score after normalizing the data along each mode.

learnt by InceptionResNetV2 on the ImageNet database may not be insightful in interpreting the frequency-time representation of the FIV dynamics.

Interestingly, a similar pattern or “structure” can be observed across multiple CNNs as demonstrated in Figs. 15–18(a). Here, samples from the desynchronized regime occupy a linear plane (which will be coined the “baseline” in further discussions), and the perpendicular distance away from this plane indicates how different the response is from the desynchronized dynamics. However, although both Xception and ResNet50V2 show the greatest separation and structure of the wake mode regimes, only the PCA plot of the former [Fig. 15(a)] was able to differentiate between the desynchronization regime, as well as lock-in responses with and without harmonic components in the fluid forcing. This, along with the highest mean fivefold weighted accuracy, motivates a more detailed analysis of the Xception model outputs, which will be discussed below.

As previously mentioned, all four lock-in regimes identified, along with the region of desynchronization, occupy distinct areas in Fig. 15. Since these regimes were identified through the time traces of

the body displacement, lift force, and vortex force, the above-mentioned observation suggests that the similarity in the FIV responses between samples is inversely correlated with their distance in the PCA plot. As such, the samples pertaining to the desynchronized region appear concentrated in the PCA plots, since the vibrations associated with this regime are consistently chaotic and negligible in amplitude. On the other end of the spectrum, the wide amplitudes and fluid forces acting on the cylinder mean that the energetic hyper-branch regime appears as a low density “cloud.” From Fig. 15(b), it can be seen that the samples can be split into two distinct groups where the hyper-branch regime occupies the right-hand side of the plot while the group of samples on the left is a conglomerate of the remaining flow regimes. This indicates that the physics driving the hyper-branch regime is distinct from the other regions and agrees well with the hypothesis by Lo et al. (2023) that this wake mode is a combined product of both VIV and the movement-induced instability (i.e., transverse galloping).

Interestingly, the patterns captured by the PCA plot in Fig. 15(a) become meaningful when an appropriate origin is established. Since the desynchronized region has no discernible wake patterns or

vibration amplitude, a “baseline plane” was established by drawing a line through the desynchronized region that represents a baseline to which each lock-in regime can be compared. First, both regimes I and III form well-defined planes that appear predominantly as linear lines on either side of the baseline. Their proximity to each other may be due to the similarities in frequency responses of both regimes where, as a consequence of wake-body synchronization, vortex shedding occurs with a predominantly singular frequency that matches the body vibration, i.e., $f_{C_y}^* \sim f_y^*$. As the aforementioned lines are not congruent, this suggests that the difference between the two planes may arise from variations in the lock-in frequency. For regime I, the body vibration frequency is invariant to U^* and remains at $f_y^* \approx 1$, while the frequency for regime III is $f_y^* \geq 1$, where the deviation away from the natural frequency in quiescent water increases with U^* . As regime III is the only lock-in regime where the f^* value deviates away from f_{nw} , it explains why it is the sole lock-in regime principally on the left of the baseline plane. While there are also samples belonging to this regime on the right-hand side of the baseline, these points occur near the transition to regime I or the hyper-branch regime, where the main lock-in frequency is $f^* \approx f_{nw}$, making them indistinguishable from a frequency perspective.

Additionally, the advantage of establishing a baseline becomes apparent when considering regime II and the hyper-branch regime, which both have strong second and third harmonic contributions in their fluid force frequency responses (i.e., C_y, C_v). By considering the orthogonal distance of the regimes on the right of the baseline plane, the magnitude of this deviation can hence be directly related to the relative strength of the harmonics. Since the strongest harmonic contributions are found in the hyper-branch regime, this regime is the furthest away from the baseline plane drawn on the lower left of Fig. 15(a). On the opposite end of the spectrum, for samples belonging to regime III on the right of the baseline, their proximity to the plane corresponds to weak harmonics in the driving forces observed intermittently in the regime.

Beyond the large-scale structures that can be seen by viewing each regime as an independent entity, the PCA plot also provides a visual representation of how structural damping affects the transition between each regime over the U^* range of interest. Two sets of

damping values were chosen from the contour plot (obtained with decreasing U^* increments) in Fig. 5(b) to illustrate the cases corresponding to the presence and suppression of the hyper-branch regime.

In the case of the former, Fig. 19 shows the evolution of the lock-in and desynchronized regimes with increasing damping ratios, when the hyper-branch regime is present. For $\zeta = 3.62 \times 10^{-2}$, the evolution of the lock-in regimes (denoted by the gray dotted arrows), except for the transition between regimes I and II, appears as distinct jumps in the PCA plot. Though increasing ζ generally decreases this separation between the regimes, the shift between regimes I and II remains almost continuous, with regime II manifesting in the PCA plot as a continuous deviation in the orthogonal direction to the baseline plane. This, in addition to the lack of a direct transition between regime I and the hyper-branch regime, indicates that regime II can be viewed as an extension of regime I that acts as an intermediary between the two regimes. The above-mentioned interpretation agrees well with the behavior observed in Lo et al. (2023), where regime II is characterized in the transverse fluid-force frequency response by a strengthening of the third harmonic with increasing U^* .

The behavior described above, where increasing the applied structural damping decreases the separation between the regimes in the PCA plot, is not only unique to responses where the hyper-branch regime is present. As shown in Fig. 20, the same trend is also observed for FIV responses where the hyper-branch regime has been suppressed. For this case, as the applied structural damping increases, regimes I and II form a loop-like structure which then collapses upon itself, resulting in only regime III remaining as the sole-surviving lock-in regime. With further increases in damping, regime III connects with the desynchronized region to form a continuous line for $\zeta = 6.30 \times 10^{-2}$, eventually leading to a PCA plot where all lock-in regimes have been completely suppressed and only the desynchronized region remains.

In summary, of all the models tested in this study, the PCA plots in Fig. 15 demonstrate that outputs of the Xception convolutional neural network are the most interpretable and provide a visual representation of the relationship between the lock-in regimes as well as the desynchronized region. The distance between samples is inversely correlated with the similarity in their FIV response, with each lock-in

09 July 2024 06:31:45

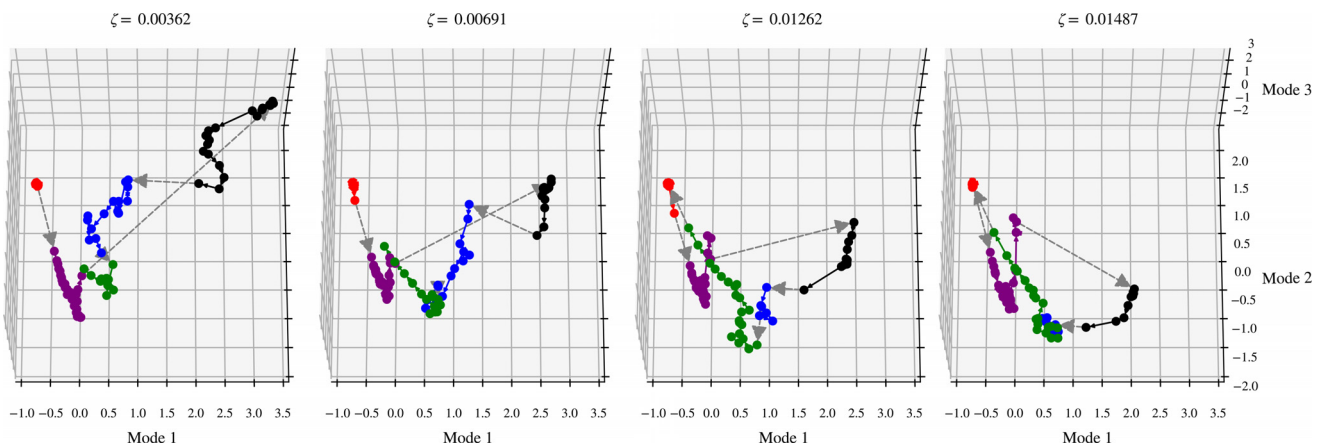


FIG. 19. PCA plots depicting the evolution of the amplitude response (with the hyper-branch regime present) as a function of reduced velocity for different structural damping ratios. The arrows show the direction of decreasing U^* , with the dotted arrows denoting the transition between the regimes. More details can be found in the legend of Fig. 15.

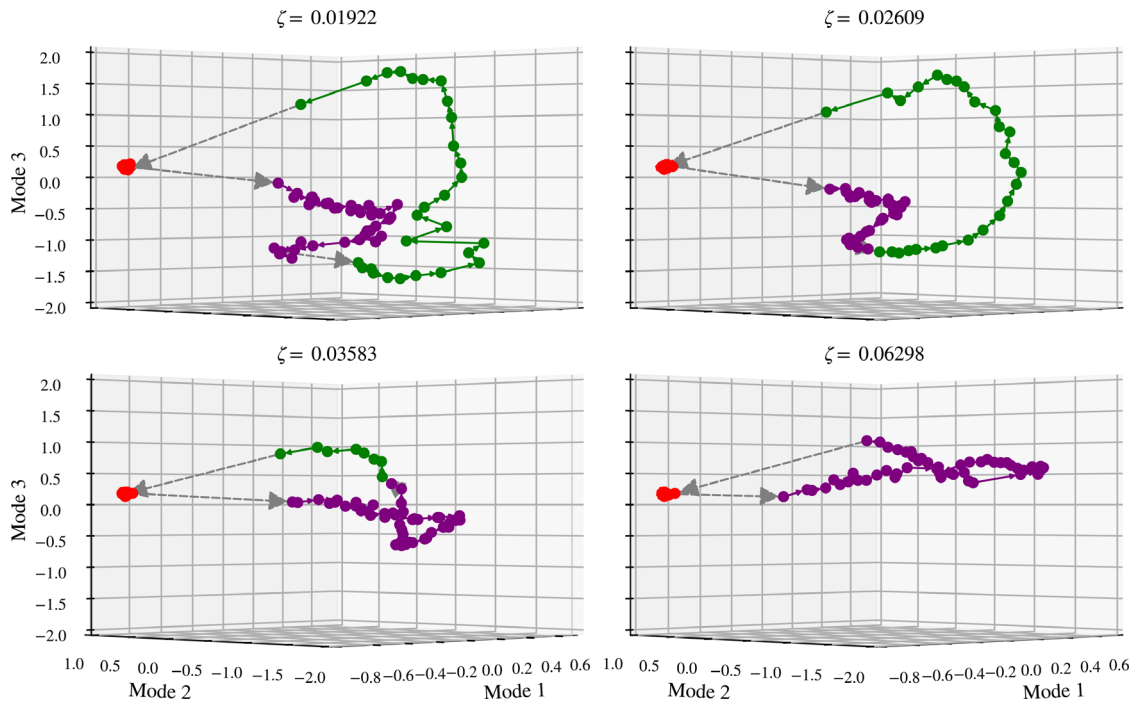


FIG. 20. PCA plots depicting the evolution of the amplitude response (with hyper-branch suppression) as a function of reduced for different structural damping ratios. The arrows show the direction of decreasing U^* , with the dotted arrows denoting the transition between the regimes. More details can be found in the caption of Fig. 15.

regime and desynchronized region appearing as distinct regions in the plot. As the hyper-branch regime appears isolated from the other flow regimes in Fig. 15(b), it suggests that the physics that drives the FIV dynamics in this regime is unique and concurs with previous assertions that the hyper-branch regime arises from manifestations of both VIV and galloping. The presence of higher order harmonics in the transverse lift force can also be observed and manifests as the orthogonal deviation of regime II and the hyper-branch regime from the linear baseline plane (whose lift force response is dominated by the natural frequency) identified in Fig. 15(a). However, we note that the patterns and trends observed in the above-mentioned PCA plots only apply to the pre-trained Xception convolutional neural network. While the other models available in Keras were also trained on the ImageNet database, the architectural differences will affect the learnt spatial features, thereby changing the model outputs that are visualized using the PCA plots. Therefore, further work will be required to understand how different models interpret the time-frequency information provided by the scalogram images.

IV. CONCLUSIONS

In conclusion, the flow-induced vibration arising from the coupling between the fluid and the elliptical cylinder is complex in nature and is dependent on multiple parameters of the structural and fluid properties. With current manual approaches to traverse this parameter space being intractable over so many variables, we propose a machine learning workflow that leverages previously collected data to both automate the wake-mode classification process prevalent in FIV studies and also to serve as an approach to visualize the relationship

between the identified flow regimes. Employing continuous wavelet transforms to obtain the time-frequency representation of the structural response of the elliptical cylinder, the image-like scalograms were then used as inputs to the pre-trained convolutional neural networks for feature extraction. The resultant output was then used to train a fully connected classifier output layer to categorize the five regimes identified by Lo et al. (2023). After forming an ensemble classifier using the five best-performing models, this approach was found to have a weighted accuracy of 99.1% on the withheld test dataset.

Furthermore, we have also shown that the visualizations of the flow regimes can be obtained by projecting the outputs of the pre-trained CNN models onto the first three principal component analysis modes. The resultant PCA plot from the pre-trained Xception model provided the most useful visual representation of how structural damping affects the transitions between the lock-in regimes as well as the desynchronized region. Specifically, it was observed that all the flow regimes occupied well-defined areas in the plot, with this distinction becoming less clear with increasing structural damping. The presence of harmonics in the driving fluid forces (i.e., C_y and C_v) was also captured in the plot, which was manifested as the orthogonal deviation to the right of the baseline plane (defined as the line drawn across through the desynchronized region).

While the amplitude and frequency information of the body displacement and driving forces (i.e., total transverse fluid force and vortex force) allow the machine learning workflow to characterize the FIV dynamics, the future work to improve the performance may include adding more channels to the scalogram images such that the drag forces and phase information (i.e., the lag between the fluid forces and the

body displacement) are also captured. Though this has utilized information about the body displacement and the transverse driving forces to great effect, the machine learning approach may benefit from a richer input space that captures more of the dynamics behind the FIV system.

Although this study is our initial implementation of data-driven techniques to characterize different FIV phenomena, we have successfully demonstrated that the proposed machine learning workflow is well suited both as a way to automate wake mode classification and as an approach to visualize the relationship between the flow regimes. While the FIV dynamics of the elliptical cylinder investigated in this study are considered predominantly stable, this data-driven approach can also be applied to systems where the competition between two different wake modes leads to intermittent switching of vibrational frequencies. As such, our proposed machine learning methodology is flexible and can be readily applied to other FIV systems, such as circular cylinders (Govardhan and Williamson, 2000) and square cylinder (Zhao *et al.*, 2018), with differing structural properties and flow conditions. Since data are constantly being generated from experimental and computational facilities globally, leveraging machine learning techniques that can harness this constant flow of information will aid in developing a more holistic understanding of the complex physics and mechanisms of fluid–structure interaction.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council Discovery Early Career Researcher Award (Grant No. DE200101650) and the Australian Research Council Discovery Project (Grant No. DP210100990).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jonathan C. C. Lo: Conceptualization (equal); Data curation (lead); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (lead). **Mark Christopher Thompson:** Conceptualization (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal). **Kerry Hourigan:** Conceptualization (equal); Funding acquisition (equal); Supervision (equal); Writing – review & editing (equal). **Jisheng Zhao:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- Abadi, M. *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems” (2015), see <http://www.tensorflow.org>.
 Abbasi, B. and Goldenholz, D. M., “Machine learning applications in epilepsy,” *Epilepsia* **60**, 2037–2047 (2019).

- Bearman, P. W., “Vortex shedding from oscillating bluff bodies,” *Annu. Rev. Fluid Mech.* **16**, 195–222 (1984).
 Blevins, R., *Flow-Induced Vibration*, 2nd ed. (Van Nostrand Reinhold Inc., New York, NY (USA), 1990).
 Bradbury, J., Merity, S., Xiong, C., and Socher, R., “Quasi-recurrent neural networks,” [arXiv:1611.01576](https://arxiv.org/abs/1611.01576) (2016).
 Brooks, P. N. H., “Experimental investigation of the aeroelastic instability of bluff two-dimensional cylinders,” Ph.D. thesis, University of British Columbia, 1960.
 Brunton, S. L., Noack, B. R., and Koumoutsakos, P., “Machine learning for fluid mechanics,” *Annu. Rev. Fluid Mech.* **52**, 477–508 (2020).
 Celin, S. and Vasanth, K., “ECG signal classification using various machine learning techniques,” *J. Med. Syst.* **42**, 241 (2018).
 Chollet, F., “Xception: Deep learning with depthwise separable convolutions,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017), pp. 1800–1807.
 Chollet, F., *Deep Learning with Python* (Manning Publications, New York, NY, 2022).
 Chollet, F. *et al.*, “Keras,” see <https://keras.io> (2015).
 Copiaco, A., Ritz, C., Fasciani, S., and Abdulaziz, N., “Scalogram neural network activations with machine learning for domestic multi-channel audio classification,” in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (IEEE, 2019), pp. 1–6.
 Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., “Imagenet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 248–255.
 Fernandes, R., Huang, L., and Vejarano, G., “Non-Audible Speech Classification Using Deep Learning Approaches,” in *International Conference on Computational Science and Computational Intelligence (CSCI)* (IEEE, 2019), pp. 630–634.
 Geron, A., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. (O’Reilly Media, Sebastopol, CA, 2019).
 Govardhan, R. and Williamson, C. H. K., “Modes of vortex formation and frequency response of a freely vibrating cylinder,” *J. Fluid Mech.* **420**, 85–130 (2000).
 Hall, P., “On the stability of the unsteady boundary layer on a cylinder oscillating transversely in a viscous fluid,” *J. Fluid Mech.* **146**, 347–367 (1984).
 Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning*, Springer Series in Statistics (Springer, New York, 2001).
 He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016a), pp. 770–778.
 He, K., Zhang, X., Ren, S., and Sun, J., “Identity mappings in deep residual networks,” [arXiv:1603.05027](https://arxiv.org/abs/1603.05027) [cs.CV] (2016b).
 Hinton, G., Srivastava, N., and Swersky, K., “Overview of mini-batch gradient descent,” in *Neural Networks for Machine Learning* (Coursera, 2012), Vol. 575, pp. 1–31, see <https://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf>.
 Hong, S., Zhou, Y., Shang, J., Xiao, C., and Sun, J., “Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review,” *Comput. Biol. Med.* **122**, 103801 (2020).
 Janke, M. and Diener, L., “EMG-to-speech: Direct generation of speech from facial electromyographic signals,” *IEEE/ACM Trans. Audio Speech Lang. Process.* **25**, 2375–2385 (2017).
 Khalak, A. and Williamson, C., “Dynamics of a hydroelastic cylinder with very low mass and damping,” *J. Fluids Struct.* **10**, 455–472 (1996).
 Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., and O’Leary, A., “PyWavelets: A Python package for wavelet analysis,” *J. Open Source Software* **4**, 1237 (2019).
 Lin, P., Hu, G., Li, C., Li, L., Xiao, Y., Tse, K., and Kwok, K., “Machine learning-based prediction of crosswind vibrations of rectangular cylinders,” *J. Wind Eng. Ind. Aerodyn.* **211**, 104549 (2021).
 Lo, J. C. C., Hourigan, K., Thompson, M. C., and Zhao, J., “The effect of structural damping on flow-induced vibration of a thin elliptical cylinder,” *J. Fluid Mech.* **974**, A5 (2023).
 Lv, Y., Sun, L., Bernitsas, M. M., and Sun, H., “A comprehensive review of nonlinear oscillators in hydrokinetic energy harnessing using flow-induced vibrations,” *Renewable Sustainable Energy Rev.* **150**, 111388 (2021).
 Menon, K. and Mittal, R., “On the initiation and sustenance of flow-induced vibration of cylinders: Insights from force partitioning,” *J. Fluid Mech.* **907**, A37 (2020).

- Morse, T. L. and Williamson, C. H. K., "Prediction of vortex-induced vibration response by employing controlled motion," *J. Fluid Mech.* **634**, 5 (2009).
- Nadeau, C. and Bengio, Y., "Inference for the generalization error," in *Advances in Neural Information Processing Systems* (MIT Press, 1999), Vol. 12.
- Narin, A., "Detection of focal and non-focal epileptic seizure using continuous wavelet transform-based scalogram images and pre-trained deep neural networks," *IRBM* **43**, 22–31 (2020).
- Navrose, Yogeswaran, V., Sen, S., and Mittal, S., "Free vibrations of an elliptic cylinder at low reynolds numbers," *J. Fluids Struct.* **51**, 55–67 (2014).
- Nemes, A., Zhao, J., Jacono, D. L., and Sheridan, J., "The interaction between flow-induced vibration mechanisms of a square cylinder with varying angles of attack," *J. Fluid Mech.* **710**, 102–130 (2012).
- Pham, T. D., "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci. Rep.* **11**, 1–11 (2021).
- Priscilla, C. and Prabha, D. P., "Credit card fraud detection: A systematic review," in *International Conference on Information, Communication and Computing Technology* (Springer, 2019), pp. 290–303.
- Salloum, R. and Kuo, C.-C. J., "ECG-based biometrics using recurrent neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2017), pp. 2062–2066.
- Shi, Z., Gao, C., Dou, Z., and Zhang, W., "Flow-induced vibration modeling of bluff bodies with data assimilation," *J. Fluids Struct.* **118**, 103866 (2023).
- Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).
- Soti, A. K., Zhao, J., Thompson, M. C., Sheridan, J., and Bhardwaj, R., "Damping effects on vortex-induced vibration of a circular cylinder and implications for power extraction," *J. Fluids Struct.* **81**, 289–308 (2018).
- Sumer, B. M. *et al.*, *Hydrodynamics around Cylindrical Structures*, Advanced Series on Ocean Engineering Vol. 26 (World Scientific, 2006).
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A., "Inception-v4, inception-ResNet and the impact of residual connections on learning," [arXiv:1602.07261](https://arxiv.org/abs/1602.07261) [cs.CV] (2016).
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," [arXiv:1512.00567](https://arxiv.org/abs/1512.00567) [cs.CV] (2015).
- Tary, J. B., Herrera, R. H., and van der Baan, M., "Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms," *Philos. Trans. R. Soc. A* **376**, 20170254 (2018).
- Teti, R., Jemielniak, K., O'Donnell, G., and Dornfeld, D., "Advanced monitoring of machining operations," *CIRP Ann.* **59**, 717–739 (2010).
- Tran, M.-Q., Liu, M.-K., and Tran, Q.-V., "Milling chatter detection using scalogram and deep convolutional neural network," *Int. J. Adv. Manuf. Technol.* **107**, 1505–1516 (2020).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., "Attention is all you need," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30.
- Vijay, K., Srinil, N., Zhu, H., Bao, Y., Zhou, D., and Han, Z., "Flow-induced transverse vibration of an elliptical cylinder with different aspect ratios," *Ocean Eng.* **214**, 107831 (2020).
- Wang, Z., Du, L., Zhao, J., and Sun, X., "Structural response and energy extraction of a fully passive flapping foil," *J. Fluids Struct.* **72**, 96–113 (2017).
- Weytjens, H. and Weerd, J. D., "Process outcome prediction: CNN vs LSTM (with attention)," in *International Conference on Business Process Management* (Springer, 2020), pp. 321–333.
- Williamson, C. and Govardhan, R., "Vortex-induced vibrations," *Annu. Rev. Fluid Mech.* **36**, 413–455 (2004).
- Witten, I. H., Frank, E., and Hall, M. A., *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. (Morgan Kaufmann, 2011).
- Wong, K. W. L., Zhao, J., Jacono, D. L., Thompson, M. C., and Sheridan, J., "Experimental investigation of flow-induced vibration of a rotating circular cylinder," *J. Fluid Mech.* **829**, 486–511 (2017).
- Zhao, J., Hourigan, K., and Thompson, M., "Flow-induced vibration of D-section cylinders: An afterbody is not essential for vortex-induced vibration," *J. Fluid Mech.* **851**, 317–343 (2018).
- Zhao, J., Hourigan, K., and Thompson, M. C., "Dynamic response of elliptical cylinders undergoing transverse flow-induced vibration," *J. Fluids Struct.* **89**, 123–131 (2019).
- Zhao, J., Leontini, J. S., Lo Jacono, D. L., and Sheridan, J., "Chaotic vortex induced vibrations," *Phys. Fluids* **26**, 121702 (2014a).
- Zhao, J., Leontini, J. S., Lo Jacono, D., and Sheridan, J., "Fluid–structure interaction of a square cylinder at different angles of attack," *J. Fluid Mech.* **747**, 688–721 (2014b).
- Zhao, J., Nemes, A., Lo Jacono, D., and Sheridan, J., "Branch/mode competition in the flow-induced vibration of a square cylinder," *Philos. Trans. R. Soc. A* **376**, 20170243 (2018).
- Zhao, L., Li, Q., Zhang, Y., Wang, H., and Du, X., "Integrating the continuous wavelet transform and a convolutional neural network to identify vineyard using time series satellite images," *Remote Sens.* **11**, 2641 (2019).