IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Automatic detection of *Beggiatoa* microbial mats in salmon farms videos

NICOLÁS MARTINEZ[1], Sebastian Moreno[2], Francisco Plaza-Vega [3], José-Emilio Castillo[4], Miguel Carrasco[5]

[1]Universidad Adolfo Ibáñez, Santiago, Chile (e-mail: nicomartigames@gmail.com)
[2]Universidad Adolfo Ibáñez, Santiago, Chile (e-mail: sebastian.moreno@uai.cl)
[3]UUniversidad Santiago de Chile, Santiago, Chile (e-mail: francisco.plaza.v@usach.cl)
[4]Universidad Adolfo Ibáñez, Santiago, Chile (e-mail: jose.emilio.castillo77@gmail.com)
[5]Universidad Diego Portales, Santiago, Chile (e-mail: miguel.carrasco@mail.udp.cl)

Corresponding author: Sebastian Moreno (e-mail: sebastian.moreno@uai.cl).

"This work was was funded by the Agencia Nacional de Investigación y Desarrollo (ANID), Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) of the Chilean government under grant number 11251120."

ABSTRACT *Beggiatoa* microbial mats are natural ecosystems present on the seabed of anoxic zones, and their presence on seabeds are linked with high contamination. For this reason, Chilean aquaculture regulations require the regular analysis of the Salmonid Farming Centers (SFC) seabeds through recorded videos. If *Beggiatoa* microbial mats are detected, the SFC is prohibited from starting a new production cycle. In this paper, we propose a Convolutional Neural Network (CNN) model based on VGG-16 for the automatic detection of *Beggiatoa* microbial mats in seabed videos. Our method involves extracting frames from underwater video footage and classifying them using a CNN. We performed an extensive hyperparameter search and evaluation using k-fold cross-validation, testing over 40 different hyperparameter combinations. The final model includes an optimized architecture of the VGG-16 network, achieving an F2-score of 0.83 ± 0.02 and a sensitivity of 0.93 ± 0.05 on the test set, indicating a failure to detect only 10% of the images with *Beggiatoa* presence. This approach aims to improve the scalability and accuracy of environmental monitoring in aquaculture by reducing the reliance on manual video analysis and minimizing human error. Our automated detection method streamlines the monitoring process, supports compliance with environmental regulations, and contributes to more sustainable aquaculture practices.

INDEX TERMS Aquaculture, *Beggiatoa* microbial mats, Chilean salmon farms, Convolutional Neural Network (CNN)

I. INTRODUCTION

Global aquaculture production is expected to surpass 106 million tonnes by 2030 worldwide, representing a 22% increase compared to 2020 [1]. Nevertheless, expected growth has slowed substantially, with annual projections declining to 2% for 2020–2030 compared to 4.2% for 2010–2020 [2]. This slowdown is driven by multiple factors, including the broader adoption and enforcement of environmental regulations, reduced availability of water and suitable production locations, and increasing outbreaks of aquatic animal diseases related to intensive production practices and marine pollution. In this context, the marine ecosystem, characterized by complex and multifaceted interactions, supports diverse economic activities, yet these can also generate cumulative environmental pressures and harm interconnected ecological systems. Among the most relevant impacts of aquaculture are organic pollution and eutrophication, where eutrophication refers to the excessive accumulation of nutrients, primarily organic nitrogen and phosphorus, along with other waste products [3, 4].

Currently, the analysis of organic matter is a primary method for evaluating marine pollution levels. This assessment is conducted in the benthic zone, the ecological region at the seabed, where the presence of organic matter serves as a proxy for pollution levels [5]. Research efforts have predominantly focused on the interactions of the physical and chemical components of the sediments [6, 7, 8]. Other studies have analyzed the benthos, developing new methods that consider biological organism markers [9, 10]. Consequently, research has aimed to understand the interactions between different benthic organisms and their responses to environmental factors and contamination levels [11]. For example,

marine pollution can alter benthic communities by increasing species mortality and favoring tolerant or opportunistic taxa, including microbial mats. Therefore, shifts in benthic biomass, abundance, and species diversity can reflect changes in pollution levels and ecosystem condition [12].

Microbial or bacterial mats are natural ecosystems that are part of the benthos, where microorganisms form communities a few millimeters thick on a surface or substrate, and can be present on the seabed [13, 14, 15, 16]. These mats are among the oldest ecosystems on Earth, with a fossil record dating back to 3.4 billion years. They are characterized as biosedimentary structures where microbes exploit various environmental niches influenced by gradients in light, redox potential, substrate concentrations for energy acquisition, and the presence of toxic compounds [16]. Studies suggest that the presence of bacterial mats on the seabed serves as a proxy of pollution influence [17, 18]. It has also been proposed that microbial mats develop in areas with chronic contamination because they thrive on the contaminants present [5, 11]. For instance, a recent study determined that microbial mats on the seabed can be used to evaluate contamination levels from salmon production in the aquaculture industry [19].

As stated before, marine farms confine a large number of fish in delimited areas or cages, generating significant amounts of feces and uneaten food. This organic load is deposited on the seabed, impacting local fauna and flora. While some organisms use the organic matter as food, the remainder decomposes, producing hydrogen sulfide (H_2S), which increases temperature and acidifies the sea [19]. This process results in the appearance of *Beggiatoa* microbial mats, bacteria that feed on H_2S [5]. Consequently, as suggested by [20], the presence of *Beggiatoa* microbial mats on the seabed may be linked to the organic matter enrichment in aquaculture areas.

In Chile, aquaculture production has averaged 1.45 million tonnes over the past five years, with 70% derived from fish production, predominantly composed of salmonid species [21]. Given the significance of this industry, both in terms of production and its social and environmental impact, the Fisheries and Aquaculture Law has established several regulations to ensure sustainable practices and environmental protection actions for the conservation, protection, repopulation, and sustainable use of fishery and aquaculture resources. For these reasons, the seabeds of the fish farms are monitored periodically, and the presence of *Beggiatoa* is considered as a pollution marker in aquaculture areas on hard and semi-hard bottoms [19, 20, 22]. Specifically, Chilean regulations mandate a visual record (underwater recording of megabenthos components) for category N°4 centers (culture centers with intensive production systems that are located in sectors of **hard or semi-hard bottoms** and whose depths are equal to or less than 60 meters). This visual record, regulated by Decree 320 of 2021 [23], allows for the evaluation of both the impact and contribution of organic matter on benthic epifauna, as well as the presence of bubbles or bacterial mats primarily composed of *Beggiatoa* bacteria. However, despite the valuable information provided by visual records, they are not without challenges. One of the main issues is the interpretation and evaluation of the videos, which in Chile is managed by the National Service of Fisheries and Aquaculture (Sernapesca) [24].

The inspection process for monitoring seabeds in aquaculture areas involves two stages. Initially, a laboratory provides videos of the Salmonid Farming Centers (SFC) seabed to Sernapesca, accompanied by a report indicating the presence or absence of microbial mats and other relevant elements. Sernapesca then reviews these reports; if they show that microbial mats are not present, Sernapesca is required to view all the videos to confirm this absence. If microbial mats are identified, the assessment is limited to the time intervals in which mats have been recorded. When microbial mats are present, the SFC is not allowed to initiate a new production cycle until normal, healthy conditions have been restored and verified in a follow-up report. Unfortunately, this review process is challenging due to the excessive amount of human hours required and its susceptibility to human error. Recent legislative changes exacerbate this issue by no longer requiring SFC reports to include the presence of microbial mats, thus necessitating Sernapesca to analyze the entire duration of all videos to determine the presence of microbial mats. This significantly increases the required human hours, highlighting the need for an automatic algorithm capable of detecting *Beggiatoa* microbial mats on the seabed. Figure 1 provides examples of these images analyzed by Sernapesca. The top row shows images with *Beggiatoa* mats, where the left image clearly displays a mat, while the presence in the other images is harder to determine. The bottom row contains images without *Beggiatoa*. The resemblance between the top and bottom images highlights how challenging it is to reliably identify *Beggiatoa* mats, emphasizing the need for sophisticated, automated methods for this purpose. To illustrate the segmentation of the *Beggiatoa* class, Figure 2 shows the distinction between regions containing *Beggiatoa* and the surrounding background. The second row presents an image with a purple tint over regions without bacterial presence while maintaining the same color level over regions containing the bacteria.

This paper presents the first model capable of detecting the presence of *Beggiatoa* in seabed videos. Specifically, we introduce a Convolutional Neural Network (CNN) as an image classification model, able to detect specific frames containing *Beggiatoa*, which can subsequently be analyzed by a specialist. Our approach includes a thorough hyperparameter tuning and evaluation process (utilizing k-fold cross-validation) to assess over 40 different hyperparameter combinations.

II. STATE OF THE ART

Object classification within images and videos has been an active area of research over the last few decades [25, 26, 27, 28, 29, 30]. This research area can be, generally speaking, separated into two areas: methods based on deep learning descriptors, and other methods that generate their own de-

FIGURE 1. Images extracted from raw video footage. The top row shows 3 images with *Beggiatoa*, being the left plot a clear example of this type of mats. The bottom row shows 3 images without *Beggiatoa*. The comparison among both rows shows, empirically, the complexity of the problem.

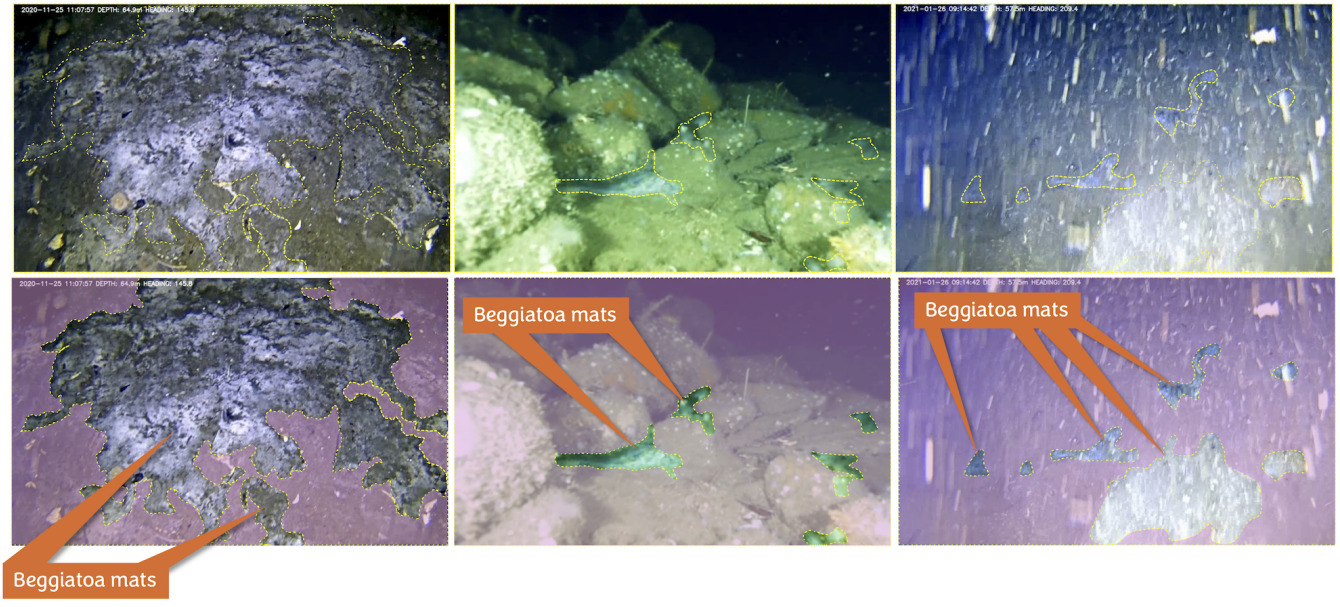

FIGURE 2. Regions with presence of *Beggiatoa* bacteria and their corresponding segmentation. Note that in the second row, a purple tint has been applied over the normal regions of the image.

scriptors based on the task. Currently, deep learning neural networks are used extensively for image classification and video analysis; including behavioral analysis, facial recognition, and autonomous driving [31, 32, 33]. There are many types of deep learning neural networks that can be used for object detection. In all cases, the process is performed through rigorous training, in which the network learns to extract important features from the images and classes. Then, the trained network is able to automatically detect the object in images not used in the training, without the need for human intervention. Unfortunately, we could not find literature related to Beggiatoa microbial mats detection in videos. Similarly, there are few papers related to Beggiatoa microbial mats detection in images. Most of the studies are more than 10 years old [14, 15, 25, 34], with only two recent studies [35, 36], from the same author (described below). This shows the importance and novelty of this work.

[34] presents one of the earliest image-based approaches for detecting and quantifying the spatial distribution of Beggiatoa mats. The authors extracted 2840 georeferenced mosaic images from ROV video footage using high-precision navigation data processed with MATISSE (IFREMER). During preprocessing, overlapping mosaic regions were removed. Segmentation was then performed using a *watershed* transform [37] combined with *relaxation-based labeling* [38], achieving a reported precision above 90%. However, the method is not fully automatic, as it still requires human intervention.

[25] focused on the automatic detection and quantification of Beggiatoa mats in images. The images were obtained

through the extraction of frames from ROV videos, and three methods were compared using the classification ratio. The first method defines seed points of a region growing algorithm as centers of regions of gradient-based watershed transformations. In the second method, human defines seed points to determine homogeneous regions. Finally, the third method is based on texture analysis estimating the similarity between different image samples through the Kullback-Leibler divergence. As mentioned in the article, the results ''are not satisfactory since the score achieved was weak''.

[15] used a method to quantify Beggiatoa mats in images, in order to study the behavior of demersal decadopods (benthic crustaceans). However, the detection is carried out under very specific and ideal conditions. All selected images had to have the camera oriented vertically downwards and with uniform luminosity. Even with these restrictions, the results of the model were unsatisfactory, arguing that they could have been affected by the lighting system. To improve the classification, the authors proposed another work with 52 new images [14]. In the selected images, alterations caused by other benthic species were avoided. First, a constant Region of Interest (RI) was defined and background extraction was performed using a *Gaussian Blurring* filter. Afterward, an image extraction was performed within the RI, which was followed by the addition of the RGB channel matrices. Then, an image enhancement process and the application of a threshold for pixel classification were performed. Finally, the percentage coverage was calculated on the resulting image and the fractal dimension was estimated using the box-counting method. The results of this process reached a Pearson correlation value of 0.67 with respect to the coverage percentage.

Note that previous studies are more than 10 years old, so the techniques used correspond to classic machine learning models or other image processing techniques. Moreover, these methodologies require human intervention or depend on an analysis to extract specific characteristics from the images, in order to obtain descriptors with which the models are trained, or software that facilitates this process. Finally, all those works were focused on the quantification of Beggiatoa mats, instead of the detection of them. Then, current methods assumes that images have Beggiatoa mats. In recent years, and due to the scarcity of labeled data for Beggiatoa, [35] used a self-training method to estimate the percentage coverage of Beggiatoa in a seabed image. The method combines Otsu thresholding [39] and Fully Convolutional Networks (FCN), obtaining a precision of 74.35%. Recently, [36] proposed the first Beggiatoa image classification method based on ensemble learning. This method consists of the combination of three models to extract features from images (one of them a Convolutional Neural Network) and four multiple classification models, obtaining values close to 91.4% in the most used metrics (accuracy, precision, recall, and F_1-score). Unfortunately, even though these results seem impressive, it seems that their methodology has multiple methodology issues, labeled as important errors by [40].

As it is explained in the methodology of [36], the images were cropped and labeled, then a data augmentation process was applied. Posterior to the data augmentation process, the data was divided into two sets training (90%) and testing (10%). As can be observed, the testing data is highly related to the training data. First, cropping the image into 128×128 pixel images implies that highly correlated data points can be in the training and testing data. For example, from a single image, even though there is no overlap, we could generate multiple images, but all of them will be related because they are generated from the same image. Second, data augmentation was applied before the separation between training and testing data. Then, the image A', generated from A, could be in the testing data while A will be in the training data, masking the real performance of the model. For example, A' could be the same image as A, but rotated in 30 degrees, as it is mentioned in [36].

III. METHODOLOGY

This paper follows the deep learning methodology for computer vision described in [41]. The methodology was adapted obtaining two stages. First we applied the data acquisition and processing, and then we apply the model evaluation and application. As shown in Figure 3, the proposed workflow comprises three main stages, data preparation, modeling, and evaluation, for detecting Beggiatoa mats from ROV videos. We proceed to barely describe each of these stages, and more details are given in each subsection.

We started analyzing the video acquisition and the manual image identification described in the reports. With this information, we extracted the frames (images) from the videos with their respective label, i.e., *Beggiatoa* mat presence (True) or *Beggiatoa* mat absence (False). In this process, we take special care to maintain a similar number of images for both classes. Then, in the model application and evaluation model subsections, we search for the best hyperparameters and neural network architecture to obtain the best possible performance. We take special care in the evaluation process. For this reason, we apply k-fold cross-validation based on groups of videos. Using this strategy, we ensure that the images from the test set are unrelated to the training data.

A. VIDEO ACQUISITION

The data provided by SERNAPESCA consists of 114 videos from the years 2020 and 2021 from 30 aquaculture centers located throughout Chile. Unfortunately, the data can not be share because of data privacy policy. The visual records were carried out by means of underwater recording of the components of the mega benthos, with special attention to the presence of filamentous bacteria (covered with visible microorganisms) and/or gas bubbles. Each video was filmed using the following procedures defined by SERNAPESCA (for detailed information, please refer to the exempt resolution number 3612-2009 of the Chilean government):

1) Equipment: Underwater recording can be done by diving or remote system. The equipment must have a wide-angle lens (120° or more) and the ability to record with

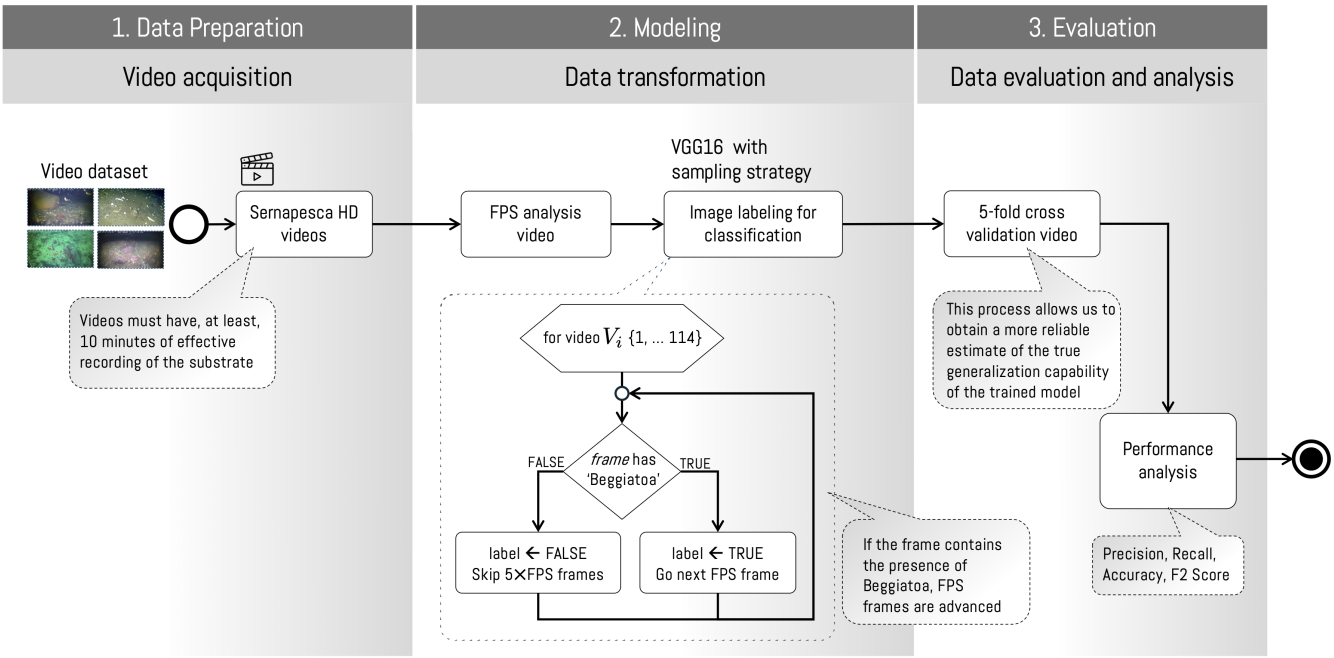

FIGURE 3. Overview of the proposed workflow for Beggiatoa mat detection from ROV videos. The pipeline comprises three stages: (1) Data preparation, where a dataset of Sernapesca HD videos is acquired; (2) Modeling, including FPS analysis and frame extraction/labeling using a sampling strategy (advance one frame when Beggiatoa is present, otherwise skip five frames) for training a VGG16-based classifier; and (3) Evaluation, where a 5-fold cross-validation is performed at the video level to prevent similar frames from appearing in both training and test sets, followed by performance analysis using Precision, Recall, Accuracy, and F2-score.

good lighting (natural or artificial) and adequate focus. In addition, there must be an auxiliary lighting system and an automatic photometric system to regulate the opening of the focus, allowing the entrance of adequate light. The cameras must be color digital cameras, with a high definition of at least HD or full HD. Periodic maintenance must be carried out on the equipment to ensure the good quality of the camera images.

2) Recording: A filming distance between the camera and background must be kept, as well as a camera drag speed such that quality filming of the background is ensured, allowing a clear and adequate observation for the correct distinction of the different components of the mega-benthos and the identification of microorganism covers and gas bubbles. The recording must not be edited, and a copy with uninterrupted recording from the surface must be provided before the dive. In the case of preliminary site characterization, the video must have, at least, 10 minutes of effective recording of the substrate (the recording time of the descent and ascent of the camera will not be considered).

B. DATA TRANSFORMATION

The automatic frame extraction process (including the label) was created based on SERNAPESCA reports and algorithm 1. Recall, SERNAPESCA experts already verified and corrected the reports with the correct label. Then, our algorithm proceeds to extract some frames with their corresponding label combining both information (video and reports).

Algorithm 1 receives a video and a report and generates the images with their corresponding labels. The algorithm starts by opening the video and extracting the Frame per Second (FPS) and all the video frames (images that make up the video). Subsequently, the video frames are inspected: if a frame shows the presence of *Beggiatoa*, the analysis proceeds by FPS frames; otherwise, it skips ahead by $5 \cdot FPS$ frames. The sampling strategy at $5 \cdot FPS$ has been widely adopted in CNN-based video learning pipelines, as it offers an effective trade-off between temporal redundancy reduction and computational efficiency while preserving sufficient visual diversity for training CNN models [42]. The change in the sampling interval is further motivated by the relevance of detecting the presence of *Beggiatoa*. Inspired by the MGSampler framework [43], which dynamically adjusts the sampling rate in response to motion events, we adapt this strategy by replacing motion with *Beggiatoa* presence. Specifically, instead of increasing the sampling process when motion is detected, the sampling strategy is dynamically adjusted based on the presence of *Beggiatoa*. In total, we generated 40,774 images, from which 19,512 have *Beggiatoa*.

C. DATA EVALUATION

To evaluate the models, we will use *k-fold* cross-validation to calculate the mean error and its standard deviation, relying on several performance metrics. *k-fold* cross-validation estimates the error distribution by dividing the data sets into k subsets with similar distribution across classes. Then $k - 1$

Algorithm 1 Image and label data extraction

Require: video and report
  open video;
  extract Frame per Second (FPS) information;
  extract all frames from video;
  while frames available **do**
    extract image from frame
    if image has *Beggiatoa* **then**
      save image with label TRUE;
      go forward to next frame (FPS);
    else
      Save image with label FALSE;
      go forward $5 \cdot FPS$;
    end if
  end while

folds are selected for training and the remaining is used for test. This process is repeated until all folds are used for testing, and the final error estimation is given by the average and standard deviation of all the estimated errors. An example of this process can be observed at Figure 4, where we selected $k = 5$.

In this work, we use *k-fold* cross-validation with 5 subsets of frames. However, we applied *k-fold* cross-validation using the videos. This avoids the separation of similar images, from the same video, into training and testing, avoiding *data-leaking* (use of test data in training data). Specifically, each video and its frames obtained in the previous step were associated with a subset (*fold*) 1, 2, 3, 4 or 5 following an alphabetical order; i.e., the first video to 1, the second to 2, the third to 3, the fourth to 4, the fifth to 5, the sixth to 1, and so on, until all videos belonged to a bin. This process can be observed in Figure 5. The top part of the figure shows the 30 different Salmonid Farming Centers, from which 114 videos were generated. As can be observed, we put the videos in different folds, and then we proceed to extract the images. Moreover, even though some videos could originate from the same center, they are not necessarily related to the same physical space, nor do they cover the exact same area. With this process, we ensure that the images of a video are specifically in a unique fold, avoiding similar images in the training and test folds.

In this paper, we used the following performance measures: *Accuracy*, *Recall*, *Precision*, and F_2-*score*. These measures are based on the confusion matrix, which is a table that shows the performance of a classification model using four values: True Positives (TP) is the number of frames with the presence of *Beggiatoa* correctly classified; True Negatives (TN) is the number of data frames without the presence of *Beggiatoa* correctly classified; False positives (FP) is the number of frames without the presence of *Beggiatoa* from data incorrectly classified as having the presence of *Beggiatoa*; and False Negatives (FN) which is the number of frames with the presence of *Beggiatoa* incorrectly classified as without the presence of *Beggiatoa*.

Accuracy (equation 1) is one of the most basic measures of model evaluation. It indicates the percentage of correctly classified frames (TP+TN), over the total number of data points. *Accuracy* ranges between 0 and 1, where a high *Accuracy* implies that the model can predict most data frames correctly.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The *Precision* (equation 2) is the number of data points with *Beggiatoa* correctly predicted over the total of frames classified with the presence of *Beggiatoa* (TP+FP). *Precision* varies between 0 and 1, where a high *Precision* implies that the majority of frames classified with the presence of *Beggiatoa* are classified correctly. However, if we have a high number of False Negatives, this values performance measure is not affected.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Similarly, the *Recall* (equation 3) is the number of data points with *Beggiatoa* correctly predicted over the total frames with the presence of *Beggiatoa* (TP+FN). *Recall* varies between 0 and 1, where a high *Recall* implies that the majority of frames with the presence of *Beggiatoa* are correctly classified. However, if we have a high number of false positives, this performance measure is not affected.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Finally, we use the F_β-*score*, a measure that harmonizes *Precision* and *Recall*, seeking a balance between the two performance using β. The value of β implies that the *Recall* is considered β times more important than *Precision*. The F_β-*score* (equation 4) varies between 0 and 1, where a high F_β-*score* implies that the model can classify the positive class and, most importantly, generate a low number of FN and FP. Although TP is associated with the class with fewer labels, we will report the F_β-*score* using both classes as TP, avoiding subjective interpretations of the errors.

$$F_\beta\text{-score} = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall} \tag{4}$$

In the current problem, it is very important to avoid FN (images where *Beggiatoa* is present but the model classifies them as without *Beggiatoa*). Consequently, *Recall* should be of greater importance than *Precision*. Therefore, we use the F-*score* with $\beta = 2$, since *Recall* is considered twice as important as *Precision*.

D. CLASSIFICATION MODELS

This paper uses the VGG-16 model as its base convolutional neural network (CNN) model [44]. VGG-16 was selected over more recent architectures due to its architectural simplicity, ability to extract high-quality generic features, relatively low computational requirements, and its role as a well-established baseline, making it particularly suitable for the

FIGURE 4. **k-fold cross-validation with $k = 5$. The 5-fold cross-validation process separates the data into 5 folds. Then, it chooses 4 folds for training (blue folds) and 1 for test (red fold), estimating a performance error for training (TrP_i) and test (TeP_i). Finally, the average and standard deviation are calculated.**

size and characteristics of the dataset under study (an image representation of this model can be observed at Figure 6). Generally speaking, this neural network can be separated into two parts, the first part extracts the features from the images (left part of the image), and the second part is a Feed-forward Neural Network (FNN, right part of the image) that learns from the features. Therefore, we start over a search for the architecture of the FNN network and its hyperparameters. Once the best model (architecture and hyperparameters) was selected, the final model was trained again using *fine tunning*. Fine tunning means that the weights of the CNN corresponding to the feature extraction process were also adjusted. Finally, a search for the classification threshold was carried out, seeking to improve the performance metrics.

1) Feed-forward neural network

Given the extracted features from the original VGG-16 (trained on Imagenet [45]), we used a Feedforward Neural Network (FNN) to classify the images. We began by searching for optimal hyperparameters for the FNN architecture, such as the number of neurons, layers, and epochs. During this hyperparameter search, we compared the performance of all learned models and selected the one with the best performance. The output layer consistently consisted of a single neuron with a sigmoid activation function.

The base architecture consists of 128 neurons for the hidden layer with ReLu activation function, followed by a *dropout* of 0.5 and the output neuron. A maximum of 20 training epochs was considered. From this base model, we try the following different combinations: a) *learning rate*: 0,001 y 0,0001; b) *batch size*: 16, 64 y 256; c) *data augmentation*: true/false data augmentation (feature-wise centering, rotation, shifting, horizontal and vertical flipping, as well as shear and zoom transformations); and d) *optimizer*: Adam and SGD. This search generated a total of 24 different models. To simplify the visualization process of the results, we show all the models in Table 1. The right column of the tables shows an identifier of each model for reference in the results section.

batch size	*optimizer*	*learning rate*	*data augmentation*	**Model**
16	Adam	0.0001	true	**1**
			false	**2**
		0.001	true	**3**
			false	**4**
	SGD	0.0001	true	**5**
			false	**6**
		0.001	true	**7**
			false	**8**
64	Adam	0.0001	true	**9**
			false	**10**
		0.001	true	**11**
			false	**12**
	SGD	0.0001	true	**13**
			false	**14**
		0.001	true	**15**
			false	**16**
256	Adam	0.0001	true	**17**
			false	**18**
		0.001	true	**19**
			false	**20**
	SGD	0.0001	true	**21**
			false	**22**
		0.001	true	**23**
			false	**24**

TABLE 1. **24 different configurations between hyperparameters and data augmentation.**

To avoid overfitting, we select the number of epochs for each model based on the loss function. For this objective, we analyze the loss function for the training data (with a subset used as validation data). We choose the number of epochs where the the difference between both curves was minimum, or if in two epochs the improvement of the loss function was lower than 0.01.

After the selection of the best model, we proceed to modify the neural network architecture. Let model **X** be the best model of all 24 configurations from Table 1. We modify the FNN architecture from model **X** trying multiple combinations of neurons and hidden layers. However, we kept the ReLu function constant and we also added a *dropout* layer with 0.5

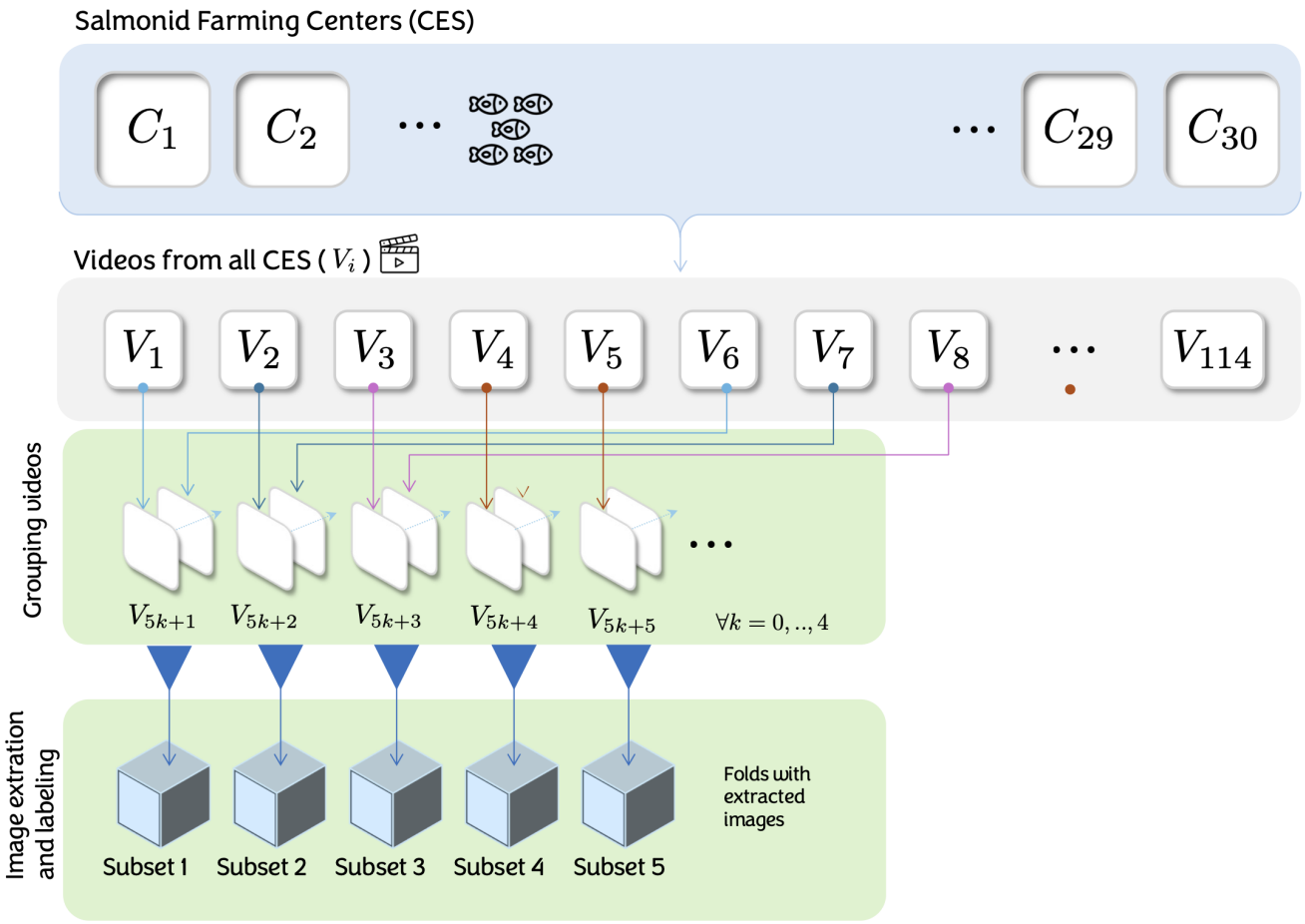

FIGURE 5. **k-fold cross-validation with video. The k-fold cross-validation process is applied over the videos, and then we proceed to extract and label the images. This process avoids the use of similar images in the training and test set, obtaining a better error estimation of the model.**

after each hidden layer. We start changing the original hidden layer to 32, 64, 128, 256, 512, and 1024 neurons. We also added a second layer using 0, half, or the same number as the first layer. For example, if the first layer has 256 neurons, we also try a second layer of 0 neurons, 128 neurons, and 256 neurons. In summary, we trained 18 different models, shown in the table 2. Note that model X.7 is equivalent to model 9 of 1.

Finally, we also apply a soft tuning process, unfreezing neurons from the VGG-16 base model used for feature extraction and incorporating them into the training process. In total, we tried 5 different models, which progressively unfreeze each of the five convolutional blocks from VGG-16 (from the last block to the first). However, the soft tuning process led to models that classified all images as negative (absence of *Begiattoa*), so we prefer to discard these details and results. This result is not surprising, because the model may learn patterns specific to the training videos that do not appear in the test data, since different videos are used for evaluation, a phenomena known as distribution shift [46].

All code was implemented using Keras and executed on an NVIDIA DGX A100 system. This GPU cluster is hosted at Universidad Adolfo Ibáñez, to which three of the authors are affiliated. In addition, considering the limited computational resources available at SERNAPESCA, we also evaluated the inference process on a standard desktop computer equipped

First layer	Second layer	Model
32	0	**X.1**
	16	**X.2**
	32	**X.3**
64	0	**X.4**
	32	**X.5**
	64	**X.6**
128	0	**X.7**
	64	**X.8**
	128	**X.9**
256	0	**X.10**
	128	**X.11**
	256	**X.12**
512	0	**X.13**
	256	**X.14**
	512	**X.15**
1,024	0	**X.16**
	512	**X.17**
	1,024	**X.18**

TABLE 2. **Different models based on the Feed-forward neural network architecture.**

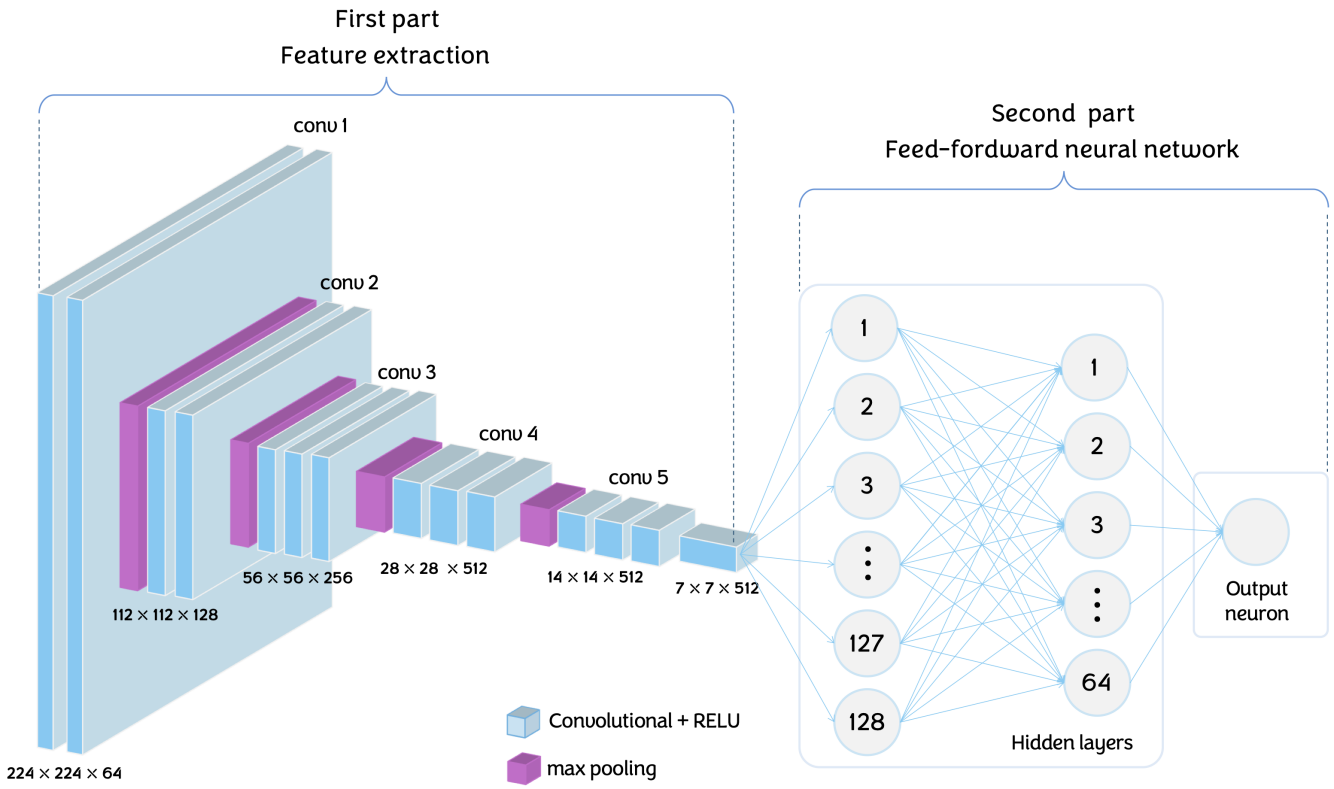

FIGURE 6. **VGG-16 architecture. The first part corresponds to the feature extraction, the second part is the feed-forward neural network (FNN). Note, each of the 7 × 7 × 512 neurons from the last block of the first part are connected to the 128 neurons of the FNN.**

with an Intel i5 CPU and 8 GB of RAM.

IV. RESULTS

In this section, we show and discuss the results of the experiments previously described. In summary, all results correspond to the mean and standard deviation of each metric, which were estimated using 5-fold cross-validation. The performance metrics are F_2-score, Precision, Accuracy, and Recall. Furthermore, to analyze overfitting and determine the number of training epochs, we analyzed the loss functions (training and test) of each learned model.

A. HYPERPARAMETER SEARCH

Figure 7 shows the results of the 24 models described in the table 1, where the average (circle) and standard deviation (bars) of the training (blue) and test (red) data are plotted. As can be observed, the performance declines significantly when using *data augmentation* techniques. We can see this in Figure 7, since in most cases, when comparing the odd models (with data augmentation) with the model that follows (same hyperparameters as the odd but without data augmentation), we notice that the second model has better performance. This could be explained because data augmentation techniques generate a set with greater diversity and variability with the aim of strengthening the model against images in different circumstances (point of view, scale, light, etc.). Then, in the case of *Beggiatoa* images, these techniques could generate circumstances that do not occur in reality, explaining the negative impact on performance, a phenomena known as label corruption [47]. For example, if an image has *Beggiatoa* in the top right corner, applying some data augmentation technique, such as zoom range or horizontal/vertical shift, eliminate the *Beggiatoa* from the image, but it keeps the original label indicating that the image has *Beggiatoa*. In this case, the label would be positive, but the true label should be negative.

Regarding the best model, we could not observe a specific trend in the rest of the hyperparameters. Besides the obvious cases (such as model 19), we can not observe, without applying a statistics test, a model that is significantly better than all others. However, we select model 20 because it has the best F_2-score average for the training and test sets, and a small standard deviation. Model 20 corresponds to a VGG-16 architecture with one hidden layer of 128 neurons (ReLu activation function), batch size 256, optimizer ADAM, learning rate 0.001, and without data augmentation.

B. ARCHITECTURE SEARCH

From the selected model (model 20), we proceed to search over different Feed-Forward Neural Network (FNN) architectures. Recall, we modified the number of neurons and the number of hidden layers, resulting in 18 new models (table 2). The results of all architectures are shown in Figure 8, where the average (circle) and standard deviation (bars) of the training (blue) and test (red) data are shown for each of the 18

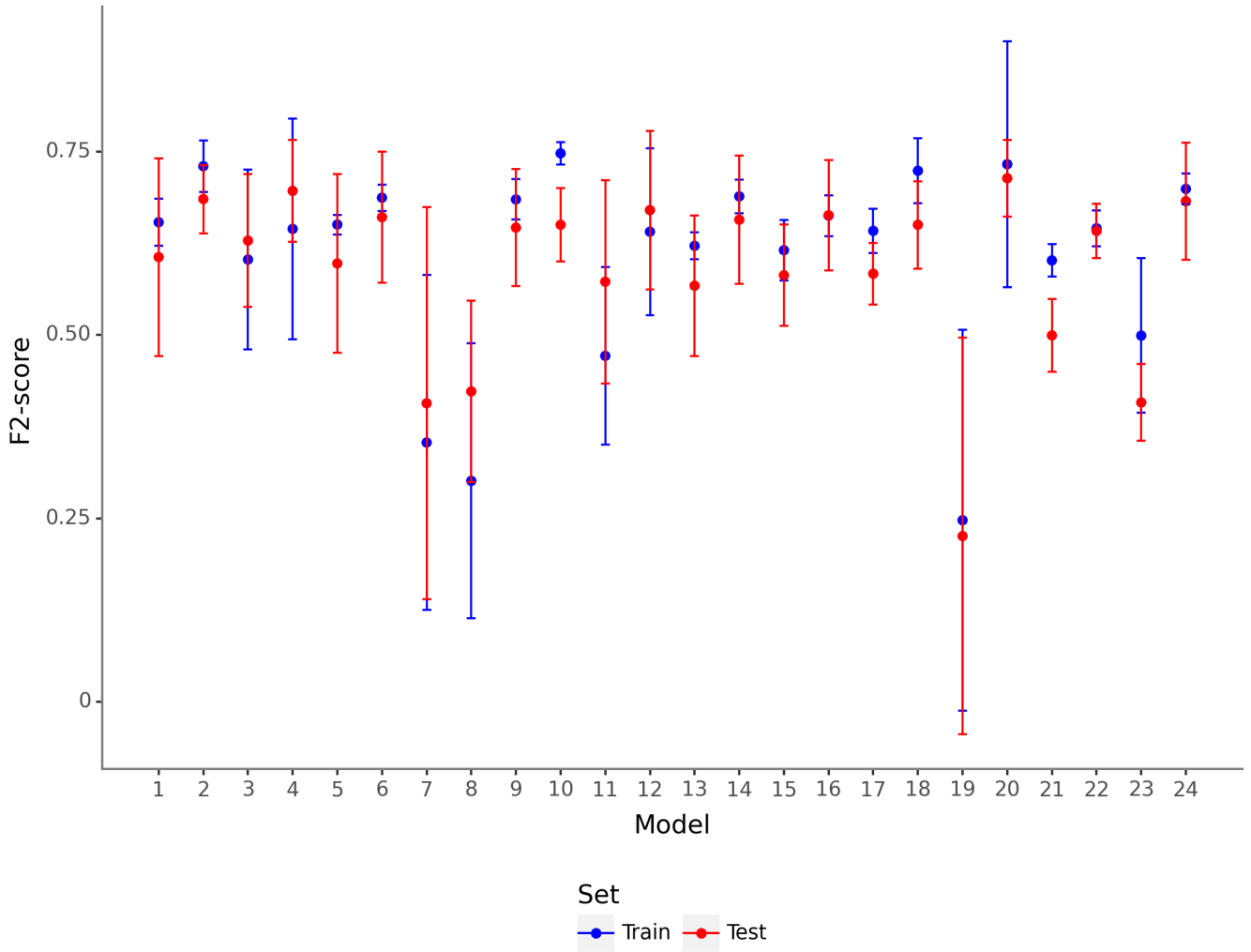

FIGURE 7. **Results for the hyperparameter selection. Starting from 1, model X is the same model than X+1, but model X was trained with data augmentation.**

models described in the table 2. In the figure, we can notice the almost zero performance of some models, such as 20.2, 20.3, and 20.6. These performances are obtained when almost all the images are predicted with the absence of *Beggiatoa*. Furthermore, the incorporation of another hidden layer for models with a single hidden layer of 32 or 64 neurons (models 20.1 and 20.4) decreases their performance (models 20.2, 20.3, 20.5, and 20.6). Analyzing the other models, we choose model 20.8 because it has the highest average performance for the training and test set, and a small standard deviation. However, it is important to note, that the differences of this model against several other models are not statistically significant, because of the high standard deviation of the other models. This high variability can be attributed to the strict evaluation protocol. Specifically, each fold contains different videos, leading to substantial differences between the training and test sets, as the data originate from different domains. While this setup provides a more reliable estimate of the model's generalization capability, it also increases performance variability. Models that adapt well to certain domains may exhibit higher error on the test fold, thereby inflating the standard deviation. As a result, given the large variability across folds, the null hypothesis of the significance analysis cannot be rejected. For this reason, we preferred to choose a model with good performance and small standard deviation.

C. ANALYSIS OF THE SELECTED MODEL

We analyze the effect of the decision threshold for model 20.8. Figure 9 shows the average precision–recall curve computed over 1,000 different threshold values (ranging from 0.0 to 1.0 in steps of 0.0001). As shown, when the threshold is set to 0.0, recall reaches 1.0 while precision remains below 0.5, indicating the imbalance toward the negative class. At a threshold of 0.5, recall remains high and a satisfactory precision is achieved. However, for threshold values greater than 0.5, precision decreases substantially; since recall is

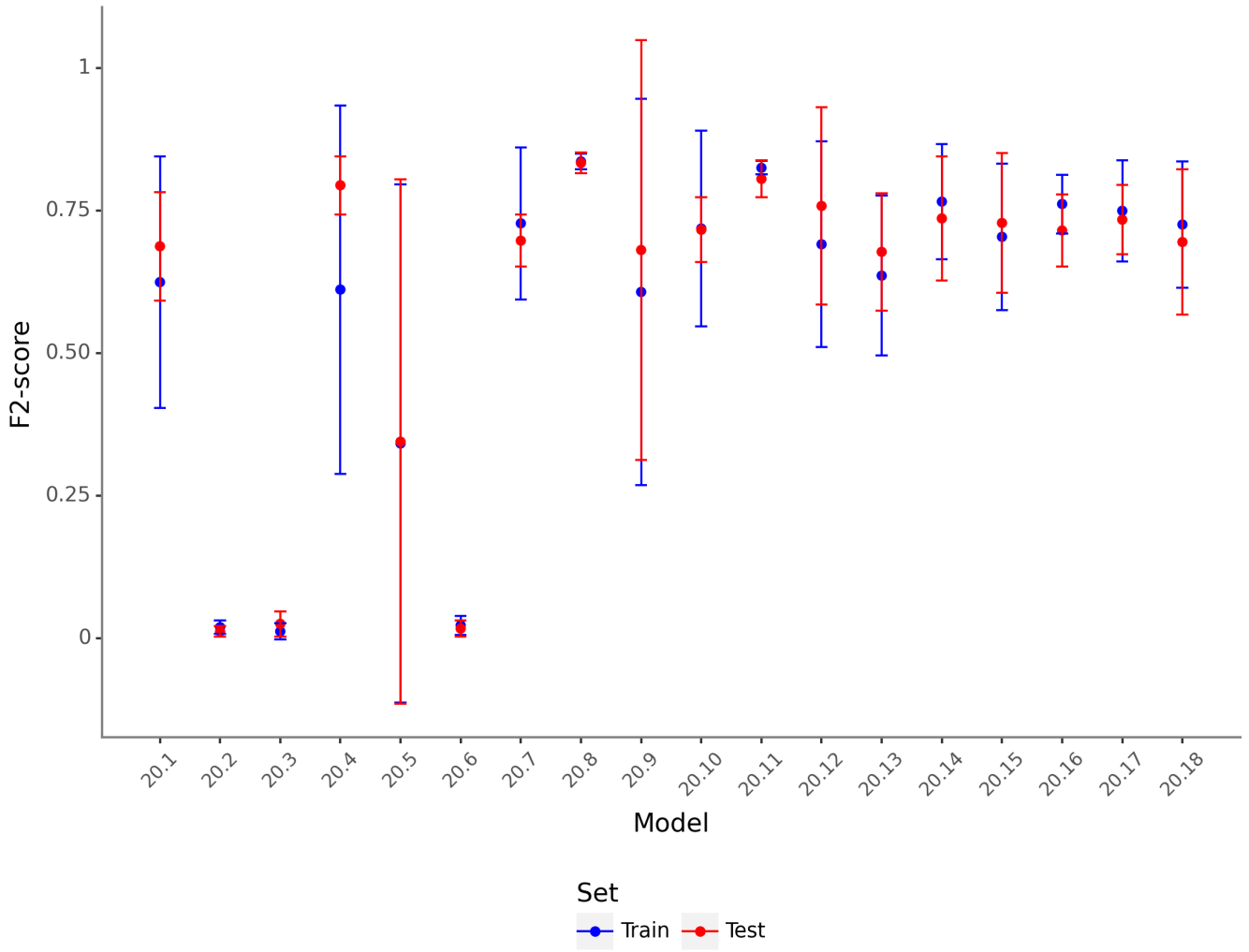

FIGURE 8. **Results for the architecture selection. Models 20.1, 20.4, 20.7, 20.10, 20.13 and 20.16 corresponds to models with a single hidden layer of 32, 64, 128, 256, 512, and 1,024 neurons, respectively.**

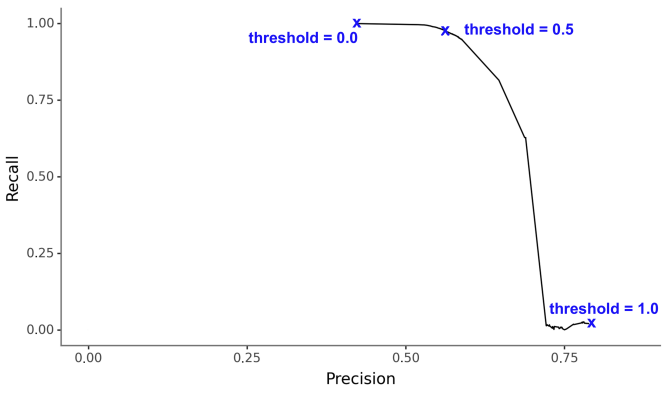

FIGURE 9. **Threshold search. Recall vs Precision curve for model 20.8.**

prioritized over precision in our evaluation, this results in a significant drop in the F2-score, setting the threshold to 0.5.

Given the final threshold, Table 3 shows the performance of the 20.8 model in terms of accuracy, recall, precision, and F2-score for both the training and test datasets. As can be observed, the small differences between training and test results indicate that the learned model generalizes well. This is particularly relevant because the videos used for testing are completely excluded from the training and validation stages and are only employed for the final evaluation. It is also important to recall that, in the current workflow, human analysis requires observing the entire video. For this reason, false positives are less critical than false negatives. If a video containing *Beggiatoa* is incorrectly classified by the model as negative (i.e., a high number of false negatives), the human analyst must review the entire video to verify the result. In contrast, if a video yields at least one correctly classified positive frame, even in the presence of many false positives, the salmon farm is closed, saving hours of works for the human analyst.

FIGURE 10. Images extracted from a test fold. The left image show a True Positive case, the middle image show a False Positive case, and the right image shows a False Negative case. The comparison among the images show, empirically, the complexity of the problem.

Metric	Training	Test
Accuracy	0.67 ± 0.04	0.72 ± 0.03
Precision	0.55 ± 0.03	0.60 ± 0.05
Recall	0.96 ± 0.02	0.93 ± 0.05
F2-score	0.84 ± 0.01	0.83 ± 0.02

TABLE 3. Results for model 20.8 for the training and test data.

Finally, Figure 10 show a qualitative analysis of the performance of the model. In the figure, three images are presented, two of which contain *Beggiatoa*. The left image shows a true positive case, where the model correctly predicts the presence of *Beggiatoa*. The middle image corresponds to a false positive case. As observed, even for a human expert it is difficult to determine that this image does not contain *Beggiatoa*, since several white segments resemble the visual appearance of the *Beggiatoa* bacteria. Finally, the right image represents one of the few false negative cases produced by the model. In this example, the *Beggiatoa* mat can be easily confused with sea shells or rocks due to their similar color and texture.

V. DISCUSSION

Deep learning and convolutional neural networks have been increasingly adopted in aquaculture for tasks such as fish classification, biomass estimation, and behavioral monitoring [48, 49, 50, 51, 52]. However, their application to seabed analysis, particularly the detection of *Beggiatoa* mats, remains highly underdeveloped. As noted in recent systematic reviews, this task is either absent or mentioned only as a prospective challenge, with no established pipelines or benchmarks currently in place [51, 52].

To date, only two contributions have proposed CNN-based methods for detecting *Beggiatoa* mats in underwater imagery. Chen et al. (2022) [35] introduced a self-supervised approach using pseudo-labels to estimate bacterial coverage, followed by an adaptive ensemble strategy to improve robustness across mat types (Chen et al., 2024 [36]). The present work builds on these efforts by delivering a fully validated system for frame-level detection in real regulatory inspection footage from salmon farming operations. The rarity of such approaches reflects a set of technical and operational challenges. First, *Beggiatoa* mats exhibit high visual ambiguity, lacking consistent shape or color, and are easily confused with similar features (e.g., sediment glare, shells). Second, annotated datasets are scarce, as expert labeling is labor-intensive and positive samples are limited. Third, underwater videos vary widely in quality, contrast, and lighting, which introduces domain shifts. Fourth, the class distribution is highly imbalanced, making learning unstable. Fifth, CNNs trained on site-specific data often fail to generalize across different depths, sediments, or recording conditions. Finally, no benchmark datasets or public evaluation protocols exist for this problem [51, 52], which limits reproducibility and comparison across methods.

A central challenge revealed by the experimental results is the presence of domain shift, more specifically covariate shift, across underwater videos. Variations in illumination, turbidity, camera trajectory, seabed composition, depth, and recording protocols induce systematic changes in the input distribution, even when the underlying task (*Beggiatoa* detection) remains unchanged. This setting corresponds to a classical covariate shift scenario, where the distribution of the inputs varies across domains while the conditional relationship between inputs and labels is preserved [53, 54]. From this perspective, the observed variability across cross-validation folds should not be interpreted as model instability, but rather as an empirical manifestation of covariate shift under realistic operational conditions. Importantly, the adopted video-level cross-validation protocol explicitly exposes the model to this variability, providing a conservative and more realistic estimate of generalization performance compared to frame-level random splits, which are known to yield overly optimistic results in the presence of domain shift [47, 54]. While such distributional shifts may limit out-of-the-box scalability across heterogeneous environments, they also highlight the relevance of future extensions incorporating domain adaptation or domain-invariant representations [46]. Despite these limitations, our results show that CNNs-when trained with expert-validated data and evaluated under cross-site conditions—can deliver high sensitivity in identifying *Beggiatoa* in operational settings. Note, the system remains conservative by design, relying on human review in the absence of positive detections. So, this hybrid approach improves scalability while preserving regulatory confidence. Future progress will require broader datasets, shared benchmarks, and architec-

tures explicitly designed for domain adaptation.

Within this context of limited prior work and structural constraints, it is important to interpret our results in relation to recent contributions, particularly those by [36, 55]. However, there are important differences with respect to these works making them hard to compare. First, [36] solves a different problem, quantification of *Beggiatoa* in sub-images of size 128×128. Second, the data and the trained model are not publicly available, being impossible to use it. Third, implementing and training this model would require manual pixel-level annotation of all 40,774 images. Fourth, even their these results seem impressive, it seems that their methodology has multiple issues labeled as important errors by [40], as was previously explained in subsection II.

From an operational and regulatory perspective, false negatives and false positives have asymmetric implications in environmental monitoring, and their impact must be interpreted in relation to the role played by the automated system within the inspection workflow rather than in isolation [56]. In the Chilean seabed inspection framework, regulatory action is triggered by the confirmed presence of *Beggiatoa* microbial mats in at least one verified video frame, whereas the absence of detections requires full-video inspection by a human expert. As a result, in the worst case scenario, a video with only false positives and no true positives will force the manual revision of the entire video (the actual procedure); but, a single true positive in the video will reduce significantly the evaluation time. Similarly, false negatives revert the process to the conservative baseline procedure already in use, without increasing regulatory risk.

Another important conceptual aspect of the proposed system is the choice of frame-level detection rather than direct video-level classification. While video-level approaches collapse temporal information into a single binary outcome, frame-level inference preserves spatial and temporal resolution, enabling fine-grained localization of potential *Beggiatoa* occurrences. This design choice is particularly relevant in inspection and monitoring scenarios, where visual traceability and explicit evidence are required to support expert validation [26, 27]. By identifying specific frames with a high probability of *Beggiatoa* presence, the system supports targeted human review instead of opaque video-level decisions. Similar frame- or image-level detection strategies have been successfully adopted in other environmental monitoring applications under a human-in-the-loop paradigm, where automated systems assist experts without replacing final decision-making authority [57]. Moreover, frame-level outputs provide a flexible foundation for future extensions, such as temporal aggregation or event-based summaries, built on interpretable and auditable visual evidence rather than end-to-end video-level predictions.

In a broader context, it is worth noting that similar automated or semi-automated systems have been developed in other marine industries. For instance, Revelas et al. (2018) [58] presented a semi-automated sediment profile imaging system to map benthic habitat conditions and seafloor deposits, primarily in offshore oil and gas contexts. Although the application domain differs, the underlying goals of benthic condition assessment and automation are closely aligned. This suggests opportunities for methodological transfer and underscores the potential of integrating automated seabed analysis tools into aquaculture monitoring frameworks.

The importance and novelty of this model can be extended to multiple uses. The presence of *Beggiatoa* mats not only could be used as indicators for pollution levels in aquaculture, but for other elements. For example, [5] highlights the negative correlation between abundance of *Zostera marina* (eelgrass) and *Beggiatoa*, associated with high sulfide levels in marine areas, caused from organic pollution and sediments from saw mills in Washington (USA), where *Beggiatoa* proliferates.

From a regulatory and governance perspective, the results of this study should be interpreted in the context of long-standing limitations in environmental monitoring of salmon aquaculture in Chile, particularly in a scenario of continued industry expansion. Previous studies have consistently highlighted that current regulatory frameworks rely on limited and often insufficient monitoring capacity, especially with respect to benthic impacts, eutrophication processes, and early-warning indicators of ecosystem degradation [59, 60, 61].

In this context, the proposed frame-level detection system contributes not by replacing regulatory decision-making, but by strengthening the operational capacity of existing inspection protocols. The identification of frames with a high probability of *Beggiatoa* presence allows inspectors to prioritize expert attention, reducing inspection time while preserving conservative safeguards. This directly addresses practical constraints faced by regulatory agencies, such as limited human resources and the increasing volume of inspection data associated with the growth of aquaculture activities.

Recent policy-oriented analyses have emphasized the need for more comprehensive, transparent, and scalable environmental monitoring systems in Chilean aquaculture, including the development of early-warning mechanisms, ecosystem-level indicators, and carrying-capacity-based management approaches [62]. The proposed methodology aligns with these recommendations by providing a reproducible and scalable tool that can be integrated into broader monitoring frameworks, supporting precautionary environmental governance. Importantly, automated screening does not relax regulatory standards but enhances their enforceability, consistency, and transparency.

VI. CONCLUSION

Beggiatoa microbial mats are natural ecosystems that are present on the seabed of anoxic zones. Its presence on the seabed under salmon production areas in the aquaculture industry is an indicator of high contamination. Currently, in Chile, the presence of *Beggiatoa* in salmon farms is carried out manually by analyzing videos of the seabed, which is not scalable and subject to human error. In this paper, we propose the use of the pre-trained VGG-16 (a convolutional neural

network) to detect the presence of *Beggiatoa* bacterial mats in images of the seabed under aquaculture areas. The development of the automatic *Beggiatoa* mats detection approach described in this article, marks a significant advancement in addressing technological and technical gaps, thereby enhancing the efficacy of SERNAPESCAs inspection processes.

In this paper, we search over multiple Convolutional Neural Network models varying different hyperparameters, such as batch size, optimizer function, and learning rate, besides trying data augmentation and soft-tunning. In addition, we also modify the architecture (number of layers and neurons) of the feed-forward network. We trained all models using 40,774 images (19,512 with *Beggiatoa*) labeled by SERNAPESCA experts. The images were extracted from 114 videos from 2020 and 2021 from 30 aquaculture centers located throughout Chile. The capacity of the model to detect the presence of *Beggiatoa* in images was empirically demonstrated, being supported by the use of *k-folds cross validation*, where each fold considers different videos to avoid mixing similar images into the training and test set. This process allows us to obtain a more reliable estimate of the true generalization capability of the trained model.

The final model corresponds to a VGG-16 architecture without data augmentation and soft-tunning with two hidden layers (128 and 64 neurons, respectively, and ReLu activation function), batch size 256, optimizer ADAM, and learning rate 0.001. The model achieved a F_2-*score* of 0.83 ± 0.02 and a sensitivity of 0.93 ± 0.05 on the test set, which implies that approximately the model fails to detect only 10% of the images with the presence of *Beggiatoa*.

From an environmental policy and regulatory perspective, the proposed system has direct implications for the management and governance of salmon aquaculture in Chile, where current regulatory frameworks remain constrained by limited monitoring capacity, particularly regarding benthic impacts and early-warning indicators of ecosystem degradation. In this context, the proposed automated detection approach is not intended to replace regulatory decision-making, but rather to strengthen the operational effectiveness, consistency, and scalability of existing inspection protocols. By reducing the time and cognitive burden associated with manual seabed video analysis, the system enables more frequent and standardized inspections under limited institutional resources, while preserving conservative safeguards through mandatory expert validation. More broadly, this study illustrates how applied machine learning can contribute to environmental governance when embedded within human-in-the-loop regulatory workflows that prioritize institutional constraints, regulatory logic, and risk-averse decision-making over purely predictive performance. In this sense, automated detection of benthic indicators such as Beggiatoa can support not only compliance monitoring, but also adaptive management strategies, including evaluations of nutrient loading, waste management practices, and the environmental feasibility of mitigation approaches such as integrated multi-trophic aquaculture.

As for future work, we consider a three-fold approach. First, the search for a better model able to detect the presence of *Beggiatoa* bacterial mats, incorporating new feature extraction characteristics and/or another type of model for classification, such as Vision Transformers. Second, the implementation of the current model for automated periodic reports preparation of *Beggiatoa* presence from videos in Sernapesca. Third, the creation of a video temporal segmentation algorithm (currently underway), based on the presence of *Beggiatoa*, using a rolling average technique in conjunction with the image classification obtained. Furthermore, as a long-term future development, this approach could be integrated into a broader framework that considers mathematical modeling and on-site analysis, being a step forward on the effective management of aquaculture. This comprehensive method has the potential to substantially improve the ability of various stakeholders, including producers and regulators, to identify suitable areas for sustainable aquacultural practices.

REFERENCES

[1] FAO. *Fisheries and Aquaculture Projections to 2030*. FAO, Rome, 2022.

[2] FAO. *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation*. FAO, Rome, 2022.

[3] David W Cole, Richard Cole, Steven J Gaydos, Jon Gray, Greg Hyland, Mark L Jacques, Nicole Powell-Dunford, Charu Sawhney, and William W Au. Aquaculture: Environmental, toxicological, and health issues. *International journal of hygiene and environmental health*, 212(4):369–377, 2009.

[4] Theodoros Mavraganis, Choremi Constantina, Markos Kolygas, Kosmas Vidalis, and Cosmas Nathanailides. Environmental issues of aquaculture development. *Egyptian Journal of Aquatic Biology and Fisheries*, 24(2):441–450, 2020.

[5] Joel K Elliott, Erin Spear, and Sandy Wyllie-Echeverria. Mats of beggiatoa bacteria reveal that organic pollution from lumber mills inhibits growth of zostera marina. *Marine Ecology*, 27(4):372–380, 2006.

[6] T. Pearson. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology: An Annual Review*, 16:229–311, 1978.

[7] T. Pearson and S. Stanley. Comparative measurement of the redox potential of marine sediments as a rapid means of assessing the effect of organic pollution. *Marine Biology*, 53(4):371–379, 1979.

[8] AJ Bale and AJ Kenny. Sediment analysis and seabed characterisation. *Methods for the study of marine benthos*, pages 43–86, 2005.

[9] D. Raffaelli. The behaviour of the nematode/copepod ratio in organic pollution studies. *Marine Environmental Research*, 23(2):135–152, 1987.

[10] D. Livingston, L. Förlin, and S. George. Molecular biomarkers and toxic consequences of impact by organic pollution in aquatic organisms. *Freshwater Bi-*

ological Association, 1994.

[11] M. Martínez-Alonso and N. Gaju. El papel de los tapetes microbianos en la biorrecuperación de zonas litorales sometidas a la contaminación por vertidos de petróleo. *Ecosistemas*, 14(2), 2005.

[12] E. Douglas, C. Pilditch, C. Kraan, L. Schipper, A. Lohrer, and S. Thrush. Macrofaunal functional diversity provides resilience to nutrient enrichment in coastal sediments. *Ecosystems*, 20(7):1324–1336, 2017.

[13] J. Harrison, R. Turner, L. Marques, and H. Ceri. Biopelículas. *Investigación y Ciencia*, page 77, 2006.

[14] J. Aguzzi, C. Costa, K. Robert, M. Matabos, F. Antonucci, S.K. Juniper, and P. Menesatti. Automated image analysis for the detection of benthic crustaceans and bacterial mat coverage using the venus undersea cabled network. *Sensors*, 11(11):10534–10556, 2011.

[15] M. Matabos, J. Aguzzi, K. Robert, C. Costa, P. Menesatti, J.B. Company, and S.K. Juniper. Multi-parametric study of behavioural modulation in demersal decapods at the venus cabled observatory in saanich inlet, british columbia, canada. *Journal of Experimental Marine Biology and Ecology*, 401(1-2):89–96, 2011.

[16] Jerónimo Pan. Microbes and marine sediments: A lifelong relationship on earth's biosphere. *Microbes: The Foundation Stone of the Biosphere*, pages 57–88, 2021.

[17] T. Pearson. The benthic ecology of loch linnhe and loch eil, a sea-loch system on the west coast of scotland. iv. changes in the benthic fauna attributable to organic enrichment. *Journal of Experimental Marine Biology and Ecology*, 20(1):1–41, 1975.

[18] C. Crawford, I. Mitchell, and C. Macleod. Video assessment of environmental impacts of salmon farms. *ICES Journal of Marine Science*, 58(2):445–452, 2001.

[19] R. Knight, J. Verhoeven, F. Salvo, D. Hamoutene, and S. Dufour. Validation of visual bacterial mat assessment at aquaculture sites through abiotic and biotic indicators. *Ecological Indicators*, 122:107283, 2021.

[20] Flora Salvo, Joseph Mersereau, Dounia Hamoutene, Rénald Belley, and Suzanne C Dufour. Spatial and temporal changes in epibenthic communities at deep, hard bottom aquaculture sites in newfoundland. *Ecological indicators*, 76:207–218, 2017.

[21] Anuarios estadísticos de pesca. Technical report, SERNAPESCA, 2018-2022.

[22] Chyrene Moncada, Christiane Hassenrück, Astrid Gärdes, and Cecilia Conaco. Microbial community composition of sediments influenced by intensive mariculture activity. *FEMS Microbiology Ecology*, 95(2):fiz006, 2019.

[23] Undersecretariat of Fisheries and Aquaculture. D.s. n° 320-2001 environmental regulation for aquaculture (updated d.s. n° 45-2021), 2021. Accessed: [10/04/2025].

[24] CIEP. Caracterización de condiciones ambientales, centros de cultivo categoría 4, región de aysén, 2010.

[25] C. Chailloux, A. Allais, P. Simeoni, and K. Olu. Automatic classification of deep benthic habitats: Detection of microbial mats and siboglinid polychaete fields from optical images on the håkon mosby mud volcano. In *Oceans*, pages 1–7. IEEE, 2008.

[26] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4489–4497, 2015.

[27] Zhong-Qiu Zhao, Peng Zheng, Shou-Tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.

[28] Thorsten Hoeser and Claudia Kuenzer. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sensing*, 12(10):1667, 2020.

[29] Licheng Jiao, Ruohan Zhang, Fang Liu, Shuyuan Yang, Biao Hou, Lingling Li, and Xu Tang. New generation deep learning for video object detection: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3195–3215, 2022.

[30] Mehul Mahrishi, Sudha Morwal, Abdul Wahab Muzaffar, Surbhi Bhatia, Pankaj Dadheech, and Mohammad Khalid Imam Rahmani. Video index point detection and extraction framework using custom yolov4 darknet object detection model. *IEEE Access*, 9:143378–143391, 2021.

[31] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning*, pages 689–696, 2011.

[32] Mohamed Othmani. A vehicle detection and tracking method for traffic video based on faster r-cnn. *Multimedia Tools and Applications*, pages 1–19, 2022.

[33] Khan Muhammad, Amin Ullah, Ali Shariq Imran, Muhammad Sajjad, Mustafa Servet Kiran, Giovanna Sannino, Victor Hugo C de Albuquerque, et al. Human action recognition using attention based lstm network with dilated cnn features. *Future Generation Computer Systems*, 125:820–830, 2021.

[34] K. Jerosch, A. Lüdtke, M. Schlüter, and G. Ioannidis. Automatic content-based analysis of georeferenced image data: Detection of beggiatoa mats in seafloor video mosaics from the håkon mosby mud volcano. *Computers & Geosciences*, 33(2):202–218, 2007.

[35] Yanyu Chen, Yunjue Zhou, Son Tran, Mira Park, Scott Hadley, Myriam Lacharite, and Quan Bai. A self-learning approach for beggiatoa coverage estimation in aquaculture. In *Australasian Joint Conference on Artificial Intelligence*, pages 405–416. Springer, 2022.

[36] Yanyu Chen, Yunjue Zhou, Mira Park, Son Tran, Scott Hadley, and Quan Bai. A novel adaptive ensemble learning framework for automated beggiatoa spp. coverage estimation. *Expert Systems with Applications*, 237:121416, 2024.

[37] Jos BTM Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, 41(1-2):187–228, 2000.

[38] Josef Kittler and John Illingworth. Relaxation labelling algorithms—a review. *Image and vision computing*, 3(4):206–216, 1985.

[39] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[40] Anastasiia Safonova, Gohar Ghazaryan, Stefan Stiller, Magdalena Main-Knorn, Claas Nendel, and Masahiro Ryo. Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 125:103569, 2023.

[41] A. Feelders, H. Daniels, and M. Holsheimer. Methodological and practical aspects of data mining. *Information & Management*, 37(5):271–281, 2000.

[42] Jihun Yoon and Min-Kook Choi. Exploring video frame redundancies for efficient data sampling and annotation in instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3308–3317, 2023.

[43] Yuan Zhi, Zhan Tong, Limin Wang, and Gangshan Wu. Mgsampler: An explainable sampling strategy for video action recognition. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1493–1502, 2021.

[44] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations*, pages 1–14, San Diego, CA, 2015. Computational and Biological Learning Society.

[45] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.

[46] Ching-Yao Chuang and Antonio Torralba. Estimating generalization under distribution shifts via domain-invariant representations. In *Proceedings of the International Conference on Machine Learning*, pages 1984–1994, 2020.

[47] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations*, 2017.

[48] Ming Sun, Xiaofen Yang, and Yinggang Xie. Deep learning in aquaculture: A review. *J. Comput*, 31(1):294–319, 2020.

[49] Daoliang Li and Ling Du. Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish. *Artificial Intelligence Review*, 55(5):4077–4116, 2022.

[50] An-Qi Wu, Ke-Lei Li, Zi-Yu Song, Xiuhua Lou, Pingfan Hu, Weijun Yang, and Rui-Feng Wang. Deep learning for sustainable aquaculture: Opportunities and challenges. *Sustainability*, 17(11):5084, 2025.

[51] Thurein Aung, Rafiza Abdul Razak, and Adibi Rahiman Bin Md Nor. Artificial intelligence methods used in various aquaculture applications: A systematic literature review. *Journal of the World Aquaculture Society*, 56(1):e13107, 2025.

[52] Sitao Liu, Zhuangzhuang Du, Guangxu Wang, Pan Zhang, Wenkai Xu, Jiaxuan Yu, and Daoliang Li. From traditional machine learning models to multimodal large models: A review of aquaculture. *Reviews in Aquaculture*, 18(1):e70111, 2026.

[53] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[54] Gabriela Csurka, editor. *Domain Adaptation in Computer Vision Applications*. Springer, 2017.

[55] Leiyu Chen, Shaobo Li, Qiang Bai, Jing Yang, Sanlong Jiang, and Yanming Miao. Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22):4712, 2021.

[56] Vagelis Plevris. Assessing uncertainty in image-based monitoring: addressing false positives, false negatives, and base rate bias in structural health evaluation. *Stochastic Environmental Research and Risk Assessment*, 39(3):959–972, 2025.

[57] Justine Boulent, Bertrand Charry, Malcolm McHugh Kennedy, Emily Tissier, Raina Fan, Marianne Marcoux, Cortney A Watt, and Antoine Gagné-Turcotte. Scaling whale monitoring using deep learning: A human-in-the-loop solution for analyzing aerial datasets. *Frontiers in Marine Science*, 10:1099479, 2023.

[58] Eugene Revelas, Brandon Sackmann, Aimee Thurlow, and Craig Jones. Mapping benthic habitat conditions and seafloor deposits using sediment profile imaging and a semiautomated image processing system. In *Offshore Technology Conference*, page D022S086R001. OTC, 2018.

[59] Alejandro H Buschmann, Felipe Cabello, Kyle Young, Juan Carvajal, Daniel A Varela, and Luis Henríquez. Salmon aquaculture and coastal ecosystem health in chile: analysis of regulations, environmental impacts and bioremediation systems. *Ocean & coastal management*, 52(5):243–249, 2009.

[60] Renato A Quiñones, Marcelo Fuentes, Rodrigo M Montes, Doris Soto, and Jorge León-Muñoz. Environmental issues in chilean salmon farming: a review. *Reviews in aquaculture*, 11(2):375–402, 2019.

[61] Moisés A Aguilera, Jaime A Aburto, Luis Bravo, Bernardo R Broitman, Rafael A García, Carlos F Gaymer, Stefan Gelcich, Boris A López, Vivian Montecino, Aníbal Pauchard, et al. Chile: environmental status and future perspectives. *World seas: An environmental evaluation*, pages 673–702, 2019.

[62] Alejandro H Buschmann, Edwin J Niklitschek, and Sandra V Pereda. Aquaculture and its impacts on the conservation of chilean patagonia. In *Conservation in Chilean*

Patagonia: Assessing the State of Knowledge, Opportunities, and Challenges, pages 303–320. Springer, New York, 2024.

NICOLÁS MARTINEZ ALFARO earned his Bachelor's Degree in Engineering Sciences in 2021 and a Master's Degree in Data Science in 2024, both from Universidad Adolfo Ibáñez, Chile. He graduated first in his class and received his Master's degree with honors. He was awarded the Pedro Luis González Scholarship, and the Honor Roll Scholarship during his undergraduate program.

He is currently working as a Software Engineer at Buk, Chile. His professional and academic interests include data science, machine learning, and applied computer vision, with a focus on developing intelligent systems and data-driven solutions.

SEBASTIÁN MORENO was born in Santiago Chile on 1982. He received the B.Sc. degree in Informatics Engineering and the M.Sc. degree in Computer Engineering from Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile, in 2004 and 2008, and his Master of Science and PhD in Computer Science from Purdue University, USA, in 2011 and 2014.

From 2022, he is an associate professor at Universidad Adolfo Ibáñez. Sebastián has been Associate director of the postgraduate academic programs (2021-2024), head of the Master of Science in Data Science (2022-2024), Master of Science in Engineering (2021-2022), and Computer Engineering (2020-2021).

Dr. Moreno research interests include machine learning, deep learning, and transfer learning. Thanks to his work, Dr. Moreno has graduated more than 15 master students, published more than 40 papers, and worked on more than 18 projects. He is also an active program committee member of the most important conferences in the machine learning area including NeurIPS, KDD, WWW, ICDM, and others.

FRANCISCO J. PLAZA-VEGA was born in Valparaíso, Chile, on June 30, 1982. He received the B.Sc. degree in Fisheries Engineering from the Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, in 2008, the M.Sc. degree in Statistics from the University of Valparaíso in 2017, the Big Data and Data Science Diploma from the Pontificia Universidad Católica de Valparaíso in 2017, and the Ph.D. degree in Statistics from the University of Valparaíso in 2023.

He is currently an Associate Professor with the Department of Mathematics and Computer Science at the Universidad de Santiago de Chile. His research interests include time series analysis, machine learning, deep learning, generative AI, stochastic processes, and statistical modeling of environmental, fisheries, and seismic systems. He has authored numerous peer-reviewed articles in journals such as Progress in Oceanography, Fisheries Research, Spatial Statistics, Communications in Statistics, and the Latin American Journal of Aquatic Research.

JOSÉ-EMILIO CASTILLO received the M.Sc. in Engineering Sciences and the Industrial Civil Engineering degree from Universidad Adolfo Ibáñez, Santiago, Chile, in 2023.

He is currently a Product Onboarding Specialist at Keirón, a B2B healthtech company, where he leads client integrations and platform adaptation processes to ensure optimal implementation and user experience. His professional background combines technical and business expertise, with previous roles in growth, sales operations, and analytics. His interests include data-driven decision-making, system integration, and the application of machine learning and analytics to improve business performance.

MIGUEL CARRASCO earned a Ph.D. with honors in informatics from the Institut des Systèmes Intelligents et de Robotique (ISIR), Pierre and Marie Curie University - Paris 6, in 2010. He also holds a Ph.D. in Engineering Sciences with a focus on Computer Science from the Department of Computer Science at Pontificia Universidad Católica de Chile, earned through a joint supervision program.

He is currently an Associate Professor at the Escuela de Ingeniería Informática y Telecomunicaciones, Universidad Diego Portales (UDP). His research interests include developing automatic algorithms based on image processing and computer vision. His primary research areas include image processing for biological applications (such as virus and pollen segmentation), computer vision for industrial inspection, the Industrial Internet of Things (IIoT), and failure prediction using pattern recognition. His past and current research includes human-computer interaction, users' gaze gesture prediction with eye trackers, automatic multiple visual inspection, and telemedicine using IoT (heart sensors). In 2007, he was awarded a scholarship from the Collège Doctoral Franco-Chilien.