Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Automation of the Van de Casteele Test via Deep Learning: Feasibility and Analysis of Instantaneous Sampling

Automação do Teste de Van de Casteele via Aprendizado Profundo: Viabilidade e Análise da Amostragem Instantânea

Salomão Soares [1], Regiane Dalazoana [2], Mariana Eiko Borba Inoue [3], Everton Gomes dos Santos [4] e Luiz Fernando Reinheimer [5]

1 Federal University of Paraná, Curitiba, Brazil. E-mail: salomao.soares@ufpr.br.
ORCID: https://orcid.org/0000-0003-4826-776X
2 Federal University of Paraná, Curitiba, Brazil. E-mail: regiane@ufpr.br.
ORCID: https://orcid.org/0000-0001-5468-0679
3 Federal University of Paraná, Curitiba, Brazil. E-mail: mariana.inoue@ufpr.br.
ORCID: https://orcid.org/0000-0002-2008-3769
4 Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil. E-mail: everton.santos@ibge.gov.br.
ORCID: https://orcid.org/0000-0002-8730-7755
5 Brazilian Institute of Geography and Statistics, Florianópolis, Brazil. E-mail: luiz.reinheimer@ibge.gov.br.
ORCID: https://orcid.org/0009-0005-6858-6604

Received: 02.2026 | Accepted: 06.2026

Abstract: The Van de Casteele Test is an important step in analyzing the quality of observations from tide gauge stations, but its manual execution is laborious. To automate this process, this research evaluated the use of a deep learning model based on YOLO, applied to images of the tide staff at the Arraial do Cabo tide gauge station, captured by low-cost digital cameras. The study also analyzed the limits of instantaneous inference, characterized by direct image measurement without the common time-averaging filters found in electronic sensors. In tide staff detection, the model showed excellent performance across various scenarios, with mAP50 of 0.995 and mAP50-95 ranging from 0.844 to 0.913. The metrological validation compared five experiments, using the conventional method as a reference. Regarding accuracy, while the Radar and Encoder sensors showed standard deviations (SD) of 0.02 m when compared to tide staff readings taken by an observer, the computer vision model converged to a SD of 0.03 m throughout the experiments. This slightly higher variability, together with an identified systematic bias, is attributed to the instantaneous capture of the sea surface, which records high-frequency noise (waves) that is filtered out by the temporal integration of level sensors and by the observer. Therefore, the results demonstrate that, given the instantaneous nature of image capture, the technique is robust and feasible for automating the Van de Casteele Test.
Keywords: Van de Casteele Test. Deep Learning. Tide Gauge. Tide Staff.

Resumo: O Teste de Van de Casteele é uma etapa importante na análise da qualidade das observações das estações maregráficas, mas sua execução manual é laboriosa. Para automatizar esse processo, esta pesquisa avaliou o uso de um modelo de aprendizado profundo baseado em YOLO, aplicado a imagens da régua de marés da estação maregráfica de Arraial do Cabo, capturadas por câmeras digitais de baixo custo. O estudo também analisou os limites da inferência instantânea, caracterizada pela medição direta de imagens, sem os filtros de média temporal comuns em sensores eletrônicos. Na detecção da régua de marés, o modelo apresentou excelente desempenho em vários cenários, com mAP50 de 0,995 e mAP50-95 variando de 0,844 a 0,913. A validação metrológica comparou cinco experimentos, tendo como referência o método convencional. Em relação à precisão, enquanto os sensores Radar e Encoder apresentaram desvios padrão (DP) de 0,02 m quando comparados aos registros da régua de marés feitos por um observador, o modelo de visão computacional convergiu para um DP de 0,03 m ao longo dos experimentos. Essa variabilidade ligeiramente maior, juntamente com um viés sistemático identificado, é atribuída à captura instantânea da superfície do mar, que registra ruídos de alta frequência (ondas) que são filtrados pela integração temporal dos sensores de nível e pelo observador. Portanto, os resultados demonstram que, considerando a natureza instantânea da captura de imagem, a técnica é robusta e viável para automatizar o Teste de Van de Casteele.
Palavras-chave: Teste de Van de Casteele. Aprendizado Profundo. Marégrafo. Régua de Marés.

1 INTRODUCTION

Since the classical definition by Gauss and Listing, the determination of sea level variations, reduced by averaging over a given period and location, has served as the starting point for countries' vertical control networks. However, these local reference surfaces only approximate the geoid or quasigeoid, due to the effect of sea surface topography and local sea level anomalies, with deviations that can reach ± 2 m relative to the global geoid (Gemael, 2019; Torge & Müller, 2012; Sánchez & Sideris, 2017).

The current understanding is that the mean sea level is not an equipotential surface of the Earth's gravity field (Torge & Müller, 2012). Still, monitoring mean sea level remains important for a wide range of scientific and operational domains. Examples include applications in traditional hydrographic work, tide forecasting, long-term sea level trends, and calibration of Satellite Altimetry (ALTSAT) (CHM, 2025; IOC, 2024; Birol et al., 2025).

According to the Intergovernmental Oceanographic Commission (IOC, 2024), sea level variability at most locations, with amplitudes typically on the order of decimeters at a decadal scale, does not allow the detection of secular trends, even with more precise sensors, without at least 50 years of data. Consequently, programs such as the Global Sea Level Observing System (GLOSS) emphasize long-term measurements with a precision of 1 cm (UNESCO, 2002). This rigorous requirement highlights the need for highly precise and reliable data, which are essential to the growing demands of contemporary science and operations.

In this context of precise data, the Van de Casteele Test (VCT), conceived for quality control and error characterization in analog tide gauges, is still used in the era of digital tide gauges (Miguez et al., 2008), remaining a fundamental step in the installation and monitoring of tide gauge stations. It is characterized by successive and simultaneous readings of the water level from the tide gauge sensor and an independent reference instrument over a complete tidal cycle (13 h). The differences between the readings of the two sensors should remain constant; however, this only occurs with a perfect tide gauge. The VCT helps determine the accuracy of the tide gauge by constructing a diagram whose shape allows different types of faults to be identified. For more details, see UNESCO (1985). When a complete collection is not possible, observations carried out over a short period of time, around high or low tide, preferably during spring tides, also allow adequate analyses (UNESCO, 2016). In addition, through the VCT, it is possible to determine the relationship between the different instrumental references (tide staff and sensor zeros), which helps correlate reference levels (Dalazoana et al., 2005).

Regular application is recommended, especially after updates or sensor modernizations, which underscores its relevance as a fundamental quality control tool. Nevertheless, the VCT remains a manual procedure performed in situ, making it a laborious and time-consuming step (Miguez et al., 2008; SONEL, 2019).

There are several possible electronic devices that can serve as independent references, ranging from another tide gauge sensor to electric probes and total stations using the trigonometric method (Miguez et al., 2008; SONEL, 2019; Silveira, 2023). The Brazilian Institute of Geography and Statistics (IBGE) uses the tide staff in the VCT because it allows water-level observation independent of any power source, recording system, or filtering mechanism, thereby serving as the main element for calibrating tide-gauge sensor observations (IBGE, 2010).

Although the VCT is fundamental for calibrating level sensors, its conventional application may be approaching its limits, as tide gauge networks modernize and expand globally, reliance on manual processes becomes a limiting factor for efficient, widespread quality control. In this regard, and according to LeCun, Bengio, and Hinton (2015), modern computer vision has evolved significantly with deep learning, enabling convolutional neural networks to automatically learn hierarchical representations directly from visual data. For example, Lima (2019) used plastic rulers to identify the advance or retreat of cracks in buildings, achieving 75.65% accuracy; Freitas (2022) sought to identify defects in road pavements, achieving accuracy of 99.89% and 95.91% for potholes and patches, respectively.

However, computer vision for reading physical tide staff is still little explored in the literature. This research innovates by integrating low-cost hardware and a single-stage detection architecture to replicate human perception in the VCT. Notably, this article is an extended version of Soares et al. (2025), presented at the I CIRF & VIII SIMGEO – Full Land Regularization & Geotechnologies. It specifically analyzes the potential of deep learning for target detection, assisting the computer vision algorithm in determining metric water levels from images captured by digital cameras. The hypothesis is that automating this workflow can reduce VCT costs, allow more frequent measurements, and help prevent late identification of instrumental faults.

This study uses images of the tide staff at the Arraial do Cabo-RJ tide gauge station. Digital cameras integrated with ESP32-CAM modules capture images at 2- and 5-minute intervals. These images evaluate four training instances based on the YOLOv8 (You Only Look Once) architecture. Inference consists of detecting the tide staff with YOLOv8 and generating bounding boxes used to extract metric attributes. A post-processing algorithm converts the pixel coordinates of the bounding box's lower edge into metric values (meters) using a second-degree polynomial transfer function. Five experiments analyzed inference performance using instantaneous images. Each measurement is processed in isolation, without temporal filtering or smoothing, to rigorously test the optical technique against the natural dynamics of sea level.

2 METHODOLOGY

This section presents the location and characterization of the tide gauge station, the materials used to obtain the images, the preparation of the dataset, and the fundamentals of deep learning with YOLOv8. The code was developed in Python for direct use in Google Colab (GOOGLE, 2026), using the hardware accelerator, the Tensor Processing Unit (TPU v2-8), during training. The adoption of this infrastructure aimed to optimize computational processing and reduce the convergence time of the four trained models. It should be noted that after this stage, the resulting weights do not require a dedicated hardware accelerator and can be used on conventional Central Processing Units (CPUs) to perform inference. This characteristic gives the method portability, enabling its application in computing environments with lower processing capacity or on edge devices.

2.1 Location and characterization of the Arraial do Cabo - RJ tide gauge station

The tide gauge station used in this research is part of the Permanent Tide Gauge Network for Geodesy (RMPG) and is located in the municipality of Arraial do Cabo, in the State of Rio de Janeiro (22°58'21"S 42°00'48"W, SIRGAS2000), code EMARC, more specifically on the premises of Porto do Forno, managed by the Municipal Port Administration Company (COMAP) and operated by IBGE since 2017. Figure 1 shows the location of EMARC.

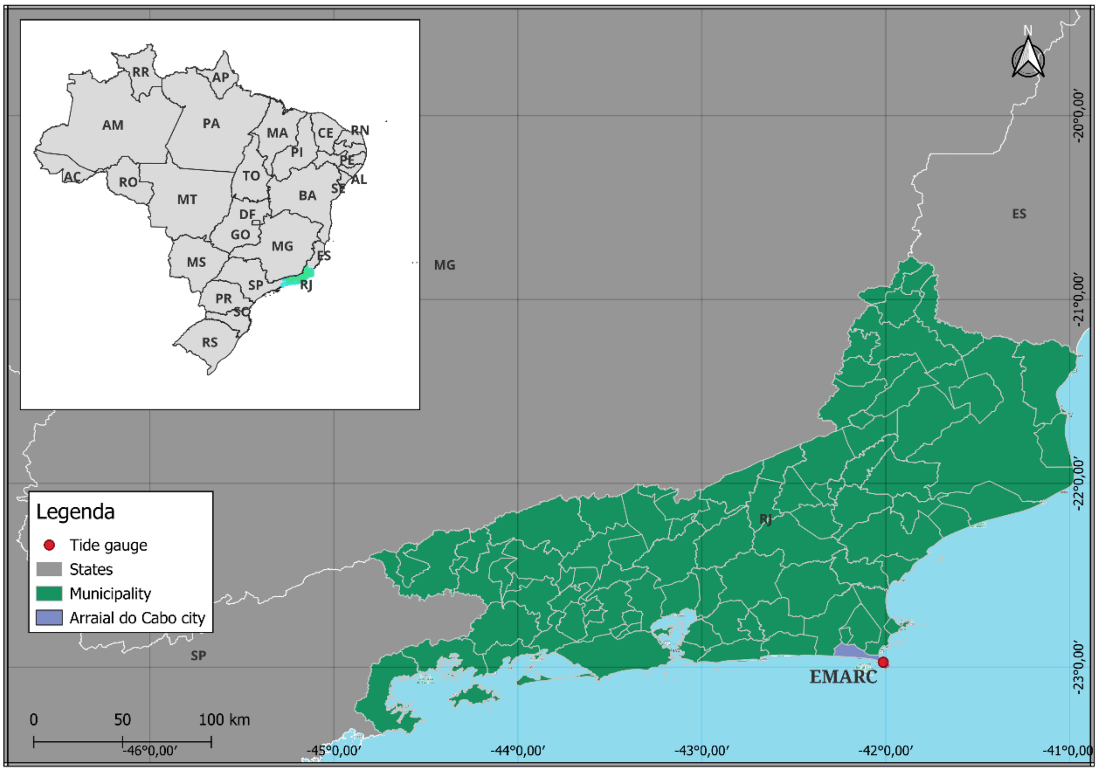
The system consists of two-level sensors: a Radar sensor, which performs measurements without contact with the water through electromagnetic waves, and an electronic shaft encoder (Encoder), which operates according to the mechanical principle of counterweight and float. In addition, it has auxiliary meteorological sensors and a tide staff (IBGE, 2021). Figure 2 shows the sensors and structures that make up the station.

2.2 Data collection

To prepare the dataset, between March 26 and 27, 2025, during a preventive inspection, two independent capture units were installed. Each unit consisted of an assembly comprising an ESP32-CAM microcontroller module, an integrated camera lens, and a 32 GB microSD card. These systems were housed in hermetic enclosures and powered by an off-grid photovoltaic system comprising a 60 W solar panel, a 30A solar charge controller for 12 V systems, and two stationary 12 V 7 Ah batteries connected in parallel. This configuration ensured energy autonomy and operational stability during the monitoring campaign. The units captured images of the tide staff at 2- and 5-minute intervals, respectively. Figure 3 presents details of the installation. The capture automation software was developed specifically for this project and loaded onto the microcontroller. At this stage of the research, the module's Wi-Fi interface was not used for data transmission, with local storage prioritized on the microSD card. On the 27th, during the spring tide, the VCT was also conducted in its conventional form, with the observer taking water level readings on the tide staff every 15 minutes between 06:00 and 19:00.

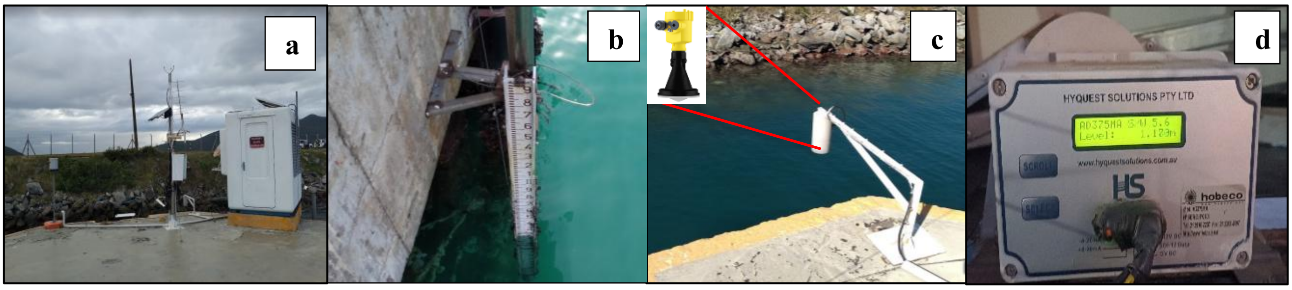
DOI: http://dx.doi.org/10.14393/rbcv78n0a-84090

Figure 1 – Location of the RMPG's Arraial do Cabo-RJ tide gauge station.



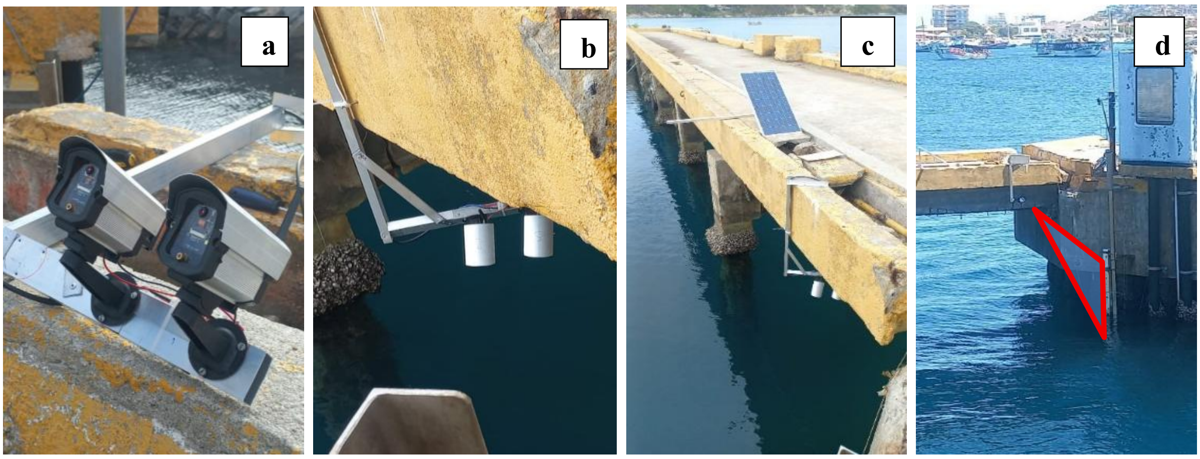
Source: The Authors (2026).

Figure 2 – a) General view of the sensors available at the tide gauge station. b) Location of the tide staff fixed on the side of the pier. c) Radar sensor. d) Encoder sensor.



Source: The Authors (2026).

Figure 3 – a) ESP32-CAM cameras. b) Detail of the camera installation. c) Installation site of the cameras and solar panel. d) Panoramic view of the image capture setup, highlighted in red.



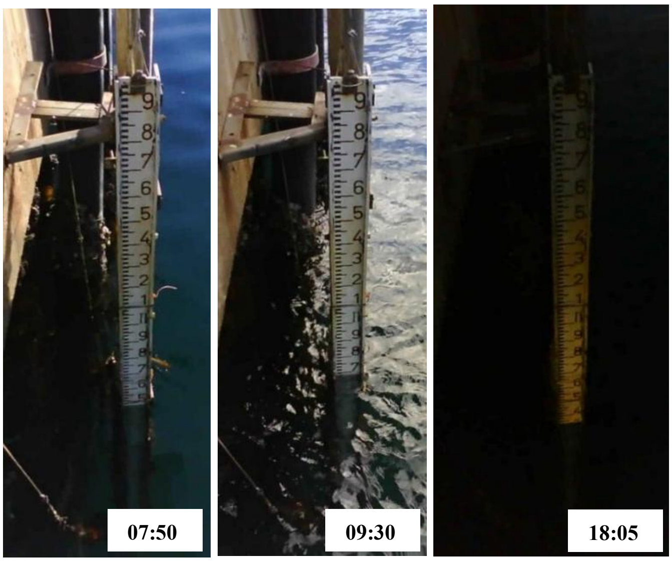
Source: The Authors (2026).

2.3 Dataset preparation

The images were inspected to verify quality, and frames with severe obstructions or lighting failures were discarded. They were then uploaded to the open-source web application Make Sense (Skalski, 2022) for manual bounding-box determination of the tide staff, with the aim of assembling the training and validation datasets. At this stage of the process, the goal was to ensure the greatest possible rigor in determining the upper and lower parts of the bounding box. This is a necessary condition to avoid false positives during the training phase, since the clear water in the region allowed the submerged initial parts to show through, which could lead to incorrect interpretations.

In addition, the image captures were carried out under different lighting conditions, which added complexity to the process. According to Del Rosso et al. (2021), a training dataset must be representative, covering all the features and phenomena to be identified, mapped, or monitored. Figure 4 presents three sample images obtained under different lighting levels.

Figure 4 – Image capture of the tide staff at different times.



Source: The Authors (2026).

The images were then divided into 3 sets: training, validation, and testing. For the training set, additional care was taken to separate images representing the widest possible range of conditions, that is, greater variability in the moments of high and low luminosity and of high and low tides, to build a clear understanding of the tide staff as the target element in the learning process. For the test and validation sets, the data were randomly split, with most allocated to the test set. This split prioritized a robust test set to ensure an external evaluation of the model's generalization capacity. The detailed organization of the sets is described in Table 1.

Table 1 – Number of digital images used in each stage of the detection process.

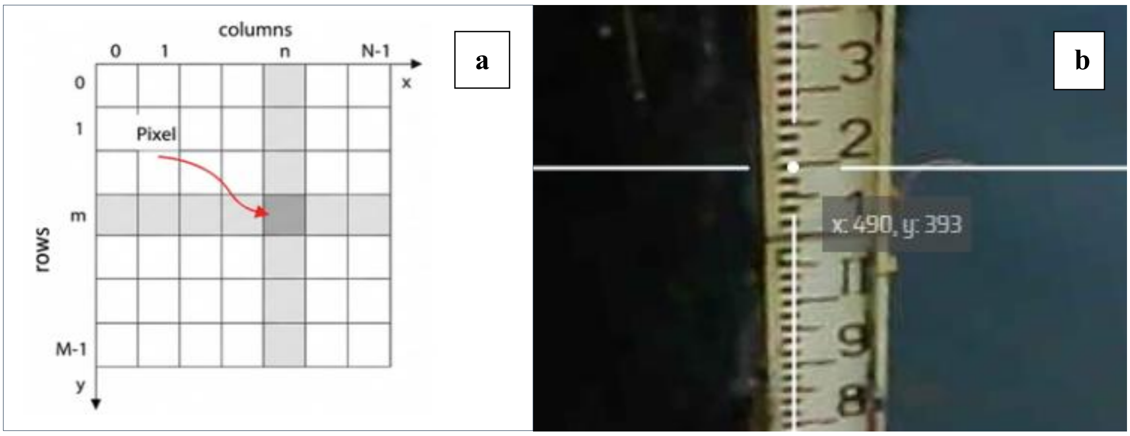
Set	Quantity	Percentage (%)
Training	72	30.2
Validation	41	17.2
Test	125	52.6
Total	238	100

Source: Adapted from Soares et al. (2025).

2.4 Determination of the value conversion function

Also, as part of data preparation, metric water level values from the tide staff were extracted for comparison with the deep learning results. The extraction of these values in the office involved a specialist visually reviewing the images, serving as the “ground truth” to validate the algorithm's accuracy. The objective was to ensure equivalent observation conditions, since in the conventional method the observer determines an average value from measurements taken over a time interval during which sea level varies continuously. Figure 5 presents the representation of a digital image and the pixel/meter relationship on the tide staff, in which, for example, the level of 2.20 m corresponds to pixel x, y (490, 393) in the digital image.

Figure 5 – a) Representation of a digital image; b) Determination of pixel coordinates on the tide staff.



Source: a) Adapted from Jähne (2005); b) The Authors (2026).

The bounding-box identification method facilitates, in the specific case of the tide staff, the use of the lower image marks, provided that the captures are strictly vertical, as in the case analyzed. When installing the digital camera, the angle and distance values were unknown. Thus, based on the principle that the tide staff and the camera are fixed, an attempt was made to determine the function that best represented the staff's constants in the images. The horizontal coordinate (x) was disregarded, since the variable of interest for the tidal cycle is strictly vertical. In all training images, the pixel values of the corresponding metric values on the tide staff were extracted, with the average of the y-values determined ($\bar{y}$), as shown in Table 2.

Analysis of the position of the values reveals the nonlinear relationship between pixels and meters. In addition, it was necessary to determine a function with extrapolation capability, since water levels could occasionally fall below the calibrated lower limit. In this research, a second-degree polynomial function was used to represent the dataset, as in Eq. (1):

$$f(x) = ax^2 + bx + c \tag{1}$$

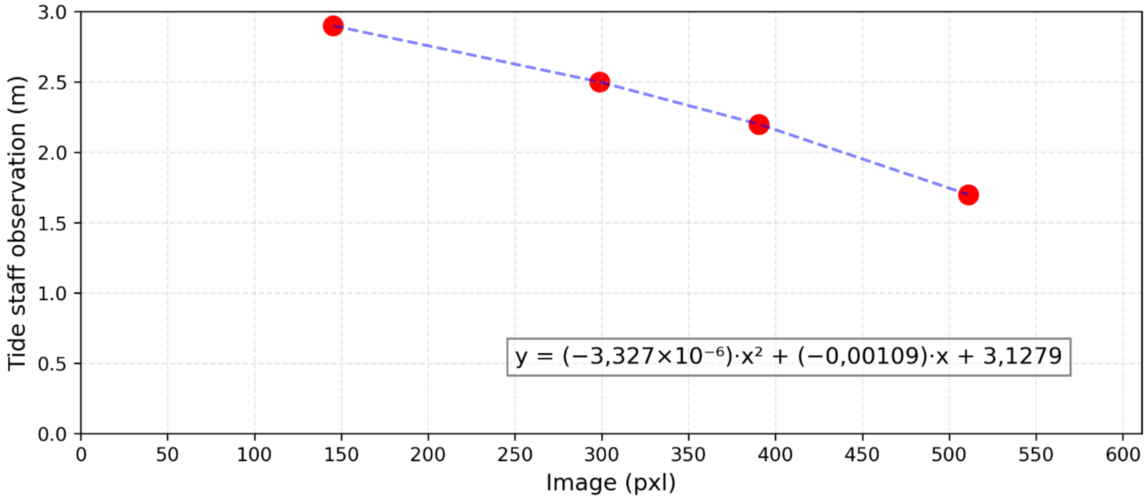
in which a, b and c are the coefficients (constants) and x represents the vertical pixel coordinate of the lower edge of the bounding box detected by the model for each image. Table 2 presents the relationship between the metric value on the tide staff and the corresponding pixel in the image, and Figure 6 presents the conversion function resulting from these values.

Table 2 – Scale relationship between the tide staff and the camera 1 images.

Tide staff (m)	**$\bar{y}$ (pxl)**
2.90	145.38
2.50	298.72
2.20	390.43
1.70	510.94

Source: Soares et al. (2025).

Figure 6 – Determination of the function for converting pixels to meters in camera 1 images.



Source: Adapted from Soares et al. (2025).

To clarify the processing flow and illustrate how algorithmic detection is converted into a physical quantity, a numerical example of the transfer function's application is presented below.

The inference of the YOLOv8n architecture provides as output the location of the object in the form of a coordinate vector $[x_1, y_1, x_2, y_2]$ in pixels. The variable of interest for this research is y_2 , which corresponds to the lower edge of the bounding box, positioned strictly at the interface between the tide staff and the water surface. Taking as an example an image in which the model detected the target with the lower vertical coordinate $y_2 = 401$ pixels, this value is isolated and applied as the independent variable y_2 in Eq. (1), using the coefficients determined by the pixel-to-metric conversion function for camera 1, as in Eq. (2) and Eq. (3):

$$y = (-3.327 \times 10^{-6}) \cdot (401)^2 + (-0.00109) \cdot (401) + 3.1279 \quad (2)$$

$$y \approx 2.16\, m \quad (3)$$

Solving the steps of the second-degree polynomial, the post-processing result indicates that the coordinate of 401 pixels in the digital image corresponds to a water level of 2.16 m on the physical scale of the tide staff. This value, estimated by computational inference, is then directly compared with the observer's reading and the records of the physical sensors (Radar and Encoder) in the VCT.

2.5 Deep learning with YOLO

Deep learning has become the most widely discussed technology today due to its results in image classification, object detection, and language processing (Srivastava et al., 2020). In this research, the detection of the tide staff plays a fundamental role and can be performed by two types of detectors: single-stage and two-stage.

Two-stage object detectors first identify regions of interest and then classify objects within them. Popular examples include the Region-based Convolutional Neural Network (RCNN). Single-stage detectors, on the other hand, construct bounding boxes and classify objects simultaneously using a single feed-forward neural network. YOLO and SSD (Single Shot MultiBox Detector) are well-established examples of this category (Taye, 2023; Zeng & Zhong, 2024).

In this research, the nano (n) variant of YOLOv8 was adopted (ULTRALYTICS, 2023). For the development of the models, the Transfer Learning technique was used (*Transfer Learning*), starting training from weights pre-trained on the COCO dataset (Lin et al., 2014). This approach allows the network to leverage basic visual features it has already learned (such as edges, shapes, and textures), thereby accelerating convergence toward the specific object of this research: the tide staff. According to Terven et al. (2023), the YOLO architecture stands out for its remarkable balance between speed and accuracy, enabling fast, reliable object identification in images.

Although training was carried out in a cloud environment (Google Colab with a TPU v2-8 hardware accelerator) to optimize processing time, the choice of the nano variant is justified by portability. The objective is to ensure that the resulting model can be implemented in the future on edge devices with limited computational resources, while maintaining low memory consumption and high real-time response speed. The main hyperparameters configured included resizing the images to 768 x 1024 pixels and running 50 or 100 epochs, depending on the experimental scenario.

2.5.1 CONFUSION MATRIX

A confusion matrix compares the actual class label with the predicted class label using test images. It is a table that presents the results of a classification model, as shown in Table 3.

Table 3 – Confusion matrix.

Predicted	Actual Positive	Actual Negative
Positive	TP	FP
Negative	FN	TN

Source: Soares et al. (2025).

Each cell contains the values for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), and it is frequently used to evaluate the performance of detection and classification models.

2.5.2 TRAINING AND VALIDATION ERROR PARAMETER

When using the YOLOv8 architecture, different loss components are monitored to assess learning performance (loss). Box loss measures the error in the prediction and localization of the boxes relative to the actual labels (ground truth). Classification loss (cls loss) represents the error in identifying the object's class. Distribution focal loss (dfl loss) refines the box geometry by modeling positional uncertainty with greater precision. These metrics are calculated for both the training data (train) and the validation data (val), enabling diagnosis of the model's fit.

Underfitting occurs when the model is too simple, or training is insufficient, resulting in high errors in both sets (training and validation). Overfitting, on the other hand, is characterized by a continuous reduction in training loss while validation loss stabilizes or increases. It indicates that the model has "memorized" the training data but fails to handle unseen data. For an ideal fit, the training and validation loss curves should converge and decrease smoothly, stabilizing at low, similar values (Mucci, 2024). The similarity in behavior between the two graphs is the main indication that the model has adequately generalized the visual features of the tide staff, which ensures the robustness of the detections.

2.5.3 EVALUATION METRICS

In object detection tasks, the following metrics are commonly used to evaluate the relationship between the model's predictions and the actual labels, providing a basis for performance evaluation and algorithm optimization (Terven et al., 2023; Zeng & Zhong, 2024; ULTRALYTICS, 2023).

Intersection over Union (IoU) is a fundamental metric in object detection, which quantifies the overlap between the predicted bounding boxes (B_p) and the actual ones (B_{gt}). It plays a fundamental role in evaluating detection accuracy. Mathematically, IoU can be defined as in Eq. (4):

$$IoU = \frac{Area\left(B_p \cap B_{gt}\right)}{Area\left(B_p \cup B_{gt}\right)} = \frac{Intersection\ Area}{Union\ \text{Area}} \tag{4}$$

Precision quantifies the proportion of true positive (TP) detections in relation to the total number of detections (TP + FP), that is, it measures how many of the positive detections are correct, as defined by Eq. (5):

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Recall quantifies the model's ability to correctly identify all the target objects present in the images, as defined by Eq. (6):

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

The Mean Average Precision (mAP) is a global metric that synthesizes precision and recall indices across multiple confidence thresholds. Its calculation is based on Average Precision (AP), which mathematically corresponds to the area under the precision-recall curve for each category, individually, as defined by Eq. (7):

$$AP = \sum_{k=1}^{m} (\mathrm{Recall}_k - \mathrm{Recall}_{k-1}) \cdot \mathrm{Precision}_k \tag{7}$$

in which m represents the total number of confidence thresholds evaluated along the curve, while $\mathrm{Precision}_k$ and Recall_k correspond to the precision and recall values obtained at the k-th threshold. In Eq. (8), the parameter n denotes the total number of classes evaluated in the dataset, taking the value of $n = 1$ in this research, since it strictly involves a single target (the tide staff):

$$mAP \ = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{8}$$

The mAP50 metric indicates the mean average precision calculated with an Intersection over Union (IoU) threshold of 0.5, that is, detection is considered a true positive when the overlap between the predicted and actual bounding boxes is at least 50%. mAP50-95, in turn, is a more rigorous parameter, representing the average of the mAP calculated over IoU intervals from 0.5 to 0.95, in increments of 0.05.

2.6 Statistical evaluation

The determination of basic statistics — mean, standard error, median, standard deviation, minimum, and maximum — as well as the 95% confidence interval, was used to present a summary of the results obtained from the office-based images and those obtained automatically through deep learning. To assess whether the difference between the observed means of the two samples was statistically significant, Student's t-test for independent samples was performed. For this purpose, the standard error of the difference between the two means (SE_d) was calculated, as defined by Eq. (9):

$$SE_d = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{9}$$

where EP_d is the standard error of the difference between the two means, S_1^2 and S_2^2 are the variances of the samples, and n_1 and n_2 are the sample sizes, represented by the number of images analyzed in the test phase.

Next, Student's t-test for independent samples was determined using Eq. (10):

$$t = \frac{|\overline{x}_1 - \overline{x}_2|}{SE_d} \tag{10}$$

in which t is Student's t statistic, $\overline{x}_1$ and $\overline{x}_2$ are the sample means, and SE_d is the standard error of the difference.

The calculation of the degrees of freedom (DF) is the final step to determine the critical value for accepting or rejecting the null hypothesis for a two-tailed 95% confidence interval, as in Eq. (11):

$$DF = n_1 + n_2 - 2 \tag{11}$$

where DF is the degrees of freedom, and $n_1$and n_2 are the sample sizes. For a calculated Student's t value lower than the critical value obtained from the Student distribution percentile table, the null hypothesis (H_0) is accepted; otherwise, H_0 is rejected. For this research, it is expected, as H_0, that there are no significant differences between the samples.

2.7 Experiments

To evaluate the robustness of the proposed model and explore the limits of inference on instantaneous images under different temporal sampling conditions, five distinct experiments were planned. The experimental strategy sought to evaluate both the generalization capacity of the model trained with 5-minute images and the accuracy of models trained only with 2-minute images.

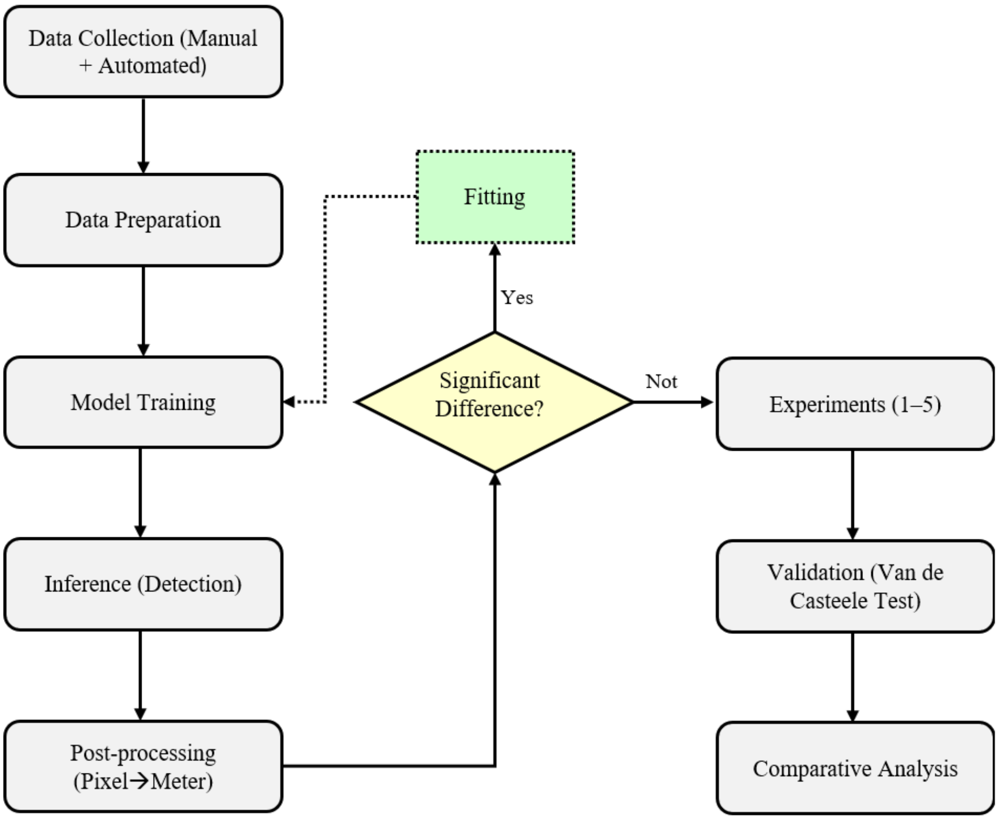
The datasets were segmented into training, validation, and testing subsets. In addition to conventional deep learning metrics (mAP), a specific subset of the test images was submitted to a rigorous metrological validation through the VCT. This was done using only data from March 27, when simultaneous control with manual tide staff readings was in place.

2.7.1 DESCRIPTION OF THE EXPERIMENTAL SCENARIOS

a) Experiment 1: This scenario used images captured at regular 5-minute intervals, collected on March 26 and 27, 2025. The model underwent training for 50 epochs. The objective was to compare automatic readings with manual readings taken in the office. For the Van de Casteele analysis, 86 images from the test set referring to the 27th were selected. The validation of this initial scenario is established as a methodological prerequisite, defining the performance standard against which the sampling and generalization variations of the subsequent experiments will be compared.
b) Experiment 2: To test the model's generalization capacity in the face of a change in the image capture interval, the weights trained in Experiment 1 were used directly on a previously unseen test set of 576 2-minute images (Camera 2, days 26 and 27). The metrological validation used 363 of these images (day 27) to verify whether the learning held up under faster captures.
c) Experiment 3: An attempt was made to increase the model's diversity by mixing data from cameras 1 and 2, with image captures at 2- and 5-minute intervals, both in training and validation. Training was extended to 100 epochs to accommodate greater data heterogeneity. The VCT was applied specifically to 317 2-minute images from the 27th.
d) Experiment 4: In this experiment, the temporal variable was isolated, using only 2-minute interval images from the 26th and 27th in all stages (training, validation, and testing). With 100 training epochs, the objective was to verify whether specialization in the 2-minute interval would result in superior performance compared to generalization or to mixed data. The Van de Casteele Test was performed with 226 images from the 27th.
e) Experiment 5: The most restrictive scenario used only 2-minute interval images, from the 27th itself, for training, validation, and testing, over 100 epochs. Although it represents a smaller data volume, it allowed analysis of the model's behavior during training and testing under identical environmental conditions. The 67 images in the test set were used in the metrological analysis.

Figure 7 illustrates the adopted methodological flow. Table 4 presents the experimental scenarios outlined and quantifies the subset of images submitted to VCT validation.

Figure 7 – Methodological flowchart, encompassing manual and automated data collection, model preparation and training, inference and post-processing, followed by evaluation of significant differences, iterative hyperparameter adjustments (such as changing the image size or learning rate, if the initial results were unsatisfactory), controlled experiments, metrological validation via the Van de Casteele Test, and comparative analysis of the results.



Source: The Authors (2026).

Table 4 – Summary of the experimental scenarios

Experiment	Training (no. of images)	Validation (no. of images)	Test (no. of images)	Total (no. of images)	Training/Validation Frequency (minute)	Test Frequency (minute)	Epochs (unit)	VCT Sample (no. of images)
1	72	41	125	238	5	5 min	50	86
2	(Exp. 1 weights)	(Exp. 1 weights)	576	689	(Same as Exp. 1)	2 min	—	363
3	388	96	317	801	2 and 5	2 min	100	317
4	280	70	226	576	2	2 min	100	226
5	226	70	67	363	2	2 min	100	67

Source: The Authors (2026).

3 RESULTS AND DISCUSSION

The presentation of the results occurs in two phases: first, the effectiveness of the proposed method (detection and metric conversion) is examined in the baseline scenario; then, the analysis is broadened to compare the various experimental scenarios and the conventional physical sensors.

3.1 Methodological analysis

The evaluation of the confusion matrix marked the beginning of the analysis of training results. Of the 41 images provided for training validation, the matrix indicated 41 true positives, with no false positives or false negatives. In other words, there were no cases in which the background (seawater and local structures) was incorrectly classified as the tide staff, nor vice versa. Therefore, as shown in Table 5, the tide-staff class was correctly classified. It should be noted that the 100% accuracy rate in the initial validation results from the operational characteristics of a fixed-purpose monitoring station. Since the camera and the tide staff remain

 DOI: http://dx.doi.org/10.14393/rbcv78n0a-84090

static, the background's geometric signature does not change, reducing the occurrence of local false positives. However, the detector's perfect performance does not imply an absence of variability: the evaluated image set was exposed to the real dynamics of the coastal environment, encompassing continuous variations in luminosity throughout the day, complex shadows, and hydrodynamic changes during a complete spring tide cycle. The robustness of the detection, therefore, reflects the YOLOv8n architecture's ability to isolate this environmental noise, focusing strictly on the high-contrast structural features of the tide staff.

Table 5 – Confusion matrix.

Actual/Predicted	Tide staff	Background
Tide staff	41	0
Background	0	0

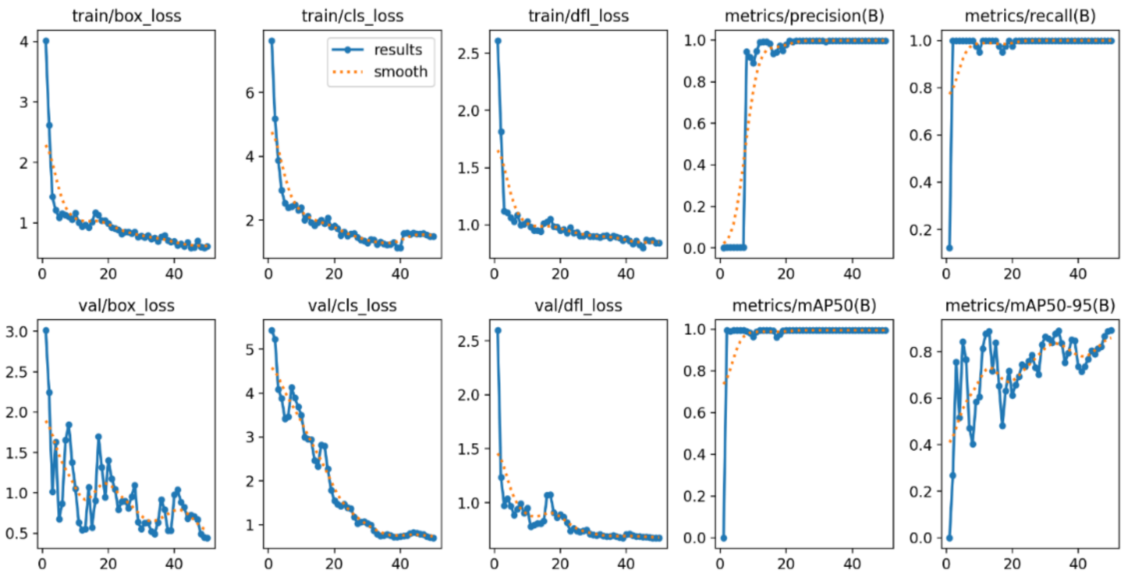
Source: Soares et al. (2025).

Figure 8 presents a graphical summary of the evolution of the losses (losses) and evaluation metrics over 50 epochs. The training graphs (train/box_loss, train/cls_loss, train/dfl_loss) show a consistent and gradual reduction, indicating that the model learned progressively. The smoothed curve (dashed orange line) shows the general trend and helps visualize the continuous improvement. For the validation graphs (val/box_loss, val/cls_loss, val/dfl_loss), greater oscillation is observed, with a downward trend, indicating good generalization. The similarity in behavior observed between the training and validation graphs suggests that there was no overfitting.

The bounding box precision and recall metrics indicated that the model adequately detected the tide staff. An mAP50 of 0.995 indicates excellent performance with an IoU of 50%, reinforcing the detection's quality. The mAP50-95 of 0.882, in turn, reflects a lower value due to the increasingly stringent IoU thresholds.

The horizontal axis of the graphs represents the number of training epochs (epochs), while the vertical axis represents the values of the loss functions (loss functions) and the performance metrics associated with the model, including localization loss (box loss), classification loss (classification loss), distribution focal loss (DFL loss), precision (precision), recall, and mAP metrics. Decreasing values of the loss functions indicate model convergence, while increasing values of precision, recall, and mAP indicate improved detection performance.

Figure 8 – Graphical summaries of training and validation.



Source: Soares et al. (2025).

The next step was to properly convert the pixel values of the lower bounding box for the 125 test images that did not participate in training. Eq. (1) was used as the pixel-to-meter conversion function for each image. The metric values identified in the images were compared with the values calculated by deep learning, and the statistical summary is shown in Table 6.

Table 6 – Statistical summary.

Statistic	Office observations (m)	Deep Learning (m)
Mean	2.14	2.15
Standard Error	0.02	0.02
Standard Deviation	0.26	0.26
Minimum	1.40	1.43
Maximum	2.46	2.49

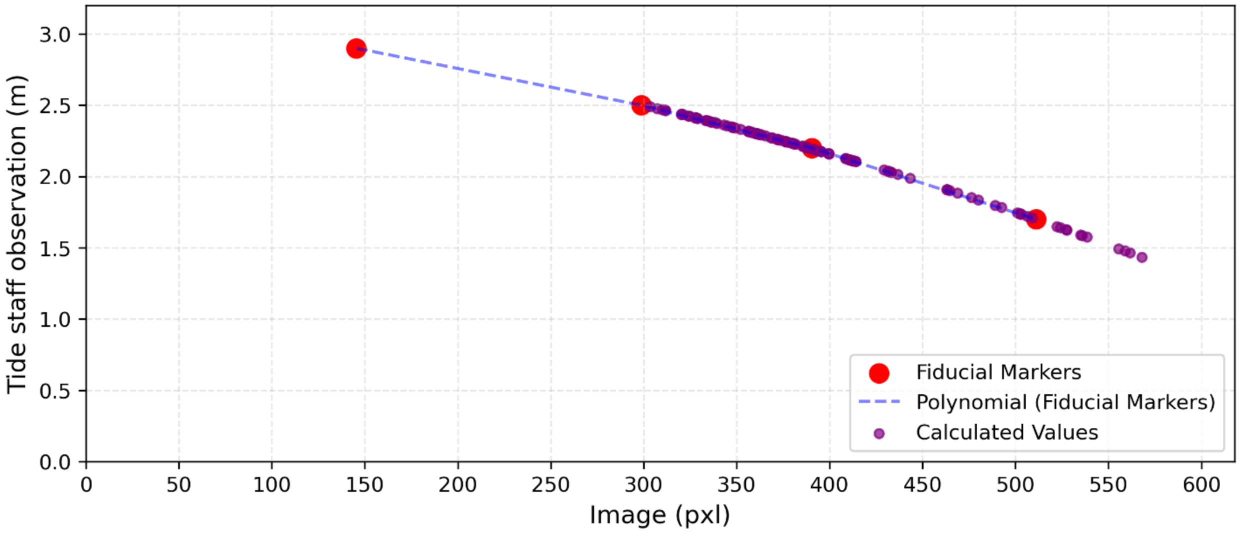
Source: Soares et al. (2025).

To verify whether the observed difference between the population means was statistically relevant, Student's t-test for independent samples was performed. Applying the methodological equations resulted in a standard error of the difference (SE_d) of 0.0329 m and a $t_{\text{calculated}} = 0.311$, associated with 248 degrees of freedom.

Comparing the obtained value with the tabulated limit for a two-tailed 95% confidence interval ($t_{\text{critical}} = 1.970$), it can be seen that $t_{\text{calculated}} < t_{\text{critical}}$. Therefore, the null hypothesis is accepted **(**H_0), confirming that there are no statistically significant differences between the water level extracted visually in the office and the values generated automatically by the inference of the YOLOv8n model integrated with the polynomial function. This result attests to the high rigor and reliability of the proposed optical method for automating the VCT workflow.

Figure 9 presents the distribution of values calculated by deep learning, along with the references that determined the pixel-to-meter conversion function. The extrapolated values maintained adequate behavior, without showing anomalies, as observed in the statistical summary.

Figure 9 – Distribution of the values calculated by deep learning.



Source: Soares et al. (2025).

Having demonstrated the model's effectiveness in Experiment 1, both in detection and metric conversion, the investigation was extended to the generalization and temporal variation scenarios (Experiments 2 through 5). At this stage, all the proposed scenarios were submitted to meteorological validation using the Van de Casteele diagram, which enabled a direct comparison between computational inference and the conventional reference method.

3.2 Comparative analysis of the experiments (Van de Casteele Test)

As demonstrated in Experiment 1, detection was highly effective. All experiments produced a confusion matrix with only true positives, and the evaluation metrics showed robust values, as shown in Table 7. Experiment 2 used the training from the first experiment. Experiment 3, with the largest training dataset, achieved the best mAP50-95.

 DOI: http://dx.doi.org/10.14393/rbcv78n0a-84090

Table 7 – Summary of the evaluation metrics

Metrics	Experiments 1 and 2	Experiment 3	Experiment 4	Experiment 5
mAP50	0.995	0.995	0.995	0.995
mAP50-95	0.882	0.913	0.844	0.874

Source: The Authors (2026).

A relevant methodological aspect concerns the positioning of the second camera. Since it was installed at a slightly different angle from the first, the value conversion function required a specific adjustment for images captured at the 2-minute frequency to reflect the relationship between the tide staff and the images, as shown in Table 8.

Table 8 – Scale relationship between the tide staff and the camera 2 images.

Tide staff (m)	$\bar{y}$ (pxl)
2.90	139.11
2.50	294.03
2.20	386.08
1.70	507.88

Source: The Authors. (2026).

Table 9 consolidates the mean (bias) and standard deviation (SD) results obtained for the level sensors compared with the conventional method, as well as the results of the five computer vision experiments. The abbreviations RAD and ENC refer, respectively, to Radar and Encoder sensors.

Table 9 – Statistical parameters of the Van de Casteele Test for the sensors, in relation to the proposed experiments and the traditional method, in meters.

Statistic	Standard		Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5	
	RAD	ENC	RAD	ENC	RAD	ENC	RAD	ENC	RAD	ENC	RAD	ENC
Mean	0.00	0.84	-0.01	0.83	-0.03	0.81	-0.04	0.81	-0.03	0.81	-0.03	0.81
SD	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03

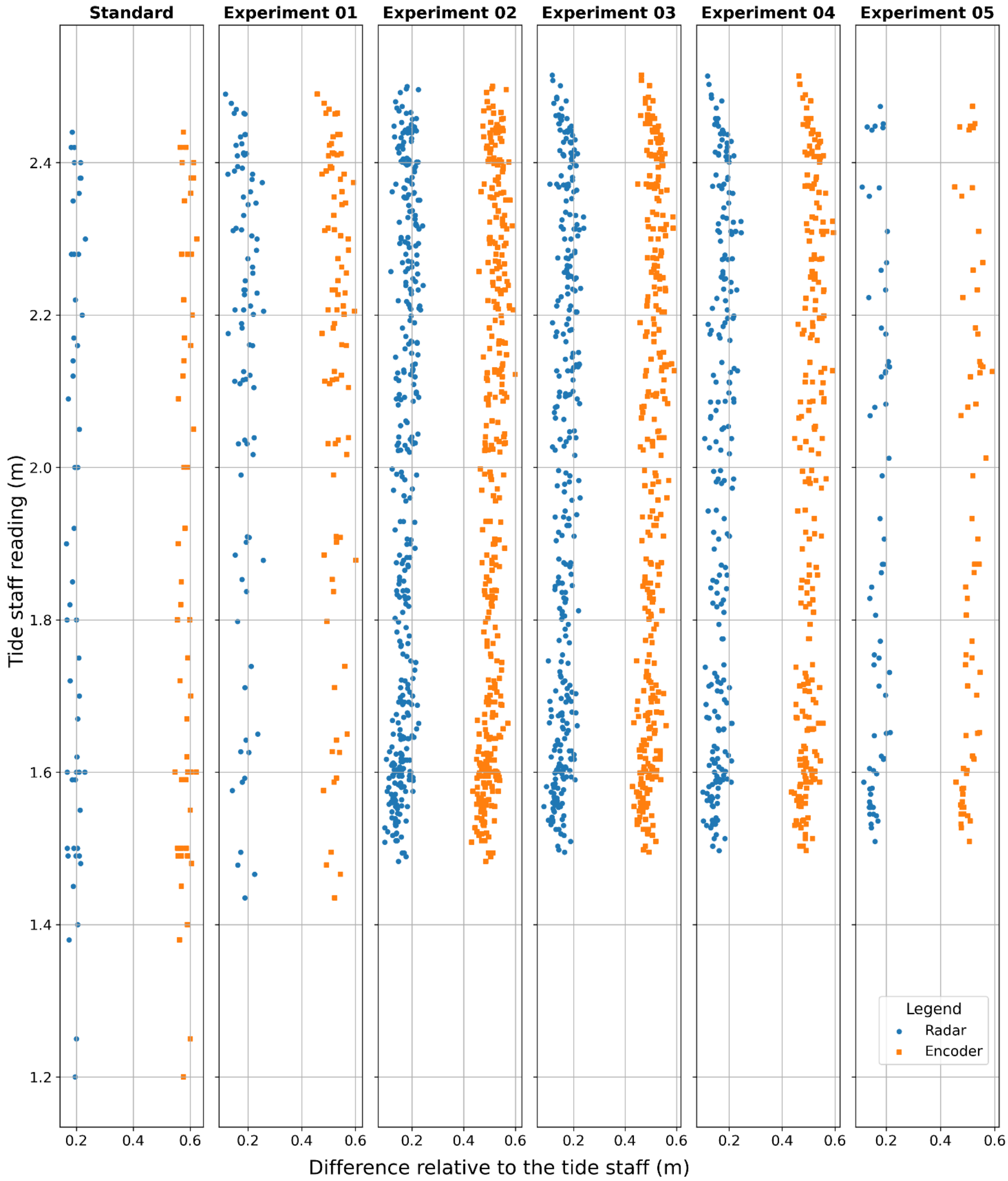
Source: The Authors. (2026).

The manual reading from the conventional method, used as a reference, showed the lowest standard deviation (0.02 m). The deep learning model results converged to a standard deviation of 0.03 m and a systematic bias of the same magnitude for both level sensors, relative to the conventional method, except in Experiment 1, which showed a slightly smaller bias. This stable behavior, even with variations in sampling frequency and changes in camera, demonstrates the robustness of the method, although it reveals the limits imposed by the instantaneous capture of images in the face of the dynamics of the water surface.

The analysis of Figure 10 shows that automated experiments generate a larger volume of data than the conventional method. However, a temporal limitation in the optical technique is observed: while the human observer recorded readings until 19:00, respecting a complete tidal cycle and reaching the 1.2 m mark on the tide staff, camera-based detection was limited to 18:00 due to reduced natural light. Also, according to Figure 10, the VCT does not indicate anomalies in the level sensors, which is a favorable indication for the test's automation process.

 DOI: http://dx.doi.org/10.14393/rbcv78n0a-84090

Figure 10 – Van de Casteele Test diagrams for each experimental scenario



Source: The Authors (2026).

4 SUMMARY AND CONCLUSIONS

The conventional Van de Casteele Test primarily serves as a reactive quality control tool, often detecting errors after data has already been compromised. Although it is recommended to carry it out at regular intervals, detecting problems such as displacements or siltation can invalidate the entire time series before diagnosis. This reactive approach tends to generate gaps or inconsistencies in the data, which requires belated corrective actions.

Given this scenario, the present research demonstrated the technical feasibility of automating the VCT through computer vision. The results of applying the YOLOv8 model indicated efficient convergence, with precision and recall metrics close to 1.0 and mAP50-95 consistently above 0.84 across all experimental scenarios. The statistical analysis confirmed the robustness of the method, with the confusion matrix showing the absence of classification errors and the Student's t-test validating the agreement between the automatic readings and the office-based reference.

Regarding metrological accuracy, the computational model showed an average standard deviation of

0.03 m, slightly higher than the 0.02 m obtained by the conventional method. This marginal difference, added to the identified systematic bias, can be attributed to the nature of the sampling: the instantaneous capture of the image fixes the instantaneous condition of the sea surface, recording high-frequency noise (waves) that is filtered out both by the level sensors, which operate through temporal integration, and by the human observer, who infers a visual average. In addition, oscillations in the experimental support structure of the cameras, which are more susceptible to wind action (vibration and movement), may have contributed to the residual variability. However, these characteristics do not invalidate the method; on the contrary, they confirm that the technique is robust for quality control, provided that the limitations of instantaneous inference are considered.

The proposed automation presents a promising approach to increasing the effectiveness of sea level monitoring. By enabling continuous, high-frequency analysis, the intelligent system can identify subtle deviations in operational behavior that precede critical failures. This enables proactive interventions, such as scheduling maintenance to clear the stilling wells before siltation invalidates the data, or recalibrating sensors before a deviation becomes significant. The cost-benefit ratio favors the automated approach, since the initial investment in low-cost hardware is offset by the elimination of costs associated with field activities, enabling a validation frequency that would be economically unsustainable with the manual method. In addition, deep learning can optimize calibration cycles by continuously adapting to the long-term performance characteristics of tide gauges.

For future work, this evaluation is intended to be expanded to other RMPG tide gauges, implementing autonomous data transmission via wireless networks. To improve data collection robustness, it is recommended to replace the aluminum support with materials more resistant to weathering and to investigate auxiliary lighting systems to overcome restrictions on nighttime monitoring. From a methodological standpoint, it is suggested that an intensive collection model be defined for specific VCT periods. This strategy would allow the temporal averaging of images, mitigating the high-frequency effects observed in this research.

Acknowledgments

The authors thank the Brazilian Institute of Geography and Statistics (IBGE) for providing the data and for granting temporary leave to the first author to pursue doctoral studies. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Author Contributions

Conceptualization: Soares, S.; Santos, E.G.; Reinheimer, L.F.; and Dalazoana, R.; Data Curation: Soares, S.; Santos, E.G.; and Reinheimer, L.F.; Formal Analysis and Investigation: Soares, S. and Dalazoana, R.; Methodology: Soares, S.; Santos, E.G.; and Reinheimer, L.F.; Supervision: Dalazoana, R.; Writing – original draft: Soares, S.; Santos, E.G.; Reinheimer, L.F.; Inoue, M.E.B.; and Dalazoana, R.; Writing – review and editing: Soares, S.; Santos, E.G.; Reinheimer, L.F.; Inoue, M.E.B.; and Dalazoana, R.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Birol, F., Bignalet-Cazalet, F., Cancet, M., Daguze, J.-A., Fkaier, W., Fouchet, E., Léger, F., Maraldi, C., Niño, F., Pujol, M.-I., & Tran, N. (2025). Understanding uncertainties in the satellite altimeter measurement of coastal sea level: insights from a round-robin analysis. *Ocean Science*, *21*, 133–150. https://doi.org/10.5194/os-21-133-2025.

Centro de Hidrografia da Marinha – CHM. (2025). *Tábuas de Maré 2025*. https://www.marinha.mil.br/chm/tabuas-de-mare?page=0.

Dalazoana, R., Luz, R.T., Lima, S.R.S., Miranda, Fr. A., Palmeiro, A.S., Miranda, Fa. A. e Freitas, S. R. C. (2005). *Controle vertical das RRNN e da posição geocêntrica do maregáfico no Porto de Imbituba*.

[Conference paper]. IV Colóquio Brasileiro de Ciências Geodésicas, Curitiba. https://artigos.ibge.gov.br/artigos-home/geodesia/2008-2005.

Del Rosso, M. P., Sebastianelli, A. P., Aparício, S. P., & Wheeler, J. P. (2021). How to create a proper EO dataset. In: *Artificial Intelligence Applied to Satellite-based Remote Sensing Data for Earth Observation*. IET. (pp. 113–130). https://doi.org/10.1049/PBTE098E_ch6.

Freitas, G. T. M. (2022). Software para identificação de defeitos na superfície de pavimentos rodoviários utilizando Deep Learning. 109 f. [Master's thesis, Universidade Federal do Ceará – UFC]. http://www.repositorio.ufc.br/handle/riufc/69600

Gemael, C. (2019). *Introdução à geodésia física*. Editora UFPR.

Google. (2026). *Google Colab* [Software]. https://colab.research.google.com/

Instituto Brasileiro de Geografia e Estatística – IBGE. (2010). *Instruções técnicas para controle geodésico de estações maregáficas – CGEM e sua vinculação vertical ao Sistema Geodésico Brasileiro – SGB*. IBGE. https://geoftp.ibge.gov.br/metodos_e_outros_documentos_de_referencia/revista_e_manuais_tecnicos/mtg11.pdf

Instituto Brasileiro de Geografia e Estatística – IBGE. (2021). *Monitoramento da variação do nível médio do mar nas estações da Rede Maregáfica Permanente para Geodesia: 2001–2020*. IBGE. https://biblioteca.ibge.gov.br/visualizacao/livros/liv101890.pdf

Intergovernmental Oceanographic Commission – IOC. (2024). *State of the Ocean Report*. IOC-UNESCO. (IOC Technical Series, 190). https://doi.org/10.25607/4wbg.

Jähne, B. (2005). *Digital Image Processing*. [S.l.]: Berlin: Springer. https://doi.org/10.1007/3-540-27563-0

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lima, J. M. C. (2019). Reconhecimento e análise de rachaduras a partir de imagens para monitoramento em regiões com atividade sísmica frequente. [Master's thesis, Universidade Federal de Alagoas - UFAL]. Programa de Pós-Graduação em Informática. Maceió, Alagoas. http://www.repositorio.ufal.br/handle/riufal/7280

Lin, T.-Y., Maire, M., Belongie, S., Lubomir Bourdev, Girshick, R., Hays, J., Perona, P., Ramanan, D., C. Lawrence Zitnick, & Piotr Dollár. (2014). Microsoft COCO: Common Objects in Context. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1405.0312

Miguez, B. M., Testut, L., & Wöppelmann, G. (2008). The Van de Casteele test revisited: an efficient approach to tide gauge error characterization. *Journal of Atmospheric and Oceanic Technology*, *25*(7). DOI: https://doi.org/10.1175/2007JTECHO554.1

Mucci, T. (2024). *Overfitting vs. underfitting*. Ibm.com. https://www.ibm.com/br-pt/think/topics/overfitting-vs-underfitting

Sánchez, L., & Sideris, M. G. (2017). Vertical datum unification for the International Height Reference System (IHRS). *Geophysical Journal International*, *209*(2), 570–586. https://doi.org/10.1093/gji/ggx025.

Silveira, L. N. (2023). *Aplicação de dados altimétricos orbitais e nivelamento hidrográfico em corpos hídricos continentais para a integração de referenciais verticais*. [Doctoral dissertation, Universidade Federal do Rio Grande do Sul]. Lume Repositório Digital. http://hdl.handle.net/10183/272222.

Skalski, P. (2022). *Make Sense* (version 1.11.0). https://www.makesense.ai/; https://github.com/SkalskiP/make-sense

Soares, S., Dalazoana, R., Inoue, M. E. B., Dos Santos, E. G., & Reinheimer, L. F. (2025). *Investigação preliminar do aprendizado profundo no teste de Van de Casteele na estação maregáfica de Arraial do Cabo/RJ*. [Conference paper]. I Congresso Internacional de Regularização Fundiária & VIII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação (I CIRF & VIII SIMGEO), Recife. https://doi.org/10.5281/zenodo.17135655.

SONEL. (2019). *Quality control of the measurements published by SHOM*. https://www.sonel.org/_Quality-control-of-the.html.

Srivastava, A., Dalvi, A., Britto, C., Rai, H., & Shelke, K. (2020). Explicit content detection using Faster R-CNN and SSD MobileNet v2. *International Research Journal of Engineering and Technology*, *7*(3). https://www.irjet.net/archives/V7/i3/IRJET-V7I31126.pdf.

Taye, M. M. (2023). Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. *Computation*, *11*(52). https://doi.org/10.3390/computation11030052.

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision. *Machine Learning and Knowledge Extraction*, *5*, 1680–1716. https://doi.org/10.3390/make5040083.

Torge, W., & Müller, J. (2012). *Geodesy* (4th ed.). Walter de Gruyter.

ULTRALYTICS. (2023). *YOLOv8*. https://docs.ultralytics.com/models/yolov8/.

UNESCO. Intergovernmental Oceanographic Commission. (1985). *Manual de medição e interpretação do nível do mar*. IOC. https://www.jodc.go.jp/info/ioc_doc/Manual/065061eb.pdf

UNESCO. Intergovernmental Oceanographic Commission. (2002). *Manual on sea level measurement and interpretation: reappraisals and recommendations as of the year 2000*. IOC. https://unesdoc.unesco.org/ark:/48223/pf0000125129?posInSet=1&queryId=f96ff20a-ffad-4228-98d2-81751401754f

UNESCO. Intergovernmental Oceanographic Commission. (2016). *Manual on sea level measurement and interpretation: radar gauges*. IOC. https://unesdoc.unesco.org/ark:/48223/pf0000246981?posInSet=2&queryId=f96ff20a-ffad-4228-98d2-81751401754f

Zeng, J., Zhong, H. (2024). YOLOv8-PD: an improved road damage detection algorithm based on YOLOv8n model. *Sci Rep* 14, 12052. https://doi.org/10.1038/s41598-024-62933-z

Main author biography

Salomão Soares was born in Rio de Janeiro, Brazil, in 1980. He received the B.Sc. degree in Cartographic Engineering from the Universidade do Estado do Rio de Janeiro (UERJ) in 2003 and the M.Sc. degree in Population Studies and Social Research from the Escola Nacional de Ciências Estatísticas (ENCE) in 2006. He is currently pursuing the Ph.D. degree in Geodetic Sciences at the Universidade Federal do Paraná (UFPR). He works as a Technologist at the Brazilian Institute of Geography and Statistics (IBGE), within the Geodesy Coordination. His research interests include vertical reference systems, satellite altimetry, and tide gauge observations.