Marine Mammal Science

ARTICLE

Evaluating the Effectiveness of a Multispecies Matching Algorithm Applied to a Novel Species, Risso's Dolphins (*Grampus griseus*)

Alexandra Vanderzee[1] | Ted Cheeseman[2,3] | Philip T. Patton[4] | John Calambokidis[1]

[1]Cascadia Research Collective, Olympia, Washington, USA | [2]Marine Ecological Research Centre, Southern Cross University, Lismore, New South Wales, Australia | [3]Happywhale.com, Santa Cruz, California, USA | [4]Migratory Bird Center, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, USA

Correspondence: Alexandra Vanderzee (avanderzee@cascadiaresearch.org)

Received: 1 November 2025 | **Revised:** 11 February 2026 | **Accepted:** 3 March 2026

Keywords: automated matching | *Grampus griseus* | multispecies algorithm | photo identification | Risso's dolphin

ABSTRACT

Automated matching algorithms are an increasingly valuable tool in marine mammal research. A multispecies matching algorithm, originally described in Patton et al. analyzes distinctive morphological characteristics on cetaceans using scarring, pigmentation patterns, and notches along the trailing edge of the dorsal fin, which can pose challenges in traditional eye-based matching due to individual variation over time. While the algorithm was trained on 24 different cetacean species, Risso's dolphins were not included. Here, we assess its effectiveness using Cascadia Research's Risso's dolphin photo-identification catalog, which consists of 1165 unique individuals from Southern California documented across 701 sightings from 2010 to 2021. The multispecies algorithm identified 90 missed matches within the catalog and was subsequently used to update the Cascadia Research catalog more efficiently with sightings collected through 2024. Our results indicate that the algorithm outperformed experienced human matchers, particularly for individuals with less distinctive features, streamlining the identification process and supporting the long-term monitoring of Risso's dolphins while showing strong potential for application to other species with similar identification challenges.

1 | Introduction

Photo-identification (photo-ID) is a foundational tool in marine mammal research, enabling researchers to track individual animals and study population trends, distribution, and movement patterns across various species (Wells and Scott 1990; Karczmarski et al. 2022). By analyzing distinct markings, dorsal fin morphology, and scar patterns, individuals can be monitored noninvasively over time, providing valuable data to inform conservation measures and management strategies (Hammond 1990; Parra et al. 2006, Robinson et al. 2012). Traditional photo-ID methods of identifying marine mammals, such as dolphins and whales, have relied on manual analysis of photographs, a process that can be both time-consuming and expensive, and subject to human error, especially in large populations with extensive datasets (Fuller et al. 2025; Tyson Moore et al. 2022; Bouma et al. 2018).

In recent years, the use of algorithm-based matching has emerged as a transformative tool in marine mammal research by automating the process of matching individuals within a catalog, offering solutions to the limitations of traditional photo-ID methods (Cheeseman et al. 2022; Patton et al. 2023; Blount et al. 2022; Maglietta et al. 2018). Recent studies have demonstrated the potential of machine learning to reliably recognize individual marine mammals across vast datasets, facilitating large-scale collaboration among researchers and citizen scientists (Cheeseman et al. 2022, Patton et al. 2023). By automating

FIGURE 1 | Map showing the Southern California Bight, where the images in this study were collected.

the matching process, these tools have the potential to improve data processing speeds, reduce error rates, and provide a more cost-effective approach to photo-ID.

Risso's dolphins (*Grampus griseus*), a gregarious, deep-diving odontocete species, are widely distributed in the waters of the Southern California Bight. Manual photo-ID methods have been successfully applied to Risso's dolphin in regions such as the Azores to quantify individual residency patterns and site fidelity (Hartman et al. 2015) and at Catalina Island to study marine mammal distribution (Shane 1994). Each Risso's dolphin has a unique pattern of scarring, coloration, and dorsal fin shape, making them ideal candidates for individual identification (Hartman et al. 2015). Subadults have much darker skin that lightens as they mature, and adult males can become almost completely white, partly due to the scarification caused by the teeth of other Risso's dolphins, markings from predation and prey interaction, and loss of pigment (Hartman et al. 2008, Hartman et al. 2015, 2016). This can make them a challenging species to match, as rapid accumulation of new scars and dorsal fin notches over time may limit the ability to reliably identify individuals across long timeframes, particularly in the case of juveniles, which accrue scars and marks at a faster rate than adults (Hartman et al. 2008, Mariani et al. 2016). In spite of rapid accumulation, Risso's dolphin scars are notably persistent over time; Fadda and Airoldi (2000) documented the longevity of individual markings over a 9-year period, while Mariani et al. (2016) found that distinct mark types defining the fin outline remained recognizable across their 11-year study.

A multispecies algorithm tool described by Patton et al. (2023) had been tested on 24 different species of cetaceans, though Risso's dolphins were not included in that initial test set. In this study, we evaluate the effectiveness of the multispecies matching algorithm described by Patton et al. (2023) in identifying individual Risso's dolphins in the Southern California Bight, using an established photo-ID catalog developed by Cascadia Research through traditional eye-based matching. The data and photographs that are part of this study were derived incidentally from two multiyear interdisciplinary research projects: the Southern California Behavioral Response Study (SOCAL-BRS), and the Tagless Behavioral and Physiological Response Study (BPRS). These projects were conducted in the Southern California Bight, primarily in the summer months of 2010–2015 (SOCAL-BRS) and from 2017 to 2021 (BPRS) to evaluate marine mammal response to mid-frequency active sonar (Southall et al. 2012; Durban et al. 2021). By analyzing the results from the multispecies algorithm, we assess the precision and potential of the algorithm to enhance traditional eye-based methods of photo identification for this species, and thus contribute to more efficient long-term monitoring of Risso's dolphins in the Southern California Bight.

2 | Methods

Most surveys, sightings, and photo-ID work were conducted from 5.9 to 6.4 m Rigid Hull Inflatable Boats (RHIBs). Larger research vessels were used occasionally, including the 20 m dive boats the R/V *Truth* and the R/V *Magician*, the 20 m motor yacht R/V *Valkyrie*, and a dedicated passive acoustic monitoring (PAM) vessel, the R/V *Bayliss*. Up to four boats operated on a single day. We conducted 499 days of nonsystematic small boat surveys over 179 different days during the months of July–October in the Southern California Bight (Figure 1) during the 2010–2014 period (Southall et al. 2012). This meant that survey effort varied due to a number of factors, including the availability of coordination with the Navy, weather, and the presence of priority species, which included toothed cetaceans and baleen whales (Southall et al. 2012). We conducted 73 days of nonsystematic small boat surveys over 27 different days during the months of October–December in the 2019–2021 period, of which the basis of operations was out of Catalina Island (Durban et al. 2021).

A single sighting of Risso's dolphins was defined as one vessel encountering a group of animals, where a group was defined as one or more individuals traveling in association with each other and exhibiting a common behavior. We tallied our effort each year by computing different vessel days of effort, since we had up to four vessels operating on a single day and each vessel could encounter a different group of animals. If multiple vessels encountered the same group of animals, these were counted as different sightings for each vessel, and if each vessel took photographs, these were compiled and the best images of each individual from that day were chosen to build the catalog. For photo identification, we used distinguishing marks and notches on either side of the dorsal region and trailing edge of each animal to identify unique individuals and track re-sightings over time. Photography for identification was typically a second priority to tagging and surface behavior monitoring. Where possible, ID photos were taken at multiple points in time while tracking a group of Risso's dolphins.

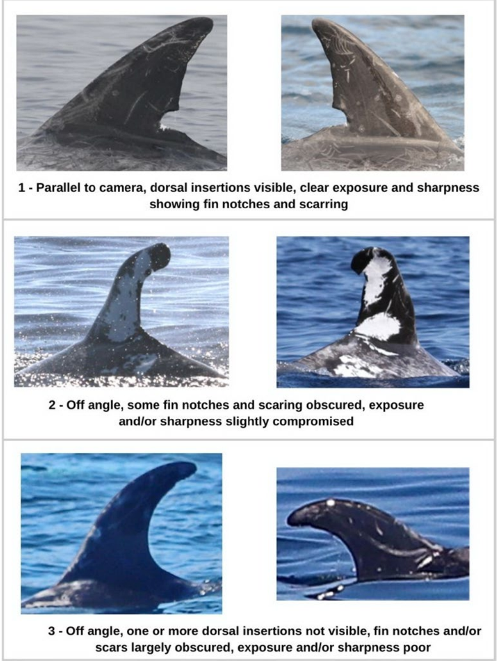

FIGURE 2 | Examples of key identification features used to compile unique IDs in the Risso's dolphin catalog, representing overall quality scores of 1, 2, and 3.

For individual photo-ID, we used high-resolution, parallel, in-focus images that showed the dorsal fin and a portion of the body around the dorsal fin. The best available image of each individual from each surfacing sequence photographed was selected and cropped into a square shape, keeping both dorsal fin insertion points visible (Figure 2). These images were imported into a database designed to reconcile large collections of dorsal fin photos into individuals that were then scored for image quality and distinctiveness. Image quality scoring consisted of assigning several categories to each image presented (Table 1). Images assigned with 1's typically represented the best, or excellent qualities of each of the categories, 2's were considered medium quality, and 3's were considered fair quality (Figure 2). Images that had multiple scores of three, or where the dorsal fin was significantly off angle, exposure dark, blurry, or the key parts of the body were obscured, were given a "poor quality" designation and not added to the catalog. Once the image quality categories were assigned, an overall mean "Image Quality" from 1 to 3 was generated for each image that reflected the scores

TABLE 1 | Image quality and distinctiveness scoring criteria.

Category	Score	Description
Image quality		
Side	Right/left	Side of animal photographed
View	Ahead/lateral/behind	Position of the animal relative to the camera
Angle	1–3	1 = parallel, 2 = slightly off angle, 3 = very off angle
ProVis	1–3	1 = both dorsal fin insertion points visible, 2 = one visible, 3 = none visible
Exposure	1–3	1 = excellent, 2 = medium, 3 = fair
Sharpness	1–3	1 = excellent, 2 = medium, 3 = fair
Overall image quality	1–3	Generated by the database based on scores above. Images scoring > 2.5 were excluded from the catalog.
Distinctiveness		
Fincat; notches in dorsal fin	1–5	1 = clean fin (no notches), 2 = 1–2 notches, 3 = 3 notches, 4 = 4+ notches, 5 = disfigured fin (missing or bent)
ScarScore	1–6	1 = minimal scarring, 0%–10%, 2 = low scarring, 10%–25%, 3 = moderate scarring, 25%–50%, 4 = high scarring, 50%–75%, 5 = heavily scarred, 75%+, 6 = abnormal pigment or disfigured

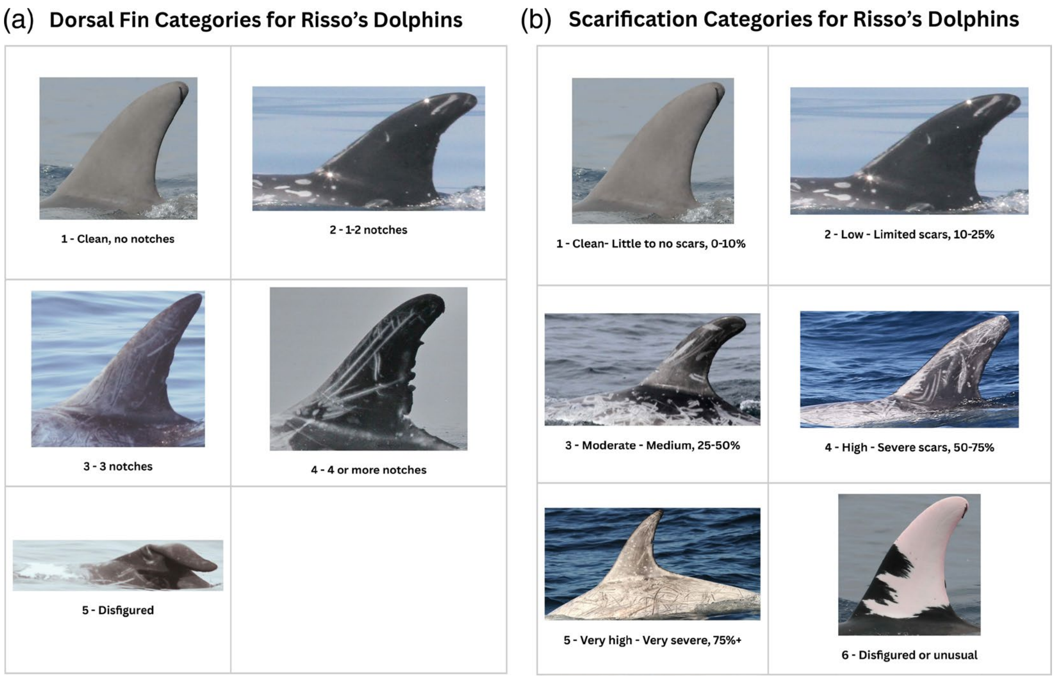

FIGURE 3 | (a and b) Distinctiveness categories for Risso's dolphins based on number of dorsal fin notches and degree of scarification.

given. No images made it into the historical catalog that scored higher than a 2.5 in overall image quality.

Distinctiveness was scored based on the number of notches in the trailing edge of the dorsal fin ("Fincat," as described in Table 1), and given a category of 1–5 (Figure 3a). Similar to Hartman et al. (2008), Risso's were also assigned to categories based on dorsal fin scarification. This scoring system ranged from 1 to 6 and categorized animals based on the amount of scarring on their dorsal fins and on the portion of the body visible around the dorsal fin (Table 1, Figure 3b). These distinctiveness designations could be subjective due to both image quality and the individual assigning the categories, depending on the angle, exposure, and sharpness of an image; some smaller notches or scars could be harder to see than others, which is why multiple images of each individual from a variety of angles were scored and compiled when possible.

The full manual and algorithm-assisted photo-ID workflows used to process these images, as well as the tests conducted as part of this study, are summarized in Figure 4. For the manual process, we used Image J to expedite the eye-based matching process, obtaining proportional measurements of dorsal fin height and width (Rasband 1997-2016) to categorize fin shape. This system was formalized after the initial catalog (spanning 2010–2014) was constructed through eye-based matching, and its performance was evaluated before integration. Accuracy tests of the measurement system, based on 836 measured images representing 46 individuals, showed that 90% of correct matches were retrieved within the first 10% of possible candidates, and nearly all correct matches were retrieved within the first 30% of possible candidates. Based on these results, catalog verification and subsequent matching efforts were streamlined by restricting searches to the top 30% of ranked potential matches to the historical catalog, substantially reducing search effort while maintaining high accuracy.

Right and left sides of images of individuals with distinctive notches were matched when possible using the trailing edge of the dorsal fin; however, right and left sides of animals with distinct scarring but no notches were counted as unique individuals since they could not be accurately paired together. Each unique individual could have between 1 and 12 images associated with that animal, in order to give the matcher a variety of angles showing different portions of the body and dorsal fin to enhance the chance of finding a match. Matching was first conducted within the same day to reconcile duplicate detections of the same individual and these reconciled individuals were then compared to an overall catalog of individuals sighted during SOCAL-BRS and BPRS. A second person would verify whether the matches made by the first person were correct, and then take any individuals not found by the first person through the catalog again to check for any missed matches.

The algorithm-assisted process involved uploading images into the dorsal fin matching algorithm platform (through happywhale.com), selecting "dorsal fin" and clicking the button "identify" to run the images against the historical catalog. The platform returned a ranked queue of potential matches for each uploaded image. Each candidate match was accompanied by a catalog ID, an algorithm-generated similarity score, and the total number of matches presented. Algorithm similarity scores for each test image were scored on a unitless 0.0–1.0 match confidence scale, where scores represent the cosine similarity between extracted feature vectors from dorsal fin regions, with 0.0 being no correlation and 1.0 being a perfect match (Cheeseman et al. 2022). The queue was ordered in descending similarity score, with the highest-ranking candidates appearing first.

Review of the results was conducted manually. Each uploaded image was compared to its corresponding list of candidate matches, and the full queue of options presented was reviewed regardless of where the correct match appeared. There was not

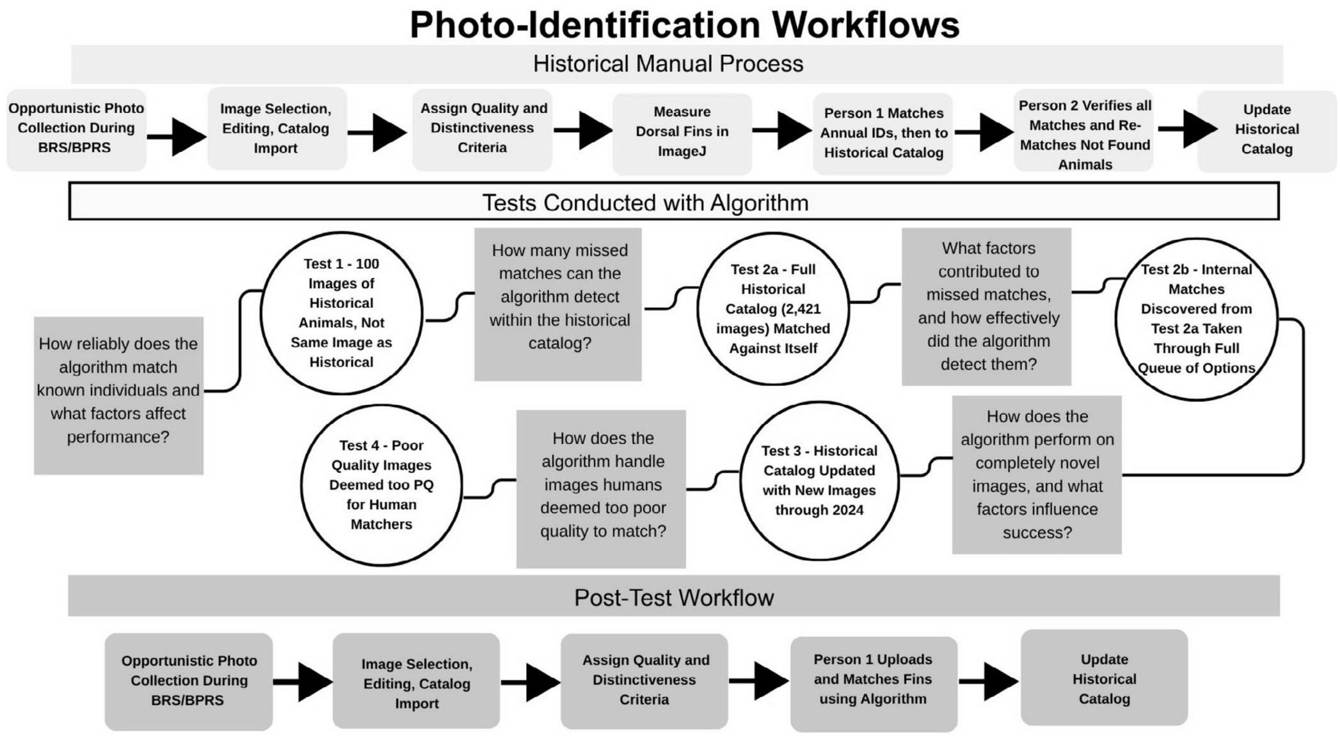

FIGURE 4 | Schematic overview of both the historical manual photo-identification workflow, the tests conducted, and the algorithm-assisted workflow.

a method to assign a match to a photo within the platform, so the results were recorded in a separate excel file, which detailed the match ID, filename of the image, the number the match appeared in the queue, the algorithm similarity score, and the number of matches presented. Any possible missed matches, defined as images that were assigned the same catalog ID by the human cataloger but where actually two distinct individuals, were also recorded, including the ID number of the missed match, the number the missed match appeared in the queue, and the corresponding algorithm similarity score.

To evaluate overall algorithm performance, the mean average precision (MAP) score was computed following the method described in Patton et al. (2023). MAP summarizes how well the algorithm ranks the correct match among the list of candidate images, where a score of 1.0 indicates that the correct match is always returned as the top-ranked candidate, while lower scores reflect cases where the correct match appears further down the ranked queue or is missed entirely (Patton et al. 2023). For each query image, a precision score was assigned based on the ranking position of its correct match (1.0 if first, 0.5 if second, down to 0 if not within the top five), and these precision scores were then averaged to calculate the MAP for that dataset. MAP provides a single, intuitive measure of how effectively the algorithm prioritizes a correct match and higher MAP values indicate that fewer images need to be reviewed to reliably identify individuals, reducing human effort while maintaining confidence in the match results.

To assess whether image characteristics influenced algorithm performance, a series of statistical tests were conducted in R version 4.3.2. One-way analyses of variance (ANOVA) were used to compare similarity scores across levels of categorical image quality and distinctiveness metrics. Assumptions of one-way ANOVA were evaluated using Shapiro–Wilk tests for normality of residuals and Levene's tests for homogeneity of variance. Where mild departures from normality were detected, results were confirmed using nonparametric Kruskal–Wallis tests, these are reported in Data S1. Linear regressions were also performed to examine the relationship between similarity score and individual image characteristics. Image metrics evaluated are detailed above in Table 1, though the count of distinct notches on the trailing edge of the fin was added (called "Fincat_Count") as an additional distinctiveness metric, and was recorded independent of the general Fincat category.

3 | Results

A total of 701 sightings of Risso's dolphins yielded 2421 images of 1165 uniquely identified animals during the 7 years of effort from 2010 to 2014, 2019, and 2021. This catalog of Risso's dolphins that was compiled from eye-based matching from 2010 to 2021 effort was given to the multispecies matching platform to serve as the basis for the tests detailed below; key results are summarized in Table 2.

3.1 | Test 1—Novel Image Matching for Known Individuals

In the first test, which involved 100 images of known individuals that were not the same image of that individual, the multispecies matching algorithm successfully identified all 100 Risso's dolphin images. Of these, 99 individuals were returned as the top-ranked match (position 1 in the queue), while one

TABLE 2 | Summary of algorithm performance and key results across tests.

Test # and purpose	Images (n)	Mean similarity score	Mean image quality	MAP (primary)	Dummy-adjusted MAP	Variable	Regression R2, t, p	ANOVA F, p
1 Novel Image Matching	100	0.72 (0.39–0.92)	1.56 (1–2)	0.991		Fincat	0.07, 2.88, 0.0049	3.85, 0.012
1 Novel Image Matching						Fincat_Count	0.04, 2.26, 0.026	2.52, 0.013
2b Missed match rank analysis	203	0.67 (0.39–0.89)	1.39 (1–2.25)	0.847		ProVis	NS	20.79, < 0.00001
2b Missed match rank analysis						Mean Image Quality	NS	9.18, 0.0028
2b Missed match rank analysis						Fincat	NS	2.94, 0.0345
3 Catalog updating	259	0.65 (0.40–0.84)	1.35 (1–2)	0.904	0.669	Fincat_Count	0.070, 2.45, 0.0171	5.99, 0.0171
4 Poor Quality Images	145	0.66 (0.45–0.99)	2.10 (2–2.5)	0.965	0.762	NS	NS	NS

Note: "NS" indicates "not significant" and blank areas indicate there is no change from what is shown above.

FIGURE 5 | Examples of opposite side matches found by the algorithm that humans had missed.

individual appeared seventh out of seven candidates. The outlier corresponded with an uncropped, unedited image, which also received the lowest similarity score in the dataset. The remaining four uncropped images all matched successfully as the top-ranked candidate. Among this dataset, four represented opposite-side matches (i.e., a left side image matched to a historical right-side image, or vice versa), yet the algorithm was still able to return the correct individual due to distinct features (Figure 5). These opposite side matches were only confirmed when there was absolute certainty in the match, meaning the shape of the dorsal fin and number of notches aligned, and there was supporting evidence of the match confirmation, such as the images being taken on the same day from the same vessel and sighting (Figure 5).

The test also identified nine missed matches. Four of these missed matches were identified directly by the algorithm, and in each case, the missed match appeared second in the queue, immediately after the correct catalog match. These four had a mean similarity score of 0.66, and exhibited mark changes, defined as changes in dorsal fin appearance over time, such as development of new notches or scars not present in earlier encounters. An additional five missed matches were identified incidentally by the human matcher during manual review of the candidate queues. One of the nine cases involved an opposite-side comparison, however, consistent dorsal ridging enabled confident identification. Overall, the missed match rate was 9%, based on the nine cases identified during the review of 100 randomly chosen individuals.

Both the categorical variable Fincat (fin distinctiveness grouped into category) and the continuous variable Fincat_Count (the total number of notches per individual) were statistically significant predictors of similarity score (Table 2), indicating that individuals with more dorsal fin notches tended to yield higher similarity scores (Figure 6). See Supporting Information for details on methods for Test 1.

3.2 | Test 2a—Full Historical Catalog

The second test involved comparing all 2421 images in the historical catalog against the catalog itself. As expected, each image returned itself as the top-ranked match, representing

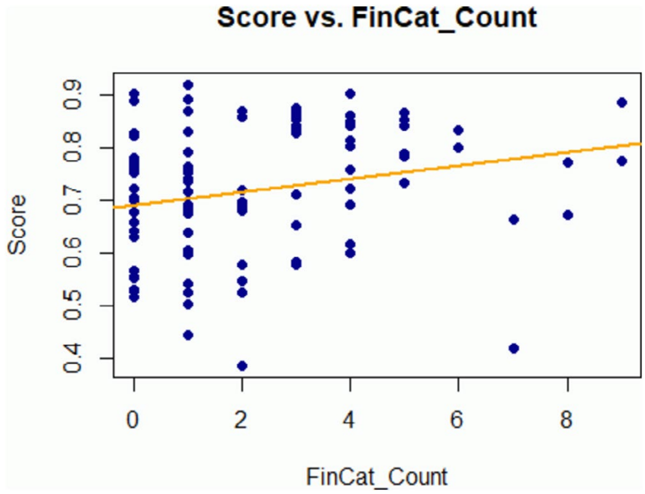

FIGURE 6 | The relationship between FinCat_Count (the number of notches in a dorsal fin) and Similarity Score with a regression line to show the trend in the data from the linear regression model.

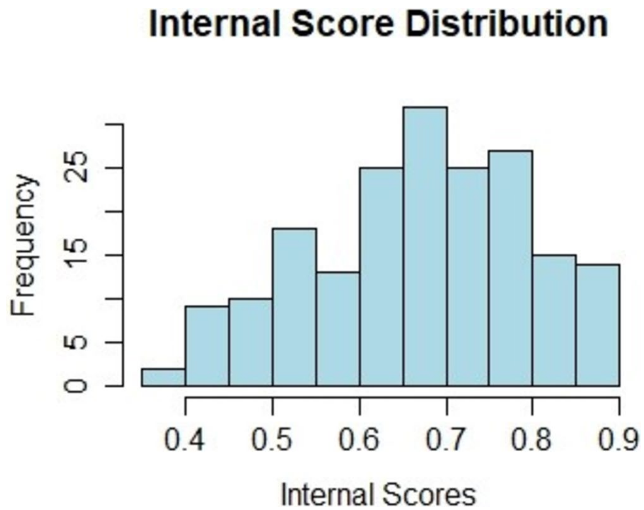

FIGURE 7 | Distribution of missed match (noted here as internals) similarity scores for 203 images.

1165 unique individuals. Within this dataset, 90 missed matches were identified, comprising a total of 203 images. The majority of these missed matches (*n* = 78 unique IDs) appeared as the second option in the queue. An additional eight unique IDs were third in the queue, two appeared fourth in the queue, and two appeared fifth in the queue.

In one case, the match was asymmetric, meaning one unique ID matched and appeared within the first five options, but the reciprocal comparison did not return the corresponding unique ID within the first five results. When the search was extended beyond the top five candidates, this match was identified at the eighth position. See Supporting Information for details on methods for Test 2a.

3.3 | Test 2b—Missed Match Rank Analysis

Test 2b focused on the ranking behavior of the missed matches identified in Test 2a. For this analysis, the 203 images associated with the 90 unique individuals were examined by reviewing the full queue of options presented by the algorithm, and recording both the position at which the missed match appeared and its corresponding similarity score. The mean algorithm similarity score for these missed matches was 0.67 (range 0.39–0.89) (Table 2), with the distribution of scores shown in Figure 7. Across these images, 71% of missed matches appeared as the second-ranked result in the algorithm queue, 8% ranked third, and 2% or fewer ranked higher in the queue (Table 3, Figure 8). The mean CRC quality score for this test was 1.39 (range 1–2.25), similar to the average quality score of the overall historical catalog (mean = 1.42, range 1–2.5) (Figure 6). Not shown in Table 3 are the handful of missed matches that came up higher than the 11th choice in the queue, totaling nine photos that matched at positions ranging from 17th to 90th. Each of these outliers occurred only once and typically corresponded with images that were off-angle, of lower quality, or were the opposite side. These had a markedly lower mean similarity score of 0.4267. To calculate the MAP score, the self-match was excluded and the first nonsame-image match was treated as the top-ranked candidate.

Analysis indicated several factors significantly influenced the algorithm similarity scores within this missed match dataset. ProVis (as defined in Table 1) had a significant effect on similarity score, indicating that as visibility decreased, similarity scores also declined. Mean image quality, calculated from the average of four image quality components (angle, ProVis, sharpness, and exposure), had a significant effect on similarity score, with lower image quality associated with reduced similarity scores (Figure 9). Similar to Test 1, Fincat showed a significant effect in ANOVA but was not significant in linear regression suggesting group-level differences without a clear trend. Unlike Test 1, Fincat_Count was not significant in this test.

Additionally, of the 203 images corresponding to missed matches identified, 46 exhibited mark changes that may have contributed to the human matcher initially missing the correct matches. 58 of these missed matches involved opposite side matches, 31 of which were of lower quality, potentially reducing detectability. Forty of the 203 images were of animals for which only a single photo was available, limiting the opportunity for the human matchers to compare multiple angles or lighting conditions during eye based matching. See Supporting Information for details on methods for Test 2b.

3.4 | Test 3—Updating Risso's Dolphin Catalog

In Test 3, CRC's survey efforts in 2022–2024 yielded 185 unique Risso's dolphin IDs across 259 images. The multispecies matching algorithm identified 50 matches, comprised of 67 images, to individuals already in the historical catalog. Of these, 58 of the matches appeared first in the queue of options presented by the algorithm, two appeared second, two appeared third, one fourth, one fifth, two in the sixth, and

TABLE 3 | Missed matches of the 203 images analyzed per location in the queue.

	Second queue	Third queue	Fourth queue	Fifth queue	Sixth queue	Seventh queue	Eighth queue	11th queue
Missed match photos	145	17	5	3	3	1	1	2
Mean quality score	0.7157 (0.4830–0.8900)	0.5954 (0.4870–0.6680)	0.5224 (0.4290–0.6610)	0.5273 (0.5050–0.5400)	0.4777 (0.466–0.4960)	0.4520 (0.4520–0.4520)	0.5450 (0.5450–0.5450)	0.5045 (0.4860–0.5230)
CRC mean quality	1.4045 (1–2.25)	1.3676 (1–1.75)	1.15 (1–1.5)	1.25 (1.25–1.25)	1.4167 (1–1.75)	1 (1–1)	1.75 (1.75–1.75)	1.5 (1.25–1.75)

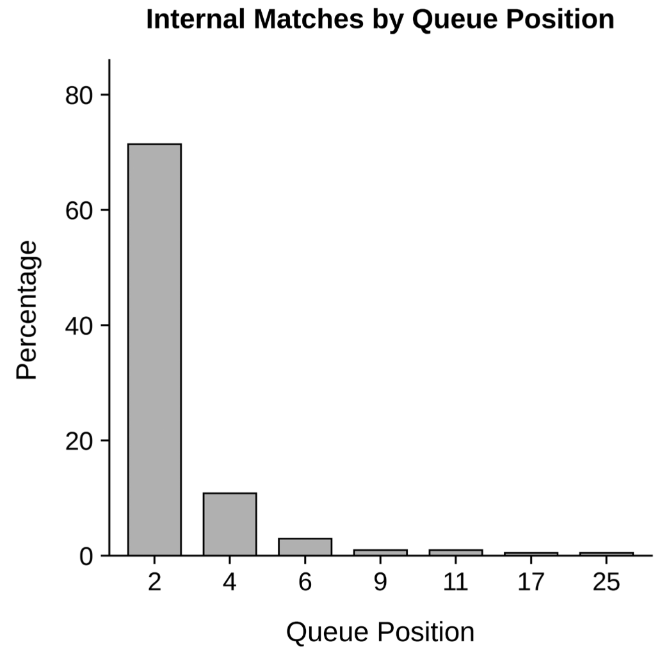

FIGURE 8 | Missed matches by queue position (shown here as internal matches).

one in the eighth position. Of the 50 matches, 27 were noted to show mark changes, such as new scratches, notches, or pigmentation marks, when compared to the historical image. Algorithm performance metrics for this test are summarized in Table 2, with detailed results below. While the MAP score for these matches is 0.9041, when recalculated across all 259 images with the inclusion of the dummy "new individual" prediction, the MAP score decreased to 0.6691, reflecting the algorithm's overall performance when both matched and unmatched cases were considered.

Linear regression analysis indicated that the number of dorsal fin notches ("Fincat_Count") was a significant predictor of similarity score (Figure 10), and ANOVA results supported this association. No other image quality or distinctiveness metrics were significantly related to the similarity score in this test. The results of Test 3 allowed Cascadia's Risso's dolphin catalog to be fully updated through 2024. Excitingly, the algorithm was able to correctly link images with year gaps as large as 14 years, even when 48 out of the 67 images involved mark changes in dorsal fin appearance. See Supporting Information for details on methods for Test 3.

3.5 | Test 4—Poor Quality Images

In Test 4, which took images of Risso's that were previously determined to be too poor quality for eye based matching, the multispecies algorithm successfully identified matches to 50 out of 145 of these images, representing 42 unique individuals. Of these, 48 appeared first in the queue of match options presented, while two of them appeared eighth in the queue of options presented. One of the matched images was noted to have undergone mark changes, and three were opposite side matches. Linear regression and ANOVA analysis did not identify any significant relationships between image quality and

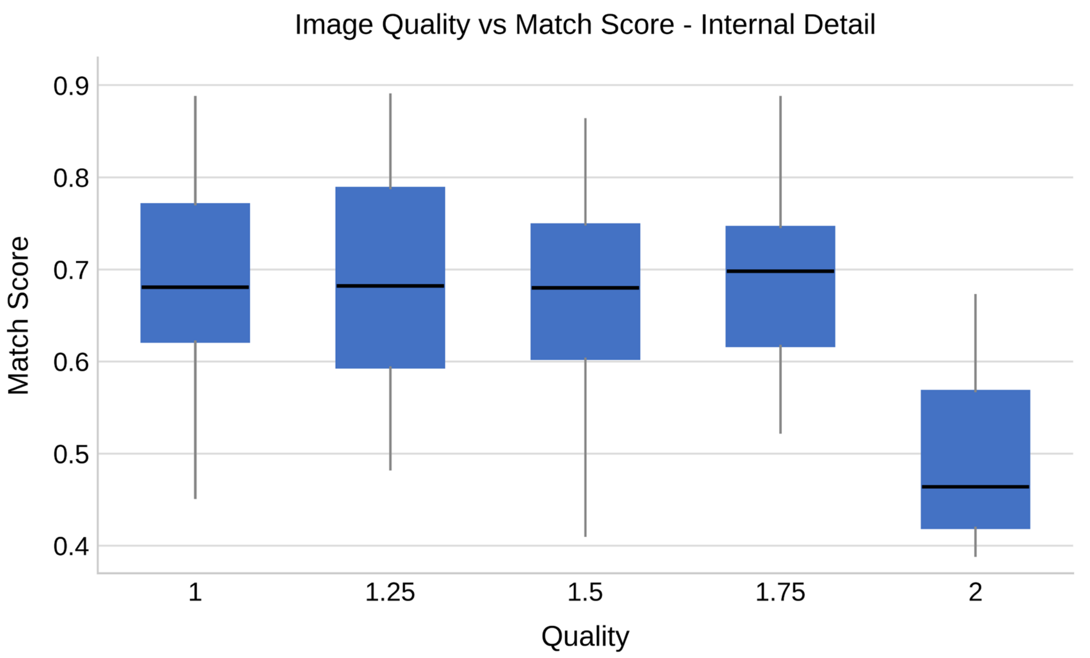

FIGURE 9 | Overall image quality (generated by Angle, Provis, Exposure, and Sharpness) in the Missed Match Rank Analysis was significant ($p=0.0028$).

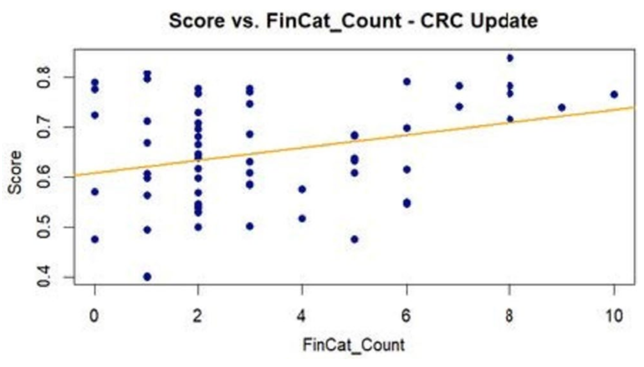

FIGURE 10 | The relationship between FinCat_Count and Score with a regression line to show the trend in the data from the linear regression model for the Updating CRC Catalog dataset.

distinctiveness metrics as related to the algorithm similarity score. See Supporting Information for details on methods for Test 4.

3.6 | Match Type Assessment Results

Overall, the algorithm and the humans performed similarly across most image types. However, the algorithm showed a slight advantage in detecting matches in images with lower similarity scores, or suboptimal quality, particularly for animals with cleaner fins and more frontal or suboptimal angles (the algorithm found 48% of images previously missed by a human at angle 2). Images missed by both the human and the algorithm (N/N) tended to have lower average quality scores, although differences were subtle and not strongly diagnostic (Table 4).

Animals with moderate to high levels of fin damage or scarring (Fincat 4+, ScarScore 3–4) were more likely to be successfully

TABLE 4 | Comparison of algorithm similarity scores between three categories of match type, where both humans and the algorithm found a match, only the algorithm found a match but humans did not, or neither humans nor the algorithm found a match.

Algorithm match	Human match	Count	Mean score	Median score
Y	Y	161	0.6884	0.694
Y	N	237	0.6725	0.670
N	N	284	0.5324	0.530

matched by both the algorithm and human, indicating distinctiveness facilitates accurate identification. Conversely, cleaner fins (Fincat or ScarScore 1–2) were more often missed by the human, with the algorithm outperforming human matchers in some of these cases. The algorithm detected 35% of clean fins that had been previously missed by humans, and 53% of animals that had a medium ScarScore but were missed by humans. The algorithm was also more successful in identifying high-quality matches (Sharpness or Exposure = 1) than human reviewers, as well as a small number of cases involving opposite side matches or marked changes not originally matched by eye. See Supporting Information for details on methods for Match Type Assessment.

4 | Discussion

The datasets compared from Cascadia's Risso's dolphin photo-ID catalog demonstrated that the multispecies algorithm was not only effective at detecting matches to known individuals already present in the catalog, but it also identified additional matches that had been missed from human matchers. This is particularly notable for Risso's dolphins given their unique morphology and progressive development of scar and pigmentation patterns over their lifetime (Hartman et al. 2008). As one of the most heavily scarred delphinids, identification depends on accurately distinguishing these dynamic and often prominent natural markings, which can change over time as animals acquire new scars and marks (Mariani et al. 2016). The algorithm's strong performance suggests positive potential for the use of this technology with Risso's dolphins over long timeframes, particularly in cases where human matchers may be challenged by cleaner fins, suboptimal angles, or image quality limitations.

Across all datasets, most correct matches appeared early in the queue of options presented, most frequently as the first or second option, while later matches typically corresponded with lower quality or off-angle images. MAP scores further confirmed the algorithm's reliability in prioritizing true matches, with an overall MAP of 0.927. For comparison, Patton et al. (2023) reported a mean MAP of 0.869 across 24 species and 39 catalogs, suggesting the algorithm's performance for Risso's dolphins is above average relative to other species. When adjusted to include a dummy "new individual" prediction for the tests that contained not found animals, MAP values decreased slightly, though it is likely not a reflection of algorithm performance, but because unmatched images were treated as potential new individuals, lowering the average precision metric. The dummy-adjusted MAP therefore provides a conservative measure of overall algorithm performance and supports the conclusion that correct matches were consistently ranked near the top of the queue. While the current dataset tested includes 1165 identified individuals, these results suggest that the algorithm is likely to remain robust as additional individuals are added, though performance for substantially larger catalogs may be influenced by image quality and fin distinctiveness.

Distinctiveness of the dorsal fin was a key determinant of algorithm similarity scores and match success across multiple tests (Test 1, Test 2b, Test 3). Individuals with more pronounced notches or irregular trailing edges achieved higher similarity scores and were more likely to be correctly matched. Both dorsal fin notches and scarification patterns were evaluated as distinctiveness metrics, but only fin notches consistently predicted algorithm success, whereas scarification was not a significant factor across tests. This trend mirrors the inherent bias seen in human-based matching, where animals with distinctive fins are more likely to be detected and re-identified, while individuals with cleaner fins may be underrepresented. This unequal detection probability is an important consideration when applying photo-ID data to downstream analysis such as mark-recapture models, where assumptions of equal capture probability can strongly influence population estimates (Rosel et al. 2011; Urian et al. 2015). Recognizing and accounting for this bias will be important for interpreting catalog-based analysis.

Image quality and fin visibility also influenced algorithm performance. The visibility of the trailing edge notches (ProVis) was a particularly strong predictor of similarity score, in addition to overall image quality, though to a lesser extent (Test 2b). These findings indicate that while general image quality matters, the visibility of the fin plays a greater role in algorithm performance. Notably, the algorithm was able to extract reliable matches from images that would typically fall below the acceptable threshold for human eye-based matching (Test 4). This ability to utilize lower quality photographs expands the pool of usable data, potentially increasing catalog completeness and the efficiency of long-term monitoring.

The comparison of algorithm and human performance highlights the potential for algorithm-based matching to both support and, in some cases, surpass traditional eye-based matching methods. While overall matching rates were similar, the algorithm demonstrated particular strengths in areas where humans missed matches, such as in cases of lower-quality images, at suboptimal angles, and with less distinctive individuals. Interestingly, the algorithm was also able to identify high-quality matches that humans missed, as well as opposite side matches and marked changes between individuals. These patterns suggest that while humans and algorithmic matching may largely overlap in performance for this species, the algorithm may be more sensitive to subtle patterns or changes, especially in individuals that may be less distinctive.

Also notable within these tests was a deviation from previous methodology used to compile images and verify matches. Prior to the algorithm's adoption, each image of a dorsal fin had to be measured manually in ImageJ to prioritize likely matches. With the algorithm's demonstrated accuracy, this step is no longer required, streamlining the data preparation process. Likewise, match verification previously relied on two people: one person to compile and match images and a second to review both matches and match "not found" animals again to ensure nothing was missed. Despite these efforts, the algorithm identified several matches that humans had missed, demonstrating that even the two-person verification process was not infallible. The results from these tests suggest sufficient confidence in the algorithm to reduce this process to a single reviewer, as a second round of manual matching no longer appears necessary.

While our evaluation focused primarily on missed matches and the image quality and distinctiveness metrics, the detection of false positives, or instances where the algorithm ranks an incorrect individual as a top match, relies on manual comparison by experienced human matchers. Consequently, users of the algorithm should be aware that while it effectively prioritizes likely matches, incorrect top-ranked assignments remain possible, particularly when adding new individuals to the catalog or when image quality is poor. An important consideration with this is the potential for bias between algorithm outputs and human decision-making. Utilizing lower quality photographs may increase false positive errors, as human reviewers may be more inclined to accept a suggested match despite low visual confidence. This underscores the importance of experienced reviewers and utilizing conservative acceptance criteria when integrating automated tools into long-term photo-ID efforts.

In the broader context of photo-ID methods, several tools have been developed to support or automate parts of the matching workflow. Cloud-based platforms such as Flukebook integrate multiple computer vision algorithms to rank candidate matches for cetacean images and manage large digital catalogs, providing ranked similarity outputs that can be reviewed manually by experts (Blount et al. 2022, Wild Me 2024). Species-targeted systems such as Dolfin, developed for Risso's dolphins, include automated photo-ID modules tuned to the characteristic features of this species' dorsal fin morphology and scarring patterns (Maglietta et al. 2018). Other tools, including R-based applications like finFindR, leverage deep learning to characterize fin edges and assist in dorsal fin matching (Thompson et al. 2021). While these platforms vary in their species scope and degree of automation, they share the goal of reducing manual effort and improving consistency in photo-ID workflows. Our evaluation demonstrates that the multispecies dorsal fin algorithm can achieve robust automated matching for a complex species like Risso's dolphins, which exhibit highly variable and progressive marking patterns. For research programs maintaining catalogs across multiple species, a unified multispecies framework offers practical advantages, including standardized data processing and consistency in matching procedures across projects and years.

This study highlights the success of the multispecies matching algorithm in both error-checking and updating Cascadia's Risso's dolphin catalog, as well as streamlining data preparation methods. By reducing the need for manual image measurements and secondary human verification, the algorithm improves the efficiency of the matching process. The results also suggest that individuals with more distinctive fins may be more likely to be matched successfully, underscoring a potential detection bias that should be considered in future applications. While human oversight remains essential, the findings indicate that a single trained matcher, supported by the algorithm, can be sufficient for catalog updates and maintenance. This shift not only improves efficiency but also strengthens confidence in the catalog's accuracy. By continuing to build on these efforts, researchers will be better equipped to track individual Risso's dolphins and enhance our understanding of their site fidelity, movement patterns, and population structure within areas like the Southern California Bight and elsewhere, which will be critical information for ongoing conservation and management efforts.

Author Contributions

Ted Cheeseman: conceptualization, methodology, resources, writing – review and editing, software. **Alexandra Vanderzee:** conceptualization, data curation, formal analysis, writing – original draft, writing – review and editing, investigation, methodology. **Philip T. Patton:** methodology, software, writing – review and editing. **John Calambokidis:** conceptualization, methodology, resources, supervision, writing – original draft, writing – review and editing, investigation.

Acknowledgments

The authors would like to acknowledge a number of people and groups that made this research possible. The SOCAL-BRS and BPRS field teams gathered data and photos, and this manuscript would not have been possible without them. Katy Foster helped collect data and photos in the field and put significant time and effort into the early stages of building the Risso's dolphin catalog. Brandon Southall and Ari Friedlander helped collect data and photos in the field and provided additional guidance throughout the early stages of the Risso's dolphin catalog. Erin Falcone gathered significant photo-ID and data in the field, built the Risso's dolphin Access database, designed the method of measuring Risso's dorsal fin proportions to categorize fin shape, and provided substantial guidance in the early stages of processing this data. Guanbo Wang and Andi Gero helped with comparing images of Risso's dolphins to update the catalog and test the multispecies matching algorithm. The SOCAL-BRS project was funded by the contracts given to Southall Environmental Associates, the Living Marine Resources Program, and the US Office of Naval Research. The BPRS project was funded by the US Office of Naval Research. The SOCAL-BRS and associated photo-ID were conducted under permit numbers 14534 and 16111 issued by the US National Marine Fisheries Service. The photo-IDs collected as part of BPRS were conducted under permit 21678 issued to John Calambokidis.

Funding

This work was supported by the Office of Naval Research, the Living Marine Resources, and the Southall Environmental Associates.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

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

References

Blount, D., S. Gero, J. Van Oast, et al. 2022. "Flukebook: An Open-Source AI Platform for Cetacean Photo Identification." *Mammalian Biology* 102: 1005–1023. https://doi.org/10.1007/s42991-021-00221-3.

Bouma, S., M. D. M. Pawley, K. Hupman, and A. Gilman. 2018. *Individual Common Dolphin Identification via Metric Embedding Learning. In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 1–6. IEEE. https://doi.org/10.1109/IVCNZ.2018.8634778.

Cheeseman, T., K. Southerland, J. Park, et al. 2022. "Advanced Image Recognition: A Fully Automated, High-Accuracy Photo-Identification Matching System for Humpback Whales." *Mammalian Biology* 102: 915–929. https://doi.org/10.1007/S42991-021-00180-9.

Durban, J. W., B. L. Southall, J. Calambokidis, et al. 2021. "Integrating Remote Sensing Methods During Controlled Exposure Experiments to Quantify Group Responses of Dolphins to Navy Sonar." *Marine Pollution Bulletin* 174: 113194. https://doi.org/10.1016/j.marpolbul.2021.113194.

Fadda, V., and S. Airoldi. 2000. "A Note on the Persistence of Dorsal Fin Scars on Risso's Dolphins Photo-Identified in the Ligurian Sea." *European Research on Cetaceans* 14: 323.

Fuller, S., S. Maggi, B. Mussi, T. Kypraios, and M. P. Pound. 2025. "Whale Vision: A Tool for Identifying Sperm Whales and Other Cetaceans by Their Flank or Fluke." *Ecological Informatics* 91: 103384. https://doi.org/10.1016/j.ecoinf.2025.103384.

Hammond, P. S. 1990. "Capturing Whales on Film–Estimating Cetacean Population Parameters From Individual Recognition Data." *Mammal Review* 20: 17–22. https://doi.org/10.1111/j.1365-2907.1990.tb00099.x.

Hartman, K. L., M. Fernandez, A. Wittich, and J. M. N. Azevedo. 2015. "Sex Differences in Residency Patterns of Risso's Dolphins (Grampus Griseus) in the Azores: Causes and Management Implications." *Marine Mammal Science* 31: 1153–1167. https://doi.org/10.1111/mms.12209.

Hartman, K. L., F. Visser, and A. J. E. Hendriks. 2008. "Social Structure of Risso's Dolphins (Grampus Griseus) at the Azores: A Stratified Community Based on Highly Associated Social Units." *Canadian Journal of Zoology* 86: 294–306. https://doi.org/10.1139/Z07-138.

Hartman, K. L., A. Wittich, J. J. Cai, F. H. Van der Meulen, and J. M. N. Azevedo. 2016. “Estimating the Age of Risso's Dolphins (Grampus Griseus) Based on Skin Appearance.” *Journal of Mammalogy* 97, no. 2: 490–502. https://doi.org/10.1093/jmammal/gyv193.

Karczmarski, L., S. C. Y. Chan, D. I. Rubenstein, et al. 2022. “Individual Identification and Photographic Techniques in Mammalian Ecological and Behavioural Research—Part 1: Methods and Concepts.” *Mammalian Biology* 102: 545–549. https://doi.org/10.1007/s42991-022-00319-2.

Maglietta, R., V. Renò, G. Cipriano, et al. 2018. “DolFin: An Innovative Digital Platform for Studying Risso's Dolphins in the Northern Ionian Sea (North-Eastern Central Mediterranean).” *Scientific Reports* 8, no. 1: 17185. https://doi.org/10.1038/s41598-018-35492-3.

Mariani, M., A. Miragliuolo, B. Mussi, G. F. Russo, G. Ardizzone, and D. S. Pace. 2016. “Analysis of the Natural Markings of Risso's Dolphins (*Grampus griseus*) in the Central Mediterranean Sea.” *Journal of Mammalogy* 97, no. 6: 1512–1524. https://doi.org/10.1093/jmammal/gyw109.

Parra, G. J., P. J. Corkeron, and H. Marsh. 2006. “Population Sizes, Site Fidelity and Residence Patterns of Australian Snubfin and Indo-Pacific Humpback Dolphins: Implications for Conservation.” *Biological Conservation* 129: 167–180. https://doi.org/10.1016/j.biocon.2005.10.031.

Patton, P. T., T. Cheeseman, K. Abe, et al. 2023. “A Deep Learning Approach to Photo–Identification Demonstrates High Performance on Two Dozen Cetacean Species.” *Methods in Ecology and Evolution* 14: 2611–2625. https://doi.org/10.1111/2041-210X.14167.

Rasband, W. S. 1997-2016. “ImageJ [Computer Software].” U. S. National Institutes of Health. https://imagej.net/ij/.

Robinson, K. P., J. M. O'Brien, S. D. Berrow, et al. 2012. “Discrete or Not So Discrete: Long Distance Movements by Coastal Bottlenose Dolphins in UK and Irish Waters.” *Journal of Cetacean Research and Management* 12: 365–371.

Rosel, P., K. Mullin, L. Garrison, et al. 2011. “Photo-Identification Capture-Mark-Recapture Techniques for Estimating Abundance of Bay, Sound and Estuary Populations of Bottlenose Dolphins Along the US East Coast and Gulf of Mexico: A Workshop Report.” NOAA Technical Memorandum NMFS-SEFSC-621.

Shane, S. H. 1994. “Occurrence and Habitat Use of Marine Mammals at Santa Catalina Island, California From 1983-91.” *Bulletin of the Southern California Academy of Sciences* 93: 13–29.

Southall, B. L., D. J. Moretti, B. Abraham, J. Calambokidis, S. L. DeRuiter, and P. L. Tyack. 2012. “Marine Mammal Behavioral Response Studies in Southern California: Advances in Technology and Experimental Methods.” *Marine Technology Society Journal* 46: 48–59. https://doi.org/10.4031/MTSJ.46.4.1.

Thompson, J. W., T. L. McDonald, L. H. Schwacke, T. R. Speakman, L. T. Quakenbush, and C. R. Smith. 2021. “Automated Recognition and Identification of Marine Mammal Dorsal Fins Using Deep Learning.” *Marine Mammal Science* 37, no. 4: 1352–1371. https://doi.org/10.1111/mms.12849.

Tyson Moore, R. B., K. W. Urian, J. B. Allen, et al. 2022. “Rise of the Machines: Best Practices and Experimental Evaluation of Computer-Assisted Dorsal Fin Image Matching Systems for Bottlenose Dolphins.” *Frontiers in Marine Science* 9: 849813. https://doi.org/10.3389/fmars.2022.849813.

Urian, K., A. Gorgone, A. Read, et al. 2015. “Recommendations for Photo-Identification Methods Used in Capture-Recapture Models With Cetaceans.” *Marine Mammal Science* 31, no. 1: 298–321. https://doi.org/10.1111/mms.12141.

Wells, R. S., and M. D. Scott. 1990. “Estimating Bottlenose Dolphin Population Parameters From Individual Identification and Capture-Release Techniques.” In *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves, 407–415. Academic Press.

Wild Me. 2024. “Flukebook [Web Application] Wild Me by Conservation X Labs.” https://www.wildme.org/flukebook.html.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** mms70159-sup-0001-Supplemental_Materials.docx.