ICES Journal of Marine Science, 2024, Vol. 81, Issue 10, 1952–1962
https://doi.org/10.1093/icesjms/fsae134
Received: 24 April 2024; **revised:** 16 August 2024; **accepted:** 11 September 2024
Advance access publication date: 7 October 2024
Original Article

Species assignment from seal diet samples using shape analyses in a machine learning framework

Monica Mion [1,*,†], **Florian Berg** [2,†], **Francesco Saltalamacchia** [1,3], **Valerio Bartolino** [1], **Johan Lövgren** [1], **Mikaela Bergenius Nord** [1], **David Gilljam** [4], **Martina Blass** [4], **Karl Lundström** [1]

[1]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Turistgatan 5, 453 30 Lysekil, Sweden
[2]Institute of Marine Research, Post Box 1870 Nordnes, 5817 Bergen, Norway
[3]Department of Biological Sciences, University of Bergen, Post Box 7803, 5020 Bergen, Norway
[4]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Skolgatan 6, SE-742 42 Öregrund, Sweden

*Corresponding author. Department of Aquatic Resources, Swedish University of Agricultural Sciences, Turistgatan 5, 453 30 Lysekil, Sweden. E-mail: monica.mion@slu.se
†Joint first authors

Abstract

The identification and quantification of prey ingested is a limiting factor in trophic ecology studies and is fundamental for assessing the impact of a predator on prey populations. Vendace (*Coregonus albula*) and whitefish (*C. lavaretus*) are two congeneric species, which are commonly preyed on by Baltic ringed seals (*Pusa hispida*). The otoliths of these two species are, however, very similar and distinguishing between them in the seal diet using visual inspection has so far been challenging. Here, otolith shape outline analyses were used in combination with machine learning techniques to discriminate between eroded vendace and whitefish otoliths from ringed seal diet samples. An experiment of *in vitro* digestion of the otoliths was performed to train a machine learning model. Our model is able to self-assign known digested otoliths back to their species of origin with >90% accuracy. Furthermore, 89% ($N = 690$) of the otoliths collected from digestive tract samples could be successfully assigned to species level, i.e. vendace or whitefish. This method is readily applicable for improved understanding of ringed seal feeding habits and predator–prey interactions, as well as large-scale applications to generate seal-predation matrix inputs for stock assessments of vendace and whitefish. Further development of the machine learning techniques to discriminate between prey species in seal and other piscivorous diets is strongly encouraged.

Introduction

Predator–prey interactions are important drivers of population dynamics in aquatic ecosystems, as they contribute to the natural mortality of fish stocks. In particular, the rate of predation is important in the assessments of interactions between piscivorous organisms and fisheries (Morisette et al. 2012). Adequate data from dietary assessments are therefore essential building blocks for the road towards an ecosystem-based management approach (ICES 2023a).

In the latter part of the 20th century, conservation efforts have led to the improvement of the status of numerous marine mammal populations worldwide (Lotze et al. 2011, Magera et al. 2013, Chasco et al. 2017). Many marine mammal species rely on fish as their main source of food. This has resulted in conflicts between piscivorous predators and fisheries, competing for the same resource (Costalago et al. 2019). This applies for the ringed seal (*Pusa hispida*) population of the northernmost part of the Baltic Sea, i.e. the Bothnian Bay, in the International Council for the Exploration of the Sea (ICES) subdivision 31, which together with the Bothnian Sea (ICES subdivision 30) forms the Gulf of Bothnia (Fig. 1). The Baltic ringed seal population decreased drastically during the 1900s due to intense hunting and, in the 1960s–1970s, due to physiological limitations, low reproduction in particular, induced by environmental pollutants (Hårding and Härkönen 1999, Kokko et al. 1999). Subsequent conservation measures, however, and decreasing levels of pollutants have enhanced a process of recovery of the ringed seal population and resulted in a continuous increase in abundance during the last three decades (Hårding et al. 2007, Härkönen et al. 2013). Following the recovery of the ringed seal population, which is the most numerous seal species in the area, seal–fisheries interactions have increased, and concerns have been raised regarding conflicts between seal conservation objectives and the interests of the fishing industry. As a consequence of the seal–fishery conflict, a regional hunting quota of ringed seals has been allowed in both Sweden and Finland in the Bothnian Bay, with ∼300–500 seals being reported shot annually in recent years (ICES 2023b).

The prey choice of ringed seals in the Gulf of Bothnia is dominated (>70% in weight) by herring (*Clupea harengus*) and vendace (*Coregonus albula*) (Suuronen and Lehtonen 2012, Lundström et al. 2014). These two species, together with salmon (*Salmo salar*), also represent the main fish species of commercial interest in the area, with the first two accounting for >90% of the total economic value of the Swedish fishery catches (Rudstam et al. 1994, Bergenius et al. 2018, Björkvik et al. 2020, Lehtonen et al. 2023).

The Bothnian Bay vendace stock has been assessed by the Swedish University of Agricultural Sciences (SLU) since the late 1990s. Since 2015, the ringed seal predation has been included as a separate source of mortality in the stock assessment but is deemed to be highly uncertain for several rea-

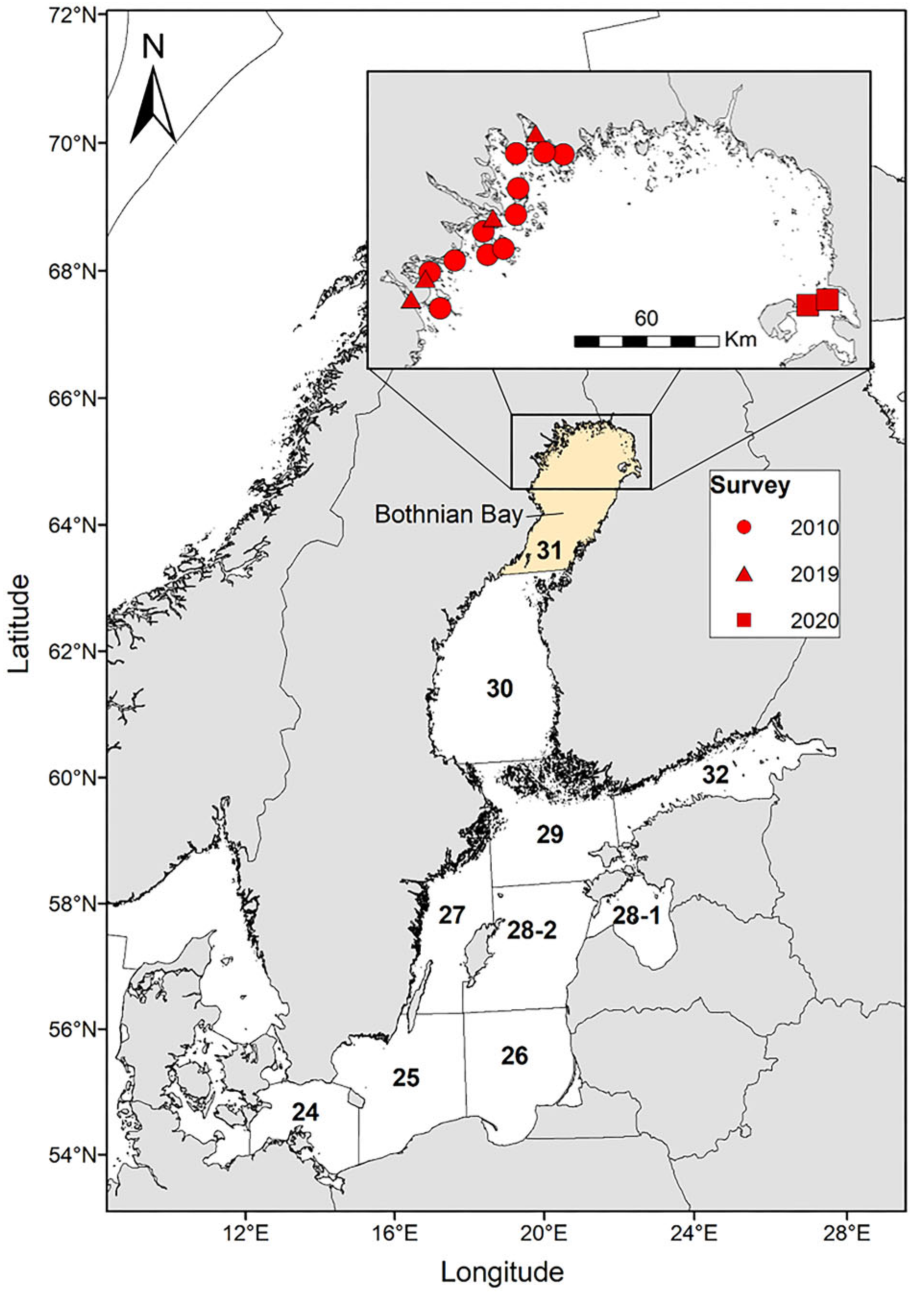

Figure 1. Geographical distribution of sampling locations for otoliths collected from whitefish during the 2010 survey and vendace during the 2019–2020 surveys. ICES subdivision numbers are reported in the map.

sons, including potential species misclassification in the seal diet samples (Fiskbarometern 2024). Improved information about the relative contribution of vendace in the ringed seal diet is therefore crucial.

Diet studies of piscivorous species (i.e. birds, mammals, and fish) traditionally rely on observations of diet sample contents to identify the species of prey and estimate prey size (Nielsen et al. 2018). Sagittal otoliths are one of the most commonly encountered prey remains in diet samples and are used for both species identification and size estimation of the consumed prey. Regression equations are routinely used to back-calculate prey body size (length and weight) based on otolith characteristics (Härkönen 1986, Leopold et al. 2001). However, otoliths found in diet samples are mostly a mixture of different species, making their identification labor-intensive. Identifying taxa with similar otolith characteristics has been difficult, and sometimes even impossible, down to the species level. Furthermore, the digestive erosion of otoliths can represent a challenge for taxonomic identification (Christiansen et al. 2004). These challenges motivate the need to look for alternative protocols and make use of new powerful approaches such as those offered by artificial intelligence and deep learning.

The species of the *Coregonus* genus have very similar otoliths (Kauhala et al. 2011). Distinguishing between vendace (*C. albula)* and common whitefish (*C. lavaretus*) otoliths in ringed seal diet samples has so far been challenging using direct observations, and may have introduced a bias to the re-

Table 1. Overview of vendace and whitefish specimens used in this study.

	Vendace				Whitefish			
	N individuals	*N* otoliths left/right	*N* images	Total length cm (median)	*N* individuals	*N* otoliths left/right	*N* images	Total length cm (median)
Original dataset	251	183/190	373	10.8–17.7 [15.2]	164	127/124	251	7.8–18.4 [13.6]
Eroded dataset	52	27/28	304	10.8–16.0 [14.3]	50	25/25	296	9.1–14.8 [13.5]
Final baseline	251	183/190	677	10.8–17.7 [15.2]	164	183/190	547	10.8–17.7 [15.2]

Total number of individuals, otoliths (left/right), and images used in the different datasets are provided. Note that individuals included in the eroded dataset were also part of the original dataset. Each otolith in the eroded dataset was photographed up to six times at different time intervals. The table also provides the range and median of the total length of the fish sampled.

sults, i.e. the number of vendace and whitefish in the seal diet may have been overestimated or underestimated.

Image-based analysis of otoliths has proven to be a successful method in differentiating otoliths of the same species belonging to different stocks (Campana and Casselman 2011), and it has also been successfully applied for developing and improving the analysis of prey remains (i.e. otoliths) in diet samples in more recent studies (de Carvalho et al. 2019, Stock et al. 2021). As morphometric and morphological parameters can be collected automatically by processing digital images of otoliths, these procedures offer several advantages over traditional methods. These automated procedures are not only faster but also more objective and accurate, ensuring precise measurements. Moreover, they are well-suited for analytical data-driven approaches, ultimately leading to improved accuracy, reduced bias, and enhanced assignment precision. Consequently, image-based analyses make the process from data collection to data analysis more efficient and reliable (Myers et al. 2020, Vabø et al. 2021), and it is expected to represent a major advantage, especially for time-consuming and prone to subjectivity data collections such as prey species identification from the analysis of stomach contents.

In this study, we applied, for the first time, otolith shape analysis in combination with machine learning techniques to discriminate otoliths collected from the digestive tracts of marine mammals. In addition, an experiment of *in vitro* digestion of the otoliths was performed to simulate the erosion process of otoliths occurring in the seal stomachs so that otolith shape variability resulting from erosion could be included in the final baseline. This novel approach represents a significant advancement in the methodology of dietary assessments by providing a more accurate and objective method for species identification. Our study has a direct relevance for fisheries management in the Bothnian Bay, especially in the context of an ecosystem approach, where a correct assessment of the vendace stock requires an accurate quantification of the seal predation. The approach proposed in our study can be readily applied to the assessment of other stocks and areas with available stomach data where seal predation plays a significant role in fish stock dynamics, as seen in other ecosystems. Improving the precision of prey species identification in the diet of seals and other marine mammals is crucial for accurately understanding and quantifying their trophodynamics. This, in turn, could support a balanced approach between conservation objectives and the interests of the fishing industry.

Materials and methods

Data collection and preprocessing

Stored dry otoliths of vendace and whitefish were available and retrieved from the archives at the Department of Aquatic Resources, SLU. All the whitefish otoliths were sampled during the vendace survey in the Gulf of Bothnia in 2010, while the vendace samples were collected during the 2019−2020 vendace surveys (Fig. 1). The vendace survey is conducted in the northwestern part of ICES subdivision 31 during the vendace trawl-fishing season in October. The samples were selected according to the size range of *Coregonus* spp. (derived from otoliths remains) usually found in the Gulf of Bothnia ringed seal digestive tracts (Lundström et al. 2014). The total fish length of the vendace and whitefish selected for this study ranged respectively from 10.8 to 17.7 cm (median 15.2 cm) and from 7.8 to 18.4 cm (median 13.6 cm, Table 1). Only otoliths without signs of damage were included in this study, and individual images of the whole left and/or right otolith were taken under reflected light with a LEICA DFC320 digital camera mounted on a LEICA stereo microscope MZ16 FA, using the software LEICA Application Suite 4.1.0 (build 1264). The otoliths were all positioned with the *Sulcus acusticus* facing down. Acquisition parameters such as shutter speed, aperture, white balance, and sensitivity (ISO) were kept uniform between all samples. Images from 251 individuals of vendace (183 left and 190 right) and 164 individuals of whitefish (124 left and 127 right) were taken. These individuals represent the original dataset (Table 1).

Furthermore, we performed an *in vitro* digestion of the otoliths to simulate the digestive erosion found in seal digestive tracts (Tollit et al. 1997, Christiansen et al. 2004). Otoliths were maintained in a solution of pH 1.5 hydrochloric acid at a constant temperature of 37°C for 30 min intervals up to 150 min ($T1 = 0$ min, $T2 = 30$ min, $T3 = 60$ min, $T4 = 90$ min, $T5 = 120$ min, $T6 = 150$ min; where T1−T2 correspond to the first erosion stage, T3−T4 correspond to the second erosion stage, and T5−T6 correspond to the third erosion stage; Fig. 2). The third erosion stage has a broad spectrum of erosion effects observed on the otoliths. This stage encompasses erosion patterns that extend beyond those observed during the *in vitro* digestion experiment. Specifically, it includes otoliths that exhibit erosion characteristics that make visual species identification difficult or impossible. The pH 1.5 was chosen based on previous experiments showing that otoliths started to erode at similar pH levels (Leopold et al. 2001, Christiansen et al. 2005, de Carvalho et al. 2019). After the completion of each experimental batch (T1, T2, T3, T4, T5, and T6), the otoliths were washed with distilled water and photographed. Otoliths from 52 vendace and 50 whitefish were chemically eroded comprising the eroded dataset (Table 1), to mimic the three different erosion stages encountered in otoliths found in seal digestive tracts (see Tollit et al. 1997 for erosion stage classification). For three vendace, both left and right otoliths of the same individuals were used resulting in 105 otoliths in total (55 otoliths for vendace and 50 otoliths for whitefish). All 102 individuals were part of the original uneroded dataset. Infor-

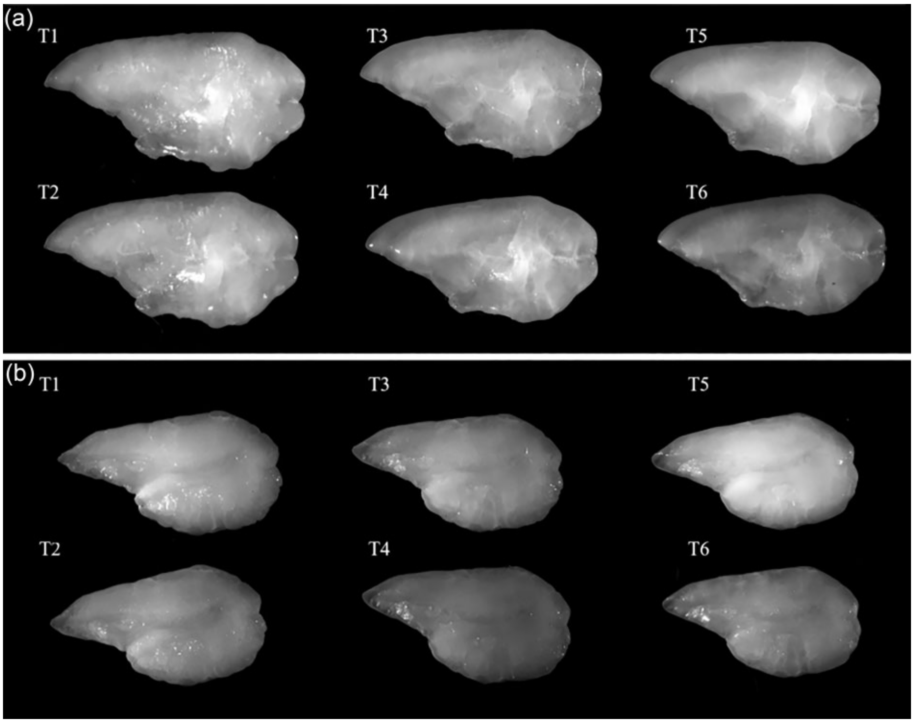

Figure 2. Pictures of (a) *C. albula* and (b) *C. lavaretus* otoliths taken at 30 min intervals between 0 and 150 min (T1 = 0 min, T2 = 30 min, T3 = 60 min, T4 = 90 min, T5 = 120 min, and T6 = 150 min). T1–T2 correspond to the first erosion stage, T3–T4 correspond to the second erosion stage, and T5–T6 correspond to the third erosion stage.

mation about the erosion stage was not used in the machine learning model given the complexity of assigning the erosion stage to digestive tract samples, characterized by subjectivity and a lack of validation (Tollit et al. 1997, 2004, Leopold et al. 1998, Byrd et al. 2020).

Lastly, otoliths found in 30 ringed seal digestive tracts and identified as *Coregonus* spp. were used in the study to assign them to the species level. In total, 778 *Coregonus* spp. otoliths found in seals collected between 2008 and 2020 in the Gulf of Bothnia were photographed.

Data analysis

The otolith shape analyses were conducted in accordance with the procedure outlined by Libungan and Pálsson (2015). The next steps were applied to all otolith images of the original and eroded dataset as well as images of otoliths found in ringed seal digestive tracts. To ensure comparability with the left otoliths and facilitate direct comparison, the right otoliths were horizontally mirrored using the GNU Image Manipulation Program (GIMP v. 2.8.22) before extracting the otolith outlines. All the following analyses were conducted in R (R Core Team 2022, version 4.2.2). The outlines of the otoliths were extracted from the images using the "shapeR" R-package developed by Libungan and Pálsson (2015). The detected outlines were overlaid on the original pictures and visually evaluated to ensure that each outline closely traced the edge of the otolith. When the outline was not accurate, the original picture was manually edited with the program GIMP to increase the contrast between the background and the otolith, or to remove organic material attached to the otolith that could disrupt the shape outline. For each otolith, the extracted X–Y coordinates of the otolith shape outline were smoothed to reduce pixel noise and transformed to 64 wavelet coefficients. The 64 wavelet coefficients were used subsequently to either discriminate vendace and whitefish otoliths, or to assign the *Coregonus* spp. from the digestive tracts of ringed seals to the species level.

Using the wavelet coefficients of the original dataset, an analysis of variance was used to investigate the potential directional asymmetry in otoliths (left–right side) within species. The results were non-significant (ANOVA, $P > .05$) and both left and right otoliths were used in the analyses.

The analytical framework was built using the "assignPOP" R-package (Chen et al. 2018) (see Fig. 3 for an overview of the framework). This framework consists of six key steps: (i) data collection, (ii) otolith shape extraction and wavelet transformation, (iii) build the baseline, (iv) evaluate the baseline, (v) build the predictive model, and (vi) species assignment. The "assignPOP" R-package provides all functions that were applied for the cross-validation and species assignment. It also allows for the application of multiple statistical classification models such as linear discriminant analysis, random forest, or support vector machines (Chen et al. 2017). First, a baseline was established, which will be used to accurately assign a species to otoliths collected from ringed seal digestive tract samples. The final established baseline consisted of wavelet coefficients extracted from the original dataset (uneroded otoliths) and the eroded dataset (*in vitro* eroded otoliths; Table 1). In total, the wavelet shape coefficients of 1224 different otolith images were included in the final baseline. The combination of the original and eroded datasets had no influence on the species assignment itself, but slightly influence the assignment probability (see Results).

A Monte-Carlo cross-validation analysis (Xu and Liang 2001) was used to test the accuracy of species assignment of the final baseline (Chen et al. 2017). The data were split

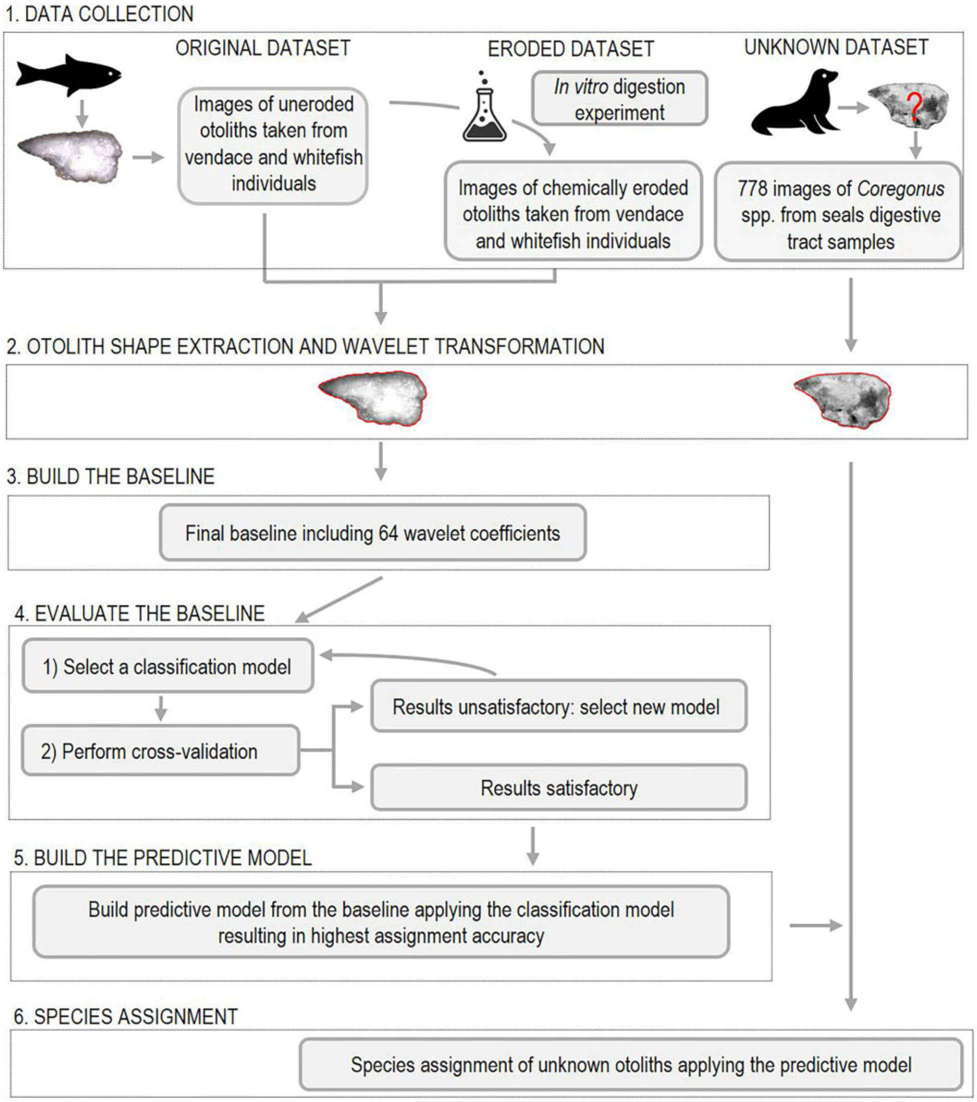

Figure 3. Overview of the analytical framework used in this study.

into training- and test-groups to assess the self-assignment accuracy of the final baseline. With the Monte-Carlo cross-validation, data were recursively split into 50, 70, and 90% to be used as training set, with the remaining samples 50, 30, and 10% allocated to the test set, respectively. The Monte-Carlo cross-validation samples new random individuals at each iteration ($n = 100$), thus does not guarantee that every individual is sampled. Therefore, a *K*-fold cross-validation (Rodriguez et al. 2010) was applied in addition (results shown in the Supplementary Material). In the *K*-fold method, samples from each group (vendace and whitefish) were randomly divided into *K* groups ($K = 3$, 4, and 5) (Chen et al. 2017). One group from each species is used as test set and the remaining *K*-1 groups are used as training sets, to estimate the self-assignment accuracy.

During the cross-validations, all classification models available in "assignPOP" were tested as suggested by Smolinski et al. (2020), who also describe the most common classification models. Among these, the support vector machine (SVM) model resulted in the highest self-assignment accuracy (~2%–3% higher) and was therefore selected to build the predictive model for assigning unknown otoliths from seal digestive tracts. The classification model was applied with its default settings and without hyperparameter tuning (Meyer et al. 2015). SVM models are particularly well-suited for classification tasks involving high-dimensional data, such as the 64 wavelet coefficients used in our study, and where the groups, such as different species, exhibit overlapping characteristics (Cristianini and Shawe-Taylor 2000, Ben-Hur et al. 2008). SVM models operate by identifying the optimal hyperplane that maximizes the margin between different groups, and achieve this by using kernel functions to project the predictive variables into a higher-dimensional feature space, enhancing the separation of groups compared to the original input space (Cortes and Vapnik 1995).

The predictive model for the species assignment was based on the final baseline using SVM as classification model. The predictive model was applied to assign the 778 otoliths found in 30 ringed seal digestive tracts. During the assignment, the predictive model estimated the posterior probability of membership (assignment probability) to each of the two baseline species (vendace and whitefish) for every individual otolith from the seal digestive tracts. Otoliths with an assignment probability <0.75 were considered as unknown species, i.e.

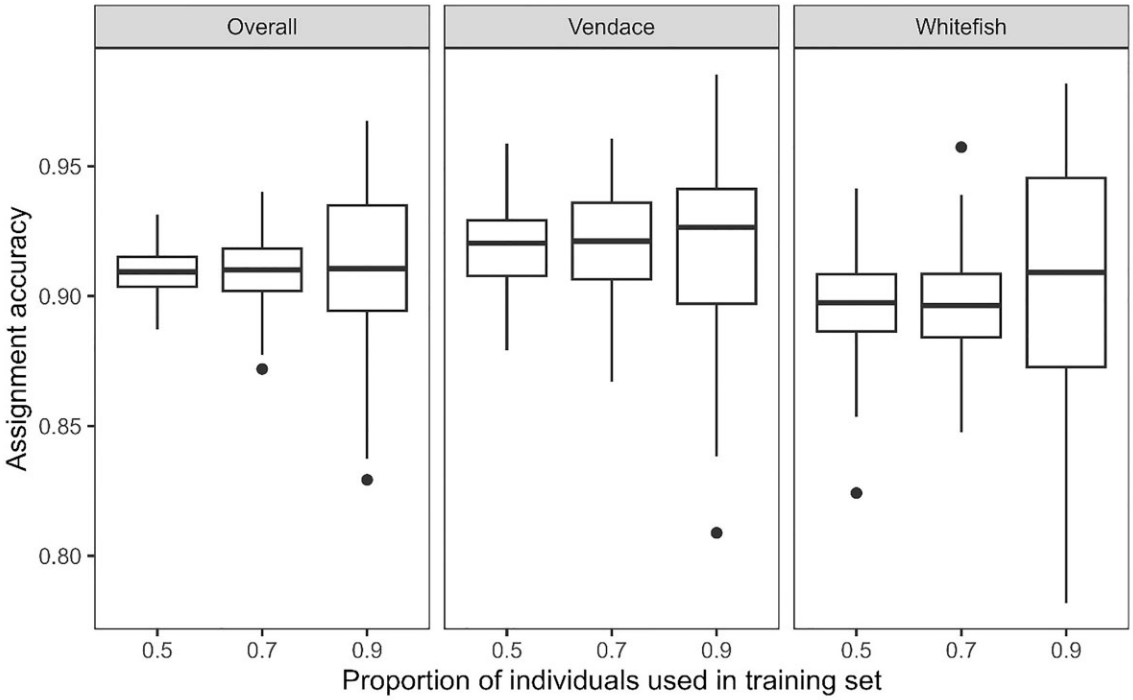

Figure 4. Self-assignment accuracies estimated through a Monte-Carlo cross-validation with random sampling of three levels of training individuals (50, 70, and 90%). Box plot details: the line within the box is the median; the top and bottom edges of the box are the 25th and 75th percentiles, respectively; the ends of the whiskers represent the minimum and maximum of non-outliers, and outliers are shown as black circles.

Coregonus spp., introducing uncertainties in the calculation of vendace and whitefish proportions in the ringed seal diet. The threshold probability of 0.75 was chosen as it provides a practical balance between minimizing misclassifications and ensuring that a sufficient number of samples could be confidently assigned to a species (Trites and Joy 2005). From a statistical point of view, the threshold probability of 0.75 indicates a situation where one outcome is three times as likely as the alternate outcome [odds = p/(1-p)]. This threshold was deemed appropriate for our study because it allows for the inclusion of a significant portion of the data, while still maintaining a reasonable level of confidence in species identification. To quantify the uncertainty of assignments, a range of min−max proportions for vendace and whitefish was calculated, i.e. proportions of unassigned otoliths were classified as either all vendace or all whitefish.

Otoliths from the *in vitro* digestion experiment were used to test if the erosion time influenced the assignment probability. Otolith wavelet coefficients of the eroded dataset were assigned to its species based on the uneroded otoliths (original dataset) using the same predictive model as above. Further, only otolith images of individuals being correctly assigned to their known species before the experiment (T1 = 0 min; 53 out of 55 vendace and 43 out of 50 whitefish) were included in the following modeling approach. A linear mixed-effects model was applied to investigate the effect of the erosion time and species and their interaction on the assignment probability. Since each individual was assigned multiple times (once for each erosion time step), the individual IDs were added as random factors. The mixed-effects model was fitted using the “lme” function within the “nlme” R-package (Pinheiro and Bates 2000). A backward selection approach was used incorporating all fixed and random effects. Non-significant model terms were excluded. First, the optimal structure of the random effects was tested using likelihood ratio test based on the models fitted by restricted maximum likelihood estimations (REML) (Zuur et al. 2009). Also based on REML fits, the fixed effects structure was optimized using marginal *F*-statistics (Pinheiro and Bates 2000). Both the random effect and the residual of the final model were normally distributed with mean of 0 and variance σ^2.

Results

The total fish length of the vendace and whitefish selected to extract the otoliths for building the baseline for the machine learning model ranged, respectively, from 10.8 to 17.7 cm (median 15.2 cm) and from 7.8 to 18.4 cm (median 13.6 cm).

The overall self-assignment accuracy of the predictive model based on uneroded and eroded otoliths combined was >90% based on both the Monte-Carlo cross-validation (Fig. 4) and the *K*-fold cross-validation (Supplementary Fig. S1) and was thus considered satisfactory for assigning a species to unknown *Coregonus* spp. otoliths from digestive tract samples.

The results of the membership probability, i.e. probability to be assigned as a vendace or whitefish species, for each otolith sampled in the Bothnian Bay ringed seal digestive tracts are visualized in Fig. 5. The predictive model was able to identify 89% ($N = 690$) of the otoliths collected from digestive tract samples either as vendace or whitefish with an assignment probability >0.75. For 11% ($N = 88$) of the otoliths the assignment probability <0.75 and these were therefore not assigned to a species. Higher threshold probabilities, such as 0.90 or 0.95, while statistically more conservative, would have significantly increased the number of unassignable samples (N of unassigned samples = 235 and N of unassigned samples = 356, >778, respectively). Overall, the diet of ringed

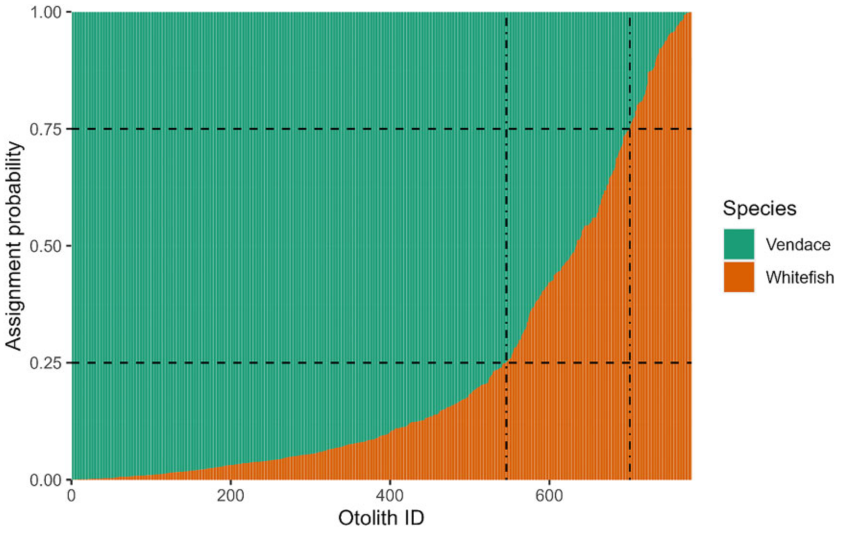

Figure 5. Membership probability of each otolith found in the sampled ringed seal digestive tracts to be assigned as vendace (green) or whitefish (orange). Horizontal dashed lines indicate the threshold probability (0.75) used for unknown individuals, which are in between the two vertical dot dashed lines. These were not considered in the estimation of diet proportions.

seals consisted of minimum 70% ($N = 545$) or maximum 81% vendace and minimum 21% ($N = 145$) and maximum 30% whitefish. The maximum ranges assume that the unassigned individuals are either classified as vendace or whitefish, respectively. The minimum proportion of vendace by seal sample varied between 0 and 100%, with an average of 70% (±20% standard deviation; Supplementary Fig. S2).

The *in vitro* digestion experiment demonstrated that the erosion time has a significantly negative effect on the assignment probability ($P < .001$; Fig. 6). The effect was the same for vendace and whitefish. However, even after an erosion time of 150 min, corresponding to considerably digested otoliths, the assignment probability was still above the used threshold probability of 0.75 for both species. Thus, the species assignment probabilities can be considered to be robust until the third erosion stage (T5–T6 = 120–150 min).

Discussion

In this study, a machine learning approach in combination with otolith shape analysis was applied for the first time to seal diet samples. This represents an innovation in the field of dietary assessments. Our research provides a readily applicable and accurate framework for the taxonomic identification of otoliths from digestive tract content analysis, representing a significant advancement in overcoming the limitations of traditional visual inspection methods. The integration of image-based analysis and machine learning in our study finds immediate application to improve the assessment of predation mortality on the vendace stock in the Bothnian Bay, which is a key factor in the assessment and scientific advice on this economically important fishery (Lövgren et al. 2022). By providing a more reliable method for species differentiation in stomach contents, our study contributes to provide a better scientific advice in support of decision-making for management of more sustainable fisheries.

Dietary assessments are crucial for understanding predator–prey interactions in aquatic ecosystems and are essential for ecosystem-based management (Pikitch et al. 2004, FAO 2008). In piscivores diet studies, only direct visual techniques such as digestive tract content analysis, as opposed to indirect techniques such as fatty-acid composition analyses or DNA metabarcoding, permit the identification of different life stages of prey and allow straightforward quantification of prey proportions and estimation of prey sizes (Nielsen et al. 2018). There are mathematical models that estimate proportions of prey using fatty acid or stable isotope compositions, which provide valuable insights into trophic relationships but may not accurately identify specific prey species or differentiate between closely related species (Iverson et al. 2018, Jackson et al. 2021). DNA-based methods, although powerful, can often only provide semi-quantitative results, making it challenging to accurately quantify prey proportions compared to direct visual analysis (Deagle et al. 2019, Mackenzie and Buckland 2022).

Otoliths retrieved from digestive tract contents have long served as a valuable tool for dietary analysis (Nielsen et al. 2018). However, relying on visual inspection for species identification presents several challenges and limitations. This method demands significant taxonomic expertise and training and can be labor-intensive and time-consuming. Moreover, it lacks a data-driven approach, making it unsuitable for routine quantification of uncertainty in assignment. These challenges are compounded by otolith degradation within the digestive system, increasing the risk of errors and potential bias in estimating the diet composition of piscivorous predators (Prime 1979, da Silva and Nielsen 1985, Jobling and Breiby 1986, Jobling 1987). Here, we provide a readily applicable accurate framework for the taxonomic identification of otoliths from digestive tract content analysis. Our application of image-based analysis and machine learning techniques to discriminate otoliths and assign species, represents a significant advancement in overcoming these limitations. By automatic extraction of morphometric parameters from otolith images, our method aims to increase accuracy, objectivity, and reliability of dietary assessments. Additionally, the *in vitro* digestion experiment to simulate the erosion stages that otoliths undergo in the seal digestive tract provided a baseline for the machine learning model. This innovative methodology allows for the discrimination between vendace and whitefish otoliths, pro-

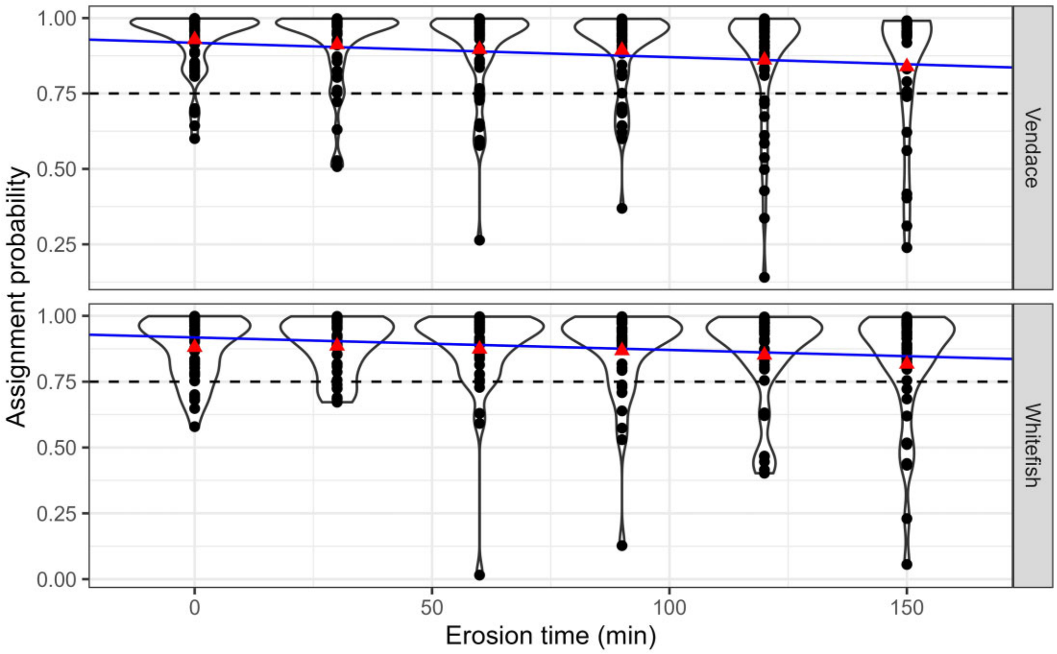

Figure 6. Violin plots of the assignment probabilities of otoliths used in the *in vitro* digestion experiment over time for both species, vendace and whitefish. Points are the individual assignment probabilities, triangles are the mean probabilities for each time step, and the solid line represents the prediction line of the linear mixed-effect model.

viding an accurate species assignment to otoliths of the genus level *Coregonus* spp.

The use of machine learning in ecological research is a growing trend (Rubbens et al. 2023), and our work contributes to this field. By applying the SVM classification model, we made an optimal use of automatically recorded shape descriptors here in the form of wavelet coefficients. This advancement can significantly speed up the labor-intensive process of manually categorizing otoliths, reducing error rates, and increasing the overall efficiency and accuracy of the process. It should be noted that in this specific case to discriminate between vendace and whitefish otoliths, the machine learning classification model provided the best results. Other more traditional classification models, such as linear discriminant analysis or random forest, were also tested but did not achieve robust self-assignment accuracy as recommended by Smolinski et al. (2020). In contrast to SVM, other models may not always be as effective in creating sharp decision boundaries, particularly in complex and high-dimensional datasets (Breiman 2001). SVM models, by focusing on the most informative data points, can achieve better classification performance, even in the presence of noisy or overlapping data (Ben-Hur et al. 2008).

The predictive model, using otolith shape coefficients and a machine learning classification model, has demonstrated remarkable self-assignment accuracy, with success rates >90% in both Monte-Carlo and *K*-fold cross-validations. This can greatly improve the accuracy of dietary assessments in piscivorous organisms. By combining uneroded and eroded otoliths in the final baseline, the overall self-assignment probabilities decrease by <1% compared with uneroded otoliths only, but it strengthens the baseline, improving the assignment probabilities of the eroded otoliths from seal digestive tracts. One caveat that should be mentioned is the temporal mismatch of baseline samples for the two species, where vendace otoliths were sampled 9–10 years later compared to whitefish otoliths. However, it has been demonstrated that otolith shape has a high temporal stability over decades (Denechaud et al. 2020). Therefore, we do not expect that the time lag will influence the presented results significantly.

The predictive model was able to identify the species of ∼89% of the otoliths from seal digestive tract samples with a high assignment probability (>0.75). This threshold probability was selected to balance the trade-off between losing too many unidentified samples and statistical power (Trites and Joy 2005). Not being able to assign ∼11% of the samples, and at the same time having the statistical power that the assigned species is three times as likely as the alternative, seems like a practical and robust setting to draw ecological conclusions on trophic interactions considering the high uncertainty that usually accompany stomach content data (Liu et al. 2005). Higher threshold probabilities would reduce the number of assigned individuals drastically, and thereby limiting our ability to draw meaningful ecological inferences from the data. This choice is particularly relevant in dietary assessments where the goal is to maximize prey species identification while maintaining confidence in the results. The threshold effectively reduces the number of unassigned otoliths, thus decreasing the uncertainty in species composition estimates, which is crucial for accurate ecological interpretations. The average proportion of otoliths assigned to vendace and whitefish (79% and 21%, respectively) in the seal diet provides valuable insights into the dietary preferences of the Baltic ringed seal in the Bothnian Bay. The application of these methods may thus affect the accuracy of the assessment of vendace in the Bothnian Bay.

Following historical challenges, including intense hunting and environmental pollutants, the Baltic ringed seal population has recovered during the last decades. This highlights the importance of monitoring the diet accurately of this key top predator in the Gulf of Bothnia ecosystem. The ability to correctly differentiate between vendace and common whitefish otoliths in the seal diet samples is crucial for both ecological understanding and fisheries management. The challenges posed by visually similar otolith characteristics have, in the past, introduced uncertainty and bias into the diet results.

It is only since 2015 that seal predation is accounted in the stock assessment of vendace in the Bothnian Bay, showing how this could represent an important driver in the temporal dynamics of this fish stock. This has a direct impact on the calculation of the proportion of vendace and whitefish in ringed seal diets, which is essential for the advice and management of the vendace stock. In the current stock assessment of vendace in the Bothnian Bay, otoliths constitute the large majority of prey remain items used for the identification of ringed seal prey species. To overcome the difficulties of correctly identifying *Coregonus* spp. as either vendace or whitefish, a ratio of 80/20% of all *Coregonus* spp. otoliths found in seal digestive traits has been used for identification of vendace and whitefish, respectively. This ratio was calculated based on preliminary results of this study presented during the latest benchmark of vendace (Lövgren et al. 2022). This study corroborates these preliminary findings and highlights the importance of correctly identifying *Coregonus* spp. in the ringed seal diet, as predation mortality from ringed seals is for some years estimated to be up to three times larger than the mortality induced by the fishery and thus having a large impact on the vendace stock dynamics. Classifying all *Coregonus* spp. as vendace would in this particular case result in an overestimation of the impact of ringed seals' predation on the stock, potentially resulting in lost fishing opportunities for the fishery.

Although frequently used to determine the diet of piscivores, such as seals, inference based on otoliths in the digestive tracts has limitations (Pierce and Boyle 1991, Bowen and Iverson 2013). First, otoliths may undergo partial erosion or complete digestion, posing difficulties in identifying prey at the species level or rendering species detection impossible (Prime 1979, da Silva and Neilson 1985, Jobling and Breiby 1986). Our results show that although the erosion time negatively affected the assignment probabilities, those were still >75% threshold also in well-digested otoliths. Thus, these methods can help assigning otoliths that are also moderately eroded, decreasing the proportion of otoliths considered unidentifiable. Second, the data's accuracy is contingent on the prey's gut passage time, providing a snapshot of the seal's diet only within a few hours or days before sampling. Consequently, the information primarily reflects prey consumed in the proximity of the collection site or during the sampling period (Prime 1979, Grellier and Hammond 2006). While this limitation may pose challenges in extensive oceanic study systems and for species engaged in long-distance foraging migrations, it is deemed less problematic in the semi-enclosed Baltic Sea (Scharff-Olsen et al. 2019).

Further analyses of additional diet samples collected regularly throughout the year are crucial to enhance the representativeness of our findings. Diet sampling has a central role in understanding trophic interactions, albeit it necessitates a substantial number of samples to yield meaningful results and enable inference drawing. Additionally, exploring potential variations in the proportions of vendace and whitefish based on various biological characteristics of seals (such as weight and age), different quarters of the year, geographical areas, and relative amounts of whitefish and vendace in the ecosystem is recommended. Such a comprehensive analysis would provide a deeper understanding of trophic interactions between these species, contributing significantly to the broader ecological understanding. Moreover, it is worth considering the potential of our approach to semi-automate prey species identification beyond *Coregonus* otoliths. By training the algorithm to identify all major prey species of ringed seals, the machine learning could potentially provide species assignments for all otoliths found in the digestive trait from a picture, not limited to vendace and whitefish.

In conclusion, this study demonstrates the utility of combining otolith shape analysis with machine learning to overcome the challenge of distinguishing between similar prey species in dietary assessments. Our approach is novel in its application of machine learning techniques to the image analysis of marine mammals' stomach contents, with direct implications for fisheries advice. By improving the accuracy of seal diet studies, particularly in the identification of *Coregonus* species, our methodology can have a direct impact on the management of fish stocks such as vendace in the Bothnian Bay. This contribution is essential for balancing the ecological role of predators like the ringed seal with the economic importance of fisheries, ensuring sustainable use of marine resources. Further research and application of this methodology are encouraged to refine and expand its utility and increase our understanding of predator–prey relationships and ecological dynamics.

Acknowledgments

We thank the many individuals who have contributed to collecting the samples on which this manuscript is based. We also thank Malin Karlsson (SLU Aqua, Lysekil) who has contributed to process the ringed seal stomach samples and gave inputs on a previous version of this manuscript.

Author contributions

M.M., F.B., F.S., and K.L. contributed to the conception and design of the study. M.M., F.S., K.L., and M.B. contributed to the data acquisition. M.M. and F.B. conducted the statistical analysis and wrote the first draft of the manuscript. K.L. provided funding. All authors contributed to the manuscript's revision and approved the final version of the manuscript.

Supplementary material

Supplementary data is available at *ICES Journal of Marine Science* online.

Conflict of interest: None declared.

Funding

V.B.'s contribution was funded by Formas research grant 2021-00826. Collection of seal samples and K.L.'s contribution were funded by Leader Fiskeområde Tornedalen Haparanda skärgård 2020 (the European Maritime, Fisheries, and Aquaculture Fund) and the Swedish Agency for Marine and Water Management.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

Ben-Hur A, Ong CS, Sonnenburg S *et al.* Support vector machines and kernels for computational biology. *PLoS Comput Biol* 2008;**4**:e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Bergenius M, Ringdahl K, Sundelöf A *et al. Atlas över svenskt kust- och havsfiske 2003–2015*. Aqua Reports. Drottningholm Lysekil Öregrund: Institutionen för akvatiska resurser, Sveriges lantbruksuniversitet, 2018. https://res.slu.se/id/publ/94302

Björkvik E, Boonstra WH, Hentati-Sundberg J *et al.* Swedish small-scale fisheries in the Baltic Sea: decline, diversity and development. In: J. Pascual-Fernández, C. Pita, M. Bavinck (eds), *Small-scale Fisheries in Europe: Status, Resilience and Governance*. Heidelberg: Springer Nature, MARE Publication Series, 2020, pp. 559–79.

Bowen WD, Iverson SJ. Methods of estimating marine mammal diets: a review of validation experiments and sources of bias and uncertainty. *Mar Mammal Sci* 2013;**29**:719–54. https://doi.org/10.1111/j.1748-7692.2012.00604.x

Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32. https://doi.org/10.1023/A:1010933404324

Byrd BL, Hohn AA, Krause JR. Using the otolith sulcus to aid in prey identification and improve estimates of prey size in diet studies of a piscivorous predator. *Ecol Evol* 2020;**10**:3584–604. https://doi.org/10.1002/ece3.6085

Campana S, Casselman J. Stock discrimination using otolith shape analysis. *Can J Fish Aquat Sci* 2011;**50**:1062–83. https://doi.org/10.1139/f93-123

Chasco BE, Kaplan IC, Thomas AC *et al.* Competing tradeoffs between increasing marine mammal predation and fisheries harvest of Chinook salmon. *Sci Rep* 2017;**7**:15439. https://doi.org/10.1038/s41598-017-14984-8

Chen K-Y, Marschall EA, Sovic MG, Fries AC, Gibbs HL, Ludsin SA. assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods Ecol Evol*. 2018; **9**: 439–446. https://doi.org/10.1111/2041-210X.12897

Christiansen JS, Gamst Moen AG, Hansen TH *et al.* Digestion of capelin, *Mallotus villosus* (Müller), herring, *Clupea harengus* L., and polar cod, *Boreogadus saida* (Lepechin), otoliths in a simulated seal stomach. *ICES J Mar Sci* 2005;**62**:86–92. https://doi.org/10.1016/j.icesjms.2004.06.022

Christiansen JS, Gildberg A, Nilssen KT *et al.* The gastric properties of free-ranging harp [*Pagophilus groenlandicus* (Erxleben, 1777)] and hooded [*Cystophora cristata* (Erxleben, 1777)] seals. *ICES J Mar Sci* 2004;**61**:287–92. https://doi.org/10.1016/j.icesjms.2004.01.002

Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;**20**:273–97. https://doi.org/10.1007/BF00994018

Costalago D, Bauer B, Tomczak MT *et al.* The necessity of a holistic approach when managing marine mammal–fisheries interactions: environment and fisheries impact are stronger than seal predation. *Ambio* 2019;**48**:552–64. https://doi.org/10.1007/s13280-018-1131-y

Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press, 2000.

da Silva J, Neilson JD. Limitations of using otoliths recovered in scats to estimate prey consumption in seals. *Can J Fish Aquat Sci* 1985;**42**:1439–42. https://doi.org/10.1139/f85-180

de Carvalho BM, Spach HL, Vaz-Dos-Santos AM *et al.* Otolith shape index: is it a tool for trophic ecology studies? *J Mar Biol Assoc UK* 2019;**99**:1675–82. https://doi.org/10.1017/S0025315419000729

Deagle BE, Thomas AC, McInnes JC *et al.* Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Mol Ecol* 2019;**28**:391–406. https://doi.org/10.1111/mec.14734

Denechaud C, Smoliński S, Geffen AJ *et al.* Long-term temporal stability of northeast Arctic cod (*Gadus morhua*) otolith morphology. *ICES J Mar Sci* 2020;**77**:1043–54. https://doi.org/10.1093/icesjms/fsz259

FAO. Fisheries Management. 2. The Ecosystem Approach to Fisheries 2.1. Best Practices in Ecosystem Modelling for Informing an Ecosystem Approach to Fisheries, FAO. Technical Guidelines for Responsible Fisheries. FAO Fisheries Technical Guidelines for Responsible Fisheries. No. 4, Suppl. 2, Add. 1. Rome: FAO, 2008. https://doi.org/10.1002/9780470015902.a0003252.pub3

Fiskbarometern. Resursöversikt [2024], Sik. 2024. https://doi.org/https://fiskbarometern.se [2024-02-15].

Grellier K, Hammond PS. Robust digestion and passage rate estimates for hard parts of grey seal (*Halichoerus grypus*) prey. *Can J Fish Aquat Sci* 2006;**63**:1982–98. https://doi.org/10.1139/F06-092

Hårding KC, Härkönen T, Helander B *et al.* Status of Baltic grey seals: population assessment and extinction risk. *NAMMCO Scientific Publ* 2007;**6**:33–56. https://doi.org/10.7557/3.2720

Hårding KC, Härkönen TJ. Development in the Baltic grey seal (*Halichoerus grypus*) and ringed seal (*Phoca hispida*) populations during the 20th century. *Ambio* 1999;**28**:619–27.

Härkönen T, Galatius A, Bräger S *et al.* Population growth rate, abundance and distribution of marine mammals. *Helsinki Commission - HELCOM* 2013;1–34. http://helcom.fi/Core%20Indicators/HELCOM-CoreIndicator Population_growth_rate_abundance_and_distribution_of_marine_mammals.pdf.

Härkönen T. *Guide to the Otoliths of the Bony Fishes of the Northeast Atlantic*. Hellerup: Danibu ApS, 1986, 256pp.

ICES. Working Group on Marine Mammal Ecology (WGMME). *ICES Scientific Rep* 2023b;**5**:123pp. https://doi.org/10.17895/ices.pub.24131736.v1

ICES. Working group on multispecies assessment methods (WGSAM; outputs from 2022 meeting). *ICES Scientific Rep* 2023a;**5**:233pp. https://doi.org/10.17895/ices.pub.22087292

Iverson SJ, Field CJ, Donohue M. Quantifying the contribution of prey species to the diet of marine predators: a comparison of methods. *Marine Ecology Progress Series* 2018;**601**:27–40. https://doi.org/10.3354/meps12787

Jackson AL, Inger R, Bearhop S *et al.* Assessing the utility of stable isotope analyses for marine predator diets. *J Mar Syst* 2021;**214**:103686. https://doi.org/10.1016/j.jmarsys.2021.103686

Jobling M, Breiby A. The use and abuse of fish otoliths in studies of feeding habits of marine piscivores. *Sarsia* 1986;**71**:265–74. https://doi.org/10.1080/00364827.1986.10419696

Jobling M. Marine mammal faeces as indicators of prey importance—a source of error in bioenergetics studies. *Sarsia* 1987;**72**:255–260. https://doi.org/10.1080/00364827.1987.10419722

Kauhala K, Kunnasranta M, Valtonen M. Hallien ravinto Suomen merialueella 2001–2007–alustava selvitys. *Suomen Riista* 2011;**57**:73–83.

Kokko H, Helle E, Lindström J *et al.* Backcasting population sizes of ringed and grey seals in the Baltic and Lake Saimaa during the 20th century. *Ann Zool Fennici* 1999;**36**:65–73.

Lehtonen TK, Gilljam D, Veneranta L *et al.* The ecology and fishery of the vendace (*Coregonus albula*) in the Baltic Sea. *J Fish Biol* 2023;**103**:1463–75. https://doi.org/10.1111/jfb.15542

Leopold MF, Damme CJG van, Philippart CJM *et al.* Otoliths of North Sea Fish—Fish Identification Key by Means of Otoliths and Other Hard Parts, Version 1.0. Amsterdam: World Biodiversity Database, CD Rom Series, ETI/NIOZ/Alterra, Biodiversity Center of ETI, University of Amsterdam, 2001.

Leopold MF, van Damme CJG, van der Veer HW. Diet of cormorants and the impact of cormorant predation on juvenile flatfish in the Dutch Wadden Sea. *J Sea Res* 1998;**40**:93–107. https://doi.org/10.1016/S1385-1101(98)00028-8

Libungan LA, Pálsson S. ShapeR: an R package to study otolith shape variation among fish populations. *PLoS One* 2015;**10**:e0121102. https://doi.org/10.1371/journal.pone.0121102

Liu C, Berry PM, Dawson TP *et al.* Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 2005;**28**:385–93. https://doi.org/10.1111/j.0906-7590.2005.03957.x

Lotze H, Coll M, Magera A *et al.* Recovery of marine animal populations and ecosystems. *Trends Ecol Evol* 2011;**26**:595–605. https://doi.org/10.1016/j.tree.2011.07.008

Lövgren J, Gilljam D, Bartolino V *et al.* Vendace in the Bothnian Bay—Benchmark Report 2021. Drottningholm, Lysekil, Öregrund, SWE: Department of Aquatic Resources, Swedish University of Agricultural Sciences, (SLU.aqua.2022.5.4-368), 2022.

Lundström K, Bergenius M, Aho T *et al.* *Födoval hos vikaresäl i Bottenviken: Rapport från den svenska forskningsjakten 2007–2009*. Aqua Reports 2014:1. Lysekil: Sveriges lant-bruksuniversitet, 23 s. 2014.

Mackenzie A, Buckland ST. Advances in the use of environmental DNA for dietary studies of aquatic predators. *Environ DNA* 2022;**4**:223–36. https://doi.org/10.1002/edn3.345

Magera A, Flemming J, Kaschner K *et al.* Recovery trends in marine mammal populations. *PLoS One* 2013;**8**:e77908. https://doi.org/10.1371/journal.pone.0077908

Meyer D, Dimitriadou E, Hornik K *et al.* *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*. Vienna: TU Wien, 2015.

Morissette L, Christensen V, Pauly D. Marine mammal impacts in exploited ecosystems: would large scale culling benefit fisheries? *PLoS One* 2012;**7**:e43966. https://doi.org/10.1371/journal.pone.0043966

Myers SC, Thorsen A, Smoliński S *et al.* An efficient protocol and data set for automated otolith image analysis. *Geosci Data J* 2020;**7**:80–8. https://doi.org/10.1002/gdj3.86

Nielsen JM, Clare EL, Hayden B *et al.* Diet tracing in ecology: method comparison and selection. *Methods Ecol Evol* 2018;**9**:278–91. https://doi.org/10.1111/2041-210X.12869

Pierce GJ, Boyle PR. A review of methods for diet analysis in piscivorous marine mammals. *Oceanogr Mar Biol* 1991;**29**:409–86.

Pikitch EK, Santora C, Babcock EA *et al.* Ecosystem-based fishery management. *Science* 2004;**305**:346–7. https://doi.org/10.1126/science.1098222

Pinheiro JC, Bates DM. *Mixed-effects Models in S and S-PLUS*. New York: Springer, 2000.

Prime JH. Observations on the digestion of some gadoid fish otoliths by a young common seal. ICES Document CM 1979/N:14. Copenhagen: International Countil for the Exploration of the Sea, 1979.

R Core Team. R: A language and environment for statistical computing. (2022). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of *K*-fold cross validation in prediction error estimation. *IEEE Trans Pattern Anal Mach Intell* 2010;**32**:569–75. https://doi.org/10.1109/TPAMI.2009.187

Rubbens P, Brodie S, Cordier T *et al.* Machine learning in marine ecology: an overview of techniques and applications. *ICES J Mar Sci* 2023;**80**:1829–53. https://doi.org/10.1093/icesjms/fsad100

Rudstam LG, Aneer G, Hilde´n M. Top−down control in the pelagic Baltic ecosystem. *Dana* 1994;**10**:105–29.

Scharff-Olsen CH, Galatius A, Teilmann J *et al.* Diet of seals in the Baltic Sea region: a synthesis of published and new data from 1968 to 2013. *ICES J Mar Sci* 2019;**76**:284–97. https://doi.org/10.1093/icesjms/fsy159

Smoliński S, Schade FM, Berg F. Assessing the performance of statistical classifiers to discriminate fish stocks using fourier analysis of otolith shape. *Can J Fish Aquat Sci* 2020;**77**:674–83. https://doi.org/10.1139/cjfas-2019-0251

Stock M, Nguyen B, Courtens W *et al.* Otolith identification using a deep hierarchical classification model. *Comput Electron Agric* 2021;**80**:105883. https://doi.org/10.1016/j.compag.2020.105883

Suuronen P, Lehtonen E. The role of salmonids in the diet of grey and ringed seals in the Bothnian Bay, northern Baltic Sea. *Fish Res* 2012;**125**:283–8. https://doi.org/10.1016/j.fishres.2012.03.007

Tollit DJ, Steward MJ, Thompson PM *et al.* Species and size differences in the digestion of otoliths and beaks: implications for estimates of pinniped diet composition. *Can J Fish Aquat Sci* 1997;**54**:105–19. https://doi.org/10.1139/f96-264

Trites AW, Joy R. Dietary analysis from fecal samples: how many scats are enough? *J Mammal* 2005;**86**:704–12. https://doi.org/10.1644/1545-1542(2005)086%5b0704:DAFFSH%5d2.0.CO;2

Vabø R, Moen E, Smolinski S *et al.* Automatic interpretation of salmon scales using deep learning. *Ecol Inform* 2021;**63**:101322. https://doi.org/10.1016/j.ecoinf.2021.101322

Xu QS, Liang YZ. Monte-Carlo cross validation. *Chemom Intell Lab Syst* 2001;**56**:1–11. https://doi.org/10.1016/S0169-7439(00)00122-2

Zuur A, Ieno EN, Walker N *et al.* *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer, 2009.

Handling Editor: Francis Juanes