Leveraging machine learning and catch commercial catch data for development of a discard bycatch mitigation plan in the Scup Fishery

October 10, 2022

Submitted to

Jeff Kaelin
Director of Sustainability and Government Relations
Lund's Fisheries, Inc.
997 Ocean Drive, Cape May, NJ 08204

Greg DiDomenico
Fisheries Management Specialist
Lund's Fisheries, Inc.
997 Ocean Drive, Cape May, NJ 08204

Prepared by

Robert Leaf, Ph.D. and Ralf Riedel, Ph.D.

Division of Coastal Sciences
School of Ocean Science and Engineering
The University of Southern Mississippi
703 East Beach Drive
Ocean Springs, MS 39564

Contents

Summary

Discards from commercial fisheries have been linked to detrimental effects on ecosystems and economies worldwide. Understanding spatial and temporal patterns of discards may assist in devising regulatory practices and mitigation strategies toward more sustainable fisheries practices. In this study, we investigate data from bycatch at sea on-board monitoring programs using a machine learning approach. Machine learning has been successful in revealing trends and patterns in various ecological applications, potentially providing a way forward as an alternative and complementary methodological approach in fisheries. We used a gradient boosting classifier for describing catch and bycatch patterns in the US Mid-Atlantic black seabass (*Centropristis striata*), summer flounder (*Paralichthys dentatus*), scup (*Stenotomus chrysops*), and shortfin squid (*Illex illecebrosus*) fishery. We found strong positive associations between the classifier and target species for catch datasets. For bycatch datasets, we found co-occurring species strong negative and positive associations, as well as associations with temperature and year of sample. From this study, we conclude that machine learning approaches to be promising in supplementing traditional methodologies, especially with the increase in data availability trends.

Rationale

A continued challenge in the management of the mid-Atlantic scup fishery is to reduce bycatch, well known in its contribution to the declines in stocks of ecologically and economically important species (Roberson and Wilcox 2022). Reducing bycatch, however, is not accomplished by the top-down imposition of regulatory measures aimed at reducing catches of unwanted species. Bycatch management need to be shaped after a holistic approach involving all stakeholders in a fishery, including government, academia, and industry at the least. Additionally, to effectively develop a strategy for bycatch reduction, new tools, based on state of the art analytical techniques, may also be of critical importance. Such tools, when adopted by managers, may offer more transparent outcomes and accepted measures for conservation of fisheries resources through reduction in bycatch.

Marine resource conservation and management has been one of the objectives of many government organizations in the United States. The Mid-Atlantic Fisheries Management Council is one of the eight councils established by the Magnuson-Stevens Fisheries Conservation and Mangement Act of 1979 that has strived to provide management advice toward sustainable fisheries in U.S. federal waters. Specific management plans are currently devised for 15 species, of which summer flounder, scup, black seabass, and shortfin squid are among the most important. One of the key factors affecting management plans is the incidence of unwanted bycatch in the fisheries targeting those four species. Data from at-sea monitoring programs are used to produce independent information sources in reference to bycatch temporal and spatial patterns by sector, harvesting gear, and stock area. Fisheries bycatch information, in turn, are used in supporting in-season management practices, ecosystem studies, and stock assessment activities.

Because bycatch mortality in most U.S. monitored fisheries is high (Bellido et al. 2011; Viana et al. 2021; Graham et al. 2022), reducing the incidence of untarget species is a management priority. The U.S. National Oceanographic and Atmospheric Administration has developed a National Bycatch Reduction Strategy with foreign and domestic partners, which includes devising harvesting gear modifications, establishment of harvest time-of-year closures, and determining area restrictions to reduce unwanted catch. Additionally, technology-based monitoring devices are research priorities to produce more accurate and larger volumes of bycatch data (Ditria et al. 2020; Khokher et al. 2021). With existing and forthcoming bycatch datasets, analytical methodologies based on machine learning (LeCun et al. 2015; Mohri et al. 2018) and suitable for big data (Mayer-Schönberger and Cukier 2014) may increasingly become center-stage in analyses supporting fishery management strategy development.

Objectives

We investigated the promise and limitatons of machine learning (ML) for analyzing temporal and spatial patterns in catches of incidentally caught living marine resources in a suite of mid-Atlantic fisheries. Our specific objectives are to

(1) provide a description of temporal and spatial patterns of bycatch in the scup, black sea bass, squid, and summer flounder fisheries, and
(2) use ML techniques to understand how gear, temporal, spatial, and environmental characteristics can be used to describe contrasts in bycatch magnitude and diversity.

Methods

In this work we used data between 1994 and 2020 obtained from the Northeast Fisheries Science Center Observer and At Sea Monitoring Program (OSMP). The OSMP collects catch data from commercial fishing vessel trips, providing independent data sources on catch biological characteristics for finfish and invertebrate marine species, such as bycatch composition and species-specific fishing mortalities. Data from OSMP are quality controlled and anonymized prior to storing for public use. Data anonymization is done to obscure information that can be traced back to vessel and individual fishers. Anonymization is done by reporting information at the levels of vessel trips, instead of tows, at quarter frequencies, and at a spatial resolution of quarter-degree-days. Anonymized OSMP data has been collected since May 1994, covering an area between latitudes 33.87^{o} and 43.05^{o} N, including coastal habitats to longitude 61.04^{o} W (Fig. 1).

Stored OSMP data fields include two keys for linking data tables, a program identifier, year and month landed, a trip identifier, haul number within a trip, statistical area code, inshore area code, gear code, a code for indicating weather the haul was observed for bycatch, latitude and longitude for where haul began, a species code, actual or estimated weight for each species, catch disposition, an indicator whether the species was dressed or round, the weight type recorded (actual or dressed), a comment field, and a field indicating the method for estimating catch weight (Table 1).

To standardize data, several columns and rows of the original dataset were omitted as a preprocessing step. Omitted data columns, reflecting metrics collected during fishing trips, were control columns, such as row identifiers, columns with little contrast, and columns with high column-to-column correlations (see below for details; Table 1). Omitted data rows included observations of frequencies lower than 0.5 percent and early observations where data collection protocols were different from current observation and recording methods. Additionally, a field of sea-surface temperature was included in the study dataset.

Model Development

The gradient boosting ensemble machine learning algorithm, a decision tree-based supervised method, was used as a classifier of bycatch weight as a function of predictors (see below). Gradient boosting was used due to its high performance compared to other ML and deep learning models (Shwartz-Ziv and Armon 2021) and its ability to capture complex non-linear dependencies at a low computational cost, especially for data with a low signal to noise ratio (Friedman 2001), as is common in fisheries-related datasets. Gradient boosting was also used for allowing transparency and interpretability of results, offered to some extent by tree-based models (Arrieta et al. 2020).

To train the models in this study, 70 percent of data rows were used as the training set and the remainder for model testing. The best number of boosting trees and their depths were determined using cross-validation. The Adaboost loss function was used for the model optimizer, decision tree stumps were the base learner, and subsampling was the regularization method. Model performance evaluation metrics were classification accuracy, recall, precision, and F-1 scores (Natekin and Knoll 2013). Because an ensemble of trees was used as the underlying algorithm for the model, interpretability of results is obscured compared to other white-box machine learning approaches (Du et al. 2019). To remedy for lack of transparency, LIME (Ribeiro et al. 2016), whereby predictor directionality with model classes are estimated, and relative predictor influence determination were used to assist in result interpretability.

Analysis for Bycatch Patterns

Data Engineering

Due to the high number of columns in the dataset obtained from OSMP, columns, also known as features, were engineered. Feature engineering was conducted to circumvent the curse of dimensionality (Alsaffar and Omar 2014), reduce data processing cost, and obtain better model learning performance (Liu et al. 2015; Li et al. 2017). Engineering consisted of feature selection and row filtering, as in the preprocessing phase above, feature modification, and feature augmentation using external data sources.

Features were selected based on domain knowledge and statistical-based methods (Li et al. 2017). Domain knowledge was obtained from counseling with data originator personnel and literature searches. Statistical-based methods were based on assessment of correlations among features and between a particular feature and the model response variable, also known as the class label (see below). Feature-to-feature correlations higher than 0.9 were processed by droping the one of the features. As a further processing, all categorical features were one-hot encoded to enable model runs (Yang et al. 2019).

The label for the boosting classifier was derived from the weight of catches. Catch weights for the OSMP data are recorded for each species on a per-haul basis. Catch weights were further transformed for this study by grouping species according to categories (Appendix I). Category weights were then logarithm transformed and dichotomized by assigning a class according to whether a value was below or above the median for that species category over the entire dataset, regardless of sampling location or time. The resulting dataset consisted of the dichotomized bycatch abundance indicator, the model class, and the features preprocessed and engineered as above.

Data Partitioning

Data from at-sea observer programs, such as OSMP, support stock assessment efforts and fisheries management plans to, among other things, reduce bycatch. Of commercially important target species, sea bass (*Centropristis striata*), summer flounder (*Paralichthys dentatus*), scup (*Stenotomus chrysops*), and short-fin squid (*Illex illecebrosus*) are the most significant. The fishery for these species is conducted using bottom otter trawls of various configurations (Shepherd and Terceiro 1994; Link et al. 2011). The discards for fishing trips targeting these species are recorded by observers for each species brought on board after a net haul. Species brought on board are subsequently either kept for commercialization or discarded overboard.

After processing the data, the dataset for this study was partitioned to construct the final datasets for analyses. The four species and two bycatch dispositions above were the basis for data partitioning, resulting in eight datasets (Table 2) analyzed independently using the modeling approach above.

Analysis for Bycatch Species Richness

Feature engineering for the boosting classifier was conducted as described above for the analysis of bycatch patterns, except for data filtering. Because species richness was the target response variable, all species, independent of frequency, were included in generating the final analysis dataset. To provide a spatial and temporal estimate of species richness, richness was calculated based on haul observations over quarter-degree square and year quarter groupings (Table 1), resulting in a smaller dataset than the original (Table 2).The label for dichotomization used in the boosting classifier for richness was based on the actual species, rather than species categories. Dichotomization was based on the median species richness taking into account the overall final dataset for species richness.

Results

General Findings

Fisheries targeting summer flounder had the largest catches and discards, followed by fisheries for sea bass, squid, and spot. Flounder fisheries had discards totalling 5,570.34 MT and retained catches 5,735.53 MT over the analysis period. The sea bass fishery kept 1081.84 MT and discarded 1323.59 MT. Fisheries for shortfin squid kept 169.91 MT and discarded 12.69 MT. Finally, for the spot fishery, 1.25 MT were discarded and 1.31 MT kept. For all four fisheries, the weight or the target species kept exceeded the weight for all other non-target species harvested. Discards, however, followed a distinct species-specific pattern. For the summer flounder and sea bass fishery, spiny dogfish (*Squalus acanthias*) comprised the majority of discards. For the spot and squid fishery, striped bass (*Morone saxatilis*) and scup were dominant (Table 3). Species richness was the highest at fishing grounds Hudson Canyon, followed by the area off No-Mans-Land, Long Island, and Wilmington Canyon (Figure 1), which in combination accounted for over half of species richness. The most abundant species either discarded or kept was summer flounder, followed by shortfin squid, scup, and butterfish (*Peprilus triacanthus*). Species richness was fairly evenly distributed, with summer flounder only accounting for less than 10 percent of the total richness (Table 4).

The unpartitioned dataset for the gradient boosting analysis consisted of 598,513 rows and 110 columns (Appendix I). Partitioned datasets were of varied sizes according to fishery type and catch disposition (Table 2).

Catch Discarded Dataset

Gradient boosting results for sea bass fishery, discarded bycatch showed temperature, year, and the species category shark to be the most important features in determining model performance (Table 2) and the association with the above-median class. Temperature and year, however, showed a dispersed signal associated with the class. The species category shark was strongest showing positive association with the above-medial class and category squids were strongly negative (Figure 2).

For the summer flounder fishery, the most important features for model performance were temperature, year, and longitude. Temperatures lower than 3.3 Celsius showed the strongest positive association with the above-medial class and species category squid the strongest negative association. Temperatures higher than 3.8, conversely, showed a strong negative association with the above-median class (Figure 3).

For the spot dataset showed temperature and year as the strongest features for model performance. Categories shark and squid were the strongest positive and negative, respectively, associations with the above-median class (Figure 4). The shortfin squid fishery showed temperature and year to be the strongest features determining model performance with species category squid as the strongest negative association and longitude larger than 70.25^{o} W as the strongest positive association with the above-median class (Figure 5).

Catch Kept Dataset

For sea bass fishery, species categories squid and sea bass were the strongest features, both positively associated with the above-median class (Figure 6). For this dataset, the species category flounder was the strongest feature determining model performance, also showing a strong positive association with the above-median class (Figure 7). The spot dataset for this disposition showed species category scup to be the most influential feature for model performance and strongest positive association with the above-median class (Figure 8). Species category squid was the strongest feature for model performance, showing the strongest positive association with the above-median class (Figure 9).

Species Richness Dataset

For the species richness dataset, temperature, year, and longitude were the strongest features toward model performance. Recent years, 2018-, had the largest positive association with the above-median class and longitude west of 71.25º W the largest negative association. Additionally, longitudes east of 73.25 to 71.75º W had the second largest positive association and years before 2009 had the second largest negative association with the above-median class (Figure 10).

Discussion

The findings of this study point to the promise of using ML approaches for describing contrasts in bycatch data for fisheries in the mid-Atlantic using abundance and taxomonic richness metrics. We show that ML approaches can assist in understanding contrasts in bycatch data from four commercially important stocks in the United States Mid-Atlantic eastern coast and that there are pronouced variations in the temporal and spatial patterns of bycatch in the region. We also extend the promise of ML approaches to data collections and tools that might further support and enhance the quality of data gathered from on-board observers programs for more effective fisheries management.

Catch Discarded Dataset

For datasets reflecting discarded catches, the above-median class for black seabass bycatch weights was positively associated with the shark and sea robin species categories, and negatively with the squid category. Larger black seabass bycatches, therefore, were associated with bycatches of species from the shark, sea robin, and squid categories, potentially reflecting black seabass co-ocurrence with the latter two fish species at least in the juvenile phase, the commercially illegal size black seabass brought on board.

The category squid was primarily negatively associated with the summer flounder above-median bycatch weight class, with the categories hake and scup following in that order. Catching higher weights of summer flounder as bycatch, as the commercially illegal-size individuals, was accompanied with lower catches of squid, hakes, and scup. A possible explanation for the negative association are interactions between gear selectivity and seasonal changes in species distribution leading to seggregation of species-specific populations of demersal fish (Shepherd and Terceiro 1994; Gabriel 1996; Link et al. 2002). Small-scale changes in habitat use within an area and season have been reported for scup and summer flounder, where one species might inhabit sandy bottoms, whereas another found in hard bottoms (Shepherd and Terceiro 1994). Such patterns of occurrence and habitat preferrences might have accounted for the observed associations in the summer flounder discards dataset.

For scup, the categories shark showed a positive and squid a negative association with the above-median bycatch weight class. A co-occurrence of sharks and scup, together with distinct habitat segreggations with squids might be expected for scup.

Finally, for the shortfin squid dataset, only the category squid was negatively associated with the above-median bycatch class. Discarded catches for the shortfin squid fishery were, therefore, the cleanest and most obvious, mostly comprising commercially illegal size species in the category squid, driving down the bycatch weights of the target shortfin squid.

Catch Kept Dataset

The analyses on the datasets for species kept for commercialization showed the ML results for feature importance and directionality to be more consistent and intuitive. For black seabass retained, categories squids, black seabass, and flounders, in that order, showed the strongest associations with the above-median class, indicating that larger black seabass catches are associated with species belonging to those categories. This latter finding shows that the black seabass fishery overlaps the high value non-target stocks of squid and flounders. At least seasonal, black seabass co-inhabit with those species as competitors for habitat and food (Musick and Mercer 1977; Garrison and Link 2000; Collette and Klein-Macphee 2002), which may at least partly explain the findings of this study. For management purposes, squid and flounders need to be taken into account in any plan to avoid potential knock on effects on the target species by overharvest of the legal non-target fish.

As for the discard datasets above, the shortfin squid and also the summer flounder fisheries was the cleanest in terms of bycatch according to the dataset of fish kept for commercialization. The more species kept belonging to the class squid, the higher the catches of shortfin squid, an obvious and expected finding. Cleaner bycatch for the squid fishery is in line with expectation. To keep catch quality, squid fishing in the study area is done mostly using large-mesh bottom trawls (Arkhipkin et al. 2015; Lowman et al. 2021). Fish trawl nets, with smaller mesh sizes, may damage the squid’s skin and mantle, thus lowering market value. Large-mesh trawls are more selective, partly explaining the clean catches reflected by the discard data for shortfin squid.

For the summer flounder and scup fisheries, the same patterns as for shortfin squid emerged, with the addition of a negative association between catches of the target species and species of the category squid. The higher the catches of squid species, therefore, the lower the catches of summer flounder and scup. The black seabass fishery, conversely, showed an inverse effect. The higher the catches of species of the squid category, the higher were the catches of black seabass, providing evidence that squid are harvested along with black seabass.

Species Richness Dataset

No clear patterns in species richness were observed from the bycatch analysis. Worth mentioning was the increase in the number of species harvested from 2018 onwards. Alternatively, a low species richness was associated with longitudes toward the western areas, offshore habitats, spanned by the dataset. This latter was expected, as offshore habitats may offer less habitat complexity and, therefore, species richness than habitats closest to shore. Overall, however, there were no indications that species richness has changed over the years or that there are any key areas of high species richness from the dataset analysed.

Machine Learning as an Alternative Analytical Framework

Although the ML analyses showed intuitive results, there were findings that departed from intuition when considering only traditional analytical approaches. The ML analyses on discard characteristics indicated that sea surface temperature and the categorical varaible year were consistently important features in classifying weight of bycatch. We note that the interpretation of these patterns should be approached with caution. Classification using the machine learning algorithm of this study arrives at an outcome by assessing the importance of features taking each individual sample (row in the data table) and predicting its class membership independently of any other sample (Deisenroth et al. 2020). When a feature is important, it is possible that the direction of effect of a feature on the class, in this study's case the below or above median, is not consistent among samples. As an example, two samples may produce a high rank for a feature importance, but in opposite directions. To remedy for this effect, assessing the directionality of a feature was needed. For the analyses on discards, feature directionality exhibited this unwanted phenomenon. Both temperature and year showed strong uncertainty as to directionality. When only looking at feature importance, considering them irrespective of their levels and direction, the sum of their effects, across all observations, was strong because of the individual contributions from each sample, making the feature an important one in determining a classification. If the objective of the ML analyses were to develop a classifier for predicting below- and above-median bycatch weights, no further interpretation would be needed. If, however, the objective is an understanding of which factors are important in determining discard patterns, as in this study, further interpretations of ML outcomes must be sought. Features temperature and year used in the discard datasets, therefore, represent variable and uncertain factors when it comes to determining discard abundance and should, therefore, be interpreted with caution.

Even with the encouraging results from the gradient boosting ML approach used in this study, suggestions for further improvements may be offered. Providing fine-grained vessel positioning may aid fisheries management decisions by better classifying movement patterns into activities associated with fishing and non-fishing practices. With the advent of affordable, off-the-shelf global positioning devices, detailed information on the spatial dynamics of fishing effort may be accurately estimated with classifiers as used in this study for small- and large-scale fisheries worldwide. Moreover, equipping vessels with cameras may also assist in estimating bycatch amounts. Camera images may be readily analyzed with computer vision approaches, such as deep learning algorithms (LeCun et al. 2015), to automate data collection, allowing for widespread coverage of bycatch data (Khokher et al. 2021). Computer vision has been successfully used in fish identification (Ditria et al. 2020), estimation of fish abundance (Tseng and Kuo 2020), and length distributions (White et al. 2006), often surpassing accuracy by human experts. Of similar promise for fisheries bycatch management are data gathered from acoustic backscatter signals for estimating the fate of discards. Fisheries acoustics are an important source for estimating stock abundance (Simmonds and MacLennan 2008; Koslow 2009; Trenkel et al. 2011) and may become a centerpiece in EAFM if predator-prey dynamics, biomass, and spatial distributions are estimated on finer resolutions and wider scales. Using ML techniques for analyzing acoustics data, especially supervised learning (Jiang et al. 2020), has shown promise in classifying bottom types in north African fisheries grounds (Sarr et al. 2021) and called the attention of managers and regulators as to the possibilities of such data in fisheries applications (Handegard et al. 2021). Acoustic data groundtruthing for labeling may initially be time consuming, but the rewards are justified, given the benefits of automation in data generation and the increasingly higher accuracy of classifiers enabling better estimates of fisheries statistics from acoustic datasets.

Machine learning approaches to analysing fisheries data will likely not replace traditional modeling methods. In combination, traditional modeling and ML may capture enough of the complexities and dynamics of ecological processes determining catch abundances to provide robust advice for sustainable harvest. A trend in augmenting the performance of traditional fisheries stock assessment and estimation models using ML has been observed recently (Pérez-Ortiz et al. 2013; Syed and Weber 2018; Kaemingk et al. 2020; Yang et al. 2020; Chan and Pan 2021), attesting to the applicability of ML algorithms to fisheries data. With the increasing prospect of automation in fisheries data collection, ML techniques may be the only feasible approach for data processing and analysis, as dataset will become larger and more complex. Automation, however, come with the cost of transparency, especially when deep learning techniques are used for classification. Because decisions based on such analysis most likely will have large ecological, economic, and social impacts, explaining the results of ML techniques in a clear and understandable way is a must. Many ML techniques are defined as opaque, whereby how results are obtained are not clearly understood. Using mechanisms for explaining the results of an analysis, as done in this study, must accompany any opaque ML technique if the benefits of this new and ever growing analytical alternative are to be fully realized.

Conclusions

The results of this study indicate that ML alternatives may successfully supplement traditional analytical approaches to fisheries research. Results from ML model runs were able to capture general expected patterns in harvest according to target species. Given the inherent uncertainty associated with fisheries data, these results are encouraging for the adoption of ML techniques to the field. The adoption of ML into the fisheries field, however, needs to be done carefully, always with the analytical objective in mind. Machine learning techniques are mostly for the objective of classification, whereas in fisheries, datasets are largely for explanation, inference, or predicting over continuous valued outcomes. Adopting ML techniques blindly, without consideration of method explainability, may be a fruitful approach is classification is the only goal. When decisions couched on understanding of ecological processes undelying analyses are the goal, a clear knowledge of how findings are derived is a must. Using ML techniques must, therefore, be used in conjunction with traditional statistical analyses or extended to add transparency, beyond only classification, to better explain model outcomes, as done in this study.

References

Alsaffar A, Omar N (2014) Study on feature selection and machine learning algorithms for Malay sentiment classification. Proceedings of the 6th International Conference on Information Technology and Multimedia 270–275

Arkhipkin AI, Rodhouse PGK, Pierce GJ, Sauer W, Sakai M, Allcock L, Arguelles J, Bower JR, Castillo G, Ceriola L, Chen C-S, Chen X, Diaz-Santana M, Downey N, González AF, Granados Amores J, Green CP, Guerra A, Hendrickson LC, Ibáñez C, Ito K, Jereb P, Kato Y, Katugin ON, Kawano M, Kidokoro H, Kulik VV, Laptikhovsky VV, Lipinski MR, Liu B, Mariátegui L, Marin W, Medina A, Miki K, Miyahara K, Moltschaniwskyj N, Moustahfid H, Nabhitabhata J, Nanjo N, Nigmatullin CM, Ohtani T, Pecl G, Perez JAA, Piatkowski U, Saikliang P, Salinas-Zavala CA, Steer M, Tian Y, Ueta Y, Vijai D, Wakabayashi T, Yamaguchi T, Yamashiro C, Yamashita N, Zeidberg LD (2015) World Squid Fisheries. Rev Fish Sci Aquac 23:92–252

Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115

Bellido JM, Santos MB, Pennino MG, Valeiras X, Pierce GJ (2011) Fishery discards and bycatch: solutions for an ecosystem approach to fisheries management? Hydrobiologia 670:317

Chan HL, Pan M (2021) Fishing trip cost modeling using generalized linear model and machine learning methods – A case study with longline fisheries in the Pacific and an application in Regulatory Impact Analysis. PLOS ONE 16:e0257027

Collette BB, Klein-Macphee G (2002) Bigelow and Schroeder's Fishes of the Gulf of Maine, Third Edition. Smithsonian Books, Washington, DC

Deisenroth MP, Faisal AA, Ong CS (2020) Mathematics for Machine Learning. Cambridge University Press,

Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM (2020) Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning. Front Mar Sci 7:

Du M, Liu N, Hu X (2019) Techniques for interpretable machine learning. Commun ACM 63:68–77

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann Stat 29:1189–1232

Gabriel L (1996) The Role of Targeted Species in Identification of Technological Interactions in Mid-Atlantic Bight Groundfish Fisheries. J Northwest Atl Fish Sci 19:11–20

Garrison LP, Link JS (2000) Dietary guild structure of the fish community in the Northeast United States continental shelf ecosystem. Mar Ecol Prog Ser 202:231–240

Graham J, Kroetz AM, Poulakis GR, Scharer RM, Carlson JK, Lowerre-Barbieri SK, Morley D, Reyier EA, Grubbs RD (2022) Commercial fishery bycatch risk for large juvenile and adult smalltooth sawfish (Pristis pectinata) in Florida waters. Aquat Conserv Mar Freshw Ecosyst 32:401–416

Handegard NO, Andersen LN, Brautaset O, Choi C, Eliassen IK, Heggelund Y, Hestnes AJ, Malde K, Osland H, Ordonez A, Patel R, Pedersen G, Umar I, Engeland TV, Vatnehol S (2021) Fisheries acoustics and Acoustic Target Classification - Report from the COGMAR/CRIMAC workshop on machine learning methods in fisheries acoustics. 2021 - 25

Jiang T, Gradus JL, Rosellini AJ (2020) Supervised Machine Learning: A Brief Primer. Behav Ther 51:675–687

Kaemingk MA, Hurley KL, Chizinski CJ, Pope KL (2020) Harvest–release decisions in recreational fisheries. Can J Fish Aquat Sci 77:194–201

Khokher MR, Little LR, Tuck GN, Smith DV, Qiao M, Devine C, O'Neill H, Pogonoski JJ, Arangio R, Wang D (2021) Early lessons in deploying cameras and artificial intelligence technology for fisheries catch monitoring: where machine learning meets commercial fishing. Can J Fish Aquat Sci 1–10

Koslow JA (2009) The role of acoustics in ecosystem-based fishery management. ICES J Mar Sci 66:966–973

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444

Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H (2017) Feature Selection: A Data Perspective. ACM Comput Surv 50:94:1-94:45

Link JS, Bundy A, Overholtz WJ, Shackell N, Manderson J, Duplisea D, Hare J, Koen-Alonso M, Friedland KD (2011) Ecosystem-based fisheries management in the Northwest Atlantic. Fish Fish 12:152–170

Link JS, Garrison LP, Almeida FP (2002) Ecological Interactions between Elasmobranchs and Groundfish Species on the Northeastern U.S. Continental Shelf. I. Evaluating Predation. North Am J Fish Manag 22:550–562

Liu Y, Tang F, Zeng Z (2015) Feature Selection Based on Dependency Margin. IEEE Trans Cybern 45:1209–1221

Lowman BA, Jones AW, Pessutti JP, Mercer AM, Manderson JP, Galuardi B (2021) Northern Shortfin Squid (Illex illecebrosus) Fishery Footprint on the Northeast US Continental Shelf. Front Mar Sci 8:

Mayer-Schönberger V, Cukier K (2014) Big Data: A Revolution That Will Transform How We Live, Work, and Think. Harper Business, Boston

Mohri M, Rostamizadeh A, Talwalkar A (2018) Foundations of Machine Learning, second edition. MIT Press,

Musick JA, Mercer LP (1977) Seasonal Distribution of Black Sea Bass, Centropristis striata, in the Mid-Atlantic Bight with Comments on the Ecology and Fisheries of the Species. Trans Am Fish Soc 106:12–25

Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. Front Neurorobotics 7:21

Pérez-Ortiz M, Colmenarejo R, Fernández Caballero JC, Hervás-Martínez C (2013) Can Machine Learning Techniques Help to Improve the Common Fisheries Policy? 278–286

Ribeiro MT, Singh S, Guestrin C (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 1135–1144

Roberson LA, Wilcox C (2022) Bycatch rates in fisheries largely driven by variation in individual vessel behaviour. Nat Sustain 1–9

Sarr J-MA, Brochier T, Brehmer P, Perrot Y, Bah A, Sarré A, Jeyid MA, Sidibeh M, El Ayoubi S (2021) Complex data labeling with deep learning methods: Lessons from fisheries acoustics. ISA Trans 109:113–125

Shepherd GR, Terceiro M (1994) The Summer Flounder, Scup, and Black Sea Bass Fishery of the Middle Atlantic Bight and Southern New England Waters. http://aquaticcommons.org/id/eprint/2693

Shwartz-Ziv R, Armon A (2021) Tabular Data: Deep Learning is Not All You Need. ArXiv210603253 Cs

Simmonds J, MacLennan DN (2008) Fisheries Acoustics: Theory and Practice. John Wiley & Sons,

Syed S, Weber CT (2018) Using Machine Learning to Uncover Latent Research Topics in Fishery Models. Rev Fish Sci Aquac 26:319–336

Trenkel VM, Ressler PH, Jech M, Giannoulaki M, Taylor C (2011) Underwater acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. Mar Ecol Prog Ser 442:285–301

Tseng C-H, Kuo Y-F (2020) Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks. ICES J Mar Sci 77:1367–1378

Viana D, de Souza MRDP, de Assis Teixeira da Silva U, Pereira DMC, Kandalski PK, Neundorf AKA, Peres D, dos Santos AT, Romão S, Moura MO, Fávaro LF, Donatti L (2021) The effect of bottom trawling time on mortality, physical damage and oxidative stress in two Sciaenidae species. Rev Fish Biol Fish 31:957–975

White DJ, Svellingen C, Strachan NJC (2006) Automated measurement of species and length of fish by computer vision. Fish Res 80:203–210

Yang KK, Wu Z, Arnold FH (2019) Machine-learning-guided directed evolution for protein engineering. Nat Methods 16:687–694

Yang S, Dai Y, Fan W, Shi H (2020) Standardizing catch per unit effort by machine learning techniques in longline fisheries: a case study of bigeye tuna in the Atlantic Ocean. Ocean Coast Res 68:

Tables

Table 1. Information collected by the NOAA Northeast Fisheries Science Center Observer and At Sea Monitoring Program.

Predictor	Data Transformation	Description
Used in this study		
Habitat	One-hot encoded; one feature	Inshore or offshore
Statistical Area	One-hot encoded; XX features	Sampling bands along the North-South direction; latitude intervals south of 34.25º N; 34.25º N -37.24º N; 37.25º N -39.24º N; 39.25º N -40.24º N; 40.25º N -41.24º N; north of 41.24º N
Quarter Degree Square	Integer	
Year	Integer	2003-2020
Quarter	Integer	1-4, for each year quarter
Latitude	Decimal	34.75, 35.25, 35.75, 36.25, 36.75, 37.25, 37.75, 38.25, 38.75, 39.25, 39.75, 40.25, 40.75, 41.25, 41.75, 42.25, 42.75º N
Longitude	Decimal	75.75, 75.25, 74.75, 74.25, 73.75, 73.25, 72.75, 72.25, 71.75, 71.25, 70.75, 70.25, 69.75, 69.25, 68.75, 68.25, 67.75, 67.25, 66.75, 66.25º W
Bycatch disposition	Alphanumeric	Catch kept or discarded; used to partition datasets
Cod mesh size	Decimal	56, 120, 133, 151 mm

Gear type	One-hot encoded; 3 features	Fish, Ruhle, Scallop,Twin
Target species	Alphanumeric	4 fisheries, used to partition datasets
Water temperature	Decimal	Obtained from NOAA buoys
Excluded from this study		
Link fields		
Program		
Nem area		
Observer Flag		
Beginning Coordinates		
Fish Disposition		
DRFlag		
Weight Type		

Table 2. Model performance metrics showing precision, recall, and F1-ratio for label above/below median value of the four fishery and two dispositions in the NE demersal finfish fishery.

Species	Catch Disposition	Number of records	Number of features	Accuracy	Precision above/below median	Recall above/below median	F1-Ratio above/below median
Sea Bass	Discarded	18,490	70	0.68	0.70/0.66	0.69/0.67	0.70/0.67
	Kept	11,435	59	0.81	0.81/0.81	0.70/0.88	0.75/0.84
Summer Flounder	Discarded	140,547	97	0.69	0.70/0.67	0.72/0.65	0.71/0.66
	Kept	68,460	88	0.79	0.75/0.81	0.68/0.86	0.71/0.83
Scup	Discarded	36,276	72	0.66	0.67/0.65	0.67/0.65	0.67/0.65
	Kept	24,084	62	0.75	0.74/0.76	0.69/0.80	0.72/0.78
Short fin squid	Discarded	193,539	110	0.70	0.73/0.67	0.78/0.60	0.75/0.63
	Kept	85,394	86	0.76	0.69/0.81	0.70/0.80	0.69/0.81
Biodiversity	n/a	10,041	85	0.73	0.74/0.73	0.73/0.74	0.74/0.73

Table 3. Catch percentages and weights by bycatch disposition for four target species in the NE US finfish fishery.

Species	Percent	Kept	Discarded
BSB as target			
Centropristis striata	32.79	664.55	124.11
Squalus acanthias	25.46	10.68	601.82
Stenotomus chrysops	15.38	257.75	112.10
Prionotus evolans	7.00	2.63	165.85
Prionotus carolinus	5.43	0.01	130.49
Paralichthys dentatus	4.64	82.96	28.76
Beringraja binoculata	1.66	7.91	32.05
Summer Flounder as target			
Paralichthys dentatus	39.82	4,311.68	190.37
Squalus acanthias	11.24	70.16	1,200.91
Beringraja binoculata	8.24	244.92	686.60
Raja eglanteria	7.19	34.36	778.25
Stenotomus chrysops	5.09	353.52	221.44
Prionotus carolinus	4.82	1.82	543.51
Lophius piscatorius	3.00	155.75	183.56
Dipturus laevis	2.48	0.76	280.11
Limulus polyphemus	2.14	23.23	219.23
Mustelus canis	1.89	49.89	163.36
Prionotus evolans	1.80	5.67	197.85
Centropristis striata	1.55	99.42	75.53
Placopecten magellanicus	1.29	31.47	114.76
Scophthalmus aquosus	1.17	4.03	128.57
Merluccius bilinearis	1.13	85.92	41.81
Spot as target			
Stenotomus chrysops	31.38	0.47	-
Morone saxatilis	17.44	-	0.45
Mustelus canis	17.39	0.04	0.41
Cynoscion regalis	12.55	0.32	-
Pomatonus saltatrix	10.18	0.26	-
Paralichthys dentatus	6.10	0.16	-
Centropristis striata	1.78	0.05	-
Squalus acanthias	1.77	-	0.05
Squid as target			
Illex Illecerosa	61.01	111.00	0.40
Scomber colias	16.92	29.87	1.02

Doeyteuthis pealeii	15.42	27.99	0.16
Stenotomus chrysops	1.81	0.58	2.71
Beringraja binoculata	1.15	0.02	2.07
Squalus acanthias	0.00	-	1.89

Table 4. Species richness from the NOAA Northeast Fisheries Science Center Observer and At Sea Monitoring Program; years 1994-2020; only species larger than 1% total data representation shown.

Species	Percent	Number
Paralichthys dentatus	9.58	61,435
Doeyteuthis pealeii	8.00	51,348
Stenotomus chrysops	6.65	42,637
Peprilus triacanthus	5.68	36,456
Centropristis striata	5.37	34,440
Lophius piscatorius	5.26	33,715
Merluccius bilinearis	4.96	31,836
Hippoglossina oblonga	3.74	23,967
Squalus acanthias	3.71	23,773
Prionotus carolinus	3.45	22,119
Beringraja binoculata	3.31	21,225
Urophycis regia	3.28	21,042
Scophthalmus aquosus	2.89	18,551
Prionotus evolans	2.67	17,143
Mustelus canis	2.55	16,337
Illex Illecerosa	1.96	12,571
Raja eglanteria	1.95	12,481
Pseudopleuronectes americanus	1.89	12,144
Pomatonus saltatrix	1.80	11,560
Homarus americanus	1.67	10,684
Urophycis chuss	1.59	10,193
Dipturus laevis	1.52	9,758
Scomber scombrus	1.16	7,421
Libinia emarginata	1.10	7,055
Limulus polyphemus	1.02	6,531

Figures

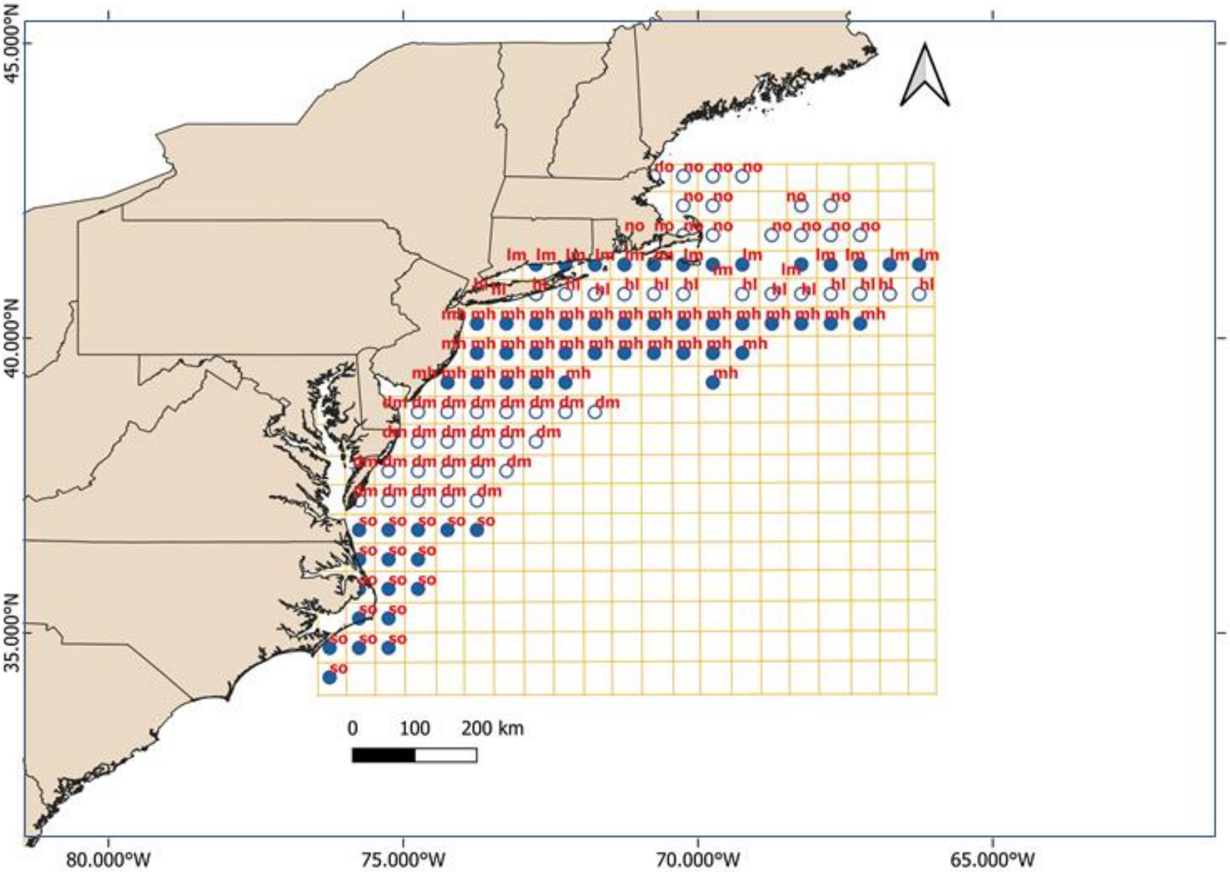

Figure 1. NOAA Northeast Fisheries Science Center Observer and At Sea Monitoring Program are of coverage.

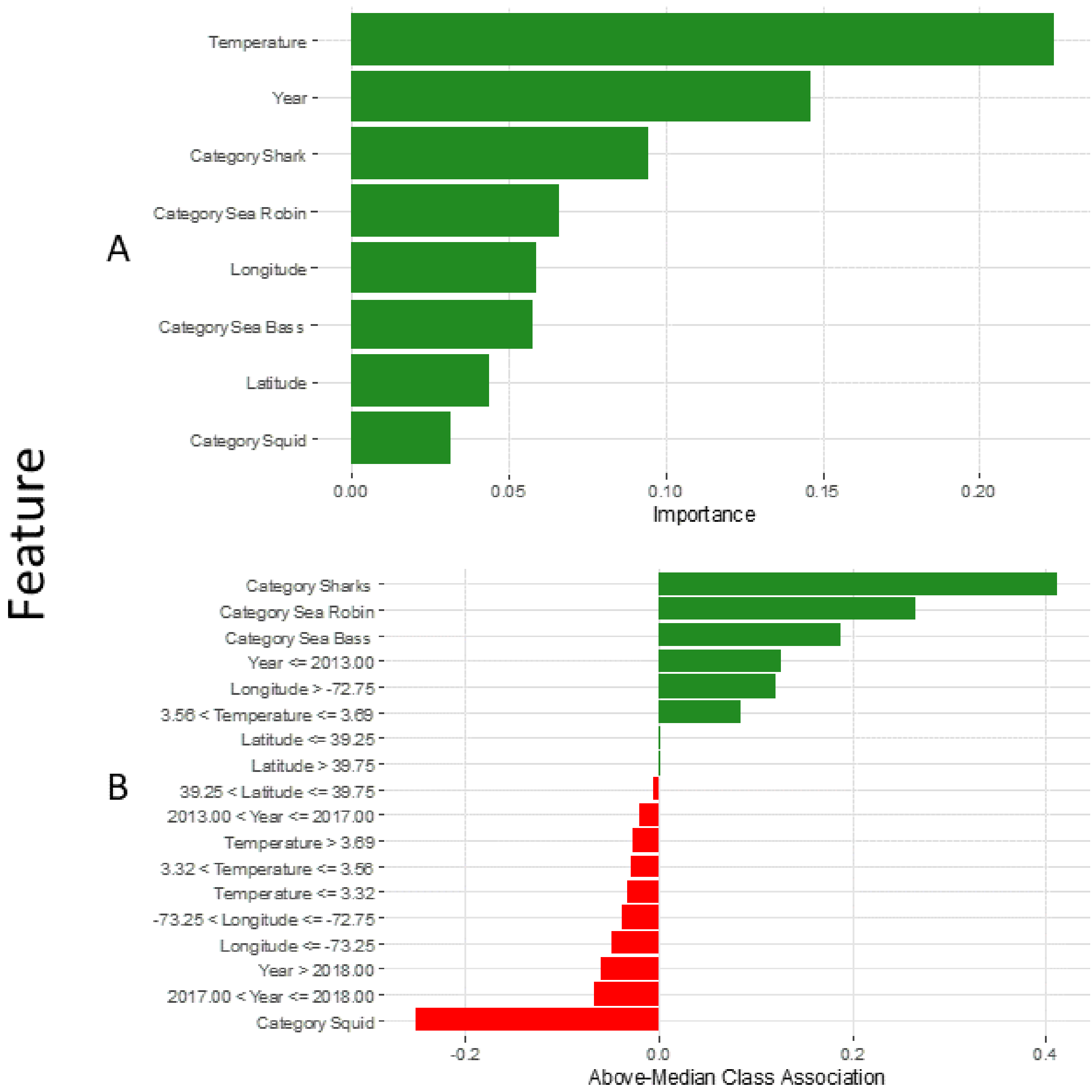

Figure 2. Feature importance after XBoost machine learning analysis for black seabass disposition discarded

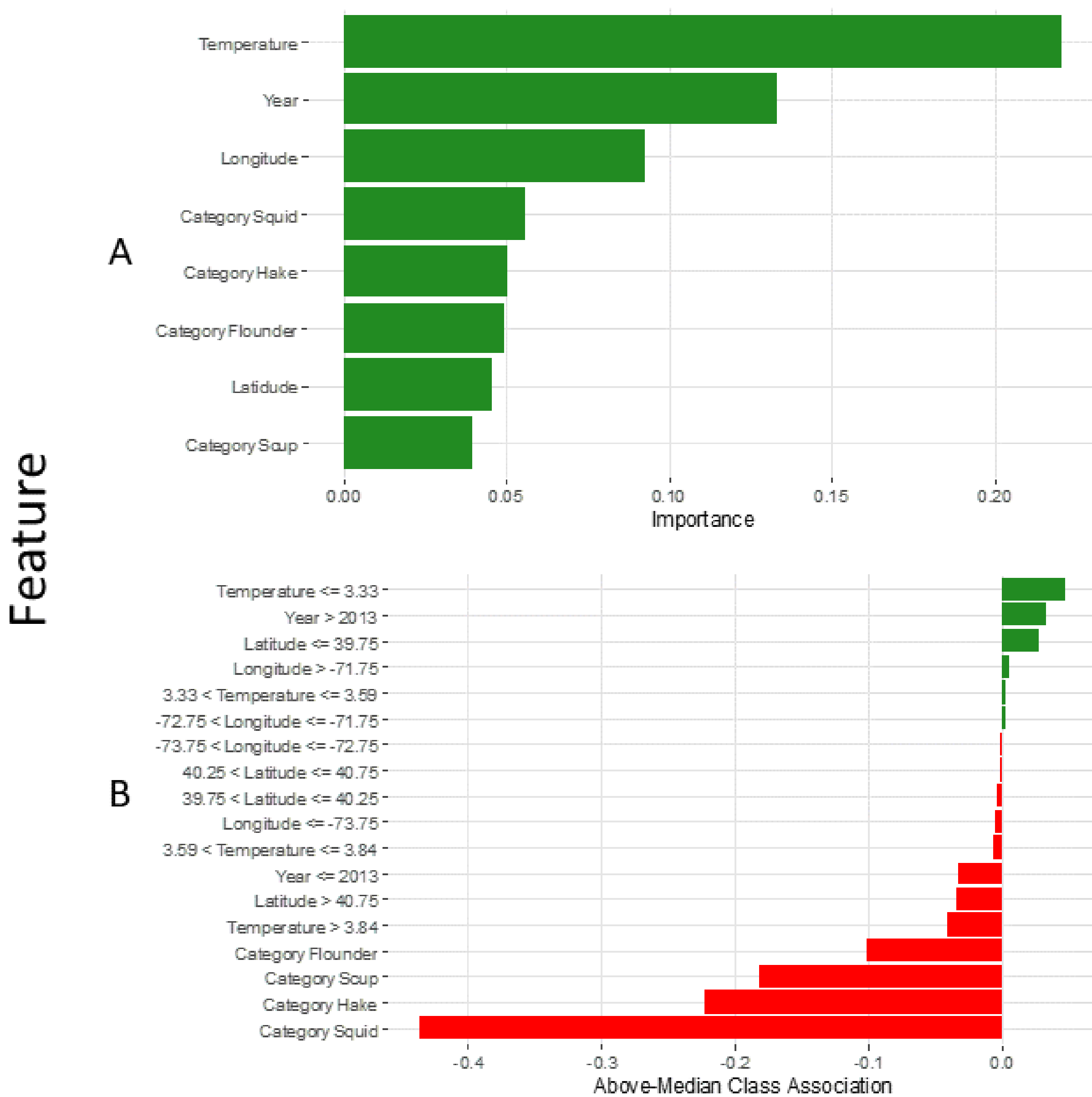

Figure 3. Feature importance after XBoost machine learning analysis for summer flounder disposition discarded

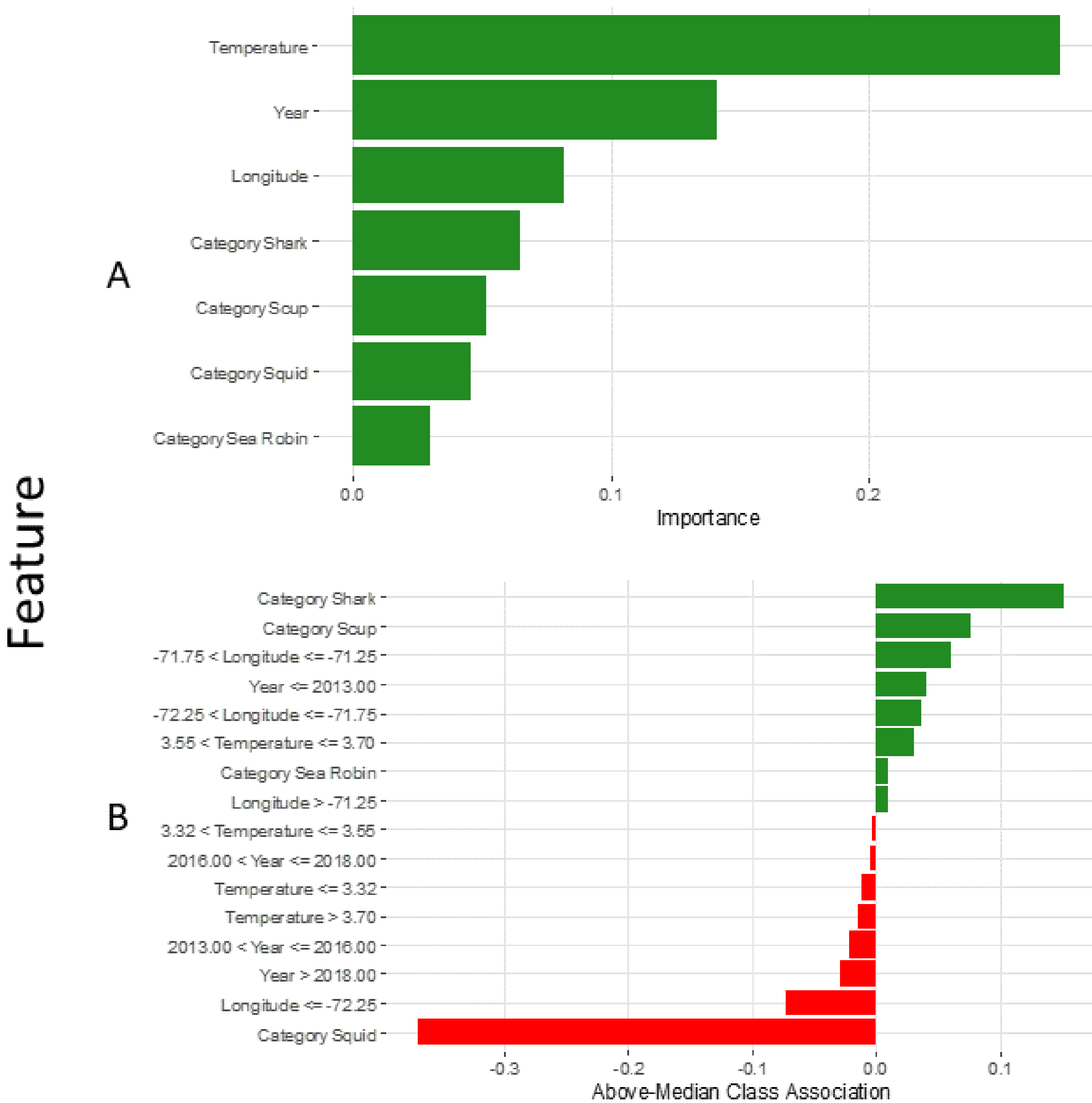

Figure 4. Feature importance after XBoost machine learning analysis for scup disposition up discarded.

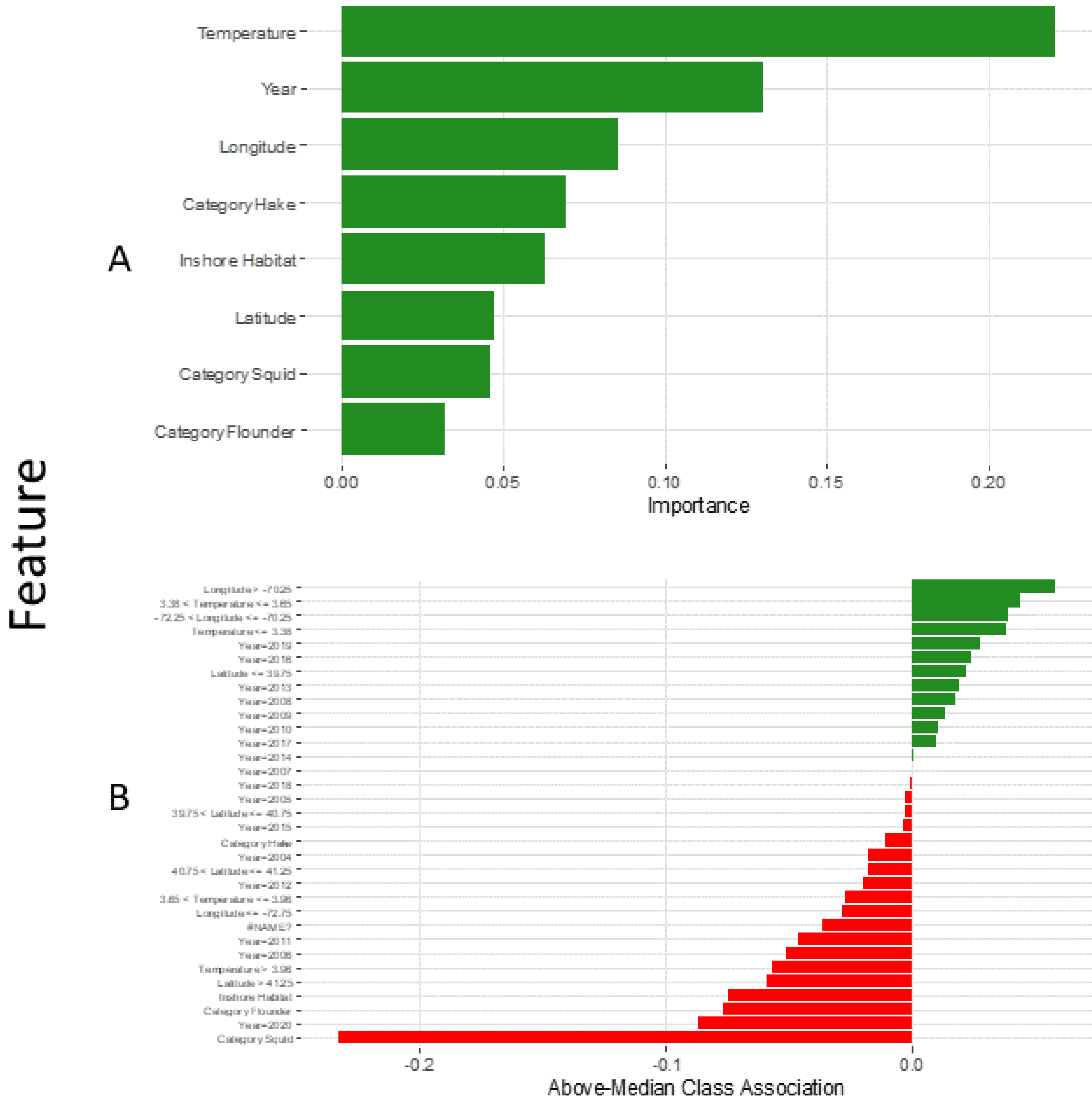

Figure 5. Feature importance after XBoost machine learning analysis for shortfin squid disposition discarded

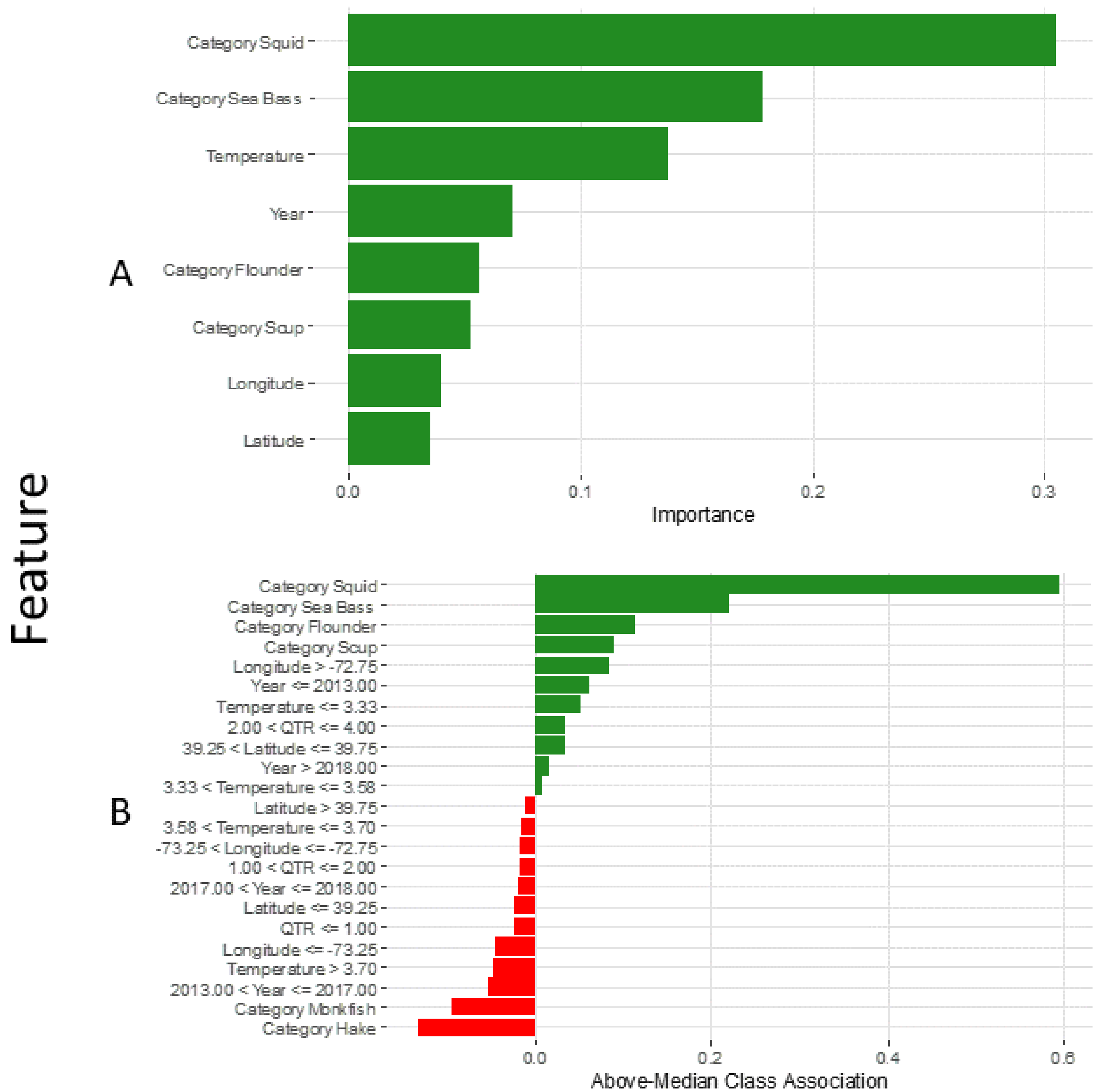

Figure 6. Feature importance after XBoost machine learning analysis for black seabass disposition kept

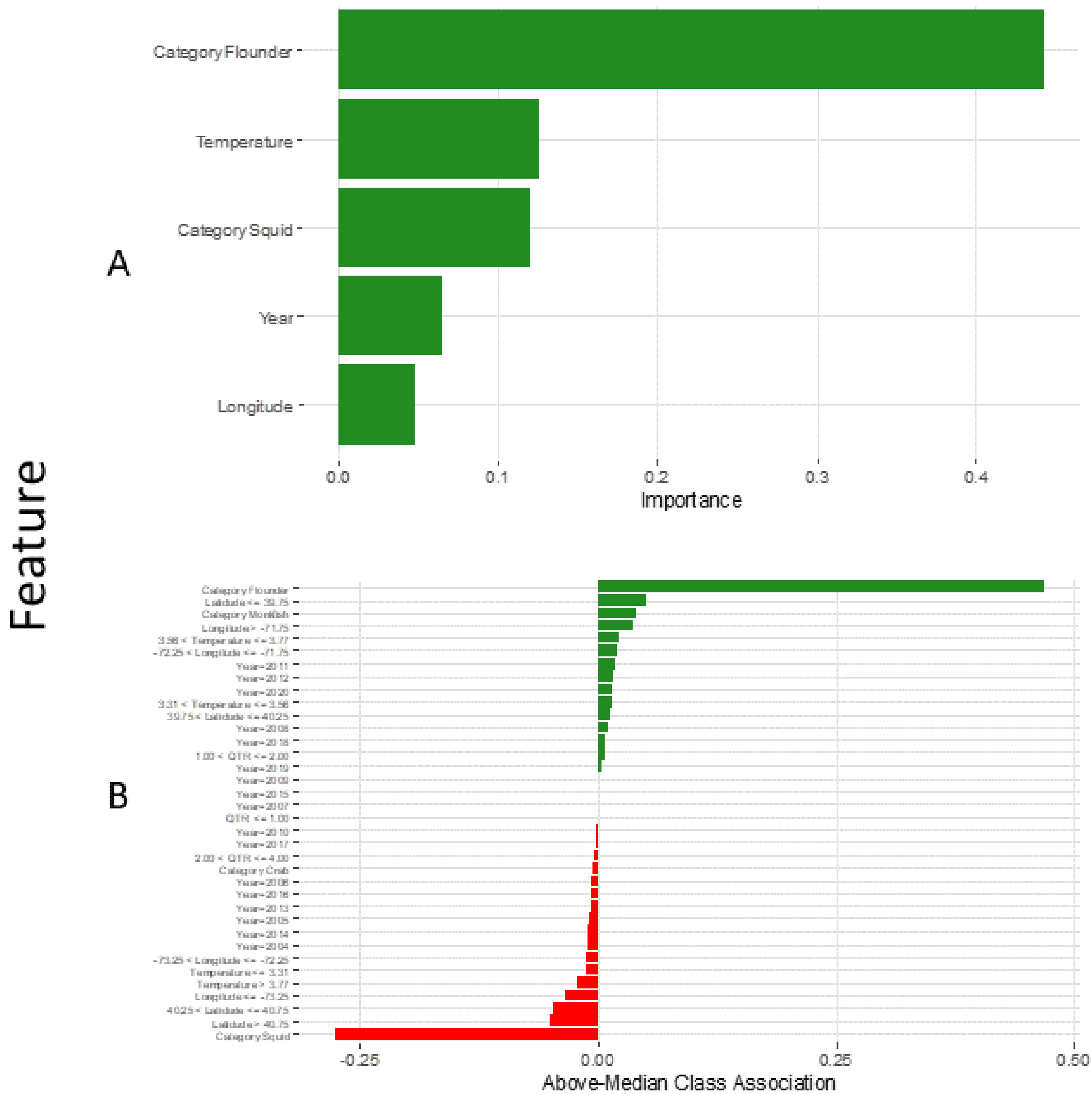

Figure 7. Feature importance after XBoost machine learning analysis for summer founder disposition kept

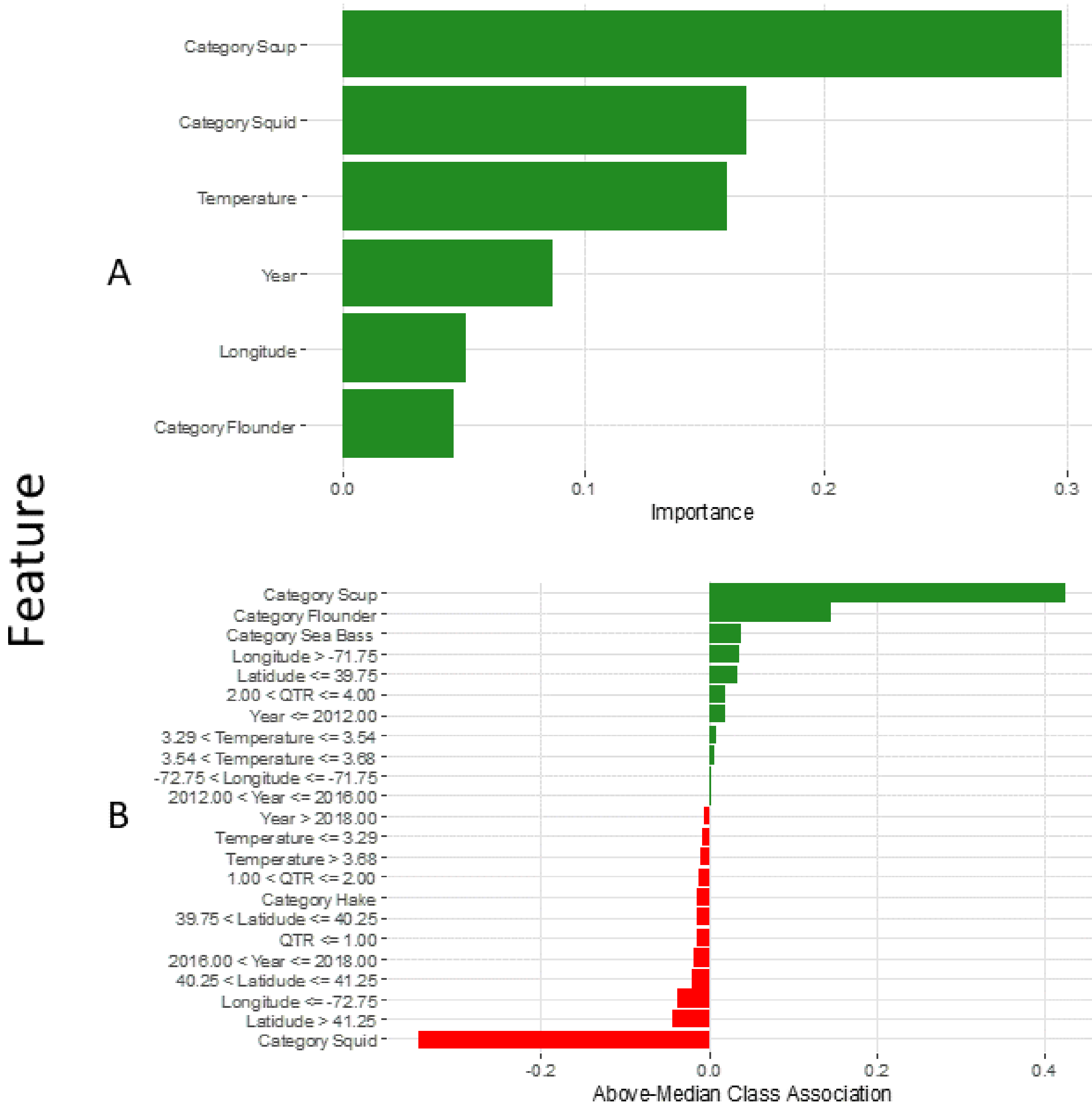

Figure 8. Feature importance after XBoost machine learning analysis for scup disposition kept

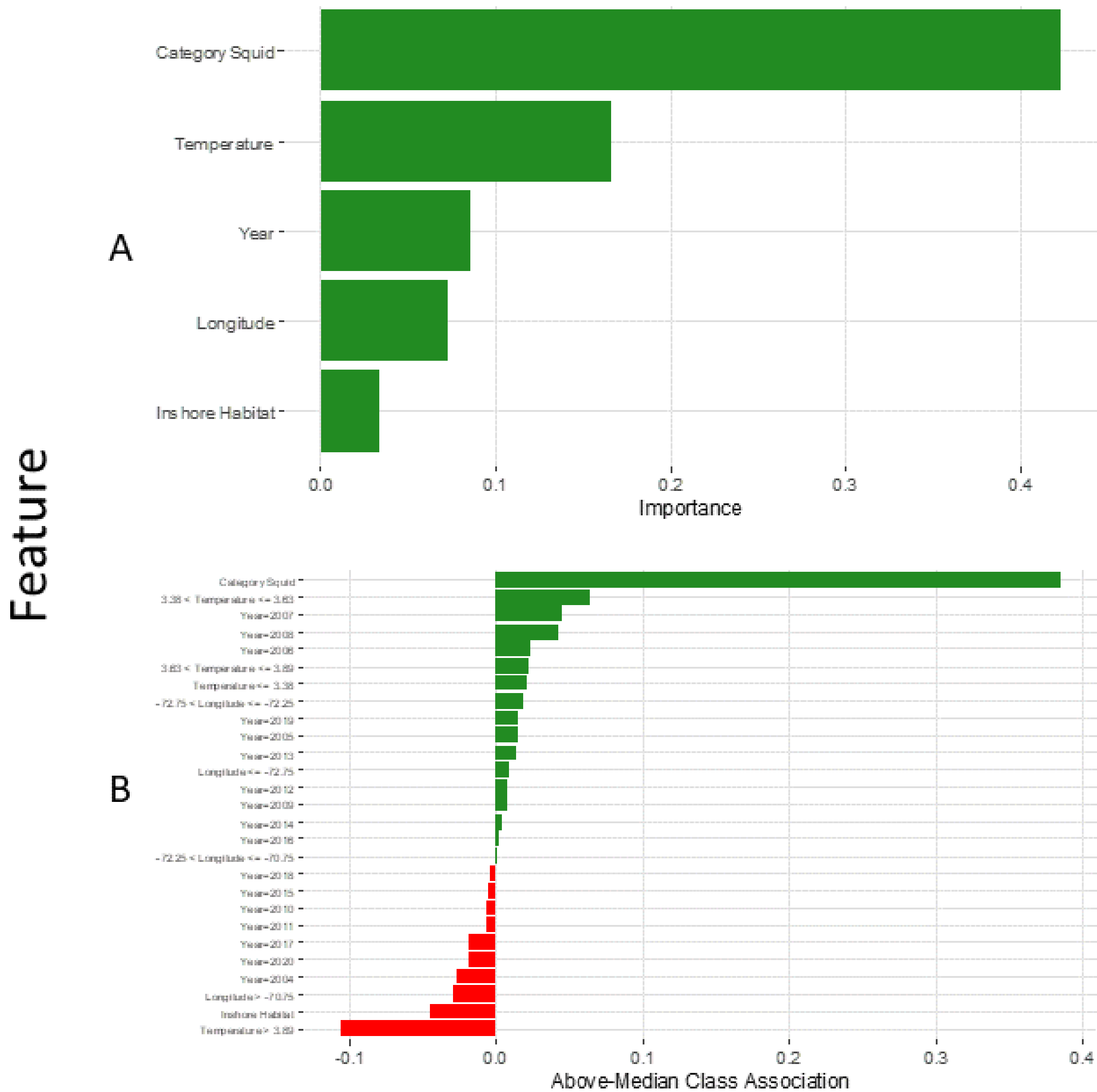

Figure 9. Feature importance after XBoost machine learning analysis for shorfin squid disposition kept

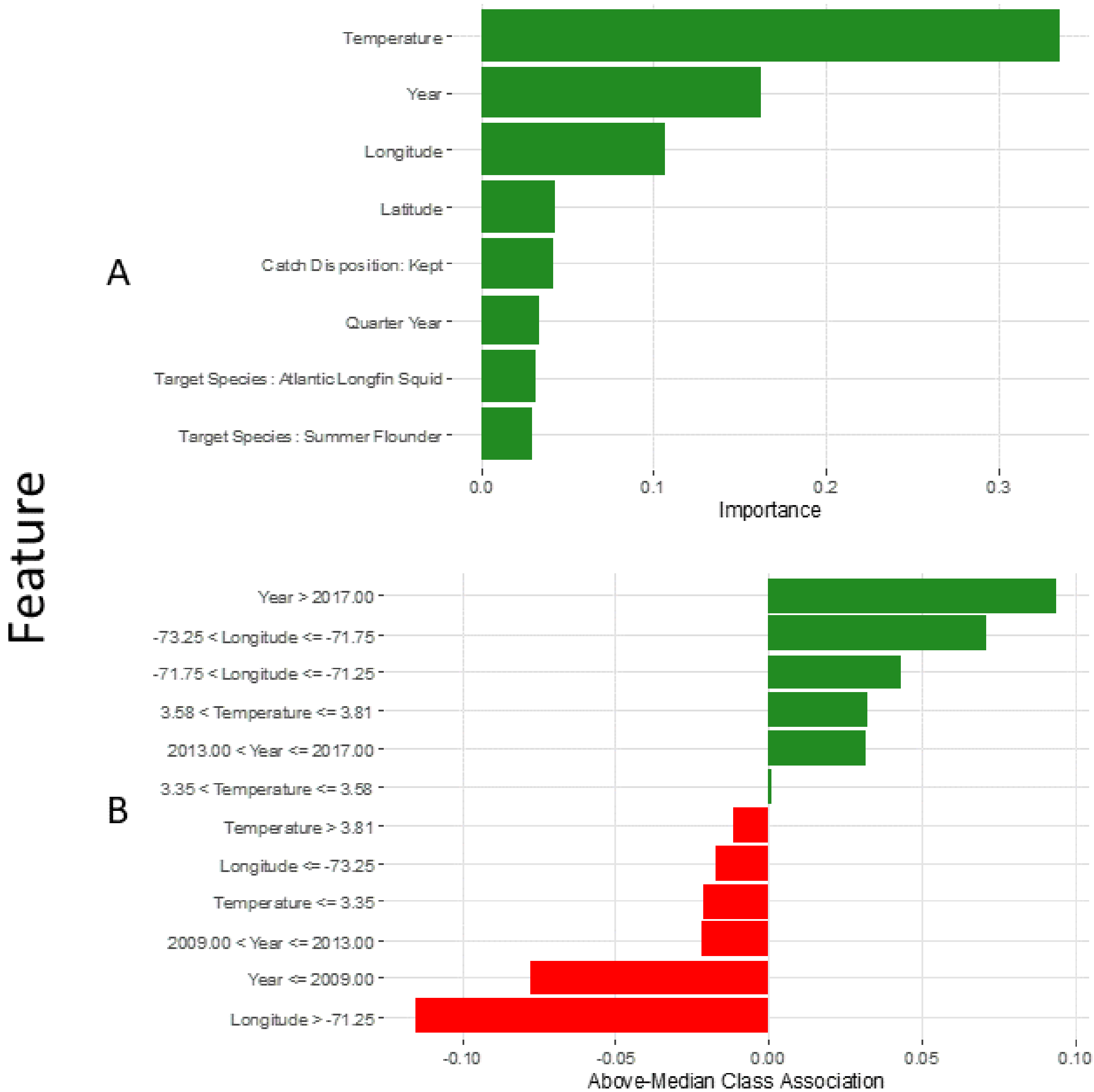

Figure 10. Feature importance after XBoost machine learning analysis for species richness.

Appendix I. NOAA Northeast Fisheries Science Center Observer and At Sea Monitoring Program species categories used in this study.

Category	Example Species
ALEWIFE	*Alosa pseudoharengus*
ANCHOVY	*Anchoa mitchilli*
ANCHOVY	*Anchoa hepsetus*
ARGENTINE	*Argentina silus*
BARRELFISH	*Hyperoglyphe perciformis*
BENTHIC INVERTEBRATE	*Littorina littorea, Mytilus edulis*
BLUEFISH	*Pomatonus saltatrix*
BLUESPOTTED CORNETFISH	*Fistularia commersonii*
BOARFISH	*Antigonia capros*
BONITO	*Sarda sarda*
BULLET MACKEREL	*Auxis rochei*
BUTTERFISH	*Peprilus triacanthus*
CLAM	*Tegillarca granosa, Spisula solidissima, Arctica islandica*
COBIA	*Rachycentron canadum*
COD	*Gadus morhua*
CODLING	*Pseudophycis bachus*
CRAB	*Callinectes sapidus, Chaceon quinquedens, Limulus polyphemus, Ovalipes ocellatus*

Category	**Example Species**
CROAKER	*Micropogonias Undulatus*
CUSK-EEL	*Tautogolabrus adspersus*
CUSK-EEL	*Brosme brosme*
CUSK-EEL	*Brosme brosme*
DORY	*Zeus faber*
DRUM	*Pogonias cromis, Sciaenops ocellatus*
EEL	*Anguilla rostrata, Nemichthys scolopaceus*
FLOUNDER	*Hippoglossoides platessoides, Hippoglossina oblonga, Paralichthys lethostigma*
GRENADIER	*Nezumia bairdii, Macrourus berglax*
HADDOCK	*Melanogrammus aeglefinus*
HAGFISH	*Myxine glutinosa*
HAKE	*Macruronus novaezelandiae, Phycis chesteri, Urophycis chuss, Merluccius australis*
HERRING	*Clupea harengus, Alosa aestivalis*
HOGFISH	*Lachnolaimus maximus*
JACK	*Caranx hippos*
KINGFISH	*Menticirrhus saxatilis, Menticirrhus americanus*
LADYFISH	*Elops saurus*
LAMPREY	*Petromyzon marinus*

Category	**Example Species**
LOBSTER	*Homarus americanus*
MACKEREL	*Scomber scombrus, Scomber colias*
MENHADEN	*Brevoortia tyrannus*
MONKFISH	*Lophius piscatorius*
MOONFISH	*Selene setapinnis*
MULLET	*Mugil caphalus*
MUMMICHOG	*Fundulus heteroclitus*
OCEAN POUT	*Zoarces americanus*
PERCH	*Morone americana*
PIGFISH	*Bodianus*
PILOTFISH	*Naucratinae ductor*
PINFISH	*Lagodon rhomboides*
POLLOCK	*Pollachius pollachius*
POMFRET	*Taractichthys longipinnis*
POMPANO	*Alectis ciliaris*
PUFFER	*Chilomycterus schoepfi, Sphoeroides maculatus*
QUAHOG	*Arctica islandica*
REDFISH	*Sebastes fasciatus*
RIBBONFISH	*Zu cristatus*
ROCKLING	*Enchelyopus cimbrius*
ROSEFISH	*Heliconlenus dactylopterus*
ROUGHY	*Gephyroberyx*

Category	Example Species
RUNNER	*Caranx crysos*
SAURY	*Scomberesox saurus*
SCAD	*Selar crumenophthalmus, Decapterus macarellus*
SCALLOP	*Argopecten irradians, Placopecten magellanicus*
SCULPIN	*Myoxocephalus octodecemspinosus*
SCUP	*Stenotomus chrysops*
SEA BASS	*Centropristis striata*
SEA POTATO	*Echinocardium cordatum*
SEA ROBIN	*Peristedion miniatum, Prionotus carolinus*
SEATROUT	*Cynoscion*
SHAD	*Alosa sapidissima, Dorosoma petense*
SHARK	*Squatina squatina, Rhizoprionodon terraenovae, Carcharias taurus*
SHEEPSHEAD	*Archosargus probatocephalus*
SHRIMP	*Pandalus borealis, Pleoticus robustus, Lysmata amboinensis*
SKATE AND RAY	*Menidia menidia, Dipturus laevis, Raja eglanteria, Leucoraja garmani, Rhinoptera bonasus*
SLENDER SNIPEFISH	*Amblyraja radiata*
SMELT	*Beringraja binoculata*

Category	Example Species
SNAIL	*Macroramphosus gracilis*
SNAKEBLENNY	*Osmerus mordax*
SNAPPER	*Lutjanus campechanus*
SNIPEFISH	*Leiostomus xanthurus*
SPOT	*Rhomboplites aurorubens*
SQUID	*Doeyteuthis pealeii, Illex Illecerosa*
STARGAZER	*Astroscopus guttatus*
STRIPED BASS	*Morone saxatilis*
TAUTOG	*Tautoga onitis*
TILEFISH	*Caulolatilus microps, Lopholatilus chamaeleonticeps*
TOADFISH	*Opsanus tau*
TUNA	*Thunnus thynnus, Euthynnus alletteratus, Thunnus albacares*
WEAKFISH	*Cynoscion regalis*
WHELK	*Busycotypus canaliculatus, Busycon carica, Sinistrofulgur perversum*
WHITING	*Merlangius merlangus*
WRECKFISH	*Polyprion americanus*
WRYMOUTH	*Cryptacanthodes maculatus*