EMBRACING MODERN METHODS IN FISHERIES; AN ENCOURAGING FIRST ATTEMPT AT USING MACHINE LEARNING TO MONITOR CATCHES IN THE DEMERSAL SHARK LONGLINE FISHERY IN SOUTH AFRICA

da Silva, C[1]., Chapman, N[2]. Rio, M[3]., Adams LA[4]., West, W[1], Booth, AJ[5], SJ Lamberth[1] & SE Kerwath[1]

[1]Department of Forestry, Fisheries and the Environment, Cape Town, South Africa

[2]WildTrust, Durban, South Africa

[3]University de Lille, Lille, France

[4]Marine Programme, South African National Biodiversity Institute, Newland, Cape Town, South Africa

[5]Department of Ichthyology and Fisheries Science, Rhodes University, Makhanda, South Africa

Abstract

Accurate fisheries catch data are essential if fisheries are to be sustainably managed. In South Africa, many fisheries have compulsory observer programmes paid for by industry, but the percentage of fishery activities observed is generally low (<5%) with poor spatial coverage. Smaller fisheries, such as the demersal shark longline fishery, have no observer coverage except for a few months in 2008/2009. This gap in observer data could, however, be filled by an electronic monitoring system (EMS). The demersal shark longline fishery has been controversial since its inception and has been the focus of negative press around allegations of high mortality rates of Endangered, Threatened and Protected (ETP) species. To improve observer coverage and to monitor ETP species, the DFFE in collaboration with WildTrust and the fishery, initiated a collaborative project to install an EMS on an active vessel. To date, 13,665 videos have been collected since December 2023 with 3,538 videos processed for still images. An initial run analysed 113 still images of sharks, batoids, and teleosts from the processed videos were uploaded to BIIGLE and annotated. A total of 337 annotations were then analysed in YOLOv 5, an object recognition algorithm. The initial model was trained using 75% of the images and tested with the remaining 25%. Unannotated images were also used to evaluate the model's performance and feasibility. Species-level identification was not feasible due to the limited number of images. The model, however, successfully differentiated batoids, sharks, and teleosts from each other with a precision of between 72 and 73%. and recorded a higher species diversity than from logbook data. The precision and rate of recall will improve with additional training images. The next stage of the project will use additional deep learning methods to automatically extract video segments of catch events, which would substantially reduce storage space and review time by analysts.

Introduction

The collection of accurate fisheries catch data is critical to ensuring sustainable management of fisheries, mitigating their environmental impacts and for managing stocks. Additionally, fisheries management authorities rely on accurate data to manage fishing through effort or harvest controls, determine appropriate licence fees, to monitor catch of Endangered Threatened and Protected species (ETP) and to manage other environmental impacts (Brown *et al.* 2021). Sustainable fisheries management relies on both fisheries-independent and fisheries-dependent data to assess and manage stocks. Fishery-independent data are gathered through scientific surveys conducted independently of commercial fishing operations, and includes information on fish abundance, distribution, size, and age structure. These data can be collected using fishing gear similar to the commercial fishing operation being assessed, acoustic surveys and tagging programmes. These data are crucial for providing unbiased estimates of fish population and ecosystem health.

On the other hand, fishery-dependent data are derived from the fishing industries, including self-reporting logbooks which collects catch, effort and bycatch data. The reliability of fishery-dependent information for assessing fish population status has often been questioned (Cotter & Pilling, 2007). Catch data from logbooks alone cannot be used to assess the stock status of fish caught by a particular fishery as they are a product of complex human behaviours involved in completing the logbooks (Branch et al. 2011). Reports on landings do not capture the full extent of all fish caught, as a significant portion of total catch may be discarded at sea (Poos et al. 2013: Uhlmann et al 2014, Ulleweit et al. 2010). This is especially true for ETP species where fishers downplay problematic bycatch interactions or species limited by quotas such as “choke” species, data are unreliable and often non-existent (Borges, 2015). Catch may be misreported (underestimated, overestimated), or not reported at all and may be exacerbated by the decentralised nature of many fishing sectors, poor monitoring and poor enforcement. Estimating total catch per species is often complicated by illegal harvest, discards and catches from subsistence fishers that are not reported (Pitcher et al. 2002). Globally it has been estimated that the overall underreporting of catch for fish populations average at 53% (Pauly and Zeller 2016).

It is therefore vital that these data are validated by Observer programmes, which independently monitor the fishing operations, effort, catch, bycatch and compliance with permit conditions or other sector-specific fishing regulations. Observer programmes allow for the verification of the accuracy of vessel-reported logbook data and may inform fisheries management. Representative coverage is, however, expensive (Suuronen and Gilman, 2020). In addition, there is an assumption made that data collected (especially catch, effort and discard of ETP species) on observed fishing trips are representative of unobserved trips (Hall, 1999). As selection of trips are often non-random and not stratified, fishing practices can be altered once an observer is on board and is termed the “observer effect” (Duarte and Cadrin, 2024). The observer effect may be intentional whereby a captain may target an area with lower ETP encounter rates, may deviate from mandatory bycatch mitigation practices (such as not use Tori lines). It could also be unintentional because carrying an observer is costly in terms of daily rate, but also in terms of losing a member of crew onboard and may result in a trip being shorter than average (Duarte and Cadrin, 2024; Faunce and Barbeaux, 2011). If this observer effect occurs frequently, it violates the assumption of representativity (Babcock et al. 2003). In general, in fisheries without 100% observer coverage, data collected from observer trips are used to scale up discards of ETP species across an entire fleet using a ratio estimator (Duarte and Cadrin, 2024: Cochran, 1977), either based on effort or landings. For observer data to be useful, it needs to be representative and stratified so that data scaling/expansion can be done by gear type or region, with confidence intervals providing estimates on variability in the observed trips within each stratum (Duarte and Cadrin, 2024). Non-representative sampling can lead to biased estimates and consequently errors in catch estimates, estimates of ETP discards and ultimately inaccurate stock assessments (Rudd and Branch, 2017).

Traditionally, monitoring compliance in longline fisheries has relied heavily on onboard human observers, who are tasked with recording catch data, bycatch incidents, and adherence to regulations. However, the challenges of deploying observers across vast oceanic areas, coupled with safety concerns and the high cost of human observation, have highlighted the need for more efficient and scalable solutions.

Observer programs are crucial for collecting reliable data on bycatch and discarding activities at sea, but their effectiveness is often limited by available resources. To enhance fisheries stock assessments, endangered species protections, and ecosystem management, these programs must be designed to meet their specific objectives, which typically require high precision and minimal bias in bycatch estimates (Babcock et al. 2003). A significant challenge is the bias introduced by non-random sampling, as many observer programs rely on vessels that volunteer or are willing to carry observers, leading to potentially unrepresentative data. Studies by Liggens et al. (1997) and Sampson (2002) highlight the importance of assessing whether the data collected are truly representative of the entire fleet. Comparing the catches of observed and unobserved vessels should be an ongoing part of any observer program, and a mandatory, randomly allocated observer program would yield more reliable results. Once observer effects and sampling bias are addressed, the necessary level of precision in bycatch estimates should be determined based on how the data will be used (Babcock et al. 2003). This process involves determining the required level of sampling effort, which can be guided by comparing with other programs or using general sample size considerations. Simulation studies suggest that coverage levels of at least 20% for common species and 50% for rare species can provide reasonably accurate bycatch estimates (Babcock et al. 2003). Observer coverage is generally low <10%, with most global discard estimates relying on costly observer programmes, which cover less than 1% of fishing activities (Benoît & Allard, 2009; Depestele et al., 2011; Poos et al., 2013; Rochet, Péronnet et al, 2002).

In the absence of representative coverage, Electronic Monitoring (EM) in longline fisheries has emerged in the past two decades as a transformative technology, revolutionizing the way fishing activities are observed, recorded, and regulated (Ames et al. 2007: Stanley et al 2011; van Helmond et al. 2020). The initial EM systems were industry-led programmes to deal with changes in fisheries management and gear theft in the British Columbia Crab fishery (Ames et al. 2007). The potential of EM system was quickly recognised as a potential panacea in fisheries challenged by poor observer coverage and monitoring of ETP species.

EM systems, which typically consist of a combination of video cameras, GPS, activity sensors and computer hardware, offer a promising alternative to traditional and costly observer programmes (van Helmond et al. 2020). These systems automatically record fishing activities, providing a continuous, tamper-proof record that can be reviewed and analysed either in real-time or post-trip. By capturing detailed footage and data, EM allows for accurate documentation of fishing practices, species identification, and compliance with fisheries management regulations (van Helmond et al. 2020). As the global demand for sustainable fishing practices intensifies, the adoption of EM in longline fisheries is increasingly seen as essential.

A comprehensive review of EM systems can be found in van Helmond et al. (2020). EM programmes have been fully adopted in many different fisheries worldwide and include trap fisheries targeting crustaceans in British Columbia, Canada, midwater trawl fisheries targeting hake in British Columbia, longline fisheries targeting Atlantic Tuna, purse seine fisheries catching tropical tuna in the Indian and Atlantic Ocean (Ames et al. 2005; van Helmond et al 2020; Michelin *et al.* 2018). At the date of the Helmond et al. (2020) review, 12 fully implemented EM systems were in place with a further 100+ EM trials underway. The biggest and longest running EM system is the British Columbia, Groundfish hook and line/Trap Catch Monitoring program with a total 200 vessels monitored. The most comprehensive multi-species, multi fishery EM programme is done by the Australian Fisheries Management Authority (AFMA) Electronic Monitoring Programme which monitors longline, hand line, gill net and trap net fisheries (Hoskin et al. 2017).

Despite such a large uptake in the use of EM systems globally, there are still considerable challenges associated with storing and analysing large volume of video captured. Analysis of footage that can amount to 2TB of video data per single fishing trip (Gerner, 2015) can be labour intensive, tedious, error prone, subjective and extremely costly (Qaio *et al.,* 2021). Most fisheries that use EM systems do not fully utilise all footage captured (Wallace et al. 2015); they either randomly select video sequences, randomly select a percentage of footage to review (5-56%); randomly select a single haul per trip or a census of video data is taken played at 8-12 times real time (van Helmond *et al.* 2020). This depends on the purpose of the monitoring; in general (but not always) where discards and bycatch of ETP species is of concern, a census of video data is done, where landings are of concern random percentage of footage is reviewed (van Helmond et al. 2020).

Recent leaps in the application of artificial intelligence (AI) technology, specifically deep learning techniques and its application to computer vision, has been shown to provide possible solutions to the costly and time-intensive bottleneck of analysing digital data. Computer vision has provided numerous applications in marine sciences such as rapid biodiversity assessment and monitoring (Mahmood *et al.* 2016), species identification (Storbeck and Daan 2001), length frequency measurement (White et al. 2006), behaviour (Papadakis et al. 2012), estimates of abundance (Ditria et al. 2020), and ecosystem classification (Piechaud et al. 2019). The application of Computer Vision has extended to the fisheries realm (Malde *et al.* 2019; Probst, 2019). A fishery-dependent collection system has been proposed by Bradley *et al*. (2019) using AI automation to reduce lags between data collection and action. However, despite EM systems being adopted and trialled globally, automated video analysist systems have not been widely applied (van Helmond et al. 2020), and if applied not yet documented in detail. To our knowledge, the first documented EM system using computer vision was trailed by Qaio *et al*. (2021), with promising results even from a small training dataset.

The use of computer vision is limited and preliminary in South Africa with various trials underway over the past two years. These initiatives include benthic surveys, species identification, behaviour of seabirds (da Silva et al. 2023). All the preliminary projects are in progress, however, there is substantial collaborations in South Africa with research groups sharing expertise.

The demersal shark longline fishery has been in place since 1991 and uses bottom set longlines to target tope (*Galeorhinus galeus*) and smoothhound sharks (*Mustelus mustelus*), over the past 30 years the number of active permitted vessels have reduced from 11 to one in 2022 (DFFE, 2023). The demersal shark longline fishery has been controversial since its inception and has been the focus of negative press around allegations of high mortality rates of Endangered, Threatened and Protected (ETP) species. In South Africa, many fisheries have compulsory observer programmes but the coverage is low (<5%) and with poor spatial coverage. Smaller fisheries, such as the demersal shark longline fishery, have been historically bereft of observer coverage except for a few months in 2008/2009. Given funding constraints and capacity issues within the Fisheries branch of the Department of Fisheries, Forestry and the Environment (DFFE) in South Africa, here we explore the use and feasibility of using open-source machine learning methods with infrastructure available at DFFE for EM data collected from the demersal shark longline fishery in South Africa.

Materials and Methods

To address the lack of observer coverage and to collect comprehensive data typically gathered by observer projects, the Department of Forestry, Fisheries and the Environment (DFFE), in collaboration with WildTrust and the fishery, initiated a collaborative project to install an Electronic Monitoring (EM) System on the last active vessel in the fleet. The EM system was initially installed in 2020. However, due to a combination of factors including the COVID-19 pandemic, vessel repairs, the listing of requiem sharks under CITES Appendix II, and interruptions caused by the Fisheries Rights Allocation Process in 2019/2020, fishing activities only commenced in December 2023.

The EM system comprises three motion-controlled Hikvision Megapixel outdoor cameras (model number DS-2XC6142FWD-IS(C)). These cameras were strategically positioned: one to capture deck activities, another on the retrieval line to monitor discards, and a third on the stern to observe seabird interactions and the use of the mandatory Tori line. The system utilizes a Hikvision Mobile 4-channel IP Network Video Recorder with 3G connectivity and a generic GPS module, along with multiple storage hardware components. Hard drives are removed and couriered every 2-3 months to ensure data integrity and continuity.

To protect the rights of the permit holder, data confidentiality contracts were established. Since December 2023, a total of 13,665 videos have been collected, with 3,538 videos processed for still images. From these, 113 still images of sharks, batoids, and teleosts were extracted using VSPlayer and subsequently uploaded to BIIGLE. In BIIGLE, the videos were annotated using the rectangle tool and classified according to the WORMS label tree into categories such as Selachii (sharks), Batoidea (skates and rays), and Osteichthyes (bony fishes).

Further classification was not completed due to the limited size of the dataset. The framework developed by Howel and Davies (2010) and Howell et al. (2023), available on [GitHub](https://github.com/DeepSeaCRU/CRUresources). This framework has been used for automating the counting of benthic invertebrate taxa (Piechaud et al. 2019, and Piechaud and Howell 2022). The framework was utilized to convert the BIIGLE annotations to the YOLO (You Only Look Once) format in R. The R script also prepared the dataset (images and annotations) for training the YOLO Convolutional Neural Networks (CNN) object detection model in Google Colab.

The dataset prepared in R was subsequently processed through the YOLOv5 object detection CNN in Google Colab, hosted by the Jupyter Notebook service, using the framework developed by DeepSeaCRU. Training was conducted using a train-from-scratch (TS) approach. Since the computational resources required to train the object detection model is executed using the free Google Colab cloud services and is independent from a local machine's hardware specifications, the entire framework can be executed in a web browser and on an entry level computer e.g. a standard government issued laptop with an Intel i5 processor and 4GB of RAM. The model's training parameters were 640x640 input pixel resolution, in batches of 16, over 100 epochs and a 75/25 percent split between the train and validation images. This dataset comprised 113 images, annotated with 185 instances of Selachii, 114 instances of Osteichthyes, and 38 instances of Batoidea.

Evaluation metrics

Two metrics were applied within the YOLOv5 framework to evaluate the accuracy of the object detector; precision and recall (Powers, 2011; Qiao et al. 2021), which are defined as

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

where a true positive refers to predicted bounding boxes that are matched to ground-truth positives, i.e. ground-truth bounding boxes for animals, while a false positive refers to the predicted bounding boxes having no overlap with any ground-truth bounding boxes. False negatives refer to the ground-truth bounding boxes missed by the detector. Precision and recall evaluate different aspects of the detector. Precision calculates the percentage of predicted bounding boxes that are relevant, whereas recall measures the percentage of the relevant bounding boxes that have been retrieved relative to the total number of ground-truth bounding boxes. The average precision (AP) score and recall rate are calculated as in Manning et al. (2008) to numerically summarize the precision of the detector. There is one adjustable parameter for the evaluation metrics in the detection framework i.e. the *Intersection over Union (IoU)* threshold. The IoU represents the proportion of overlap between a predicted object and its corresponding ground-truth. For evaluation purposes, successful detections were determined to be true if the *IoU* was above an *IoU* threshold set to 0.5.

A confusion matrix (Stehman, 1997) was used via the Howel and Davies (2010) framework was used to evaluate the performance of the catch detector in terms of correctly identifying animal within the three categories.

Results and Discussion

Random observation of videos showed that the EMS system installed on the vessel recorded a higher species diversity than logbooks. In general, data in logbooks reflect high value catch species such as tope, smoothhound and requiem sharks (*Carcharhinus obscurus* and *C. brachyurus*), and live release of ETP species of concern such as sandtiger sharks (*Carcharias taurus*) and smooth hammerhead sharks (*Sphyrna zygaena*). Live release of unwanted endemic shysharks such as pyjama sharks (*Poroderma africanum*), leopard catsharks (*P. pantherinum*) and other unwanted teleosts such as white seacatfish (*Galeichthys feliceps*) was not recorded in logbooks. Species identification from reviewing footage alone was possible for most species with the exception of some of the cryptic requiem sharks and a few skates and possibly teleosts. Given the preliminary status of this study, analytical comparison of logbook footage versus EM footage was not possible.

These preliminary results do not evaluate the use of various CNN frameworks as many of the commercially available proprietary CNN frameworks may not be feasible for use in South Africa due to large costs. Results from the YOLOv5 CNN frame by frame object detection results from the testing data set are shown in Table 1. Precision (p) measures the proportion of true positive predictions out of all positive predictions made by the model. In general, the precision was highest for Selachi (0.729, n= 185 annotations), followed by Batoidea (0.733, n=38 annotations) with the least precision for Osteichthyes (0.724, n = 114 annotations). This is likely due to the low number of training images for Batoidea and high variation in colour and shape for Osteichthyes. The model showed fairly consistent precision across all classes, with values around 0.729 and 0.733, this indicates that a good proportion of the models' positive predictions are correct. Recall (R) measures the proportion of true positive predictions out of all actual positives. Overall recall was high at 0.747, Recall was highest for Batoidea at 0.857, indicating the model is very good at detecting them when they are present. This is likely related to their distinctive shape. Recall was lower for Selachii (0.717) and lowest for Osteichthyes (0.667).

The Mean Average Precision at IoU threshhold (mAP50) (Table 1) gives an overall indication of the model's detection accuracy. It is an average precision score at a specific Intersection over IoU threshold of 0.5. Detection accuracy was highest for Batoids (0.732), followed by Selachii (0.712) and then Osteichthyes (0.711). Scores suggest that the model performs reasonably well across all classes, with mAP50 values of 0.717 to 0.76, indicating decent detection accuracy.

The Mean Average Precision across 0.5 to 0.95 IoU threshold (mAP50-95) (Table 1) is a more stringent metric than mAP50 and is calculated across different IoU thresholds from 0. 5 to 0.95. It reflects how well the model performs across varying levels of strictness in overlap between the predicted and the true bounding boxes. mAP50-95 was highest for Batoidea (0.345), then Selachii (0.312), followed by Osteichthyes (0.293). The lower mAP50-95 compared to mAP50 suggest that the model's performance decreases as the IoU threshold increases, indicating that it may struggle with precise localisation of the objects.

The model, overall, shows good performance in detecting and classifying the Batoidea, with a high recall and reasonable precision, which is promising for accurately identifying his class on this fishing vessel. Selachii and Osteichthyes show slightly lower performance, particularly in recall for Osteichthyes, which implies that the model sometimes fails to detect these instances. Lower mAP50-95 values highlight that while the model is generally accurate, it might struggle with precise object localisation, particularly under stricter conditions. These preliminary results suggest that this deep learning model can be effectively used in an EM programme to monitor and classify species on this vessel. However, further improvements are needed to improve the recall for Osteichthyes as the model sometimes fails to detect them.

Table 1. Yolov5 model training results output for the EM footage

Class	Images	Instances	P	R	mAP50	mAP50-95
Selachii	29	60	0.729	0.717	0.712	0.312
Batoidea	29	7	0.733	0.857	0.732	0.345
Osteichthyes	29	21	0.724	0.667	0.711	0.293
All	29	88	0.729	0.747	0.718	0.317

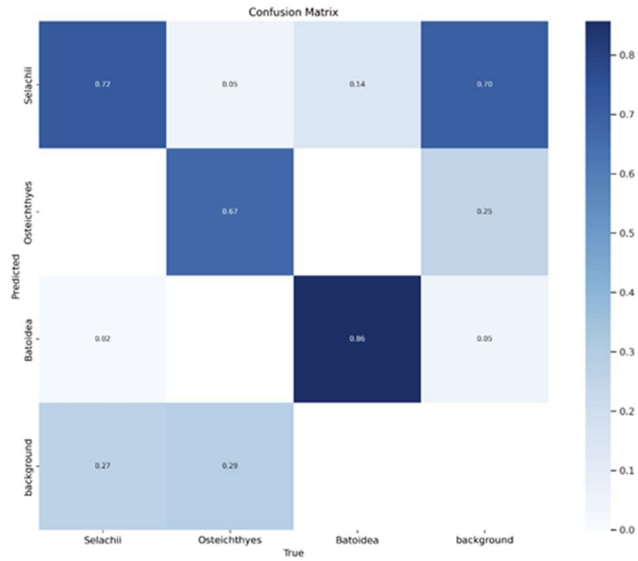

Fig 1. Confusion Matrix for Model Performance Evaluation.

The confusion matrix (Figure 1) visualises the performance of the YOLOv5 model across four classes: ‘Selachii’, ‘Osteichthyes’, ‘Batoidea’, and ‘background’. The matrix highlights the true positive rates along the diagonal and the misclassification rates in the off-diagonal cells, providing a comprehensive overview of the model’s accuracy and areas for improvement. Diagonal elements represent the cases where the predicted class matches the true class i.e. true positives. Batoidea had the highest percentage of true predictions (85%), followed by Selachii (72%) and then Osteichthyes (64%). The off diagonal elements represent misclassifications where the predicted class did not match the true class, 5% of Osteichtheys were incorrectly predicted as Selachii, 14% of Batoidea, were mistakenly identified as Selachii and 2% of Selachii were incorrectly identified as Batoidea and 70% of the background was wrongly predicted as Selachii. The model performed well with Batoidea and Selachii classes, as indicated by the high values on the diagonal (0.85 and 0.72), respectively. Osteichthyes and the background are more frequently misclassified, with noticeable confusion between Osteichthyes and other classes. The background class also shows significant misclassification, especially being confused with Selachii.

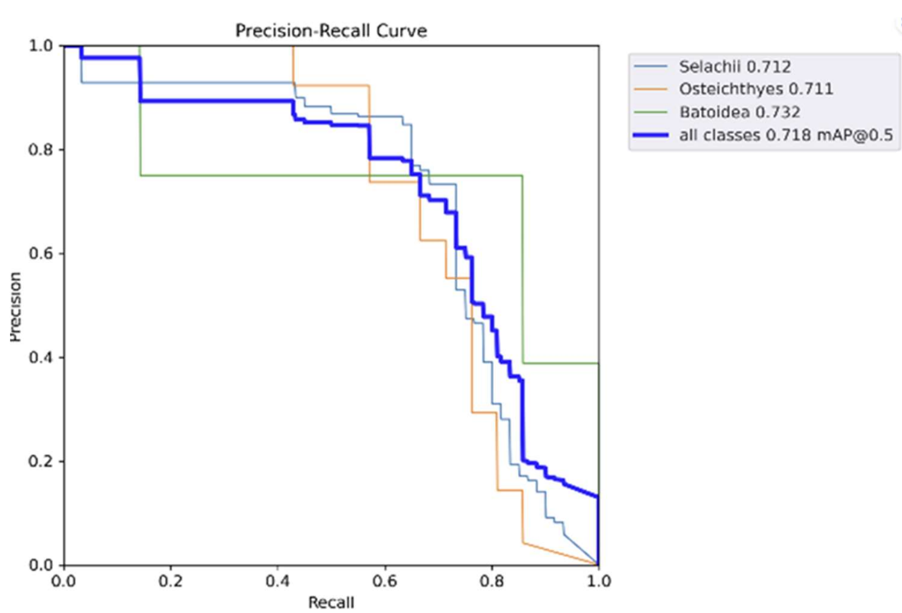

Figure 2: The precision-recall curve for three classes: Selachii, Osteichthyes, and Batoidea, and a combined line for all classes. The Mean Average Precision (mAP) scores set at a threshold of 0.5 are 0.712 for Selachii, 0.711 for Osteichthyes, 0.732 for Batoidea and 0.718 for all classes combined

Precision-Recall Curve (Figure 2), a crucial tool in evaluating the performance of classification models, particularly in scenarios where the classes are imbalanced. This curve plots precision (on the y-axis) against recall (on the x-axis) at various threshold settings, providing a comprehensive view of the trade-off between these two metrics.

Recall measures the proportion of actual positives that are correctly identified by the model (ranging from 0-1, where 1 means that all predicted positives are correct), specifically “Of all the actual positive instances, how many were correctly predicted as positive?” Precision measures the proportion of positive predictions that are actually correct (ranging from 0-1 where1 means that all predicted positives are correct), specifically “Of all the instances predicted as positive, how many were actually positive?”

The Precision-Recall Curve includes three lines representing different classes evaluated by the classification model. For Selachii, the Average Precision (AP) Score is 0.712. this summarises the precision-recall curve as the weighted mean of precisions achieved at each threshold, with an increase in recall from the previous threshold sued as the weight. An AP score of 0.712 indicates a relatively good performance for this class. For Osteichthyes, the AP score was 0.77, similar to Selachii, the score reflects the model’s ability to balance precision and recall. The close AP scores for Selachii and Osteichthyes suggest comparable performance for these classes. Batoidea had a higher AP score of 0.723; indicating that the model performs better in distinguishing Batoidea from other classes. This is not surprising given their morphometric differences to both the Selachii and Osteichthyes.

The Precision-Recall Curve is particularly useful in datasets with imbalanced datasets, this is the case with this small data set with 185 Selachii, 114 Osteichthyes and 38 Batoidea. Model performance was high for Batoidea especially given the low comparative number of annotations used for training. The Curve also assists with choosing optimal threshold that balances precision and recall according to the specific requirements of the application. In this case, high recall will be prioritised to ensure that all positive cases are identified, even at the cost of lower precision.

Conclusion and recommendations

Although EM has been widely trailed and fully implemented in many fisheries worldwide, automation of the process is still lacking (van Helmond et al. 2020). In this study, we piloted deep learning methods to fill this gap using open-source deep learning methods accessible to marine scientists and few resources. This preliminary proof of concept use of deep learning methods applied to electronic data has shown promising results even with a small training dataset (n=113 images) that represented 36 longline sets from the EM system installed on the demersal shark longline vessel. Species-level identification was not feasible due to the limited number of images. The model, however, successfully differentiated batoids, sharks, and teleosts from each other with a precision of between 72 and 73% and recorded a higher species diversity than from logbook data. The deep learning-based object detectors worked well for this task despite the small training set. This may be due to the small number annotation per class. The precision and rate of recall will improve with additional training images, including images with only background and training the model for longer. Due to limitations within BIIGLE and privacy issues, people were excluded from annotations. Therefore, the model includes people as backgrounds, often the confusion with backgrounds is as a result of white gumboots worn by fishers or white clothing. In addition, fishers use a white plastic bin to store small sharks which obscured sharks and rays. Going forward, background images will include people with wearing various personal protective fishing gear, background with no catch and backgrounds with or without catch in different weather conditions day and night. The use of other annotation software will be trialled to include fishers and fish in the same frame and identify “catch events” similar to what was used by Qaio *et al.* 2021. Once model performance is improved with the addition of a larger representative training set, we will investigate using additional deep learning methods to automatically extract video segments of catch events, which would substantially reduce storage space and review time by analysts. The intention of this study is to assist analysts to review EM to generate accurate catch reports efficiently. Current limitations of the study are that it constantly records images, due to errors during setup with the trigger from the main line retrieval capstain which would have ensured that only active fishing was recorded. This was not initially seen as a problem as Rights Holders record start of fishing time, however if time is not recorded accurately in in logbook it adds considerable processing time. Since this fishery operates with a single vessel, it is not necessary to extend the training set to other vessels for training. In the future these methods may be extended to the Large Pelagic Longline fleet in South Africa as several Rights Holder have installed cameras already on their vessels for security reasons. In conclusion, the deep learning methods applied to the EM system on the demersal shark longline vessel shows great promise and is feasible due to the ability to use open-source models on government issued infrastructure. It has the potential to enhance transparency and accountability of a fishery that has been in the media frequently due to allegations of ETP bycatch and also supports the conservation of marine ecosystems by enabling more effective monitoring of bycatch, including ETP species.

Acknowledgements

The authors would like to thank the support from WildTrust for funding the EM system on the vessel and the Rights Holders for allowing us to use their vessel for this project.

Data availability

The data underlaying this article is held at DFFE South Africa and has been used with permission. The data underlying this article cannot be shared publicly due to the in-confidence nature of on-vessel footage.

References:

Ames, R.T., 2005. The efficacy of electronic monitoring systems: a case study on the applicability of video technology for longline fisheries management (p. 64). Seattle, WA: International Pacific Halibut Commission.

Babcock, E.A., Pikitch, E.K., & Hudson, C.G., 2003. How much observer coverage is enough to adequately estimate bycatch? Pew Institute of Ocean Science. 〈https://oceana.org/reports/how-much-observer-coverage-enough-adequately-estimate-bycatch〉

Benoît, H.P. and Allard, J., 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? *Canadian Journal of Fisheries and Aquatic Sciences*, 66(12), pp.2025-2039.

Borges, L., 2015. The evolution of a discard policy in Europe. *Fish and fisheries*, 16(3), pp.534-540.

Bradley, D., Merrifield, M., Miller, K.M., Lomonico, S., Wilson, J.R. and Gleason, M.G., 2019. Opportunities to improve fisheries management through innovative technology and advanced data systems. *Fish and fisheries*, 20(3), pp.564-583.

Branch, T.A., Jensen, O.P., Ricard, D., Ye, Y. and Hilborn, R., 2011. Contrasting global trends in marine fishery status obtained from catches and from stock assessments. *Conservation biology*, 25(4), pp.777-786.

Brown, C.J., Desbiens, A., Campbell, M.D., Game, E.T., Gilman, E., Hamilton, R.J., Heberer, C., Itano, D. and Pollock, K., 2021. Electronic monitoring for improved accountability in western Pacific tuna longline fisheries. *Marine Policy*, *132*, p.104664.

Cochran, W.G., 1977. Sampling Techniques, 3rd Edition. John Wiley & Sons, New York.

Cotter, A.J.R. and Pilling, G.M., 2007. Landings, logbooks and observer surveys: improving the protocols for sampling commercial fisheries. *Fish and Fisheries*, 8(2), pp.123-152.

da Silva, C., Samaai, T., Kerwath, S., Adams, L., Watson, K., Bernard, A., van der Heever, G., Angel, A., Schoombie, S., Frainer, G. and Franken, M.L., 2023. Leaping into the future: Current application and future direction of computer vision and artificial intelligence in marine sciences in South Africa. *Research Ideas and Outcomes*, 9, p.e112231.

Depestele, J., Vandemaele, S., Vanhee, W., Polet, H., Torreele, E., Leirs, H. and Vincx, M., 2011. Quantifying causes of discard variability: an indispensable assistance to discard estimation and a paramount need for policy measures. *ICES Journal of Marine Science*, 68(8), pp.1719-1725.

Ditria E, Lopez-Marcano S, Sievers M, Jinks E, Brown C, Connolly R (2020) Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning. *Frontiers in Marine Science* 7 https://doi.org/10.3389/fmars. 2020.00429

DFFE (Department of Forestry, Fisheries and the Environment) 2023. Status of the South African marine fishery resources 2023. Cape Town: DFFE

Duarte, D. and Cadrin, S.X., 2024. Review of methodologies for detecting an observer effect in commercial fisheries data. *Fisheries Research*, *274*, p.107000.

Faunce, C.H., Barbeaux, S.J., 2011. The frequency and quantity of Alaskan groundfish catcher-vessel landings made with and without an observer. *ICES J. Mar. Sci*. 68 (8), 1757–1763..

Gerner, M. 2015. Cost effective monitoring in Australia's tuna longline fisheries. In International Workshop on the Application of Electronic Monitoring Systems in Tuna Longline Fisheries, Kaohsiung City, Taiwan.

Hall, M.A., 1999. Estimating the ecological impacts of fisheries: What data are needed to estimate bycatches? Proceedings of the International Conference on Integrated Fisheries Monitoring, 175–184. https://www.fao.org/3/x3900e/x3900e04.htm〉

Hosken, M., Williams, P. and Smith, N., 2017. A brief update on ER and EM progress in the region. *WCPFC–Scientific committee thirteenth regular session, 9–17 August 2017, Rarotonga, Cook Islands*.

Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher RB (2016) Automatic annotation of coral reefs using deep learning. OCEANS 2016 MTS/IEEE Monterey https://doi.org/10.1109/oceans.2016.7761105

Howell, K. and Davies, J., 2010. Deep-sea species image catalogue. DeepSeaCRU. Marine Biology and Ecology Research Centre, University of Plymouth. On-line version, 2, p.2016.

Howell, KL., Bridges, A.E.H., Davies, J., Parimbelli, A., Piechaud, N., 2023. An Ecologist's Guide to BIIGLE. University of Plymouth. 46pp. https://doi.org/10.5281/zenodo.7728927

Malde, K., Handegard, N.O., Eikvil, L. and Salberg, A.B., 2020. Machine intelligence and the data-driven future of marine science. *ICES Journal of Marine Science*, 77(4), pp.1274-1285.

Michelin, M., Elliott, M., Bucher, M., Zimring, M. and Sweeney, M., 2018. Catalyzing the growth of electronic monitoring in fisheries. California Environmental Associates and the Nature Conservancy.

Papadakis V, Papadakis I, Lamprianidou F, Glaropoulos A, Kentouri M (2012) A computer-vision system and methodology for the analysis of fish behavior. *Aquacultural Engineering.* 46: 53-59. https://doi.org/10.1016/j.aquaeng.2011.11.002

Pauly, D. and Zeller, D., 2016. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature communications*, 7(1), p.10244.Piechaud N, Hunt C, Culverhouse P, Foster N, Howell K (2019) Automated identification of benthic epifauna with computer vision. *Marine Ecology Progress Series*. 615: 15-30. https://doi.org/10.3354/meps12925

Piechaud, N. and Howell, K.L., 2022. Fast and accurate mapping of fine scale abundance of a VME in the deep sea with computer vision. *Ecological Informatics.* 71. https://doi.org/10.1016/j.ecoinf.2022.101786

Pitcher, T.J., Watson, R., Forrest, R., Valtýsson, H.Þ. and Guénette, S., 2002. Estimating illegal and unreported catches from marine ecosystems: a basis for change. *Fish and Fisheries*, 3(4), pp.317-339.

Poos, J.J., Aarts, G., Vandemaele, S., Willems, W., Bolle, L.J. and Van Helmond, A.T.M., 2013. Estimating spatial and temporal variability of juvenile North Sea plaice from opportunistic data. *Journal of Sea Research*, 75, pp.118-128.

Probst, W.N., 2020. How emerging data technologies can increase trust and transparency in fisheries. *ICES Journal of Marine Science*, 77(4), pp.1286-1294.

Rochet, M.J., Péronnet, I. and Trenkel, V.M., 2002. An analysis of discards from the French trawler fleet in the Celtic Sea. I*CES Journal of marine Science*, 59(3), pp.538-552.

Rudd, M.B., Branch, T.A., 2017. Does unreported catch lead to overfishing? *Fish and Fisheries*, 18 (2), 313–323. https://doi.org/10.1111/faf.12181.

Stanley, R.D., McElderry, H., Mawani, T. and Koolman, J., 2011. The advantages of an audit over a census approach to the review of video imagery in fishery monitoring. *ICES Journal of Marine Science*, 68(8), pp.1621-1627.

Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1), pp.77-89.

Suuronen, P., Gilman, E., 2020. Monitoring and managing fisheries discards: New technologies and approaches. *Marine Policy*, 116, 103554. https://doi.org/10.1016/j. marpol.2019.103554.

Liggens, G. W., M. J. Bradley, S. J. Kennelly. 1997. Detection of bias in observer-based estimates of retained and discarded catches from a multispecies trawl fishery. *Fisheries Research* 32:133-147.

Sampson, D. 2002. Final Report to the Oregon Trawl Commission on Analysis of Data from the At-Sea Data Collection Project. Oregon State University. Newport, Oregon.
http://www.onid.orst.edu/~sampsond/projects/edcp/

Uhlmann, S.S., Van Helmond, A.T., Kemp Stefánsdóttir, E., Sigurðardóttir, S., Haralabous, J., Bellido, J.M., Carbonell, A., Catchpole, T., Damalas, D., Fauconnet, L. and Feekings, J., 2014. Discarded fish in European waters: general patterns and contrasts. *ICES Journal of Marine Science*, 71(5), pp.1235-1245.

Ulleweit, J., Stransky, C. and Panten, K., 2010. Discards and discarding practices in German fisheries in the North Sea and Northeast Atlantic during 2002–2008. *Journal of applied ichthyology*, 26, pp.54-66.

van Helmond, A.T., Mortensen, L.O., Plet-Hansen, K.S., Ulrich, C., Needle, C.L., Oesterwind, D., Kindt-Larsen, L., Catchpole, T., Mangi, S., Zimmermann, C. and Olesen, H.J., 2020. Electronic monitoring in fisheries: lessons from global experiences and future opportunities. *Fish and Fisheries*, 21(1), pp.162-189.

White DJ, Svellingen C, Strachan NJ (2006) Automated measurement of species and length of fish by computer vision. Fisheries Research 80: 203-210. https://doi.org/ 10.1016/j.fishres.2006.04.009

Qiao, M., Wang, D., Tuck, G.N., Little, L.R., Punt, A.E. and Gerner, M., 2021. Deep learning methods applied to electronic monitoring data: automated catch event detection for longline fishing. *ICES Journal of Marine Science*, 78(1), pp.25-35.