© Volume 2, Issue 10, October 2026 | JATIR
e-ISSN: 2839-5786, p-ISSN: 3139-8464

A Comparative Evaluation of Machine Learning- AI Based Fish Catch Prediction for Ratnagiri Coastal Fisheries

[1]Vrunali Surve, [2]Swara Redij, [3]Deepak Patil, [4]Prof. Atiya Kazi
[1,2,3,4]*Department of Information Technology, Finolex Academy of Management and Technology, Ratnagiri, India*

***Abstract*—Marine fishing is a main source of income in Ratnagiri district, yet most decisions about when and where to fish still rest on personal experience. This paper reviews AI and machine learning work in fisheries, covering vessel tracking, temporal catch prediction, ML-based fishing prediction, data standardization, environmental data and AI-based monitoring. Most of these studies deal with a single task, one region or one data type, or stop before forecasting catch quantity. We also summarize a Ratnagiri project that links historical catch records with Google Earth Engine environmental variables for zone-wise prediction using XGBoost and district-wise prediction using Random Forest and LSTM. In the zone-wise experiment XGBoost gave an MAE of 14,212.93 t, an RMSE of 23,218.80 t and an R^2 of 0.46, and a walk-forward check of a gradient boosting model over 2018–2021 gave an MAE of 8,761.36 t and an R^2 of 0.51 with large errors in individual zones. The test sets are very small, so the results are indicative at best. Catch values for future years are model-based forecasts, not observed or guaranteed figures.**

***Index Terms*—Fish catch prediction, machine learning, XGBoost, Random Forest, LSTM, Google Earth Engine, Ratnagiri fisheries.**

I. INTRODUCTION

Ratnagiri is one of the leading marine fishing districts of Maharashtra. Landing centers such as Mirkarwada, Nate/Jaitapur, Burondi and Dabhol keep a large fishing community employed and feed the local economy. Even so, the everyday calls a fisherman has to make, such as when to go out and where to cast the net, mostly come down to experience. Fish availability shifts with the season and with sea conditions like surface temperature, salinity, rainfall and wind. Not knowing what to expect from a trip makes planning hard, both for the fishermen and for the authorities who manage the fishery.

Long-term landing records already exist. According to the project presentation, the Central Marine Fisheries Research Institute (CMFRI) has gathered fish landing data for Ratnagiri over 1982–2024, but nobody had used that record to build an ML-based catch prediction system for the district. Remote-sensing platforms such as Google Earth Engine (GEE) now supply environmental variables that can be paired with these catch records. Machine learning can pick up the relationship between such inputs and the quantity landed, although how useful the result is depends on how openly the data, the evaluation and the limits of the findings are reported.

AI work in fisheries covers a lot of ground. Much of it deals with vessel tracking, data standardization, image-based monitoring, bycatch risk or aquaculture, and comparatively little tries to forecast catch quantity for one coastal region. In this paper we go through the available studies, compare their methods and weaknesses, point out the gap, and describe how the Ratnagiri project tries to fill it with zone-wise and district-wise prediction. We also say plainly what the current evaluation can support and what it cannot.

II. LITERATURE REVIEW

A. Fisheries Vessel Tracking, Trajectory and Anomaly Analysis

Rodríguez et al. [1] used Automatic Identification System (AIS) records with anomaly detection to flag suspicious fishing behavior. This helps with monitoring, but it predicts neither catch quantity nor fishing locations. Brandoli et al. [2] moved from multiple-aspect trajectories toward prediction by combining AIS data, environmental data and machine learning, but their approach is tied to one region. Adibi et al. [3] fused vessel trajectories, landing reports and environmental data for the Northern Adriatic Sea. They then used Random Forest to predict monthly catch per unit effort, and it beat a naive baseline by about 13%. The edited volume in which this work appears [4] concentrates on trajectory analysis and does not build a catch-prediction model. All of these depend on vessel-tracking data, which we do not have for Ratnagiri, where the records are landing data.

B. Temporal Catch Prediction

Gislason et al. [5] proposed CATCH, a Convolutional-LSTM approach for temporal catch hotspots. It captures temporal patterns well but is limited to one region and has no real-time deployment. It shows that LSTM-type models suit temporal data, which is why we use an LSTM at the district level.

C. Machine Learning for Fishing and Catch Prediction

Brito et al. [6] optimized fishing predictors with XGBoost to raise target-species catches while limiting harm to marine biodiversity. They did not forecast future catch. Gao et al. [16] built a boosted regression tree model from satellite sea surface temperature and sea surface height. It predicted chub mackerel fishing grounds in the Yellow Sea and East China Sea (AUC 0.897 on 2011 test data), but it gives fishing-ground presence, not catch quantity. Rahman et al. [17]

compared linear regression, decision tree and random forest on 18 years of climatic data to predict marine fish landings in Malaysia. Random forest performed best (R^2 of 0.86 on testing). This supports using environmental variables for landing prediction, but the study is limited to Malaysian states and uses no boosting or sequence models. A 2025 article on advances in marine fisheries prediction with machine learning [7] also appears in the project's reference list. The supplied material does not give its details, so we cite it only as a sign of continuing research interest.

D. Fisheries Data Standardization

Pennino et al. [8] used a multi-output artificial neural network to standardize heterogeneous fisheries data. Records arrive from different sources in different formats, so this matters. The study is about standardizing data, not forecasting catch. The Ratnagiri project builds preprocessing and prediction into one pipeline.

E. Environmental and Oceanographic Data in Fisheries

Mahfuz and Slawig [9] applied a Conditional Variational Autoencoder to predict steady states of a marine ecosystem model, which is unrelated to catch prediction. Gietzmann-Sanders et al. [10] built a deep-learning depth model for Chinook salmon bycatch risk, so their focus is bycatch mitigation. A study on geographic information systems for fisheries resource management [11] underlines how useful spatial data is. None of these forecasts catch for a coastal district like Ratnagiri. In our project, the environmental variables (rainfall, SST, salinity, temperature and wind speed) come from Google Earth Engine.

F. AI-Based Fisheries Monitoring and Aquaculture

Elstner et al. [12] introduced Ocean Ruler, a computer-vision method for estimating catch size from images at small-scale fisheries. It does not predict abundance or future catch. Alimadadi et al. [13] paired Deep Deterministic Policy Gradient reinforcement learning with a decision tree to improve interpretability for fish farmers, which applies to fish farming only. Shete et al. [18] combined IoT sensor data with an ensemble of Random Forest, Gradient Boosting and XGBoost to predict dissolved oxygen in aquaculture (R^2 of about 0.84–0.86). It shows that tree-based boosting works well on environmental data, but the setting is pond aquaculture, not marine catch. A 2025 article [14] shows how broadly AI is now used in fisheries and aquaculture.

G. Research Trends in Fisheries Prediction

The field has moved from vessel tracking and anomaly detection [1], [2], [3] to temporal and tree-based prediction [5], [6], [16], [17], and on to data standardization [8], image-based monitoring [12], ecosystem modeling [9] and IoT-based aquaculture analytics [18]. Recent titles also point to growing interest in large language models for fisheries management [15]. Most studies concentrate on one task, one kind of data or a different setting. Forecasting catch quantity for a particular region from landing records and environmental data remains rare.

III. COMPARATIVE ANALYSIS

Table 1 compares the reviewed studies using the methods and limitations recorded in the project presentation.

TABLE 1 COMPARISON OF REVIEWED STUDIES

Study	Method	Limitation / gap
Rodríguez et al. [1]	AIS data, anomaly detection	Does not predict catch quantity or fishing locations
Brandoli et al. [2]	AIS + environmental data + ML	Limited to one geographic region
Adibi et al. [3]	AIS + landing + environmental data, Random Forest	Built on vessel trajectories; Northern Adriatic Sea only
Tserpes et al. (Eds.) [4]	Semantic trajectory analysis	Does not develop a catch-prediction model
Gislason et al. [5]	Convolutional LSTM	Region-specific; lacks real-time deployment
Brito et al. [6]	XGBoost	Does not provide future catch prediction
Gao et al. [16]	Boosted regression tree	Predicts fishing-ground presence, not catch quantity
Rahman et al. [17]	Linear regression, decision tree, random forest	Malaysia only; no boosting or sequence models
Pennino et al. [8]	Multi-output ANN	Focuses on data standardization, not forecasting
Mahfuz and Slawig [9]	CVAE	Not intended for fish-catch prediction
Gietzmann-Sanders et al. [10]	Deep learning	Focuses only on bycatch mitigation
Elstner et al. [12]	Computer vision	Does not predict abundance or catch quantity
Alimadadi et al. [13]	DDPG + decision tree	Applicable only to fish farming
Shete et al. [18]	RF + Gradient Boosting + XGBoost ensemble	Aquaculture water quality, not marine catch

The table makes it clear that the studies tackle quite different problems: vessel behavior [1]–[4], temporal hotspots [5], predictor optimization [6], fishing-ground and landing prediction [16], [17], data standardization [8], ecosystem modeling [9], bycatch risk [10], image-based monitoring [12] and aquaculture analytics and decision support [13], [18]. Only a handful use ML to predict

something related to catch, and those are limited by region, data type or deployment. As far as the project material summarizes them, none of the reviewed studies forecasts catch quantity for Ratnagiri from historical landing data and environmental variables.

IV. RESEARCH GAP AND PROJECT CONTRIBUTION

The review points to four gaps. To begin with, many studies never forecast catch quantity. Next, those that do use ML for catch-related prediction are tied to one region or have not been deployed. Third, studies that bring in environmental data often fail to connect it with catch records. Finally, none of the reviewed work is aimed at Ratnagiri's coastal fishermen or gives them an easy-to-use decision-support tool. The Ratnagiri project addresses these gaps in four ways:

- Region-specific focus: it uses Ratnagiri district data and four zones (Burondi, Dabhol, Mirkarwada and Nate Jaitapur).
- Data integration: it merges historical catch records with GEE-based environmental variables.
- Two prediction levels: zone-wise forecasting with XGBoost, and district-wise forecasting with Random Forest and LSTM.
- Decision support: a dashboard in which the user selects a zone and obtains a forecast for the upcoming financial year.

V. SYSTEM ARCHITECTURE OVERVIEW

Fig. 1 shows the overall architecture of the proposed machine learning AI based fish catch prediction system for Ratnagiri coastal fishermen. The design runs two data workflows side by side and brings them together before any model is trained. The architecture is described layer by layer below, and Table 2 summarizes the role of each layer.

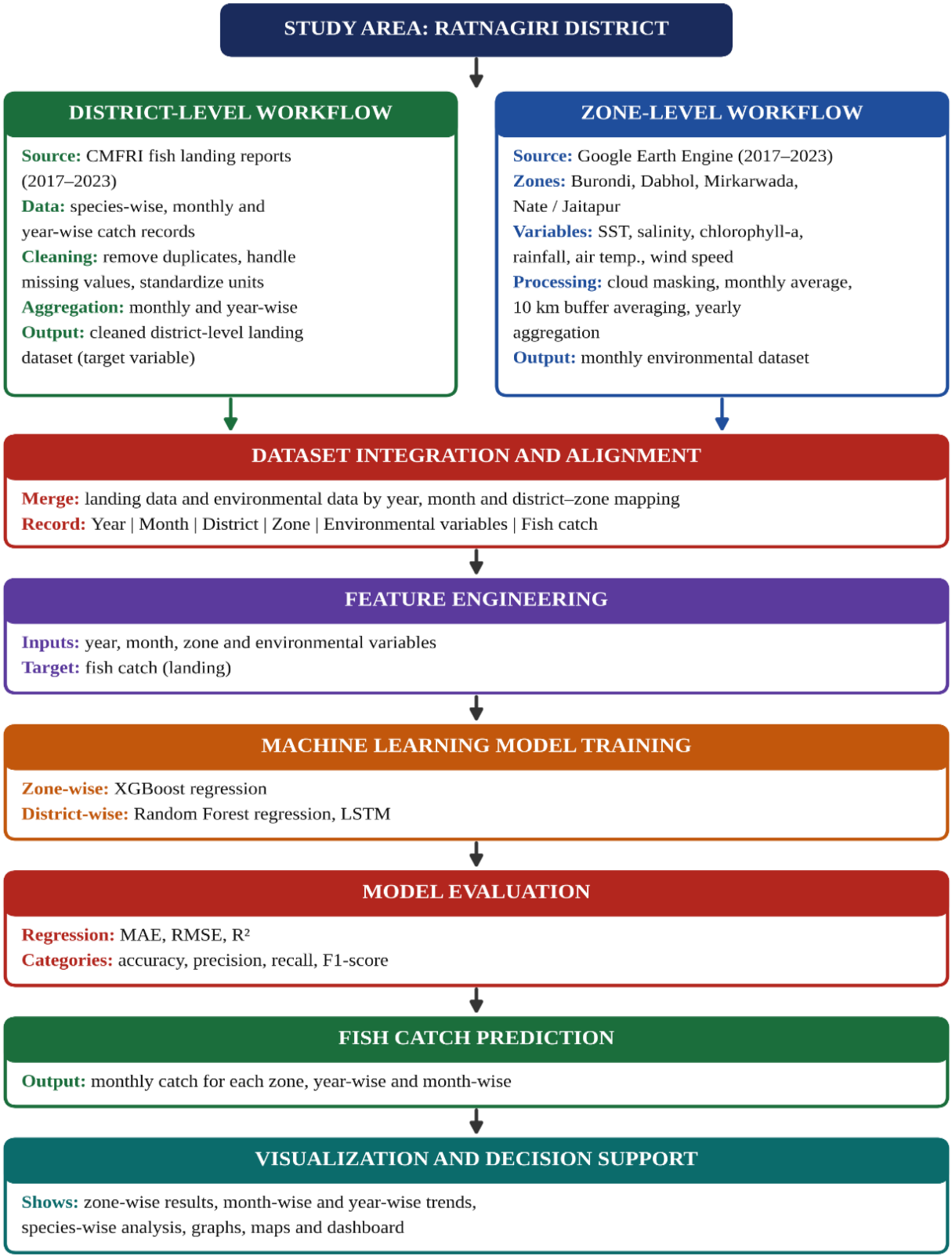

Fig. 1. System architecture of the machine learning AI based fish catch prediction system for Ratnagiri coastal fishermen using historical landing data.

A. District-Level Workflow (Fish Landing Data)

The left branch handles the catch records, which form the target variable. Landing data come from the CMFRI fish landing reports for Ratnagiri district and are described by species-wise catch, monthly landings and year-wise records. The raw records are cleaned by removing duplicates, handling missing values, standardizing units and aggregating them monthly and yearly. The output of this branch is a cleaned district-level fish landing dataset, carrying year and month, that serves as the target variable for the models.

B. Zone-Level Workflow (Environmental Data)

The right branch builds the environmental inputs. Four coastal zones are defined: Burondi, Dabhol, Mirkarwada and Nate / Jaitapur. For each zone and month, GEE supplies sea surface temperature (SST), salinity, chlorophyll-a, rainfall, air temperature and wind speed. These raw layers go through cloud masking and quality checks, monthly averaging, spatial averaging within a 10 km buffer and year-wise aggregation. The result is a monthly environmental dataset for every zone, with year and month attached.

C. Dataset Integration and Alignment

The two branches meet in the integration layer. District-level landing records and zone-level environmental data are merged using year, month and the district–zone mapping. Each merged record holds the year, month, district, zone, the environmental variables and the fish catch. Aligning the two sources in time and place is what allows the models to learn how sea conditions relate to the quantity landed.

D. Feature Engineering, Model Training and Evaluation

The input features are the year, the month, the zone and the environmental variables, and the target is the fish catch (landing). XGBoost is the zone-wise model, and Random Forest and LSTM are the district-wise models. The architecture lists accuracy, precision, recall and F1-score as evaluation measures; as explained in Section VI-G, these apply once catch is grouped into categories, while the regression experiments in this paper are judged mainly with MAE, RMSE and R^2.

E. Prediction, Visualization and Decision Support

The trained models predict monthly fish catch for each zone in Ratnagiri district, year-wise and month-wise. The final layer turns these predictions into something a fisherman or an officer can use: zone-wise prediction results, month-wise and year-wise trends, species-wise analysis drawn from the district data, and graphs, maps and a dashboard.

TABLE 2 ARCHITECTURE LAYERS AND THEIR ROLES

Layer	Main components	Role
District-level workflow	CMFRI landing reports; species-wise, monthly and yearly catch	Provides the cleaned catch data used as the target variable
Zone-level workflow	GEE: SST, salinity, chlorophyll-a, rainfall, air temperature, wind speed; four zones	Provides monthly environmental inputs for each zone
Dataset integration	Merge by year, month and district–zone mapping	Aligns catch and environment into one dataset
Feature engineering	Year, month, zone, environmental variables	Prepares model inputs; catch is the target
Model training	XGBoost (zone-wise); Random Forest, LSTM (district-wise)	Learns the catch–environment relationship
Model evaluation	MAE, RMSE, R^2; accuracy, precision, recall, F1 for categories	Measures how well each model predicts
Prediction	Monthly catch per zone	Produces year-wise and month-wise forecasts
Visualization and decision support	Zone-wise results, trends, species analysis, maps, dashboard	Presents results to fishermen and authorities

The diagram shows the full design, which works at monthly resolution and up to 2023. The experiments completed so far use annual catch records for 2017–2022 (Section VII), because catch is available as an annual figure per zone. Moving to monthly catch data is one of the future directions discussed in Section IX.

VI. DATA AND METHODOLOGY

A. Data Sources

The project draws on two groups of data (Table 3). The district-level data come from CMFRI [19] and the Maharashtra Fisheries Department [20] and cover species-wise landings and annual production for Ratnagiri district. Zone-level catch figures are taken from the Maharashtra Fish Production Report [21], while zone-level environmental data come from Google Earth Engine [22].

TABLE 3 DATASETS USED IN THE PROJECT

Item	District-level	Zone-level
Source	CMFRI; Maharashtra Fisheries Dept.	Maharashtra Fish Production Report; Google Earth Engine
Content	Species-wise landings; annual production	Zone-wise catch; monthly environmental data
Period	Historical records, 2017–2022	Monthly environmental data, 2017–2022
Area	Ratnagiri district	Burondi, Dabhol, Mirkarwada, Nate Jaitapur
Use	Species analysis; district-wise prediction	Zone-wise catch prediction

B. Dataset Integration

The raw dataset in the zone-wise notebook holds 288 monthly records (four zones over 2017–2022). Catch values exist for 2017–2021 only. The 48 missing values are the 2022 months, where no observed catch is available, and we left them unfilled. Catch is recorded as an annual figure per zone and simply repeated across the months of that year. We aggregated the monthly environmental data to annual values and merged them with annual catch, which gave 20 zone-year records (four zones, five financial years).

C. Data Preprocessing

Preprocessing involved standardizing column names, converting values to numeric form, checking for duplicates (there were none) and dealing with missing values. Gaps in the environmental data were filled by linear interpolation within each zone, falling back on the zone median where needed. Catch values were never imputed. Financial years were sorted by their starting year so that the time order stays intact.

D. Environmental Parameters

The environmental variables are rainfall, sea surface temperature (SST), salinity, temperature and wind speed, and chlorophyll-a was also considered in the project. Annual rainfall is the sum of the monthly values, and the other variables are annual means. All of these are historical measurements. No measurements exist for a future year, so the environmental inputs for that year are estimated from past information (Section VIII) and should be read as estimates, not observations.

E. Feature Engineering

The zone-wise annual model takes the zone (one-hot encoded) and the annual environmental variables as inputs, with annual catch in tonnes as the target.

F. Machine Learning Models

Zone-wise prediction uses XGBoost (300 estimators, learning rate 0.03, depth 2). The walk-forward validation reported in Section VII-B uses a scikit-learn Gradient Boosting Regressor (100 trees, learning rate 0.05, depth 2, Huber loss). District-wise prediction uses Random Forest and an LSTM network. The LSTM has a single 32-unit layer, dropout of 0.2 and a small dense head, and was trained with the Adam optimizer and mean-squared-error loss on short sequences of annual data. Fig. 2 shows the flow of the zone-wise XGBoost experiment from data collection to the forecast and dashboard.

Fig. 2. Flow diagram of the zone-wise XGBoost fish catch prediction.

G. Model Evaluation

We used a time-based split so that models are always tested on a later year than the ones they were trained on. Financial years 2017–2020 (16 records) formed the training set, and the latest year,

2021–22 (four zone records), was kept for testing. Since fish catch in tonnes is a continuous quantity, MAE, RMSE and R^2 are the main metrics. Accuracy, precision, recall and F1-score only make sense once catch is turned into predefined categories, so they are reported separately from the regression results. As a second check, an expanding-window (walk-forward) validation over 2018–2021 is reported in Section VII-B.

VII. PROJECT RESULTS

A. Zone-Wise XGBoost Result

Table 4 gives the XGBoost result on the latest-year test set. The model reached an MAE of 14,212.93 t, an RMSE of 23,218.80 t and an R^2 of 0.51.

TABLE 4 ZONE-WISE XGBOOST TEST RESULT (LATEST YEAR)

Model	MAE (t)	RMSE (t)	R^2
XGBoost	14,212.93	23,218.80	0.51

Important limitation. The test set holds only four zone observations, so an R^2 calculated from four points is unstable. None of this establishes statistical significance, general validity or production-level reliability.

B. Walk-Forward Validation of the Zone-Wise Model

Table 4 rests on a single held-out year. To see how the zone-wise model behaves over several years, we also ran an expanding-window (walk-forward) validation. For every test year from 2018 to 2021, the model was trained only on the years before it (one year of data for the first fold, four years for the last) and then predicted that year's catch in all four zones, which gives 16 zone-year predictions. This experiment used a scikit-learn Gradient Boosting Regressor (100 trees, learning rate 0.05, depth 2, Huber loss), with the zone one-hot encoded and rainfall, SST, salinity, air temperature and wind speed as inputs. Pooled over the 16 predictions, it gave an MAE of 8,761.36 t, an RMSE of 13,582.90 t and an R^2 of 0.51. These numbers cannot be compared directly with Table 4, because the model and the test design are different. Table 5 splits the error by zone, and Figs. 3 to 7 show the actual and predicted catch behind these numbers.

TABLE 5 WALK-FORWARD ERROR BY ZONE (2018–2021)

Zone	Mean actual (t)	Mean predicted (t)	MAE (t)	Mean error (%)
Burondi	5,612.25	9,104.41	5,153.76	105.75
Dabhol	18,109.75	13,224.24	4,885.51	26.68
Jaitapur-Nate	6,094.00	12,771.73	6,677.73	117.72
Mirkarwada	46,814.75	39,606.08	18,328.43	34.56

Two things stand out. The absolute error is largest in Mirkarwada, simply because it lands the most fish, while the percentage error is largest in Burondi and Jaitapur-Nate, where the average error is

roughly the size of the actual catch itself (106% and 118%). The pooled R^2 should also be read with care: catch differs by an order of magnitude between zones, so a good part of the R^2 comes from the model telling the large zone from the small ones, not from following year-to-year change inside any one zone. The figures below show that year-to-year behavior directly.

C. All Zones Combined

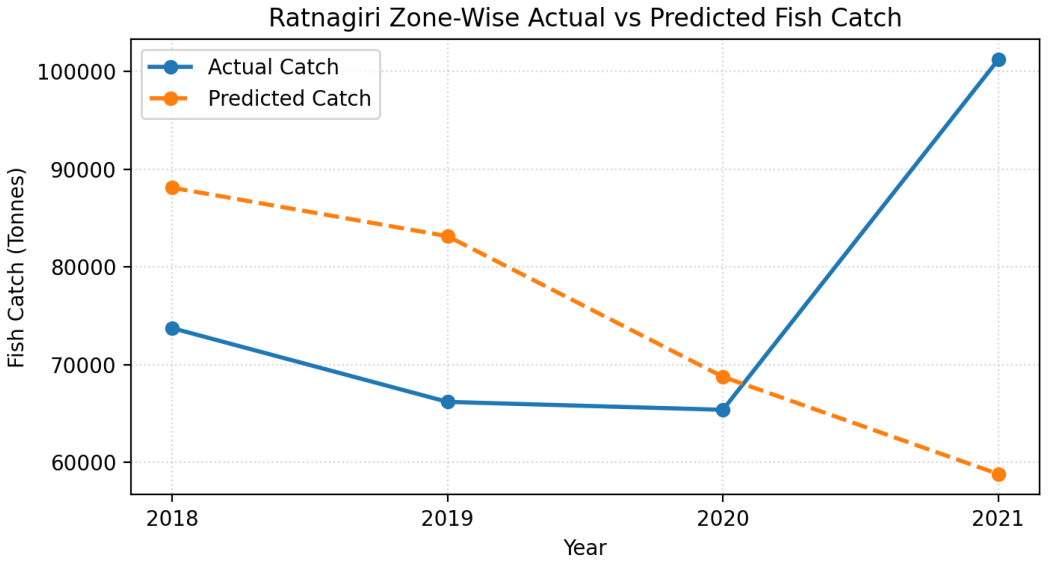

Fig. 3. Actual and predicted fish catch for the four zones combined, 2018–2021 (walk-forward validation).

Fig. 3 adds up the four zones for each year. The predicted total falls steadily from 88,120 t in 2018 to 58,800 t in 2021. The actual total goes the other way at the end: it drops from 73,738 t to 65,374 t between 2018 and 2020, then jumps to 101,228 t in 2021. The model therefore over-predicts in 2018 (by about 14,400 t) and 2019 (about 17,000 t), comes close in 2020 (68,754 t against 65,374 t) and under-predicts by about 42,400 t in 2021.

This pattern occurs for two reasons. First, the 2021 jump belongs almost entirely to one zone. Mirkarwada rose from 30,829 t to 79,073 t (about 156%), while the other three zones together fell from 34,545 t to 22,155 t (about 36%). The model has no way of knowing this, since the environmental readings of 2021 show nothing specific to Mirkarwada, and a tree-based model cannot predict a value far above any it saw in training (the highest Mirkarwada catch before 2021 was 47,010 t). Second, the near match in 2020 is a coincidence of cancelling errors, not a sign of accuracy: that year Mirkarwada was over-predicted by 9,439 t and Jaitapur-Nate by 4,198 t, while Dabhol was under-predicted by 6,934 t and Burondi by 3,323 t, and these happened to nearly offset each other. The combined graph is therefore the most flattering view of the model, and the zone-wise graphs that follow are the more honest one.

D. Burondi

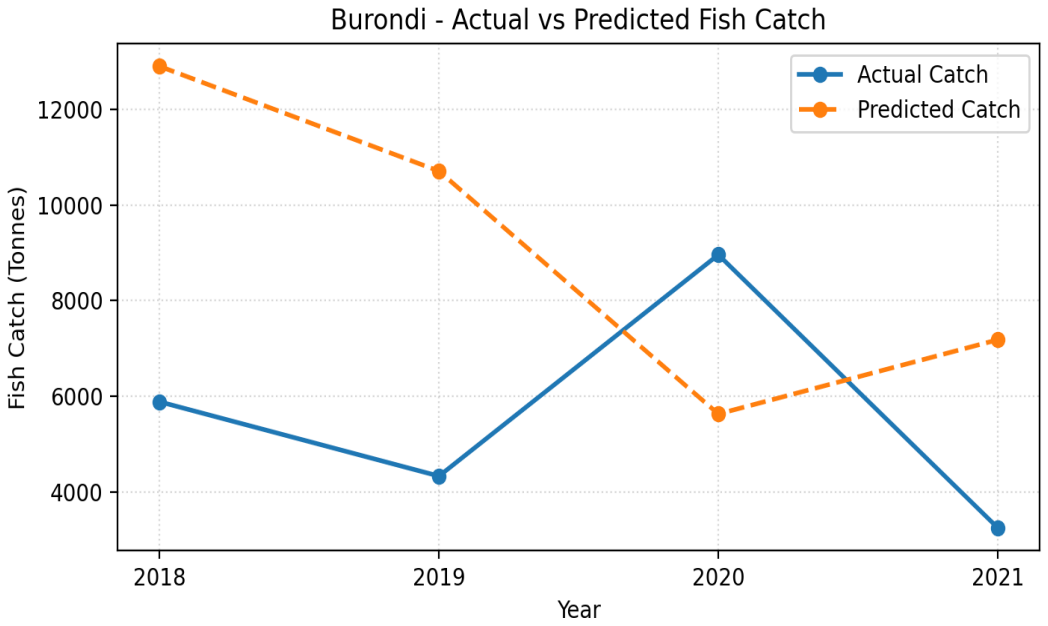

Fig. 4. Burondi: actual and predicted fish catch, 2018–2021.

In Burondi (Fig. 4), the actual catch moves in a zigzag between 3,261 t and 8,959 t (5,890, 4,339, 8,959 and 3,261 t in 2018 to 2021). The predicted catch starts far above it (12,892 t in 2018 and 10,705 t in 2019), crosses it between 2019 and 2020, and ends slightly above it again in 2021 (7,185 t). In 2020 and 2021 the predicted line moves opposite to the actual one: the model predicts a fall in 2020, when catch rose to its highest value of the series, and a rise in 2021, when catch fell to its lowest.

The large early errors occur because the first folds see very little data. In 2018 the model was trained on 2017 alone (four records), and it returned exactly the same value for Burondi and Dabhol (12,892.30 t), which suggests it could not separate the two zones. This is plausible, since their environmental readings are almost identical (in 2017, SST of 28.38 °C and 28.39 °C, and annual rainfall of 2,410 mm and 2,556 mm). As more years enter the training set, the Burondi prediction moves toward the zone's true level and the error shrinks from 7,002 t in 2018 to 3,323 t in 2020. The jump to 8,959 t in 2020 is not anticipated, because no environmental variable in the data changes enough to signal it. Since Burondi's catches are small, even these moderate absolute errors amount to a mean percentage error of about 106%.

E. Dabhol

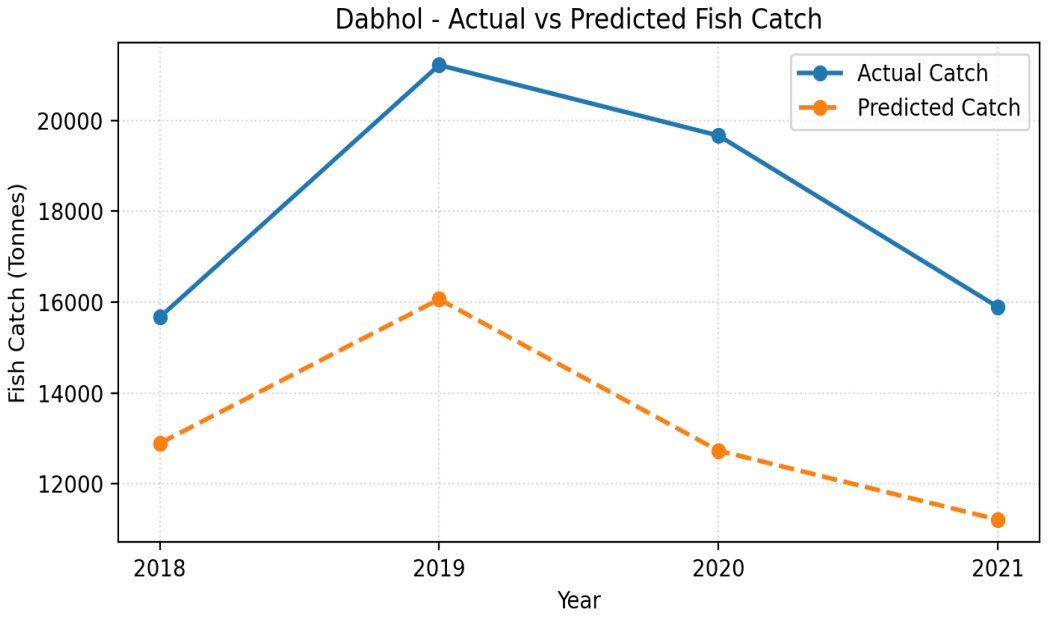

Fig. 5. Dabhol: actual and predicted fish catch, 2018–2021.

Dabhol (Fig. 5) is the best-behaved zone. The predicted line has the same shape as the actual one: both rise from 2018 to 2019 (15,666 t to 21,215 t actual; 12,892 t to 16,062 t predicted) and both fall in 2020 and 2021. Its mean percentage error, 26.7%, is the lowest of the four zones. The weakness is the level. The model under-predicts in every year, by 2,774 t in 2018 and by 4,681 to 6,934 t in the later years, so the predicted line sits below the actual one throughout.

The shape is reproduced because the model carries a rough idea of how big Dabhol is relative to the other zones, and the environmental inputs add a little year-to-year movement. The level is too low most likely because Dabhol's environmental readings closely resemble those of the smaller zones, mainly Burondi. Each fold therefore blends in some of the smaller catches of those zones, and the 2020 and 2021 predictions (12,730 t and 11,213 t) end up below the smallest Dabhol catch ever recorded (15,596 t). The model was also always trained on the past, and Dabhol's catch stepped up in 2019, so its predictions lag behind that higher level.

F. Jaitapur-Nate

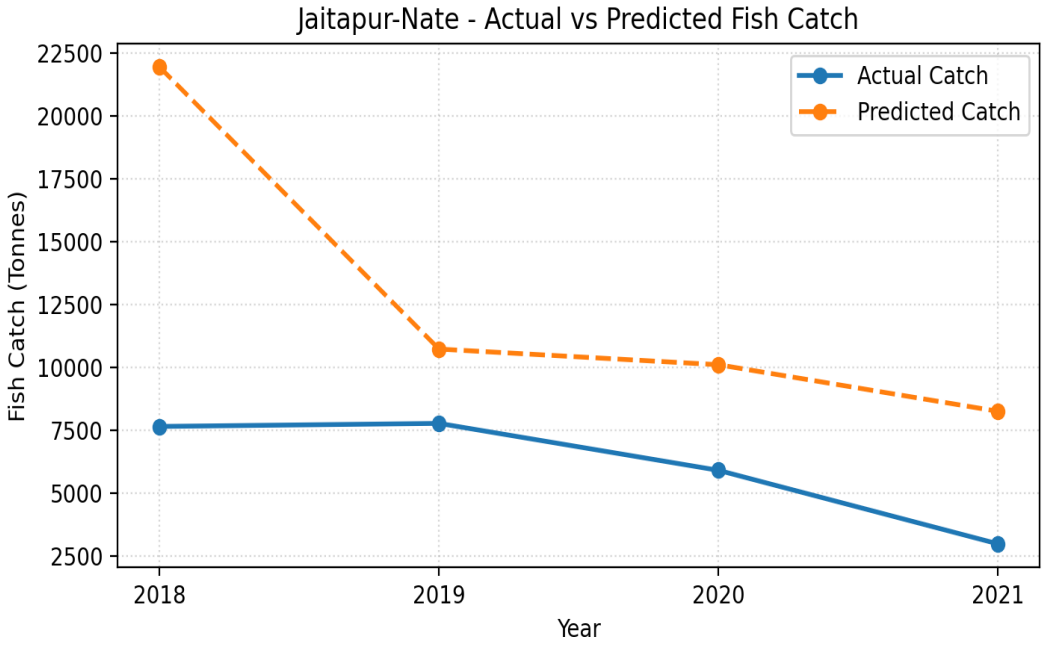

Fig. 6. Jaitapur-Nate: actual and predicted fish catch, 2018–2021.

In Jaitapur-Nate (Fig. 6) the actual catch declines almost steadily, from 7,666 t in 2018 to 3,000 t in 2021, with a flat stretch in 2018–2019 (7,666 t and 7,788 t). The predicted catch also declines, from 21,954 t to 8,269 t, so the direction is right, but the prediction is above the actual catch in every year. The biggest single error in the whole experiment, 14,288 t, is the 2018 prediction for this zone, and the gap narrows afterwards (2,955 t in 2019, 4,198 t in 2020 and 5,269 t in 2021). The 2018 error occurs because the model was trained on 2017 only. Jaitapur-Nate had the warmest SST of the four zones that year (28.62 °C), closest to Mirkarwada (28.56 °C), the zone with by far the largest catch, so the model most likely assigned it a Mirkarwada-like catch. With more years of data the model learns that this zone is small, and the error falls. The remaining over-prediction in the later years arises because the actual catch keeps falling to values below anything seen before (3,000 t in 2021, against an earlier minimum of 5,922 t), and a tree-based model does not predict below the range of its training targets very well. The steady decline itself seems to come from something the environmental variables do not capture, since rainfall and SST move up and down over these years without following it.

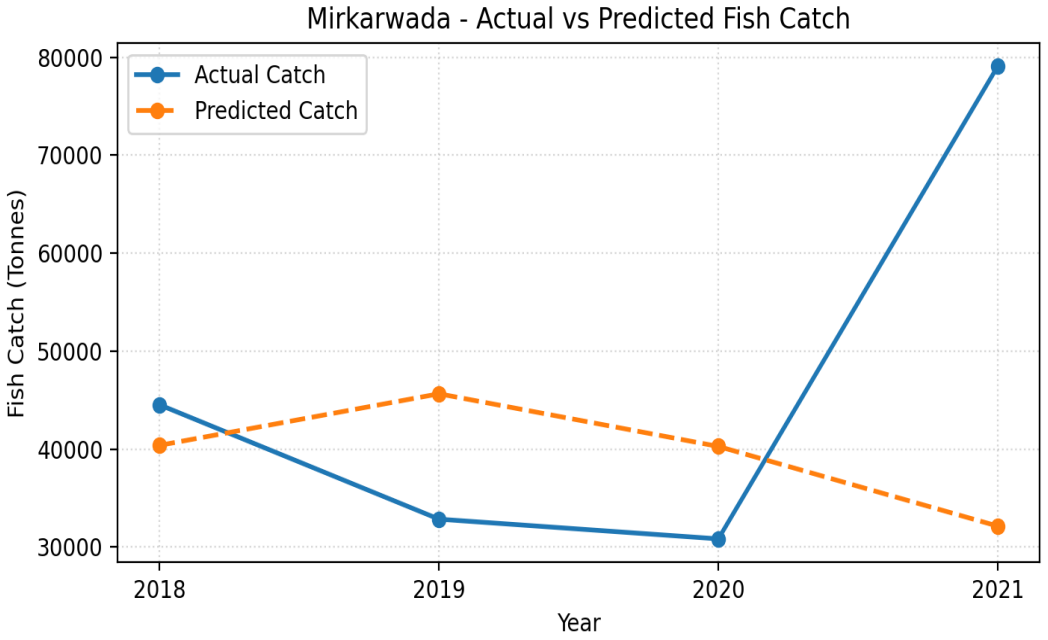

Fig. 7. Mirkarwada: actual and predicted fish catch, 2018–2021.

Mirkarwada (Fig. 7) is the largest zone and carries most of the total error. The model is close in 2018 (40,382 t predicted against 44,516 t actual), then over-predicts in 2019 (45,641 t against 32,841 t) and 2020 (40,268 t against 30,829 t), when the actual catch was falling. In 2021 the actual catch leaps to 79,073 t, nearly 1.7 times the highest earlier value (47,010 t), while the prediction falls to 32,133 t. The 2021 error of 46,940 t is the largest absolute error in the experiment and, on its own, accounts for most of the zone's MAE of 18,328 t.

The 2019–2020 over-prediction occurs because the model has mainly learned that Mirkarwada is a high-catch zone, and since its catches in 2017 and 2018 were high (47,010 t and 44,516 t), it keeps predicting values of that size while the actual catch declines. The 2021 miss is of a different kind. The value lies far outside everything the model had seen, and the environmental readings for that year were similar in all four zones (for example, SST rose in every zone), so nothing in the inputs pointed to a Mirkarwada-specific surge. Factors that are not in the dataset, such as fishing effort, vessel numbers, species mix or changes in how landings were reported, may have played a role, but this cannot be established from the present data.

H. Predicted Catch for 2022

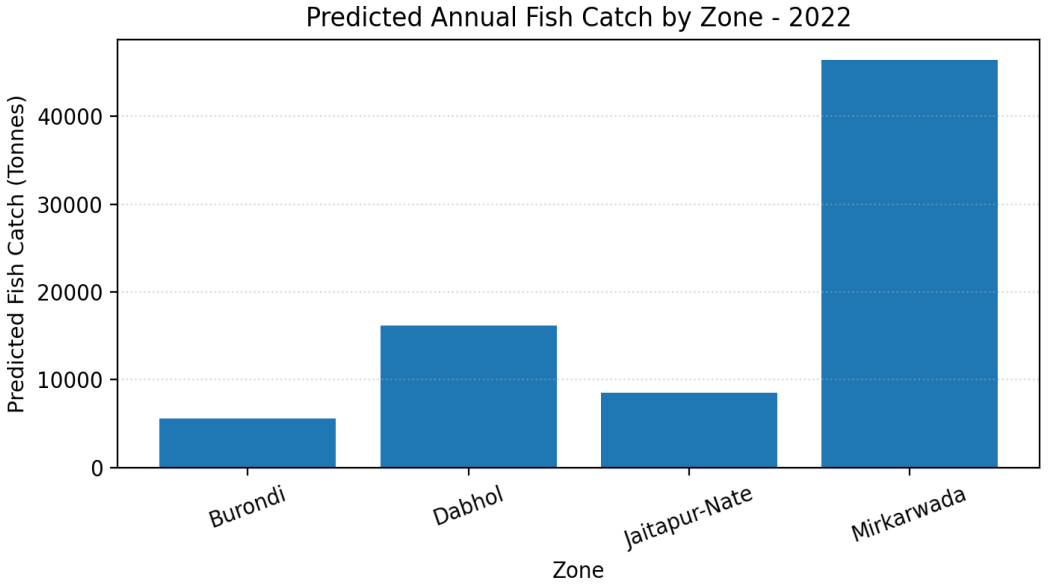

Fig. 8. Predicted annual fish catch by zone for 2022 (final model trained on 2017–2021).

After validation, the model was retrained on all 20 zone-year records (2017–2021) and used to predict 2022 (the financial year 2022–23). Fig. 8 and Table 6 show the result: about 5,632 t for Burondi, 16,178 t for Dabhol, 8,516 t for Jaitapur-Nate and 46,439 t for Mirkarwada, or about 76,765 t in total. The 2022 environmental inputs are the annual values derived from Google Earth Engine for that year (SST between 28.33 °C and 28.66 °C, rainfall between 1,943 mm and 3,028 mm). No observed 2022 catch exists yet, so these predictions cannot be checked, and for years beyond the environmental record the inputs would have to be estimated as well.

TABLE 6 PREDICTED 2022 CATCH COMPARED WITH EARLIER YEARS

Zone	2021 actual (t)	2022 predicted (t)	Change vs 2021 (%)	2017–2021 mean (t)
Burondi	3,261	5,631.55	+72.69	5,604.6
Dabhol	15,894	16,177.99	+1.79	17,607.0
Jaitapur-Nate	3,000	8,516.34	+183.88	7,307.2
Mirkarwada	79,073	46,438.94	−41.27	46,853.8

The percentage changes in Table 6 look dramatic, but they should not be read as forecast growth or decline. The 2022 predictions sit close to each zone's five-year average, and the predicted total (76,765 t) is within about 1% of the average annual total for 2017–2021 (about 77,373 t). The big changes against 2021 arise because 2021 itself was unusual: within the 2017–2021 records it was the lowest year for Burondi and Jaitapur-Nate and the highest for Mirkarwada. This behavior follows from how the final model works. About 93% of its feature importance sits on the zone

indicators (Mirkarwada 0.81, Dabhol 0.11) and only about 7% on the five environmental variables, with rainfall the largest of them (about 0.05). In practice the model predicts each zone's typical level and adjusts it only slightly for the environment, which also explains why it follows the zones' scale well and their yearly swings poorly.

I. Why the Predictions Deviate

Taken together, Figs. 3 to 7 point to four causes that are consistent with the data, although with 16 test points none of them can be proven. (1) Very little training data: the first fold learns from four records, and the early errors are the largest. (2) Weak environmental signal: zone identity carries most of the model's importance, and neighboring zones have almost identical SST, salinity and rainfall, so the inputs do not help separate them. (3) No extrapolation: tree-based models rarely predict beyond the range of their training targets, so the 2021 peak in Mirkarwada and the 2021 lows in Burondi and Jaitapur-Nate were missed. (4) Missing drivers and coarse data: catch is recorded once a year and repeated across months, the environmental inputs are annual averages, and fishing effort, vessel numbers and reporting changes are not in the dataset. Longer records, monthly catch, an effort variable and expanding-window validation with more years are the most direct ways to address these points (Section IX).

J. District-Wise Status

The district-wise models (Random Forest and LSTM) were set up on annual data for 2017–2022, training on 2017–2021 and holding out 2022. With only one held-out year, R^2 is undefined for the Random Forest, and the LSTM was scored on the same sequences it was trained on. For these reasons we do not report district-level performance figures here.

VIII. SYSTEM FRAMEWORK AND DECISION SUPPORT

The framework has two parts. In the zone-wise part, the user only picks a zone and the system forecasts catch for the coming financial year (2022–23 in the notebook). In the district-wise part, historical district catch and environmental variables are used to forecast for Ratnagiri as a whole. A dashboard displays the predictions along with seasonal trends and catch analytics.

It helps to keep four kinds of information apart: historical observed data (catch records), environmental measurements (GEE variables for past periods), model inputs (processed features), and forecasted environmental conditions used for the next year. The next year’s environment has not been observed yet, so it is estimated from recent historical values. The resulting catch is therefore a model prediction, not an observed or guaranteed quantity, and its uncertainty also carries the error in those estimated inputs.

IX. CHALLENGES AND FUTURE DIRECTIONS

Several challenges remain. The zone-wise dataset is small (20 annual records), the test set has just four observations, and because catch is recorded annually, seasonal behavior is not modeled

directly. The district-level evaluation needs more held-out years. Future environmental conditions have to be estimated, which adds uncertainty of its own. Zone names and data sources also differ and have to be reconciled carefully.

In future work we would like to add more years and species-wise records, use rolling or expanding time-based validation, report uncertainty intervals, compare the district-wise models on enough held-out data, and bring in real-time environmental inputs. Testing the system in the field with fishermen and fisheries authorities, and offering a simple multilingual interface, would make it more useful in practice.

X. CONCLUSION

This review looked at AI and ML studies related to fisheries and found that most deal with vessel tracking, data standardization, monitoring, bycatch or aquaculture, and very few with region-specific catch forecasting. The Ratnagiri project tries to close that gap by combining historical catch records with GEE-based environmental variables at both zone and district levels, within the architecture described in Section V. In the completed zone-wise experiment, XGBoost gave an MAE of 14,212.93 t, an RMSE of 23,218.80 t and an R^2 of 0.46. With a test set of only four observations, though, these results show promise and not established accuracy, and predictions for future years remain model-based forecasts that rest on estimated environmental conditions. The walk-forward results confirm this caution: pooled errors were moderate, but individual zones were over- or under-predicted by large margins, most visibly Mirkarwada in 2021.

ACKNOWLEDGMENT

The authors are grateful to their project guide, the industry mentor and the faculty of the Department of Information Technology, Finolex Academy of Management and Technology, for their guidance and support. We also thank CMFRI, the Maharashtra Fisheries Department and Google Earth Engine, and acknowledge the research papers that helped us understand existing approaches.

REFERENCES

[1] J. P. Rodríguez, X. Irigoien, C. M. Duarte, and V. M. Eguíluz, “Identification of suspicious behavior through anomalies in the tracking data of fishing vessels,” EPJ Data Science, vol. 13, Art. no. 23, 2024.

[2] B. Brandoli et al., “From multiple aspect trajectories to predictive analysis.”

[3] P. Adibi et al., “Predicting fishing effort and catch using semantic trajectories and machine learning,” in Multiple-Aspect Analysis of Semantic Trajectories (MASTER 2019), Lecture Notes in Computer Science, vol. 11889. Cham: Springer, 2020, pp. 91–105, doi: 10.1007/978-3-030-38081-6_7.

[4] K. Tserpes, C. Renso, and S. Matwin, Eds., Multiple-Aspect Analysis of Semantic Trajectories, Lecture Notes in Computer Science, vol. 11889. Cham: Springer, 2020.
[5] H. Gislason et al., "Convolutional-LSTM approach for temporal catch hotspots (CATCH)."
[6] J. Brito et al., "Strategic optimization of fishing predictors enhances target species catches while minimizing marine biodiversity impact."
[7] "Advances in marine fisheries prediction using machine learning techniques," Fisheries Research, 2025.
[8] M. S. Pennino et al., "Using a multi-output neural network model to standardize heterogeneous fisheries data," Journal of Ocean University of China, 2025, doi: 10.1007/s11802-025-6136-0.
[9] S. M. Mahfuz and T. Slawig, "Prediction of steady states in a marine ecosystem model by a machine learning technique," International Journal of Information Technology, 2026.
[10] M. Gietzmann-Sanders et al., "A depth model for informing Chinook salmon bycatch risk in the Gulf of Alaska," Fisheries and Aquatic Sciences, 2026.
[11] "Geographic information systems for fisheries resource management," GeoJournal, vol. 87, 2022.
[12] J. Elstner et al., "Ocean Ruler – An image-based, AI-driven approach to small-scale fisheries monitoring and catch size estimation," Marine Biology, 2025.
[13] A. Alimadadi et al., "Enhancing interpretability and explainability for fish farmers."
[14] "Artificial intelligence and machine learning applications in fisheries and aquaculture," Current Fisheries Reports, 2025.
[15] "Large language models (LLMs) for fisheries management research: Understanding potential and navigating risks," Reviews in Fish Biology and Fisheries, vol. 36, no. 1, 2026.
[16] F. Gao, X. Chen, W. Guan, and G. Li, "A new model to forecast fishing ground of Scomber japonicus in the Yellow Sea and East China Sea," Acta Oceanologica Sinica, vol. 35, no. 4, pp. 74–81, 2016, doi: 10.1007/s13131-015-0767-8.
[17] L. F. Rahman, M. Marufuzzaman, L. Alam, M. A. Bari, U. R. Sumaila, and L. M. Sidek, "Application of machine learning to investigate the impact of climatic variables on marine fish landings," National Academy Science Letters, 2022, doi: 10.1007/s40009-022-01110-0.
[18] R. P. Shete, C. Amith Shekhar, Y. V. Mahajan, A. M. Bongale, D. Dharrao, et al., "IoT-driven ensemble machine learning model for accurate dissolved oxygen prediction in aquaculture," Discover Internet of Things, vol. 5, Art. no. 94, 2025, doi: 10.1007/s43926-025-00201-w.
[19] Central Marine Fisheries Research Institute (CMFRI), fish landing data repository. [Online]. Available: https://eprints.cmfri.org.in/19094/
[20] Department of Fisheries, Government of Maharashtra. [Online]. Available: https://fisheries.maharashtra.gov.in/
[21] Maharashtra Fish Production Report, zone-wise fish catch data.
[22] Google Earth Engine. [Online]. Available: https://earthengine.google.com/