Supplementary Materials

This is supplementary materials for the manuscript "A data-driven predictive modelling of delousing treatment events in Norwegian salmon farming using machine learning".

1 Data Pre-processing

To obtain the salinity data for a given farm site $\hat{S_{ft}}$, we used the following procedure:

1. For each station in the available hydrological stations dataset, get latitude and longitude position
2. Calculate the geodesic distance between the farm site and hydrological station using `GeoPy` package - https://geopy.readthedocs.io/en/stable
3. Return the station `id` with the shortest distance
4. Get the mean salinity for a specific week.
   1. If the data were available, calculate mean salinity for depth less than 50 m.
   2. If no data were found, then estimate the value $\hat{S_{ft}} \sim Normal(\mu_4, \sigma_4)$, such that $\mu_4 = 30$ and $\sigma_4 = 4.22$.

Table S1: Description of dataset features used in the training of machine learning models.

Feature	Type	Description
year	numeric	2020, 2021, 2022, 2023, 2024
week	numeric	Week number $(1, \ldots, 52)$
capacity	numeric	Production capacity of a locality in tonne
purposes	categorical	Purpose of granted licences (e.g., commercial, research, development, exhibition, education)
productionTypes	categorical	Production types include food fish, broodstock fish, green licences (A, B, C)
orgCount	numeric	Number of organization managing the locality
licenses	numeric	Number of licences granted for a locality
lat	numeric	Latitude coordinate of a locality
lon	numeric	Longitude coordinate of a locality
vesselVisits	numeric	Number of vessels visited a locality

Feature	Type	Description
hasPd	bool	Binary number indicating the presence (1) or absence (0) of pancreas disease
hasIla	bool	Binary number indicating the presence (1) or absence (0) of infectious salmon anemia (ISA) disease
avgAdultFemaleLice	numeric	Average count of female sea lice
avgMobileLice	numeric	Average count of mobile sea lice
avgStationaryLice	numeric	Average count of stationary sea lice
avgAdultFemaleLicePrev	numeric	Average count of female sea lice previous week
avgMobileLicePrev	numeric	Average count of mobile sea lice previous week
avgStationaryLicePrev	numeric	Average count of mobile sea lice previous week
seaTemperature	numeric	Sea temperature of a locality for a given week
salinity	numeric	Estimated weekly salinity of a locality
fish_stock	numeric	Estimated weekly fish stock of a locality
biomass	numeric	Estimated weekly biomass of a locality
feed_amount	numeric	Estimated weekly feed amount of a locality
cleaner_fish_stock	numeric	Estimated weekly cleaner fish stock of a locality
nearbySites_count	numeric	Number of nearby farm sites within radius 10 km
cumLiceCount	numeric	Total lice count (avgAdultFemaleLice + avgMobileLice + avgStationaryLice)
avgFemaleLice_last_month	numeric	Average female lice count in the past month
avgFemaleLice_nearby	numeric	Average female lice count of nearby farm sites
maxFemaleLice_nearby	numeric	Maximum female lice count of nearby farm sites
exceedLiceLimit_nearby	numeric	Number of nearby farm sites exceeding lice limit
treatmentCount_nearby	numeric	Number of treatment applied at nearby farm sites

2 Spatial-Temporal Analysis

Additional analysis of treatment events from 2020 to 2024 consists of the following figures: Figure S1, Figure S2, Figure S3, Figure S4, and Figure S5.

3 ML training

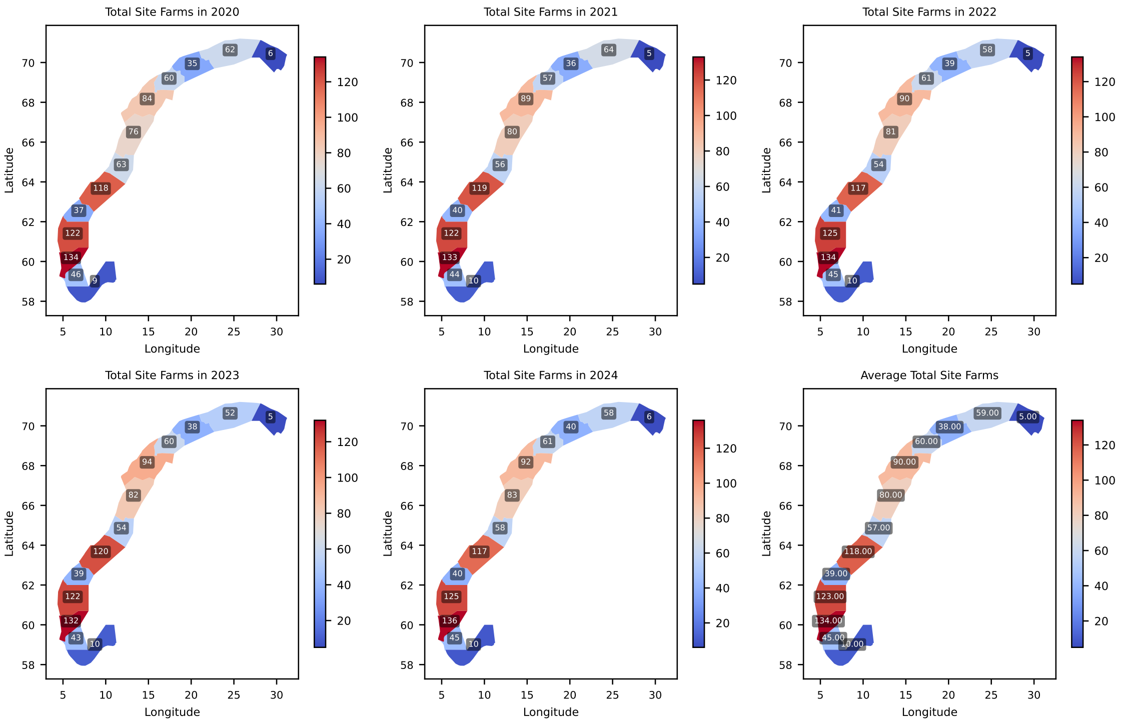

Figure S1: The number of site farms across production areas from 2020 to 2024.

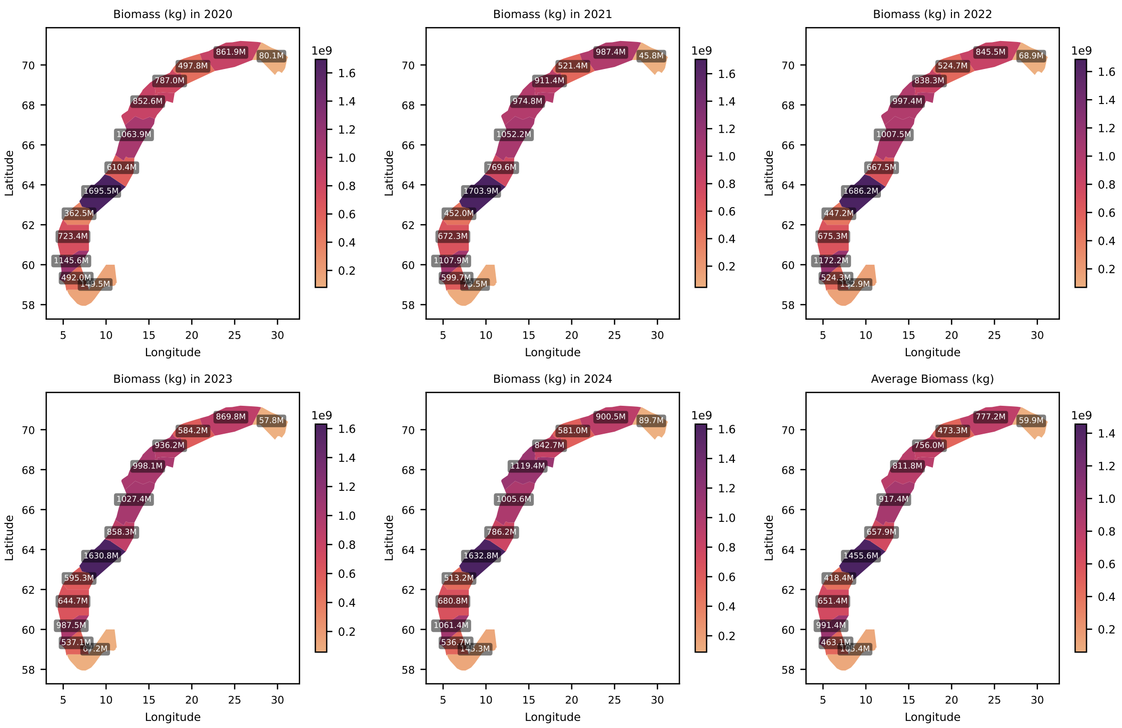

Figure S2: The number of site farms across production areas from 2020 to 2024.

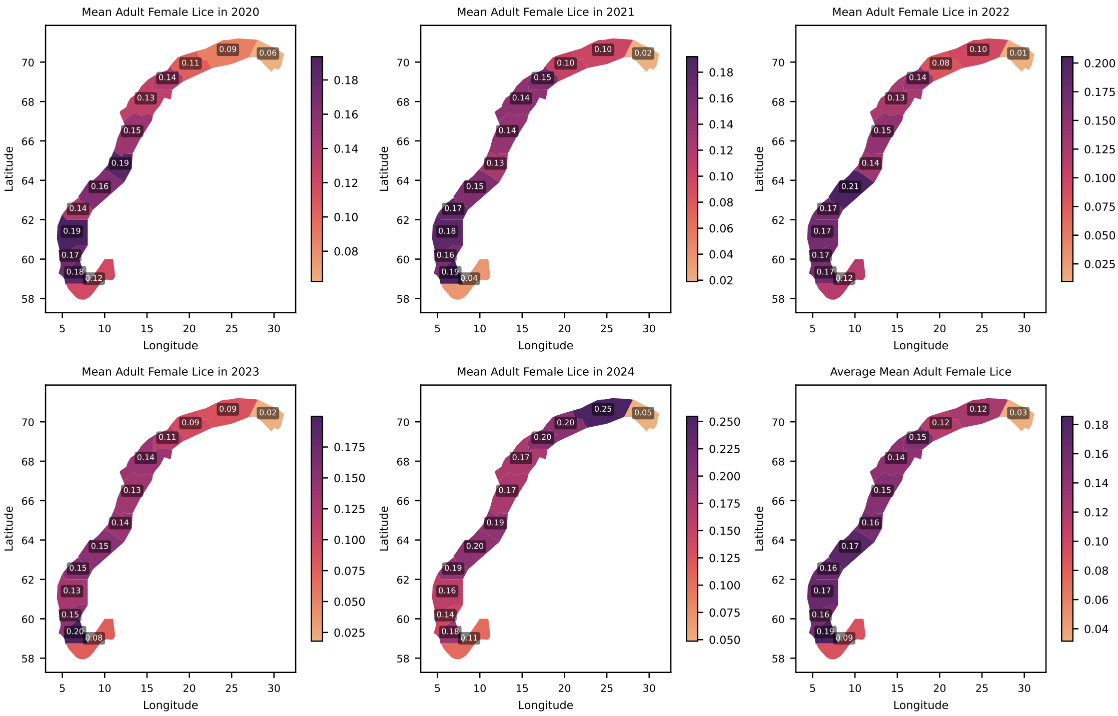

Figure S3: The average count of adult female lice across production areas from 2020 to 2024.

Table S2: A comparative training performance of the models across production areas using the best performing sampling method.

Area	Baseline	Metric	AB	CB	GB	LGB	LR	RF	XGB
1	0.101351	AP	0.24	0.28	0.22	0.24	0.25	0.23	0.25
		PR-AUC	0.48	0.57	0.45	0.48	0.51	0.54	0.49
		ROC-AUC	0.77	0.84	0.74	0.76	0.77	0.82	0.77
		Precision	0.30	0.33	0.29	0.31	0.38	0.26	0.33
		Recall	0.65	0.80	0.58	0.62	0.62	0.80	0.63
		F1	0.41	0.46	0.38	0.41	0.43	0.39	0.43
2	0.093652	AP	0.19	0.17	0.20	0.19	0.26	0.20	0.20
		PR-AUC	0.52	0.51	0.51	0.49	0.56	0.51	0.51
		ROC-AUC	0.75	0.73	0.76	0.74	0.80	0.76	0.76
		Precision	0.21	0.19	0.23	0.22	0.31	0.22	0.24
		Recall	0.81	0.81	0.76	0.74	0.78	0.78	0.77
		F1	0.33	0.31	0.36	0.34	0.45	0.35	0.36
3	0.153669	AP	0.32	0.29	0.30	0.31	0.39	0.29	0.31
		PR-AUC	0.59	0.57	0.58	0.59	0.64	0.58	0.59
		ROC-AUC	0.77	0.74	0.76	0.77	0.80	0.75	0.77
		Precision	0.36	0.32	0.33	0.34	0.47	0.32	0.35
		Recall	0.80	0.78	0.80	0.81	0.77	0.81	0.80
		F1	0.49	0.46	0.47	0.48	0.58	0.46	0.48
4	0.165419	AP	0.30	0.36	0.30	0.36	0.37	0.29	0.37
		PR-AUC	0.59	0.60	0.58	0.63	0.61	0.58	0.63

Continued on next page

Table S2: A comparative training performance of the models across production areas using the best performing sampling method.

Area	Baseline	Metric	AB	CB	GB	LGB	LR	RF	XGB
		ROC-AUC	0.74	0.66	0.74	0.63	0.77	0.74	0.64
		Precision	0.33	0.75	0.34	0.85	0.45	0.33	0.85
		Recall	0.81	0.35	0.79	0.28	0.73	0.79	0.30
		F1	0.47	0.46	0.47	0.42	0.55	0.46	0.44
5	0.126212	AP	0.29	0.27	0.28	0.29	0.27	0.28	0.30
		PR-AUC	0.56	0.57	0.54	0.57	0.54	0.55	0.57
		ROC-AUC	0.76	0.76	0.74	0.76	0.74	0.75	0.77
		Precision	0.34	0.31	0.33	0.35	0.32	0.33	0.35
		Recall	0.75	0.80	0.72	0.75	0.74	0.75	0.77
		F1	0.46	0.44	0.45	0.47	0.44	0.46	0.48
6	0.153167	AP	0.34	0.36	0.34	0.35	0.41	0.34	0.36
		PR-AUC	0.63	0.59	0.61	0.62	0.65	0.61	0.63
		ROC-AUC	0.78	0.66	0.78	0.79	0.81	0.77	0.63
		Precision	0.37	0.72	0.38	0.39	0.47	0.37	0.88
		Recall	0.86	0.36	0.81	0.83	0.81	0.83	0.27
		F1	0.51	0.45	0.52	0.53	0.59	0.51	0.41
7	0.164884	AP	0.35	0.32	0.33	0.35	0.37	0.32	0.36
		PR-AUC	0.59	0.56	0.58	0.59	0.62	0.58	0.65
		ROC-AUC	0.76	0.74	0.75	0.75	0.77	0.75	0.63
		Precision	0.44	0.39	0.40	0.44	0.46	0.37	0.92
		Recall	0.68	0.68	0.72	0.68	0.74	0.76	0.26
		F1	0.54	0.50	0.51	0.53	0.55	0.50	0.40
8	0.132850	AP	0.28	0.25	0.26	0.27	0.31	0.26	0.29
		PR-AUC	0.52	0.49	0.53	0.52	0.57	0.53	0.62
		ROC-AUC	0.73	0.71	0.74	0.73	0.76	0.74	0.60
		Precision	0.36	0.32	0.31	0.34	0.38	0.31	0.94
		Recall	0.63	0.62	0.72	0.65	0.72	0.72	0.19
		F1	0.46	0.41	0.44	0.44	0.49	0.44	0.32
9	0.097730	AP	0.19	0.17	0.18	0.20	0.20	0.18	0.20
		PR-AUC	0.47	0.49	0.49	0.54	0.48	0.49	0.54
		ROC-AUC	0.72	0.70	0.73	0.56	0.73	0.72	0.56
		Precision	0.23	0.19	0.21	0.87	0.24	0.20	0.87
		Recall	0.69	0.77	0.75	0.13	0.68	0.76	0.13
		F1	0.34	0.30	0.33	0.23	0.36	0.32	0.23
10	0.102072	AP	0.19	0.20	0.20	0.21	0.21	0.19	0.21
		PR-AUC	0.42	0.48	0.49	0.56	0.48	0.47	0.49
		ROC-AUC	0.68	0.73	0.73	0.57	0.72	0.72	0.74
		Precision	0.28	0.24	0.24	0.89	0.27	0.22	0.27
		Recall	0.51	0.69	0.70	0.14	0.65	0.70	0.67
		F1	0.36	0.36	0.36	0.24	0.37	0.34	0.38
11	0.070806	AP	0.12	0.10	0.12	0.10	0.12	0.13	0.11
		PR-AUC	0.38	0.44	0.41	0.38	0.43	0.41	0.45

Continued on next page

Table S2: A comparative training performance of the models across production areas using the best performing sampling method.

Area	Baseline	Metric	AB	CB	GB	LGB	LR	RF	XGB
		ROC-AUC	0.66	0.64	0.67	0.63	0.67	0.69	0.65
		Precision	0.15	0.11	0.14	0.12	0.13	0.17	0.12
		Recall	0.59	0.75	0.65	0.61	0.70	0.61	0.77
		F1	0.23	0.19	0.23	0.20	0.22	0.26	0.20
12	0.061053	AP	0.15	0.16	0.16	0.16	0.13	0.17	0.16
		PR-AUC	0.49	0.50	0.47	0.48	0.45	0.49	0.48
		ROC-AUC	0.75	0.75	0.75	0.74	0.71	0.77	0.75
		Precision	0.18	0.18	0.19	0.18	0.16	0.20	0.19
		Recall	0.80	0.80	0.73	0.76	0.72	0.78	0.76
		F1	0.28	0.29	0.30	0.29	0.25	0.31	0.29
13	0.042301	AP	0.20	0.21	0.17	0.19	0.10	0.18	0.18
		PR-AUC	0.38	0.37	0.23	0.41	0.20	0.24	0.35
		ROC-AUC	0.68	0.65	0.60	0.70	0.60	0.62	0.65
		Precision	0.26	0.26	0.19	0.21	0.08	0.17	0.21
		Recall	0.47	0.46	0.24	0.57	0.29	0.29	0.46
		F1	0.32	0.28	0.21	0.26	0.12	0.21	0.25

Table S3 shows the detailed hyperparameter settings for each ML model. Random sampling was used to generate parameter sampling space. Some of the parameter value were based on Bentéjac, Csörgő, and Martínez-Muñoz (2021), while others were adjusted according to our own modifications. The **`quniform`** refers to a function that generates random value under uniform distribution, with rounding.

$$quniform(a, b, q) \sim round(uniform(a, b)/q) * q$$

Meanwhile, the **`choice`** function will randomly return a value from the given list. The weighting parameter choice were transformed as follows:

- `'w1'` $= 1$,
- `'w2'` $= 2$,
- `'w_sqrt_ratio'` $= \sqrt{\frac{\sum \text{majority class}}{\sum \text{minority class}}}$,
- `'w_ratio'` $= \frac{\sum \text{majority class}}{\sum \text{minority class}}$

The hyperparameter tuning was performed using the hyperdrive job run in the Azureml. Further information can be seen in https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters?view=azureml-api-2.

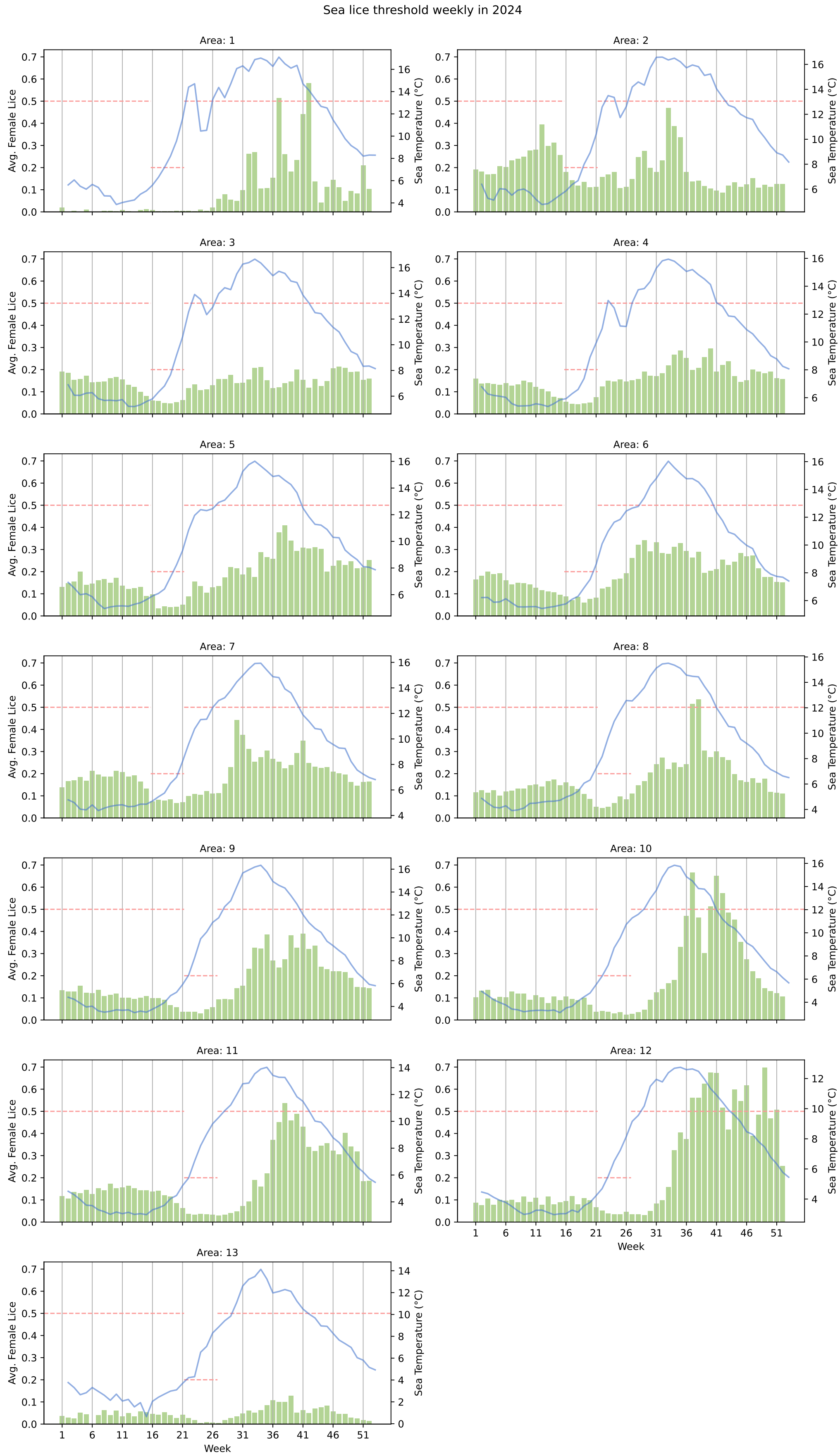

Figure S4: Weekly average count of adult female lice across production areas in 2024, with the threshold indicated by a red line. The blue line indicates the sea temperature.

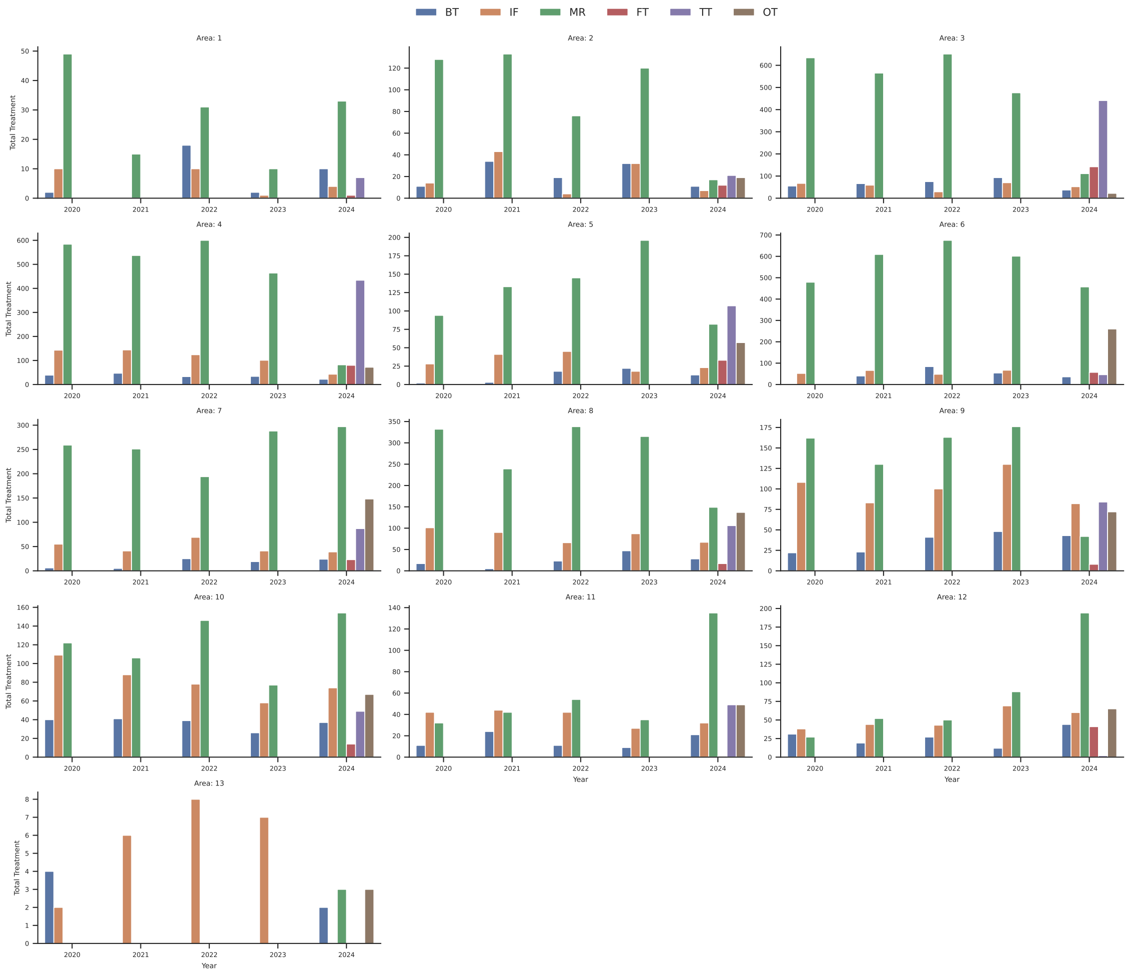

Figure S5: Distribution of treatment types across production areas from 2020 to 2024. BT: Bath Treatment, IF: In Feed treatment, MR: Mechanical Removal, FT: Freshwater Treatment, TT: Thermal Treatment, OT: Other treatment.

Table S3: Hyperparameter settings for each ML model.

Hyperparameter	Initial value	Random sampling value
AB		
n_estimators	100	`quniform(100, 1000, 50)`
learning_rate	0.01	`uniform(0.001, 0.3)`
random_state	42	`quniform(1, 50, 5)`
CB		
iterations	10	`quniform(100, 1000, 100)`
max_depth	3	`quniform(2, 10, 1)`
learning_rate	0.01	`uniform(0.001, 0.3)`
subsample	0.5	`uniform(0.01, 1)`
l2_leaf_reg	3	`quniform(2, 12, 2)`
bagging_temperature	1	`uniform(0, 1)`
scale_pos_weight	1	`choice('w1', 'w2', 'w_sqrt_ratio', 'w_ratio')`
GB		
n_estimators	100	`quniform(100, 1000, 100)`
learning_rate	0.01	`uniform(0.001, 0.3)`
max_depth	3	`quniform(2, 10, 1)`
subsample	1.0	`uniform(0.01, 1)`
max_features	'sqrt'	`choice('log2', 'sqrt')`
min_samples_split	3	`quniform(2, 20, 2)`
min_samples_leaf	1	`quniform(1, 70, 5)`
LGB		
n_estimators	100	`quniform(100, 1000, 100)`
num_leaves	50	`quniform(20, 50, 1)`
learning_rate	0.01	`uniform(0.001, 0.3)`
max_depth	3	`quniform(2, 10, 1)`
subsample	0.5	`uniform(0.01, 1)`
reg_alpha	0.15	`uniform(0.001, 1)`
reg_lambda	0.05	`uniform(0.01, 2)`
scale_pos_weight	1	`choice('w1', 'w2', 'w_sqrt_ratio', 'w_ratio')`
LR		
penalty	'l2'	`choice('l1', 'l2')`
C	1.0	`loguniform(0.01, 100)`
max_iter	100	`quniform(100, 1000, 100)`
class_weight	'balanced'	``choice(`balanced','w1', 'w2', 'w_sqrt_ratio','w_ratio')``
RF		
n_estimators	100	`quniform(100, 1000, 100)`
max_depth	3	`quniform(2, 10, 1)`
criterion	'gini'	`choice('gini', 'entropy', 'log_loss')`
max_features	'sqrt'	`choice('log2', 'sqrt')`
min_samples_split	3	`quniform(2, 20, 2)`
min_samples_leaf	5	`quniform(1, 70, 5)`
class_weight	'balanced'	``choice(`balanced','balanced_subsample','w1', 'w2', 'w_sqrt_ratio', 'w_ratio')``
XGB		
n_estimators	100	`quniform(100, 1000, 100)`
learning_rate	0.01	`uniform(0.001, 0.3)`
max_depth	3	`quniform(2, 10, 1)`
subsample	0.5	`uniform(0.01, 1)`
gamma	0.1	`uniform(0.01, 1)`
reg_alpha	0.15	`uniform(0.001, 1)`
reg_lambda	0.05	`uniform(0.01, 2)`
scale_pos_weight	1	`choice('w1', 'w2', 'w_sqrt_ratio', 'w_ratio')`

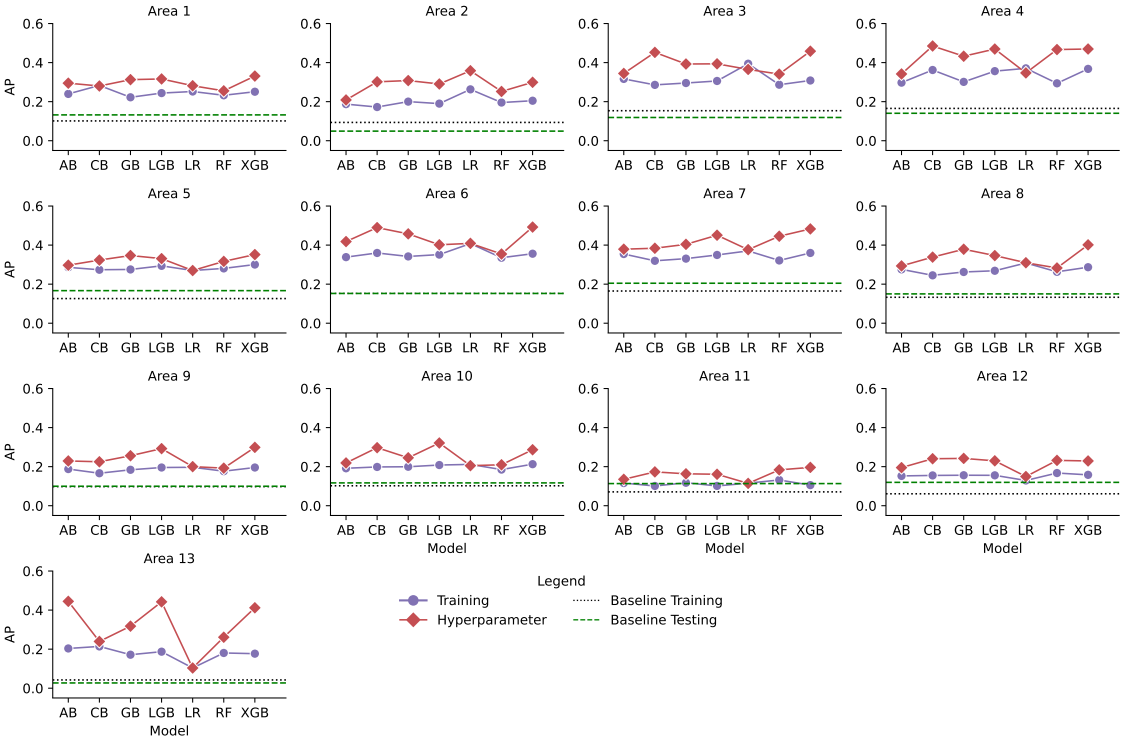

Figure S6: A comparison of model performance after hyperparameter tuning.

4 ML Testing

Table S4: A comparative testing performance of the models across production areas using the best hyperparameter settings.

Area	Baseline	Metric	AB	CB	GB	LGB	LR	RF	XGB
1	0.132231	AP	0.43	0.31	0.47	0.46	0.41	0.22	0.37
		PR-AUC	0.74	0.58	0.75	0.72	0.65	0.62	0.72
		ROC-AUC	0.84	0.78	0.80	0.82	0.79	0.73	0.81
		Precision	0.51	0.37	0.66	0.60	0.55	0.23	0.45
		Recall	0.79	0.77	0.65	0.71	0.67	0.92	0.75
		F1	0.62	0.50	0.65	0.65	0.60	0.37	0.56
2	0.049107	AP	0.14	0.29	0.35	0.28	0.37	0.20	0.32
		PR-AUC	0.41	0.59	0.59	0.60	0.60	0.49	0.63
		ROC-AUC	0.82	0.81	0.80	0.81	0.69	0.83	0.85
		Precision	0.16	0.41	0.52	0.40	0.91	0.23	0.40
		Recall	0.87	0.66	0.64	0.66	0.38	0.79	0.75
		F1	0.27	0.50	0.57	0.50	0.53	0.36	0.52
3	0.118981	AP	0.28	0.50	0.42	0.44	0.44	0.31	0.52
		PR-AUC	0.65	0.73	0.71	0.73	0.72	0.60	0.73
		ROC-AUC	0.80	0.84	0.85	0.86	0.84	0.81	0.84
		Precision	0.31	0.64	0.49	0.51	0.54	0.35	0.67
		Recall	0.87	0.73	0.82	0.83	0.78	0.83	0.73
		F1	0.46	0.68	0.61	0.63	0.64	0.49	0.70
4	0.140523	AP	0.40	0.54	0.55	0.56	0.48	0.54	0.56
		PR-AUC	0.69	0.78	0.77	0.77	0.75	0.76	0.78
		ROC-AUC	0.84	0.77	0.80	0.80	0.84	0.80	0.80
		Precision	0.45	0.86	0.79	0.82	0.58	0.79	0.82
		Recall	0.85	0.55	0.64	0.62	0.78	0.62	0.62
		F1	0.58	0.67	0.71	0.71	0.67	0.70	0.71
5	0.166778	AP	0.28	0.33	0.40	0.35	0.35	0.35	0.39
		PR-AUC	0.57	0.63	0.59	0.60	0.66	0.63	0.59
		ROC-AUC	0.72	0.77	0.77	0.79	0.78	0.79	0.78
		Precision	0.29	0.36	0.51	0.38	0.39	0.39	0.49
		Recall	0.87	0.84	0.66	0.85	0.80	0.85	0.70
		F1	0.44	0.50	0.58	0.53	0.53	0.53	0.58
6	0.152408	AP	0.51	0.58	0.54	0.46	0.46	0.37	0.54
		PR-AUC	0.77	0.80	0.78	0.75	0.75	0.64	0.79
		ROC-AUC	0.85	0.83	0.84	0.86	0.84	0.83	0.80
		Precision	0.60	0.75	0.68	0.51	0.53	0.39	0.76
		Recall	0.80	0.71	0.74	0.87	0.81	0.91	0.64
		F1	0.69	0.73	0.71	0.64	0.64	0.55	0.69
7	0.204707	AP	0.46	0.49	0.49	0.57	0.50	0.55	0.59
		PR-AUC	0.74	0.78	0.76	0.80	0.77	0.79	0.82
		ROC-AUC	0.82	0.83	0.82	0.82	0.81	0.81	0.79
		Precision	0.50	0.55	0.56	0.73	0.60	0.71	0.84
		Recall	0.86	0.83	0.81	0.70	0.74	0.69	0.61
		F1	0.63	0.66	0.66	0.71	0.67	0.70	0.71

Continued on next page

Table S4: A comparative testing performance of the models across production areas using the best hyperparameter settings.

Area	Baseline	Metric	AB	CB	GB	LGB	LR	RF	XGB
8	0.150129	AP	0.35	0.41	0.45	0.41	0.41	0.33	0.47
		PR-AUC	0.67	0.70	0.68	0.70	0.66	0.65	0.74
		ROC-AUC	0.79	0.80	0.74	0.80	0.80	0.78	0.74
		Precision	0.40	0.51	0.73	0.51	0.51	0.38	0.80
		Recall	0.80	0.73	0.51	0.72	0.73	0.81	0.50
		F1	0.53	0.60	0.60	0.60	0.60	0.51	0.62
9	0.099737	AP	0.34	0.33	0.37	0.39	0.30	0.23	0.41
		PR-AUC	0.61	0.62	0.58	0.63	0.57	0.57	0.63
		ROC-AUC	0.75	0.79	0.73	0.72	0.79	0.78	0.72
		Precision	0.54	0.44	0.65	0.73	0.37	0.27	0.76
		Recall	0.56	0.66	0.49	0.46	0.72	0.80	0.47
		F1	0.55	0.53	0.56	0.56	0.49	0.40	0.58
10	0.116799	AP	0.32	0.37	0.36	0.34	0.25	0.31	0.37
		PR-AUC	0.54	0.60	0.60	0.61	0.56	0.56	0.60
		ROC-AUC	0.77	0.68	0.72	0.68	0.75	0.76	0.70
		Precision	0.42	0.79	0.62	0.70	0.30	0.41	0.71
		Recall	0.67	0.38	0.47	0.38	0.72	0.63	0.43
		F1	0.51	0.51	0.54	0.49	0.43	0.50	0.54
11	0.113304	AP	0.21	0.39	0.41	0.44	0.28	0.39	0.42
		PR-AUC	0.49	0.63	0.64	0.66	0.58	0.68	0.61
		ROC-AUC	0.69	0.72	0.76	0.80	0.78	0.78	0.74
		Precision	0.28	0.73	0.63	0.60	0.33	0.57	0.71
		Recall	0.58	0.45	0.57	0.66	0.77	0.62	0.51
		F1	0.38	0.56	0.60	0.63	0.46	0.59	0.59
12	0.119571	AP	0.34	0.36	0.37	0.35	0.32	0.38	0.30
		PR-AUC	0.61	0.63	0.63	0.57	0.53	0.67	0.52
		ROC-AUC	0.83	0.71	0.71	0.71	0.83	0.82	0.70
		Precision	0.38	0.65	0.67	0.61	0.35	0.46	0.51
		Recall	0.85	0.45	0.45	0.46	0.88	0.75	0.47
		F1	0.53	0.53	0.54	0.53	0.50	0.57	0.49
13	0.026906	AP	0.03	0.15	0.15	0.15	0.03	0.06	0.15
		PR-AUC	0.08	0.15	0.27	0.26	0.08	0.30	0.27
		ROC-AUC	0.50	0.66	0.66	0.66	0.50	0.58	0.66
		Precision	0.00	0.40	0.40	0.40	0.00	0.25	0.40
		Recall	0.00	0.33	0.33	0.33	0.00	0.17	0.33
		F1	0.00	0.36	0.36	0.36	0.00	0.20	0.36

Table S5: Average AP lift and standard deviation of the models using the testing dataset.

Model	AP lift
AB	2.52 ± 0.72
CB	3.51 ± 1.17
GB	3.73 ± 1.30
LGB	3.61 ± 1.09
LR	3.03 ± 1.53
RF	2.74 ± 0.69
XGB	3.74 ± 1.19

References

Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2021. "A Comparative Analysis of Gradient Boosting Algorithms." *Artificial Intelligence Review* 54 (3): 1937–67. https://doi.org/10.1007/s10462-020-09896-5.