Non-Invasive Fish Species Classification Using Deep Learning:

A Hydroacoustic Approach for Sustainable Ecological Monitoring

Scarlett(Yi) Yang

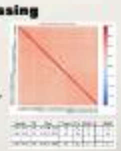
Introduction

Problem: Ecological monitoring of inland fisheries
- Traditional methods (trawling/netting)
- Hydroacoustic solution

Goal
- Build a deep-learning model to classify fish species between Lake Trout and Smallmouth Bass, with a high accuracy (80%) using frequency-related features

Impact
- Enable sustainable, scalable ecological monitoring.

Data & Preprocessing



Methodology

Results

LSTM Performance
- 73.8% accuracy (LOPO validation)

Key Insight
- Sequential modeling improves accuracy over traditional ML.

Limitation:
- Target accuracy (80%) unmet due to small dataset.

Conclusion

Limitations & Future Work

THANK YOU!

Prezi

Introduction

Problem: Ecological monitoring of inland fisheries.

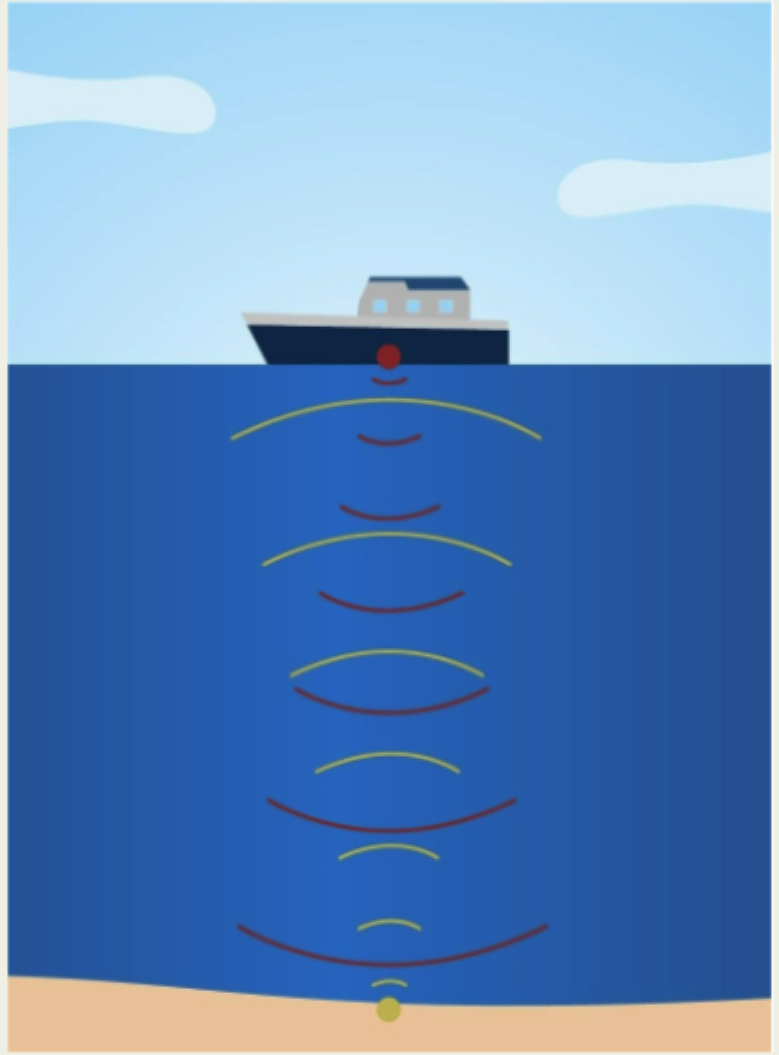
- Traditional methods (trawling/netting)
- Hydroacoustic solution

Goal:

- Build a deep-learning model to classify fish species between Lake Trout and Smallmouth Bass, with a high accuracy (80%) using frequency-related features.

Impact:

- Enable sustainable, scalable ecological monitoring.



Data & Preprocessing

Dataset:

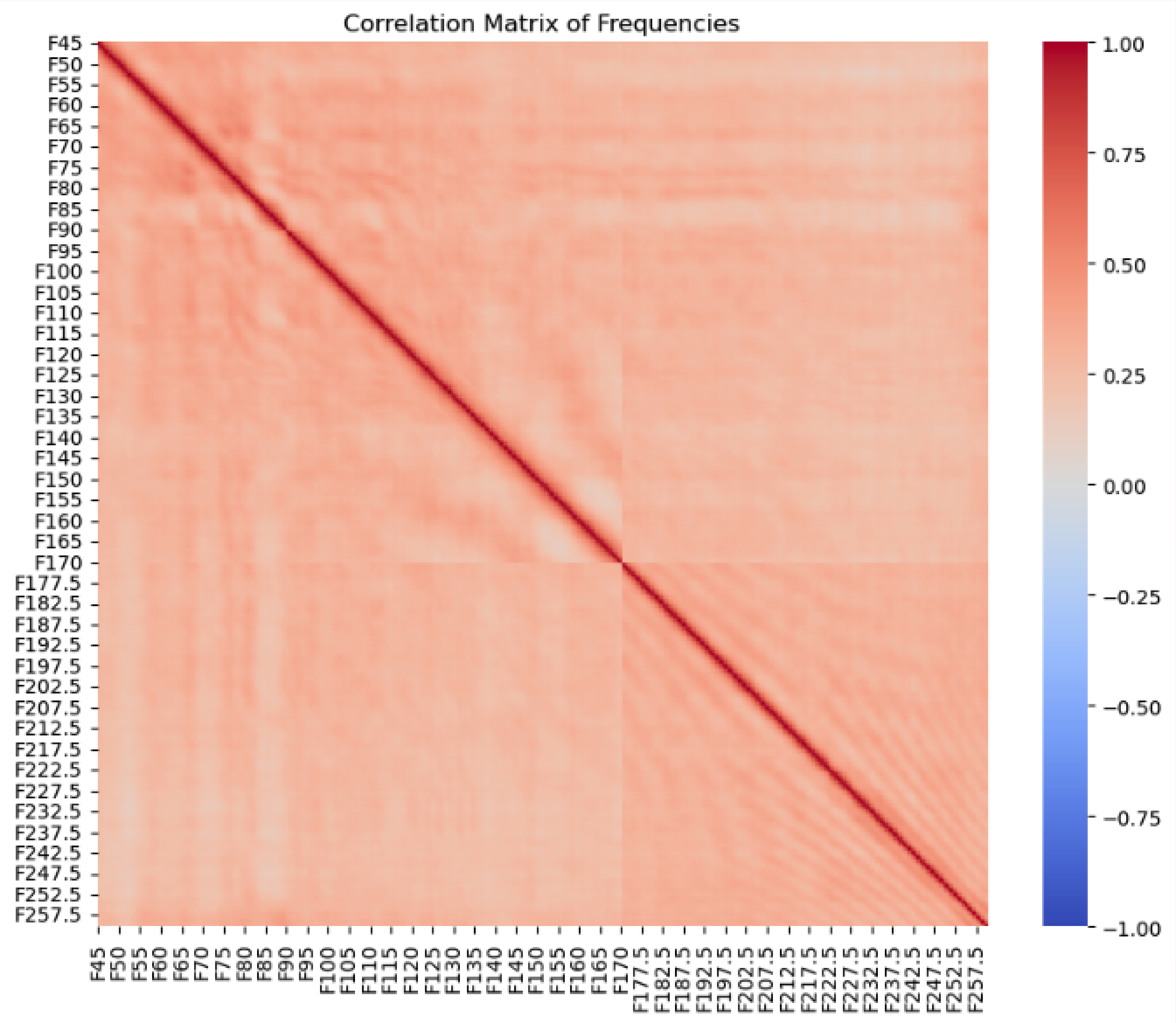
- 6,085 observations
  - 3,828 from 9 LT
  - 2,257 from 7 SMB
- 426 frequency features (45–260 Hz).

Preprocessing:

- Removed inactive fish and missing values.
- Standardized features (z-score)
- reduced dimensionality (PCA + Random Forest -> 10 key features).

Validation:

- Leave-One-Pair-Out (LOPO) to prevent data leakage.
- Stratified K-fold with K = 5.



Species	ID	Time	Ping#	F45	F45.5	...	F260
Lake Trout	10	15:11:06.1020	P_1	X_1	Y_1	...	Z_1
Lake Trout	10	15:12:07.2030	P_2	X_2	Y_2	...	Z_2
⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
Lake Trout	10	15:20:12.5060	P_n	X_n	Y_n	...	Z_n

Methodology

Machine Learning (Baselines):

- Logistic Regression (66% accuracy), XGBoost (72.8%).
- Failed to capture sequential dependencies.

Deep Learning (LSTM):

- Architecture: 2-layer LSTM (64/32 units), masking layer for padding.
- Input: 3D sequences (timesteps × features).
- Training: Adam optimizer, 10 epochs, batch size 2.

Method Description	LOPO Accuracy	K-Fold Accuracy
LSTM with 64/32 hidden units, masking layer, and Adam optimizer	0.738	0.700

Method Description	LOPO Accuracy
Logistic Regression with top features selected by Random Forest	0.658
Logistic Regression with PCA-reduced features	0.663
XGBoost using top features selected by Random Forest	0.718
XGBoost using PCA-reduced features	0.728
XGBoost with SMOTE applied to statistical features	0.75
Random Forest with SMOTE applied to statistical features	0.688

Results

LSTM Performance:

- 73.8% accuracy (LOPO validation).

Key Insight:

- Sequential modeling improves accuracy over traditional ML.

Limitation:

- Target accuracy (80%) unmet due to small dataset.

Limitations & Future Work

Limitations:

- Small dataset (16 fish), class imbalance (63% LT vs. 37% SMB).
- Computational cost and hyperparameter sensitivity.

Future Work:

- Expand dataset with more fish and environmental diversity.
- Test Fourier transforms or transformer models.
- Field validation in natural habitats.

Conclusion

Key Takeaway:

- LSTMs outperform ML models by leveraging time-series dependencies.
- Framework enables non-harmful, scalable ecological monitoring.

Next Steps:

- Improve accuracy with larger datasets and model refinement.

THANK YOU!

https://github.com/blackchocspyyy/FishSpecies

Click here to redirect to the Github.