MetalFin - Artificial Intelligence Final Project
CSI 4130
Nate Oberdier
November 24th, 2025

Abstract:

Heavy metal contamination in seafood,particularly Mercury (Hg), Lead (Pb), Arsenic (As), and Cadmium (Cd), poses a major public health concern. While research shows that contamination varies widely among species, regions, and diets, consumers often receive generalized warnings that lead to unnecessary avoidance of nutrient-dense fish. My project, MetalFin, develops a machine-learning model capable of predicting heavy metal concentrations in fish using features such as species, size, water habitat, trophic level, location, and feeding behavior. The model additionally estimates safe consumption limits for individuals based on toxicity thresholds and personal factors. For personal research purposes, this report also explores the interesting claim that cooked fish results in significantly higher mercury absorption (92%) compared to raw fish (8%), a value sourced from the writings of a controversial nutritionist, Aajonus Vonderplanitz, also with peer-reviewed scientific literature. This integrated AI and health analysis ultimately aims to promote informed, rather than fearful, seafood choices.

Introduction:

Heavy metals have almost always been traced inside fish from aquatic ecosystems. These metals have accumulated over time through industrial discharge, mining, agricultural runoff, and natural geological activity. Fish have been able to exceed the safe level of heavy metals consumption from these intoxicated living conditions. These high risking foods have led others to avoid seafood altogether, despite the high nutritional value that some offer.

The goal of my project was to collect a large amount of data ranging from multiple different sources, train a model on all of this data, estimate average values for each individual seafood, and all users to understand how they can be potentially impacted by the heavy metals in their consumption.

A big issue that consumers assess when deciding what fish is deemed safe to consume is the amount of heavy metals accumulating inside the fish. The problem with this is that humans do not fully absorb all heavy metals that are present in these seafood reports. This data misleads consumers to avoid seafood despite some heavy metals being almost completely unabsorbed.

Related Work:

I used 4 main sources to estimate the absorption rates of each of the four heavy metals. From studies, I found that mercury is mostly absorbed with a rate of 80-100%, followed by arsenic with 80-95%, lead with 5-15%, and cadmium with 3-8%. Another thing to consider is how cooking seafood can determine how the heavy metals are absorbed, mainly mercury.

My main drive for this project was to use Aajonus Vonderplanitz's findings in mercury absorption rates between raw and cooked swordfish. His studies and claims provide evidence that mercury is only absorbed 8% when raw and up to 92% when cooked. This claim can make or break someone's decision to consume seafood when it comes to a concerning mercury level. Although his study was not peer reviewed or widely discussed, I decided to include it into my project for personal interests in the topic to provide contrast to highly scientific studies and go with a more controversial take instead.

There are many models in the past that have tried predicting specific heavy metals such as mercury in tuna, lead/cadmium in freshwater fish, and classification models to predict pollutants in aquatic ecosystems. They all focus on one element yet none dive deep enough to evaluate all 4 main heavy metals such as my own.

Data:

I used multiple different datasets with 1000 different entries to provide enough variation for the model to try to find the best average value for each seafood selection. The information used were sourced from NOAA seafood contamination datasets, FAO global fishery species reports, FDA total diet/sampling data, and species metadata. I used Python scripts to merge the datasets with ChatGPT assistance and put together a master dataset to use for my model. It saved me a lot of time because I would have to otherwise manually import each dataset and combine them all together with correct formatting and clean the entire dataset. The dataset consisted of the following metrics: species, location, habitat, diet, length_cm, weight_g, mercury_mg_kg, lead_mg_kg, arsenic_mg_kg, and cadmium_mg_kg. These values were more than enough to train my model and provide valuable information to users.

Methods:

MetalFin is a probabilistic model that combines both regression and classification, completely written in Google Colab. Random Forest Regression was used to find the highest mercury levels in each seafood item so that the model can determine which value should be set for each of the 4 heavy metals within each fish. I also tested these results against Gradient Boosting Regressor, XGBoost Regressor, and Linear Regression which was used as a baseline. Random Forest had the best results which was used in the final version of my regression model.

The classification model used is Random Forest Classifier which separates each seafood item into 1 of 3 different categories labeled low, medium, and high, determining how much heavy metals they contain. I also tested Support Vector Machine, Logistic Regression, and K-Nearest Neighbors. Random Forest happened to have the best results for the classification model as well which is why it was also used in the final version.

The code was preprocessed into 70% training, 15% testing, and 15% validation. This allowed the models to accurately be trained and properly tested with a reasonable amount of validation. Any missing values were replaced with median values so that other metric values were still taken into consideration. Overly contaminated fish data was removed to reduce skewed information for a fish that would usually have lower heavy metals. Many fish species were combined together to get a general result rather than having over complications. Many salmon species were grouped together for this reason.

The absorption rates were already stated in related work. They provided that mercury is mostly absorbed with a rate of 80-100%, followed by arsenic with 80-95%, lead with 5-15%, and cadmium with 3-8%. I used Aajonus' research to compose a simple tweak to mercury's absorption rate which states that when consuming raw fish, you only absorb 8% of mercury and dispose of the rest. This allowed me to implement this percentage into my project and alter many species safety concerns. I then used my findings of absorbable heavy metals and compared the amounts to datasets which determine how safe it is to consume specific amounts of heavy metals. This made it possible to classify each seafood/fish into 3 different categories: Safe, Risky, and Dangerous.

After I coded the models for my project, I implemented an AI tool that helped users compare and examine the heavy metals inside of fish based on their species, where it was raised, how it will be consumed, and an amount(with custom units). I made it possible for users to compare as many combinations of seafood at the same time with a button that can clear past search results.

Experiments & Results:

After running my processed data, the Random Forest Regression model was evaluated on MAE, RMSE, and R^2. The validation set had MAE: 0.1367 mg/kg, RMSE: 0.2197 mg/kg, and R^2: 0.6384. The testing set had MAE: 0.1422 mg/kg, RMSE: 0.2476 mg/kg, and R^2: 0.5591. These metrics demonstrate that this regression model has a high amount of variance when it comes to heavy metals within fish. There was a wide variation of results for both sets showing good amounts of generalization for validation and test scores.

The Random Forest Classifier model was evaluated on precision, recall, f1-score, and support. The purpose of this model was to sort the fish into 3 different categories indicating the safety level of consumption. The validation and test score were very similar showing strong performance in low risk and high risk seafood classification. Medium risk classification had a relatively lower precision score meaning it was more difficult to accurately assign this risk level correctly.

Once training was complete, averages within species and metal ratio metrics were computed for more accurate results. This made biological factors more accurate, such as predators having higher heavy metal content. This also significantly improved clarity and results in the AI tool. It assigned most smaller fish into low risk categories and larger fish into higher risks.

The model and tool were now both complete and now it was time to do testing on how the metrics line up. Before taking Aajonus' absorption rate of mercury into account, many fish species were either risky or dangerous to consume on a weekly basis. This could've potentially scared users from eating specific fish from the high amounts of mercury present. After converting the raw preparation into a 8% absorbed rate, almost all of the fish were safe to eat on a weekly basis, including almost 2 pounds of swordfish. This was very surprising to find as swordfish is one of the most heavy metal contaminated fish to consume, as we've been told.

Although mercury is less of a concern for raw preparations of certain seafoods, arsenic is still 95% absorbable with high values in almost every fish. In almost every occurrence, the fish is unsafe to eat based on high levels of absorbable arsenic. We've been told to avoid certain types of fish our entire lives because of high mercury levels which aren't even the main concern. Surprisingly enough, arsenic is the biggest worry when consuming raw forms of seafood which would be the last thing to consider for many consumers.

Conclusion:

MetalFin predicts the contaminant levels of all seafood within the combined dataset through machine learning techniques with many different metrics and combinations. This was done through ecological, biological, and production-type metadata found in datasets relating to seafood and fish. My Random Forest Regression model was able to achieve a high performance with a R^2 validation value of 0.6384, and test value of 0.5591, proving that the data can predict contamination to a degree. The Random Forest Classification model also efficiently categorizes low and high risk categories of seafood through common heavy metal patterns. Medium risk contamination categories result in lower precision but eventually overlap to provide meaningful structure in the model.

MetalFin provides a useful source for consumers to interact and interpret seafood safety with more transparency. It takes into account controversial and experimental information to decrease scare factors that would otherwise steer them away from consuming seafood. It allows them to fully understand that not all seafood is dangerous to consume, raw variations have much lower levels of mercury, risk is entirely species dependent, and machine learning can be used to inform and influence better dietary choices.

The most impactful result from this project was how low the absorbable rates of heavy metals were despite how many seafoods are labeled as "dangerous" for being contaminated. Half of heavy metals are minimally absorbable, slightly reducing the risk for contaminants inside of seafood. The main concern for any consumer of raw seafood should be arsenic due to its high levels of absorbability with an average rate of 95%. The other rates of absorption for raw preparation are the following: mercury with 8%, lead with 10%, and cadmium with 5%. These values are so significantly low which can potentially change the consideration of seafood consumption. When considering what seafood to purchase, the amount of arsenic should always be considered based on the findings of MetalFin. This project does not advise others to consume raw food, but only to educate different properties and absorption rates of raw and cooked seafood.

Citations & Acknowledgements:

Agency for Toxic Substances and Disease Registry. (2000). Toxicological Profile for Lead. U.S. Department of Health and Human Services. https://www.atsdr.cdc.gov/toxprofiles/tp13.pdf

Barone, G., Storelli, A., Garofalo, R., Mallamaci, R., & Storelli, M. M. (2022). Residual Levels of Mercury, Cadmium, Lead and Arsenic in Some Commercially Key Species from Italian Coasts (Adriatic Sea): Focus on Human Health. Toxics, 10(5), 223. https://doi.org/10.3390/toxics10050223

Charette, T., Bueno Dalto, D., Rosabal, M., Matte, J. J., & Amyot, M. (2021). Assessment of In Vitro Bioaccessibility and In Vivo Oral Bioavailability as Complementary Tools to Better Understand the Effect of Cooking on Methylmercury, Arsenic, and Selenium in Tuna. Toxics, 9(2), 27. https://doi.org/10.3390/toxics9020027

Clarkson, T. W., & Magos, L. (2006). The Toxicology of Mercury and Its Chemical Compounds. Critical Reviews in Toxicology, 36(8), 609–662. https://doi.org/10.1080/10408440600845619

EFSA Panel on Contaminants in the Food Chain. (2009). Scientific opinion on mercury in food. EFSA Journal, 7(10), 1351. https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2009.1351

EPA. (2021). EPA-FDA fish advice: Technical information (archived snapshot). U.S. Environmental Protection Agency. https://19january2021snapshot.epa.gov/fish-tech/epa-fda-fish-advice-technical-information_.html

EPA. (2024). Fish tissue data collected by EPA and partners. U.S. Environmental Protection Agency. https://www.epa.gov/choose-fish-and-shellfish-wisely/fish-tissue-data-collected-epa-partners

Food and Drug Administration. (2014). Mercury levels in commercial fish and shellfish (1990–2012). U.S. Food and Drug Administration. https://www.fda.gov/food/environmental-contaminants-food/mercury-levels-commercial-fish-and-shellfish-1990-2012

Garofalo, L., Sala, M., Focardi, C., Pasqualetti, P., Delfino, D., D'Onofrio, F., Droghei, B., Pasquali, F., Nicolini, V., Galli, F. S., Scaramozzino, P., Ubaldi, A., Russo, K., & Neri, B. (2025). Monitoring of Cadmium, Lead, and Mercury Levels in Seafood Products: A Ten-Year Analysis. Foods, 14(3), 451. https://doi.org/10.3390/foods14030451

Gelfond Laboratory at Stony Brook University. (2012). Seafood mercury database. https://www.stonybrook.edu/commcms/gelfond/_pdf/Seafood%20Mercury%20Database.pdf

Järup, L., & Åkesson, A. (2009). Current status of cadmium as an environmental health problem. Toxicology and Applied Pharmacology, 238(3), 201–208. https://doi.org/10.1016/j.taap.2009.04.020

Li, X., et al. (2025). Global patterns of mercury in seafood: A harmonized open dataset. Scientific Data, 12, Article 45. https://doi.org/10.1038/s41597-025-04570-3

Vonderplanitz, A. (2006). Mercury in fish: Do we absorb it? https://aajonus.net/mercury-in-fish-do-we-absorb-it