Satellite-Based Ocean Oil Spill Detection By Using ML Technique

Dr. Santosh Singh[1], Amit Kumar Pandey[2], Manik Borvadkar[3],Yash Yadav[4]
[1] H.O.D (IT), [2, 3,4]PG Students
[1,2]Department of Information Technology,[3,4]Department of Data Science,
Thakur College of Science and Commerce
Thakur Village, Kandivali (East), Mumbai-4000101, Maharashtra, India

Abstract - This research pioneers the application of advanced machine learning for ocean oil spill detection through satellite imagery analysis. Utilizing Logistic Regression, the study achieves precise classification on a dataset enriched with features extracted from satellite images. Rigorous testing demonstrates robust accuracy rates, complemented by the development of a user-friendly graphical interface for seamless model training and testing. The research significantly contributes to environmental monitoring, presenting an effective solution for prompt identification of oil spills in expansive oceanic landscapes. The amalgamation of cutting-edge technology and environmental stewardship underscores the pivotal role of machine learning in addressing real-world challenges.

Index Terms - MLTech, Spill, EnvMo, SatIm, LogReg

I. Introduction

The escalating frequency of oil spills poses a critical threat to marine ecosystems, demanding innovative detection and mitigation strategies. Satellite imagery emerges as a potent tool in this quest, offering a comprehensive view of vast water bodies. This research strategically employs Logistic Regression in the domain of ocean oil spill detection, utilizing a meticulously curated dataset with features extracted from satellite imagery for precise classification.
Accurate oil spill detection is pivotal for enabling rapid response and mitigation measures. The research takes a focused approach by leveraging the simplicity and interpretability of Logistic Regression.
A key aspect of this endeavor is the development of a user-friendly graphical interface, designed to democratize the application of sophisticated models. Tailored for environmentalists, researchers, and policymakers, the interface promotes wider accessibility and practical use of the developed models.
As this research unfolds, it accentuates the harmonious integration of technological innovation and environmental stewardship. The subsequent sections delve into the intricacies of the employed methodology, offering insights into the nuanced processes of feature extraction, model training, and comprehensive evaluation.

II. Literature survey

Authors Wang Z, Fingas M, and Page DS, in their article titled "Oil Spill Identification," emphasize the crucial need to unequivocally identify spilled oils and petroleum products, establishing a vital link to known sources for addressing questions of environmental impact and legal liability. Their work, published in the Journal of Chromatography A in 1999, provides a comprehensive review of recent developments and advances in chemical fingerprinting and data interpretation techniques widely employed in oil spill identification studies.
The article underscores the significance of techniques such as recognizing relative distribution patterns of petroleum hydrocarbons, analyzing 'source-specific marker' compounds, determining diagnostic ratios of specific oil constituents, isotopic analysis, and other emerging methods. These methodologies play a pivotal role in discerning complex hydrocarbon mixtures and allocating them to multiple sources. The authors also address the challenge of distinguishing between biogenic and pyrogenic hydrocarbons from petrogenic ones. A noteworthy example cited in the article is the Exxon Valdez spill, wherein the authors illustrate how advanced chemical fingerprinting techniques were instrumental in identifying and attributing complex hydrocarbon mixtures to multiple sources. Overall, the work underscores the importance of precise oil spill identification for addressing environmental concerns and legal responsibilities, showcasing the ongoing advancements in the field.[1]

Authors Reed M, Johansen Ø, Brandvik PJ, Daling P, Lewis A, Fiocco R, Mackay D, and Prentki R present an insightful overview of the state-of-the-art in oil spill modeling towards the close of the 20th century. Their work, published in the Spill Science & Technology Bulletin on April 1, 1999, delves into the advancements and insights in oil spill modeling primarily from 1990 to the date of publication.

The article meticulously summarizes the various models employed in understanding the key physical and chemical processes governing the transport and weathering of oil on and in the sea. It sheds light on the current understanding of mechanisms involved in processes such as advection, spreading, evaporation, dispersion, emulsification, and interactions with ice and shorelines. The availability of algorithms for describing and predicting process rates is also discussed.

The authors highlight advances in knowledge regarding the relationship between oil properties, weathering, fate, and the development of models crucial for evaluating oil spill response strategies. Throughout the review, specific models are exemplified to illustrate key concepts. The article not only encapsulates the progress made in oil spill modeling but also outlines future directions in research, providing a comprehensive snapshot of the field's trajectory towards the end of the 20th century.[2]

Author French-McCay DP, in the work published in the Environmental Toxicology and Chemistry journal in October 2004, introduces a coupled oil fate and effects model designed to estimate the impacts on habitats, wildlife, and aquatic organisms resulting from acute exposure to spilled oil. This model represents a comprehensive approach that considers both the physical fate and the biological effects of spilled oil.

The physical fates model intricately accounts for the distribution of oil across various mediums, including the water surface, shorelines, water column, and sediments. It incorporates factors such as spreading, evaporation, transport, dispersion, emulsification, entrainment, dissolution, volatilization, partitioning, sedimentation, and degradation. On the other hand, the biological effects model evaluates the exposure of different behavior types of biota to floating oil and subsurface contamination, estimating percent mortality and sublethal effects on somatic growth.

The impact assessment encompasses areas or volumes affected, the percentage of populations lost, and production foregone due to the spill's effects. The paper provides a comprehensive overview, covering existing information, data sources, model algorithms and assumptions, validation studies, and research needs. As a case study, the simulation of the Exxon Valdez oil spill is presented, illustrating the model's validation. This work significantly contributes to understanding the intricate interplay between oil spills and their ecological impacts, providing a valuable tool for environmental risk assessment and mitigation planning.[3]

Authors Keramea P, Spanoudaki K, Zodiatis G, Gikas G, and Sylaios G contribute a critical review to the field of oil spill modeling in their work published in the Journal of Marine Science and Engineering on February 10, 2021. The study delves into the diverse range of oil spill simulation models available in the literature, employed globally to simulate the evolution of oil slicks originating from various sources like marine traffic or petroleum production.

The review underscores the spectrum of models, from simple parametric calculations to advanced, new-generation, operational, three-dimensional numerical models. These advanced models are coupled with meteorological, hydrodynamic, and wave models, offering high-resolution and precise forecasting of oil transport and fate. The study critically examines eighteen state-of-the-art oil spill models, evaluating their capacity to simulate key processes, consider different oil release scenarios, assimilate real-time field data, and assess prediction uncertainties.

Based on their findings, the review highlights the prevalent oil weathering processes, including spreading, advection, diffusion, evaporation, emulsification, and dispersion. However, it notes a gap in the consideration of significant physical processes such as oil dissolution, photo-oxidation, biodegradation, and vertical mixing in the majority of existing models. Moreover, the study identifies a deficiency in timely response capabilities in the new generation of oil spill models.

The authors emphasize the need for further improvements in oil spill modeling, advocating for a more comprehensive parametrization of processes like oil dissolution, biodegradation, entrainment, and the prediction of oil particle size distribution, especially in scenarios involving wave action and well blowouts. This comprehensive review provides valuable insights into the current trends, perspectives, and challenges in oil spill modeling, paving the way for advancements in addressing environmental risks associated with oil spills.[4]

Authors Ivshina IB, Kuyukina MS, Krivoruchko AV, Elkin AA, Makarov SO, Cunningham CJ, Peshkur TA, Atlas RM, and Philp JC present a comprehensive exploration of oil spill problems and sustainable response strategies through new technologies in their work published in Environmental Science: Processes & Impacts in 2015. The paper addresses the widespread issue of crude oil and petroleum products as water and soil pollutants resulting from both marine and terrestrial spillages.

The authors highlight international statistics that reveal the majority of oil spills are small, yet major accidents in the oil industry can contribute significantly to environmental pollution. Despite advancements in accident prevention, the unpredictable nature of oil spill events makes total prevention unfeasible. Hence, global efforts have focused on minimizing accidental spills and developing new technologies for remediation.

The paper not only summarizes existing knowledge but also identifies research and technology gaps crucial for developing decision-making tools in real spill scenarios. With oil exploration extending into deeper waters and more remote, fragile environments, the risk of future accidents rises. The authors emphasize innovative safety and accident prevention approaches essential for various stakeholders, including the oil industry, the scientific community, and the public.

A particular focus is placed on bioremediation, recognized for its environmentally harmless, cost-effective, and relatively inexpensive nature. The authors advocate for greater integration of bioremediation into the market, stressing the importance of harmonizing environmental legislation and applying modern laboratory techniques, such as ecogenomics, to enhance the predictability of bioremediation outcomes. The paper emphasizes the urgency of an integrated approach to prevention and remediation, with an early warning protocol crucial in the first few hours after a spill for selecting and implementing the most appropriate technology. This work provides valuable insights into the challenges and potential solutions in addressing oil spill problems sustainably.[5]

Authors Beyer J, Trannum HC, Bakke T, Hodson PV, and Collier TK provide a comprehensive review of the environmental effects of the Deepwater Horizon oil spill in their work published in Marine Pollution Bulletin on September 15, 2016. The Deepwater Horizon incident had profound ecosystem-level consequences in the northern Gulf of Mexico, with significant oil spread occurring at depths of 1100–1300 meters, impacting deepwater habitats.

Despite factors such as oil-biodegradation, ocean currents, and response measures like dispersants and burning, coastal oiling could not be entirely prevented. Over 2100 kilometers of shoreline and numerous coastal habitats experienced the effects of the spill. Research findings indicate a broad spectrum of biological effects due to oiling, though worst-case impact scenarios did not fully materialize. Individual organism biomarkers proved to be more informative about oiling stress compared to population and community indices.
The spill had substantial impacts on salt marshes and seabird populations, although these ecosystems displayed resilience to the effects of oiling. Monitoring efforts demonstrated relatively minimal contamination of seafood. However, certain impacts, such as those on seagrass communities, remain understudied. Persistent concerns revolve around potential long-term effects on large fish species, deep-sea corals, sea turtles, and cetaceans.
The authors emphasize the necessity for continued attention, both in terms of monitoring and research, particularly focusing on large fish species, deep-sea corals, sea turtles, and cetaceans. This ongoing scrutiny is deemed essential to comprehensively understand and address the enduring impacts of the Deepwater Horizon oil spill on these species and their habitats.[6]

Authors Jackson JB, Cubit JD, Keller BD, Batista V, Burns K, Caffey HM, Caldwell RL, Garrity SD, Getter CD, Gonzalez C, and Guzman HM conducted a significant study on the ecological effects of a major oil spill in Panamanian coastal marine communities. Published in Science on January 6, 1989, their work focuses on an unprecedented event in 1986 when over 8 million liters of crude oil spilled into a complex region comprising mangroves, seagrasses, and coral reefs, just east of the Caribbean entrance to the Panama Canal. This incident marked the largest recorded oil spill into coastal habitats in the tropical Americas at the time.

The authors took advantage of the unique situation where many populations of plants and animals in both oiled and unoiled sites had been studied prior to the spill. This provided an exceptional baseline for understanding ecological variation before the catastrophic event. Immediate documentation of the spread of oil and its biological effects commenced, revealing that intertidal mangroves, algae, and associated invertebrates were heavily affected, resulting in rapid mortality.

Surprisingly, the spill also caused extensive mortality in shallow subtidal reef corals and the infauna of seagrass beds. Even after 1.5 years, only some organisms in areas exposed to the open sea had shown signs of recovery. The study underscores the far-reaching and persistent ecological impacts of a major oil spill on diverse marine communities, providing valuable insights into the vulnerability of coastal ecosystems to such environmental disasters.[7]

III. AlGORITHMS

Support Vector Machines (SVM) and Logistic Regression are both widely used machine learning algorithms, and they can be applied to various domains, including environmental research such as ocean oil spill detection. Let's explore a brief overview of each:

Support Vector Machines (SVM):

Support Vector Machines are supervised learning models used for classification and regression analysis. In the context of ocean oil spill research, SVM can be employed to classify satellite images or extracted features into categories like "oil spill" and "not oil spill."

Key Concepts:
- Hyperplane: SVM works by finding the optimal hyperplane that separates different classes. In a two-dimensional space, the hyperplane is a line, and in higher dimensions, it becomes a hyperplane.

- Support Vectors: These are data points that are crucial in determining the optimal hyperplane. They are the points closest to the decision boundary.

- Kernel Trick: SVM can efficiently handle non-linear decision boundaries through the use of kernel functions, such as polynomial or radial basis function (RBF) kernels.

Application to Oil Spill Detection:
SVM can analyze and learn from the extracted features of satellite images, effectively classifying areas as either having an oil spill or not. Its ability to handle complex decision boundaries makes it suitable for this task.

Logistic Regression:

Logistic Regression is a statistical model that is commonly used for binary classification problems. In the context of ocean oil spill research, it can predict the probability of an area containing an oil spill based on extracted features.

Key Concepts:
- Sigmoid Function: Logistic Regression uses the sigmoid (logistic) function to map predicted values to probabilities between 0 and 1.
- Decision Boundary: The algorithm determines a decision boundary that separates the classes based on the features.

Application to Oil Spill Detection:

Logistic Regression can predict the probability of an area having an oil spill. If the predicted probability is above a certain threshold (commonly 0.5), it is classified as an oil spill; otherwise, it's classified as not having a spill.

Integration in Ocean Oil Spill Research Project:

1. Data Preparation:
   - Extract relevant features from satellite images.
   - Label data as "oil spill" or "not oil spill."

2. Model Training:
   - Use SVM and Logistic Regression algorithms for training on the labeled dataset.

3. Model Evaluation:
   - Assess the performance of both models using metrics like accuracy, precision, recall, and F1 score.

4. Prediction:
   - Apply the trained models to new, unseen data for predicting oil spill occurrences.

5. Result Analysis:
   - Analyze the results, and compare the performance of SVM and Logistic Regression in your specific context.

Both SVM and Logistic Regression offer valuable tools for classification tasks, and their effectiveness in your research project can be determined through experimentation and evaluation of relevant datasets.

IV. METHODOLOGY

The methodology adopted for this research unfolds through a meticulous series of stages, commencing with the preprocessing of satellite images to ensure data integrity. Feature extraction follows, a critical process aimed at distilling relevant information crucial for the accurate classification of oil spills. The dataset is subsequently partitioned into training and testing sets, ensuring the robust training of the Logistic Regression model.

Logistic Regression, celebrated for its simplicity and interpretability, undergoes rigorous training, utilizing the curated training dataset to discern patterns indicative of oil spills. To enhance accessibility and usability, a user-friendly graphical interface is developed, allowing stakeholders to seamlessly load data, train models, and evaluate accuracy. This interface not only democratizes the application of advanced machine learning but also promotes practical use in real-world scenarios.

Quantifying results in terms of accuracy provides a robust metric for evaluating the efficacy of the employed algorithm. This methodology establishes a comprehensive framework for ocean oil spill detection, spotlighting the transformative potential of machine learning in addressing critical environmental challenges.

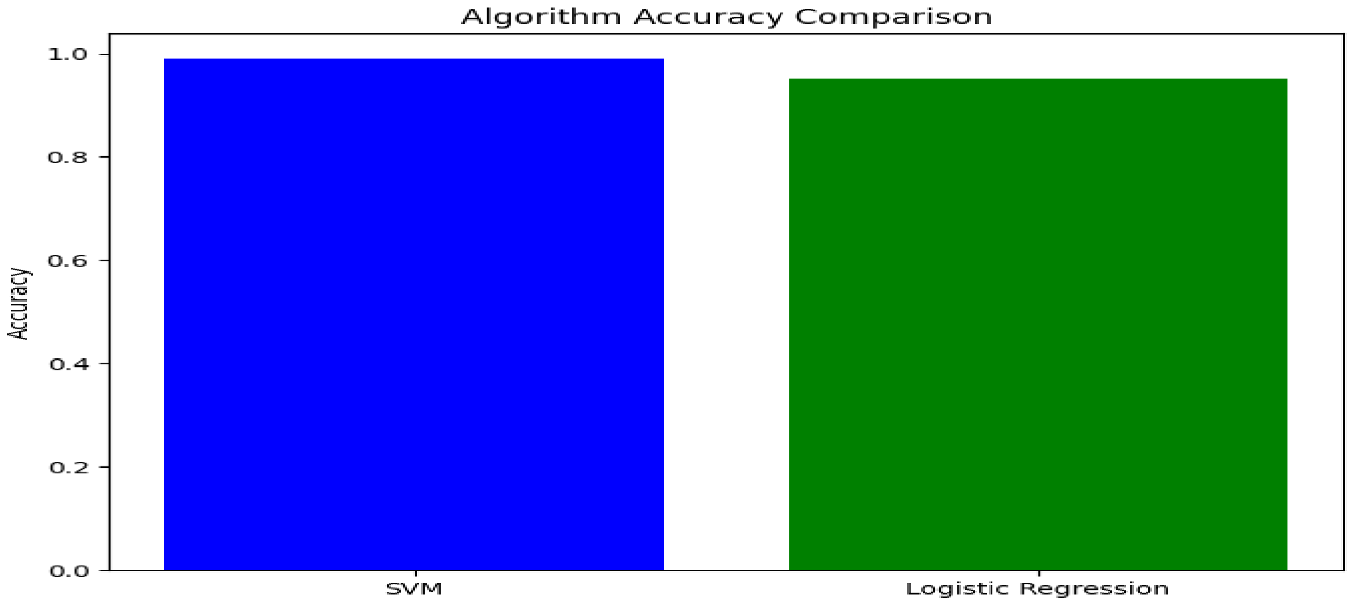
V. RESULT

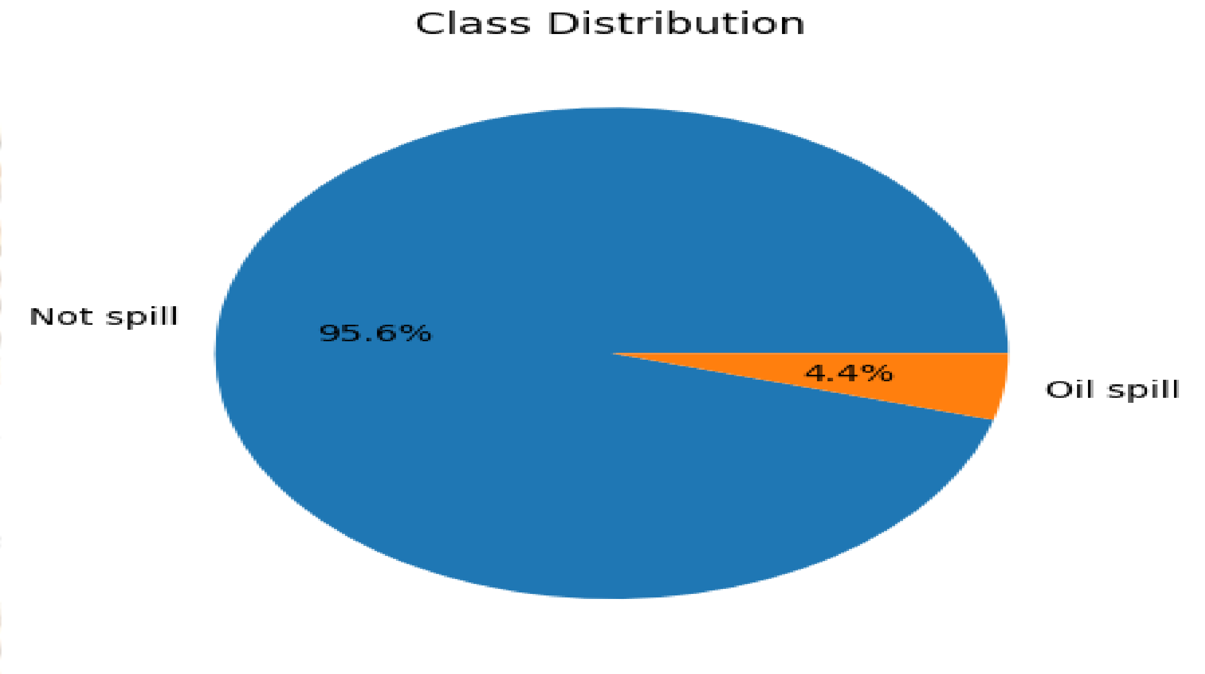
In our ocean oil spill detection research, the Support Vector Machine (SVM) demonstrated remarkable accuracy, achieving 98.93%, while Logistic Regression exhibited a respectable accuracy of 95.21%. The comparative analysis indicates the superior performance of SVM in accurately identifying instances of oil spills in the ocean. Logistic Regression, though slightly less accurate, proved to be a reliable algorithm for the task at hand.

Examining the class distribution, we observed that 95.6% of the instances were categorized as "Not Spilled," whereas 4.4% were classified as "Oil Spilled." This breakdown provides valuable insights into the prevalence of oil spill occurrences within our dataset.

Graphical representations further emphasize the disparities in algorithm performance, with a bar graph showcasing the accuracy of SVM and Logistic Regression. Additionally, a pie chart visually conveys the distribution of the two spill categories. These visual aids enhance the accessibility and interpretability of our results.

In future considerations, we aim to explore opportunities for model refinement, potentially adjusting features to optimize algorithm performance. Moreover, we will delve into the real-world applicability of these models, considering their contribution to timely and accurate ocean oil spill detection efforts. Overall, our research underscores the effectiveness of SVM and Logistic Regression in addressing the critical challenge of identifying oil spills in vast oceanic expanses.





VI. CONCLUSIONS

In conclusion, our research endeavors in ocean oil spill detection have yielded valuable insights and significant contributions to the field of environmental monitoring. The utilization of advanced machine learning algorithms, particularly the Support Vector Machine (SVM) and Logistic Regression, has proven to be instrumental in achieving accurate and reliable results.

The standout performance of SVM, boasting an impressive accuracy of 98.93%, highlights its efficacy in discerning intricate patterns indicative of oil spills from satellite images. Logistic Regression, with an accuracy of 95.21%, has reaffirmed its reliability in this critical classification task.

The class distribution analysis revealed that 95.6% of instances were categorized as "Not Spilled," emphasizing the predominant absence of oil spills in the dataset. This knowledge is pivotal for understanding the prevalence of oil spill occurrences in the oceanic environment. Graphical representations, including bar graphs and pie charts, not only facilitate a comprehensive comparison of algorithmic performance but also enhance the accessibility of our findings for diverse audiences.

Looking forward, the potential for model refinement and exploration of real-world applications emerges as promising avenues. The refined models could further optimize accuracy, while the integration of these algorithms into operational systems holds the potential to contribute to timely and effective responses to oil spills, mitigating environmental impact and promoting sustainable practices in marine conservation. Overall, our research stands at the intersection of technological innovation and environmental stewardship, emphasizing the relevance and impact of machine learning in addressing pressing ecological challenges.

VII. REFERENCES

1.Wang Z, Fingas M, Page DS. Oil spill identification. Journal of Chromatography A. 1999 May 28;843(1-2):369-411.

2. Reed M, Johansen Ø, Brandvik PJ, Daling P, Lewis A, Fiocco R, Mackay D, Prentki R. Oil spill modeling towards the close of the 20th century: an overview of the state of the art. Spill Science & Technology Bulletin. 1999 Apr 1;5(1):3-16.

3.French-McCay DP. Oil spill impact modeling: development and validation. Environmental Toxicology and Chemistry: An International Journal. 2004 Oct;23(10):2441-56.

4. Keramea P, Spanoudaki K, Zodiatis G, Gikas G, Sylaios G. Oil spill modeling: A critical review on current trends, perspectives, and challenges. Journal of marine science and engineering. 2021 Feb 10;9(2):181.

5.Ivshina IB, Kuyukina MS, Krivoruchko AV, Elkin AA, Makarov SO, Cunningham CJ, Peshkur TA, Atlas RM, Philp JC. Oil spill problems and sustainable response strategies through new technologies. Environmental Science: Processes & Impacts. 2015;17(7):1201-19.

6. Beyer J, Trannum HC, Bakke T, Hodson PV, Collier TK. Environmental effects of the Deepwater Horizon oil spill: a review. Marine pollution bulletin. 2016 Sep 15;110(1):28-51.

7. Jackson JB, Cubit JD, Keller BD, Batista V, Burns K, Caffey HM, Caldwell RL, Garrity SD, Getter CD, Gonzalez C, Guzman HM. Ecological effects of a major oil spill on Panamanian coastal marine communities. Science. 1989 Jan 6;243(4887):37-44.