COMPUTATIONAL FRAMEWORK AND ML-DRIVEN MODELING FOR PREDICTION OF OPTIMAL BIOPLASTIC COMPOSITION FOR THE DEVELOPMENT OF ECO-FRIENDLY POLYMER NANOCOMPOSITE BIOPLASTICS FROM AGRICULTURAL RESIDUES

Project Reference No.: *49S_MSC_0185*

College : St Joseph's University, Bengaluru
Branch : Department of Computer Science
Guide(s) : Dr. Mrinmoyee Bhattacharya
Dr. A M Bojamma
Student(s) : Mr. Harikrishna Alungal Prakas
Mr. Bharath B M
Mr. Jonathan Moncy Mathew
Mr. Tanishq Mani Maran Raj

Keywords:

Bioplastics, Agricultural Residues, Machine Learning, Bayesian Optimization, Food Packaging, Life Cycle Assessment, Polymer Nanocomposites, Sustainable Materials

Introduction:

1. Pollution issue is worsening day in day out. We should actually devise ways of preparing plastics that are eco-friendly. We produce more than 400 million tons of plastic per annum. The greater part of this plastic is produced out of such materials as polypropylene, polyethylene and PET. These plastics are not good to the environment since they are not easily degraded.
2. Farming wastes such as rice husks, banana peels and sugarcane bagasse are a lot and can be used to produce plastics. It is one method of utilizing waste and creating something new and useful. It also works well on the environment since it is waste less and will help us produce things.
3. The traditional method of creation in plastics is to experiment with a myriad of different factors until we get one that works. This is time consuming and is also costly. Machine learning can be used to assistance to make new plastics cheaper and faster. Machine learning will be able to analyse massive data. A prediction of the new plastic characteristics. It is an assurance that we do not need to initiate numerous experiments in the laboratory.
4. Random Forest type of algorithm are highly effective predictors of what a new plastic will possess. The best process to make a new plastic can be easily identified within a short period of time when we apply these algorithms using such processes as Bayesian optimization and genetic algorithms.
5. The present project concerns creating a computer program which is based on machine learning so that we can use farm waste to produce new plastics. We desire to have safe, to use, with food, plastics, and are good, to the environment. It is also in the process of making a web site that will simplify the process of using this program by researchers and designers. This will assist in making packaging quicker.

Objectives:

1. **To collect the data on waste-derived bioplastics and analyze it:**
   To systematize, gather and appraise the obtained data about bioplastics derived out of agricultural waste materials such as rice husks, banana peels and sugarcane bagasse in their chemical composition and property with regard to their application in food packaging.
2. **Evaluation of the material properties using different machine learning models**
   To produce and train higher quality machine learning models, including the Random Forest, and Multitask Learning, to predict the qualities of bioplastics, including tensile strength, elongation, oxygen permeability, water vapour permeability, and transparency, qualities of superior quality.
3. **To optimize bioplastic formulations and processing conditions**
   To use optimization techniques, including Bayesian optimization, genetic algorithms, and grid search to determine the best combinations of raw materials, additives, fillers, and processing parameters that result in improved bioplastic performance.
4. **To assess the environment sustainability of the fabric compared with other traditional plastics.**
   To conduct a comparative analysis of the green impact of the ready bioplastic compounds to the traditional petroleum-sourced plastics, e.g., polypropylene (PP), polyethylene (PE) and polyethylene terephthalate (PET), regarding such variables as biodegradability, carbon footprint and lifecycle impact**.**

Methodology:

1. The proposed framework begins with collection and preparation of a dataset of 200 to 500 formulations. These bioplastic formulations come from published literature, experimental studies and material databases like PolyInfo and Polymer Genome. Each record in the dataset includes information like the type of residue used such as rice husk, banana peel or sugarcane bagasse as well as the ratios of plasticizer and filler and the conditions used to process the bioplastic formulations. The dataset also includes the properties of the formulations like tensile strength, elongation at break oxygen permeability, water vapor permeability and transparency.
2. Missing values are then filled in with a method known as k-NN to prepare the dataset ready to be used and the k-NN method is also used to remove outliers using the IQR method. The data is then normalized using Min-Max normalization and categorical variables are encoded using one- encoding.
3. The models of machine learning are then trained on the data: Random Forest, and a Multitask Learning model. The Multitask Learning model is capable of making predictions of all five properties of the formulations simultaneously based on the interrelations between the properties. . All the models are tested using 5fold cross-validation and their performance is measured using metrics like R^2, RMSE and MAE. The goal is to get an R^2 value of than 0.85. SHAP analysis is also used to understand how each feature of the formulations affects the predicted properties.
4. The machine learning models that are validated are then used to optimize the formulations. This is done using three strategies: Bayesian Optimization, Genetic

Algorithms and Grid Search. The goal of the optimization is to maximize the tensile strength and transparency of the formulations while minimizing their permeability.
5. The optimized bioplastic formulations are then compared to plastics like PP, PE and PET using a Life Cycle Assessment. The Life Cycle Assessment looks at the Global Warming Potential -Renewable Energy Use and Eutrophication Potential of the bioplastic formulations.

Result and Conclusion:

The objective of the project is to produce a list of 200-600 bioplastic formulations prepared using agricultural waste. This data set will aid in developing models that are capable of making predictions on five major properties of bioplastics:

* Tensile Strength
* Elongation At Break
* Oxygen Permeability
* Water Vapour Permeability
* Transparency.

1. All properties are supposed to have their R^2 value more than 0.85. The project is based on a form of machine learning known as Multitask Learning. The method will assist to enhance precision due to the consideration of relationship patterns among various properties.
2. A method such as SHAP is also applied in the project to learn the influence of various ingredients, such as the type of waste, the concentration of plasticizers and their filler ratio on the properties of bioplastics. This is with an aim of discovering the food packaging standards of the bioplastic formulations. Based on Bayesian Optimization and Genetic Algorithms, the project reduces the number of possible formulations to only 5 to 10 promising formulations. It is hoped that such optimized bioplastics would have a difference in environmental characteristics with standard plastics made of petroleum. The project will evaluate the effects of these bioplastics in terms of Life Cycle Assessment.
3. In conclusion this project demonstrates that learning to make bioplastics with the use of agricultural waste with machine learning is a strong strategy. Through the use of data-driven prediction, optimization, and environmental analysis the project will accelerate the discovery of eco- bioplastic formulations to food packaging.
4. The work serves a purpose of minimizing plastic pollution and increasing the use of a circular economy. It assists in bridging it, between large-scale industrial production and waste to create sustainable packaging.

References:

- I. Malashin, D. Martysyuk, V. Tynchenko, A. Gantimurov, A. Semikolenov, V. Nelyub, and A. Borodulin, "Machine Learning-Based Process Optimization in Biopolymer Manufacturing: A Review," Polymers, vol. 16, no. 23, p. 3368, Nov. 2024, doi: 10.3390/polym16233368.
- A. O. Adeola, R. Senga, and P. N. Nomngongo et al., "Sustainable Valorization of Agricultural Waste into Bioplastic and Its End-of-Life Recyclability for Biochar Production: Economic Profitability and Life Cycle Assessment," Chemosphere, vol. 370, Dec. 2024, doi: 10.1016/j.chemosphere.2024.143941.

- K. Yadav and G. C. Nikalje, "Comprehensive Analysis of Bioplastics: Life Cycle Assessment, Waste Management, Biodiversity Impact, and Sustainable Mitigation Strategies," PeerJ, vol. 12, e18013, Sep. 2024, doi: 10.7717/peerj.18013.
- R. Fujita, Y. Amamoto, and J. Kikuchi, "Bayesian Optimization of Biodegradable Polymers via Machine Learning Using Low-Field NMR Relaxation Features," npj Materials Degradation, vol. 9, p. 72, 2025, doi: 10.1038/s41529-025-00613-7.
- C. Uribe-Wandurraga, R. Palencia et al., "Sustainable Production of High-Performance Bioplastics from Agricultural and Industrial Biomass Waste by Integrating Deep Eutectic Solvent (DES) Pretreatment and Acetylation Processes," ACS Omega, vol. 10, no. 11, pp. 10949–10961, Mar. 2025, doi: 10.1021/acsomega.4c09059.
- O. D. Ogundipe, F. Adeyemo et al., "Statistical Modeling and Optimization of Bioplastic Synthesis from Waste Corn Using Polynomial Regression Analysis," Green Energy and Environmental Technology, IntechOpen Journals, vol. 7, Nov. 2024, doi: 10.5772/geet.20240012.
- I. P. Malashin, D. Martysyuk, V. S. Tynchenko, A. Gantimurov, A. Semikolenov, V. Nelyub, and A. Borodulin, "Deep Learning for Property Prediction of Natural Fiber Polymer Composites," Scientific Reports, vol. 15, no. 1, p. 27837, Jul. 2025, doi: 10.1038/s41598-025-10841-1.
- E. Champa-Bujaico, P. García-Díaz, and A. M. Díez-Pascual, "Machine Learning for Property Prediction and Optimization of Polymeric Nanocomposites: A State-of-the-Art," International Journal of Molecular Sciences, vol. 23, no. 18, p. 10712, Sep. 2022, doi: 10.3390/ijms231810712.
- S. Klébert et al., "Machine Learning-Driven Property Predictions of Polymer Composites," Composites Part B: Engineering, 2025, doi: 10.1016/j.compositesb.2025.112233.
- P. Singh, A. Sharma, A. Rathee, V. Puri, H. Chopra, M. Nagpal, M. Kaur, and T. Malik, "Revolutionizing Packaging: Bioplastics for Superior Food and Pharmaceutical Solutions," Journal of Bioactive and Compatible Polymers, vol. 39, no. 6, Nov. 2024, doi: 10.1177/09673911241305184.

Future Scope:

The current study can be improved in different directions to enhance both the robustness and practical applicability of the proposed framework for bioplastic development.

1. Dataset can be expanded using diverse agricultural waste sources

Further research would include adding to the datasets other types of lignocellulose and starch-based agricultural residues (wheat straw, coconut coir, corn starch) into the dataset. This growth will enhance the generalization and prediction of the machine learning models of a larger range of bioplastic formulations and material characteristics.

2. Incorporation with real-time experimental data

This research could also be developed with incorporating formulated machine learning models to real-time data collection systems of experiments in the laboratory. This form of integration would facilitate adaptive learning and dynamic optimization of bioplastic formulations to provide on-the-fly decisions to changing processing conditions and material compositions in response to continuously generated experimental data.

3. Integration of safety and biodegradability assessment models

The predictive models of toxicity and biodegradability will be also prioritized in the next research in order to allow meeting the food safety and environmental standards. Including these parameters in the optimization framework, the proposed study will help design bioplastics, which are not just high-performing but also safe in contact with food and are environmentally safe.