2025 Term Paper Project

Machine Learning Analysis of Natural Surface Structures

Objective: The aim of this project is to show the applications of machine learning algorithms to the analysis of natural surface structures, such as those found on leaves, shells, feathers, seeds, and insect bodies. These surfaces often display unique patterns and textures that inspire innovative engineering solutions. By studying these structures, we can discover recurring patterns, predict functional properties, and identify similarities between different types of surfaces. In this project, you will apply multiple machine learning methods to the same dataset and demonstrate how different approaches could provide insights about engineering solutions.

The project consists of three main elements. The first element is **classification**, where you will apply classical machine learning algorithms such as decision trees, logistic regression, support vector machines, and k-nearest neighbors to categorize surface patterns based on extracted features. These features may include texture descriptors, geometric metrics, or other quantitative characteristics derived from the dataset. Your goal is to train, validate, and compare classifiers, and evaluate their performance using accuracy, precision, recall, and confusion matrices.

The second element is **prediction**, which estimates a continuous property of the surface, such as fractal dimension or roughness index, from geometric and texture features. You will apply regression methods such as linear regression, polynomial regression, ridge regression, lasso regression, and support vector regression. Model performance will be evaluated using statistical measures such as R^2, root mean squared error (RMSE), and mean absolute error (MAE).

The third element is **clustering**, in which you will use unsupervised learning algorithms such as k-means, hierarchical clustering, or DBSCAN to group surface structures based on feature similarity. You will assess clustering quality using metrics such as silhouette scores or the Davies-Bouldin index and compare the resulting groupings to the known classifications obtained from the first element. This comparison will help you understand the differences between supervised and unsupervised learning outcomes in practical applications.

You may acquire your dataset in one of three ways. First, you can use a publicly available dataset from sources such as Kaggle, the UCI Machine Learning Repository, or the Open Images Dataset. Search terms such as texture, leaf, or surface patterns may be useful in finding relevant data. Second, you may collect your own data by photographing natural objects with visible surface patterns and extracting features using Python libraries such as OpenCV or scikit-image. Third, you may generate synthetic data using fractal pattern generators or texture synthesis tools, which allow you to control specific geometric and texture parameters for experimental purposes. Regardless of your approach, your dataset should contain sufficient variation to allow for meaningful analysis using classification, prediction, and clustering methods.

Bonus points: Although the focus of the course is on classical machine learning, you may optionally incorporate **deep learning methods** such as convolutional neural networks (CNNs) or autoencoders into your project. For example, CNNs can be used for automatic feature extraction from images, while autoencoders can assist with dimensionality reduction or denoising. Students who successfully integrate deep learning concepts into their project will receive bonus points, provided they clearly explain the methods used, compare them with classical approaches, and discuss the benefits and limitations observed.

Flexibility and Student Responsibility: This project description is intended as a *guideline*. Students are free to select *any suitable methods, algorithms, or datasets* as long as the focus remains on the analysis of *natural surface structures* or nature-inspired pattern extraction. You may expand your work beyond the three main elements if you wish, for example by integrating additional feature extraction techniques, applying alternative algorithms, or experimenting with different evaluation metrics. It is your responsibility to explore, test, and implement various coding options using Python and appropriate libraries. Creativity, methodological rigor, and clear documentation will be rewarded.

Your final report (8-12 pages) should include an introduction explaining the motivation for analyzing surface structures, a detailed description of the dataset and features, the methods used in each element, the results obtained, and a discussion comparing the outcomes from classification, prediction, and clustering. If applicable, include an additional section describing any deep learning methods employed and their performance relative to classical approaches.

Examples of Fractals in Nature