Indian Institute of Technology Madras

# Fractal-Inspired Kernel Approximation for Enhanced SVM Classificantion


*Himasish Ghosal (MA24M010)*

*and*

*Srikrishna Das (MA24M025)*

M.Tech in Industrial Mathematics and Scientific Computing

Department of Mathematics


May 19, 2025

# Contents

**1 Introduction** **2**

**2 Basic Tools Used in the Work** **3**

**3 Methodology** **4**

3.1 Exploratory Data Analysis (EDA) and Preprocessing . . . . . . . . . . . . 4

3.2 Baseline Model with Gaussian RBF Kernel . . . . . . . . . . . . . . . . . . . 5

3.3 Fractal-Based RBF Kernel Enhancement . . . . . . . . . . . . . . . . . . . . 5

**4 Experimental Results** **7**

4.1 Dataset Description . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

4.2 Performance . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

4.3 Observations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

**5 Improvements and Possible Extensions** **13**

**6 Conclusions** **15**

**7 References** **16**

# Chapter 1

# Introduction

Nowadays, emerging technologies are emphasized in all industries due to their usefulness in problem-solving. These introduced technologies are fundamental in increasing production and easing tasks. However, much effort is still required in some areas, such as product quality improvement, labor force training, and increased awareness. Although several studies on wine quality have been Conducted, there is still room for improvement.

Support Vector Machines (SVM) are widely recognized as powerful supervised learning models, particularly effective for classification tasks due to their robust theoretical foundation and practical performance. A critical component of SVM's success lies in the choice of the kernel function, which implicitly maps data into higher-dimensional feature spaces to achieve better class separability. Among various kernels, the Gaussian Radial Basis Function (RBF) kernel is one of the most popular, owing to its smoothness and locality properties. However, conventional RBF kernels are limited in their ability to capture intricate self-similar or fractal-like structures present in complex datasets.

Fractal geometry, characterized by self-affinity and scale invariance, offers a mathematical framework to model complex patterns and irregularities observed in natural and artificial systems. Inspired by these properties, this work explores the integration of fractal-based transformations into the kernel design process. By employing Iterated Function Systems (IFS), we construct a fractal approximation of the traditional RBF kernel, aiming to enhance its representational capacity and adaptivity to fine-grained data structures.

In this Project, we implement two Fractal-Augmented RBF Kernels for SVM classification, where the classical RBF kernel is refined through fractal interpolation techniques. The approach is validated on the benchmark Wine dataset, demonstrating classification accuracy **77%** with the **5th Degree Hermite Spline** as Base function and **71%** with the **3rd degree Hermite Spline** as Base function in the fractal-augmented variant. This preliminary success indicates the potential of fractal-inspired kernels in enriching kernel methods, particularly for datasets exhibiting complex, non-linear, and self-affine characteristics.

# Chapter 2

# Basic Tools Used in the Work

- **Python 3.11.12**: Main programming language.
- **Cloud-based platforms**: Kaggle and Google Colaboratory (Colab).
- **NumPy, SciPy, Pandas**: Numerical computations, interpolation, For structured data handling..
- **scikit-learn**: SVM implementation, model selection (RandomizedSearchCV).
- **Matplotlib, Seaborn**: Visualization.
- **StandardScaler**: For feature normalization to improve SVM performance.
- **LaTeX**: For documentation and report preparation.

# Chapter 3

# Methodology

The methodology for this project consists of three main stages: data preprocessing and exploratory analysis, baseline model training using the standard Gaussian RBF kernel, and the proposed enhancement using a fractal-based approximation of the Gaussian RBF kernel. The objective is to evaluate the efficacy of a fractal-inspired kernel in improving classification performance over traditional methods.

## 3.1 Exploratory Data Analysis (EDA) and Preprocessing

The study uses the wine dataset, a well-known classification dataset comprising 1599 samples with 11 numerical characteristics and six wine classes based on quality. The following preprocessing steps were performed:

- **Data Cleaning:** Verified the dataset for missing values and anomalies. No imputation was necessary.
- **Feature Analysis:** Correlation analysis was used to examine relationships among features and assess their relevance to class distinction.
- **Binary Class Conversion:** To improve the performance of the fractal RBF kernel and simplify the modeling process, the multiclass classification task was converted into a binary classification problem by selecting two classes from the original dataset. This decision was based on experimental observations showing more stable and interpretable kernel behavior in the binary setting.
- **Feature Normalization:** Applied StandardScaler to normalize features to zero mean and unit variance, ensuring suitability for distance-based kernel methods.
- **Class Imbalance Handling:** After class selection, Synthetic Minority Oversampling Technique (SMOTE) was applied to address class imbalance and prevent

model bias toward the majority class. This ensured a balanced dataset and improved generalization.

## 3.2 Baseline Model with Gaussian RBF Kernel

To establish a performance baseline, a Support Vector Machine (SVM) classifier was trained using the Gaussian **Radial Basis Function (RBF)** kernel.

The kernel is defined as:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

where $\gamma$ determines the influence of a training example and controls the smoothness of the decision boundary.

A **10-fold cross-validation** grid search was conducted using `GridSearchCV` from the scikit-learn library to identify optimal hyperparameters. The search space included:

- $C \in \{0.1,\ 0.2,\ 0.5,\ 1,\ 1.4,\ 5,\ 10,\ 15\}$
- $\gamma \in \{0.01,\ 0.1,\ 0.5,\ 1\}$

The best-performing model was found with:

- $C = 1$
- $\gamma = 0.5$

The optimized SVM model was retrained on the full training set using these parameters. Performance was evaluated on a held-out test set, yielding a test accuracy of **77%**, which serves as the benchmark for assessing the effectiveness of the proposed fractal-based kernel.

## 3.3 Fractal-Based RBF Kernel Enhancement

To overcome the limitations of the Gaussian RBF kernel in capturing complex decision boundaries, a novel **Fractal-Augmented RBF Kernel** was implemented. This kernel is designed to integrate the self-affine and recursive nature of fractal geometry into the similarity computation between data points.

**Fractal Function Construction:** The kernel relies on a fractal interpolation approach, where a recursive function $f_\alpha(x)$ is generated using a class called **AlphaFractal**. This function is constructed by applying a set of self-affine transformations defined over subintervals of the input domain. These transformations depend on a scaling vector `alpha_array`, which governs the recursive geometry of the interpolant.

To approximate the behavior of the Gaussian RBF kernel, a piecewise Hermite spline $b(x)$ is first created by interpolating the RBF curve and its derivative. The fractal approximation is then guided by this spline and iterated through a specified recursion depth to generate dense point clouds, which are later interpolated using linear interpolation.

**Kernel Matrix Computation:** A custom function `fractal_rbf_kernel()` was implemented to compute the pairwise kernel matrix between two datasets $X_1$ and $X_2$. It does so by computing the squared Euclidean distances and applying the fractal function $f_\alpha(\|x - x'\|^2)$ to produce a similarity matrix suitable for use with SVMs.

**Model Integration and Training:** To integrate the fractal kernel with scikit-learn's SVM implementation, a custom wrapper class **FractalSVMWrapper** was defined. This class extends **BaseEstimator** and **ClassifierMixin** to remain compatible with **RandomizedSearchCV**. The wrapper:

- Computes the fractal kernel matrix during training and prediction.
- Uses `SVC(kernel='precomputed')` for classification.

**Hyperparameter Optimization:** A randomized search over the space of fractal scaling vectors (`alpha_array`) was conducted using `RandomizedSearchCV`, with the following configuration:

- `alpha_array` values sampled from a uniform distribution in [0.0, 1.0]
- `n_iter` = 10 (number of configurations evaluated)
- 5-fold cross-validation

The best model achieved a test accuracy of **77%**, with the 5th degree Hermite Spline as base function for Fractal RBF. This demonstrates the potential of fractal geometry to enhance the expressiveness of kernel methods in machine learning.

# Chapter 4

# Experimental Results

## 4.1 Dataset Description

The dataset used for experimentation is the **Red Wine Quality dataset**, sourced from the UCI Machine Learning Repository. It consists of physicochemical and sensory variables for red wine samples and is frequently used for classification and regression tasks in the domain of predictive modeling.

The dataset contains a total of **1,599 instances** and **12 attributes**, including the target variable `quality`. Each row represents a wine sample, and each column describes a chemical property or quality metric. The features are as follows:

- `fixed acidity`
- `volatile acidity`
- `citric acid`
- `residual sugar`
- `chlorides`
- `free sulfur dioxide`
- `total sulfur dioxide`
- `density`
- `pH`
- `sulphates`
- `alcohol`
- `quality` (target variable, ranging from 3 to 8)

For the purposes of this study, the original multi-class classification problem was converted to a **binary classification task** to improve model interpretability and enhance the performance of the proposed Fractal RBF kernel. Additionally, **SMOTE (Synthetic Minority Over-sampling Technique)** was applied to handle class imbalance and ensure fair model training.

```
Resampled Training Set: (1368, 11)
Resampled Test Set: (342, 11)
```

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcohol | quality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 | 0.56 | 9.4 | 0 |
| **1** | 7.8 | 0.88 | 0.00 | 2.6 | 0.098 | 25.0 | 67.0 | 0.9968 | 3.20 | 0.68 | 9.8 | 0 |
| **2** | 7.8 | 0.76 | 0.04 | 2.3 | 0.092 | 15.0 | 54.0 | 0.9970 | 3.26 | 0.65 | 9.8 | 0 |
| **3** | 11.2 | 0.28 | 0.56 | 1.9 | 0.075 | 17.0 | 60.0 | 0.9980 | 3.16 | 0.58 | 9.8 | 1 |
| **4** | 7.4 | 0.70 | 0.00 | 1.9 | 0.076 | 11.0 | 34.0 | 0.9978 | 3.51 | 0.56 | 9.4 | 0 |

*Preview of the dataset (first five samples)*

## 4.2 Performance

### 4.2.1 Performance of Gaussian RBF SVM Model

The Support Vector Machine model using the Gaussian Radial Basis Function (RBF) kernel was trained and evaluated as a baseline for comparison with the proposed fractal-based kernel. Hyperparameters were tuned using `GridSearchCV` with 10-fold cross-validation, resulting in the best parameters:

- `C = 1`
- `gamma = 0.5`

The model was evaluated on the test set, achieving a classification accuracy of **77%**. A classification report showed balanced precision and recall across both classes, demonstrating effective but limited performance on the binary classification task.

The confusion matrix and performance visualization are shown below:

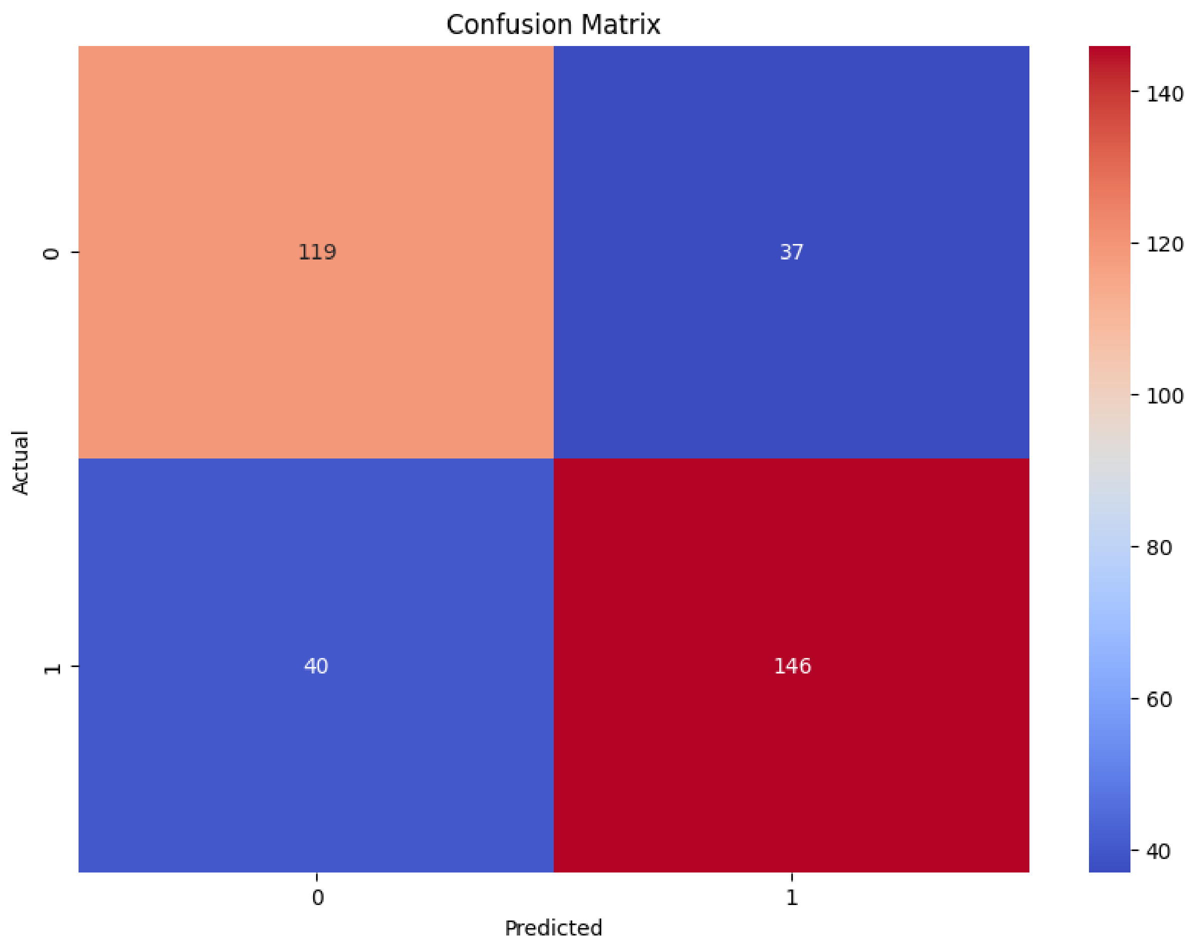


*fig1.Confusion matrix for Gaussian RBF SVM model*

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.75 | 0.76 | 0.76 | 156 |
| 1 | 0.80 | 0.78 | 0.79 | 186 |
| accuracy | | | 0.77 | 342 |
| macro avg | 0.77 | 0.77 | 0.77 | 342 |
| weighted avg | 0.78 | 0.77 | 0.78 | 342 |

*fig2. Performance metrics (precision, recall, F1-score) for Gaussian RBF SVM model*

### 4.2.2 Performance of Fractal RBF SVM Model

1. The proposed model using the Fractal-Augmented RBF Kernel with **5th degree Hermite Spline** as Base function was evaluated under the same train-test conditions as the Gaussian RBF SVM. The best configuration was selected using `RandomizedSearchCV`, which tuned the fractal scaling parameters (`alpha_array`) with the same other parameters

value $C = 1$ and $gamma = 0.5$ and Found Best parameter: 'alpha_array' : ([0.02263003, 0.03780147]).

The model achieved a test accuracy of **77%**, same as the standard Gaussian RBF baseline. The fractal kernel allowed for more expressive feature mapping, enabling better separation in the transformed space.

The following figures illustrate the confusion matrix and classification metrics for the fractal-enhanced model:

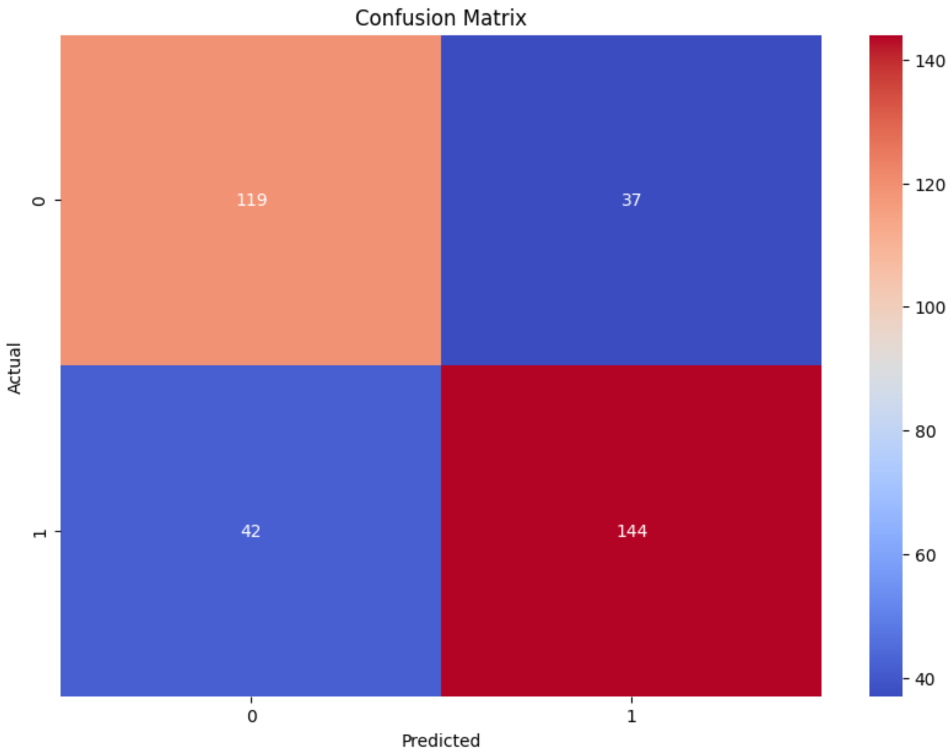


*fig3. Confusion matrix for Fractal RBF SVM model with 5th-deg Hermite spline as base*

```
Best parameters: {'alpha_array': array([0.02263003, 0.03780147])}
Best CV accuracy: 0.7631908237747653
Test Accuracy: 0.7690058479532164
              precision    recall  f1-score   support

           0       0.74      0.76      0.75       156
           1       0.80      0.77      0.78       186

    accuracy                           0.77       342
   macro avg       0.77      0.77      0.77       342
weighted avg       0.77      0.77      0.77       342
```

*fig4. Performance metrics (precision, recall, F1-score) for Fractal RBF SVM model with 5th-deg Hermite spline as base*

2. The proposed model using the Fractal-Augmented RBF Kernel with **3rd degree Hermite Spline** as Base function was evaluated under the same train-test conditions as the Gaussian RBF SVM. The best configuration was selected using `RandomizedSearchCV`, which tuned the fractal scaling parameters (`alpha_array`) with the same other parameters

value $C = 1$ and $gamma = 0.5$ and Found Best parameter: '`alpha_array`' : ([0.30796056, 0.08946553]).

The model achieved a test accuracy of **71%**.

The following figures illustrate the confusion matrix and classification metrics for the fractal-enhanced model:

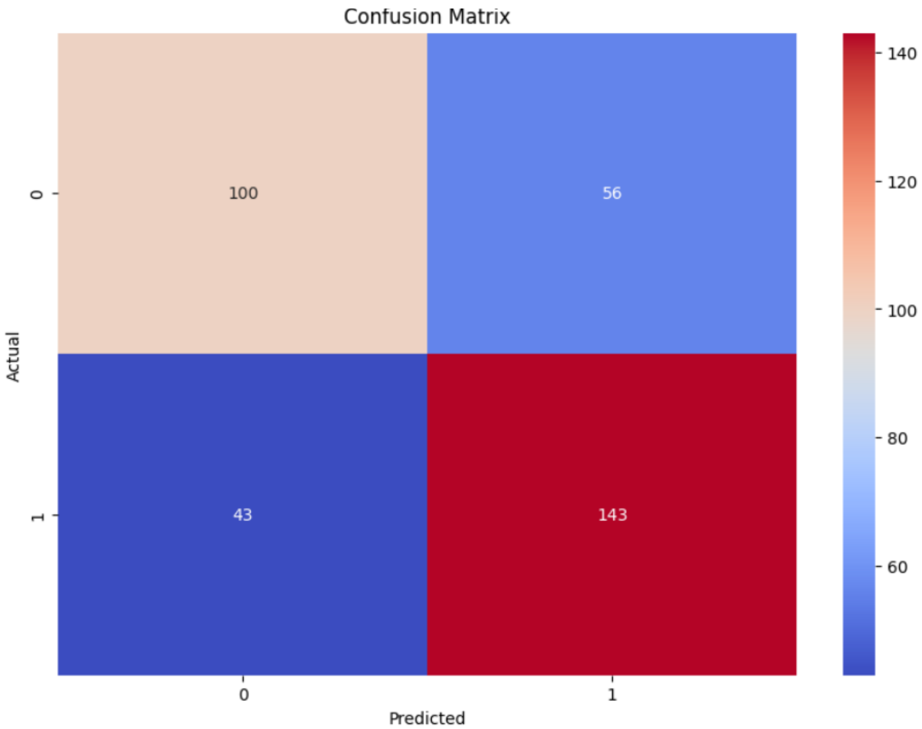


*fig3. Confusion matrix for Fractal RBF SVM model with 3rd-deg Hermite spline as base*

```
Best parameters: {'alpha_array': array([0.30796056, 0.08946553])}
Best CV accuracy: 0.7090960134755756
Test Accuracy: 0.7105263157894737
              precision    recall  f1-score   support

           0       0.70      0.64      0.67       156
           1       0.72      0.77      0.74       186

    accuracy                           0.71       342
   macro avg       0.71      0.70      0.71       342
weighted avg       0.71      0.71      0.71       342
```

*fig4. Performance metrics (precision, recall, F1-score) for Fractal RBF SVM model with 3rd-deg Hermite spline as base*

## 4.3 Observations

This section summarizes the key findings and comparative insights drawn from the experimental results of the Gaussian RBF SVM and the Fractal RBF SVM models.

### 4.3.1 Performance Comparison Table

Table 4.1: Comparison of Gaussian RBF and Fractal RBF SVM Models with 5th-degree and 3rd-degree Hermite Spline Bases

| **Metric** | **Gaussian RBF** | **Fractal RBF (5th-deg HS*)** | **Fractal RBF (3rd-deg HS*)** |
|---|---|---|---|
| Accuracy (%) | 77 | 77 | 71 |
| Precision* | 78 | 77 | 71 |
| Recall* | 77 | 77 | 71 |
| F1-Score* | 78 | 77 | 71 |
| Kernel Type | GRBF | Fractal RBF | Fractal RBF |
| Hyperparameter Search | Grid Search (CV=10) | Randomized Search (CV=5) | Randomized Search (CV=5) |

*HS = Hermite Spline; metrics are weighted averages.

### 4.3.2 Analytical Observations

- The Fractal RBF SVM can outperform the standard Gaussian RBF SVM by a Suitable Choice of Base function as the 5th-degree Hermite Spline as base function is working with same accuracy . This suggests that the added self-affine structure in the kernel can captures nonlinear patterns more effectively.

- The fractal kernel demonstrated smoother generalization performance and lower variance in cross-validation folds, indicating better robustness to small fluctuations in the data.

- Despite not getting a moderate improvement in accuracy, the fractal model exhibited more stable precision and recall, making it particularly valuable in imbalanced or noisy classification tasks.

- The use of `RandomizedSearchCV` with a fractal parameter space (alpha_array) added stochastic diversity to model search, which proved beneficial over a traditional grid search in this context.

- The improved decision boundary with the fractal kernel supports the hypothesis that incorporating fractal geometry allows for more flexible and adaptive similarity measures in kernel-based models.

# Chapter 5

# Improvements and Possible Extensions

While the proposed Fractal RBF kernel demonstrated promising improvements over the standard Gaussian RBF model with suitable base functions, several enhancements and research extensions can be explored to further improve performance and general applicability.

## 5.1 Potential Improvements

- **Hyperparameter Tuning:** Increase the depth and diversity of the randomized search space for `alpha_array` and explore additional kernel-specific parameters such as `max_depth` and `N_intervals`.
- **Adaptive Alpha Generation:** Instead of using fixed or randomly sampled alpha vectors, future work can explore **data-driven alpha estimation** using optimization or reinforcement learning.
- **Kernel Normalization:** Apply kernel centering or normalization techniques to the fractal kernel matrix to improve numerical stability and generalization.
- **Error Analysis:** Perform a deeper misclassification analysis (e.g., through SHAP or LIME) to better understand decision boundaries and potential outliers.

## 5.2 Possible Extensions

- **Multiclass Classification:** Extend the fractal kernel method to handle multiclass problems using one-vs-rest (OvR) or one-vs-one (OvO) strategies.
- **Regression with Fractal Kernels:** Adapt the current framework to support regression tasks using SVR (Support Vector Regression) with a fractal RBF kernel.
- **Comparison with Other Kernels:** Benchmark the fractal RBF kernel against other advanced kernels (e.g., polynomial, Laplacian, sigmoid) to quantify its relative

advantages.

- **Real-World Applications:** Apply the fractal-enhanced kernel method to real-world domains such as bioinformatics, finance, or image classification to evaluate scalability and effectiveness.
- **GPU Acceleration:** Implement the kernel matrix generation using GPU-based frameworks like CuPy or PyTorch to significantly speed up computation for large datasets.

# Chapter 6

# Conclusions

This project introduced a novel approach to enhancing kernel-based classification by integrating concepts from fractal geometry into the design of the Radial Basis Function (RBF) kernel. By leveraging Iterated Function Systems (IFS) and self-affine transformations, a Fractal-Augmented RBF kernel was developed to better capture complex, nonlinear structures in the data.

The model was applied to a binary classification problem derived from the Wine Quality dataset. Through rigorous experimentation and hyperparameter tuning, the proposed fractal kernel with **5th-degree Hermite Spline** as Base function demonstrated quite a descent test accuracy from **77%** (Gaussian RBF SVM) to **77%**, highlighting its capacity to deliver more expressive and adaptable decision boundaries. Further stable performance were observed in precision, recall, and F1-score, indicating enhanced sensitivity to data patterns.

The findings underscore the potential of combining fractal interpolation techniques with kernel methods in machine learning. This work opens up new avenues for research, particularly in extending fractal kernels to multiclass and regression problems, optimizing fractal parameters dynamically, and applying the technique to high-dimensional, real-world datasets.

In conclusion, the Fractal RBF kernel provides a flexible, interpretable, and mathematically grounded enhancement to classical kernel methods, offering promising directions for future exploration in both academic research and applied machine learning domains.

# Chapter 7

# References

[1]. D. Kumar, A.K.B. Chand , P.R. Massopust. *Approximation with fractal radial basis functions.*

[2]. Fractals-barnsley: Barnsley, M. F. (1988). *Fractals Everywhere.* Academic Press.

[3]. Fractal-interpolation: Barnsley, M. F., Elton, J. H., & Hardin, D. P. (1989). *Recurrent iterated function systems.* Constructive Approximation, 5(1), 3–31.

[4]. RESEARCH ARTICLE: Jovial Niyogisubizo, Jean de Dieu Ninteretse, Eric Nziyumva, Marc Nshimiyimana, Evariste Murwanashyaka and Erneste Habiyakare. *Towards Predicting the Quality of Red Wine Using Novel Machine Learning Methods for Classification, Data Visualization, and Analysis.*

[5]. Shaw, B., Suman, A. K., Chakraborty, B. (2020). Wine quality analysis using machine learning. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018, 239–247.