**Unsupervised Learning Methods for Functional Data with Missing Observations:

An Application to Alzheimer’s Disease Analysis**

Abstract

Alzheimer’s disease is a progressive neurodegenerative disorder that significantly affects cognitive functions and quality of life. Early identification of disease patterns is crucial for timely intervention and improved patient management. However, real-world biomedical data—particularly functional data—often suffer from missing observations, irregular sampling, and high dimensionality. These challenges limit the effectiveness of traditional supervised learning approaches.

In this project, we investigate **unsupervised learning methods for functional data with missing observations**, focusing on Alzheimer’s disease–related datasets. The proposed framework applies advanced preprocessing techniques, functional data representation, and unsupervised clustering methods to discover latent structures and meaningful disease-related patterns without relying on labeled data. Experimental analysis demonstrates that unsupervised approaches can effectively capture intrinsic variations in functional signals, even in the presence of missing values. The results indicate that this methodology can support exploratory medical analysis and contribute to early-stage disease understanding.

---

1. Introduction

Alzheimer’s disease (AD) is one of the most prevalent neurological disorders worldwide and poses a growing challenge to healthcare systems. The disease progression is gradual and complex, often characterized by subtle functional changes that are difficult to detect at early stages.

Functional data—such as longitudinal cognitive scores, neuroimaging-derived signals, or time-dependent clinical measurements—provide rich information about disease evolution. However, such data frequently contain **missing observations** due to patient dropout, sensor failure, or irregular clinical visits. Conventional machine learning techniques often struggle with incomplete functional datasets.

To address these limitations, this project explores **unsupervised learning techniques** that do not require labeled outcomes and can robustly operate on incomplete functional data. By focusing on Alzheimer’s disease, this work aims to demonstrate how unsupervised methods can reveal hidden structures, patient subgroups, and disease-related patterns that are otherwise difficult to observe.

---

2. Related Work

Previous research in Alzheimer’s disease analysis has largely relied on supervised learning approaches, including classification and regression models, which require labeled datasets. While effective, these methods depend heavily on annotation quality and data completeness.

Recent studies have emphasized:

- Functional data analysis (FDA) for longitudinal medical data
- Missing data imputation techniques
- Clustering-based disease subtyping
- Dimensionality reduction for high-dimensional biomedical signals

However, fewer works focus on combining **functional data modeling**, **missing observation handling**, and **unsupervised learning** within a unified framework. This project addresses this gap by emphasizing exploratory analysis and structure discovery rather than prediction alone.

---

3. Dataset and Problem Formulation

3.1 Functional Data Description

The dataset used in this project consists of functional observations related to Alzheimer’s disease. Each subject is represented by time-dependent or curve-based measurements that reflect cognitive or neurological changes over time.

3.2 Missing Observations

Missing values arise due to:

- Irregular patient follow-ups
- Data acquisition errors
- Incomplete clinical records

The presence of missing data necessitates robust preprocessing and modeling strategies that preserve underlying functional characteristics.

---

4. Methodology

The proposed framework consists of the following stages:

4.1 Data Preprocessing

- Handling missing observations using interpolation or smoothing techniques
- Normalization and alignment of functional curves
- Removal of noisy or inconsistent measurements

4.2 Functional Representation

Functional data are represented using basis functions or smooth approximations to transform discrete observations into continuous functional forms.

4.3 Unsupervised Learning Techniques

The following unsupervised methods are applied:

- Functional clustering to identify patient subgroups
- Distance-based similarity measures adapted for incomplete functional data
- Dimensionality reduction techniques to visualize latent structures

4.4 Pattern Discovery

Clusters and latent components are analyzed to interpret disease-related trends and variations among subjects.

---

5. Experimental Results and Analysis

The experimental evaluation demonstrates that:

- Unsupervised learning methods successfully group subjects with similar functional patterns
- Meaningful clusters emerge despite missing observations
- Identified structures align with expected disease progression behaviors

Visual analysis and clustering outcomes indicate that the proposed approach is effective for exploratory Alzheimer’s disease analysis.

---

6. Discussion

The results highlight the potential of unsupervised learning in medical data analysis, particularly when labeled data are unavailable or incomplete. By focusing on functional data, this approach captures temporal dynamics that are often ignored in traditional feature-based models.

The ability to handle missing observations makes the framework suitable for real-world clinical datasets, where perfect data completeness is rarely achievable.

---

7. Limitations and Future Work

Although the proposed framework performs effectively, several limitations remain:

- Sensitivity to preprocessing and smoothing choices
- Scalability to very large datasets
- Lack of direct clinical validation

Future work may include:

- Integration of deep unsupervised models
- Advanced imputation techniques
- Hybrid semi-supervised approaches
- Validation with real clinical outcomes

---

8. Conclusion

This project presents an unsupervised learning framework for functional data with missing observations, with a specific application to Alzheimer’s disease analysis. The results demonstrate that meaningful patterns and patient subgroups can be identified without labeled data, even in the presence of incomplete observations.

The proposed methodology offers a valuable tool for exploratory medical research and has the potential to support early disease understanding and decision-making in healthcare analytics.