

Case Study: Rare Genetics

Medical Data Science Application

Learn about the core concepts of **case studies** and their application in medical data science.



Key Concepts

Theoretical background and core principles of case studies



Medical Application

Real-world use cases in medical environments



Performance Evaluation

Effectiveness measurement and improvement methodology



Practical Examples

Implementation using real-world data

Key Points

- Approach considering the specificity of medical data
- Efficient learning in limited data environments

- Design ensuring safety and reliability

Rare Genetics Case Study Overview

Approximately 7,000 rare genetic diseases are known worldwide, and while individual disease incidence rates are low, they collectively affect about 6-8% of the population, making them an important public health issue. This case study covers data science approaches for diagnosing and treating rare genetic diseases, and teaches methods for building effective analysis and predictive models in limited data environments.

Real-World Application Examples

For certain rare genetic diseases, there may be only a few hundred patients worldwide. In such situations, traditional large-scale data-driven machine learning approaches are difficult to apply. Therefore, innovative approaches using transfer learning, meta-learning, and few-shot learning techniques are necessary.

Research Methodology

1. Data Collection & Integration

Collect and integrate information from various sources (genomic data, clinical records, imaging data) to construct comprehensive patient profiles.

2. Feature Extraction & Selection

Identify key biomarkers and genetic variants related to diseases from high-dimensional genomic data and apply dimensionality reduction techniques.

3. Model Development & Validation

Use ensemble techniques and regularization methods suitable for small datasets to prevent overfitting and

4. Clinical Application & Evaluation

Validate developed models in actual clinical settings and evaluate performance using metrics such as sensitivity,

improve generalization performance.

specificity, and positive predictive value.

Learning Objectives & Expected Outcomes

Through this case study, learners will understand the unique challenges of rare disease research and develop the ability to perform effective analysis even in limited data situations. In particular, the following practical skills can be cultivated:

- ▶ **Data Augmentation Techniques:** Methods to effectively expand training data from limited samples
- ▶ **Domain Knowledge Integration:** Strategies to incorporate biological and clinical knowledge into models
- ▶ **Ensuring Interpretability:** Building explainable AI models that clinicians can trust
- ▶ **Ethical Considerations:** Compliance with patient privacy protection and data security

Future Outlook

The use of AI and data science in the field of rare genetic diseases continues to expand. Through data sharing via multi-institutional collaborative networks, advances in synthetic data generation technology, and the introduction of privacy-preserving technologies such as federated learning, accurate diagnosis and personalized treatment will become available to more patients.