

Proposal for Implementing "Multilabel Classification with Group Testing and Codes"

Group Members: Suryansh Patidar - 22B1036
Krish Rakholiya - 22B0927
Omkar Ramesh Shirpure - 22B0910

1 Paper Selected

"Multilabel Classification with Group Testing and Codes" by *Shashanka Ubaru and Arya Mazumdar*

2 Objective

The goal of this project is to implement and evaluate the group testing (GT)-based multilabel classification (MLC) approach proposed in the paper. This method efficiently reduces the label space using binary matrix transformations and leverages error correction techniques to improve classification accuracy.

3 Datasets

We will use publicly available multilabel datasets from *The Extreme Classification Repository*, such as:

- **RCV1-2K**: A text classification dataset with sparse labels.
- **EURLex-4K**: A legal document dataset with multilabel annotations.
- **AmazonCat-13K**: A large-scale product categorization dataset.

4 Implementation Plan

1. **Preprocessing**: Load datasets and convert label vectors into sparse binary representations.
2. **Group Testing Matrix Construction**: Generate binary matrices using *concatenated Reed-Solomon codes* or *expander graphs*.
3. **Encoding Labels**: Use Boolean OR operation to reduce label dimensions.
4. **Training**: Train m binary classifiers using the transformed labels.
5. **Prediction & Decoding**: Predict reduced labels and recover full label vectors using the proposed decoding algorithm.

5 Evaluation Strategy

To validate our implementation, we will compare our results against baseline methods such as:

- **Compressed Sensing (CS)-based MLC**
- **Embedding-based approaches (e.g., SLEEC)**

Metrics for evaluation:

- **Hamming Loss** (Lower is better)
- **Precision@k** (Higher is better)
- **Computational Efficiency** (Training & inference time)

6 Expected Outcome

We aim to demonstrate that the GT-based approach achieves **lower prediction error and better scalability** compared to other multilabel classification techniques.