Reproducible Machine Learning Results' Checklist in Bioinformatics

Hamidreza Heidarzadeh, Tristan Glatard

Department of Computer Science and Software Engineering

Concordia University, Montréal, Québec, Canada

1 Abstract

1.1 Background

1.2 Results

1.3 Conclusion

2 Introduction

3 Materials and Methods

3.1 Dataset

3.2 Features

3.3 Models

3.3.1 Data Preprocessing

3.3.1.1 5-Fold Cross-Validation on a shuffled dataset

3.3.1.2 5-Fold Cross-Validation out of each class

3.3.1.3 Subsampling

3.3.2 The Model

3.3.2.1 7-class-based model

3.3.2.2 8-class-based model

3.3.3 The Algorithm(SVM)

3.3.3.1 Gamma and Cost

3.3.3.2 Thresholding

3.4 Models' Evaluation

3.4.1 Metrics from the paper

3.4.2 Micro vs Macro

4 Results

5 Discussion

6 Conclusion

References