Anticipating Spectrogram Classification Error With Combinatorial Coverage Metrics

2023 IEEE Cognitive Communications for Aerospace Applications Workshop (CCAAW)

Tyler Cody, Ph.D.

Laura Freeman, Ph.D.

Virginia Tech National Security Institute

Academic Mission

- I am a systems theorist and machine learning engineer.
- I develop and apply systems theory to bridge systems engineering and artificial intelligence.

My mission is to develop rigorous first-principles for engineering AI-heavy systems.

"In order to improve your game, you must study the endgame before everything else, for whereas the endings can be studied and mastered by themselves, the middle and the opening must be studied in relation to the endgame."

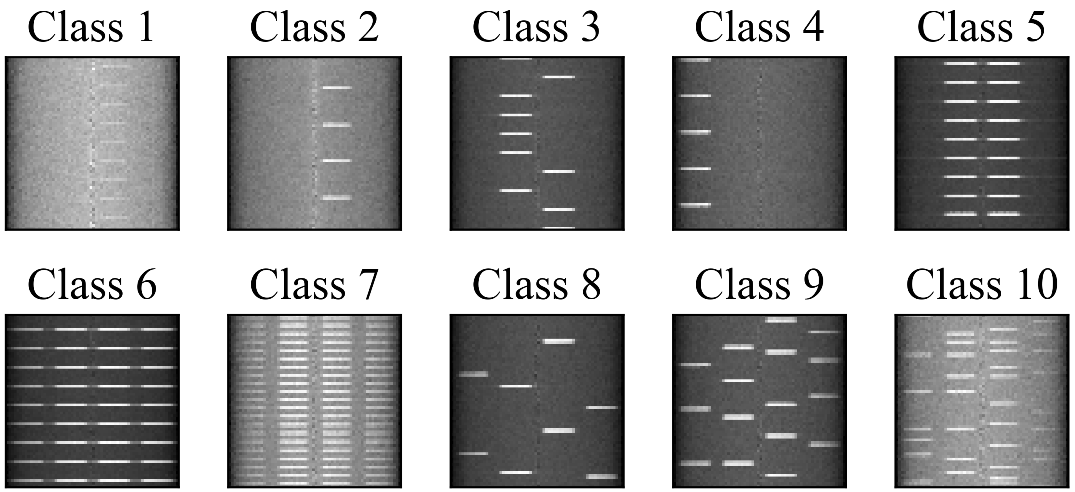
- Jose Raul Capablanca

Executive Summary

Combinatorial coverage metrics have a high correlation with accuracy on batches of spectrograms, therefore they can be used to anticipate error in spectrogram classification.

Deep Spectrogram Classification

- Variety of applications in literature including to coexist with and discern primary user, spectrum sharing scenarios, spectrum occupancy reconstruction
- Also there exists advanced deep architectures for supervised problem



F. Paisana, A. Selim, M. Kist, P. Alvarez, J. Tallon, C. Bluemm, A. Puschmann, and L. DaSilva, "Context-aware cognitive radio using deep learning," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2017, pp. 1–2.

Preliminaries on Coverage

Definition 1. t-way Combinatorial Coverage.
Consider a universe with k factors such that $\mathcal{U}$ is the set of all valid k-way value combinations. Let $\mathcal{U}^t$ be the set of valid t-way combinations. Given a set of data $D \subseteq \mathcal{U}$, let D^t define the set of t-way value combinations appearing in D. The t-way combinatorial coverage of D is

$$CC^t(D) = \frac{|D^t|}{|\mathcal{U}^t|},$$

where $|D|$ denotes the cardinality of D.

SDCC is alternatively measures the proportion of valid t-way value combinations that appear in one set relative to another set. SDCC is defined in the following [12].

Preliminaries on Coverage

Definition 2. t-way Set Difference Combinatorial Coverage. *Let D_A and D_B be sets of data, and ${D_A}^t$ and ${D_B}^t$ be the corresponding t-way sets of data. The set difference ${D_B}^t \backslash {D_A}^t$ gives the value combinations that are in ${D_B}^t$ but that are not in ${D_A}^t$. The t-way set difference combinatorial coverage is*

$$SDCC^t(D_B, D_A) = \frac{|{D_B}^t \setminus {D_A}^t|}{|{D_B}^t|}.$$

Method

We propose a method to learn a latent space for spectrograms that is meaningful for coverage analysis of classification accuracy. Our method consists of three steps.

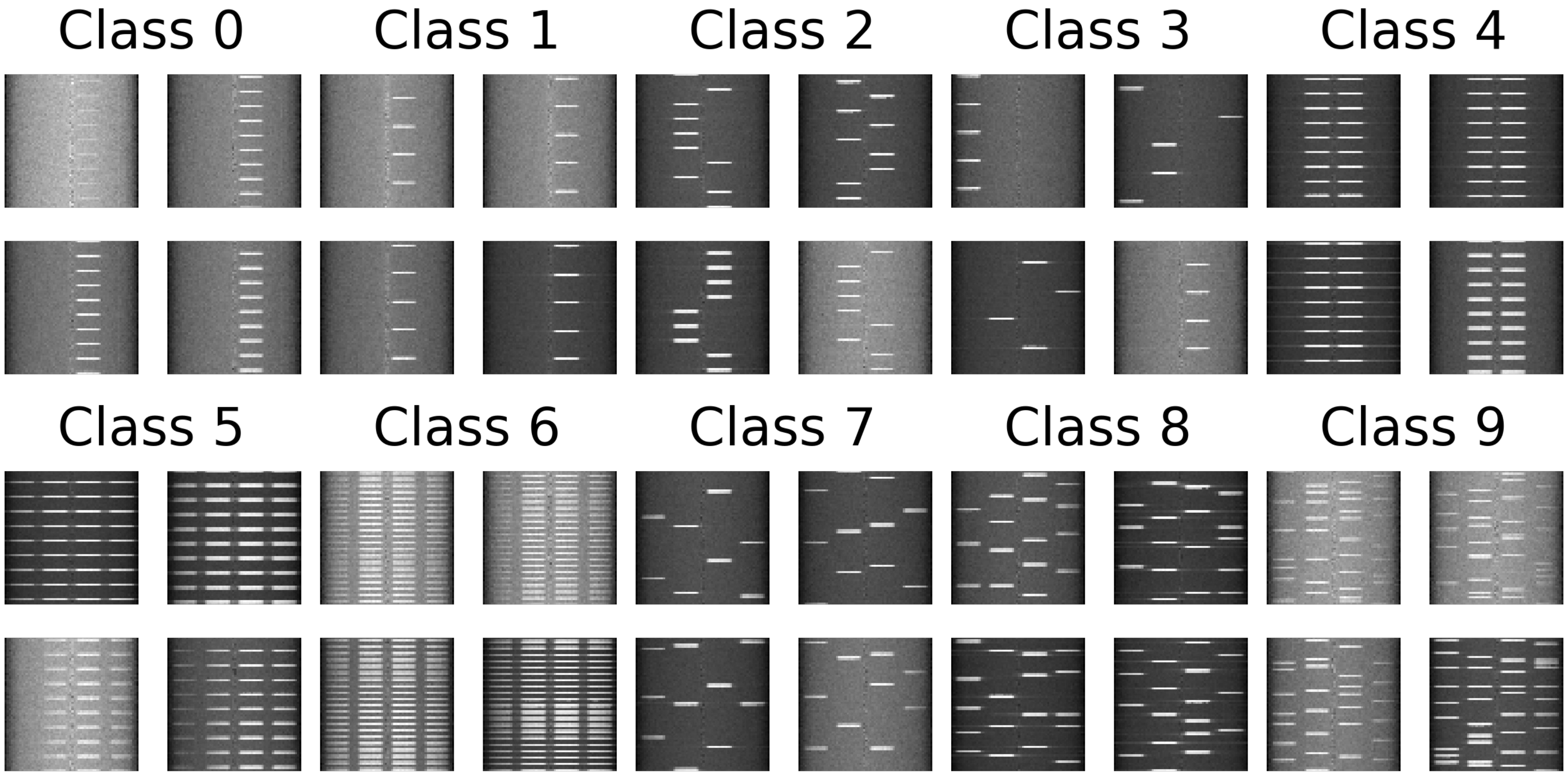
1) First, use a deep autoencoder to encode spectrograms from pixel-space to a lower-dimensional representation.
2) Second, use metric learning techniques for dimension reduction to further reduce the dimensionality of the encoded spectrograms.
3) Last, to discretize the data, apply k-means clustering to the metric learning spaces and to the reconstruction error of the autoencoder.

Metric Learning Methods

- We use PCA, NCA, and LMNN
  - PCA is principal component analysis
  - NCA is a variation on k-nearest neighbors (KNN) classification that directly maximizes a variant of leave-one-out performance
  - LMNN is a variation on KNN classification that learns a Mahalanobis distance

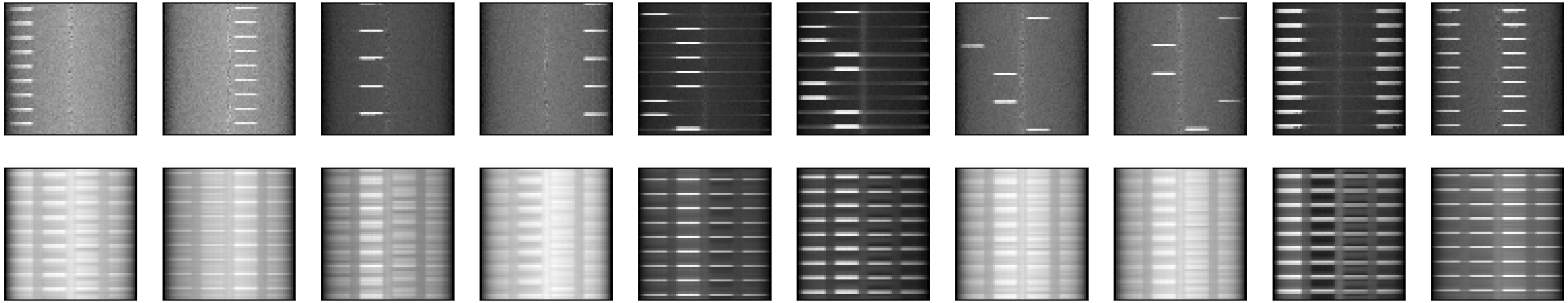
Cody, Tyler, and Laura Freeman. "Metric Learning Improves the Ability of Combinatorial Coverage Metrics to Anticipate Classification Error." *2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. IEEE, 2023.

Data

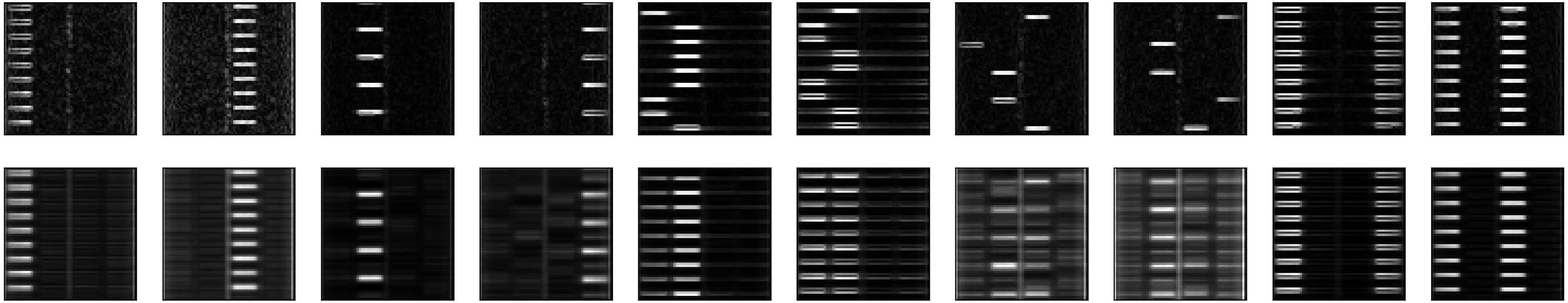

Class 0
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9

Encoding

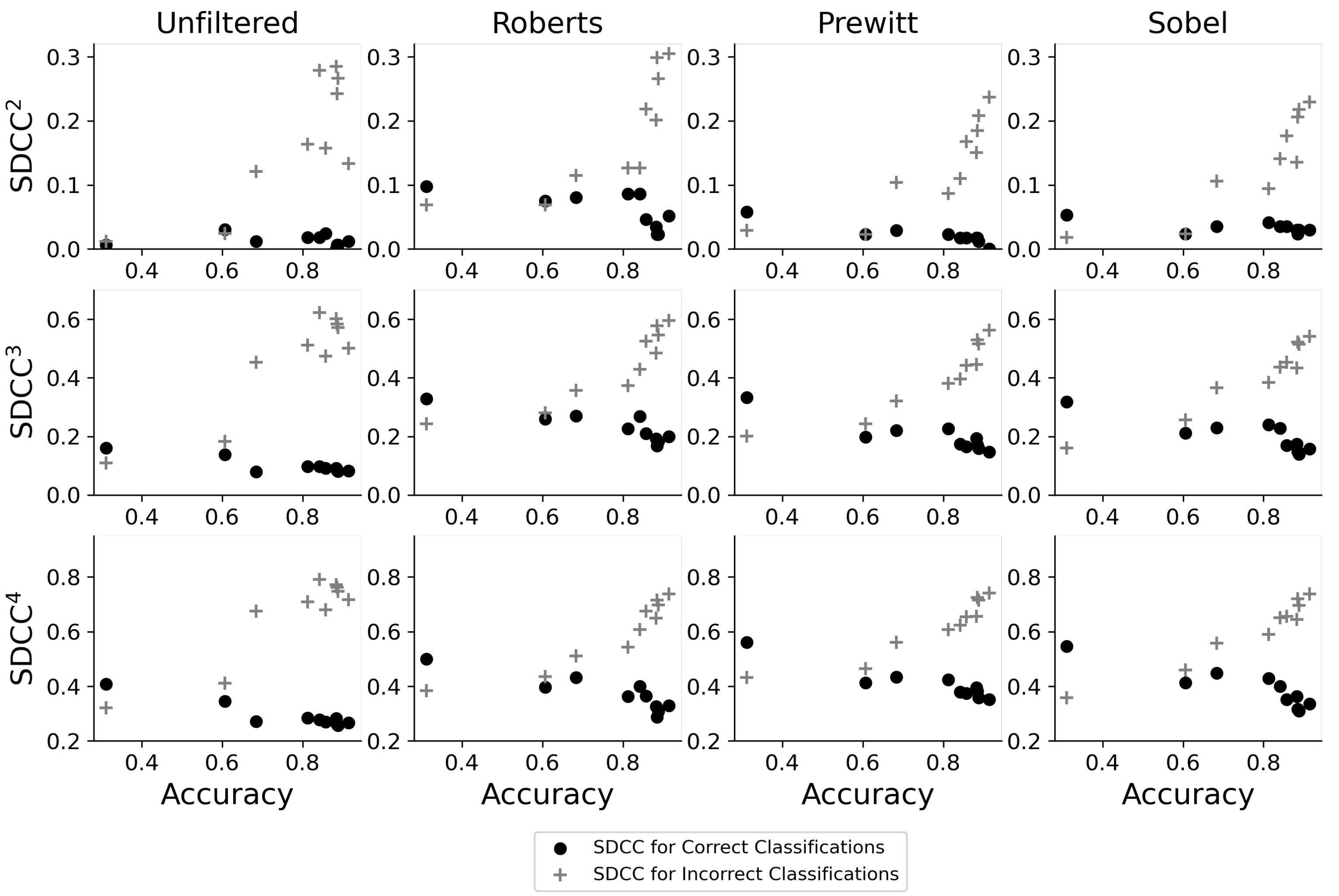
Without filter



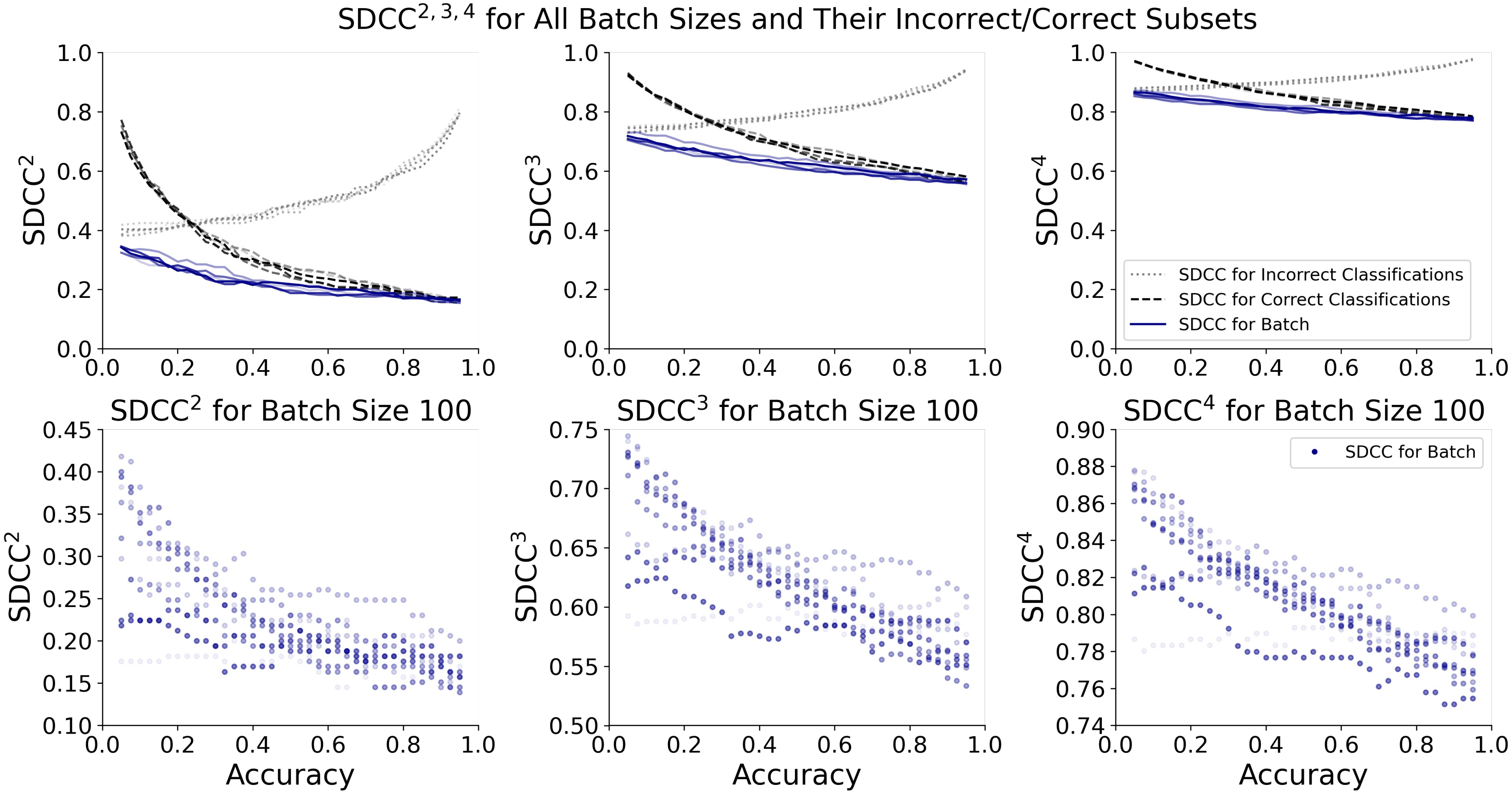
With filter



Cross-Validation Results



Batch Results Plot



Batch Results Table

TABLE I
MEAN AND STANDARD DEVIATION OF PEARSON CORRELATIONS BETWEEN ACCURACY AND SDCC

Batch Size	t=2	t=3	t=4
100	(-0.87, 0.16)	(-0.94, 0.03)	(-0.95, 0.04)
200	(-0.84, 0.28)	(-0.80, 0.58)	(-0.77, 0.58)
300	(-0.93, 0.02)	(-0.95, 0.03)	(-0.93, 0.08)
400	(-0.92, 0.07)	(-0.96, 0.04)	(-0.97, 0.02)
500	(-0.90, 0.11)	(-0.94, 0.07)	(-0.95, 0.05)

Acknowledgements

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) under Contract HQ003419D0003. The Systems Engineering Research Center (SERC) is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology. Any views, opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense nor ASD(R&E).

Contact

Tyler Cody, Ph.D.

Research Assistant Professor

Intelligent Systems Division, Virginia Tech National Security Institute

tcody@vt.edu

LinkedIn, ResearchGate as “Tyler Cody”