Developing Natural Language Processing and Supervised Learning Techniques to Classify Mars Tasks

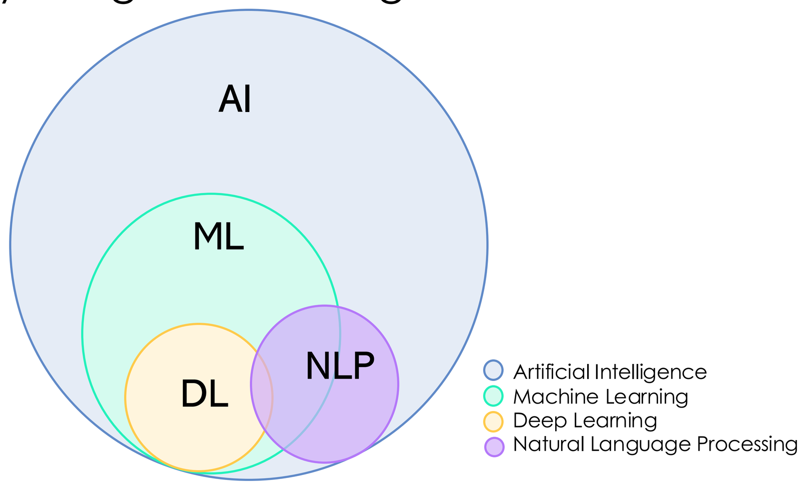
Adam Kurth[1], Hunter Rehm[2] ,Mona Matar[3]

[1]Arizona State University, Tempe, AZ, [2]HX5 LLC., Cleveland, OH, [3]NASA Glenn Research Center, Cleveland, OH

Introduction

In preparation for long-duration Mars missions, understanding astronaut tasks is critical. We employ Natural Language Processing (NLP) and deep learning (DL) to classify and analyze each Mars task, for better insight into their associated physiological and cognitive demands.

__What is NLP and DL?__ NLP is a field of machine learning (ML) concerned with the interaction between computers and human language, used for understanding text data. DL is primarily concerned with learning abstract patterns in the data.



Data

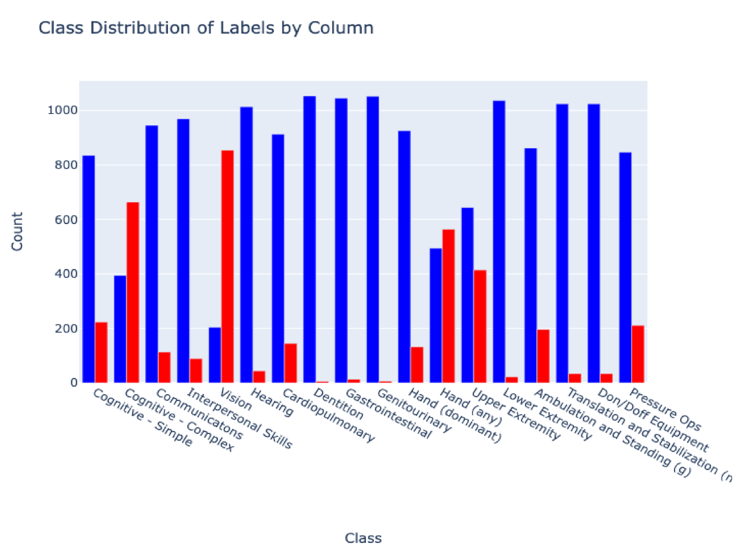
- *__Mars Tasks (MT)__*: 1,058 text descriptions and objectives of specific tasks that astronauts may perform during a Mars mission [1].
  e.g., *"Descend crater wall while carrying hand tools and wearing surface EVA suit to conduct geological research."*
- *__Human System Task Categories (HSTC)__*: 18 categories of human systems with description [2].
  e.g., *Cognition-Simple/Complex, Vision*.

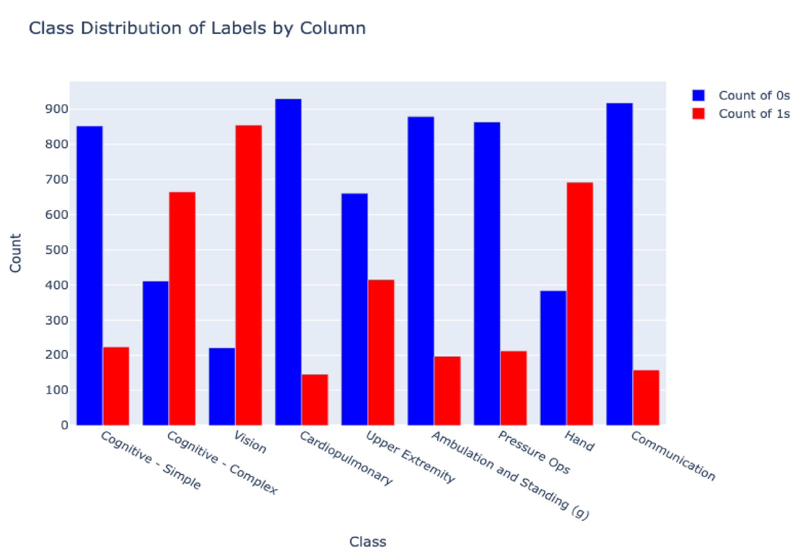
MT are dataset samples, HSTC are features. We trained our model using a binary table where a value of 1 indicates the involvement of an HSTC in the MT, and 0 otherwise. Those tables were generated by the Clinical Science Team from the Science Integration Office.

Challenges

- Severe class imbalance. The number of 0's (majority) and 1's (minority) are vastly unequal.
  Solution: Combine and drop categories with too few examples, resulting in 9/18 HSTC.
- Too little data. DL requires as much data as possible to classify HSTC.
  Solution: implementing an exponential scoring into the preprocessing step to create a sampling distribution for augmentation.

Before — DROP/MELT → After: 9/18 HSTC





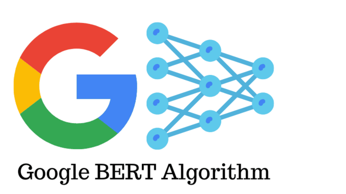
Augmentation

Select minority HSTCs to generate more examples of MT.

- By calculating the frequency, normalizing, and applying an exponential weighting factor to create sampling distribution to augment.
- *Easy Data Augmentation (EDA)* technique is used for efficient augmentation [4][5]. Train, validation, and test splits are correctly separated before augmentation.

Implementation

Google's *Bidirectional Encoder Representation for Transformers* (BERT) is based on the transformer architecture (DL) [4]. The BERT model is used to capture abstract semantic and syntactic uses of the MT [5].



https://creatives.me/2022/09/22/what-is-google-bert/

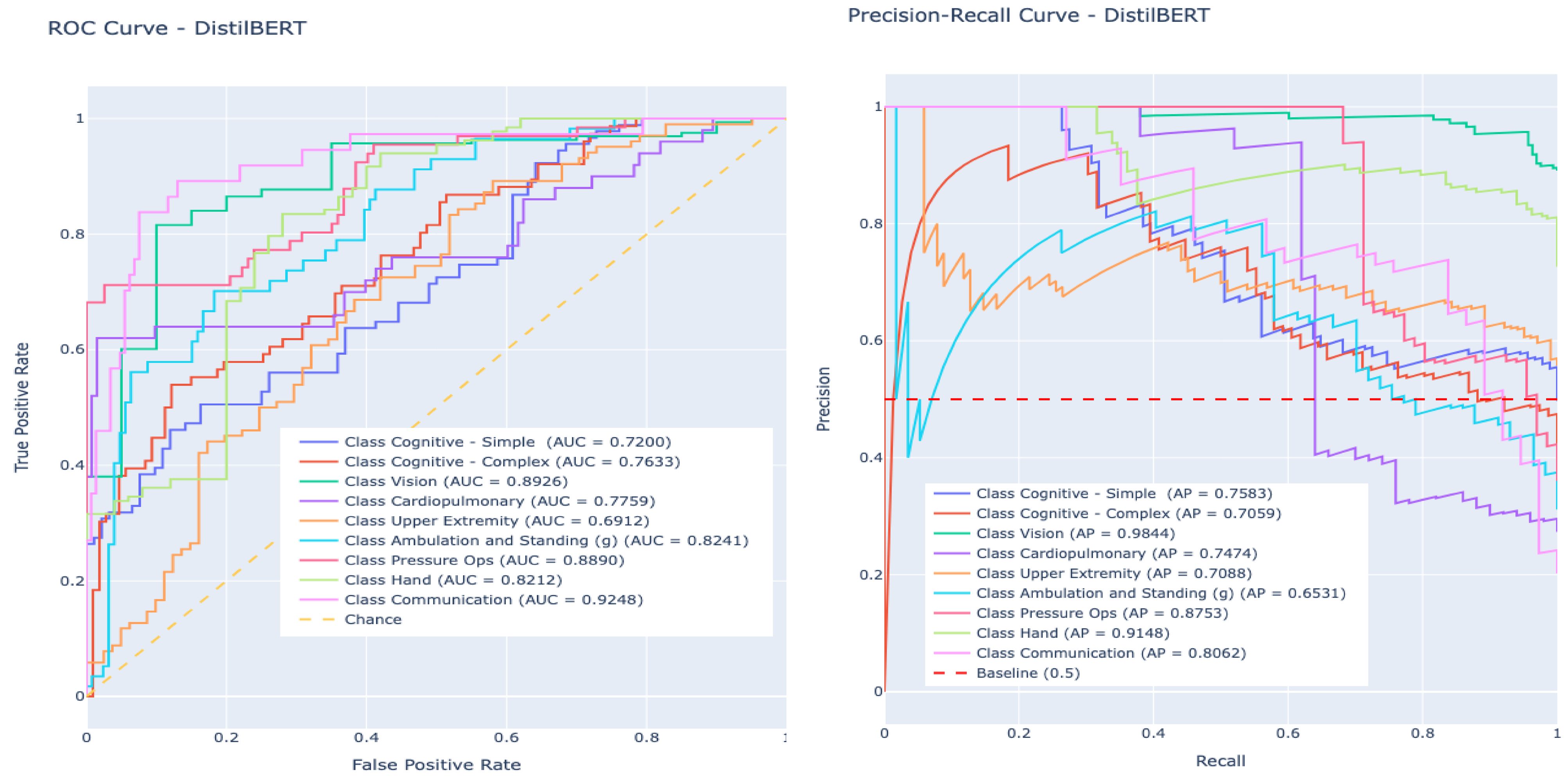
Training

1. Use MT descriptions (only 9/18) and assess model performance (MT only)
2. Use MT and HSTC descriptions and assess performance (MT + HSTC)
3. Preliminary results embedding for unsupervised learning using MT.

Results for Supervised Learning

Results show that LLMs perform best when predicting the involvement of vision, pressure ops, hand, and communication categories in the execution of Mars tasks. That is high AUC in the ROC and PR curves, as well as the F1 score.

Supervised BERT (MT + HSTC) using 9/18 HSTC

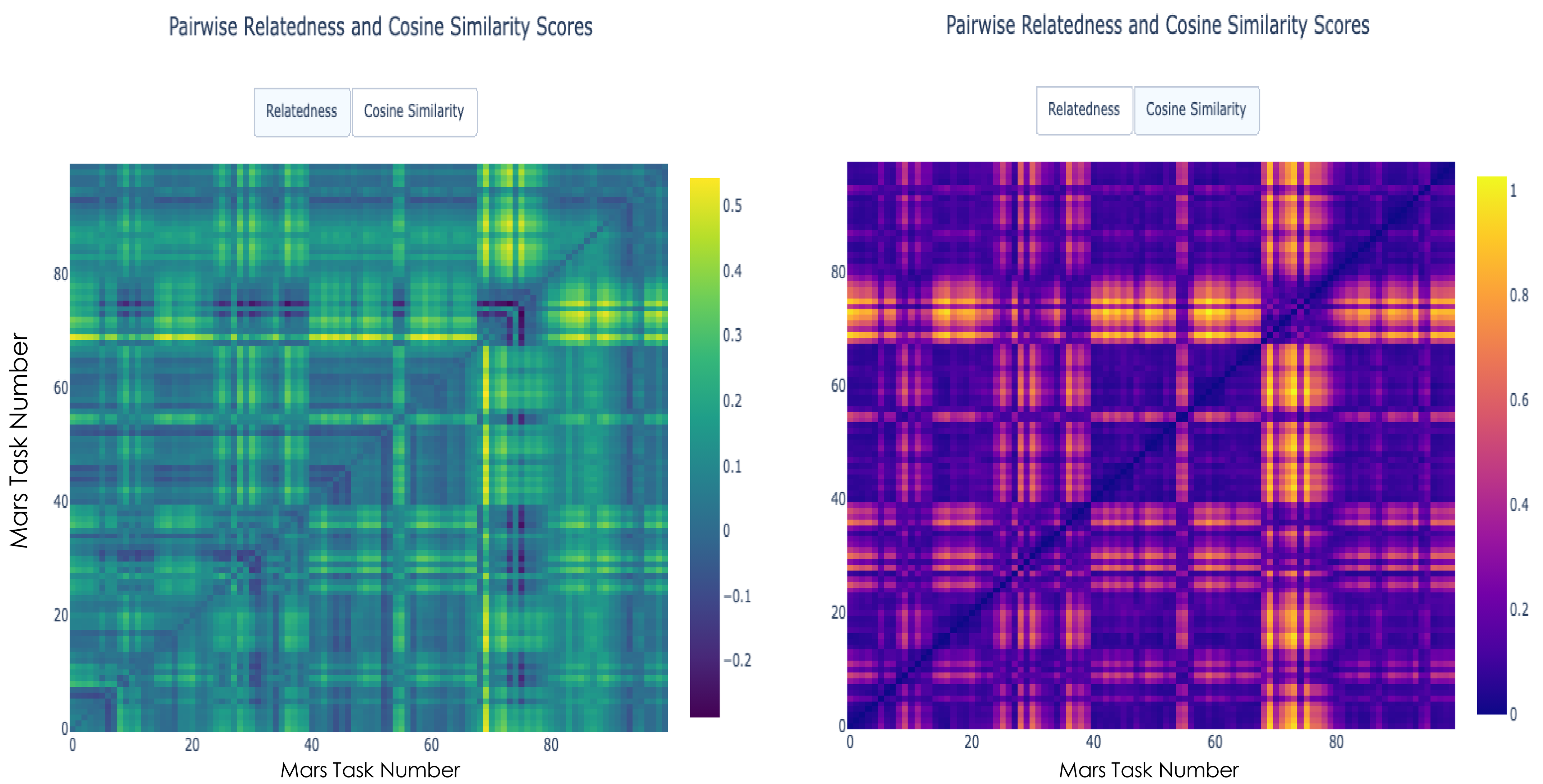


Classification Report

	precision	recall	f1-score	support
Cognitive – Simple	0.73	0.51	0.60	91
Cognitive – Complex	0.54	0.78	0.63	76
Vision	0.89	1.00	0.94	163
Cardiopulmonary	0.96	0.46	0.62	50
Upper Extremity	0.66	0.84	0.74	102
Ambulation and Standing (g)	0.63	0.58	0.61	57
Pressure Ops	0.85	0.71	0.78	66
Hand	0.83	0.96	0.89	133
Communication	0.62	0.89	0.73	37

Future Work: Unsupervised Learning

To tailor BERT to be able to predict without training on data, we plan to compare and cluster similar MT together by extracting the meaning of the tasks without the HSTC labels. By comparing two (or more) sentences with each other we can use a pairwise *relatedness* metric for clustering [6]. This, and traditional cosine similarity, was used in the preliminary stages of development for distributed document representation [6].



Acknowledgements

The authors are grateful for the support of NASA's Human Research Program. Project funding was provided by the Science Integration Office.

References:

[1] Stuster, J. et al . "Human exploration of Mars: Preliminary lists of crew tasks." (2018). [2] Fernandez W. et al (2023) *Journal of Space Safety Engineering* 10, no. 2: 231-238. [3] Devlin J. et al (2019). arXiv preprint arXiv:1810.04805. [4] Avinash. (2024, February 29). *Decoding Bert: The NLP model that understands language like never before(LLM)-generative AI*. Medium. https://medium.com/@avinashmachinelearninginfo/decoding-bert-the-nlp-model-that-understands-language-like-never-before-llm-generative-ai-501190fab127 [5] joyasree78. (2023, July 4). Is GPT Group of models decoder only model. OpenAI Developer Forum. https://community.openai.com/t/is-gpt-group-of-models-decoder-only-model/286586/2

[6] Dikshit, V. (2019, July 18). Bert for unsupervised text tasks. Medium. https://medium.com/ether-labs/bert-for-unsupervised-text-tasks-fa6e9ce5d133