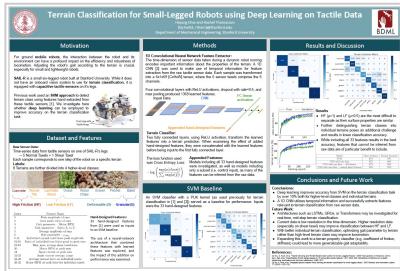




CS 229 projects, Spring 2020

All project posters and reports



Terrain Classification for Small Legged Robots Using Deep Learning on Tactile Data

General Machine Learning

Hojung Choi, Rachel Thomasson

[\[report\]](#) [\[poster\]](#)

Application of machine learning methods to identify and categorize radio pulsar signal candidates

Physical Sciences

Serena Debesai, Carmen Gutierrez, Nazli Ugur Koyluoglu

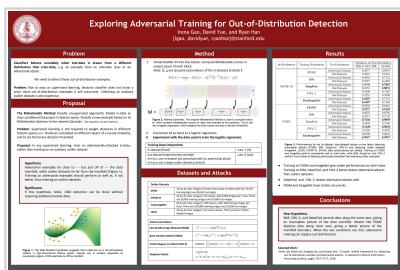
[\[report\]](#)

Using Machine Learning Models to Predict S&P500 Price Level and Spread Direction

Finance & Commerce

Alex Fuster, Zhichao Zou

[\[report\]](#)



Exploring Adversarial Training for Out-of-Distribution Detection

General Machine Learning

Irena Gao, Ryan Han, David Yue

[\[report\]](#) [\[poster\]](#)

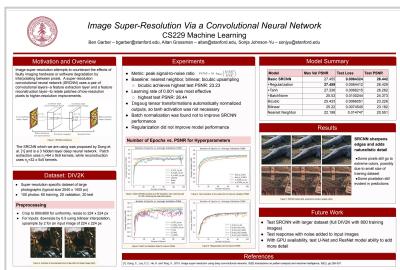


Image Super-Resolution Via a Convolutional Neural Network

Computer Vision

Ben Garber, Aitan Grossman, Sonja Johnson-Yu

[\[report\]](#) [\[poster\]](#)

Measuring and Incorporating Correlations in Generative Adversarial Networks

General Machine Learning

Vishesh Gupta

[\[report\]](#)

Classification of Legal Text

Finance & Commerce

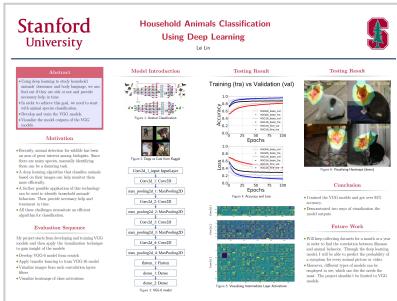
Krithika Iyer

[\[report\]](#)

Pancreatic cancer prognosis using clinical and radiomic data

Life Sciences

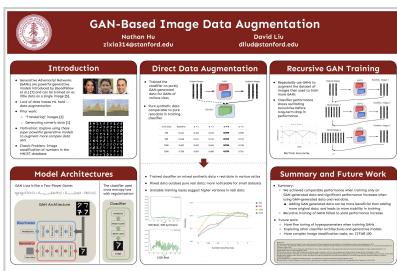
Arash Jamalian

[\[report\]](#)

Household Animals Classification Using Deep Learning

Computer Vision

Lei Lin

[\[report\]](#) [\[poster\]](#)

GAN-Based Image Data Augmentation

Computer Vision

David Liu, Nathan Hu

[\[report\]](#) [\[poster\]](#)

Semi-supervised Segmentation of Brain MRI Images

Life Sciences

Ali Mottaghi

[\[report\]](#)

Using LSTM and SARIMA Models to Forecast Cluster CPU Usage

General Machine Learning

Langston Nashold, Rayan Krishnan

[\[report\]](#)

Predicting a Decline in Patient Reported Outcomes for Cancer Patients on Chemotherapy

Life Sciences

Nicolai Ostberg, Dylan Peterson

[\[report\]](#)

Vision-Based Precision Pose Estimation For Autonomous Formation Flying

General Machine Learning

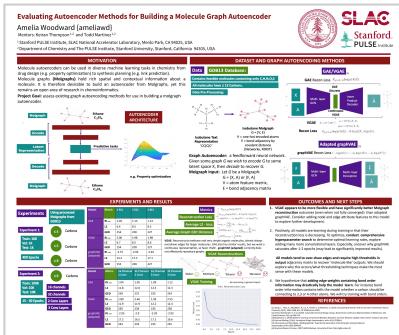
Rohan Punnoose

[\[report\]](#)

CS229: Final Report

Finance & Commerce

Kun Qian, Xingzi Xu, Emily You

[\[report\]](#)**Evaluating Autoencoder Methods for Building a Molecule Graph Autoencoder**

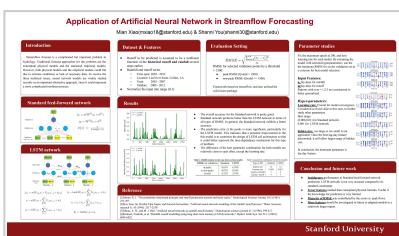
General Machine Learning/Physical Sciences

Amelia Woodward

[\[report\]](#) [\[poster\]](#)**Machine Learning based classification for Sentimental analysis of IMDb reviews**

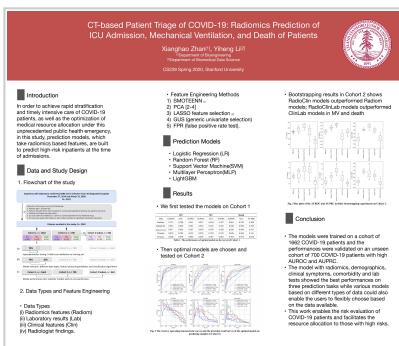
Natural Language

Chun-Liang Wu, Song-Ling Shin

[\[report\]](#)**Application of Artificial Neural Network in Streamflow Forecasting**

Physical Sciences

Mian Xiao, Shanni You

[\[report\]](#) [\[poster\]](#)**CT-based Patient Triage of COVID-19: Radiomics Prediction of ICU Admission, Mechanical Ventilation, and Death of Patients**

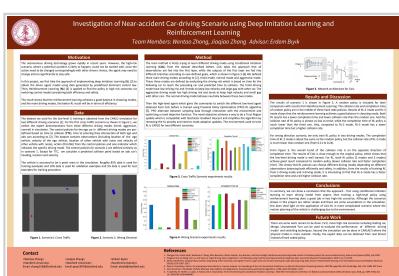
Life Sciences

Xianghao Zhan, Yiheng Li

[\[report\]](#) [\[poster\]](#)**Data Driven Solutions and Discoveries in Mechanics Using Physics Informed Neural Network**

Physical Sciences

Qi Zhang, Yilin Chen, Ziyi Yang

[\[report\]](#)**Investigation of Near-accident Car-driving Scenario using Deep Imitation Learning and Reinforcement Learning**

Theory & Reinforcement Learning

Wentao Zhong, Jiaqiao Zhang

[\[report\]](#) [\[poster\]](#)

Collaborative Filtering on Keywords Recommendation for Clinical Trial Records

Xiao Zhou¹

¹Stanford Center for Clinical Data Development

Abstract

Since after its discover similar clinical trials can be found in the same document due to the similarity of medical terms used in the abstracts, but the document may not contain all the necessary medical terms. Thus, it is necessary to recommend similar clinical trials based on the medical terms used in the abstracts and associated documents. The recommended results will be more accurate and useful. This paper proposes a collaborative filtering algorithm based on keywords to recommend similar clinical trials. The proposed algorithm can find the most similar clinical trials by calculating the similarity between the medical terms in each trial record.

Keywords

- > Document
- > Medical Record
- > Collaborative Filtering
- > Similarity

Model Development

The hypothesis is that for MDSI terms i , j and k , the similarity between the i -th and j -th term is proportional to the number of times the i -th term appears in the k -th document. The performance is measured by the average precision (AP) of the top 10 recommended results. AP is calculated as the sum of the precision at rank k for each query divided by the number of non-zero queries. The AP is calculated as follows:

$$AP = \frac{1}{|Q|} \sum_{q \in Q} \frac{\sum_{k=1}^{|R_q|} \min(\text{rel}_q(k), k)}{\sum_{k=1}^{|R_q|} k}$$

TF-IDF Utility Matrix

The TF-IDF of words terms in each trial record is calculated by the formula below:

$$tf_{ij} = \frac{\text{tf}_{ij}}{\sum_{j=1}^n \text{tf}_{ij}}$$

and implemented in a utility matrix U in Matlab as follows:

$$U = \sum_{i=1}^m u_i u_i^T$$

Model Training and Test

The training document were 50% of $m = 6$ clinical trials, and the test set was the remaining 50%. The performance is measured by the average precision (AP) of the top 10 recommended results. AP is calculated as the sum of the precision at rank k for each query divided by the number of non-zero queries. The AP is calculated as follows:

$$AP = \frac{1}{|Q|} \sum_{q \in Q} \frac{\sum_{k=1}^{|R_q|} \min(\text{rel}_q(k), k)}{\sum_{k=1}^{|R_q|} k}$$

Online Usage Data Recovery

The online usage data recovery is based on "nearest neighbor" estimation of a record vector, which is calculated as follows:

$$\hat{u}_i = \frac{1}{k} \sum_{j=1}^k u_j$$

Implementation

The whole data array matrix U is divided into two parts: U_{train} and U_{test} . The development matrix U_{train} and the test matrix U_{test} are used to calculate the precision of the predicted results.

Conclusion

This paper proposes a collaborative filtering algorithm based on keywords to recommend similar clinical trials. The proposed algorithm can find the most similar clinical trials by calculating the similarity between the medical terms in each trial record.

Notice that, it is incomplete.

Collaborative Filtering on Keywords Recommendation for Clinical Trial Records

General Machine Learning

Xiao Zhou

[report] [poster]