SHRAY MATHUR

Email: shray@mit.edu
Mobile: +91 8826828362
Website: https://shray64.github.io/

ACADEMIC DETAILS

Bachelor of Engineering (Honors) 2021
Computer Science with Minor in Data Science **CGPA 8.5**
Birla Institute of Technology and Science, Pilani, India

High School 2017
Delhi Public School, Gurgaon, India **94.4% (Gold Medal)**

Technical Proficiency:
Python, Tensorflow, PyTorch, Keras, R, Matlab, Java, C, ArcMap

PUBLICATIONS AND RESEARCH PAPERS

1. "Predictability of U.S Regional Extreme Precipitation Occurrence Based on Large Scale Meteorological Patterns (LSMPs)" by Gao X., **Mathur S**. – **Journal of Climate** - https://doi.org/10.1175/JCLI-D-21-0137.1
2. "Integrated Models of Machine Learning for Design of a Crop Recommendation System for Rajasthan, India" by Goel L., **Mathur S.**, Jindal A., Gandhi A. - **Journal of the Indian Society of Remote Sensing** (submitted, under review)
3. "YouTop200: A Most-Watched Video Object Segmentation Dataset" by Wei D., [et al, including **Mathur S.**] – IEEE International Conference on Computer Vision **(ICCV) 2021** (submitted, under review)
4. "Interpretable Machine Learning for Satellite Based Remote Sensing" by Jeppesen J., **Mathur S**., Jacobsen R. – **Remote Sensing of Environment** (in progress)
5. "Design and Implementation of a Crop Recommendation System Using Nature Inspired Intelligence for Rajasthan" by Goel L., Jindal A., **Mathur S.,** Gandhi A. – **Deep Learning for Sustainable Agriculture** (accepted)

RESEARCH PROJECTS

Research Intern, Massachusetts Institute of Technology – Cambridge, MA, USA May 2020 – Present

Working with the MIT Joint Program on the Science and Policy of Global Change on classifying, identifying, and predicting synoptic-scale circulation patterns associated with extreme precipitation events using deep learning methods.

Research Project 1:
Extreme Precipitation Event Prediction based on Large Scale Meteorological Patterns (LSMPs) for the Pacific Coast California and the Midwestern United States. Our work on this projected has been accepted at the Journal of Climate.

- Trained machine learning models with 1979-2005 data and tested on 2006-2019 NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA 2) reanalysis data.
- Examined combinations of atmospheric variables like wind speed (zonal and meridional components), air temperature, specific and relative humidity, total precipitable water, integrated vapor transport and Dew-point temperature.
- Techniques like oversampling, undersampling and SMOTE were used to balance the data set.
- Applied Deep learning methods like Neural Network Ensembles and Convolutional Neural Networks (CNNs).

- Across 24 CNN schemes, highest Gilbert Skill Scores (GSS) of 0.32 for PCCA and 0.26 MWST were achieved.
- Interannual variability of extreme events was well captured with temporal correlations significant at the 0.01 level and RMSEs less than 2 days for PCCA.

Publication:
"Predictability of U.S Regional Extreme Precipitation Occurrence Based on Large Scale Meteorological Patterns (LSMPs)" by Gao X., **Mathur S**. – **Journal of Climate** - https://doi.org/10.1175/JCLI-D-21-0137.1

Research Project 2:
Cluster and visualize atmospheric forcing patterns from climate reanalysis data products using a neural network based **unsupervised machine learning** approach.

- Three separate reanalysis products are used in the study, specifically – ECMWF's ERA5, NASA's MERRA2 and NCEP's NARR.
- Self-organizing maps (SOMs), a type of artificial neural network, are used to perform unsupervised clustering and identify dominant atmospheric circulation patterns.
- Exponential decaying neighborhood function and learning rate parameter are used to identify a two-dimensional lattice of explanatory codes to which individual days are assigned and clustered.
- Composite weather patterns formed at individual SOM nodes are compared to derive insights regarding the datasets.

Research/Scientist Assistant, University of Texas at Austin – Austin, TX, USA Jan 2021 – May 2021

Worked with the Bureau of Economic Geology at UT Austin on the use of machine learning techniques for high spatio-temporal soil moisture estimates from spaceborne GNSS-R data using NASA's Cyclone Global Navigation Satellite System (CYGNSS).

- Gridded and harmonized the CYGNSS L1 v3.0, TxSON in-situ soil moisture and SMAP datasets.
- Conducted a three-way grid-wise sensitivity analysis between CYGNSS, TxSON and SMAP.
- Demonstrated significant shifts (at the 0.01 level) in the spatio-temporal relationship between CYGNSS-derived surface reflectivity and soil moisture.
- Built an Artificial Neural Network to retrieve soil moisture estimates at 9km and 3km grid resolutions and established an optimal set of input features including surface reflectivity, SP incidence angle, surface roughness, NDVI, Elevation and Depth to Restrictive Layer.
- Retrieved soil moisture values at an *R* of 0.763 and *RMSE* of 0.034.

Undergraduate Research Thesis:
"Investigating CYGNSS-derived surface reflectivity for estimating soil moisture in Texas" – by **Mathur S.**
https://shray64.github.io/files/UT/Thesis.pdf

Research Intern, Harvard University – Cambridge, MA, USA. Sept 2020 – May 2021

Working at the Visual Computing Group on 3D instance segmentation for bio-medical images and 2D video object segmentation.

Research Project 1:
Built a semi-automatic segmentation pipeline to develop a novel dataset – **YouTop200** – for Video Object Segmentation (VOS). Our work on this project has been submitted to the IEEE International Conference on Computer Vision (ICCV) 2021.

- Worked with over 200 most-watched full-length YouTube videos across 10 genres with 431K annotated instance masks
- Propagated masks using a Space-Time Memory Network. Took a coarse-to-fine approach to mask propagation: from keyframes to 1FPS to 6FPS.
- Used Volumetric Segmentation Annotation Software (VAST) to examine and correct segmentation results at each stage of the semi-automatic pipeline.

Publication:
"YouTop200: A Most-Watched Video Object Segmentation Dataset" – IEEE International Conference on Computer Vision 2021 **(ICCV)** (submitted, under review)

Research Project 2:
Leverage unlabeled data to develop a **Semi-Supervised Learning** approach for **3D segmentation** of Electron Microscopy volumes.

- Used a pre-trained Residual Symmetric 3D U-Net model to generate pseudo-labels for additional unlabeled volumes.
- Retrained the model using labelled and pseudo-labeled volumes to improve adapted rand scores on test volumes.
- Converted binary foreground probability maps, instance contours and signed distance transform to instance masks via the watershed segmentation algorithm.

Research Intern, Aarhus University – Aarhus, Denmark **June 2020 – Sept 2020**

Worked in the Department of Electrical and Computer Engineering at Aarhus University in Aarhus, Denmark in the area of **Explainable AI**. My work involved identifying optimal methods for describing and visualizing the decision process for a range of machine learning classifiers used for classifying crop types using multi-temporal satellite radar data from Sentinel 1.

- Applied Logistic Regression and Neural Networks for crop type classification of 12 crop types based on 44,333 fields in Jutland and Funen, Denmark.
- Used temporal patterns in VV, VH backscattering to determine the crop types.
- Linear Discriminant Analysis (LDA) was used to create a smaller set of uncorrelated features to handle multi-collinear features and Permutation Importance, Shapley Explanations and LIME were used to identify the most important LDA components for a given crop type.
- Layer-wise Relevance Propagation (LRP) was used to explain the predictions proposed by the neural network. Other techniques like DeConvNet, Guided BackProp and DeepLIFT were also studied.
- Consistent explanations for feature relevance from the different explainability techniques were observed.

Research Paper:
"Interpretable Machine Learning for Satellite Based Remote Sensing" – by Jeppesen J., Mathur S., Jacobsen R. – **Remote Sensing of Environment** (in progress)

Research Project, Birla Institute of Technology and Science – Pilani, India **Sept 2019 – Apr 2020**

Worked on **Machine Learning based models for crop recommendations** for districts in the Indian state of Rajasthan using satellite data and imagery from ResourceSat (India) and Landsat 8(USA) satellites.

- Satellite data for different districts of Rajasthan was downloaded from the USGS earth explorer website.

- Features used for classification of crop type were soil type, soil pH, soil moisture, land type, NDVI, NDWI, rainfall and temperature.
- Used a quantification model for soil texture based on a continuum of bare soils, obtained by overlapping multi-temporal satellite images to help classify soils into categories based on particle size.
- The blue, green, red and SWIR bands of Landsat 8 images were found to be the most suitable bands for estimating soil pH.
- Land Surface Temperature was calculated using the Thermal bands of Landsat 8 data and the NDVI was calculated using the Red and NIR bands.
- Built a correlation between Land Surface Temperature and NDVI using scatter plots to generate area specific constants used in calculation of the Soil moisture index which was then used to estimate soil moisture.
- Compared the performance of different machine learning models – KNN, SVM, Random Forest, Multi-Layer Perceptron – and observed that the Random Forest model achieved the highest accuracy.

Publications:

"Integrated Models of Machine Learning for Design of a Crop Recommendation System for Rajasthan, India" by Goel L., **Mathur S.**, Jindal A., Gandhi A. - **Journal of the Indian Society of Remote Sensing** (submitted, under review)

“Design and Implementation of a Crop Recommendation System Using Nature Inspired Intelligence for Rajasthan” by Goel L., Jindal A., **Mathur S.,** Gandhi A. – **Deep Learning for Sustainable Agriculture** (accepted)

Citizenship: USA