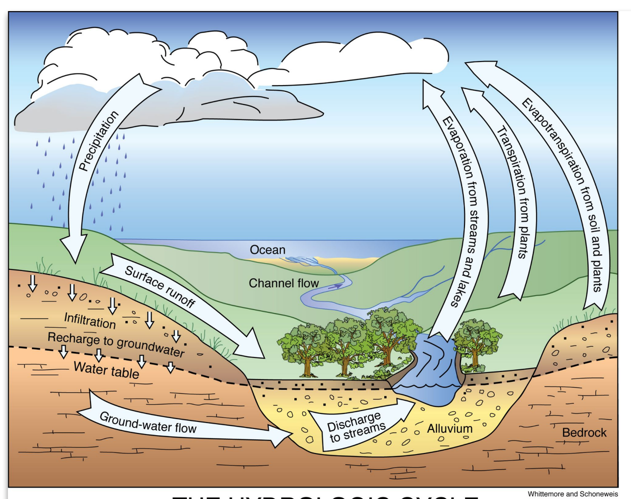
IMPROVING STREAMFLOW PREDICTIONS WITH VISION TRANSFORMERS

Kshitij Tayal[1], Arvind Renganathan[2], Dan Lu[1]
[1]Oak Ridge National Laboratory, [2]University of Minnesota
tayalk@ornl.gov, renga@umn.edu, lud1@ornl.gov

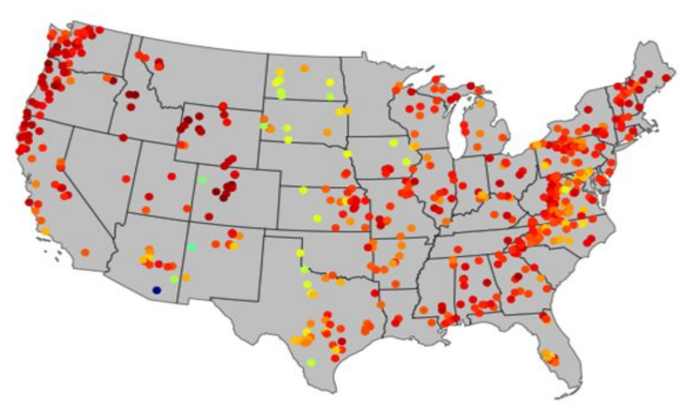
Climate Change AI | Oak Ridge National Laboratory | University of Minnesota | ICLR

Introduction

- Accurate streamflow prediction is crucial to understand climate impacts on water resources and develop effective adaptation strategies.



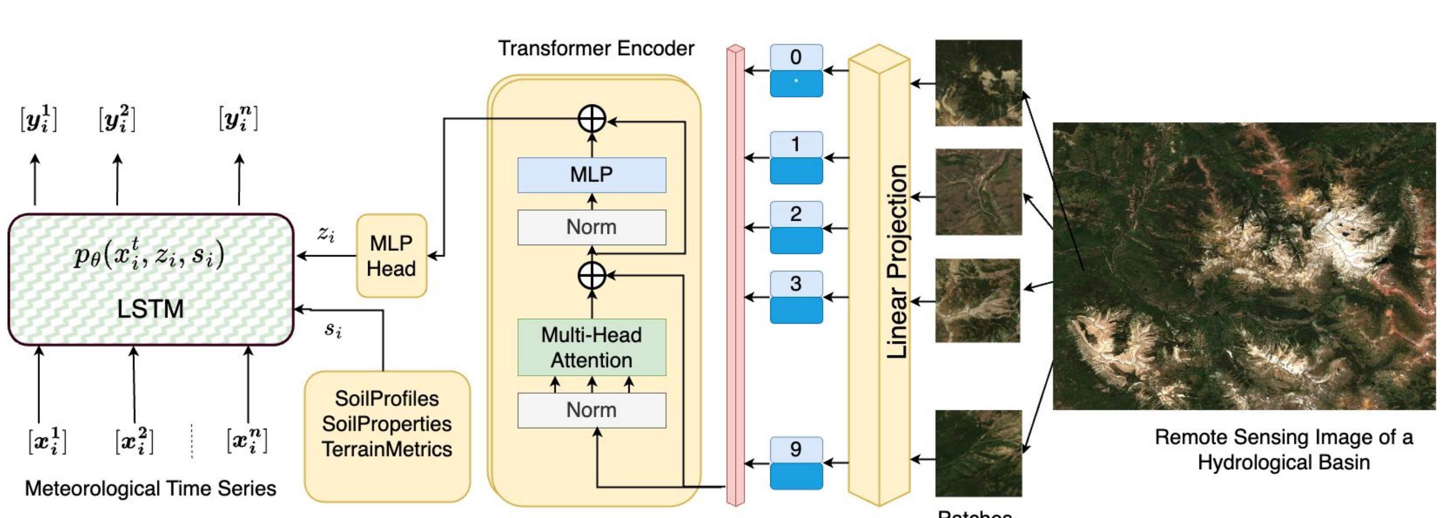
THE HYDROLOGIC CYCLE


BASINS IN UNITED STATES

- A global Long Short-Term Memory (LSTM) model, using data from multiple basins, can enhance streamflow prediction, but these global models require detailed basin attributes, yet acquiring them remains a challenge.
- To overcome this, we introduce the Geo-ViT-LSTM model, a novel approach that enriches LSTM predictions by integrating basin attributes derived from remote sensing with a vision transformer.

Proposed Approach: Geo-ViT-LSTM

- Geo-ViT-LSTM centers on two main aspects: (1) Multi-modal feature extraction, using ViT to transform RS imagery into spatially aware feature representations, highlighting key terrain, land cover, and vegetation health indicators, and (2) Integrated modeling, wherein the RS features are combined with static attributes and meteorological sequences within the LSTM model.
- ViT segments the remote sensing image into equal-sized sub-images, which are linearly embedded and enhanced with positional embeddings to encode their spatial locations. These tokens are then processed through transformer encoder blocks.
- An MLP head aggregates the transformer outputs to incorporate global image information, which is then concatenated with existing static attributes and fed into a 32-hidden unit LSTM.



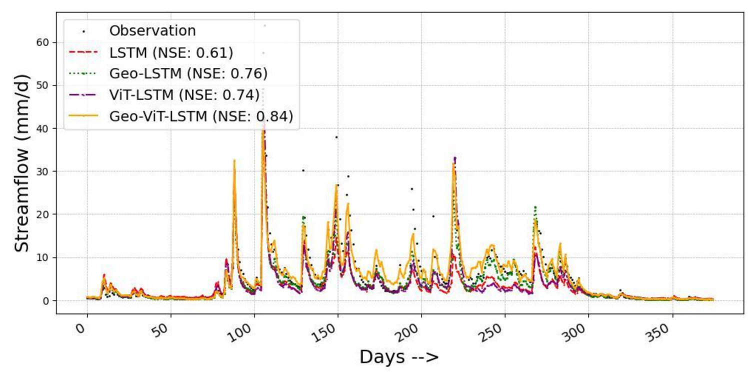
Geo-ViT-LSTM

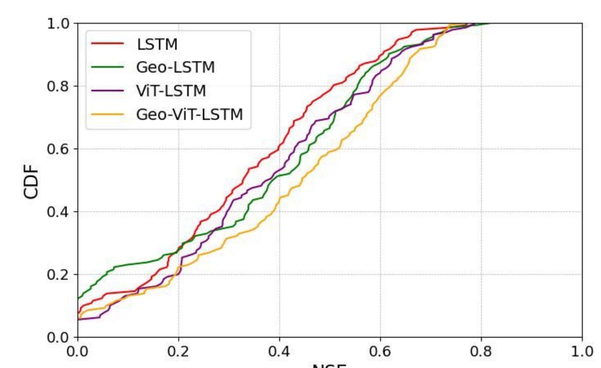
Dataset Construction

- CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) dataset: Includes daily weather data, streamflow observations, and 27 basin attributes from 531 catchments across the US.
- Sentinel-2 images for all 531 basins downloaded via Google Earth Engine, with cloud masking and median compositing.

Experiments and Results

- Benchmarked against three baselines: LSTM, Geo-LSTM, and ViT-LSTM.
- Two testing strategies: Temporal Test (testing on 400 training basins during the testing period) and Spatiotemporal Test (testing on remaining 131 basins during the same testing period).
- Geo-ViT-LSTM achieved the best performance on both cases, with an 11% improvement over the next best method in terms of lower RMSE and higher NSE values.
- Cumulative Distribution Function (CDF) plot demonstrates that a larger segment of the test dataset achieved superior NSE scores with Geo-ViT-LSTM.





Conclusion & Future Work

- Geo-ViT-LSTM marks a significant advancement in land surface modeling, providing a more comprehensive tool for managing water resources under climate change.
- The vision transformer effectively incorporates context attributes and enhances model accuracy by leveraging the spatial relationships inherent in the RS data.
- Future work aims to extend the methodology for dynamic monitoring from multiple snapshots of RS images and broaden the application to include other water- and carbon-related variables that depend on land surface attributes.