**CRIS Project: Potential Research Questions and Pertinent Information**

**Scientific Background**

*Using eXplainable Artificial Intelligence for Climate-Adaptive Crop Production.* Climate variability and change affect crop production by increasing the uncertainty of precipitation and temperature with critical implications for crop yield (Kukal and Irmak, 2018). Accurate short-range predictions of hydroclimatic factors and crop yield along with a robust understanding of long-term water sustainability challenges are critical for producers, policymakers, and researchers to increase the resilience of agricultural production through timely adaptation to risk from climate variability. Novel eXplainable Artificial Intelligence (XAI) frameworks are cutting-edge data-driven tools to derive insights for climate-adaptive crop production through rigorous data-driven understanding of climate impacts on agricultural water availability and effectiveness of different cropping systems. Explainability refers to agreement of predicted hydroclimatic factors and trends with real-world observations (i.e., fidelity) and ability to offer simple, transparent explanation of data-driven insights (i.e., interpretability) to facilitate information transfer to stakeholders (Chakraborty et al., 2021). Unraveling the underlying reasons why an Artificial Intelligence (AI) model makes certain predictions is a key challenge in the hydroclimatic and agroclimatic domains and is likely to encourage the adoption of AI by practitioners and stakeholders for high stake hydrological and agricultural decisions geared towards long-term sustainability and resilience of agricultural production. AI-driven hydroclimatic modeling approaches have advanced significantly in the past decade with a growing number of applications of artificial neural networks, support vector machines, adaptive neuro-fuzzy inference systems, regression trees, and random forest (Ardabili et al., 2019). These non-explainable or black-box models have been applied to project precipitation (Kang et al., 2020; Pham et al., 2020), temperature (Fan et al., 2021; Jeong et al., 2021), evapotranspiration (Ashrafzadeh et al., 2020; Ferreira & da Cunha, 2020), runoff (He at al., 2020; Feng at al., 2021), groundwater recharge (Sahoo et al., 2017; Kordestani et al., 2019; Miraki et al., 2019), and agroclimatic risks (Chemura, Schauberger, and Gornott, 2020; Shin, Kim, and Ha, 2020). Yet, there is a critical need for reliable, intelligent XAI frameworks to assess the long-term changes in the hydrological and agricultural processes. Reliable, high-accuracy predictions of agroclimatic risks under future climate conditions require: (i) development of robust AI data imputation methods based on sequential transfer learning theory to fill-in long data gaps from different data sources, and thus, overcome the limitations of the existing off-the-shelf imputation models for handling long stretches of continuous missing data; (ii) automatic synthesis and selection of suitable exogenous and endogenous climatic and temporal features over longer periods to accurately train the model for reliable projection of extreme hydroclimatic conditions (e.g., severe droughts or floods); (iii) algorithmic improvements through the application of hybrid probabilistic AI models – e.g. custom boosting models (Chen and Guestrin, 2016; Ke et al., 2017; Chakraborty and Elzarka, 2019a and 2019b) hybridized with a natural gradient approach (Duan et al., 2019; Başağaoğlu et al., 2021) – to enhance model accuracy and minimize and quantify prediction uncertainties; (iv) apply post-hoc explainability layers on top of inherently interpretable AI models (Chakraborty et al., 2021) to enhance the models’ trustworthiness and derive insights regarding the underlying physical processes; and (v) analyze the insights with visual tools and interpret the critical inflection points of hydroclimatic and agroclimatic features beyond which the agroclimatic risks and crop yields either increase or decrease. Thus, allowing decisionmakers and stakeholders to contemplate and prepare for all potential future scenarios, and overcome the natural barriers by leveraging the power of meaningful and trustworthy projections and insights.

**Sub-objectives: Approach and Research Procedures**

**Sub-objective 1.A:** Develop an XAI framework for agroclimatic risk assessment and adaptation in the Southern Great Plains, including the Fort Cobb Reservoir Experimental watershed (FCREW).

**Rationale Statement:** The climate of the Southern Great Plains is conducive to high levels of agricultural productivity if adequate water resources are available to cope with extreme climate variability. Understanding future agroclimatic threats to crop production is essential for adaptive agricultural water management and cropping systems. At present, there is no XAI framework for high-accuracy predictions of agroclimatic factors that affect crop production and agricultural water sustainability. The XAI framework is needed to test the effectiveness of adaptation plans and conservation measures in the Southern Great Plains using various available data streams.

**Research Goal 1.A.1:** Apply AI techniques to predict agroclimatic factors (e.g., reference crop evapotranspiration, actual evapotranspiration, soil moisture, runoff, surface water discharge, groundwater recharge, and groundwater levels).

**Research Goal 1.A.2:** Test the predictive power of XAI to provide short-range (e.g., season-ahead) forecast of crop yield and variability of agricultural water resources. Test the explainability power of XAI thru conditional probability assessments of the inflection points unraveled by the XAI, as described by Chakraborty et al. (2021).

**Research Goal 1.A.3:** Assess future agroclimatic risks to crop production in the Southern Great Plains by analyzing the XAI-derived inflection points beyond which the crop yield decreases and evaluating the chances of occurrence of such inflection points in the future.

**Development of XAI Framework Modules:** Modules that will be developed include automated AI-based imputation of missing agroclimatic data if necessary, deep feature synthesis and selection, novel hybridized AI models to characterize the inherent relationship between the agroclimatic variables, facilitate probabilistic predictions of extreme events and quantify prediction uncertainties. If necessary, custom grid-search cross-validated boosting models will be used to impute continuous stretches of missing local climate data from nearby stations via sequential transfer-learning technique. An innovative deep feature synthesis (Kanter et al., 2015) will be implemented to generate temporally lagged features to automatically select the most relevant features based on data autocorrelations and correlations with other data. The automatic selection of the most relevant lags for both the exogenous and endogenous variables will be done via a greedy search technique coupled with cross-validation heuristics by minimizing the squared error loss function. This technique will eliminate the need for trial-and-error procedures to select the best input for AI modeling. A hybrid boosting technique and natural gradient optimization will be used to reduce the prediction uncertainties via probabilistic predictions instead of deterministic predictions. The AI hyperparameters will be tuned using a k-fold cross validation process to produce more accurate models. The tuned models will be tested on an unseen portion (10-30%) of the original dataset and the performance will be measured using the correlation coefficient (*R2*) and the root mean square error (RMSE) metrics. We will apply a post-hoc model agnostic representation of feature importance, where the impact of each feature on the model will be calculated using a tool called SHAP (SHapley Additive exPlanations) based on Shapley values (Roth, 1988; Štrumbelj and Kononenko, 2014; Lundberg and Lee, 2017; Lundberg et al., 2020). In other words, SHAP integrated with the AI models yields the XAI framework to reveal how much each feature contributes - either positively or negatively - to the target and the average marginal contribution of each feature value to the corresponding target value. The XAI-generated explanations can be categorized as global (i.e., summarized relevance of the input features in the model) or local (i.e., based on individual predictions).

**Data, analysis, and interpretation:** The XAI framework will be built, trained, and tested on extensive hydroclimatic and agroclimatic data at different spatial and temporal resolutions, including: historical climate data from mesoscale climate monitoring networks, airports, climate data clearing houses (e.g., TerraClimate, gridMET, PRISM, and MERRA2); computed time series of hourly, daily, and monthly reference evapotranspiration using the Penman-Monteith equation and daily actual evapotranspiration from available eddy covariance measurements; available moisture data; surface water discharge data from USGS gaging stations and groundwater data; available groundwater recharge estimates from USGS and based on distributed watershed modeling using calibrated Soil and Water Assessment Tool (SWAT) model; representative concentration pathway (RCP) projections of climate change from Coupled Model Intercomparison Project Phase 5 (CMIP5); and the new 21st century shared socio-economic pathway (SSP) projections of climate change from CMIP6. These datasets will be utilized to generate accurate hydroclimatic and agroclimatic future projections using the trained and tested XAI frameworks described in the previous section. The proposed XAI framework could help us characterize the relationship between the hydroclimatic and agroclimatic processes, i.e., how they interact and influence each other. Such characterizations can be trusted by the end-users if the accuracy of the base AI models is high on the testing data and the explanations are tested via conditional probabilities. Higher accuracies and conditional probabilities signify that the model can recognize the unique underlying processes, and thus, its predictions and explanations can be trusted, respectively.

**Contingencies:** None

**Collaborations:** Dr. Debaditya Chakraborty, University of Texas at San Antonio, San Antonio, TX will develop the explainable AI modeling framework. Dr. Ali Mirchi, Oklahoma State University, Stillwater, OK, will assess short-range future agroclimatic risks (See letters of collaboration).

**Literature Cited**

Ardabili, S., Mosavi, A., Dehghani, M. and Várkonyi-Kóczy, A.R., 2019, September. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In International Conference on Global Research and Education (pp. 52-62). Springer, Cham.

Ashrafzadeh, A., Kişi, O., Aghelpour, P., Biazar, S.M. and Masouleh, M.A., 2020. Comparative study of time series models, support vector machines, and GMDH in forecasting long-term evapotranspiration rates in northern Iran. Journal of Irrigation and Drainage Engineering, 146(6), p.04020010.

Başağaoğlu, H., Chakraborty, D. and Winterle, J., 2021. Reliable evapotranspiration predictions with a probabilistic machine learning framework. Water (Status: Accepted).

Chakraborty, D. and Elzarka, H., 2019a. Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. Energy and Buildings, 185, pp.326-344.

Chakraborty, D. and Elzarka, H., 2019b. Advanced machine learning techniques for building performance simulation: a comparative analysis. Journal of Building Performance Simulation, 12(2), pp.193-207.

Chakraborty, D., Başağaoğlu, H. & Winterle, J. Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. Expert Systems with Applications 170, 114498 (2021).

Chemura, A., Schauberger, B. and Gornott, C., 2020. Impacts of climate change on agro-climatic suitability of major food crops in Ghana. PloS one, 15(6), p.e0229881.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Duan, T., Avati, A., Ding, D.Y., Basu, S., Ng, A.Y. and Schuler, A., 2019. Ngboost: Natural gradient boosting for probabilistic prediction. arXiv preprint arXiv:1910.03225.

Fan, Y., Krasnopolsky, V., van den Dool, H., Wu, C.Y. and Gottschalck, J., 2021. Using Artificial Neural Networks to Improve CFS Week 3-4 Precipitation and 2-Meter Air Temperature Forecasts. Weather and Forecasting.

Feng, Z.K., Niu, W.J., Tang, Z.Y., Xu, Y. and Zhang, H.R., 2021. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. Journal of Hydrology, p.126062.

Ferreira, L.B. and da Cunha, F.F., 2020. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. Computers and Electronics in Agriculture, 178, p.105728.

He, X., Luo, J., Li, P., Zuo, G. and Xie, J., 2020. A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. Water Resources Management, 34(2), pp.865-884.

Jiang, S., Xiao, R., Wang, L., Luo, X., Huang, C., Wang, J.H., Chin, K.S. and Nie, X., 2019. Combining Deep Neural Networks and classical time series regression models for forecasting patient flows in Hong Kong. IEEE Access, 7, pp.118965-118974.

Jeong, S., Park, I., Kim, H.S., Song, C.H. and Kim, H.K., 2021. Temperature Prediction Based on Bidirectional Long Short-Term Memory and Convolutional Neural Network Combining Observed and Numerical Forecast Data. Sensors, 21(3), p.941.

Kang, J., Wang, H., Yuan, F., Wang, Z., Huang, J. and Qiu, T., 2020. Prediction of Precipitation Based on Recurrent Neural Networks in Jingdezhen, Jiangxi Province, China. Atmosphere, 11(3), p.246.

Kanter, J.M. and Veeramachaneni, K., 2015, October. Deep feature synthesis: Towards automating data science endeavors. In 2015 IEEE international conference on data science and advanced analytics (DSAA) (pp. 1-10). IEEE.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. In Advances in neural information processing systems (pp. 3146-3154).

Khashei, M. and Bijari, M., 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing, 11(2), pp.2664-2675.

Kordestani, M.D., Naghibi, S.A., Hashemi, H., Ahmadi, K., Kalantar, B. and Pradhan, B., 2019. Groundwater potential mapping using a novel data-mining ensemble model. Hydrogeology journal, 27(1), pp.211-224.

Lundberg, S. and Lee, S.I., 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874.*

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, *2*(1), pp.56-67.

Miraki, S., Zanganeh, S.H., Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H. and Pham, B.T., 2019. Mapping groundwater potential using a novel hybrid intelligence approach. Water resources management, 33(1), pp.281-302.

Pham, B.T., Le, L.M., Le, T.T., Bui, K.T.T., Le, V.M., Ly, H.B. and Prakash, I., 2020. Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 237, p.104845.

Roth, A.E. ed., 1988. *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press.

Sahoo, S., Russo, T.A., Elliott, J. and Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resources Research, 53(5), pp.3878-3895.

Shin, J.Y., Kim, K.R. and Ha, J.C., 2020. Seasonal forecasting of daily mean air temperatures using a coupled global climate model and machine learning algorithm for field-scale agricultural management. Agricultural and Forest Meteorology, 281, p.107858.

Štrumbelj, E. and Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, *41*(3), pp.647-665.

Xu, W., Peng, H., Zeng, X., Zhou, F., Tian, X. and Peng, X., 2019. A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. Applied Intelligence, 49(8), pp.3002-3015.

Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, pp.159-175.