California Wildfires: Predicting Wildfire Spread Using Machine Learning

Gwen Squires
Dept. of Mathematics | Southeast Missouri State University
Graduate Mentor: Daniel Breininger | Faculty Advisor: Dr. Nezamoddin N. Kachouie
Florida Institute of Technology

Introduction

- **4 million** acres burned, 2020
- **2.6 million** acres burned, 2021
- **$148.5 billion** in costs, 2018
- **30%** emissions increase, 2020
- **50,000** premature deaths, 2008-18

McKinney Fire in Klamath National Forest, CA. (CNN, 2022)

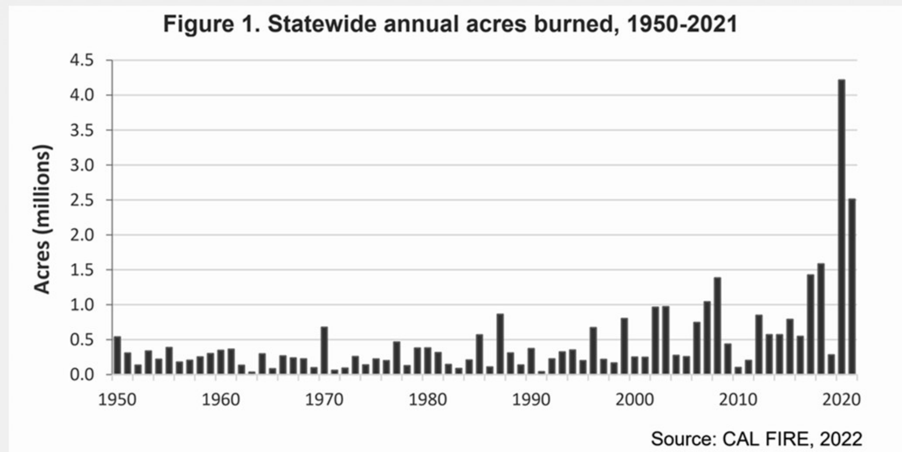
FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Previous Work


Environmental Impact Assessment Review

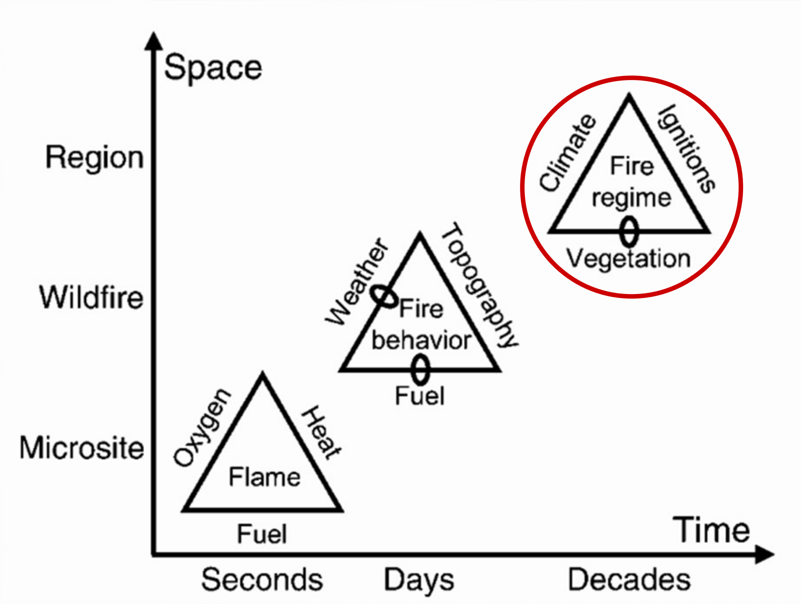
Estimating the probability of wildfire occurrence in Mediterranean landscapes using Artificial Neural Networks

Climate change and the eco-hydrology of fire: Will area burned increase in a warming western USA?



Acreage burned by year. (OEHHA, 2022)

Purpose



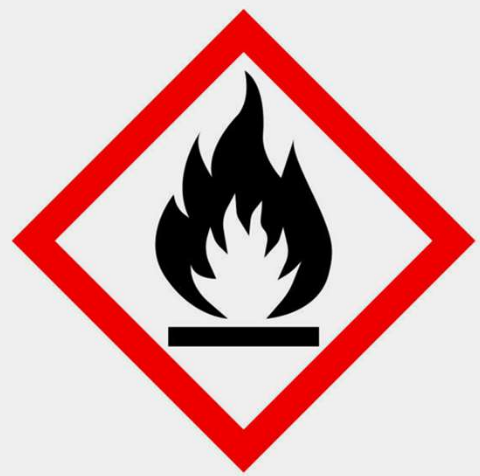
Spatial and temporal scales. (Parisien & Moritz, 2009)

Ignition vs. Spread

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Objective

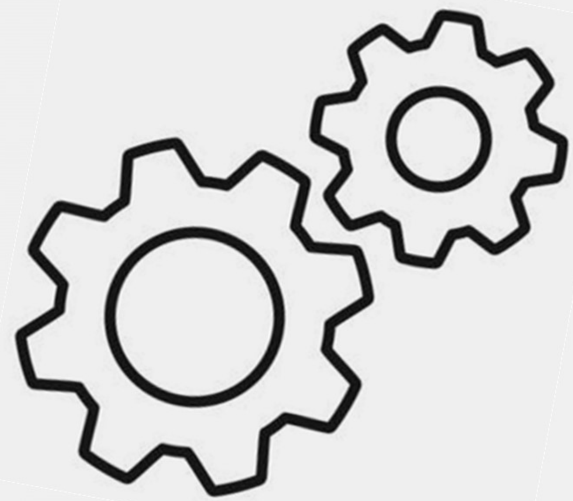
Determine the flammability of California's landscapes by predicting whether fire will spread given an ignition occurs at a specified time and location.



FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

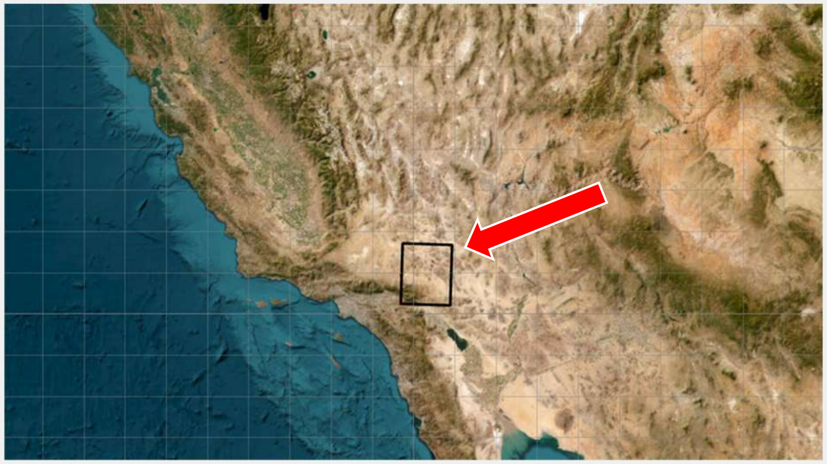
Procedure

- Data Processing
- Time Series Analysis
- Logistic Regression
- Neural Network



Data Sources

- TerraClimate
- Spatial wildfire occurrence data for the United States
- California Vegetation - WHR13 Types



Region of study, California, USA.

Data Description

- ➔ Actual Evapotranspiration
- ➔ Climate Water Deficit
- ➔ Potential Evapotranspiration
- ➔ Precipitation
- ➔ Runoff
- ➔ Soil Moisture
- ➔ Downward Surface Shortwave Radiation
- ➔ Maximum Temperature
- ➔ Minimum Temperature
- ➔ Vapor Pressure
- ➔ Wind Speed
- ➔ Vapor Pressure Deficit
- ➔ Palmer Drought Severity Index

Original observation

Difference

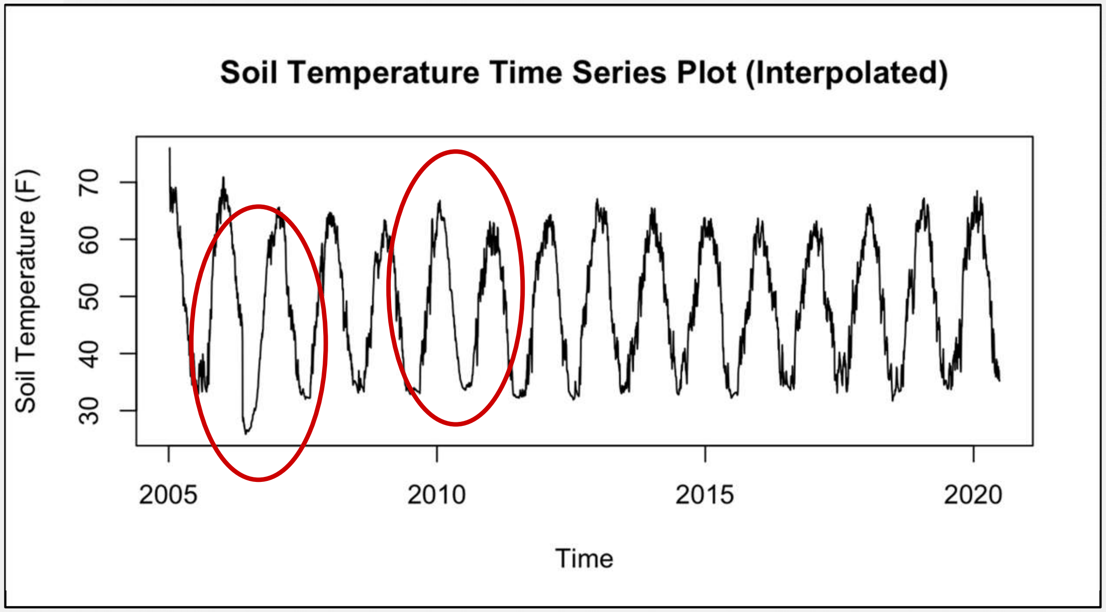
Anomaly

Anomaly Lags

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

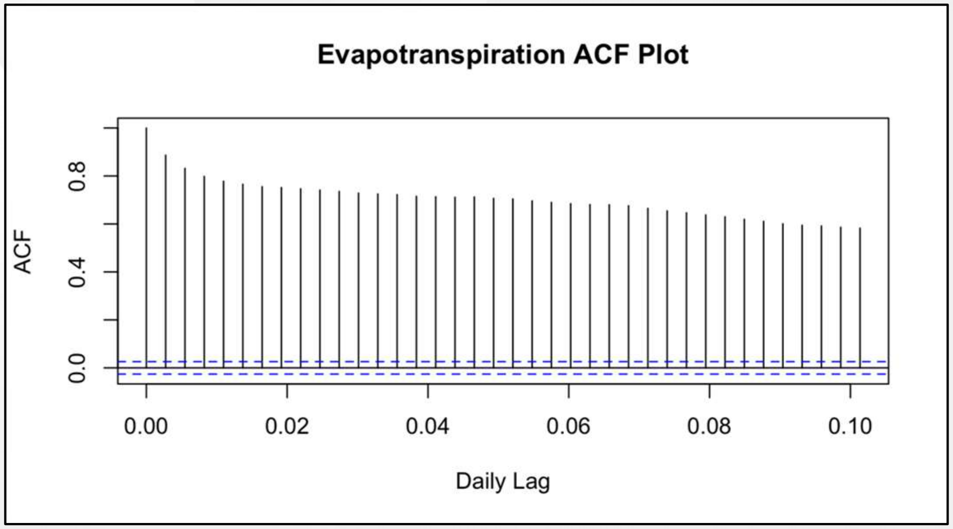
Time Series Analysis

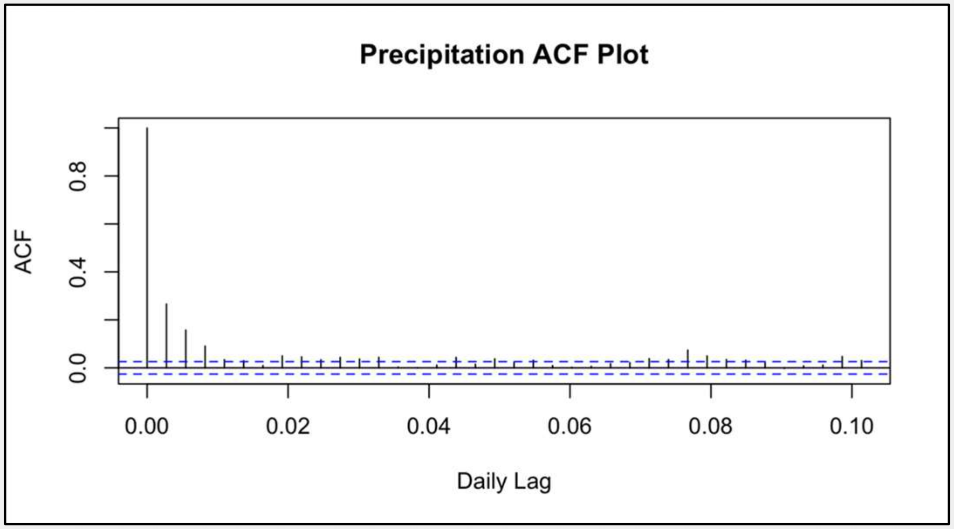
- California Irrigation Management Information System (CIMIS) data



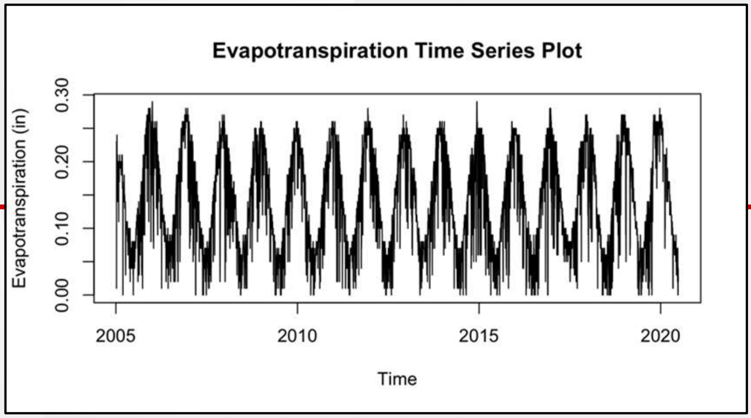
Time Series Analysis Cont.

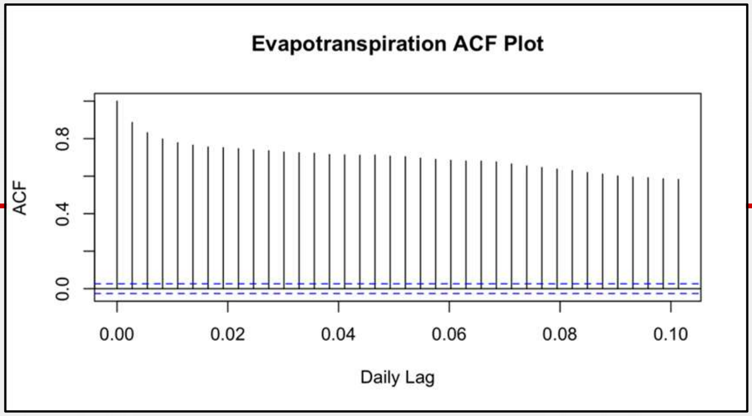
- Checking for stationarity

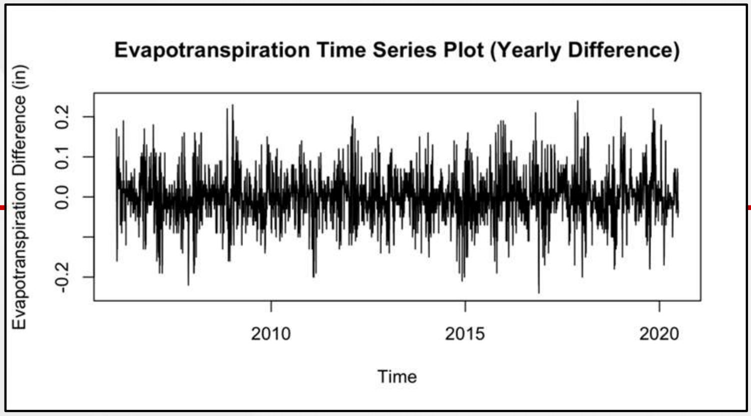


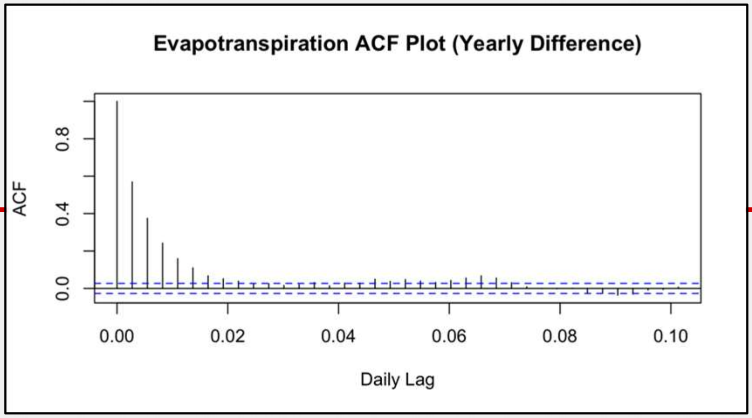


Time Series Analysis Cont.





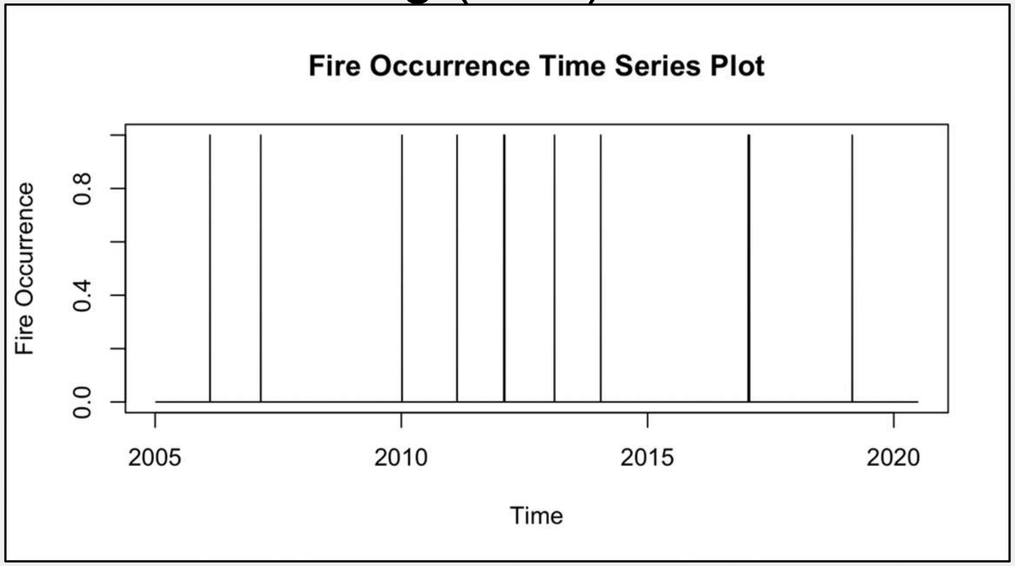




FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Time Series Analysis Cont.

- Autoregressive (AR) model
- Autoregressive Distributed Lag (ADL) model



Feature Reduction

x_1

⋮

$x_{n\text{-}1}$

x_n

$$\log_e \frac{p}{1-p} = \beta_0 + \beta_1 x_1$$

Keep
($p \leq 0.05$)

Remove
($p > 0.05$)

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Feature Reduction Cont.

Univariate Logistic Regression

LARGE SET

stepAIC

SMALL SET

Feature Reduction Cont.

LARGE SET

↓

Subset

↓

Train-Test Split

SMALL SET

↓

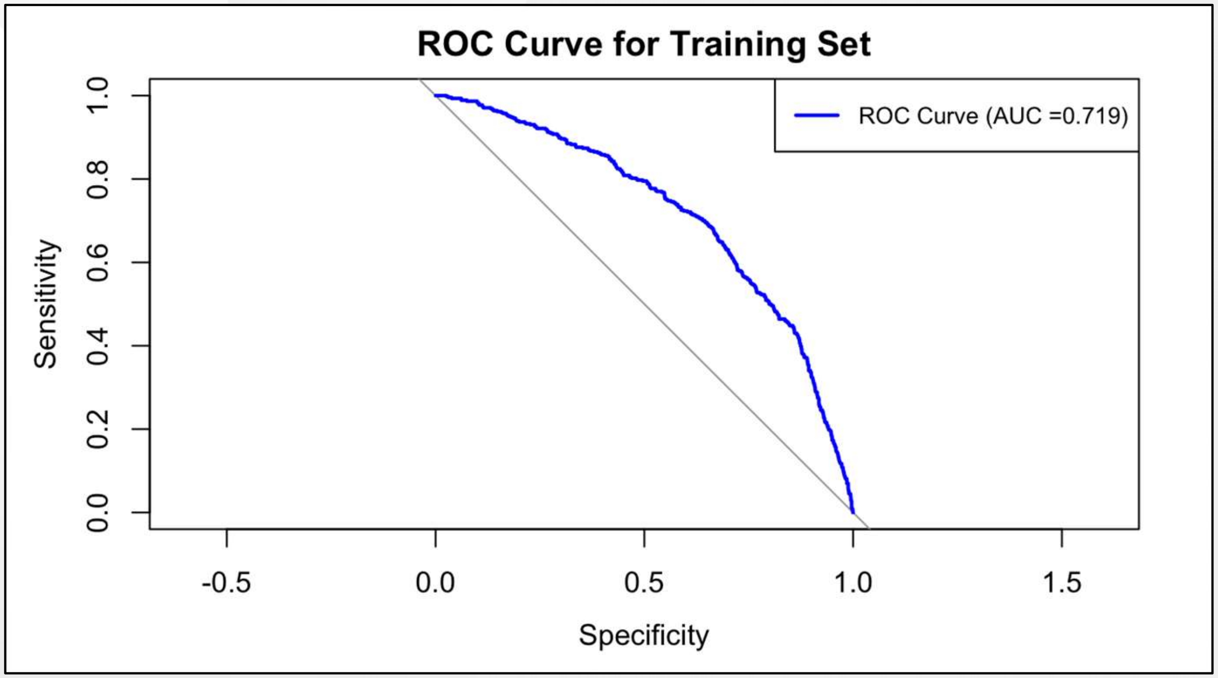
Subset

↓

Train-Test Split

Logistic Regression

Large Variable Set

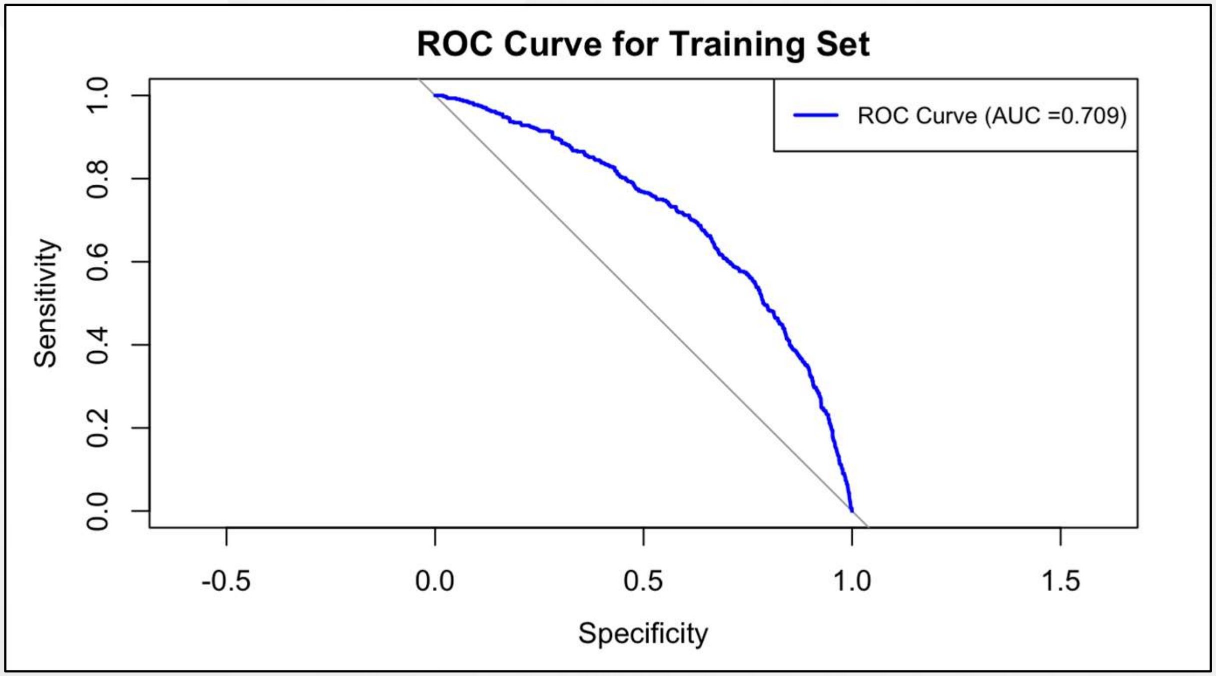


Accuracy	**0.633**
Sensitivity	**0.612**
Specificity	**0.653**
AUC	**0.719**

Optimal Cutoff: 0.214

Logistic Regression Cont.

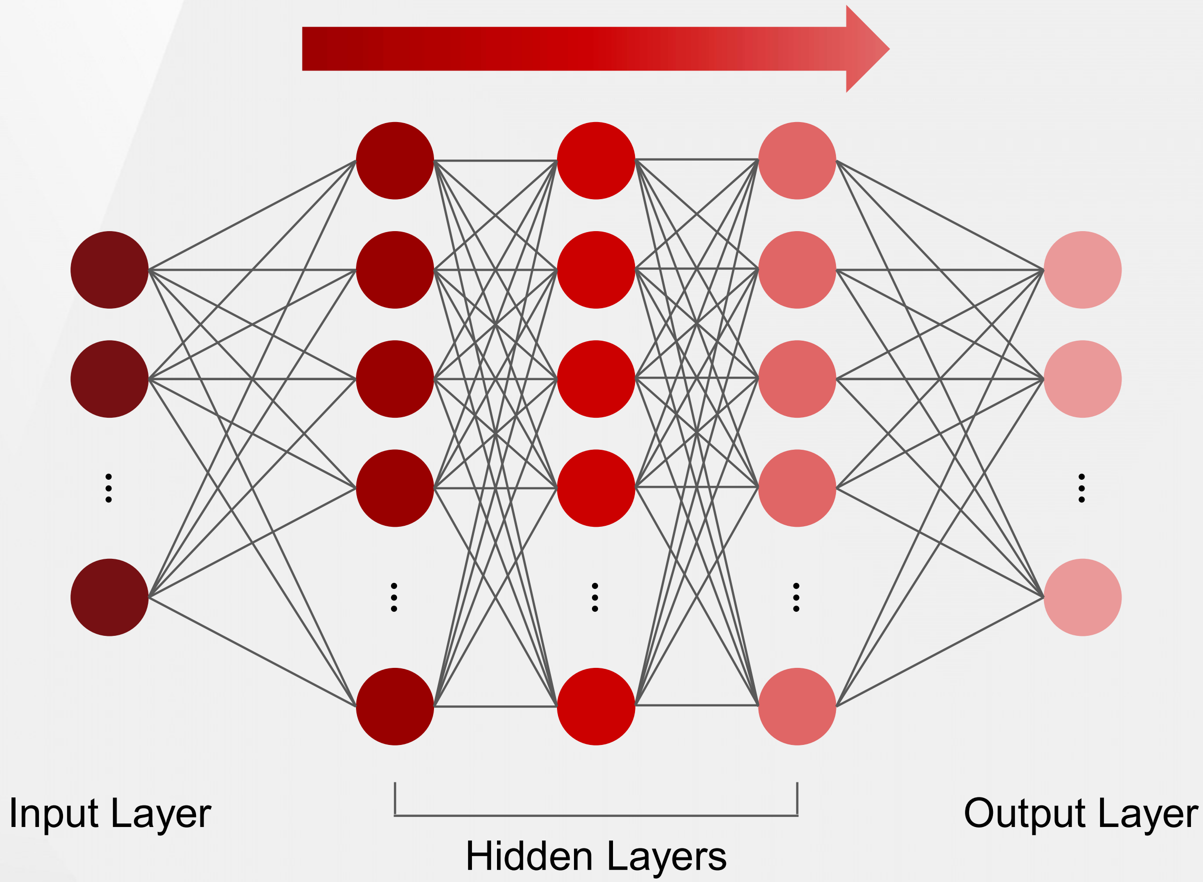
Small Variable Set



Accuracy	**0.643**
Sensitivity	**0.674**
Specificity	**0.612**
AUC	**0.709**

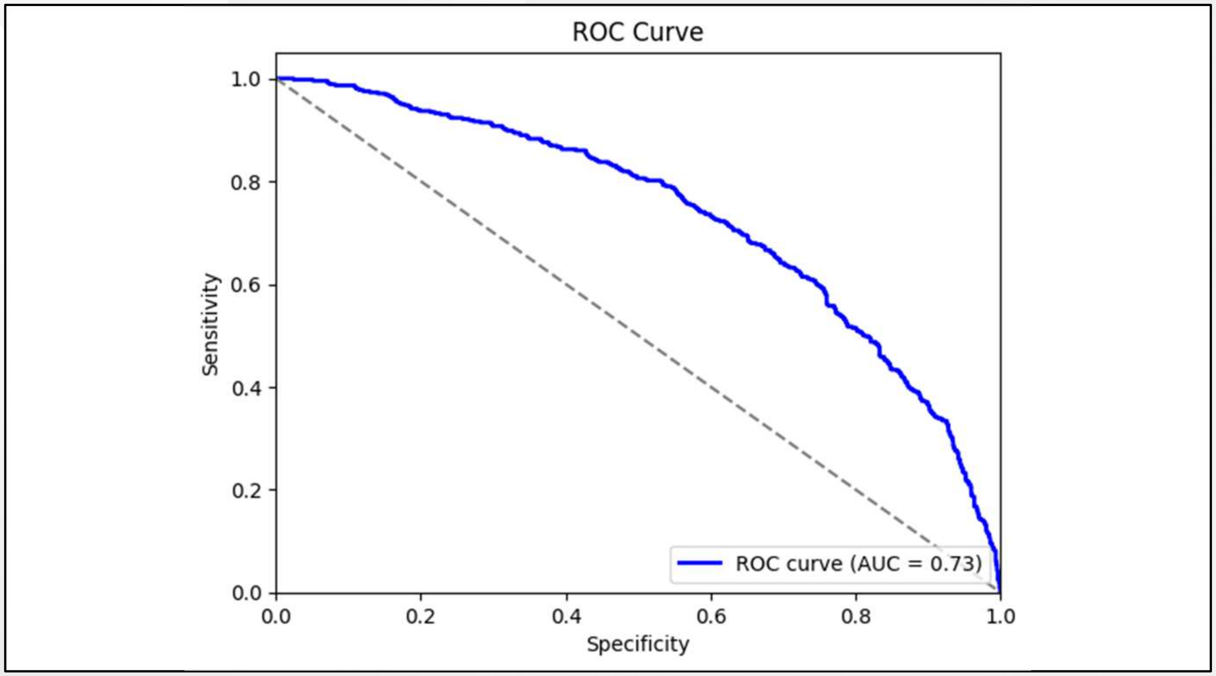
Optimal Cutoff: 0.203

Feed-Forward Neural Network



Feed-Forward Neural Network Cont.

Large Variable Set

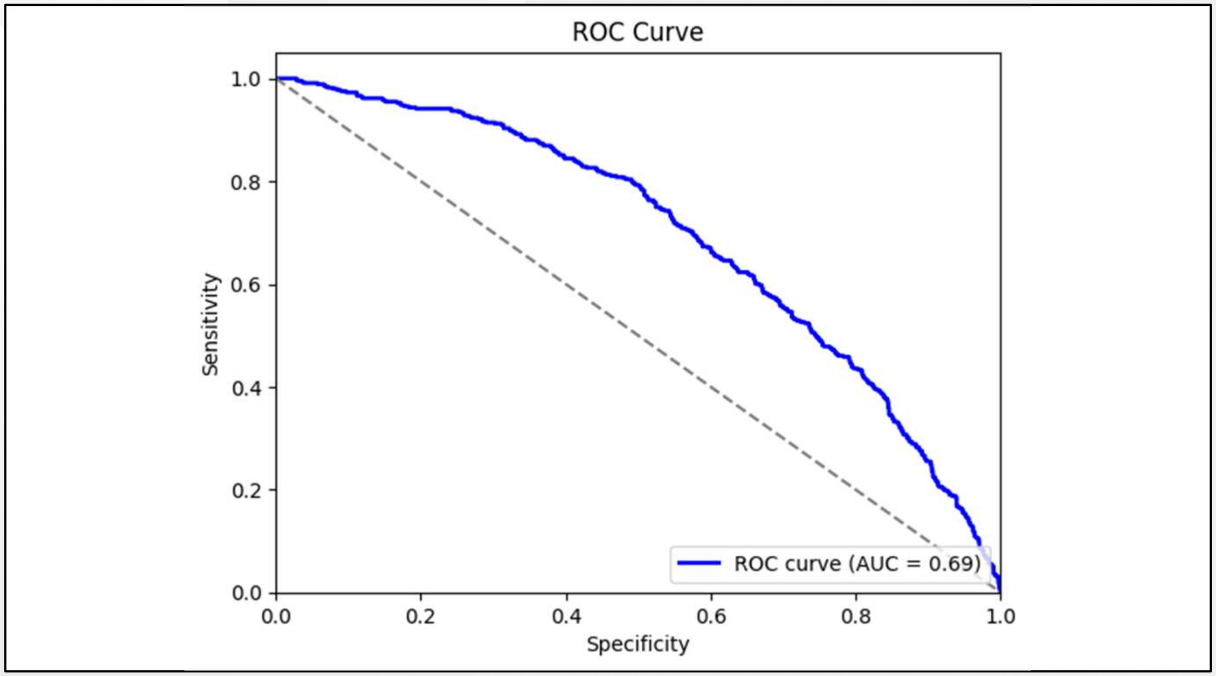


Accuracy	**0.740**
Sensitivity	**0.640**
Specificity	**0.840**
AUC	**0.733**

Optimal Cutoff: 0.398

Feed-Forward Neural Network Cont.

Small Variable Set



Accuracy	**0.680**
Sensitivity	**0.800**
Specificity	**0.560**
AUC	**0.691**

Optimal Cutoff: 0.335

Results

	Large Logistic	Small Logistic	Large NN	Small NN
Accuracy	0.633	0.643	**0.740**	0.680
Sensitivity	0.612	0.674	0.640	**0.800**
Specificity	0.653	0.612	**0.840**	0.560
AUC	0.719	0.709	**0.733**	0.691

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Discussion


Strong winds. (Delbert, 2022)


Former wetland near Tulelake, California. (NPR, 2022)

Conclusion

- *Summary*
  - Neural networks outperformed logistic regression models
  - Subset approach disregards time series element
- *Future work*
  - Other neural networks
  - New variable combinations and selection methods
  - Need for complementary ignition model

Fighting a wildfire. (WHO, 2024)

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Thank you!

Questions?

References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015 [Dataset]. Climatology Lab. https://www.climatologylab.org/terraclimate.html
- Esri. (2024). California Multi-Source Vegetation Layer [Data set]. ArcGIS. Retrieved May 20, 2024, from https://www.arcgis.com/home/item.html?id=b7ec5d68d8114b1fb2bfbf4665989eb3
- Halofsky, J. E., Peterson, D. L., & Harvey, B. J. (2020). Changing wildfire, changing forests: The effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. Fire Ecology, 16(4). https://doi.org/10.1186/s42408-019-0062-8
- Office of Environmental Health Hazard Assessment [OEHHA]. (2022). Wildfires. In *Indicators of Climate Change in California*. California Environmental Protection Agency. https://oehha.ca.gov/media/downloads/climate-change/document/04wildfires.pdf
- Parisien, M.A., & Moritz, M. A. (2009). Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecological Monographs, 79*(1), 127–154. https://doi.org/10.1890/07-1289.1
- Pham, K., Ward, D., Rubio, S., Shin, D., Zlotikman, L., Ramirez, S., ... & Jiang, X. (2022, December). California wildfire prediction using machine learning. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 525-530). IEEE.
- Satir, O., Berberoglu, S., & Donmez, C. (2016). Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. Geomatics, Natural Hazards and Risk, 7(5), 1645-1658. https://doi.org/10.1080/19475705.2015.1084541
- Short, K. C. (2022). Spatial wildfire occurrence data for the United States, 1992-2020 (FPA_FOD_20221014) (6th ed.). Fort Collins, CO: Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2013-0009.6

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY®

:)

:)

Data

-12 months for 29 years; 348 months; Jan 1992-Dec 2020

-217,500 total rows of observations (348 x 625)

-25x25 region = 625 cells (16 km^2)

-region with habitat variability; has enough fires to be relevant to study

-near LA and important parks

:)

aet: (Actual Evapotranspiration, monthly total), units = mm   water leaving soil
def: (Climate Water Deficit, monthly total), units = mm   pet minus aet
pet: (Potential evapotranspiration, monthly total), units = mm   water could transpire
ppt: (Precipitation, monthly total), units = mm
q: (Runoff, monthly total), units = mm   not absorbed by soil
soil: (Soil Moisture, total column - at end of month), units = mm   water in soil
srad: (Downward surface shortwave radiation), units = W/m2   sunlight
tmax: (Max Temperature, average for month), units = C
tmin: (Min Temperature, average for month), units = C
vap: (Vapor pressure, average for month), units = kPa   humidity
ws: (Wind speed, average for month), units = m/s
vpd: (Vapor Pressure Deficit, average for month), units = kpa   dryness
PDSI: (Palmer Drought Severity Index, at end of month), units = unitless   neg=dry
fires: total number of fires, count
fire_total: sum of fire area in given month and given cell
habitat: specific habitat classification (most common specific habitat in that cell)
habitat_g: general habitat classification (most common general habitat in that cell)
cell: concatenated lon0 and lat0 with comma separator
fire_events: 1 when at least one natural fire; 0 when no natural fire
fire_spread: 1 when fire area >0.1 acres; 0 when smaller or none

FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

Logistic including lag anomalies

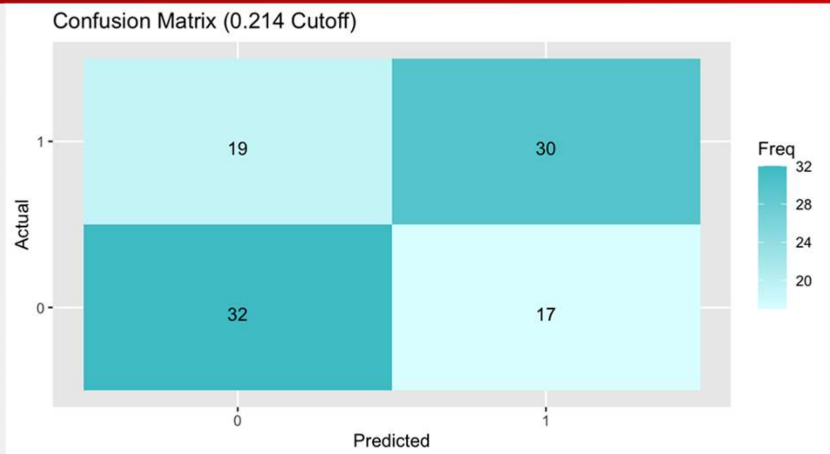
-90-10 train-test split
-all variables; reduce based on p-val >0.05
-remove:
#"q"
#"diff_aet"
#"diff_ppt"
#"diff_q"
#"diff_soil"
#"diff_PDSI"
#"anom_q"
#"anom_srad"
#"lag_anom_srad"
-reduce with stepAIC to:
#Step: AIC=2028.84
aet + def + ppt + srad + vap + ws + vpd + PDSI + diff_pet + diff_tmin + diff_vap + diff_vpd + anom_aet + anom_def + anom_pet + anom_ppt + anom_vap + anom_ws + anom_vpd + anom_PDSI + lag_anom_def + lag_anom_pet + lag_anom_vap
-optcutoff on train data;
confusion/accuracy/pred for test data

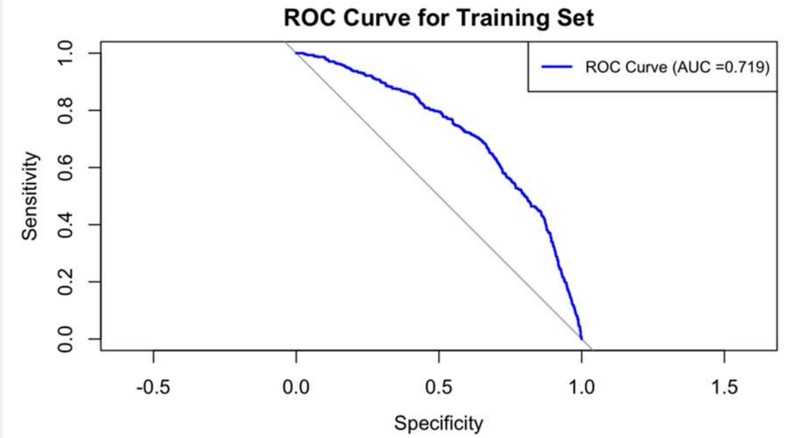
-STEP model has higher accuracy, sensitivity
-FULL model has higher specificity, AUC

Logistic including lag anomalies (LARGE)

-90-10 train-test split
-all variables; reduce based on p-val >0.05
-remove:
#"q"
#"diff_aet"
#"diff_ppt"
#"diff_q"
#"diff_soil"
#"diff_PDSI"
#"anom_q"
#"anom_srad"
#"lag_anom_srad"

-full model results:
-using optcutoff (YJS): 0.2136121
#Accuracy : 0.6327
#Sensitivity : 0.6122
#Specificity : 0.6531
#AUC : 0.719

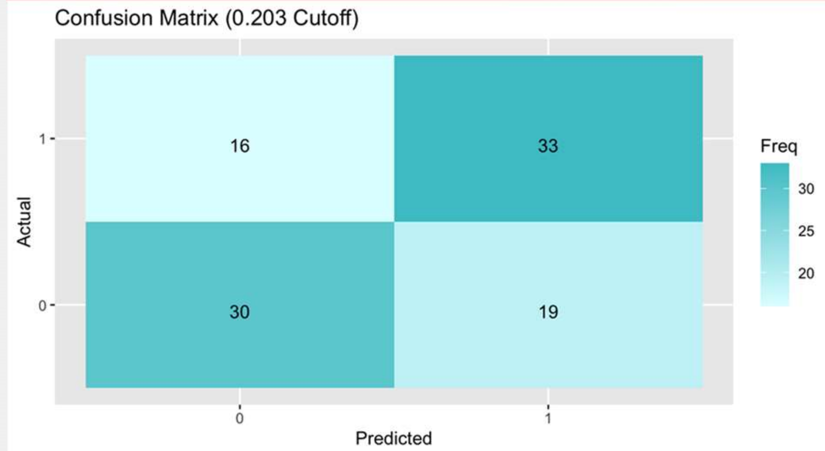


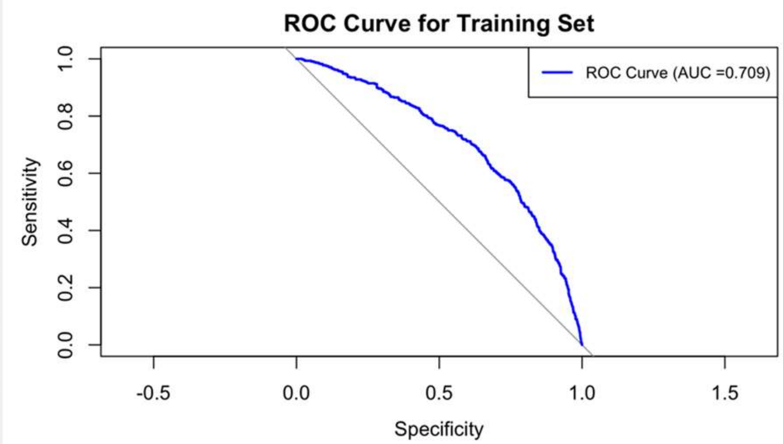


Logistic including lag anomalies (SMALL)

-90-10 train-test split
-all variables; reduce based on p-val >0.05
-remove:
#"q"
#"diff_aet"
#"diff_ppt"
#"diff_q"
#"diff_soil"
#"diff_PDSI"
#"anom_q"
#"anom_srad"
#"lag_anom_srad"
-reduce with stepAIC to:
#Step: AIC=2028.84
aet + def + ppt + srad + vap + ws + vpd + PDSI + diff_pet + diff_tmin + diff_vap + diff_vpd + anom_aet + anom_def + anom_pet + anom_ppt + anom_vap + anom_ws + anom_vpd + anom_PDSI + lag_anom_def + lag_anom_pet + lag_anom_vap
-optcutoff on train data;
confusion/accuracy/pred for test data

-step model results:
-using optcutoff (YJS): 0.2031919
#Accuracy : 0.6429
#Sensitivity : 0.6735
#Specificity : 0.6122
#AUC : 0.709



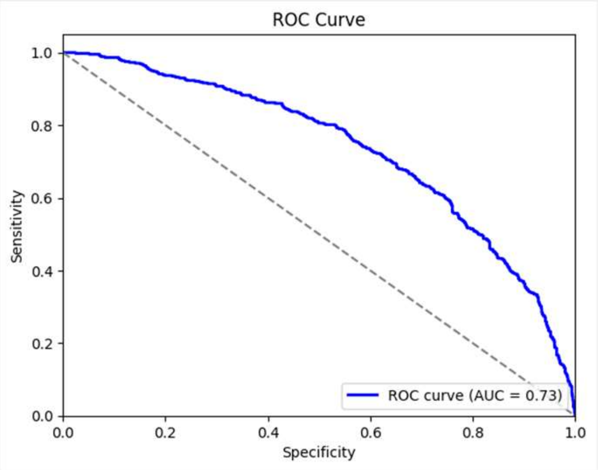


NN including lag anomalies (LARGE)

-90-10 train-test split
-all variables; reduce based on p-val >0.05
-remove:
#"q"
#"diff_aet"
#"diff_ppt"
#"diff_q"
#"diff_soil"
#"diff_PDSI"
#"anom_q"
#"anom_srad"
#"lag_anom_srad"

-activation: leaky_relu
-5 layers: 80, 60, 40, 20, 1
-binary focal cross-entropy (alpha=0.1)
-optimizer: adam
-normalized
-30 epochs, batch size 10

-full model results:
-using optcutoff (YJS): 0.39797372
#Accuracy : 0.740
#Sensitivity : 0.640
#Specificity : 0.840
#AUC : 0.733



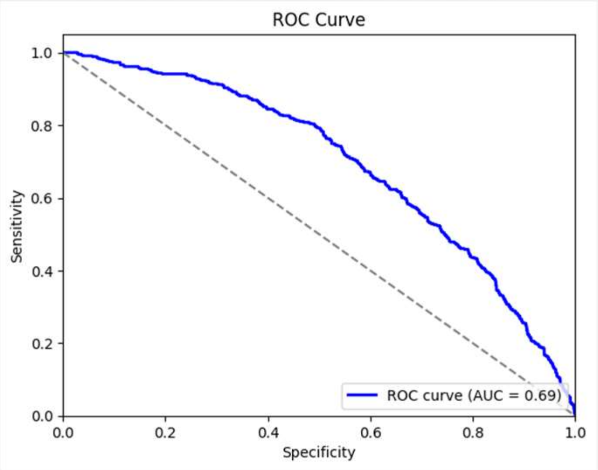
FLORIDA TECH
FLORIDA'S STEM UNIVERSITY

NN including lag anomalies (SMALL)

-90-10 train-test split
-all variables; reduce based on p-val >0.05
-remove:
#"q"
#"diff_aet"
#"diff_ppt"
#"diff_q"
#"diff_soil"
#"diff_PDSI"
#"anom_q"
#"anom_srad"
#"lag_anom_srad"
-ALSO reduce with stepAIC to:
#Step: AIC=2028.84
aet + def + ppt + srad + vap + ws + vpd + PDSI + diff_pet + diff_tmin + diff_vap + diff_vpd + anom_aet + anom_def + anom_pet + anom_ppt + anom_vap + anom_ws + anom_vpd + anom_PDSI + lag_anom_def + lag_anom_pet + lag_anom_vap
-optcutoff on train data;
confusion/accuracy/pred for test data

-full model results:
-using optcutoff (YJS): 0.33526808
#Accuracy : 0.680
#Sensitivity : 0.800
#Specificity : 0.560
#AUC : 0.691

-activation: leaky_relu
-5 layers: 80, 60, 40, 20, 1
-binary focal cross-entropy (alpha=0.1)
-optimizer: adam
-normalized
-30 epochs, batch size 10



FLORIDA TECH
FLORIDA'S STEM UNIVERSITY