IgniteAI

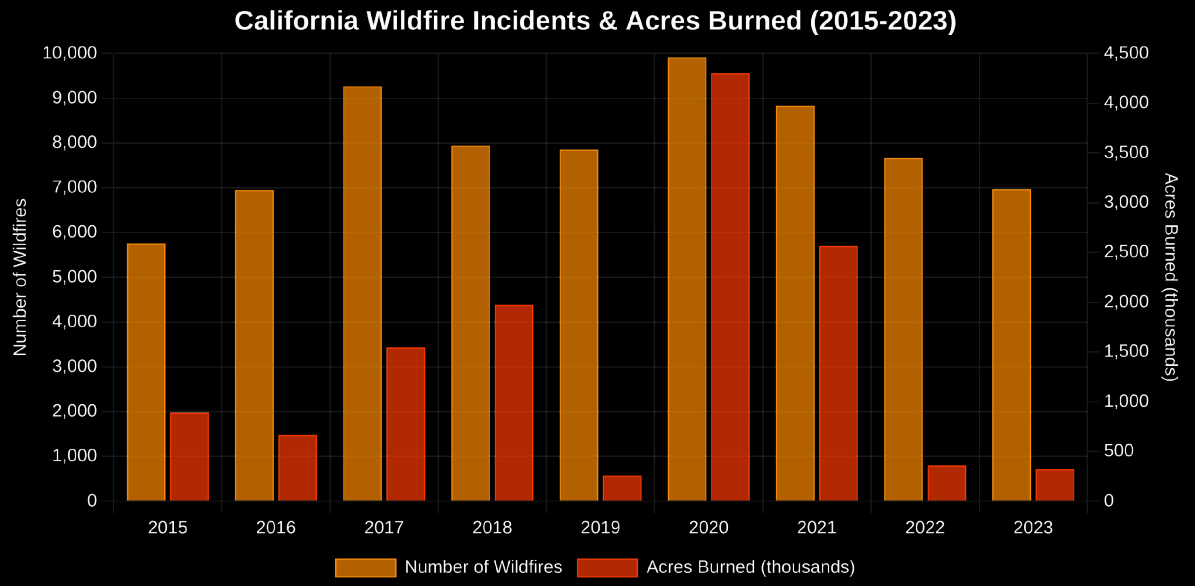
Predicting Wildfire Risk in California Using Machine Learning

Umair Habib | Austin Lee | Mukhammadali Fardaev | Bianka Paul

Introduction

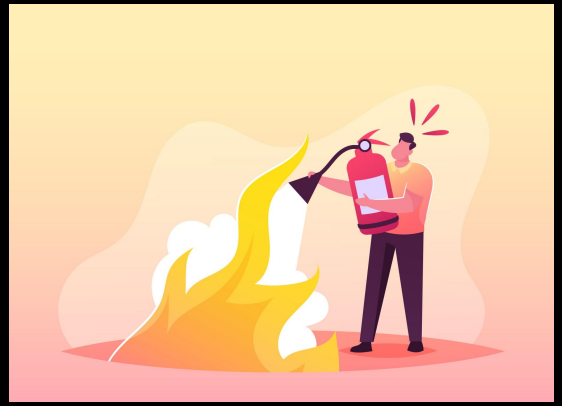
IgniteAI is focused on predicting wildfire risk in California using weather station data and historical fire events.

- Analyze 216,645+ climate records and 354,319+ fire events
- Develop predictive models for early wildfire detection
- Support resource allocation and policy decisions
- Create reproducible framework for environmental prediction



Motivation

Wildfire prediction presents both an urgent environmental challenge and a complex machine learning problem.



- **Environmental Impact**

  Increasing frequency and severity of wildfires threatens ecosystems and communities

- **Technical Challenge**

  Complex interplay of meteorological factors requires advanced ML approaches

- **Real-world Application**

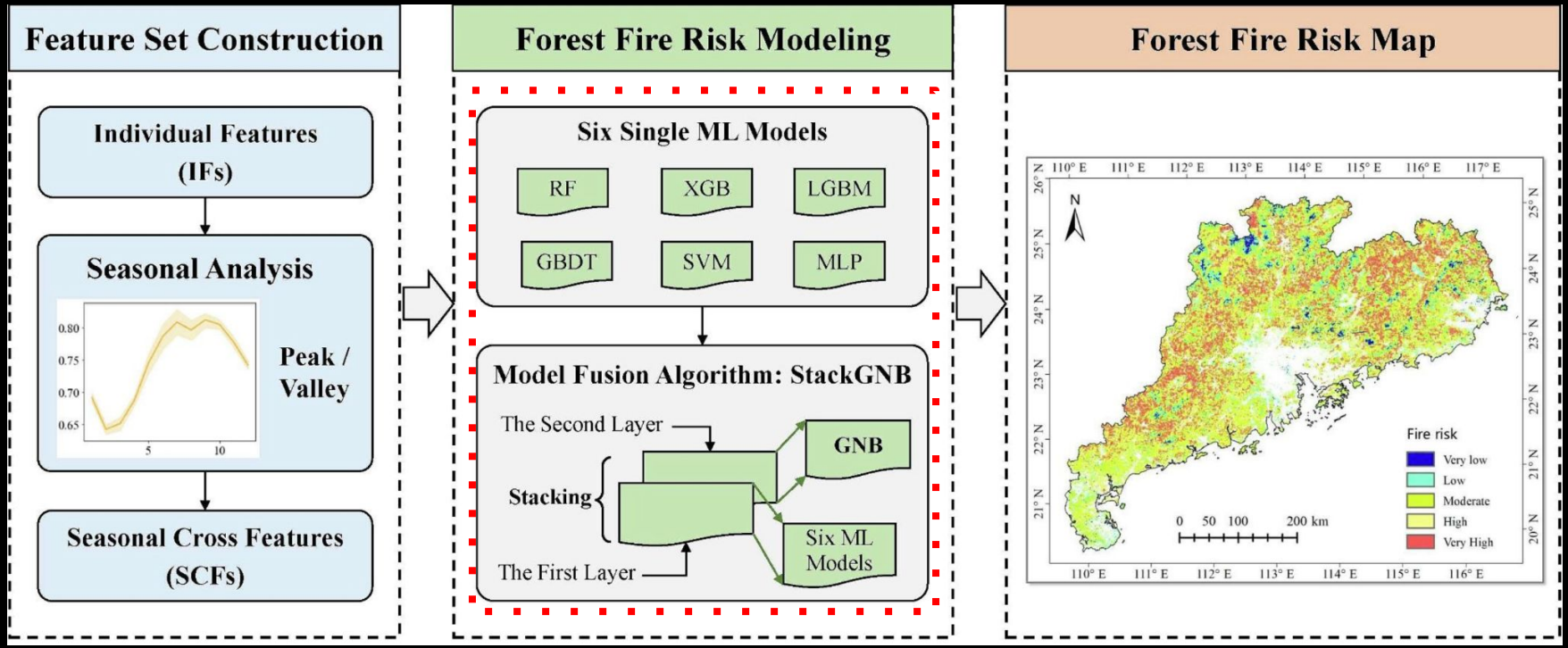
  Supports risk assessment, resource allocation, and policy decisions

Related Work

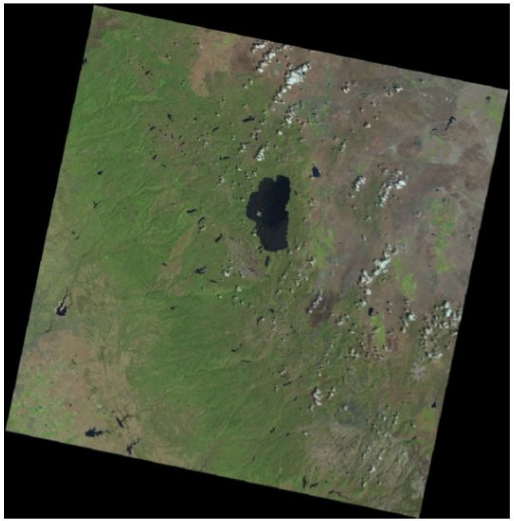
COMPANY

You could enter a subtitle here if you need it

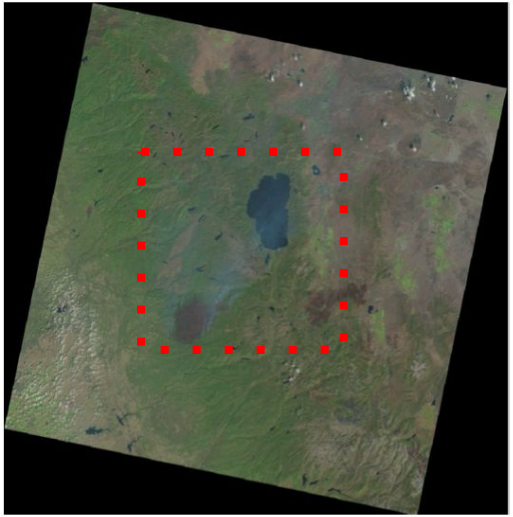
Stacking models



Predicting California wildfire risk with deep neural networks



(a) Satellite photo centered on South Lake Tahoe one month before the 2021 Caldor Fire (grassy lower left)



(b) Satellite photo centered on South Lake Tahoe during the 2021 Caldor Fire (smoke and burn region clearly visible)

Figure 1: Two dataset images

Takeaways

Class imbalance major obstacle for wildfires

Imagenet not mapped onto actual wildfires

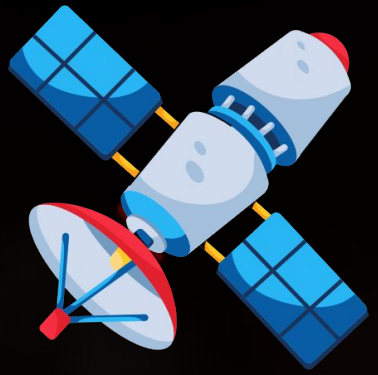
Accuracy	Precision	Recall
0.733	0.04	0.004

Data

COMPANY

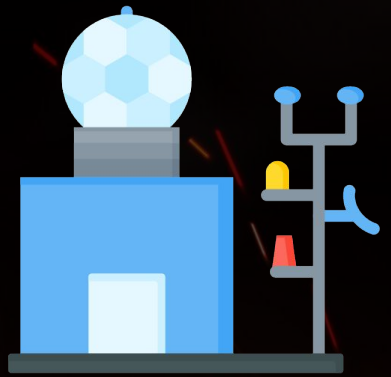
You could enter a subtitle here if you need it

Period: 2021-2024



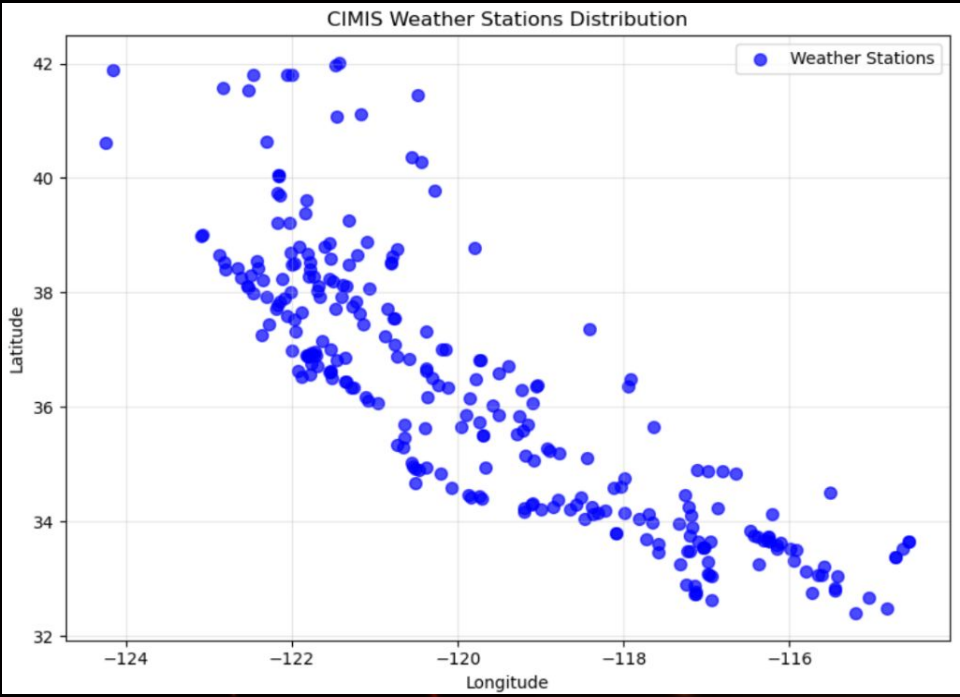
VIIRS Satellite Fire Data

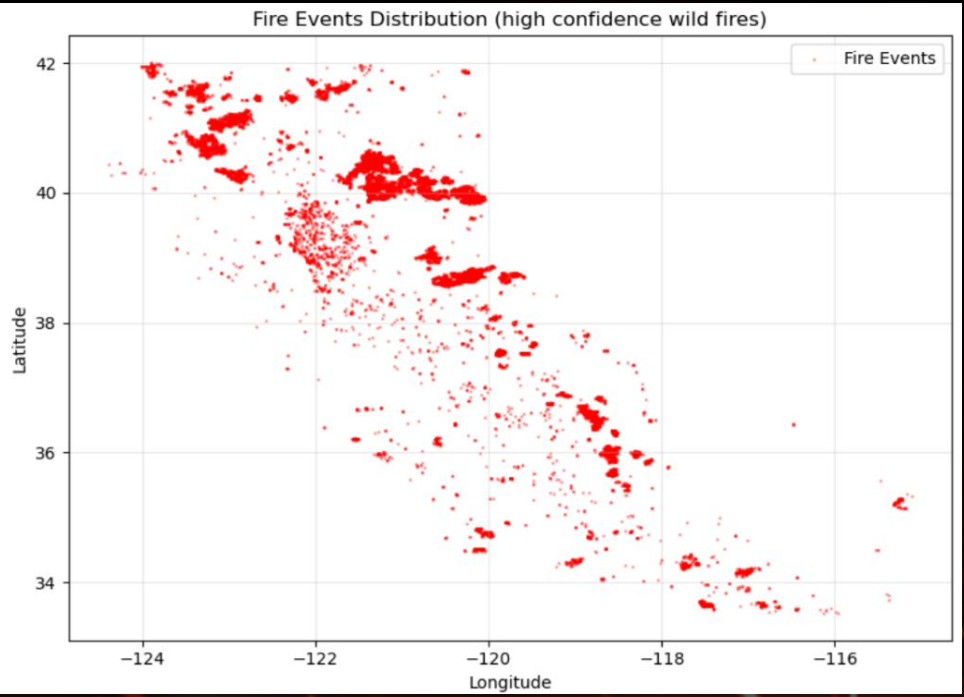
354,319 Fire events recorded across California



CIMIS Weather Station Data

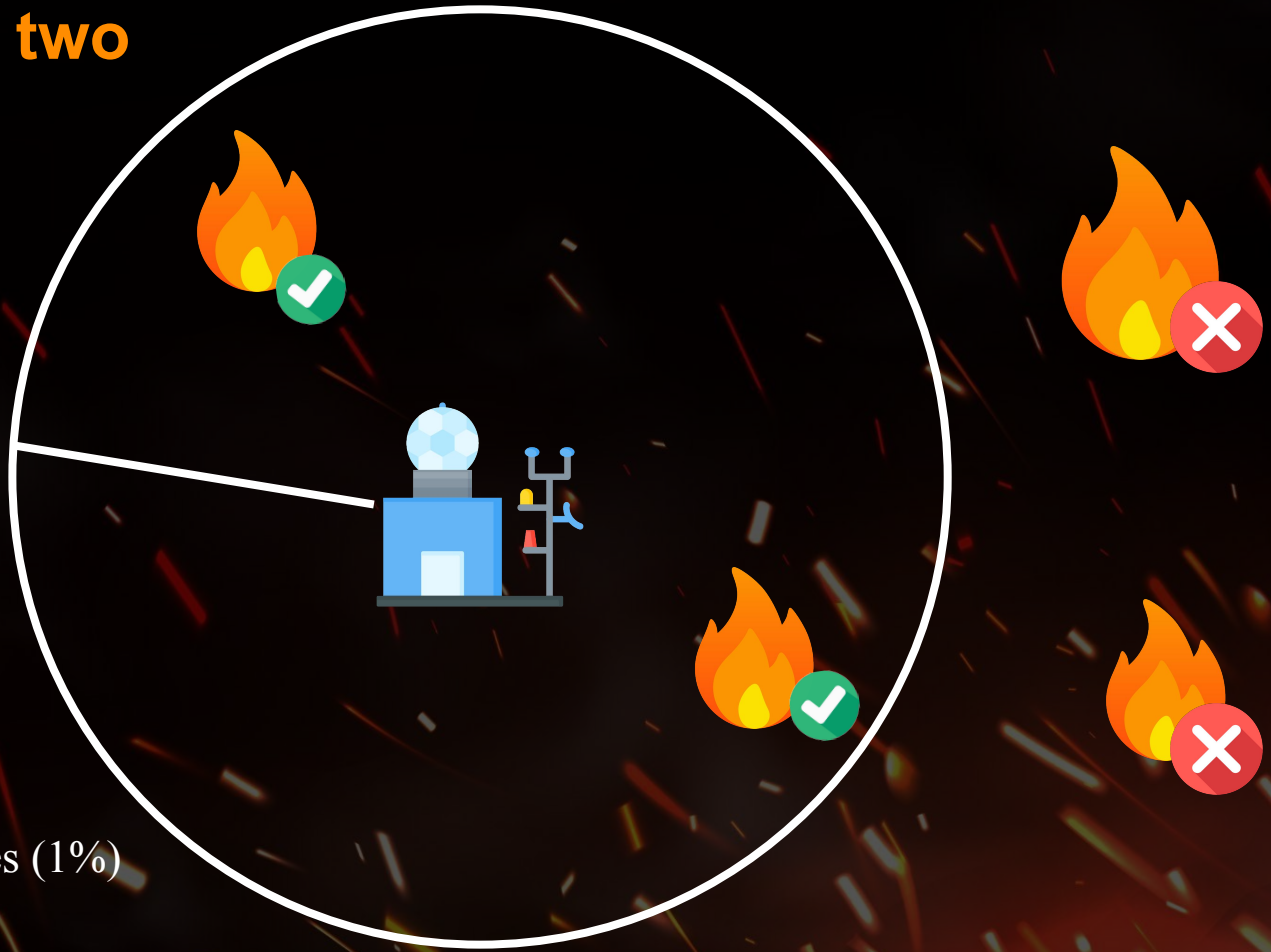
216,645 Climate records from California weather stations


CIMIS Weather Stations Distribution
Weather Stations
Latitude
Longitude
42
40
38
36
34
32
−124
−122
−120
−118
−116


Fire Events Distribution (high confidence wild fires)
Fire Events
Latitude
Longitude
42
40
38
36
34
−124
−122
−120
−118
−116

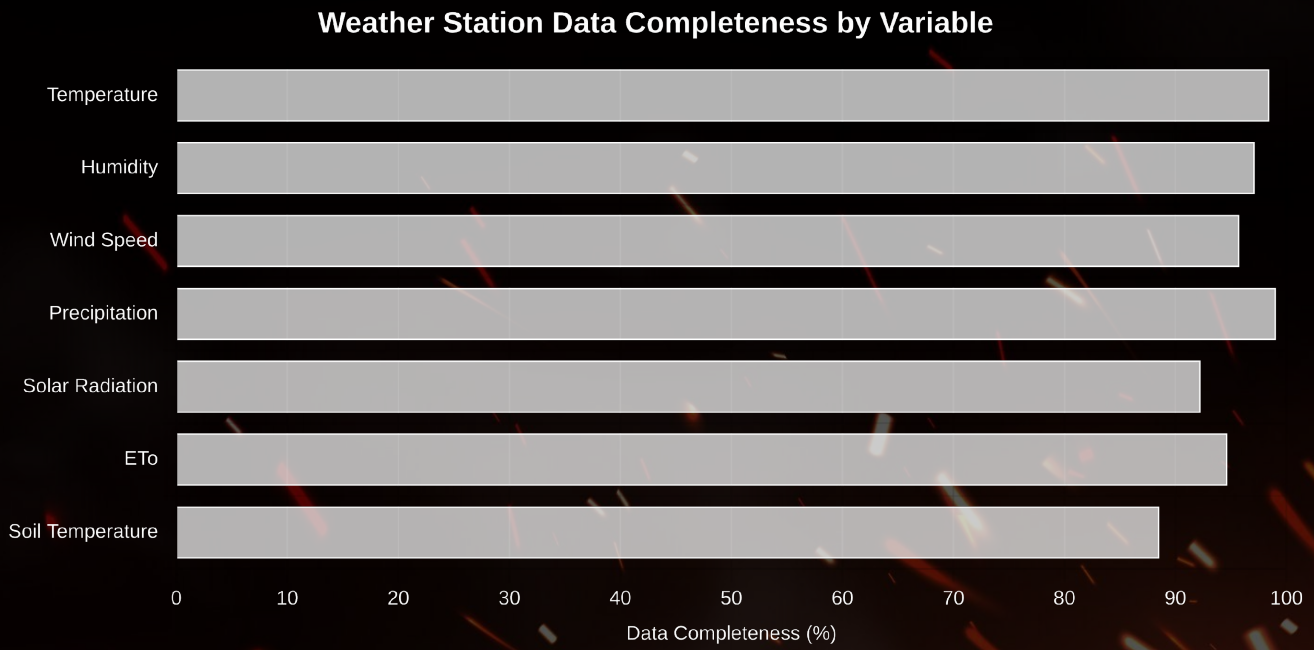
Combining two datasets

20 miles



- 202,732 rows
- 15 columns
- 2,367 wildfires (1%)

Features



- 7 day trailing average
- prior day result

Models

COMPANY

You could enter a subtitle here if you need it

Base logistic regression model fits

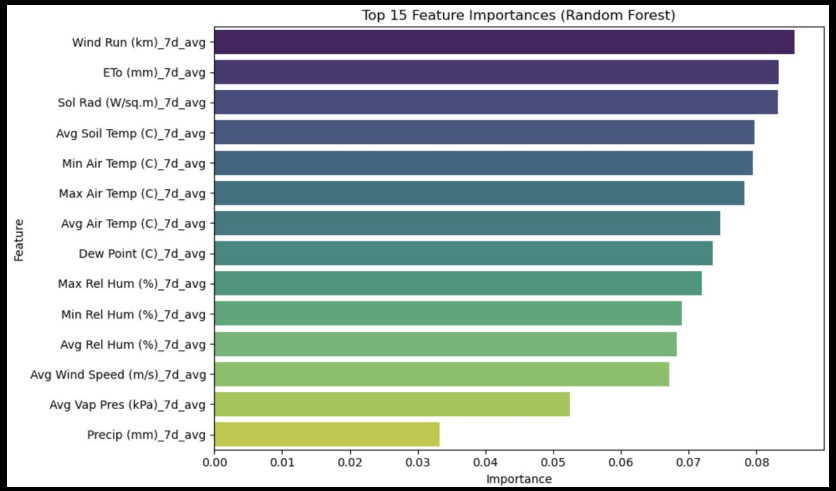
Full Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	0.99	36298
1	0.00	0.00	0.00	436
accuracy			0.99	36734
macro avg	0.49	0.50	0.50	36734
weighted avg	0.98	0.99	0.98	36734

- Accuracy rate is extremely high due to majority ‘no fire’ class prediction.
- The model can not predict any wildfire occurrences.
- There is need for more complex models.

Can’t see the fire through the trees

	precision	recall	f1-score	support
0	0.99	1.00	0.99	36307
1	1.00	0.05	0.10	427
accuracy			0.99	36734
macro avg	0.99	0.53	0.55	36734
weighted avg	0.99	0.99	0.98	36734



At a wildfire precision of 100%, we may reliably act when the model predicts wildfire

However, at a recall at 5%, we are only able to action 5% of the time

For real world application of predicting wildfires, we would actually prefer higher recall in which we would potentially bear the cost of false alarms over being unprepared in the event of a wildfire.

Neural network unlocks additional wildfire predictions but at a cost

Metric	Value
Hidden Layers	[50, 25, 10, 10]
Activation	ReLU
Optimizer	Adam (LR=0.001)
Output Layer	Sigmoid
Loss Function	Binary Cross Entropy

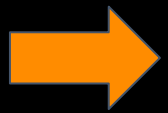
```
Threshold: 0.015
[[30925  5382]
 [  268   159]]
              precision    recall  f1-score   support

           0       0.99      0.85      0.92     36307
           1       0.03      0.37      0.05       427

    accuracy                           0.85     36734
   macro avg       0.51      0.61      0.48     36734
weighted avg       0.98      0.85      0.91     36734
```

≠

Despite adding depth and class weighting to our neural network, the model struggled to learn meaningful wildfire patterns – achieving only 54% accuracy.

Metric	Value
Hidden Layers	**[100, 50, 25, 25, 10, 10]**
Activation	ReLU
Optimizer	Adam (LR=0.001)
Output Layer	Sigmoid
Loss Function	Binary Cross Entropy

Class Weights	'Balanced'



Even with six hidden layers, adaptive learning, and a balanced loss function, the model performed just slightly better than random chance

Conclusion

Study Summary

- DataSet
  - 202,732 observations
  - 2,367 fire events
  - 1.1% class imbalance
- 4 ML Models used

Key Result

- Random Forest (best performer)
- Positive predictive capabilities
- Extreme class imbalance

Limitations & Future Works

- **Current:** Temporal constraints, 20-mile spatial resolution, class imbalance
- **Future:** Longer historical data, topographical features, fire weather indices
- **Goal:** Enhanced prediction accuracy for operational fire management deployment