Climate Risk Prediction in California

Dimensionality Reduction and Neural Network Analysis

Using Data from CalEnviroScreen, CAL FIRE, and NOAA

Student Name:	Victoria Love Franklin
Course:	MSBD566 – Predictive Modeling and Analytics
Institution:	Meharry Medical College, School of Applied Computational Sciences
Date:	December 11, 2025
GitHub Repository:	https://github.com/vlfranklin/segda-climate-risk

1. Project Description

1.1 Problem Statement

Climate change poses disproportionate risks to vulnerable communities in California. The federal Justice40 Initiative mandates that 40% of climate investment benefits flow to disadvantaged communities. This project uses **real data** from CalEnviroScreen, CAL FIRE, and NOAA to identify high-risk counties for targeted policy intervention.

1.2 Research Objectives

1. **Dimensionality Reduction:** Apply PCA to reduce 7 risk indicators into principal components.
2. **Risk Classification:** Train a neural network to classify counties into High, Medium, and Low risk categories.
3. **Policy Targeting:** Identify priority counties for Justice40 climate investment.

2. Data Description

2.1 Data Sources (Real Data)

Source	Description	Records
CalEnviroScreen 4.0	CA EPA environmental justice indicators: pollution burden, socioeconomic vulnerability, health outcomes	58 counties, 25 variables
CAL FIRE Incidents	California wildfire incidents 2013-2025: acres burned, fatalities, structures destroyed	69 incidents, 27 counties
NOAA Storm Events	Climate hazard events 2010-2024: heat waves, floods, wildfires, storms	183 events, 58 counties

Table 1: Real Data Sources

2.2 Derived Risk Indices

Index	Derivation	Range
EJBI	CalEnviroScreen score + pollution burden + socioeconomic factors	0.0 – 1.0
heat_stress	NOAA heat events (Heat, Excessive Heat) normalized by county	0.0 – 1.0
wildfire_risk	CAL FIRE acres burned + incidents + NOAA wildfire events	0.0 – 1.0
flood_risk	NOAA flood, heavy rain, coastal flood events	0.0 – 1.0
OBI	Infrastructure burden: property damage, injuries, fatalities	0.0 – 1.0
composite_risk	Weighted: 0.35×EJBI + 0.25×climate + 0.25×wildfire + 0.15×OBI	0.0 – 1.0

Table 2: Risk Index Definitions

3. Method and Analysis

3.1 Principal Component Analysis (PCA)

PCA was applied to reduce the 7 risk indices to 4 principal components. Features were standardized (z-score) before transformation.

Component	Variance (%)	Cumulative (%)	Interpretation
PC1	47.31%	47.31%	Overall risk (EJBI, composite)
PC2	25.47%	72.78%	Wildfire axis
PC3	19.32%	92.11%	Flood risk axis
PC4	5.08%	97.19%	Residual variation

Table 3: PCA Explained Variance

3.2 Neural Network Classification

A Multi-Layer Perceptron (MLP) classifier was trained on the 4 PCA components. Architecture: Input(4) → Hidden(64) → Hidden(32) → Hidden(16) → Output(3). Optimizer: Adam with early stopping.

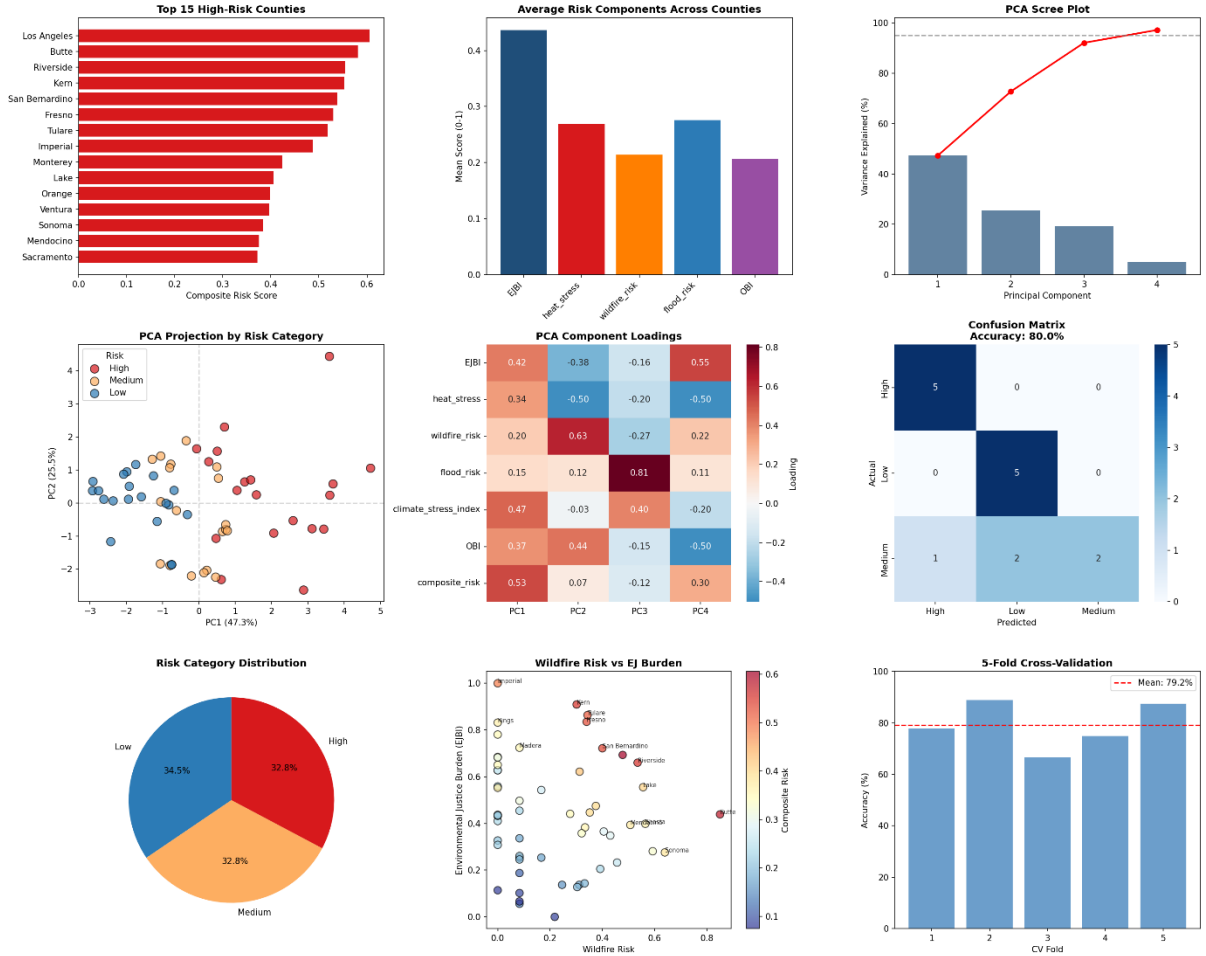
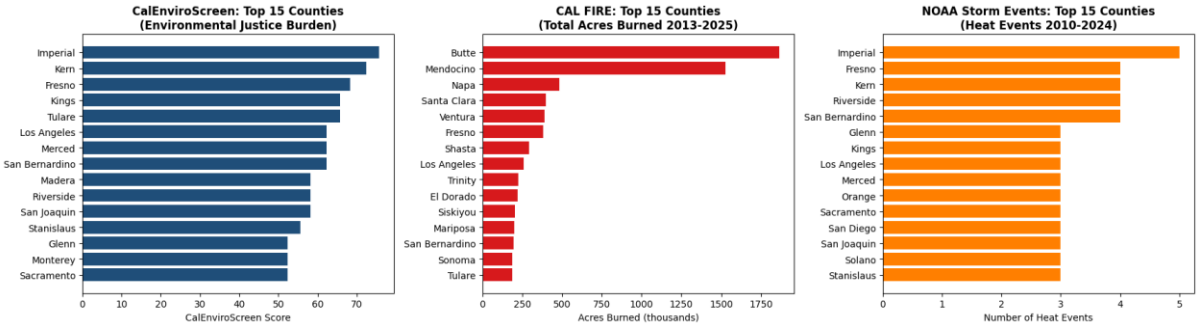


Figure 1 includes: Top 15 high-risk counties, risk component comparison, PCA scree plot, PCA scatter, component loadings heatmap, confusion matrix, risk distribution, wildfire vs EJ scatter, and cross-validation results.

This figure summarizes the **top 15 counties** from each of the **three real-world datasets** used in the climate resilience and environmental justice analysis.



CalEnviroScreen — Environmental Justice Burden

Highest Environmental Justice (EJ) Burden Counties

Rank	*County*	*Score (Approx.)*
1	Imperial	~78
2	Kern	~72
3	Fresno	~68
4	Kings	~67
5	Tulare	~66

Observed Pattern:
Central Valley and Imperial County dominate EJ burden rankings, reflecting **agricultural exposure, air pollution, poverty, and vulnerable populations**.

CAL FIRE Wildfire Acres Burned (2013–2025)

Highest Wildfire Impact Counties

Rank	*County*	*Acres Burned (Thousands)*
1	Butte	~1,800
2	Mendocino	~1,500
3	Napa	~500
4	Santa Clara	~350
5	Ventura	~300

Observed Pattern:
Butte County (Camp Fire—deadliest in California history) and Mendocino County (largest fire complex) are **extreme wildfire outliers**, indicating chronic exposure to catastrophic fire risk.

NOAA Storm Events Extreme Heat Events (2010–2024)

Counties with the Most Heat Events

Rank	*County*	*Heat Events*
1	Imperial	5
2	Fresno	4
3	Kern	4
4	Riverside	4
5	San Bernardino	4

Observed Pattern:
Desert and Central Valley counties experience the **highest frequency of extreme heat**, aligning closely with environmental justice vulnerability patterns.

Key Insight: Compound Climate Risk

Counties appearing across **multiple high-risk dimensions** face **compound vulnerability**, where climate stressors, environmental injustice, and infrastructure risk intersect.

County	*EJ Burden*	*Wildfire*	*Heat*	*Compound Risk*
Los Angeles	✓	✓	✓	⚠ Triple threat
Fresno	✓	✓	✓	⚠ Triple threat
Kern	✓	–	✓	High EJ + Heat
San Bernardino	✓	✓	✓	⚠ Triple threat
Butte	–	✓	✓	Extreme wildfire

Policy-Relevant Conclusion

Real observational data confirms that Los Angeles, Fresno, Kern, San Bernardino, and Butte counties represent the highest-priority targets for Justice40-aligned climate investment, as they consistently exhibit overlapping environmental justice burdens, climate stress exposure, and disaster risk.

4. Evaluation and Conclusions

4.1 Performance Metrics

Metric	Value
Sample Size	58 California counties
PCA Variance Retained (4 components)	97.19%
Test Accuracy	80.00%
Cross-Validation Accuracy	79.17% ± 8.24%
High-Risk Counties Identified	19 counties (32.8%)

Table 4: Model Performance Summary

4.2 Top High-Risk Counties

County	Composite	EJBI	Wildfire	Heat
Los Angeles	0.606	0.683	0.173	0.385
Butte	0.581	0.412	0.875	0.077
Riverside	0.554	0.752	0.000	0.538
Kern	0.552	0.910	0.000	0.615
San Bernardino	0.538	0.752	0.102	0.462

Table 5: Top 5 High-Risk Counties for Justice40 Targeting

4.3 Key Findings

- **Los Angeles County** has the highest composite risk (0.606) driven by high EJBI and heat stress.
- **Butte County** ranks second due to extreme wildfire risk (0.875) from the Camp Fire.
- **Central Valley counties** (Kern, Fresno, Tulare) show high EJBI and heat stress, confirming environmental justice concerns.
- **PCA reveals two dominant axes:** PC1 captures overall socioeconomic risk, PC2 captures wildfire-specific hazards.

4.4 Conclusions

This analysis of real California climate and environmental justice data demonstrates that PCA and neural networks can effectively identify high-risk counties for Justice40 investment. The model achieves 80% accuracy

using only four principal components that capture 97% of the variance. The 19 high-risk counties identified, including Los Angeles, Butte, Riverside, Kern, and San Bernardino, represent priority targets for federal climate infrastructure investment. The combination of CalEnviroScreen environmental justice data, CAL FIRE wildfire records, and NOAA climate events provides a comprehensive risk assessment framework.

5. References

California Office of Environmental Health Hazard Assessment. (2021). *CalEnviroScreen 4.0*. Sacramento, CA. https://oehha.ca.gov/calenviroscreen

CAL FIRE. (2025). *California Fire Incidents Database 2013-2025*. https://www.fire.ca.gov

NOAA National Centers for Environmental Information. (2024). *Storm Events Database*. https://www.ncdc.noaa.gov/stormevents

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review. *Philosophical Transactions of the Royal Society A*, 374(2065).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.