Golisano College of Computing and Information Sciences

Spatiotemporal Wildfire Prediction using Evolutionary Algorithms

From Passive Risk Mapping to Active Swarm Deployment

Adit Dhall
MS Artificial Intelligence
Email: ad6449@rit.edu

Motivation & Research Question

- **Motivation:** Traditional grid-based risk models require exhaustive scanning (0(N)), which is too computationally expensive for autonomous drones with limited onboard processing.
- **Research Question 1(Hybrid Accuracy):** Can a heuristic Genetic Algorithm (GA) that combines physics with historical fires replicate the predictive accuracy of a Deep Learning model using only local, sparse sampling?
- **Research Question 2(Efficiency):** Does a physics informed evolutionary approach significantly reduce computational complexity compared to total area scanning.

Data Pipeline (GEE Augmentation)

- **Base Data:** Historical wildfire events (2016-2021) with meteoritical data were sourced from the US Wildfire Dataset on Kaggle **[4].**
- **Feature Selection:** Selected features based on established literature:
  - **Spatial (Lat/Lon):** Foundational for fire susceptibility modeling **[1].**
  - **Fuel (NDVI):** Vegetation density added via Google Earth Engine (MODIS), a critical driver of fire spread **[2].**
  - **Terrain (Slope/Elevation):** Topographic factors extended via SRTM DEM to model physical fire behavior **[3].**
- **Single Feature Vector:** This table demonstrates the integration of disparate data sources.

Feature	Value	Source
Latitude	39.5527	Base Dataset [4]
NDVI (Fuel)	0.2954	Google Earth Engine
Slope	22°	Google Earth Engine
ERC (Risk)	68	US Forest Service
Label	Fire(1)	Ground Truth

Table 1: Data Snippet – Single Feature Vector

Methodology: The Hybrid Model

- **Baseline Model:** A Random Forest Classifier trained on 11 features to generate "Ground Truth" risk map.
- **Genetic Algorithm:** Physics based fitness function to hunt for high risk zones.
  - **Fitness Function:** $\mathcal{J}_x = \mathrm{Fuel}_{\mathrm{NDVI}}^2 \times \mathrm{Weather}_{\mathrm{ERC}} \times \mathrm{Slope}_{\mathrm{Factor}} \times (1 + \mathrm{History\_weight} \cdot e^{-d})$

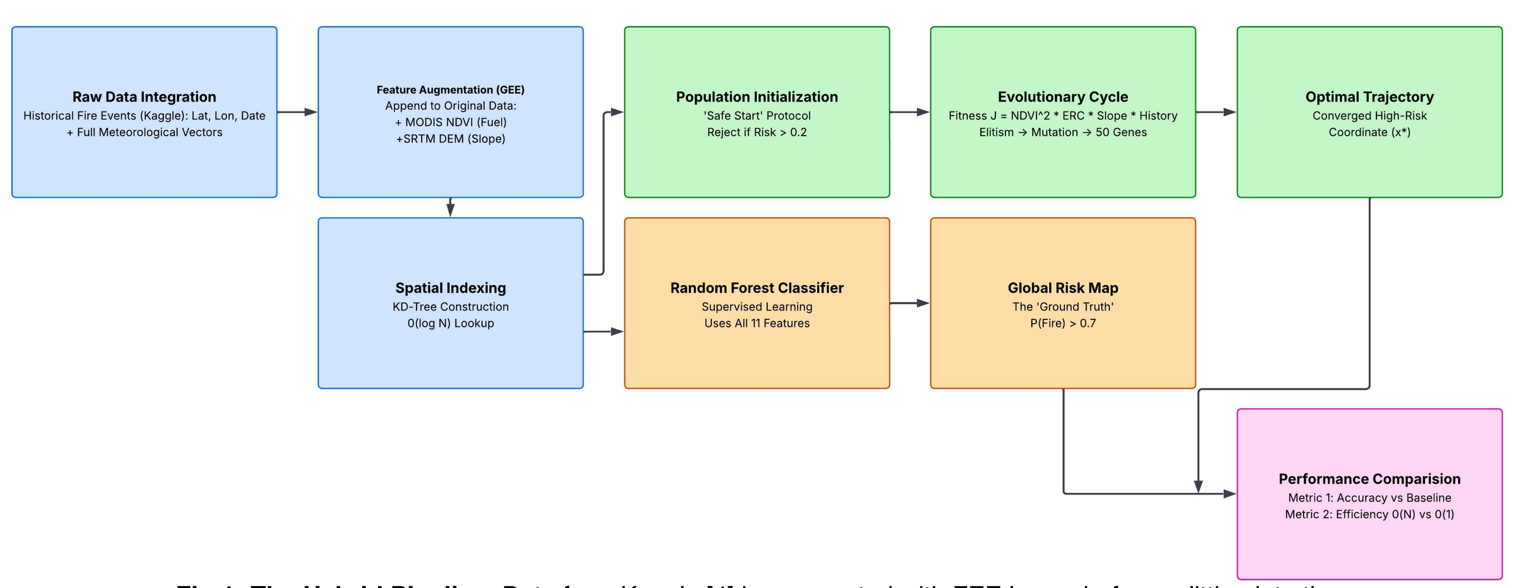


Fig 1: The Hybrid Pipeline. Data from Kaggle [4] is augmented with FEE layers before splitting into the Baseline (Random Forest) and the Heuristic (GA)

Parameter	Value	Role
Population Size	300	Ensures high-density search coverage.
Generation	50	Sufficient for convergence (stopping condition).
Mutation Rate	0.3	Aggressive variance to escape local optima.
Elitism	Top 3	Preserves the best found solution

Table 2: GA Hyperparameters

Results

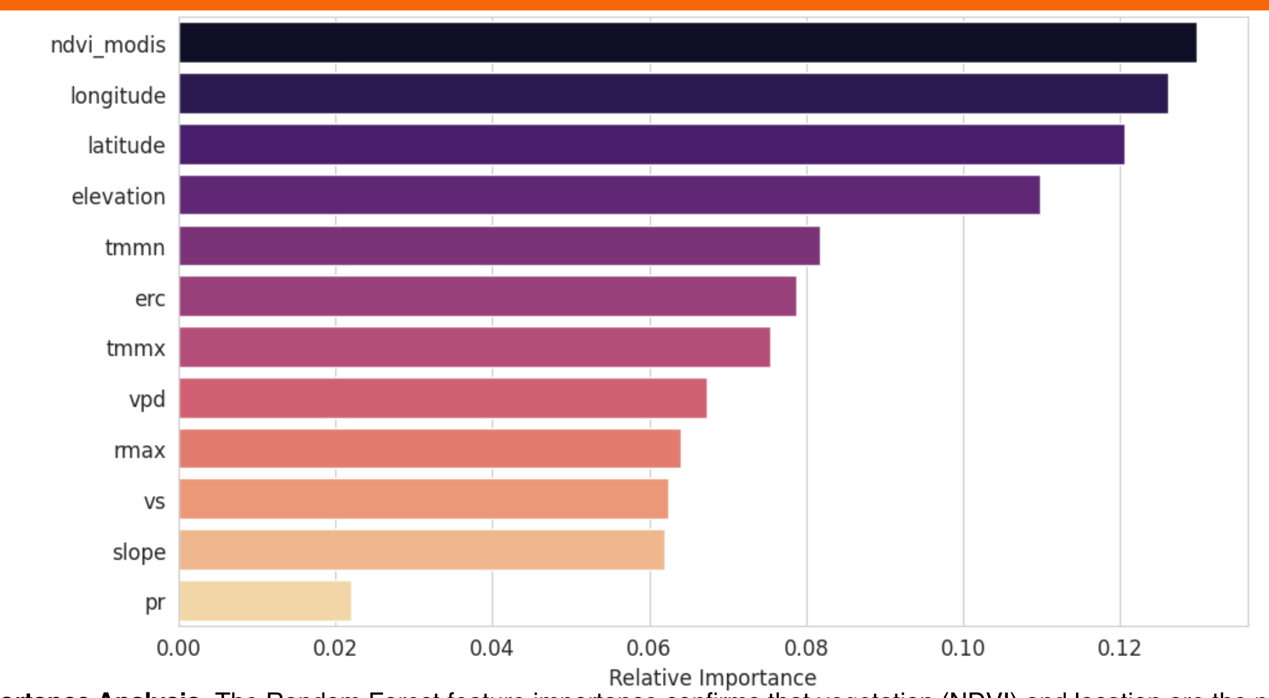


Fig 2: Feature Importance Analysis. The Random Forest feature importance confirms that vegetation (NDVI) and location are the primary drivers of fire risk, validating the weights used in the GA's physics equation.

RQ1 Answer: Predictive Accuracy (Macro vs Micro)

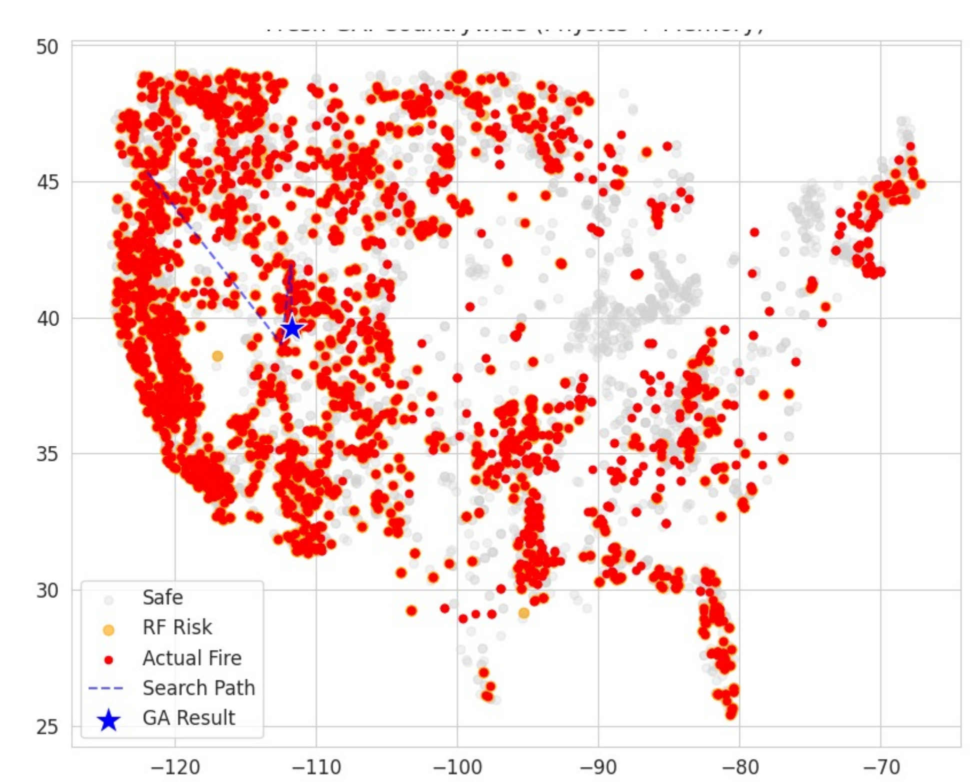
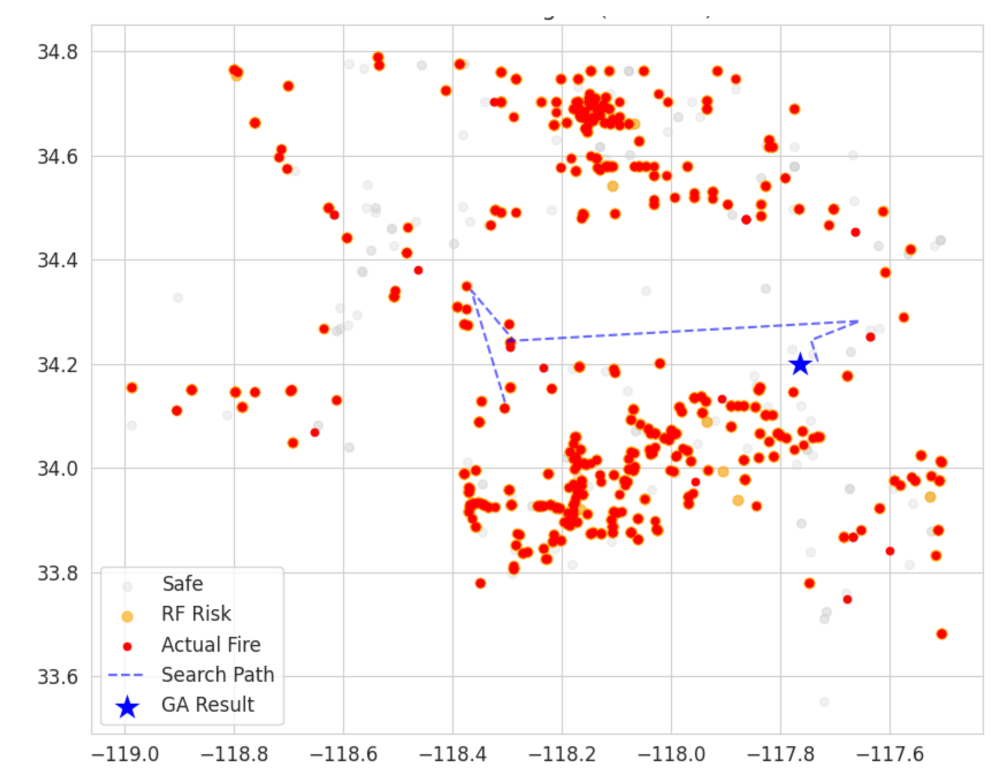


Fig 3: Macro Search Behavior. The GA (Blue Dashed Line) successfully navigates from a random safe start towards the high risk fire zone, guided solely by the fitness gradient



- **Fig 4: Micro Region Precision (LA Case Study).** While the Grid model highlights all possible risk areas (orange clouds), the GA identifies optimal target near the actual fire ignition point.

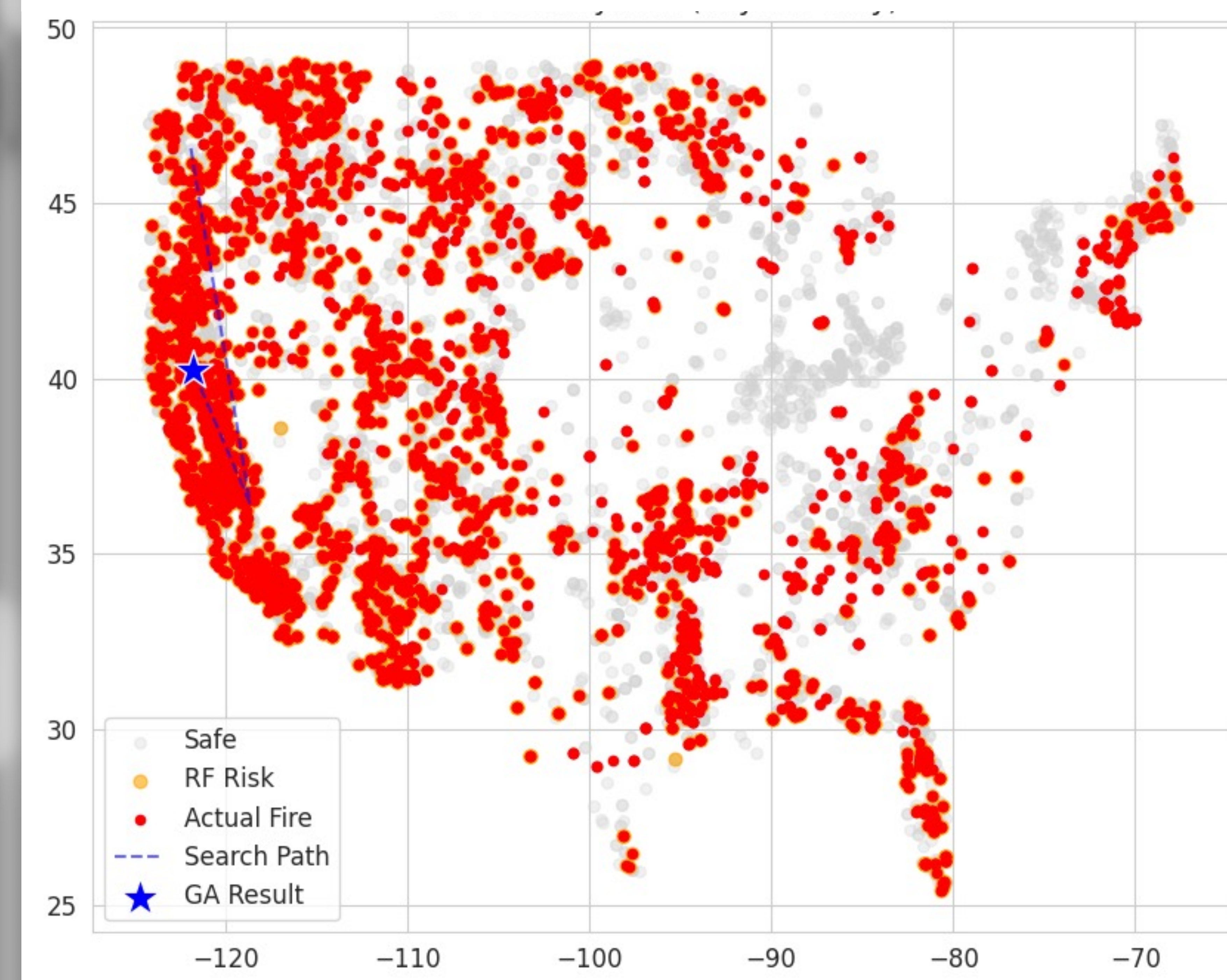
RQ2 Answer: Predictive Accuracy (Macro vs Micro)

Metric	Baseline(Random Forest)	Genetic Algorithm
Search Method	Exhaustive Grid Scan	Stochastic Heuristic Path
Complexity	0(N) (Scales with map size)	0(1) (Fixed by Parameters)
Evaluation	16,812 (Dataset Size)	15,000 (300 x 50)
Scalability	Linear Growth (Slown on large maps)	Constant (Fast on any map)

Table 3: Complexity Comparison

Discussion

- **Physics Informed Success:** The "Square Fuel" term (NDVI^2) in the fitness function successfully prevented the algorithm from getting trapped in sparce urban areas, which is a common issue in heuristic search **[1]**
- **Ablation Study:** Tested a variant with zero historical bonus in the fitness function, relying solely on physics (Fuel/Slope)
  - Results: As shown in **fig 5** below, the agent still converged on the fire zone/ high risk zone. This proves the system detects actual fire and is not reaching high risk zone due to historical fire data. This makes it robust for changing climate regimes.
- **Operational Viability:** The results demonstrate that a drone swarm could identify ignition zones in real time by sensing local conditions (slope/fuel) rather than downloading massive dataset



- **Fig 5: Physics only Search.** Even without historical data, the GA (blue path) correctly identifies the risk zone based purely on vegetation and topography, validating the model for unmapped regions.

Conclusions

- **RQ1 Answered:** The Genetic Algorithm successfully approximated the Random Forest's prediction, consistently converging on high-risk clusters.
- **RQ2 Answered:** The evolutionary approach achieved this accuracy with <1% of the computational overhead required by the baseline model.
- **Impact:** This establishes a viable framework for autonomous wildfire scouting in disconnected, bandwidth constrained environments.

References:

•**[1]** Hong, H., et al. (2018). Applying genetic algorithms to set the optimal combination of forest-fire-related variables. *Science of The Total Environment*, 630, 1044-1056.
•**[2]** Mabdeh, N., et al. (2022). Forest fire susceptibility assessment and mapping using support vector regression. *Sustainability*, 14(15), 9446.
•**[3]** Razavi-Termeh, S. V., et al. (2025). Optimizing ensemble learning for satellite-based multi-hazard monitoring. *Scientific Reports*, 15, 15381.
•**[4]** FireCastRL. (2024). *US Wildfire Dataset (2014-2025)* [Data set]. Kaggle. https://www.kaggle.com/datasets/firecastrl/us-wildfire-dataset