Project Outline

Wildfire Ignition Forecasting

Anatole Lobenko[1], Ameya Dinesh Kurme[1], Hai Hoang Nguyen[1], Hayden Kulle[1], Joshua Nicolaus[1], and Yeungbin Lee[1]

[1]Team 13

November 29, 2025

Submitted to
Data and Web Science Group
Prof. Dr. Sven Hertling
University of Mannheim

1 Problem Definition and Goals

Wildfire ignition prediction aims to estimate whether a fire will ignite at a specific location and date. We treat this as a binary classification task with the target variable `wildfire_bin` $\in \{0, 1\}$.

We use 10-day aggregated meteorological and drought conditions (temperature, humidity, wind, precipitation, radiation, fuel-moisture indices), complemented by terrain (`elevation_m`, `slope_deg`) and grouped NLCD land-cover features. These form the final tabular input for all models.

Our objectives are to:

1. Train and compare Logistic Regression, Random Forest, and XGBoost.
2. Identify the most informative features for ignition risk.
3. Evaluate performance—with emphasis on high recall—and analyse remaining prediction errors.

2 Dataset

2.1 Overview

We use the US Wildfire Dataset from Kaggle (Lab 2025), which combines IRWIN ignition records with GRIDMET daily meteorological observations for the period 2014–2025. Each sample is derived from a 75-day sequence around an ignition event; from this sequence we compute the 10-day averages that serve as model inputs.

After preprocessing, the dataset contains 126,800 labelled samples, including 50,720 ignition events and 76,080 synthesised non-ignition cases, providing broad spatial and temporal coverage. Table 2.1 shows the dataset's features and their descriptions

Table 2.1: Feature structure of the raw dataset prior to preprocessing.

Category	Variables
Spatial / Temporal	`latitude`, `longitude`, `datetime`
Target	`wildfire_bin` (binary ignition label, 0/1)
Precipitation	`pr`
Humidity	`rmax`, `rmin`
Temperature	`tmmn`, `tmmx`
Wind	`vs`
Radiation	`srad`
Moisture / Atmosphere	`sph`, `vpd`
Fire danger / fuel	`bi`, `fm100`, `fm1000`, `erc`, `etr`, `pet`

2.2 Data Assessment

The dataset contains real ignition samples and around 76k synthetic non-ignition samples, which closely match real records in feature ranges. A small number of rows where all features took the sentinel value `32767`—invalid GRIDMET frames—were removed, as were a few duplicates. The data extend through 2025, supporting a chronological train–validation–test split.

2.2.1 Temporal Distribution

Clear temporal shifts occur between the training years (2014–2021) and the test period (2022–2025). The ignition rate rises from about 20% in the training period to 58% in 2022–2025, with 2024 reaching nearly 86%. Feature distributions also shift, with later ignitions occurring more often in south-western, lower-elevation, flatter, and more human-influenced areas. These changes affect calibration and contribute to the temporal error patterns discussed later.

2.2.2 Outliers

IQR-based outlier detection shows that precipitation (`pr`) has the highest outlier rate ($\sim$7%), while other meteorological variables have fewer than 4%. These values represent realistic environmental extremes and were retained. Ignition frequency is slightly lower among outlier rows (18.7%) than non-outlier rows (21.2%), indicating that isolated extreme values do not increase ignition likelihood.

2.2.3 Correlation

Strong correlation clusters ($|\rho| > 0.9$) occur among fuel-dryness indices (`erc`, `fm100`, `fm1000`), evapotranspiration variables (`etr`, `pet`), and humidity-related measures (`rmin`, sph, vpd). These relationships informed later feature-selection steps to reduce redundancy.

3 Preprocessing and Feature Engineering

3.1 Preprocessing

Rows in which all meteorological and drought variables took the sentinel value 32767 were removed, as they represent invalid GRIDMET frames rather than real observations. A small number of duplicate entries—mostly among synthetic non-ignition samples—were also dropped. Outliers and correlations were assessed during data inspection, but no features were removed at this stage; all variables were kept for later engineering and selection.

3.2 Feature Engineering

Short-term environmental conditions were summarised by averaging each meteorological and drought variable over the 10 days preceding the reference date. Additional feature groups include rain-derived indicators (accumulated rainfall and time since last rain), terrain attributes from LANDFIRE (USGS and LANDFIRE Program 2021), grouped NLCD land-cover classes (Multi-Resolution Land Characteristics Consortium 2021), temporal encodings (cyclic day-of-year and weekday), and local fire activity within 50 km (Fire50km_past_count).

3.3 Feature Selection

Feature selection focused on reducing redundancy and improving stability. Technical identifiers were removed, and highly correlated engineered variables within the same meteorological or drought-related groups were simplified by retaining only representative predictors. Low-variation or unstable features, as well as overlapping temporal encodings and unnecessarily complex interactions, were removed to preserve interpretability. The final feature groups used for training are listed in Table 3.1.

Table 3.1: Final feature set used for modelling.

Feature group	Features (examples)
Spatial & terrain	`latitude`, `longitude`, `elevation_m`, `slope_deg`
Rain-based indicators	`pr_max`, `pr_sum`, `since_prev_pr`
10-day meteorological averages	`rmax_mean_rt_50_59`, `rmin_mean_rt_50_59`, `sph_mean_rt_50_59`, ...
Drought–terrain interactions	`erc_over_fm1000`, `erc_over_fm1000_slope`
Temporal (cyclical)	`doy_sin`, `doy_cos`
Land-cover (grouped NLCD)	`nlcd_fuel_continuity`, `nlcd_canopy_height_class`, ...
Target variable	`wildfire_bin`

4 Machine Learning Approaches

4.1 Baseline Model

As a simple reference, we include a temperature-based rule: `wildfire_bin` = 1 if the 10-day maximum temperature exceeds 300 K, and 0 otherwise. This threshold was chosen based on the temperature distribution, where ignitions occur slightly more often at higher values.

4.2 Model Selection

We evaluate three standard classifiers for binary wildfire prediction:

- **Logistic Regression (non-symbolic)**: A linear baseline model that offers interpretability and establishes a reference for learned classifiers.
- **Random Forest (symbolic)**: A tree-based ensemble capturing nonlinear relationships and interactions in tabular data.
- **XGBoost (symbolic)**: A gradient-boosted decision-tree model known for strong performance and effective modelling of complex feature interactions.

5 Evaluation

5.1 Experimental Setup

Models are trained on 2014–2021 data with a 20% validation split and evaluated on 2022–2025 to approximate a real forecasting scenario. Logistic Regression uses z-score normalisation, while tree-based models operate on the original feature scales. The class distribution is imbalanced and shifts over time (training ∼4:1, test ∼1.4:1); since the split is chronological, stratification is not possible. Logistic Regression and Random Forest apply `class_weight="balanced"` to compensate, whereas XGBoost handles imbalance through its internal loss.

Performance is measured using accuracy, precision, recall, F1-score, and ROC–AUC, with ROC–AUC used for hyperparameter tuning due to its threshold-independence and robustness under imbalance. ROC and precision–recall curves further illustrate ranking behaviour across thresholds.

5.2 Hyperparameter Tuning

All models were tuned using the same validation split, selecting the configuration with the highest ROC–AUC. For each method, we explored a focused hyperparameter range and highlight the best values in **bold**.

Table 5.1: Hyperparameter search ranges and selected best values.

Model	Parameter	Range / Best Value
Logistic Regression	Regularisation strength (`logreg_C`)	$\{0.01, 0.1, 1, \mathbf{10}\}$
Random Forest	`n_estimators`	$\{100, \mathbf{200}, 300\}$
	`max_depth`	$\{\text{None}, 10, 20, \mathbf{30}\}$
	`min_samples_split`	$\{2, 5, \mathbf{10}\}$
	`min_samples_leaf`	$\{1, \mathbf{2}, 4\}$
	`max_features`	$\{"\mathbf{sqrt}", "log2"\}$
XGBoost	`max_depth`	$\{4, \mathbf{5}, 6\}$
	`min_child_weight`	$[4, \mathbf{6}, 13]$
	`subsample`	$[0.5, 1.0]$ (best: $\mathbf{0.926}$)
	`colsample_bytree`	$[0.5, 1.0]$ (best: $\mathbf{0.814}$)
	`reg_lambda`	$\{0, 0.1, 0.5, 1, 5, 10, 50, 100, \mathbf{200}\}$
	`reg_alpha`	$\{0, 0.1, 1, 5, 10, 50, \mathbf{100}\}$
	`learning_rate`	$\{0.01, 0.03, 0.05, \mathbf{0.07}, 0.1\}$

6 Model Performance Results

6.1 Results

Because missing an ignition is far more costly than raising a false alarm, models are evaluated both at their default threshold and at a validation-selected high-recall threshold (targeting recall ≥ 0.8). The results for all models on the 2022–2025 test set are shown in Table 6.1.

Table 6.1: Model performance on the 2022–2025 test set.

Model	Threshold	Accuracy	Precision	Recall	F1
Baseline (Temp > 300K)	–	0.490	0.616	0.325	0.426
Logistic Regression	Base	0.561	0.264	0.655	0.377
Logistic Regression	High recall	0.369	0.232	0.914	0.370
Random Forest	Base	0.478	0.830	0.128	0.223
Random Forest	High recall	0.630	0.630	0.880	0.734
XGBoost	Base	0.592	0.702	0.468	0.561
XGBoost	High recall	0.623	0.646	0.801	0.715

Across all models, threshold adjustment mainly affects the precision–recall trade-off, while accuracy and F1 remain relatively stable. Among the learned models, XGBoost shows the most consistent performance across thresholds.

Compared to the baseline rule (accuracy 0.49, recall 0.33), the high-recall XGBoost model substantially improves recall (0.80 vs. 0.33) while maintaining higher precision and overall F1. This makes it far more effective for avoiding missed ignitions.

For these reasons, XGBoost is selected as the final model. The high-recall threshold reduces false negatives without severely degrading precision, and regime-specific thresholding further improves operational cost. Calibration analysis shows that XGBoost tends to underestimate absolute ignition probabilities, so its outputs are best interpreted as relative risk scores rather than calibrated likelihoods.

6.2 Tree Interpretation

Figure 6.1 shows an example tree from the XGBoost model. The top split is based on a fuel-dryness indicator, meaning the model first distinguishes days with very dry fuel conditions from those with more moisture. Subsequent branches rely heavily on recent rainfall patterns—such as how much rain fell in the past days or how long it has been since the last rain—reflecting the strong influence of short-term moisture deficits on ignition likelihood.

Deeper splits incorporate factors such as burning conditions, humidity, elevation, and land-cover type, adjusting the prediction according to regional and terrain differences.

Leaf values show the final risk: positive leaves correspond to dry, warm, fire-prone conditions, while negative leaves represent wetter or less favourable settings for ignition. Overall, the tree demonstrates that XGBoost learns intuitive rules linking dryness, recent weather, and landscape characteristics to ignition risk.

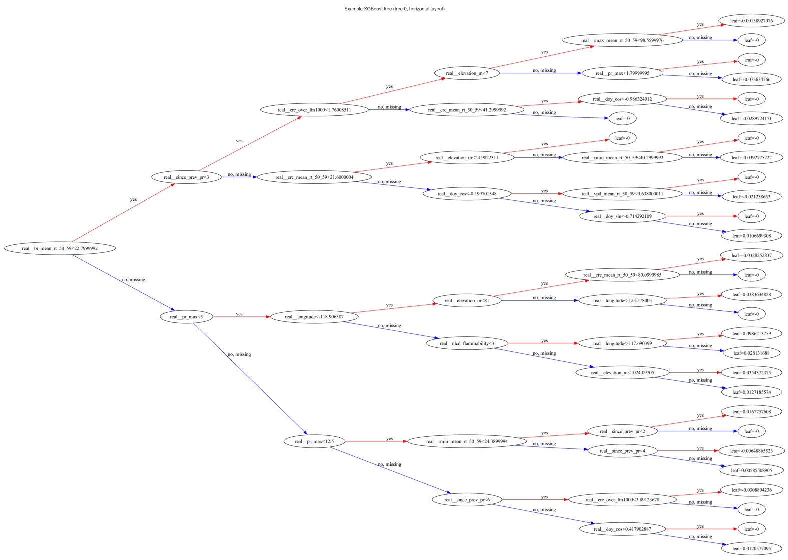
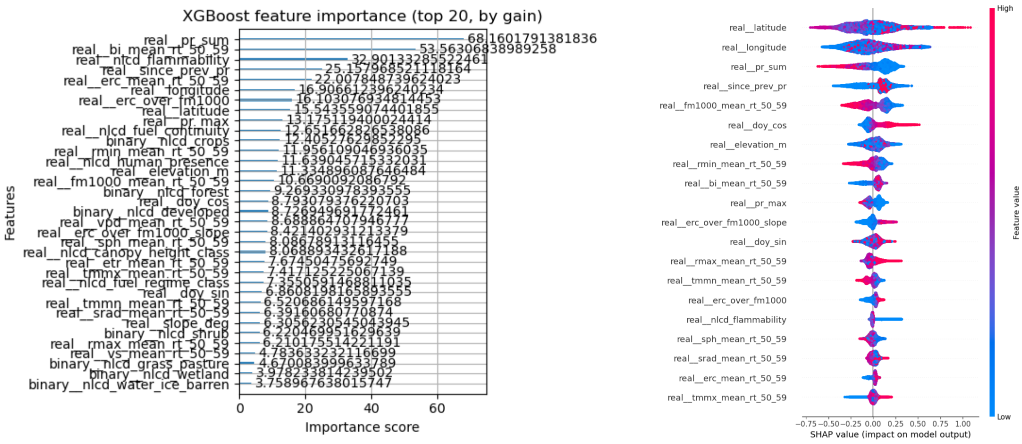


Figure 6.1: Example XGBoost tree structure (Tree 0).

6.3 Feature Importance

We analyse the XGBoost model using both gain-based feature importance and SHAP values.



(a) Gain-based feature importance. (b) SHAP summary plot.

Figure 6.2: Feature importance analysis for the high-recall XGBoost model.

Both analyses indicate that the model relies primarily on short-term moisture and fuel-dryness signals: recent precipitation, days since rain, and fuel-moisture indices are consistently the strongest predictors. Fire-weather variables and spatial coordinates also contribute, reflecting clear regional and environmental patterns. SHAP values show that dry, hot, low-humidity conditions push predictions toward ignition, while wetter conditions reduce risk. Overall, the model captures physically intuitive relationships between weather, landscape, and ignition likelihood.

7 Error Analysis

We analyse errors for the high-recall XGBoost model, which reflects our priority of minimising missed fires.

7.1 Temporal Errors

As shown in Figure 7.1, the temporal error pattern closely follows the changing fire-label distribution. In 2022–2023, when fires are relatively rare, the model produces mainly false positives due to the recall-oriented threshold. From early 2024 onward, the fire rate rises sharply, reducing both false positives and overall errors: most days are true fire days, and false negatives become the dominant error type. The improved performance in 2024–2025 is therefore driven by the unusually high fire prevalence rather than better model generalisation.

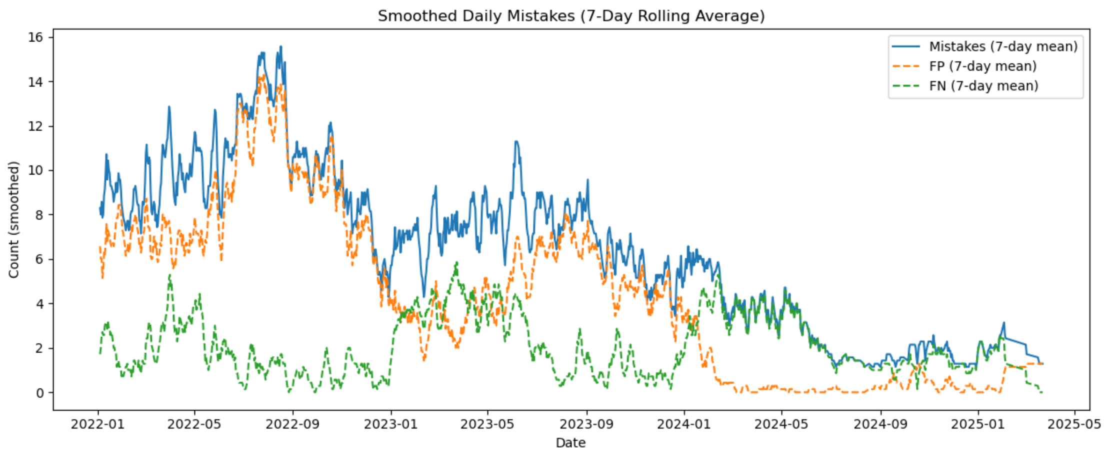


Figure 7.1: Smoothed daily mistakes (7-day rolling average) for the high-recall XGBoost model.

7.2 Spatial Errors

Figure 7.2 shows that false positives are widespread across the western and central U.S., where many non-fire days resemble typical ignition conditions. False negatives appear mainly in the Northeast and Midwest, indicating weaker generalisation to regions whose ignition patterns differ from the western-dominated training signal. This spatial contrast reflects strong performance in common fire environments and reduced accuracy in less typical ignition settings.

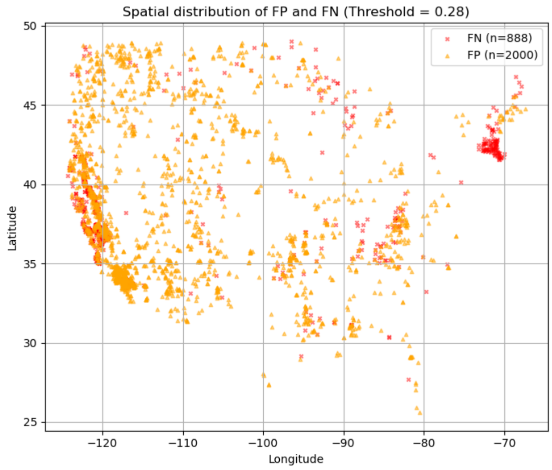


Figure 7.2: Spatial distribution of false positives and false negatives.

7.2.1 Error Regimes and Implications

Clustering the false negatives reveals two main regimes: (1) urban and peri-urban West Coast areas, where ignitions receive intermediate scores below the decision threshold, and (2) high-elevation shrub and forest regions, which show distinct humidity and precipitation patterns and produce high but sub-threshold scores. These regimes align with those targeted by the regime-specific thresholds used in Section 6.1, which reduce false negatives and expected cost without retraining the model.

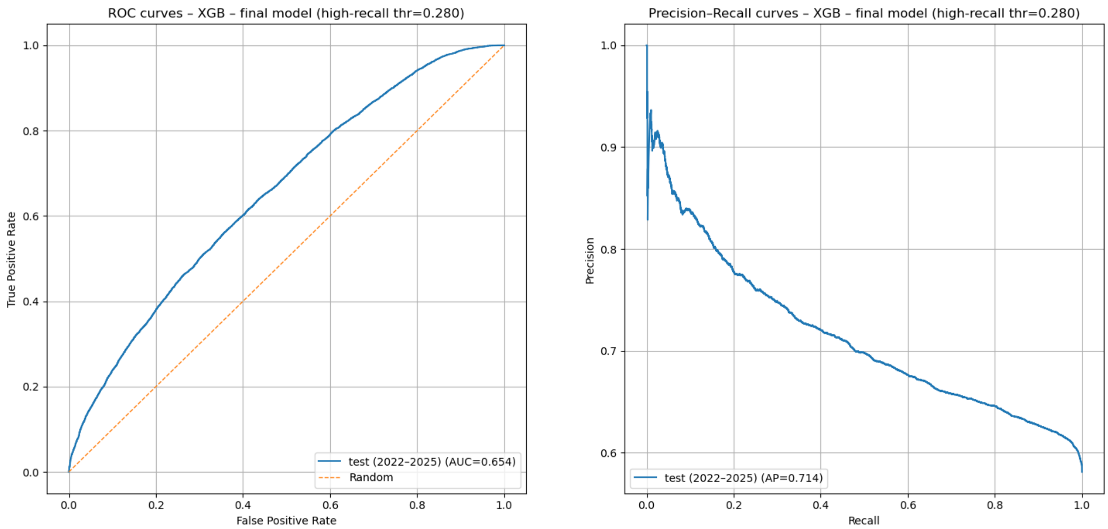
7.3 Comparison to State of the Art

State-of-the-art studies such as EWXS (Liu et al. 2025), the Montesinho model (Dong et al. 2022), and North American susceptibility mapping (Khan et al. 2025) typically focus on smaller regions or coarser temporal resolutions. In contrast, our model predicts ignition nationwide using daily aggregated meteorology, terrain, and land cover.

Across the literature, tree-based ensembles consistently deliver the strongest results. EWXS reports near-perfect accuracy (∼99%, AUC 0.983), while the Montesinho model achieves moderate performance (∼82% accuracy, AUC ∼0.80). These findings mirror our results, where XGBoost provides the most stable and competitive performance among the tested models. Although numerical comparison is not directly meaningful, our model follows the established trend that gradient-boosted trees remain a robust and practical choice for wildfire-ignition prediction.

Appendix

A.1 Additional Figures



(a) ROC curve (AUC = 0.654).

(b) Precision–Recall curve (AP = 0.714).

Figure A.1: ROC and Precision–Recall curves for the high-recall XGBoost model.

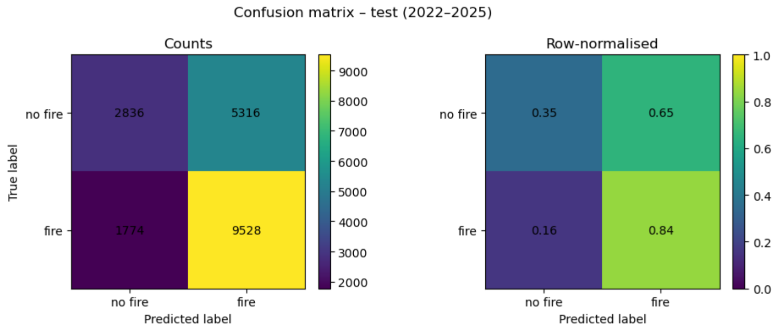


Figure A.2: Confusion matrix for the high-recall XGBoost model on the 2022–2025 test set.

Bibliography

Dong, Z., L. Ribeiro, and J. Santos (2022). Wildfire prediction model based on spatial and temporal characteristics: A case study of a wildfire in portugal's montesinho natural park. *Sustainability 14*(16), 10107.

Khan, M. A., C. Saha, and M. T. Rahman (2025). Machine learning-based wildfire susceptibility mapping: A GIS-integrated predictive framework. *Applied Sciences 15*(22), 12188.

Lab, F. R. (2025). Us wildfire dataset. `https://www.kaggle.com/datasets/firecastrl/us-wildfire-dataset`. Kaggle dataset. Accessed: 2025-11-29.

Liu, T., Y. Zhou, W. Chen, and Z. Xu (2025). Tackling the wildfire prediction challenge: An explainable ai model combining XGBoost with SHAP (ewxs). *Forests 16*(4), 689.

Multi-Resolution Land Characteristics Consortium (2021). Nlcd 2021 land cover database. `https://www.mrlc.gov/data/statistics/national-land-cover-database-2021-nlcd2021-statistics-2021`. Accessed: 2025-11-29.

USGS and LANDFIRE Program (2021). Landfire terrain data. `https://www.landfire.gov/data/FullExtentDownloads`. Accessed: 2025-11-29.

Ehrenwörtliche Erklärung

Ich versichere, dass ich die beiliegende Bachelor-, Master-, Seminar-, oder Projektarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und in der untenstehenden Tabelle angegebenen Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

I certify that the attached bachelor's, master's, seminar, or project thesis has been completed without the assistance of third parties and without the use of resources other than those specified, and that all passages taken either verbatim or in content from the sources used have been clearly identified as such. This work has not been submitted to an examination board in the same or similar form. I am aware that a false declaration may have legal consequences.

Declaration of Used AI Tools

Tool	Purpose	Where?	Useful?
ChatGPT	Rephrasing	Throughout	+
ChatGPT	LaTeX Trouble Shooting	Throughout	+
ResearchGPT	Summarization of related work	State of the Art Studies	+
ChatGPT	Pipeline Building and Understanding	Models	++
ChatGPT	Visuals Creation	Figure 7.2, 7.1	+
ChatGPT	references.bib code generation	Bibliography	++

Unterschrift
Mannheim, den 30.11.2025

Anatole Lobenko   Hayden Kulle   Joshua Nicolaus

Ameya Dinesh Kurme   Yeungbin Lee

Hai Hoang Nguyen