

From forecast skill to economic value: sub-hourly wildfire potential forecasting across Australian regions

Alberto Ardid^{A,*} , Anthony Power^B, Andres Valencia^B, H. Grant Pearce^C , Shana Gross^D, David Dempsey^A and Matthias Boer^E

For full list of author affiliations and declarations see end of paper

*Correspondence to:

Alberto Ardid
 Civil and Environmental Engineering,
 University of Canterbury, Christchurch,
 New Zealand
 Email: alberto.ardid@canterbury.ac.nz

Received: 9 September 2025

Accepted: 10 February 2026

Published: 25 March 2026

Cite this: Ardid A *et al.* (2026) From forecast skill to economic value: sub-hourly wildfire potential forecasting across Australian regions. *International Journal of Wildland Fire* **35**, WF25221. doi:[10.1071/WF25221](https://doi.org/10.1071/WF25221)

© 2026 The Author(s) (or their employer(s)). Published by CSIRO Publishing on behalf of IAWF.

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License ([CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/))

OPEN ACCESS

ABSTRACT

Background. Wildfire risk is rising under climate change, yet most operational forecasts rely on daily indices that miss rapid weather shifts preceding ignition. Accurate, timely and economically justified forecasts are critical for early warning and resource allocation. **Aims.** This study evaluates a sub-hourly machine-learning (ML) forecasting system for fire potential using weather-station data from three Australian regions (Sunshine Coast, Brisbane, Hobart), and quantifies its economic value using a cost–loss framework. **Methods.** ML classifiers were trained on sub-hourly Automatic Weather Station (AWS) data and benchmarked against the Fire Behaviour Index (FBI). Forecast discrimination was assessed using True Positive Rates (TPRs) and False Positive Rates (FPRs), as well as a Potential Economic Value (PEV) analysis, complemented by learning-curve tests. **Key results.** The ML model improved forecast skill over the FBI by 10–30%, with the ML system doubling potential savings relative to the FBI. Learning-curve diagnostics indicated that stable performance is achieved after ~30 fire events. **Conclusions.** The ML forecasting framework demonstrates predictive skill and measurable economic value, highlighting its potential for scalable, cost-effective early-warning systems. **Implications.** By linking forecast skill directly to financial and operational outcomes, this approach provides a quantitative basis for prioritising investment in timely, data-driven fire-warning tools across regions with limited data and constrained budgets.

Keywords: Australia, Automatic Weather Station (AWS) data, cost–loss framework, early-warning systems, Fire Behaviour Index (FBI), machine learning (ML), Potential Economic Value (PEV), transfer machine learning, wildfire potential forecasting, sub-hourly forecasting.

Introduction

Increasing wildfire activity raises a practical question for land managers and forecasters worldwide: can warning systems keep pace with the sub-hourly weather changes that govern ignition and initial fire spread, and deliver forecasts accurate enough to inform real-time decisions? Satellite-derived records show more frequent burn days and higher-intensity fires since 2000 across south-eastern Australia, western North America, southern Europe and the boreal zone (Jolly *et al.* 2015; Boer *et al.* 2020; Bowman *et al.* 2020). However, most operational fire-danger systems still rely on once-daily updates that fail to resolve the rapid meteorological fluctuations critical to ignition, underlining a growing mismatch between fire behaviour and forecast accuracy (Harris and Lucas 2019). This discrepancy is expected to intensify as climate change, land clearing and continued housing growth in the wildland–urban interface (WUI) amplifies exposure and losses (Westerling *et al.* 2006; Abatzoglou and Williams 2016; Radeloff *et al.* 2018; Abram *et al.* 2021).

Beyond predictive accuracy, this trend raises a second challenge: can we measure the economic value of improved forecasts and their capacity to reduce losses? Forecasting systems are not judged by accuracy alone – emergency agencies must decide when the cost of mobilising crews, aircraft and public warnings is justified by the expected reduction in damages. Without a clear measure of economic benefit, even technically superior models can struggle to gain operational traction. Establishing a framework that links forecast performance to real-world

savings is therefore essential for integrating machine-learning (ML) forecasts into policy and preparedness planning.

Background

Traditional daily empirical indices remain the foundation of operational fire-danger warnings. In Australia, the McArthur Forest Fire Danger Index (FFDI) combines temperature, humidity, wind speed and a drought factor into a single value (McArthur 1967; Noble et al. 1980), with analogous systems used elsewhere such as the Fire Weather Index (FWI) and National Fire Danger Rating System (NFDRS) (Cohen and Deeming 1985; Van Wagner 1987). Although these indices can be computed hourly, operational decisions are still tied to daily aggregates, which miss rapid sub-daily dynamics – humidity crashes, synoptic wind shifts, and delayed heatwave effects – that often precede ignition (Harris and Lucas 2019).

The 2022 Australian Fire Danger Rating System (AFDRS) replaced the FFDI with the fuel-specific Fire Behaviour Index (FBI) and a simplified four-level public scale (Hollis et al. 2024). While the Bureau of Meteorology (BoM) now computes the FBI hourly on a ~ 1.5 km grid (Matthews et al. 2019; Sauvage et al. 2024), only a single daily rating is released publicly, likely to maintain message consistency and avoid overwhelming non-specialist audiences. However, this creates a gap between the rapidly evolving fire weather conditions captured in agency-level data and the information available for operational decision-making. Sub-hourly ML forecasts, such as those developed here, are intended to bridge this gap by providing fire management agencies and emergency responders with higher-frequency risk assessments that can inform pre-positioning of resources, crew readiness, and tactical suppression decisions, complementing rather than replacing the simplified public-facing daily ratings.

Meanwhile, several decades of dense weather-station records and Earth-observation data have created the ‘big-data’ environment in which ML models thrive (Jain et al. 2020; Di Giuseppe et al. 2025). Modern artificial intelligence (AI) systems already rival conventional numerical weather prediction in both speed and skill (Allen et al. 2025; Price et al. 2025). Applied to fire danger, ML methods can ingest multi-scale meteorological inputs and generate probabilistic forecasts every few minutes at negligible cost, offering a pathway to near real-time warnings where sufficient data coverage exists. However, transferability between regions has been a frequently noted weakness of fire-danger ML models (Shadrin et al. 2024).

Recent Australian and international studies confirm this shift, from daily or spatial fire-risk mapping toward high-frequency probabilistic forecasting, using a range of ML and deep-learning approaches (e.g. Gibson et al. 2020; Nur et al. 2023; Abohaia et al. 2025a; Henderson and Chakraborty 2025; Palk et al. 2025; Partheepan et al. 2025; Xu et al. 2025). These advances collectively underscore the transition toward data-driven, dynamically updated fire-potential models such as developed here.

Despite advances in predictive skill using data-driven models, the operational adoption of ML fire forecasts ultimately depends on demonstrating economic value. Fire-management agencies operate under constrained budgets, and any new forecasting tool must justify its cost by quantifying tangible benefits. The Potential Economic Value (PEV) framework (Murphy 1985; Busker et al. 2023; A. Ardid, D. Dempsey, G. McDonald, S. Cronin, unpubl.) provides a rigorous basis for this assessment by comparing the realised benefits of a forecast with those of a perfect forecast and a no-forecast baseline. Expressed in relative monetary terms, PEV links technical skill directly to decision-making efficiency and resource allocation. This study therefore places equal emphasis on forecast performance and its economic utility, using PEV as an objective metric to evaluate the operational relevance of sub-hourly fire-potential forecasts across contrasting Australian climates. Cost assumptions are informed by recent post-fire accounting in Australia ($\sim$ US\$500 ha $^{-1}$; Dexter and Macleod 2017) and average US federal wildfire expenditures ($\sim$ US\$2.9 billion year $^{-1}$; McWethy et al. 2019).

Our study

This study addresses two main aspects of sub-hourly wildfire potential forecasting: (i) evaluating predictive performance across distinct Australian fire-climate regimes, and (ii) quantifying the economic value of those forecasts using a cost-loss framework. We build on the ML forecasting system of Ardid et al. (2025a), extending it from a single Sunshine Coast site to nine Automatic Weather Stations (AWSs) spanning three contrasting environments – subtropical coastal Sunshine Coast, peri-urban Brisbane and cool-temperate Hobart (Fig. 1a–c). Together, these regimes capture similar gradients in temperature climate, fuel structure and synoptic drivers that characterise other fire-prone regions, such as Mediterranean southern Europe and the chaparral-to-conifer belts of the western US (e.g. Moritz et al. 2014; Jones et al. 2024; Di Giuseppe et al. 2025). Using almost 50 years of sub-hourly weather observations and 126 documented fires across the three regions, we train Random-Forest classifiers that issue ignition-probability forecasts every 30 min, benchmarking skill against the FBI and testing cross-region transferability. In contrast to prior Australian ML studies focused on spatial fire-susceptibility or daily severity mapping (e.g. Gibson et al. 2020; Nur et al. 2023; Partheepan et al. 2025), our framework targets sub-hourly wildfire potential derived directly from AWS data, enabling faster operational updates and preliminary tests of regional transferability.

The model workflow/pipeline is illustrated in Fig. 1d. Each forecast cycle ingests 30-min AWS observations – temperature, relative humidity, wind speed and other variables – transformed into 10-day rolling windows from which more than 700 statistical and spectral features are extracted. A Random-Forest classifier with 300 trees is trained on these features to distinguish ‘fire’ from ‘non-fire’ windows, and the

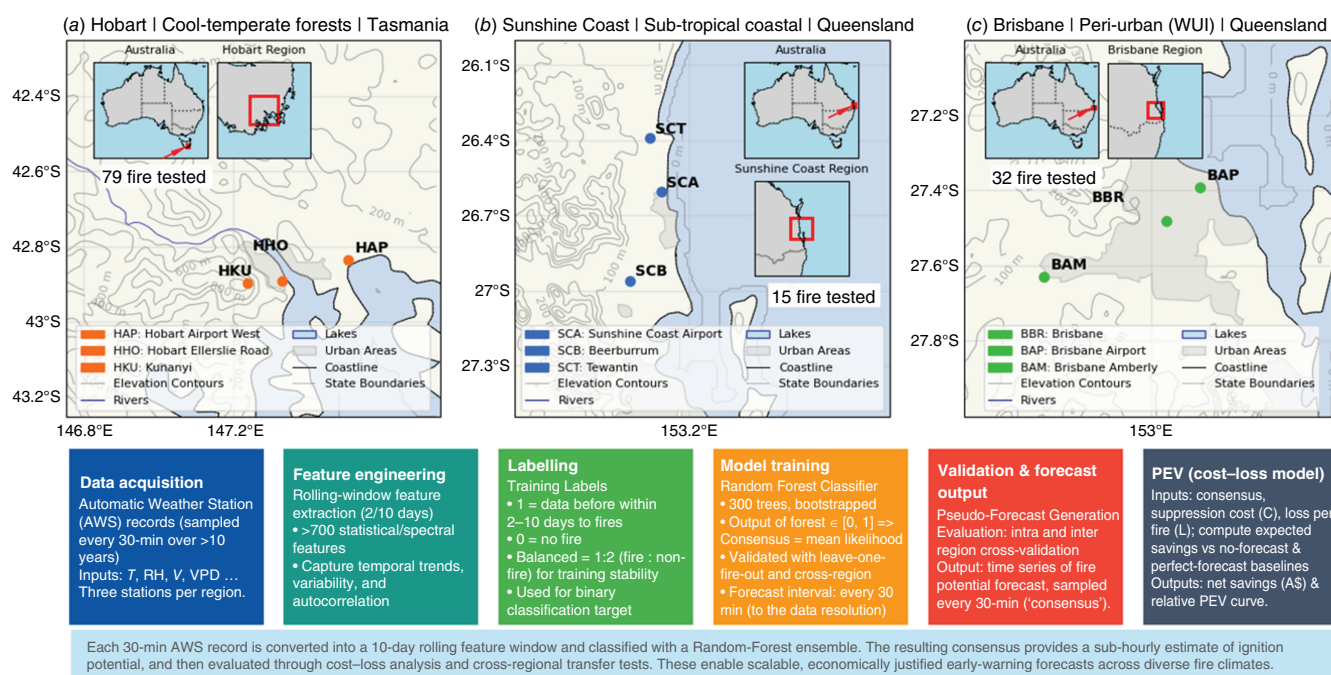


Fig. 1. Spatial context and end-to-end workflow for the sub-hourly fire-potential forecasting framework. (a–c) Automatic Weather Stations (AWSs) and regional extents for the three study areas: (a) Hobart (Tasmania), (b) Sunshine Coast (Queensland), and (c) Brisbane (Queensland). Filled circles mark the AWS used for model development; inset map locates the regions within Australia. Basemap layers show elevation contours, rivers, coastline and urban areas for geographic reference. Fires tested per region are indicated. Methodological pipeline is shown in colored boxes below the maps. Each 30-min AWS record is transformed into a 10-day rolling window from which over 700 statistical and spectral features are extracted. Windows are labelled as ‘fire’ (if an ignition occurs within the next 2–10 days) or ‘non-fire’ and used to train a 300-tree Random-Forest classifier that produces a likelihood forecast every 30 min. Validation combines leave-one-fire-out and cross-region tests to assess transferability. Forecast outputs feed an economic cost-loss module designed to estimate the Potential Economic Value (PEV) and net savings of the forecasting system over extended periods. The PEV analysis is not an operational decision model, but rather an evaluation framework used to quantify the long-term economic benefits and trade-offs of different forecasting approaches under varying uncertainty conditions.

ensemble ‘consensus’ (the mean across all trees) represents the sub-hourly forecast output (fire potential).

To evaluate operational and economic value, forecast outputs are passed to a cost-loss framework that quantifies the potential benefits of early warnings relative to suppression and damage costs. This analysis estimates the PEV and net savings of the forecasting system over extended periods, comparing both the ML model and the FBI against a baseline scenario with no forecast guidance. Importantly, the PEV framework is not an operational decision model but an evaluation tool that expresses the long-term economic benefits and trade-offs of different forecasting approaches under varying uncertainty conditions.

Methods

Study sites and data streams

We analysed nine AWSs distributed across three fire-climate regimes: subtropical Sunshine Coast, peri-urban Brisbane and temperate Hobart. These stations span gradients in

elevation, distance to coast, fuel type and exposure risk, including high-risk WUI zones. Each AWS provides 30-min meteorological records including temperature (T), relative humidity (RH), wind speed (V), dew point, pressure and rainfall. From these time series, we derived vapour-pressure deficit (VPD) and the Fine Fuel Moisture Code (FFMC). Fire ignitions were identified from official incident reports, yielding over 125 documented fire events. Time span and event counts per region are as follows: Sunshine Coast (2002–2017, $n = 15$ fires), Brisbane (2000–2018, $n = 32$ fires) and Hobart (2005–2022, $n = 79$ fires).

For consistency, both the ML forecasts and the FBI were computed using the same hourly AWS observations (temperature, humidity, wind speed and drought factor), ensuring that comparisons reflect model performance differences rather than input discrepancies. We did not use BoM forecasted weather or public deterministic ratings (BoM 2024); all FBI values in this study were recalculated directly from observed AWS data.

To ensure a fair comparison between models, we computed the FBI using the fuel type most consistent with the dominant vegetation and documented fire environments in

each region. Inspection of the event catalogue showed that ignitions on the Sunshine Coast and in Hobart occurred largely within heathland, shrubland and mixed dry-eucalypt vegetation. We therefore used the Heathland FBI for both regions, as it best reflected the fuels involved in recorded ignitions. In Brisbane, where fires were predominantly associated with dry sclerophyll forests and eucalypt woodlands, we used the Forest FBI. Although the ML models rely solely on observed weather variables and are not linked to any prescribed fuel model, we benchmarked them against the FBI variant most representative of fire-prone fuels at each site to provide an appropriate, fuel-specific operational comparison.

The three regions represent distinct fire–climate regimes that differ in their temperature, humidity and moisture-dynamics. The Sunshine Coast is a subtropical coastal environment characterised by rapid sub-daily humidity changes, frequent sea-breeze wind shifts and diurnal moisture variability during the fire season (Ardid *et al.* 2025a). Brisbane's peri-urban fringe experiences a warmer, drier subtropical climate with moderate diurnal humidity swings and a high proportion of anthropogenic ignitions (AIC 2020). Hobart, by contrast, has a cool-temperate climate where rapid fuel drying is driven by episodic foehn-like wind events and heatwave plateaus (Sharples *et al.* 2010). These different regimes influence how quickly fine fuels dry and re-moisten, and therefore the degree to which fast-changing meteorological conditions contribute to ignition likelihood.

Feature engineering and forecast target

To characterise pre-ignition signals at multiple timescales, each meteorological stream was segmented into rolling windows of 10 days (see Ardid *et al.* 2025a for sensitivity analysis over the window length), producing backward-looking feature vectors. We extracted 700+ time-series features per window using 'tsfresh' (Christ *et al.* 2018), including statistical moments, spectral coefficients, autocorrelations and change-based metrics. A binary classification target was assigned to each 30-min timestamp, where label = 1 indicated that a fire occurred within a future look-ahead window (2, 5 or 10 days), and label = 0 otherwise. Due to strong class imbalance (Tang *et al.* 2020), fire-positive windows are outnumbered by non-fire windows by ~1:50 in the raw record. Later, label-0 windows were randomly under-sampled to achieve ~1:2 balance during training.

Model architecture and training

We implemented an ensemble of Decision Tree classifiers using scikit-learn (Pedregosa *et al.* 2011). Each tree was trained on a balanced resample of the feature matrix to equalise the number of fire and non-fire windows (~1:2). A total of 300 trees were trained independently, and their probabilistic outputs were averaged to produce a consensus forecast representing the mean ignition likelihood (fire

potential). This configuration follows the ensemble architecture described in Dempsey *et al.* (2022) and Ardid *et al.* (2022, 2025b, 2025c), which demonstrated reliable classification under class imbalance and varying forecast horizons in similar forecasting contexts. Hyperparameters for the Decision Trees were fixed based on empirical performance and computational efficiency reported in Ardid *et al.* (2025a), and no further tuning was applied to avoid overfitting.

Out-of-sample evaluation was ensured using a cross-validation strategy that withholds entire fire events or stations for testing, with all remaining fires contributing to training, and class imbalance was mitigated through random under-sampling applied independently for each Decision Tree (see the following subsections). This design maintains independence between training and testing windows while capturing local meteorological variability; and maintains sensitivity to rare ignition events while limiting overfitting to background conditions.

Cross-validation and transfer learning

Two cross-validation strategies were employed. (i) Tailored models were evaluated using a leave-one-fire-out scheme at each station, where an 8-week window centred on each fire (15 days before and after the event) was excluded from both feature selection and model training, mimicking pseudo-operational forecasting. (ii) Generalised models followed a leave-region-out approach, in which models were trained on all stations from two regions and tested on the third. In both cases, feature ranking and model fitting were restricted to the training data to avoid information leakage. Within each fold, under sampling was applied independently for every Decision Tree, and repeated across 300 trees so that nearly all non-fire windows contributed to the ensemble, preserving the background variability of weather conditions. Model evaluation was always conducted on the full, naturally imbalanced held-out records (see codes in Supplementary Material).

Evaluation metrics

As explained, model evaluation is conducted through a time-blocked cross-validation on the complete, naturally imbalanced test record. For each fire, training and testing produce an out-of-sample consensus record spanning approximately 1 month around the fire date. To evaluate model performance during non-fire periods, a pseudo out-of-sample record is generated by training a single Random Forest model using all fires while undersampling the training set to match the duration of the fire periods. In practice, this setup remains effectively out-of-sample, as the model has been exposed to less than 5% of the data later used for testing. The resulting non-fire consensus record is concatenated with the fire-specific out-of-sample records to produce a continuous

model output over the entire record to be tested. Feature scaling and model fitting remain confined to the training years to prevent information leakage. This approach yields an unbiased estimate of model reliability under realistic event frequencies.

Model discrimination is summarised using threshold-free discrimination metrics (e.g. Receiver Operating Characteristic (ROC) curves, alert-rate versus detection trade-offs and Precision-Recall (PR) curves; see [Ardid *et al.* 2025a](#)). Here we present model discrimination quantified using the True Positive Rate (TPR) and False Positive Rate (FPR), defined as: $TPR = \text{True Positives (TP)} / \text{True Positives} + \text{False Negatives (FN)}$, $FPR = \text{False Positives (FP)} / \text{False Positives} + \text{True Negatives (TN)}$. In this context, positive windows correspond to pre-fire periods (label = 1) preceding documented ignition events, and negative windows correspond to all other non-fire periods (label = 0) within the held-out test years. Thus, the TPR denominator equals the total number of fire windows, and the FPR denominator equals the total number of non-fire windows. [Fig. 2](#) provides a schematic example of how TP, FP, FN and TN are identified from the sub-hourly consensus time series relative to each fire event and threshold.

In this context, these metrics (TPR, FPR) were selected over accuracy or precision because they are threshold-independent and directly define the ROC curves used throughout this study. Moreover, they align with the cost–loss framework employed for the PEV analysis, in which TPR reflects avoided fire losses (successful detections) and FPR represents unnecessary response costs (false alarms). Accuracy and precision are strongly influenced by the event imbalance and are therefore less informative for assessing operational value.

Cost–loss for forecast utility

We estimated PEV using the binary cost–loss formulation originally proposed by [Murphy \(1985\)](#) and widely applied in ensemble-weather forecasting ([Richardson 2000](#); [Mylne 2002](#)) and wildfire-risk studies ([Gunasekera *et al.* 2006](#); [Abt](#)

[et al. 2008](#); [Dale 2010](#); [Lüthi *et al.* 2021](#); [Sadowska *et al.* 2021](#); [Crowley *et al.* 2023](#)). Let $C = \text{A\$50,000}$ (cost of a precautionary deployment) and $L = \text{A\$10 million}$ (loss from an uncontrolled fire). These values represent the trade-off between taking preventive action (C) and suffering major losses from a missed fire (L), allowing forecast performance to be evaluated in economic terms relevant to operational decision-making.

This formulation assumes that successful intervention completely prevents the loss. In practice, suppression only reduces losses with a certain success probability, so the framework can be adjusted for partial-mitigation scenarios: a correctly forecast fire still incurs a smaller loss (e.g. A\$5 million) or, more generally, an expected loss of $(1 - P)L$, where P is the probability that the warning avoids or limits damage. Either adjustment lowers absolute PEV values but does not affect the relative ranking of models, since the same scaling applies to all systems.

The TPR and FPR quantify the key trade-offs that drive forecast utility in this framework. In simple terms, a higher TPR means that more real fires are detected in advance, thereby reducing potential losses, whereas a lower FPR means fewer unnecessary responses and lower operational costs. Together, these two rates determine the POV, describing how often an action correctly prevents a loss versus how often it incurs an unnecessary cost.

Representative cost and loss parameters were selected to reflect realistic suppression and damage estimates for Australasian contexts – approximately US\$33,500 (A\$50,000) per precautionary deployment and US\$6.7 million (A\$10 million) per uncontrolled fire – consistent with operational values reported by fire management agencies ([AFAC 2021](#); [FENZ 2021](#); [NZFOA 2022](#); [Marshall *et al.* 2023](#)). Sensitivity tests (Supplementary Text S2) show that although absolute PEV values depend on these assumptions, the relative performance ranking of models remains stable.

We emphasise TPR and FPR rather than accuracy or precision because these rates are threshold-independent

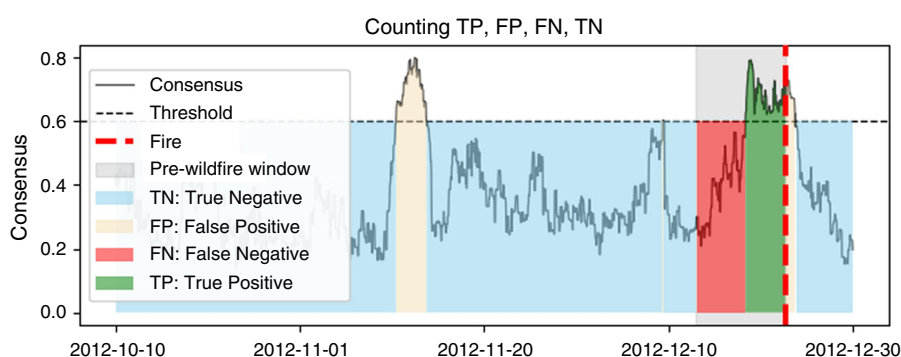


Fig. 2. Illustration of how True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN) are counted in the time series of model consensus probabilities. The dashed horizontal line marks the probability threshold used to issue a ‘fire’ forecast. The shaded red dashed line marks the observed fire time. Pre-fire windows (grey) are used for evaluation, with green showing correctly forecast fires (TP), red showing missed fires (FN) and blue showing correct non-fire predictions (TN). Yellow shading marks false alarms (FP).

and align directly with the cost–loss structure: TPR represents avoided losses and FPR represents unnecessary expenditures. Accuracy-based metrics are less informative under the strong class imbalance ($\approx 1:50$) present in sub-hourly ignition data.

With absolute event counts, the expected forecast cost E_p is defined as:

$$E_p = C \times (\text{FPR} \times D) + L \times ((1 - \text{TPR}) \times T) \quad (1)$$

where,

C = cost of a false alarm (precautionary action when no fire occurs; \$ per event)

L = loss from a missed fire (fire occurs without warning; \$ per event)

T = total number of documented fires in the record (count)

D = total number of days in that record (days)

TPR = true positive rate (fraction of fires correctly detected; dimensionless, 0–1)

FPR = false positive rate (fraction of no-fire days with false alarms; dimensionless, 0–1).

The reference cases are:

$$E_{\text{no-f}} = L \times T \quad (2)$$

$$E_{\text{pe-f}} = 0 \quad (3)$$

where, Eqn 2 represents the cost of never taking precautionary action (all fires are missed) and Eqn 3 represents perfect forecasting (no false alarms, no missed fires). PEV is then:

$$\text{PEV} = \frac{E_{\text{no-f}} - E_p}{E_{\text{no-f}} - E_{\text{pe-f}}} \quad (4)$$

Uncertainty is assessed through a Monte Carlo analysis that perturbs both C and L by $\pm 20\%$, generating an ensemble of PEV curves and optimal thresholds. The full procedure is implemented in the companion script ‘pev_3regions_monte_carlo.py’, available in the project repository.

Results

In this section, we report the performance of our ML forecasts relative to the FBI across the three fire–climate regions: Sunshine Coast (subtropical), Brisbane (peri-urban) and Hobart (temperate). For each region, model outputs combine fire-specific out-of-sample consensus records with the pseudo out-of-sample non-fire record to yield a continuous forecast over the full testing period. All metrics are computed on these held-out, naturally imbalanced records without any re-balancing at evaluation. This setup ensures that the reported discrimination, reliability and economic value

metrics reflect realistic operational event frequencies rather than balanced experimental conditions.

For each region, we evaluate model skill using TPR, false FPR, PEV and projected net savings under operational cost assumptions. The optimal classification threshold for each model is selected retrospectively to maximise PEV, reflecting how an operator might calibrate decision rules based on historical performance. Results are based on tailored ML models trained with local AWS data and assessed using pseudo-operational leave-one-fire-out cross-validation. Full performance metrics for both models are summarised in Table 1.

Hobart (temperate Tasmania)

Fire seasons in temperate forests are shorter, but heatwave plateaus and foehn-like wind events can trigger sudden escalations (Dowdy and Pepler 2018; Harris and Lucas 2019). The FBI, when evaluated at its optimal threshold of 24.99, would correctly forecast/predict 43% of recorded ignitions in Hobart (FPR = 0.101), yielding a PEV of 0.239 and net savings of A\$35.9 million, assuming that correctly forecast fires are always successfully avoided. These figures thus represent upper-bound estimates contingent on idealised response effectiveness. In comparison, our sub-hourly ML tailored model, at a calibrated threshold of 0.84, correctly forecasts 75.5% of ignitions (FPR = 0.135), improves PEV to 0.502 and nearly doubles the net savings to A\$75.4 million (Table 1).

During a representative Hobart fire season, the FBI values fluctuate moderately and reach isolated peaks but do not consistently rise ahead of fire events (Fig. 3a, b; see full record in Supplementary Fig. S1). In contrast, the ML consensus score (i.e. the mean predicted fire potential across the Random-Forest ensemble) exhibits a progressive increase leading up to the ignition cluster in early 2013 and remains elevated until the last fire in late May. This sustained high forecast confidence provides an early-warning signal. Notably, the consensus score briefly dips just before the final ignition. This brief dip may reflect a genuine prediction failure. It could also arise from uncertainty in the reported ignition time, as incident and fire-database records of ignition timing are known to differ from satellite- or model-derived estimates by hours or more in some cases (Benali et al. 2016), emphasising the importance of precise fire time records. Regardless, the broader signal quality remains high. In contrast, FBI values prior to these fires tend to be lower and more variable, with fewer sustained elevations. This pattern, observed across much of the Hobart catalogue, suggests that while the FBI captures some ignition-relevant conditions, it may miss others that are more consistently flagged by the ML model (see Supplementary Fig. S1 for a complete time series of consensus probabilities and FBI outputs across the full Hobart record).

The ROC analysis (Fig. 3c) confirms these patterns: the tailored ML model achieves an area under the curve (AUC)

Table 1. Optimal operating points for fire prediction models across three Australian regions.

Region	Model	Optimal threshold	TPR	FPR	PEV	Net savings (A\$ million)
Sunshine Coast	FBI	19.53	0.38	0.08	0.23	34.5
Sunshine Coast	ML tailored	0.6	0.796	0.077	0.651	97.7
Sunshine Coast	ML generalised	0.58	0.682	0.135	0.392	39.3
Brisbane	FBI	12.97	0.611	0.109	0.407	61.1
Brisbane	ML tailored	0.57	0.718	0.158	0.421	63.2
Brisbane	ML generalised	0.52	0.568	0.137	0.379	38.1
Hobart	FBI	24.99	0.429	0.101	0.239	35.9
Hobart	ML tailored	0.84	0.755	0.135	0.502	75.4
Hobart	ML generalised	0.71	0.654	0.129	0.384	35.3

The table presents performance metrics for both the Fire Behaviour Index (FBI) and Machine Learning (ML) tailored and generalised models, evaluated at thresholds that maximise Potential Economic Value (PEV). Reported values are averages based on an assumed fire suppression cost of US\$6.7 million (A\$10 million) per fire and an alert cost of US\$33,500 per day (A\$50,000) (conversions from AUD to USD using an exchange rate of A\$1 = US\$0.67). Metrics include True Positive Rate (TPR), False Positive Rate (FPR), PEV and net savings. The optimal threshold corresponds to the decision point that best balances false alarms and missed events for each region-model pair.

of 0.88, outperforming both the FBI (AUC = 0.80) and the generalised ML model (AUC = 0.78). Consistent with this, the PEV curves (Fig. 3d) show that tailored ML provides substantially higher economic value, nearly doubling the net savings of the FBI. In contrast, the generalised model offers only a modest improvement over the FBI, and in some cases performs comparably or slightly worse in cost terms. This reflects the fact that the FBI already captures the dominant ignition drivers; therefore, only locally trained ML models yield clear operational gains. The small difference between FBI and generalised ML indicates that much of the transferable meteorological structure is captured, but fully realising the benefits of ML forecasting requires calibration to local fire records.

This consistency across regional datasets demonstrates encouraging generalisation capability, suggesting that the key meteorological precursors of ignition are broadly shared (these aspects of transferability and data sufficiency are examined further in the Discussion section).

Brisbane fringe (peri-urban Queensland)

Brisbane's WUI experiences higher baseline ignition rates driven by anthropogenic activity and afternoon convective storm outflows (Sharples *et al.* 2016). In this region, the FBI already performs moderately well: at its optimal threshold of 12.97, it achieves a TPR of 61.1% and a PEV of 0.407, resulting in projected savings of A\$61.1 million. The ML model further improves detection to 71.8% (TPR), but with a higher FPR of 0.158, about 50% greater than the FBI's FPR, lifting PEV only slightly to 0.421 and net savings to A\$63.2 million (Table 1).

While the FBI captures key ignition drivers such as sustained dryness and diurnal variability, the room for added ML benefit appears smaller than in other regions. However, Brisbane's peri-urban environment contains a heterogeneous mix of dry sclerophyll forest, grass-shrub mosaics and highly variable anthropogenic ignition sources operating at fine spatial scales. Even though we used the FBI fuel type most representative of the region (Forest FBI), a single fuel model cannot capture this spatial and temporal variability in fuel structure and ignition context. In contrast, the ML model learns directly from high-frequency meteorological patterns without requiring an explicit fuel-type assumption, allowing it to retain some predictive advantage despite the smaller absolute gain. This highlights that ML models may offer complementary value in complex, mixed-fuel landscapes where standard fuel-specific indices have limited ability to represent local variability.

Results indicate that FBI values in Brisbane fluctuate daily and often show short-lived peaks that do not consistently precede ignition events (Fig. 4, Supplementary Fig. S2). The ML consensus forecast also shows variability but more frequently exhibits a sustained ramp-up in probability prior to fires (Fig. 4). For example, ignitions in May, July and August 2006 are each preceded by rising ML probabilities over multiple time steps, while FBI values during these periods appear more erratic and show transient spikes without clear lead-in structure. While both models produce noise, the ML signal tends to show greater temporal coherence ahead of fire events, enhancing its potential value for anticipatory decision-making.

The ROC analysis confirms that both models achieve reasonable discrimination, with the ML model yielding an

AUC of 0.89 compared to 0.81 for the FBI, reflecting a stronger ability to distinguish between fire and non-fire periods. The PEV analysis indicates that this improvement translates to a modest gain in operational value, with the ML model yielding net savings of A\$63 million versus \$61 million for the FBI.

Sunshine Coast (costal sub-tropical Queensland)

Daily fire danger in the Sunshine Coast is often driven by fast midday humidity collapses and abrupt sea-breeze wind

shifts (e.g. Hanley et al. 2005). In this highly dynamic environment, the FBI underperforms: its optimal threshold (19.53) correctly predicts only 38.0% of ignitions, with a FPR of 0.080, yielding a PEV of 0.230 and estimated net savings of A\$34.5 million. The ML model substantially improves on this, reaching a TPR of 79.6% while maintaining a similar FPR (FPR = 0.077).

Results over the full record indicate that ML forecasts frequently show gradual increases in the days leading up to ignition events (Supplementary Fig. S3). This pattern is particularly visible during the 2016–2017 fire season

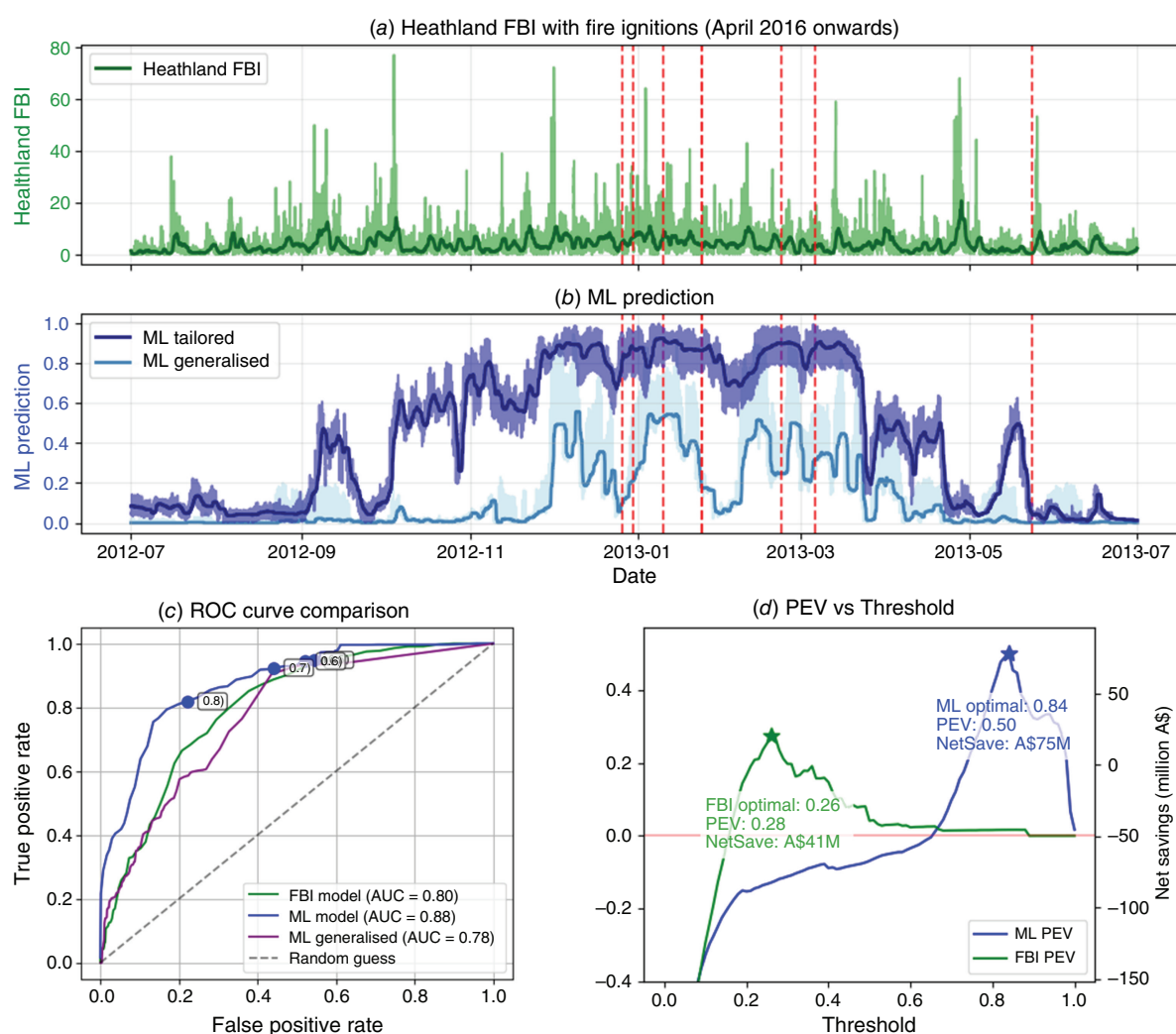


Fig. 3. Forecast skill and economic value comparison between machine-learning (ML) and Australian Fire Danger Rating System (AFDRS)-Fire Behaviour Index (FBI) models for Hobart. (a) Time series of the AFDRS-FBI for Hobart from July 2012 to July 2013, with documented fire ignitions marked by red dashed lines. (b) Sub-hourly ML forecast probability using tailored (dark blue) and generalised (light blue) models over the same period. Shaded bands show $\pm 1\sigma$ confidence interval from bootstrap ensemble. ML predictions capture rising fire potential days to weeks before ignition events. (c) Receiver Operating Characteristic (ROC) curves comparing FBI (green), ML tailored (purple) and ML generalised (blue) models. ML tailored achieves the highest area under the curve (AUC) (0.88), outperforming both FBI (0.80) and ML generalised (0.78). (d) Potential Economic Value (PEV) versus probability threshold. ML model achieves optimal PEV = 0.84 (net savings: A\$75 million) at threshold 0.60; FBI model peaks at PEV = 0.26 (net savings: A\$41 million). ML forecasts offer higher discrimination and operational value across all thresholds.

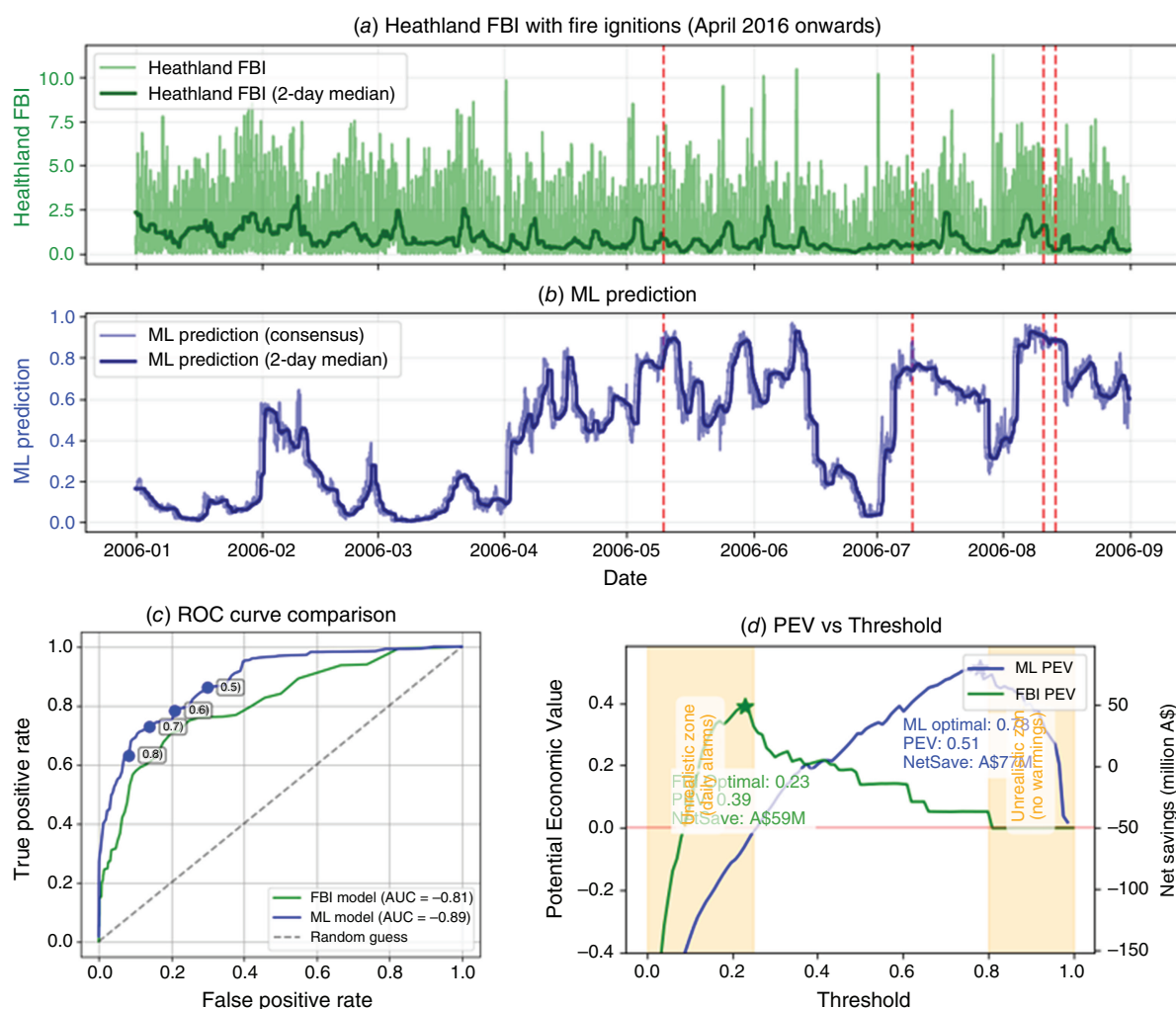


Fig. 4. Forecast skill and economic value comparison between machine-learning (ML) and Australian Fire Danger Rating System (AFDRS)-Fire Behaviour Index (FBI) models for Brisbane. (a) Time series of the AFDRS-FBI (heathland configuration) over a representative 1-year period, with documented fire ignitions marked by red dashed lines. (b) Sub-hourly ML forecast probabilities generated by tailored (dark blue) and generalised (light blue) models across the same period. Shaded bands represent $\pm 1\sigma$ confidence intervals derived from bootstrap ensembles. ML models often anticipate elevated fire potential days to weeks prior to ignition events. (c) Receiver Operating Characteristic (ROC) curves comparing FBI (green) and ML model (blue). The tailored ML model consistently achieves the highest discrimination performance, outperforming both the FBI index and the generalised ML model. (d) Potential Economic Value (PEV) as a function of decision threshold. The ML model delivers substantially higher operational value across thresholds compared to the FBI benchmark, indicating superior utility for cost-effective fire preparedness and response decisions.

(Fig. 5), where the ML consensus values rise ahead of most documented fires and remain elevated through late summer. During the same period, FBI values tend to fluctuate and show fewer extended peaks aligned with fire activity. For instance, the November and December fires coincide with sustained high ML probabilities, whereas FBI increases are less pronounced.

The ROC analysis indicates that the ML model achieves an AUC of 0.95, compared to 0.73 for the FBI, suggesting stronger ability to distinguish between fire and non-fire conditions. The corresponding PEV curves show that this translates into higher decision value, with the ML model

reaching a maximum PEV of 0.65 and projected savings of A\$98 million, at an optimal operating threshold of 0.6. For comparison, the FBI achieves a PEV of 0.20 and estimated savings of A\$30 million.

Discussion

Our results show that our sub-hourly ML forecasts can potentially improve early warning in Australia's fire prone forest regions, particularly in regions where traditional indices have difficulty capturing rapid changes in fire-weather conditions.

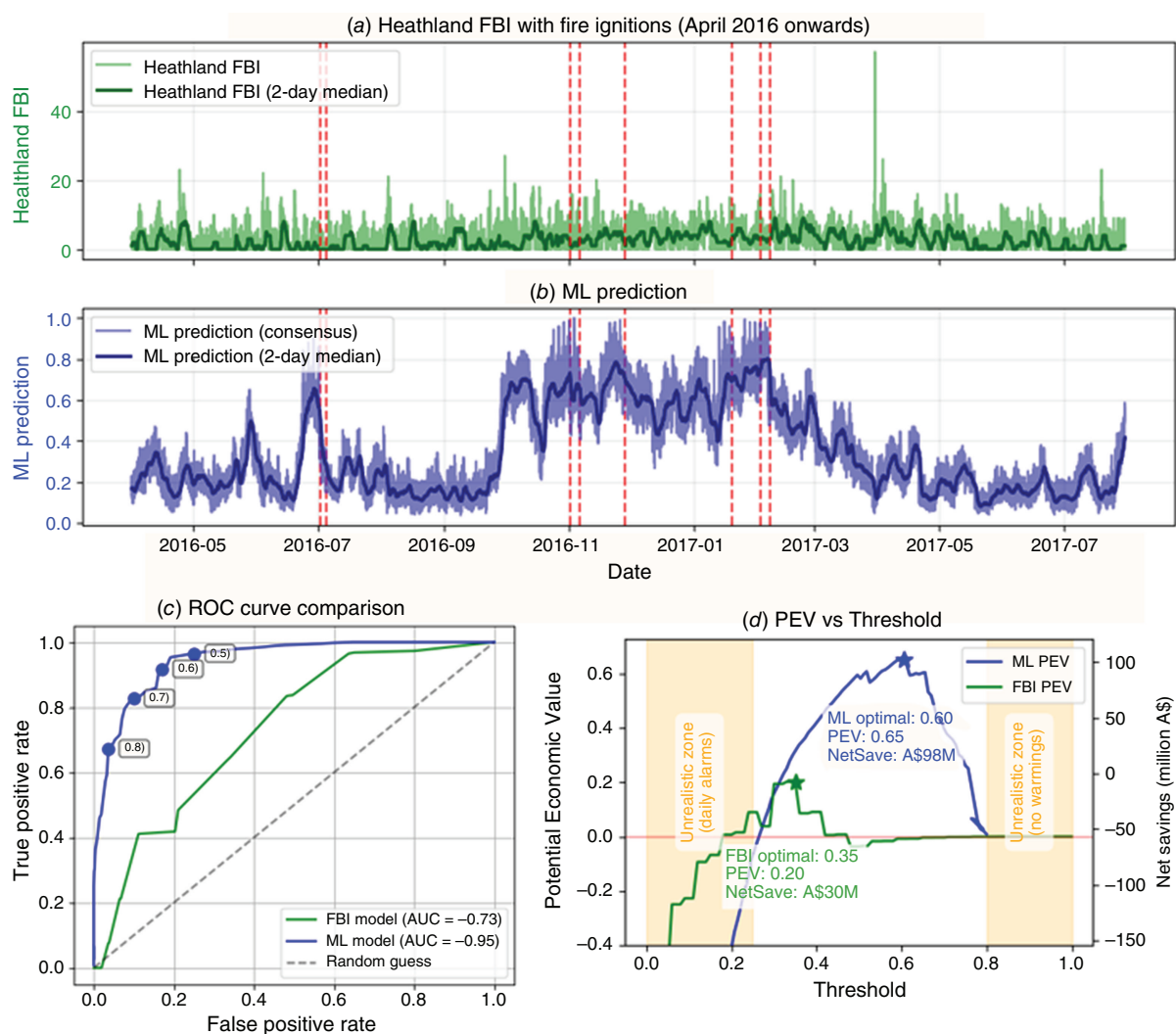


Fig. 5. Forecast skill and economic value comparison between machine-learning (ML) and Australian Fire Danger Rating System (AFDRS)-Fire Behaviour Index (FBI) models for Sunshine Coast. (a) Time series of the AFDRS-FBI (heathland configuration) over a representative 1-year period, with documented fire ignitions marked by red dashed lines. (b) Sub-hourly ML forecast probabilities generated by tailored (dark blue) and generalised (light blue) models across the same period. Shaded bands represent $\pm 1\sigma$ confidence intervals derived from bootstrap ensembles. ML models often anticipate elevated fire potential days to weeks prior to ignition events. (c) Receiver Operating Characteristic (ROC) curves comparing FBI (green) and ML model (blue). The tailored ML model consistently achieves the highest discrimination performance, outperforming both the FBI index and the generalised ML model. (d) Potential Economic Value (PEV) as a function of decision threshold. The ML model delivers substantially higher operational value across thresholds compared to the FBI benchmark, indicating superior utility for cost-effective fire preparedness and response decisions.

While the benefits of ML in handling high-frequency data and generating probabilistic outputs are acknowledged in the literature, this study supports the evidence of the transferability of ML models across contrasting climates, and their capacity to outperform existing indices under certain conditions. Nevertheless, the present framework should be viewed as a complementary forecasting layer, not a replacement for existing fire danger systems, pending validation under a wider range of fuel and climate regimes.

The degree to which our ML gains transfer across sites is uneven. Where existing indices fail to track fast, moisture-

driven swings, such as on the Sunshine Coast and to a minor degree in Hobart, the ML model delivers large improvements. By contrast, at Brisbane, where traditional metrics already capture most ignition drivers, the incremental benefit is modest. This variability argues for location-specific performance checks and reminds us that practical roll-out still hinges on data-quality issues (e.g. inaccurate fire incident data, such as imprecise ignition times) and on agency-specific workflows, risk tolerance and budget constraints, and an underlying evaluation of when the model may be relevant to new regions.

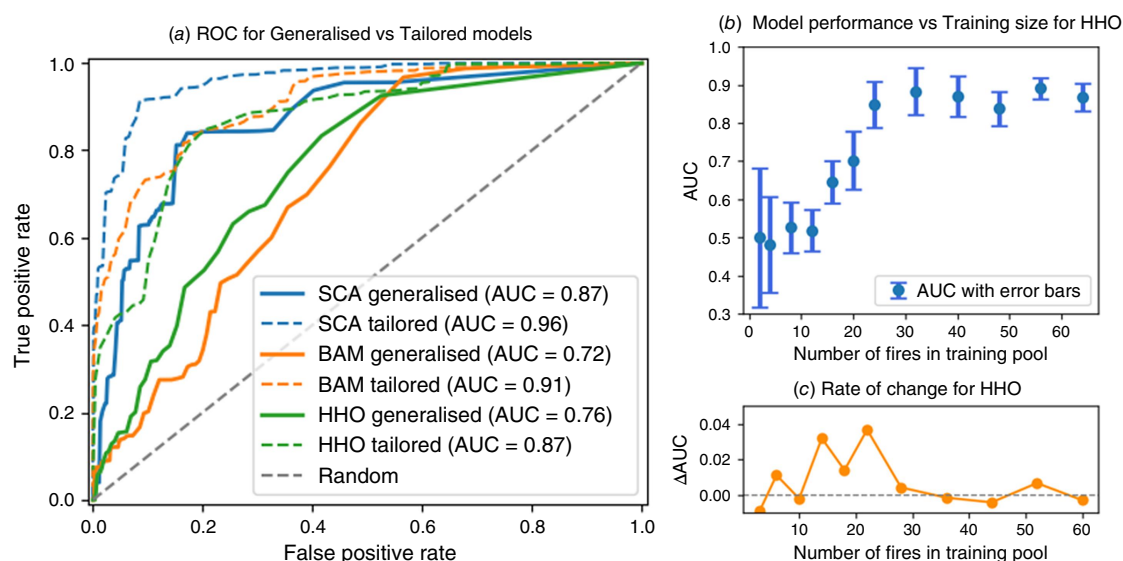


Fig. 6. Model generalisation and ergodic learning-curve behaviour. (a) Receiver Operating Characteristic (ROC) curves comparing tailored (dashed lines) and generalised (solid lines) Random-Forest models for three representative stations: Sunshine Coast Airport (SCA), Brisbane Airport (BAM) and Hobart Domain (HHO). Tailored models consistently outperform generalised counterparts, though generalised models retain strong discrimination (e.g. SCA AUC = 0.87 vs 0.96). (b) Learning curve for HHO showing AUC versus number of fire events used in training. Median AUC increases rapidly until ~30 fires, after which gains plateau. Error bars indicate interquartile range across 100 bootstrap replicates. (c) Rate of change (first derivative) of AUC from panel (b), confirming performance saturation beyond 32 events. This ergodic convergence, meaning that statistical patterns in past events are representative of future behaviour, suggests that approximately 30 well-documented fires are sufficient for model stabilisation and transferability.

While the models show consistent discrimination and positive economic value, these results should be interpreted within the limits of the available dataset and experimental design. Dataset size constrains model generality, sample counts remain modest in some regions, limiting the diversity of meteorological–fuel combinations captured during training, and the cost–loss analysis assumes fixed suppression effectiveness and decision ratios, simplifying regional differences in asset value and response capability. These simplifications do not alter the relative model comparisons but highlight the need for broader data coverage, refined parameterisation and improved uncertainty quantification. Moreover, the current framework forecasts ignition likelihood but not subsequent fire spread, meaning wind speed and other growth-related variables may be underweighted. Future extensions could integrate wind-driven and satellite-derived fuel-moisture metrics, as well as physics-based energy-balance features, and extend testing to additional regions representing distinct pyroclimates (e.g. Cunningham *et al.* 2024) to further enhance model applicability and transferability.

Model generalisation and data sufficiency

The cross-regional experiments highlight both the promise and limits of model generalisation. When trained on stations from two fire–climate regimes and tested on the third, the

sub-hourly classifier retained discrimination, in some cases only slightly below locally tailored models (e.g. Fig. 3c). As shown in Fig. 6a, tailored models improved skill by about 0.07 AUC on average, indicating that while regional calibration improves performance, much of the predictive structure transfers between sites.

Fig. 6b shows that model skill at Hobart (HHO) increases rapidly with additional training data and levels off after roughly 30 fires, with the rate-of-change analysis (Fig. 6c) confirming this plateau for the events contained in our historical record. Importantly, this behaviour reflects the limits of the dataset rather than a guarantee that future conditions will resemble the past. Fire–weather relationships are non-stationary, and novel combinations of fuel, climate, or ignition drivers may lie outside the variability captured here. The ML framework can be incrementally updated as new, potentially novel events occur, allowing the model to expand its feature space over time.

Two supplementary analyses support this interpretation. Variability-accumulation curves (Supplementary Fig. S8) show that cumulative variability in T, RH and V converges toward the full-dataset variability after ~20–30 fires. Box-plot comparisons (Supplementary Fig. S9) further indicate that fires occur across broad weather ranges – typically 40–80% of the full climatological spectrum – rather than under narrowly clustered conditions. Together, these results

suggest that a modest archive of ~30 well-documented fires captures most of the meteorological variability relevant for sub-hourly forecasting, enabling initial deployment in data-limited regions while retaining the ability to improve as new events are added.

How our findings extend fire-risk forecasting research

A growing body of work now applies ML to wildfire prediction. Global forecasts coupling satellite and meteorological data have demonstrated skill at daily-to-weekly timescales (Di Giuseppe *et al.* 2025; Illarionova *et al.* 2025), while deep-learning ensembles for fire susceptibility mapping have achieved strong classification performance, though typically focused on static risk layers rather than short-term, weather-driven ignition potential (Bjånes *et al.* 2021; Shmuel and Heifetz 2022). Our sub-hourly, station-based wildfire potential framework complements these advances by resolving rapid humidity drops and wind shifts that precede ignition, showing that finer temporal granularity yields substantial gains even over the AFDRS's new fuel-specific FBI (Hollis *et al.* 2024).

The PEV analysis showed that ML-based fire-potential forecasts consistently outperformed the FBI across all study regions. Importantly, these results hold not only under static cost-loss assumptions but also when those assumptions are varied to reflect operational uncertainty. A Monte Carlo sensitivity analysis – sampling $\pm 20\%$ variation in suppression and alert costs – confirmed that PEV estimates for the ML model remained stable and continued to exceed those of the FBI (Supplementary Fig. S7; Supplementary Text on Monte Carlo). Few previous studies have translated improved forecast skill into monetary terms (e.g. Dexter and Macleod 2017; DAE 2021); by applying the PEV framework to wildfire warnings, our analysis links technical forecast improvements directly to operational savings. This coupling of forecast accuracy, uncertainty analysis, and economic value highlights the robustness and practical significance of the results. This economic resilience, combined with better classification performance, supports the ML model's suitability for operational use in anticipatory fire management frameworks that must function under uncertain or evolving cost structures.

Conclusion

Sub-hourly ML fire potential forecasts rooted in standard weather-station data close a long-standing temporal gap in wildfire warning systems. By showing that such models can be adapted across contrasting climates and evaluated in cost-loss terms, this study advances them from proof-of-concept toward potential decision-support tools, recognising that operational deployment will still require region-specific calibration and validation (as novel events happen).

The workflow (calibrate, monetise) offers a blueprint for integrating probabilistic AI into operational risk management while maintaining transparency and economic accountability. Coupling these forecasts with dynamic fuel-moisture inputs, impact-based response models and user-behaviour testing is the logical next step. Until then, the takeaway is simple: when fire-weather shifts by the hour, our warning tools must move just as quickly; and light-weight, station-level ML makes that possible.

Supplementary material

Supplementary material can be accessed from the article page online.

References

- Abatzoglou JT, Williams AP (2016) Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences* 113(42), 11770–11775. doi:10.1073/pnas.1607171113
- Abohaia Z, Elkhoully A, Barachi ME, Al-Khatib O (2025a) Explainable AI-driven wildfire prediction in Australia: SHAP and feature importance to identify environmental drivers in the age of climate change. *Fire* 8(11), 421. doi:10.3390/fire8110421
- Abohaia Z, Elkhoully A, Barachi ME, Al-Khatib O (2025b) Regional prediction of fire characteristics using machine learning in Australia. *Fire* 8(8), 330. doi:10.3390/fire8080330
- Abram NJ, Henley BJ, Sen Gupta A, Lippmann TJ, Clarke H, Dowdy AJ, *et al.* (2021) Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Communications Earth & Environment* 2(1), 8. doi:10.1038/s43247-020-00065-8
- Abt KL, Huggett Jr. RJ, Holmes TP (2008) Designing economic impact assessments for USFS wildfire programs. In 'The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species'. (Eds TP Holmes, JP Prestemon, KL Abt) pp. 151–166. (Springer) doi:10.1007/978-1-4020-4370-3_7
- AFAC (2021) Strategic Directions for Fire and Emergency Services in Australia and New Zealand 2022–2026. Available at <https://www.afac.com.au/resources/strategic-directions-for-fire-and-emergency-services-in-australia-and-new-zealand-2022-2026>
- AIC (2020) Understanding bushfire: trends in deliberate vegetation fires in Australia. Available at https://www.aic.gov.au/sites/default/files/2020-07/tbp027_04_qld.pdf
- Allen A, Markou S, Tebbutt W, Requeima J, Bruinsma WP, Andersson TR, *et al.* (2025) End-to-end data-driven weather prediction. *Nature* 641, 1172–1179. doi:10.1038/s41586-025-08897-0
- Ardid A (2025) Datastreams and Features for Sub-Hourly Data-Driven Wildfire Potential Forecast in Australia (three regions) [Dataset]. Zenodo. doi:10.5281/zenodo.15692845
- Ardid A, Dempsey D, Caudron C, Cronin S (2022) Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. *Nature Communications* 13(1), 2002. doi:10.1038/s41467-022-29681-y
- Ardid A, Valencia A, Power A, Boer MM, Taturji M, Gross S, Dempsey D (2025a) Sub-hourly forecasting of fire potential using machine learning on time series of surface weather variables. *International Journal of Wildland Fire* 34, WF24113. doi:10.1071/WF24113
- Ardid A, Dempsey D, Caudron C, Cronin S, Kennedy B, Girona T, *et al.* (2025b) Ergodic seismic precursors and transfer learning for short term eruption forecasting at data scarce volcanoes. *Nature Communications* 16(1), 1758. doi:10.1038/s41467-025-56689-x
- Ardid A, Barth A, Dempsey D, Manga M, Cronin SJ (2025c) Forecasting eruptions at steamboat geyser: time scales, differentiability, and detectability of seismic precursors through data-driven methods. *Journal of Geophysical Research: Machine Learning and Computation* 2(2), e2024JH000569. doi:10.1029/2024JH000569

- Benali A, Russo A, Sá AC, Pinto RM, Price O, Koutsias N, Pereira JM (2016) Determining fire dates and locating ignition points with satellite data. *Remote Sensing* 8(4), 326. doi:10.3390/rs8040326
- Bjånes A, De La Fuente R, Mena P (2021) A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics* 65, 101397. doi:10.1016/j.ecoinf.2021.101397
- Boer MM, Resco de Dios V, Bradstock RA (2020) Unprecedented burn area of Australian mega forest fires. *Nature Climate Change* 10(3), 171–172. doi:10.1038/s41558-020-0716-1
- BoM (2024) Fire Weather Services Overview. (Bureau of Meteorology) Available at <http://www.bom.gov.au/weather-services/fire-weather-centre/fire-weather-services/>
- Bowman DM, Kolden CA, Abatzoglou JT, Johnston FH, van der Werf GR, Flannigan M (2020) Vegetation fires in the Anthropocene. *Nature Reviews Earth & Environment* 1(10), 500–515. doi:10.1038/s43017-020-0085-3
- Busker T, de Moel H, van den Hurk B, Aerts JC (2023) Impact-based seasonal rainfall forecasting to trigger early action for droughts. *Science of The Total Environment* 898, 165506. doi:10.1016/j.scitotenv.2023.165506
- Christ M, Braun N, Neuffer J, Kempa-Liehr AW (2018) Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package). *Neurocomputing* 307, 72–77. doi:10.1016/j.neucom.2018.03.067
- Cohen JD, Deeming JE (1985) 'The National Fire-Danger Rating System: Basic equations.' (USDA Forest Service Research Paper RM-273)
- Crowley K, Miller A, Richardson R, Malcom J (2023) Increasing damages from wildfires warrant investment in wildland fire management. Office of Policy Analysis Report R-2023-001. (U.S. Department of the Interior) doi:10.13140/RG.2.2.27598.23361
- Cunningham CX, Williamson GJ, Nolan RH, Teckentrup L, Boer MM, Bowman DM (2024) Pyrogeography in flux: reorganization of Australian fire regimes in a hotter world. *Global Change Biology* 30(1), e17130. doi:10.1111/gcb.17130
- DAE, Deloitte Access Economics (2021) 'Special report: Update to the economic costs of natural disasters in Australia.' (Australian Business Roundtable for Disaster Resilience & Safer Communities)
- Dale L (2010) The true cost of wildfire in the Western US. Available at <https://academiccommons.columbia.edu/doi/10.7916/c490-r123>
- Dempsey DE, Kempa-Liehr AW, Ardid A, Li A, Orenia S, Singh J, et al. (2022) Evaluation of short-term probabilistic eruption forecasting at Whakaari, New Zealand. *Bulletin of Volcanology* 84(10), 91. doi:10.1007/s00445-022-01600-5
- Dexter B, Macleod D (2017) What is the True Cost of Forest Fire Management on Public Land in Victoria? Available at <https://forestfirevictoria.org.au/wp-content/uploads/2020/07/Appendix-8.pdf>
- Di Giuseppe F, McNorton J, Lombardi A, Wetterhall F (2025) Global data-driven prediction of fire activity. *Nature Communications* 16(1), 2918. doi:10.1038/s41467-025-58097-7
- Dowdy AJ, Pepler A (2018) Pyroconvection risk in Australia: climatological changes in atmospheric stability and surface fire weather conditions. *Geophysical Research Letters* 45(4), 2005–2013. doi:10.1002/2017GL076654
- FENZ (2021) The Cost of Fire in New Zealand. Available at <https://www.fireandemergency.nz/assets/Documents/Files/Report-193-The-Cost-of-Fire-in-New-Zealand.pdf>
- Gibson R, Danaher T, Hehir W, Collins L (2020) A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sensing of Environment* 240, 111702. doi:10.1016/j.rse.2020.111702
- Gunasekera, D, Mills, G, & Williams, M (2006). Economic impact of fire weather forecasts. In '2006 Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere Proceedings RMRS-P-42CD, Vol. 42.' (Eds C Aguirre-Bravo, PJ Pellicane, DP Burns, D Sidney) pp. 633–637. (US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO)
- Hanley DE, Cunningham P, Goodrick SL (2005) Interaction between a wildfire and sea-breeze front. In 'Proceedings First EastFIRE Conference'. (George Mason University: Fairfax, VA)
- Harris S, Lucas C (2019) Understanding the variability of Australian fire weather between 1973 and 2017. *PLoS One* 14(9), e0222328. doi:10.1371/journal.pone.0222328
- Henderson K, Chakraborty RK (2025) A machine learning predictive model for bushfire ignition and severity: the study of Australian black summer bushfires. *Decision Analytics Journal* 14, 100529. doi:10.1016/j.dajour.2024.100529
- Hollis JJ, Matthews S, Fox-Hughes P, Grootemaat S, Heemstra S, Kenny BJ, Sauvage S (2024) Introduction to the Australian fire danger rating system. *International Journal of Wildland Fire* 33(3), WF23140. doi:10.1071/WF23140
- Illarionova S, Shadrin D, Gubanov F, Shutov M, Tasuev U, Evteeva K, et al. (2025) Exploration of geo-spatial data and machine learning algorithms for robust wildfire occurrence prediction. *Scientific Reports* 15(1), 10712. doi:10.1038/s41598-025-94002-4
- Jain P, Coogan SC, Subramanian SG, Crowley M, Taylor S, Flannigan MD (2020) A review of machine learning applications in wildfire science and management. *Environmental Reviews* 28(4), 478–505. doi:10.1139/er-2020-0019
- Jolly WM, Cochrane MA, Freeborn PH, Holden ZA, Brown TJ, Williamson GJ, Bowman DM (2015) Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications* 6(1), 7537. doi:10.1038/ncomms8537
- Jones MW, Veraverbeke S, Andela N, Doerr SH, Kolden C, Mataveli G, et al. (2024) Global rise in forest fire emissions linked to climate change in the extratropics. *Science* 386(6719), eadl5889. doi:10.1126/science.adl5889
- Lüthi S, Aznar-Siguan G, Fairless C, Bresch DN (2021) Globally consistent assessment of economic impacts of wildfires in CLIMADA v2.2. *Geoscientific Model Development* 14, 7175–7187. doi:10.5194/gmd-14-7175-2021
- Marshall E, Elliot-Kerr S, McColl-Gausden SC, Penman TD (2023) Costs of preventing and suppressing wildfires in Victoria, Australia. *Journal of Environmental Management* 344, 118606. doi:10.1016/j.jenvman.2023.118606
- Matthews S, Sauvage S, Grootemaat S, Hollis J, Kenny B, Fox-Hughes P (2019) Implementation of models and the forecast system for the Australian Fire Danger Rating System. In 'Proceedings of the 6th International Fire Behaviour and Fuels Conference. Vol. 29', Sydney, Australia.
- McArthur AG (1967) Fire behaviour in Eucalypt forests. Available at <https://www.cabidigitallibrary.org/doi/full/10.5555/19670604037>
- McWethy DB, Schoennagel T, Higuera PE, Krawchuk M, Harvey BJ, Metcalf EC, et al. (2019) Rethinking resilience to wildfire. *Nature Sustainability* 2(9), 797–804. doi:10.1038/s41893-019-0353-8
- Moritz MA, Batllori E, Bradstock RA, Gill AM, Handmer J, Hessburg PF, et al. (2014) Learning to coexist with wildfire. *Nature* 515(7525), 58–66. doi:10.1038/nature13946
- Murphy AH (1985) Decision making and the value of forecasts in a generalized model of the cost-loss ratio situation. *Monthly Weather Review* 113(3), 362–369. doi:10.1175/1520-0493(1985)113<0362:DMATVO>2.0.CO;2
- Mylne KR (2002) Decision-making from probability forecasts based on forecast value. *Meteorological Applications* 9(3), 307–315. doi:10.1017/S1350482702003043
- New Zealand Forest Owners Association (2022) 'Quantifying Forest Industry Investment in Fire Risk Management Study.' (NZFOA: Auckland, New Zealand) Available at https://www.nzfoa.org.nz/images/FOA_Fire_Investment_Report_A4_FA_Web_spreads.pdf
- Noble IR, Bary GAV, Gill AM (1980) McArthur's fire-danger meters expressed as equations. *Australian Journal of Ecology* 5, 201–203. doi:10.1111/j.1442-9993.1980.tb01243.x
- Nur AS, Kim YJ, Lee JH, Lee CW (2023) Spatial prediction of wildfire susceptibility using hybrid machine learning models based on support vector regression in Sydney, Australia. *Remote Sensing* 15(3), 760. doi:10.3390/rs15030760
- Palk M, Knappmann K, Voss S, Jovanovic R (2025) Robust prediction of wildfire spread in Australia. *IEEE Access* 13, 132703–132723. doi:10.1109/ACCESS.2025.3592124
- Partheepan S, Sanati F, Hassan J (2025) Modelling bushfire severity and predicting future trends in Australia using remote sensing and machine learning. *Environmental Modelling & Software* 188, 106377.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. (2011) Scikit-learn: machine learning in Python. *The Journal of Machine Learning Research* 12, 2825–2830.
- Price I, Sanchez-Gonzalez A, Alet F, Andersson TR, El-Kadi A, Masters D, et al. (2025) Probabilistic weather forecasting with machine learning. *Nature* 637(8044), 84–90. doi:10.1038/s41586-024-08252-9

- Radeloff VC, Helmers DP, Kramer HA, Mockrin MH, Alexandre PM, Bar-Massada A, et al. (2018) Rapid growth of the US wildland–urban interface raises wildfire risk. *Proceedings of the National Academy of Sciences* 115(13), 3314–3319. doi:10.1073/pnas.1718850115
- Richardson DS (2000) Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* 126(563), 649–667. doi:10.1002/qj.49712656313
- Sadowska B, Grzegorz Z, Stępnicka N (2021) Forest fires and losses caused by fires—an economic approach. *WSEAS Transactions on Environment and Development* 17, 181–191. doi:10.37394/232015.2021.17.18
- Sauvage S, Fox-Hughes P, Matthews S, Kenny BJ, Hollis JJ, Grootemaat S, et al. (2024) Australian Fire Danger rating system research prototype: a climatology. *International Journal of Wildland Fire* 33(3), WF23144. doi:10.1071/WF23144
- Shadrin D, Illarionova S, Gubanov F, Evteeva K, Mironenko M, Levchunets I, et al. (2024) Wildfire spreading prediction using multi-modal data and deep neural network approach. *Scientific Reports* 14(1), 2606. doi:10.1038/s41598-024-52821-x
- Sharples JJ, Mills GA, McRae RH, Weber RO (2010) Foehn-like winds and elevated fire danger conditions in southeastern Australia. *Journal of Applied Meteorology and Climatology* 49(6), 1067–1095. doi:10.1175/2010JAMC2219.1
- Sharples JJ, Cary GJ, Fox-Hughes P, Mooney S, Evans JP, Fletcher MS, et al. (2016) Natural hazards in Australia: extreme bushfire. *Climatic Change* 139, 85–99. doi:10.1007/s10584-016-1811-1
- Shmuel A, Heifetz E (2022) Global wildfire susceptibility mapping based on machine learning models. *Forests* 13(7), 1050. doi:10.3390/f13071050
- Tang X, Machimura T, Li J, Liu W, Hong H (2020) A novel optimized repeatedly random undersampling for selecting negative samples: a case study in an SVM-based forest fire susceptibility assessment. *Journal of Environmental Management* 271, 111014. doi:10.1016/j.jenvman.2020.111014
- Van Wagner CE (1987) 'Development and structure of the Canadian forest fire weather index system.' No. 35. pp. viii + –37. (Forestry Technical Report - Canadian Forestry Service)
- Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW (2006) Warming and earlier spring increase western US forest wildfire activity. *Science* 313(5789), 940–943. doi:10.1126/science.1128834
- Xu Z, Li J, Cheng S, Rui X, Zhao Y, He H, et al. (2025) Deep learning for wildfire risk prediction: integrating remote sensing and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing* 227, 632–677. doi:10.1016/j.isprsjprs.2025.06.002

Data availability. All analyses in this study were conducted using Python 3.10. The codebase and processed datasets used to generate the results are publicly available under an Massachusetts Institute of Technology (MIT) Licence at <https://doi.org/10.5281/zenodo.15692845> (Ardid 2025) and includes all relevant materials needed to reproduce the analyses presented. The shared repository provides the data and codes to train ML models used to develop and evaluate a sub-hourly wildfire potential forecasting system across the three studied Australian regions. It contains core Python scripts for model training, regional cross-validation and consensus prediction generation. It also includes raw and processed weather station records, fire ignition datasets and pre-computed feature time series. A dedicated module is included for economic evaluation using the PEV framework. This module implements all necessary procedures to compute ROC curves, threshold optimisation, net savings and summary statistics. The code is released under the MIT Licence, which permits free use, modification and distribution of the software, provided that the original licence and copyright notice are retained. However, any third-party use of the software or data for commercial purposes is prohibited without explicit permission from the corresponding author. Users are also expected to adhere to the non-commercial terms associated with the datasets and outputs provided. If you use the code or data, please cite the associated paper. For questions or collaboration inquiries, contact the corresponding author at alberto.ardid@canterbury.ac.nz.

Conflicts of interest. The authors declare that they have no conflicts of interest.

Declaration of funding. This research did not receive any specific funding.

Declaration of use of AI. Generative artificial intelligence (AI) tools were used in a limited capacity during the preparation of this manuscript. ChatGPT (OpenAI) was used to assist with language editing, writing review and the preparation of some figure layouts. The AI tool was not used for data collection, data analysis, model development, or interpretation of the scientific results. All analyses, modelling procedures and conclusions presented in this study were designed, implemented and verified by the authors. The authors take full responsibility for the content of the manuscript.

Author affiliations

^ACivil and Environmental Engineering, University of Canterbury, Christchurch, New Zealand.

^BCovey Associates, Sunshine Coast, Qld, Australia.

^CFire & Emergency New Zealand (FENZ), Christchurch, New Zealand.

^DBioeconomy Science Institute, Scion Group, Christchurch, New Zealand.

^EHawkesbury Institute for the Environment, Western Sydney University, Sydney, NSW, Australia.