

ROC-곡선 분석을 활용한 산불위험예보 등급 분류 기준 개선

안수정 · 한송희 · 김현수 · 장미나 · 이예은 · 안희영 · 허수정 

국립산림과학원 산불연구과

Improvement of Forest Fire Danger Classification Criteria Using ROC-Curve Analysis

Sujung Ahn, Songhee Han, Hyunsoo Kim, Mina Jang, Yeeun Lee,
Heeyoung Ahn and Sujung Heo 

Division of Forest Fire Research, National Institute of Forest Science, Seoul 02455, Korea

요약: 본 연구는 최근 10년 간(2015-2024) 일 별, 지역 단위 별 산불위험지수와 실제 산불 발생 데이터를 활용하여 현행 산불위험예보 등급 분류 기준의 개선 방안을 제시하였다. 현행 기준은 산불위험지수의 임계값을 넘는 면적(픽셀 수)이 70% 이상을 차지할 때 해당 지역을 특정 등급으로 분류하고 있다. 현행 기준에 따르면 “매우높음” 등급이 전체의 0.06%에 불과하고, “낮음” 등급이 70.54%로 과도하게 높은 특성을 보인다. 또한, 등급 별 산불 발생확률이 “높음”(19.9%) > “매우높음”(8.1%)으로 나타나 등급 체계의 논리적 일관성이 부족함을 확인하였다. 이에 ROC 곡선 분석 및 정확도 분석을 통해 산불위험등급 최적 분류 기준을 도출하였다. “매우 높음”은 위험지수 86 이상인 면적(D4) \geq 15%이면서 위험지수 66 이상인 면적(D3+D4) \geq 60% 일 경우, “높음”은 위험지수 66 이상인 면적(D3+D4) \geq 60%이면서 “매우높음”의 조건은 충족하지 못하는 경우, “다소높음”은 위험지수 51 이상인 면적(D2+D3+D4) \geq 40%인 경우, “낮음”은 그 외의 경우로 변경하였다. 개선된 기준은 등급이 높아질수록 산불 발생률이 증가하는 단조 증가를 보였으며(“낮음” 2.1% < “다소높음” 9.7% < “높음” 15.4% < “매우높음” 27.4%), “매우높음” 등급의 발령 횟수를 증가시켜 산불위험지수의 실효성을 높일 수 있을 것으로 기대된다.

Abstract: This study proposes improvements to the refined classification criteria for forest fire danger forecasting by using daily regional forest fire danger index forecasts and observed fire occurrence data over a 10-year period (2015-2024). Under the existing criteria, a region is assigned a grade when areas (number of pixels) exceeding the forest fire risk index threshold account for at least 70% of the total area. This approach results in an excessively high proportion of “Low” grades (70.54%) and a very small proportion of “Very High” grades (0.06%). Moreover, the fire occurrence probabilities by grade—High (19.9%) > Very High (8.1%)—reveal a lack of logical consistency in the current scheme. To address this issue, optimal classification criteria were derived using ROC-curve and accuracy analyses. The revised criteria define “Very High” as cases in which areas with a risk index \geq 86 (D4) account for at least 15% and areas with a risk index \geq 66 or above (D3+D4) account for at least 60%; “High” refers to cases in which areas with a risk index of 66 or above (D3+D4) account for at least 60% but not meeting the ‘Very High’ criterion; “Moderately High” (hereafter “Moderate”) includes cases in which areas with a risk index of 51 or above (D2+D3+D4) account for at least 40%. “Low” includes all remaining cases. The revised classification shows a monotonic increase in observed fire occurrence rates with increasing grade (Low 2.1% < Moderate 9.7% < High 15.4% < Very High: 27.4%), demonstrating improved logical coherence and predictive accuracy. By increasing the issuance of “Very High” warnings, the revised criteria are expected to enhance the operational effectiveness of forest fire danger forecasting.

Key words: forest fire danger rating system, danger classification criteria, ROC curve, recall, precision

* Corresponding author
E-mail: sjheoo@korea.kr

ORCID
Sujung Heo  <https://orcid.org/0000-0002-9155-3414>

서론

산림청 국립산림과학원에서 제공하는 국가산불위험예보시스템(<https://forestfire.nifos.go.kr>)은 2003년부터 운영되고 있으며, 기상, 임상, 지형 조건을 기반으로 산불위험지수 실황과 예보를 산출하여 제공한다. 이 시스템은 로지스틱 회귀모형을 활용한 기상요소와 산불 발생 간의 관계 분석을 기반으로 구축되었다(Lee et al., 2005; Won et al., 2016). 권역별 산불 발생 알고리즘을 거쳐(Won et al., 2011; 2016), 현재는 봄철과 가을철의 기상 변화에 따른 전국 통합 산불 발생확률 모형이 개발되어 운영되고 있다(Won et al., 2018). 국가산불위험예보시스템에서는 효과적인 산불 예방을 위해 지역별로 산불위험등급을 “매우높음”, “높음”, “다소높음”, “낮음”의 네 등급으로 분류하여 제공하고 있다.

산불위험지수는 0-100까지의 값으로 산출되며, 값의 범위에 따라 4단계(“낮음” 51 미만, “다소높음” 51~66 미만, “높음” 66~86 미만, “매우높음” 86 이상)로 나누어진다. 지역별로 각 단계에 해당하는 면적(픽셀수) 비율(각각 D1, D2, D3, D4)을 산정하고, 특정 단계 이상에 해당하는 비율의 합계가 70% 이상일 경우 해당 지역의 산불위험등급으로 산정한다.

이와 같은 현행 산불위험예보 등급 분류 기준에는 다음과 같은 구조적 문제점이 있다. 첫 번째는 특정 지역이 가장 높은 등급인 “매우높음”으로 분류되는 기준이 위험지수 86이상인 픽셀의 비율이 70%이상으로 매우 엄격하기 때문에 “매우높음” 등급이 발령되는 경우가 매우 드물고, “낮음” 등급이 상대적으로 자주 발령된다는 점이다. 이에 따라 “낮음” 등급에서 발생하는 산불이 “매우높음” 등급에서 발생하는 산불보다 많다는 문제점이 언론 등을 통해 제기된 바 있다(Newsis, 2019).

현행 산불등급 기준은 지난 2007년부터 행정적 필요에 따라 70%라는 단순한 임계값을 일괄 적용하고 있다. 하지만, 이는 지역별 산불 발생 특성 변화 등에 따른 통계적 분석이나 지속적인 검증 단계 없이 관행적으로 적용되고 있는 상태이므로, 현재 산불발생 경향을 충분히 반영하지 못하는 한계가 있다.

외국의 경우, 각국의 기후·연료·운영체계에 따라 분류 방식과 임계값 설정에 차이가 있다. 캐나다 시스템은 최종 산불기상지수(Fire Weather Index) 값을 기준으로 위험을 “낮음(Low)”에서 “극심(Extreme)”까지 구분하며, 일반적으로 산불기상지수가 약 25 이상이면 ‘높음(High)’, 50 이상이면 ‘극심(Extreme)’ 수준으로 간주된다. 다만, 각 주(州)나 지역의 기후 특성에 따라 절대값보다는 백분위를 이용한 지역 맞춤형 임계값을 적용한다. 미국은 위험도를 “낮음(Low)-보통(Moderate)-높음(High)-매우높음(Very High)-극심(Extreme)”의 다섯 단

계로 분류한다. 전국 단일 기준은 없으며, 각 지역의 연료모델과 기후조건에 따라 통계적 백분위를 기준으로 등급 경계를 설정하는 것이 일반적이다. 호주는 온도, 습도, 풍속, 건조인자 등을 이용해 0-100 이상의 값을 산출하며, 산불위험지수(Forest Fire Danger Index, FFDI) 값에 따라 “보통(Moderate)-높음(High)-매우 높음(Extreme)-재앙적(Catastrophic)”으로 구분하며, 보통 $FFDI \geq 50$ 을 ‘매우 높음’ 경계로, $FFDI \geq 100$ 을 ‘재앙적’으로 정의한다. 이처럼 외국의 위험등급기준은 과거 화재발생 통계와 지수값의 관계를 바탕으로 상대적 임계값을 설정하며, 단순 수치 구분을 넘어 정책적 대응지침(예: 대피·화재금지 조치 등)과 연계되어 운영되고 있다.

분류 체계의 최적 임계점을 찾는 방법은 다양하다. 데이터 자체의 분포에 따라 구분하는 균등 간격, 분위수 간격, Jenks Natural Breaks 등이 있고, 통계적 검증을 기반으로 하는 ROC 곡선 분석, 최소 P-value 탐색, Decision Tree 방법 등이 있다. 또한 필요한 경우에는 정책적 결정을 통해 국내/국외에서 자주 쓰이는 임계값(예: 대형산불 100 ha 이상 등)을 기반으로 구분할 수도 있다.

ROC 분석은 의료 진단에서 민감도와 특이도의 최적 균형점을 찾기 위해 개발되었으나(Swets, 1988), 현재는 자연재해 예측 분야에서 핵심적 성능 평가 도구로 자리 잡았다. Jolliffe and Stephenson(2012)은 기상예보에서 ROC 분석이 예측 성능의 객관적 평가와 최적 임계값 결정에 필수적이라고 강조하였다. 특히 AUC 0.7 이상일 때 실용적 가치를 갖는다는 Hanley and McNeil(1982)의 기준은 산불위험 분류 기준의 적절성을 판단하는 정량적 척도를 제공한다.

Hastie et al.(2009)은 분류 문제에서 조건부 확률 분석 $P(\text{Fire} / \text{Grade}_i)$ 이 각 범주의 실제 위험도를 정량화하는 핵심 도구라고 강조하였다. 특히 Flach(2012)는 위험도 등급에서 단조성 조건이 만족되지 않으면 분류 체계의 논리적 일관성이 훼손된다고 지적하였다. 단조성 조건이란 위험등급이 높아짐에 따라 산불 발생률이 높아진다는 조건이다.

$$Risk(\text{Level}_1) \leq Risk(\text{Level}_2) \leq \dots \leq Risk(\text{Level}_n) \quad (1)$$

따라서 산불위험등급 분류 체계에서 산불 발생률이 “높음” 등급에서 가장 높게 나타나도록 논리적 일관성을 갖출 필요가 있다.

본 연구는 기존 연구들이 주로 단일 통계 기법이나 전문가 경험에 의존한 것과 달리, 기술통계 → ROC 분석 → 정확도 분석 → 정확도, 논리적 일관성, 실용적 가치를 만족시키는 최적 임계값 탐색의 순차적인 체계를 통해 각 방법론의 한계를 상호 보완하고자 한다. 최신 산불 발생 데이터를 활용한 통합적 접근을 통해 현행 분류 기준의 문제점을 정량적으로

진단하고, 통계학적 방법론을 적용하여 개선된 등급 분류 기준을 제시하는 것을 목적으로 한다.

연구 방법

1. 연구데이터

본 연구는 2015-2024년(10년간) 국가산불위험예보시스템에서 산출된 일별, 지역 단위(전국 17개 광역시도)별 산불위험지수 현황값 61,408건 및 산림청 산불피해대장의 산불 발생 데이터를 활용하였다. 본 연구에서 사용한 주요 변수는 각 위험지수의 단계에 해당하는 면적(픽셀 수)의 비율, 산불 발생 건수, 피해 면적 등을 활용하였다.

2. 분석 방법

1) 기술통계 분석

현행 산불위험등급 산정 기준(Figure 1, Table 1) 적용 시

등급별 분포와 실제 산불 발생률을 산출하여 기준의 적절성을 평가하였다. 등급별 분포 분석 각 위험등급의 빈도와 상대적 비중을 계산하여 등급 간 균형을 평가하였다. 특히 극값(“매우높음”, “낮음”) 등급의 분포가 실용적 범위에 있는지 검토하였다.

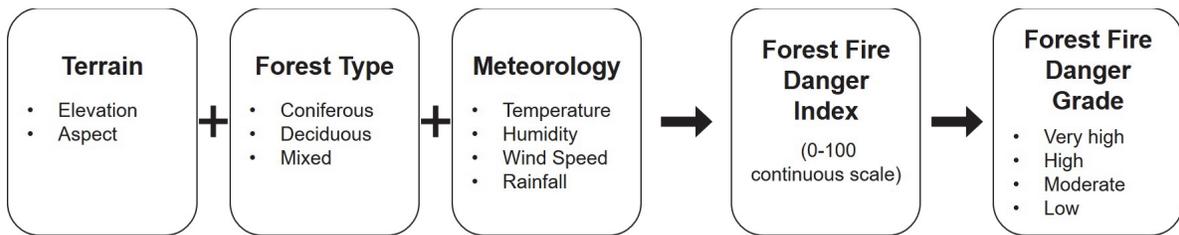
$$P(Grade_i) = \frac{n_i}{N} \tag{2}$$

여기서 $P(Grade_i)$ 는 등급 i 의 비율, n_i 는 등급 i 에 해당하는 관측치 수, N 은 전체 관측치 수이다.

각 등급 조건 하에서의 산불 발생확률을 계산하여 등급별 실제 위험도를 정량화하였다:

$$P(Fire | Grade_i) = \frac{n_{fire,i}}{n_i} \tag{3}$$

여기서 $n_{fire,i}$ 는 등급 i 에서 산불이 발생한 관측치 수이다.



Current Criteria(70% Threshold)	Improved Criteria(Optimized)
Very high: D4 ≥ 70%	Very high: D4 ≥ 15 AND D3+D4 ≥ 60%
High: D3+D4 ≥ 70%	High: D3+D4 ≥ 60%
Moderate: D2+D3+D4 ≥ 70%	Moderate: D2+D3+D4 ≥ 40%
Low: Otherwise	Low: Otherwise

Note: D1: index < 51; D2: 51 ≤ index < 66; D3: 66 ≤ index < 86; D4: index ≥ 86
 Area percentage (%) refers to the proportion of grid cells (pixels) in each category within a region.
 * Source: National Institute of Forest Science, Korea Forest Service

Figure 1. Overview of forest fire danger index calculation (Source: National Institute of Forest Science, Korea Forest Service).

Table 1. Criteria of regional forest fire danger classification.

Forest Fire Danger Class (등급명)	Forest Fire Danger Index (산불위험지수)	Area (%) (면적 비율)
Very high (매우높음)	≥ 86 (D4)	≥ 70%
High (높음)	≥ 66 (D3+D4)	≥ 70%
Moderate (다소높음)	≥ 51 (D2+D3+D4)	≥ 70%
Low (낮음)	Else (그 외)	

*D1 : area (number of pixels) where the danger index < 51
 D2 : area where 51 ≤ the danger index < 66
 D3 : area where 66 ≤ the danger index < 86
 D4 : area where 86 ≤ the danger index

산불위험등급 순서와 실제 발생률 순서의 일치성을 검증하기 위해 단조성(Monotonicity) 테스트를 수행하였다. 이상적인 분류 체계에서는 다음과 같이 산불위험도가 높아질수록 산불 발생률이 증가하는 단조성 조건이 만족되어야 한다.

$$P(\text{Fire} | \text{Low}) < P(\text{Fire} | \text{Moderate}) < P(\text{Fire} | \text{High}) < P(\text{Fire} | \text{Veryhigh}) \quad (4)$$

2) ROC 곡선 분석

각 등급별로 산불 발생 예측 성능을 평가하기 위해 ROC(Receiver Operating Characteristic) 곡선을 구성하고 AUC(Area Under Curve)를 계산하였다. 산불 발생과 비발생을 이진 분류 문제로 접근하여 각 위험 단계의 예측 성능을 평가하였다. 즉 산불의 발생과 비발생을 예측하는 모델에서 “매우높음”, “높음”, “다소높음”의 경우는 산불의 발생을 예측한 경우로, “낮음” 등급의 경우는 산불의 비발생을 예측한 경우로 구분하였다.

위험지수 단계별 면적 비율(낮음 D1, “다소높음” D2, “높음” D3, “매우높음” D4) 및 조합(D3+D4, D2+D3+D4)에 대해 다양한 임계값에서 민감도(Sensitivity 또는 Recall)와 특이도(Specificity)를 계산하였다.

* 민감도(True Positive Rate):

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

* 1-특이도(False Positive Rate):

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

여기서 TP(True Positive)는 실제 산불 발생을 올바르게 예측한 경우, FN(False Negative)은 산불 발생을 놓친 경우, FP(False Positive)는 산불이 발생하지 않았는데 발생으로 예측한 경우, TN(True Negative)은 비발생을 올바르게 예측한 경우이다.

ROC 곡선 하 면적(AUC)을 계산하여 예측 성능을 정량화하였다.

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (7)$$

Youden's J statistic을 활용하여 분류정확도가 최대가 되는 임계값을 도출하였다. 특정 임계값에서의 민감도(Sensitivity)와 특이도(Specificity)의 합을 최대화하여 계산한다. J가 1에 가까울수록 모델의 성능이 완벽에 가깝다는 것을 의미하며, J=0 일 경우 무작위 분류와 동일한 성능을 나타낸다고 볼 수 있다.

$$J = \max(TPR + TNR - 1) \quad (8)$$

$$*TNR(\text{True Negative Rate}) = 1 - FPR \quad (9)$$

3) 최적 임계값 탐색

현행 분류 체계의 임계값과 ROC-곡선법에서 도출한 통계적 최적 임계값 사이에서 임계값을 5% 단위로 조정하면서 최적 임계값을 찾는 과정을 반복수행하였다. 최적 임계값을 구하기 위해 다음 세 가지 조건을 동시에 만족하는 임계값 조합을 탐색하였다.

1. 실용성(Practicality): 산불위험등급이 높아짐에 따라 각 등급별 발령 횟수가 순차적으로 감소하도록 배분
2. 논리적 일관성(Monotonicity): 산불 발생률이 등급 순서에 따라 순차적으로 증가(단조증가)
3. 예측 성능(Statistical Accuracy): 정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity), F1-score 등 예측 성능을 종합적으로 만족시키는 최적 조합 선택

이 중, 실용성과 논리적 일관성을 만족하는 임계값을 우선 선별 후, 이 중에서 예측 성능을 최대화하는 임계값을 선택하는 방식으로 최적 임계값을 도출하였다.

4) 통계적 유의성 및 효과크기 검증

현행 기준과 개선된 기준에 대해 각각 카이제곱 검정과 Cramér's V를 계산하여 등급 체계와 실제 산불발생 간의 연관성 강도를 정량적으로 비교하였다. 이를 통해 개선된 기준이 산불위험과 실제 발생 간의 관계를 더 잘 반영하는지 평가하였다.

카이제곱 검정은 두 범주형 변수 간의 독립성을 평가하는 통계적 방법으로, 본 연구에서는 산불위험등급(명목척도)과 산불발생 여부(이진변수) 간의 연관성을 검증하는 데 활용하였다.

$$\chi^2 = \sum [(O_{ij} - E_{ij}) / E_{ij}] \quad (10)$$

여기서 O_{ij} 는 관측빈도, E_{ij} 는 기대빈도이다.

카이제곱 통계량은 표본 크기에 민감하게 반응하여, 표본이 클 경우 실질적으로 중요하지 않은 차이도 통계적으로 유의하게 나타날 수 있다(Cohen, 1988). 이러한 한계를 보완하기 위해 Cramér's V를 효과크기 측정치로 사용하였다. Cramér's V는 카이제곱 통계량을 표준화한 값으로, 0에서 1 사이의 값을 가지며 표본 크기의 영향을 받지 않는다(Cramér, 1946). 계산식은 다음과 같다.

$$V = \sqrt{\chi^2 / (n \times \min(r-1, c-1))} \quad (11)$$

여기서 n 은 전체 관측치 수, r 은 행의 수, c 는 열의 수이다.

Cramér's V 의 해석 기준은 자유도에 따라 달라지며, 본 연구의 경우 자유도가 3(4개 등급 - 1)이므로 Cohen(1988)이 제시한 다음 기준을 적용하였다.

- $0.00 \leq V < 0.10$: 매우 약한 연관성 (Negligible association)
- $0.10 \leq V < 0.20$: 약한 연관성 (Weak association)
- $0.20 \leq V < 0.30$: 중간 연관성 (Moderate association)
- $0.30 \leq V < 0.40$: 강한 연관성 (Strong association)
- $V \geq 0.40$: 매우 강한 연관성 (Very strong association)

결 과

1. 현행 기준의 문제점 진단

현행 산불위험예보 등급 분류 기준을 2015-2024년 데이터에 적용한 결과, 다소간의 구조적 불균형이 확인되었다. 첫 번째로, “매우높음” 등급은 전체 61,371건 중 단 36건(0.06%)에 불과하여 실용적 가치가 제한적임을 보였다. 반면 “낮음” 등급은 43,293건(70.54%)으로 과도하게 높은 비율을 차지하여 예보시스템의 변별력이 저하되었다. 이는 Weiss(2004)가 지적한 불균형 데이터 문제의 전형적 사례에 해당한다.

또 다른 문제는 등급별 실제 산불발생률이 논리적 순서와 일치하지 않는다는 점이다. “높음” 등급의 발생률이 19.87%로 “매우높음” 등급의 8.11%보다 2.4배 높게 나타나 분류 체계의 문제점을 드러냈다(Table 2). 통계적 유의성 검정 결과, 카이제곱 검정에서 $\chi^2 = 3,302.6$ ($p < 0.001$)으로 등급과 산불발생 간 유의한 관계가 확인되었으나, 등급 순서와 위험도 순서의 불일치로 인해 실질적 예측력은 제한적이었다. 이는 Flach(2012)가 강조한 단조성 조건에 부합되지 않아, 현행 기준의 논리적 일관성이 부족하다는 것을 알 수 있다.

각 위험 단계에 대한 ROC 분석에서 “다소높음” 이상 단계 면적 비율의 조합(D2+D3+D4)이 가장 높은 예측 성능을 보

였다(AUC = 0.743). 이는 Hanley and McNeil(1982)이 제시한 임상적 유용성 기준 0.7을 상회하는 수치로, 해당 조합의 실용적 가치를 시사한다. “매우높음” 단계의 면적 비율(D4) 단독으로는 상대적으로 낮은 성능을 보여(AUC = 0.627) 현행 기준에서 “매우높음” 등급 판정에서 “매우높음” 단계의 면적 비율(D4) $\geq 70\%$ 조건의 한계를 확인할 수 있다.

2. 최적 임계값 도출

각 단계 이상에 해당하는 면적, 즉 D4, D3+D4, D2+D3+D4 각각의 변수의 경우에 대한 ROC 곡선을 도출하였다(Figure 2). 이 ROC 곡선을 활용한 Youden's J statistic을 통해 통계적 최적값을 도출한 결과, D2+D3+D4 48%, D3+D4 2%, D4 1%가 도출되었다(Table 3). 이는 현행 기준의 일괄적인 임계값(70%)이 통계적 최적점과 상당한 괴리가 있음을 의미한다.

그러나 통계적 최적점에 해당하는 이 임계값(48, 2, 1)을 그대로 적용할 경우 “높음” 단계 이상(D3+D4)이 2%만 초과해도 “높음” 등급이 발령되어 “높음” 등급 발령 빈도가 과도하게 높아지고(31.7%), “다소높음” 발령 빈도가 너무 적어져(3.3%), 실용성이 떨어지는 문제가 발생한다(Table 4). 또한 D4의 경우 최적 임계값에서 민감도가 36.5%로 낮아 단독으로 활용하기에는 성능이 저하된다(Table 3).

따라서, ROC 분석에서 도출된 통계적 최적값과 현행 기준(70%)의 사이에서 최적 임계값을 도출하는 것이 필요하다. 최적 임계값 도출을 위해 다음 조건을 만족하는 임계값 조합을 5% 단위로 탐색하였다.

다양한 임계값 조합에서 발령 빈도, 산불 발생률, 예측 성능을 비교한 결과(Table 4, Figure 3), (40, 60, 15) 조합은 발령 빈도의 점진적 감소(Very High 3.8% < High 9.8% < Moderate 27.2% < Low 59.2%), 산불 발생률의 단조증가(2.1% \rightarrow 9.7% \rightarrow 15.4% \rightarrow 27.4%)를 만족하면서도 높은 예측 성능(민감도 81.0%, $\chi^2 = 3,969.7$)을 달성하여 최적 조합으로 선정되었다. 특히 (48, 2, 1) 조합 대비 발령 빈도의 실용성과 논리적 일관성 측면에서 우위를 보였다. 이에 따라, D2+D3+D4 40%, D3+D4 60%, D4 15%가 최적 임계값으로 도출되었다.

Table 2. Distribution and fire occurrence rate under the current classification criteria.

Class	Cases	Proportion(%)	Fire Occurrence Rate(%)	Remarks
Very high	36	0.06	8.33	Too rare to be operationally useful
High	6,571	10.71	19.86	Higher rate than “Very high”
Moderate	11,471	18.69	11.52	Acceptable level
Low	43,293	70.54	3.03	Too frequent

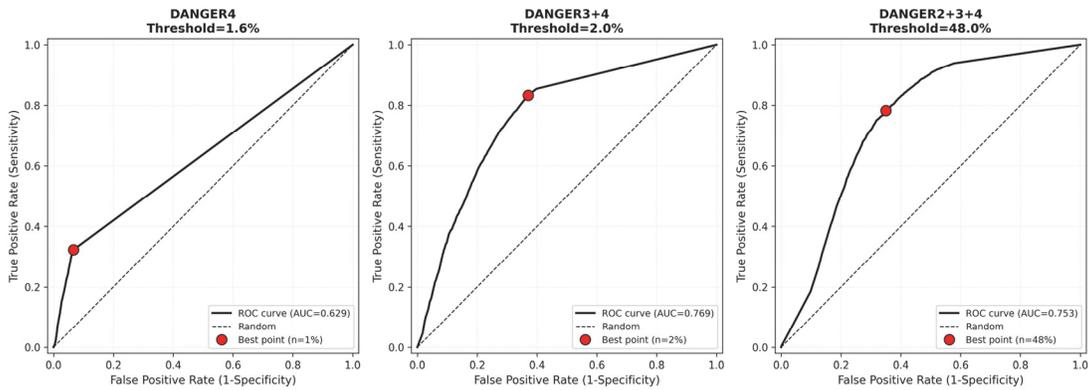


Figure 2. ROC Curves and optimal threshold points for each danger combination (D4, D3+D4, D2+D3+D4).

Table 3. Optimal thresholds derived from Youden's J Statistic.

Danger Level Combination	Optimal threshold (%)	Youden's J	Sensitivity (%)	Specificity (%)
D4	1	0.257	36.5	89.9
D3+D4	2	0.464	83.2	63.2
D2+D3+D4	48	0.433	78.4	64.9

Table 4. Changes in warning frequency, fire occurrence rate, and prediction accuracy by threshold combination.

Threshold Combination (% ≥ 51 , ≥ 66 , ≥ 86)	Warning frequency (%)				Monotonic decrease	Fire occurrence rate (%)				Monotonic increase	Performance			
	Very high	High	Moderate	Low		산불발생률 (Low→Moderate→High→Very high)	Accuracy	Sensitivity	Specificity		F1-Score	χ^2		
48, 2, 1	8.3	31.7	3.3	56.7	X	1.7 → 3.7 → 10.3 → 25.0	O	0.612	0.852	0.595	0.220	4741.7		
50, 10, 5	6.1	26.4	5.7	61.9	X	2.2 → 5.3 → 12.1 → 25.7	O	0.656	0.788	0.647	0.227	4307.0		
45, 20, 10	4.7	23.2	11.5	60.6	X	2.2 → 6.8 → 13.3 → 26.5	O	0.644	0.796	0.634	0.223	4181.5		
40, 40, 15	3.9	16.8	20.1	59.2	O	2.1 → 8.4 → 14.6 → 27.1	O	0.631	0.810	0.619	0.220	4082.9		
40, 60, 15	3.8	9.8	27.2	59.2	O	2.1 → 9.7 → 15.4 → 27.4	O	0.631	0.810	0.619	0.220	3969.7		
35, 65, 20	3.2	8.9	30.4	57.5	O	1.9 → 9.6 → 16.6 → 27.8	O	0.617	0.827	0.602	0.217	3959.3		

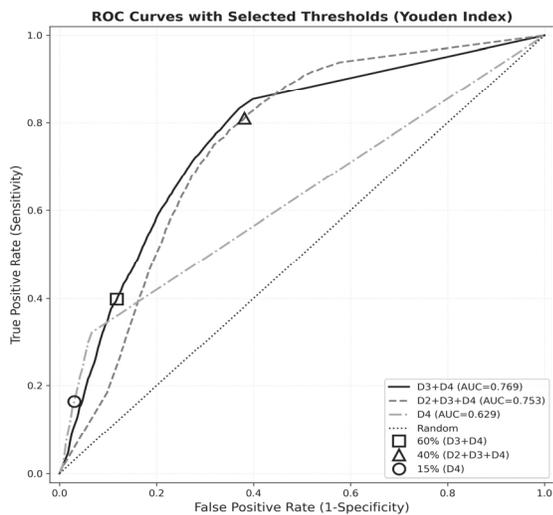


Figure 3. Visual representation of the decision point that simultaneously meets monotonicity, practicality, and statistical accuracy criteria (Optimal Combination: 40, 60, 15).

3. 새로운 분류 기준 적용

최적 임계값 도출에 따라, 다음과 같은 개선된 분류 기준을 도출하였다. “매우높음”을 분류할 때 D4는 단독으로 사용하기에는 민감도가 낮으므로, 높음을 분류할 때 사용하는 D3+D4 조건에 추가하여 사용하였다.

최종적으로 도출한 산불위험등급 분류 기준은 “매우 높음”은 위험지수 86 이상인 면적(D4) $\geq 15\%$ 이면서 위험지수 66 이상인 면적(D3+D4) $\geq 60\%$ 일 경우, “높음”은 위험지수 66 이상인 면적(D3+D4) $\geq 60\%$ 인 경우, “다소높음”은 위험지수 51 이상인 면적(D2+D3+D4) $\geq 40\%$ 인 경우, 낮음은 그 외의 경우에 발령된다(Table 5).

최종적으로 선택된 산불위험등급 최적 임계값이 ROC-곡선상의 통계적 최적값(Figure 2)과 보이는 차이는 Figure 3과 같다. 이는 도출된 최적 임계값이 산불위험등급 발령 빈도, 산불 발생률, 예측 성능 등 여러 논리적·기능적 조건을 만족

Table 5. Improved criteria of regional forest fire danger class.

Forest Fire Danger Class	Index Range	Area (%) (Number of pixels(%))
Very high	≥ 86 (D4)	≥ 15%
	≥ 66 (D3+D4)	≥ 60%
High	≥ 66 (D3+D4)	≥ 60%
Moderate	≥ 51 (D2+D3+D4)	≥ 40%
Low		Else

시키는 값으로서, 단순히 민감도와 특이도의 1:1 직선에서 가장 먼 지점이 선택되는 단순 통계적 최적값과는 다소간의 차이가 생길 수 있다는 점을 보여준다.

1) 개선 효과 검증

실증분석을 바탕으로 도출한 개선된 분류 기준을 적용한 결과, 등급별 분포의 극단적 불균형 문제가 크게 개선되었다. “매우높음” 등급이 2,346건(3.82%)으로 기존 36건 대비 65.2 배 증가하여 실용적 활용이 가능한 수준에 도달했으며, “낮음” 등급도 36,289건(59.13%)으로 11.41%p 감소하여 과도한 안전 편향이 완화되었다(Table 6).

가장 중요한 개선 사항은 등급별 발생률의 논리적 순서 확립이다. 개선된 기준에서는 “낮음” 2.06% < “다소높음” 9.70% < “높음” 15.37% < “매우높음” 27.41%로 더 높은 위험도에서 산불이 더 많이 발생하는 완벽한 단조증가를 달

성하였다(Table 6). 이는 현행 기준에서 관찰된 등급 역전 현상을 완전히 해결한 것으로, 분류 체계의 논리적 일관성을 확보했다.

2) 통계적 유의성 검증

카이제곱 검정 결과, 현행 기준과 개선된 기준 모두 산불위험등급과 산불 발생 간에 통계적으로 매우 유의한 연관성을 보였다($p < 0.001$). 그러나 카이제곱 통계량의 크기에서 의미 있는 차이가 관찰되었다(Table 7).

- 현행 기준: $\chi^2 = 3,302.6$ ($p < 0.001$)

- 개선된 기준: $\chi^2 = 3,969.7$ ($p < 0.001$)

개선된 기준의 카이제곱 통계량은 현행 기준 대비 667.1 증가(20.2% 증가)하여, 등급과 산불 발생 간의 통계적 연관

Table 6. Comparison of grade distribution and fire occurrence before and after improvement.

Grade	Current criteria	Improved criteria	Change	Fire occurrence Rate
Very High	0.06% (36 cases)	3.82% (2,346 cases)	× 65.2 increase	27.41%
High	10.71% (6,571 cases)	9.80% (6,017 cases)	- 0.91%p	15.37%
Moderate	18.69% (11,471 cases)	27.24% (16,719 cases)	+ 8.55%p	9.70%
Low	70.54% (43,293 cases)	59.13% (36,289 cases)	- 11.41%p	2.06%

Table 7. Comparison of predictive performance between current and improved criteria.

Metric	Current Criteria	Improved Criteria	Change	Interpretation
Accuracy	0.727	0.631	-0.095	Slight drop in overall accuracy
Sensitivity	0.667	0.810	+0.138	Significant improvement in fire detection
Precision	0.145	0.127	-0.018	Slight drop in precision
Specificity	0.731	0.619	-0.111	Slight increase in false alarms
F1-Score	0.239	0.220	-0.019	Minor reduction
AUC	0.713	0.746	+0.033	Slight increase (4.6%)
Youden Index	0.398	0.429	+0.031	Slight increase (7.8%)
χ^2	3,302.6	3969.7	+667.1	Significant increase (20.2%)
Cramér's V	0.232	0.254	+0.022	Improvement within moderate level

성이 강화되었음을 보여준다. 이는 개선된 등급 체계가 실제 산불 발생 패턴을 더 정확하게 반영한다는 것을 의미한다.

카이제곱 통계량은 표본 크기에 민감하므로, 실질적 연관성의 강도를 평가하기 위해 효과 크기 측정치인 Cramér's V를 계산한 결과는 아래와 같다.

- 현행 기준: $V = 0.232$ (중간 연관성, Moderate association)
- 개선된 기준: $V = 0.254$ (중간 연관성, Moderate association)

Cramér's V는 0.232에서 0.254로 0.022 증가(9.5% 증가)하여, 중간 연관성(Moderate association) 범주 내에서 실질적 향상을 보였다. Cohen(1988)의 해석 기준에 따르면, 자유도가 3인 경우 $0.20 \leq V < 0.30$ 범위는 중간 연관성으로 분류되며, 두 기준 모두 이 범주에 속하나 개선된 기준에서 상한에 더 근접한 값을 보였다(Table 7).

Cramér's V의 향상은 두 가지 중요한 의미가 있다. 첫째, 개선된 등급 체계가 산불 발생을 더 효과적으로 변별한다는 것을 보여준다. 둘째, 표본 크기의 영향을 제거한 순수한 연관성 강도의 증가를 나타내므로, 실질적으로 의미 있는 개선이 이루어졌음을 확인할 수 있다.

특히 현행 기준에서 관찰된 등급 역전 현상("높음" 19.9% > "매우높음" 8.1%)이 개선된 기준에서는 완벽한 단조증가("낮음" 2.1% < "다소높음" 9.7% < "높음" 15.4% < "매우높음" 27.4%)로 전환됨에 따라, 통계적 연관성도 함께 향상된 것으로 해석된다.

종합하면, 카이제곱 검정과 Cramér's V 분석 모두 개선된 분류 기준이 통계적으로 유의하며 실질적으로도 의미 있는 수준의 성능 향상을 달성했음을 입증한다. 이는 개선된 기준이 산불위험예보시스템의 예측력과 신뢰성을 실질적으로 높일 수 있음을 시사한다.

고 찰

본 연구에서 도출한 개선된 기준은 기존의 재해 위험도 분류 이론과 높은 일치성을 보인다. Kaplan and Garrick(1981)의 위험 정의에서 강조한 "발생 가능성"과 "결과의 심각성"을 D3+D4 조합과 D4 단독 조건으로 각각 구현함으로써, 고전적 위험 이론의 현대적 적용 사례를 제시했다.

Flach(2012)의 단조성 원칙 관점에서 현행 기준의 등급 역전 현상은 분류 체계의 근본적 결함을 의미한다. 본 연구에서 개선된 기준이 완벽한 단조증가를 달성한 것은 이론적 요구

사항을 충족하는 동시에 실용적 가치를 크게 향상시킨 성과로 해석된다. 이는 기존 연구들이 주로 예측 정확도에만 집중한 것과 달리, 분류 체계의 논리적 일관성까지 동시에 고려한 차별화된 접근이다.

본 연구의 실무적 기여는 "매우높음" 등급의 실용성 회복이다. 기존 0.06%에서 3.82%로의 증가는 단순한 수치 변화를 넘어 예보 시스템의 작동성 회복을 의미한다. Mason and Graham(2002)이 지적한 바와 같이 환경 위험 예측에서 극값 등급의 희소성은 의사결정자의 관심도 저하와 대응 체계 약화를 야기한다. 개선된 기준은 이러한 문제를 해결하여 실질적인 위험 관리 도구로서의 기능을 회복시켰다.

정책적 관점에서 등급별 발생률의 단조증가 달성은 대국민 신뢰도 향상에 직접적으로 기여할 것으로 예상된다. 현행 기준에서 "높음"보다 "매우높음"의 발생률이 낮은 모순적 상황은 예보의 신뢰성을 저하시키는 요인이었다. Jolliffe and Stephenson(2012)의 기상예보 신뢰성 연구에 따르면, 등급 체계의 논리적 일관성은 대중의 예보 수용도와 직결되며, 이는 궁극적으로 예방 효과로 이어진다.

본 연구는 재해 위험도 분류 분야에서 여러 이론적 기여를 제공한다. 첫째, 다중 방법론의 통합적 검증 체계를 통해 기존의 단편적 접근을 넘어선 종합적 분류 기준 개발 프레임워크를 제시했다. 둘째, 실증데이터 기반의 객관적 임계값 도출을 통해 전문가 경험에 의존하던 기존 관행에서 벗어난 과학적 접근을 구현했다. 셋째, 시간적 변동성을 체계적으로 고려한 적응적 분류 체계의 가능성을 입증했다.

앞으로 산불위험예보는 단순한 예보 전달을 넘어 산불예방과 피해저감까지 연결되어야 한다. 현행 산불위험예보에서 산불의 환경적 인자만을 고려하는 한계를 뛰어넘어, 향후 연구에서는 인구밀도, 도로밀도, 건물 분포 등 인위적 인자를 반영함으로써, 산불위험예보의 정밀도를 높일 뿐 아니라, 피해 예측까지도 가능한 산불위험예보시스템을 구축할 수 있도록 추가적인 연구가 필요하다. 또한 기상에 선행적인 관계를 가지지 않는 산림의 연료수분 변화 특성까지도 고려하여, 더욱 정확하고 정밀한 산불위험예보를 생산하는 연구가 필요하다고 하겠다.

결 론

본 연구는 현행 산불위험예보 등급 분류 기준에 존재하는 구조적 문제를 해결하기 위해 수행되었다. 최근 10년간의 실증데이터를 분석한 결과, 현행 기준은 70%라는 면적 기준으로 인해 "매우높음" 등급이 희소하고, "낮음" 등급이 큰 비중을 차지하는 편중 현상을 보였으며, "높음" 등급의 산불 발생

률이 “매우높음” 등급보다 높은 역전 현상이 발생하여 분류 체계의 논리적 일관성이 부족한 상황이었다. 본 연구는 ROC 곡선 분석을 적용하여 민감도와 특이도의 균형점을 반영하고, 논리적 일관성 등 여러 기준을 만족시키는 최적 분류 기준을 개발하였다.

개선된 기준 적용 결과, “매우높음” 등급의 발생 횟수는 실질적인 등급으로 활용할 수 있는 수준으로 증가하였고, “낮음” 등급은 적정 수준까지 감소하였다. 무엇보다도, 등급이 높아질수록 산불 발생확률이 높아지는 단조증가를 달성하여 이를 통해 분류 체계의 논리적 일관성을 획득하였다. 또한, “매우높음” 등급 발령 횟수가 실질적으로 활용 가능한 수준으로 증가하여, 예보의 실효성을 높일 수 있다.

본 연구 결과는 현행 국가산불위험예보시스템의 등급체계를 개선하는 데 직접적으로 활용될 수 있다. 산불위험등급 분류 기준 최적화를 통한 예보의 정확도와 논리적 일관성 개선은 예보에 대한 신뢰성을 높여 국민의 산불예보 활용 증가와 산불예방 효과로 이어질 것이다. 특히, 인력, 장비, 예산 등 한정된 진화자원을 진정한 고위험 상황에 집중적으로 투입할 수 있게 되어, 예방 효과의 극대화 및 조기 대응 효율성을 높일 수 있을 것이다. 본 연구를 통해 개선된 분류 기준의 도입은 산불예보 정확도 향상과 재해 관리 선진화에 기여하고, 국민의 생명과 재산을 산불로부터 보호하는 데 기여할 것으로 기대된다.

감사의 글

본 연구는 산림청 국립산림과학원 산림과학연구사업 <인공지능 기반 차세대 산불위험예보 알고리즘 개발(과제번호 FE0500-2025-02-2025)>의 지원으로 수행되었습니다.

References

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, U.S.A.

Cramér, H. 1946. *Mathematical methods of statistics*. Princeton University Press, Princeton, U.S.A.

Flach, P. 2012. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, Cambridge, U.K.

Hanley, J.A. and McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1): 29-36.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. *The elements*

of statistical learning: data mining, inference, and prediction. Springer, New York, U.S.A.

- Jolliffe, I.T. and Stephenson, D.B. (Eds.). 2012. *Forecast verification: a practitioner's guide in atmospheric science* (2nd ed.). Wiley-Blackwell, Chichester, U.K.
- Kaplan, S. and Garrick, B.J. 1981. On the quantitative definition of risk. *Risk Analysis* 1(1): 11-27.
- Lee, S. Y., Won, M. S. and Han, S. Y. 2005. Developing of Forest Fire Occurrence Danger Index Using Fuel and Topographical Characteristics on the Condition of Ignition Point in Korea. *Transactions of Korean Institute of Fire Science and Engineering* 19(4): 75-79.
- Mason, S.J. and Graham, N.E. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society* 128(584): 2145-2166.
- Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240(4857): 1285-1293.
- Weiss, G.M. 2004. Mining with rarity: a unifying framework. *ACM SIGKDD Explor. Explorations Newsletter* 6(1): 7-19.
- Won, M. S., Jang, K. C. and Yoon, S. H. 2018. Development of the National Integrated Daily Weather Index (DWI) Model to Calculate Forest Fire Danger Rating in the Spring and Fall. *Korean Journal of Agricultural and Forest Meteorology* 20(4): 348-356.
- Won, M. S., Yoon, S. H. and Jang, K. C. 2016. Developing Korean Forest Fire Occurrence Probability Model Reflecting Climate Change in the Spring of 2000s. *Korean Journal of Agricultural and Forest Meteorology* 18(4): 199-207.
- Won, M. S., Yoon, S. H., Koo, K. S. and Kim, K. H. 2011. Spatio-Temporal Analysis of Forest Fire Occurrences during the Dry Season between 1990s and 2000s in South Korea. *Journal of Korean Association of Geographic Information Studies* 14(3): 150-162.
- Youden, W.J. 1950. Index for rating diagnostic tests. *Cancer* 3(1): 32-35.

Manuscript Received : November 11, 2025

First Revision : December 9, 2025

Accepted : December 10, 2025