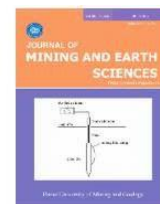




Journal of Mining and Earth Sciences

Website: <https://jmes.humg.edu.vn>



Drought risk prediction using remote sensing data and machine learning: A case study in Quang Tri province



Trang Thu Thi Tran ^{1,*}, Hung Le Trinh ², Thao Phuong Thi Do ¹, Ha Mai Nguyen ¹, Phu Van Le ²

¹ Hanoi University of Mining and Geology, Hanoi, Vietnam

² Le Quy Don Technical University, Hanoi, Vietnam

ARTICLE INFO

Article history:

Received 05th May 2025

Revised 13th Aug 2025

Accepted 23th Aug 2025

Keywords:

Drought index,
Drought,
Machine learning,
Quang Tri province,
Remote sensing.

ABSTRACT

Quang Tri Province was located in Vietnam's North Central Coast, frequently experiences severe climatic conditions including prolonged droughts. These conditions adversely affect agricultural production and local livelihoods. This study proposes a drought risk prediction model by integrating remote sensing data with machine learning techniques to support water resource management and mitigate drought-related damages. Remote sensing data from Landsat and Sentinel-2 satellites covering the period from 2016 to 2025 were collected and processed using the Google Earth Engine (GEE) cloud platform. Drought indices including NDVI, LSWI, NDWI, MSI, NDDI, SAVI, VCI, TCI, TVDI, and VHI were computed to assess temperature conditions, soil moisture, and vegetation health. Based on the TVDI index, data were categorized into five drought severity levels: no drought, mild drought, moderate drought, severe drought, and extreme drought. A training dataset was constructed by randomly selecting pixel samples representing each drought level. Three machine learning models including Random Forest (RF), Support Vector Machine (SVM), and Gradient Tree Boosting (GTB) were employed to classify and predict drought risk. The GTB model achieved the highest accuracy with an overall accuracy of 91.67% and a Kappa coefficient of 0.89, outperforming both SVM and RF models. The drought risk distribution map generated using the GTB model clearly highlights areas with high drought risk, particularly in the eastern and central regions of Quang Tri Province. In addition, 2019 and 2022 recorded severe and extreme drought areas exceeding 90% of the total area of the region.

Copyright © 2025 Hanoi University of Mining and Geology. All rights reserved.

*Corresponding author

E - mail: tranthithutrang@humg.edu.vn

DOI: 10.46326/JMES.2025.66(5).02



Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <https://tapchi.humg.edu.vn>



Ứng dụng viễn thám và kỹ thuật học máy trong dự báo nguy cơ hạn hán, thử nghiệm tại tỉnh Quảng Trị

Trần Thị Thu Trang ^{1*}, Trịnh Lê Hùng ², Đỗ Thị Phương Thảo¹, Nguyễn Mai Hà ¹, Lê Văn Phú ²

¹ Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

² Đại học Lê Quý Đôn, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

Quá trình:

Nhận bài 05/5/2025

Sửa xong 13/8/2025

Chấp nhận đăng 23/8/2025

Từ khóa:

Chỉ số hạn,
Dữ liệu viễn thám,
Hạn hán,
Học máy,
Tỉnh Quảng Trị.

TÓM TẮT

Tỉnh Quảng Trị nằm ở khu vực Bắc Trung Bộ Việt Nam, thường xuyên chịu ảnh hưởng của điều kiện khí hậu khắc nghiệt, đặc biệt là hạn hán kéo dài, gây tác động tiêu cực đến sản xuất nông nghiệp và cuộc sống của người dân. Nghiên cứu đề xuất mô hình dự báo nguy cơ hạn hán bằng cách tích hợp dữ liệu viễn thám và kỹ thuật học máy, qua đó hỗ trợ công tác quản lý tài nguyên nước và giảm thiểu thiệt hại do hạn hán. Dữ liệu viễn thám từ các vệ tinh Landsat và Sentinel-2 trong giai đoạn 2016-2025 được thu thập và xử lý trên nền tảng điện toán đám mây Google Earth Engine (GEE). Tiếp theo, các chỉ số hạn hán bao gồm NDVI, LSWI, NDWI, MSI, NDDI, SAVI, VCI, TCI, TVDI và VHÍ được tính toán để phản ánh tình trạng nhiệt độ, độ ẩm đất và sức khỏe thảm thực vật. Dựa trên chỉ số TVDI, dữ liệu được phân loại thành năm mức độ hạn hán: không hạn, hạn nhẹ, hạn trung bình, hạn nặng và hạn rất nặng. Tập dữ liệu huấn luyện cho mô hình học máy được xây dựng bằng cách chọn ngẫu nhiên các điểm ảnh đại diện cho từng mức độ hạn hán. 03 mô hình học máy gồm Random Forest (RF), Support Vector Machine (SVM) và Gradient Tree Boosting (GTB) được áp dụng để phân loại và dự báo nguy cơ hạn hán. Kết quả cho thấy mô hình GTB đạt độ chính xác cao nhất với chỉ số độ chính xác tổng thể và hệ số Kappa lần lượt là 91,67% và 0,89, cao hơn so với SVM và RF. Bản đồ phân bố nguy cơ hạn hán được xây dựng từ mô hình GTB phản ánh rõ ràng các khu vực có nguy cơ hạn hán cao, đặc biệt là ở phía đông và trung tâm tỉnh Quảng Trị. Bên cạnh đó, năm 2019 và 2022 ghi nhận diện tích hạn nặng và rất nặng vượt quá 90% tổng diện tích khu vực.

© 2025 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

*Tác giả liên hệ

E - mail: tranthithutrang@humg.edu.vn

DOI: 10.46326/JMES.2025.66(5).02

1. Mở đầu

Việt Nam là một trong những quốc gia chịu ảnh hưởng nặng nề bởi biến đổi khí hậu, đặc biệt là hạn hán (Phan và Ngô, 2013). Trong những năm gần đây, tình trạng thiếu nước, khô hạn ngày càng trở nên phổ biến và nghiêm trọng, đặc biệt tại khu vực miền Trung và Tây Nguyên, nơi có lượng mưa phân bố không đều và thời gian mùa khô kéo dài (Dũng và nnk., 2016; Phùng và Nguyễn, 2023). Mặc dù Việt Nam đã có nhiều nỗ lực trong việc quản lý tài nguyên nước và thích ứng với biến đổi khí hậu, song thiệt hại do hạn hán gây ra cho sản xuất nông nghiệp, đời sống dân cư và hệ sinh thái vẫn ở mức đáng báo động (Khảm và Khánh, 2022). Ngoài ra, các đợt hạn hán gần đây đã làm hàng chục nghìn hecta cây trồng bị ảnh hưởng, hàng trăm nghìn người dân thiếu nước sinh hoạt, đặc biệt nghiêm trọng tại các tỉnh ven biển miền Trung như Quảng Trị, Hà Tĩnh và Bình Thuận (Khảm và Khánh, 2022; Trịnh và Đào, 2019).

Hạn hán là hiện tượng tự nhiên có quá trình diễn biến phức tạp, bị chi phối bởi nhiều yếu tố khí hậu, thổ nhưỡng, độ ẩm đất, lớp phủ bề mặt và hoạt động của con người (He và nnk., 2025). Việc giám sát và dự báo hạn hán theo phương pháp truyền thống chủ yếu dựa vào các trạm đo khí tượng - thủy văn, tuy nhiên mạng lưới trạm phân bố không đồng đều và không đủ độ phủ cho các khu vực miền núi, bán sơn địa. Điều này dẫn đến hạn chế trong việc đánh giá mức độ hạn hán ở quy mô không gian rộng (Mallick và nnk., 2025; Oyarzabal và nnk., 2025).

Với sự phát triển của công nghệ địa không gian, đặc biệt là dữ liệu viễn thám và trí tuệ nhân tạo, việc giám sát hạn hán đã có những bước tiến đáng kể (Tamrakar và nnk., 2024; Zamani và nnk., 2025). Dữ liệu ảnh vệ tinh cho phép theo dõi sự thay đổi của thực vật, độ ẩm và nhiệt độ bề mặt đất trong thời gian dài và trên diện rộng. Các chỉ số hạn hán đã được ứng dụng hiệu quả trong việc đánh giá mức độ khô hạn và sức khỏe thảm thực vật (Kogan, 1995; Quiring và Ganesh, 2010; Mukherjee và nnk., 2018). Bên cạnh đó, các công trình nghiên cứu như của Kogan (1997), Serban và Maftai (2025) đã ứng dụng viễn thám để xác định hạn hán và nguy cơ hạn hán. Tuy nhiên, việc chỉ sử dụng chỉ số ảnh đơn lẻ dẫn tới thiếu khả năng phản ánh đầy đủ các yếu tố phức tạp và mối quan hệ phi tuyến giữa các yếu tố hạn hán.

Trong thời gian gần đây, cùng với sự phát triển mạnh mẽ của trí tuệ nhân tạo, nhiều nghiên cứu đã ứng dụng các mô hình học máy và học sâu để nâng cao hiệu quả giám sát và dự báo hạn hán, đặc biệt trong các môi trường địa lý phức tạp như vùng đồi núi, miền Trung Việt Nam (Liu và nnk., 2018; Ha và nnk., 2024). Các mô hình như Random Forest (RF), Support Vector Machine (SVM), Gradient Tree Boosting (GTB) và các mạng học sâu (Deep Neural Networks - DNN) đã cho thấy tiềm năng lớn trong việc dự báo hạn hán ở các quy mô khác nhau, với độ chính xác cao hơn các phương pháp truyền thống (Elbeltagi và nnk., 2023; Wu và Chen, 2017).

Bài báo này đề xuất mô hình xây dựng bản đồ dự báo nguy cơ hạn hán tại khu vực tỉnh Quảng Trị bằng cách tích hợp dữ liệu viễn thám, GIS và mô hình học máy. Ba thuật toán học máy phổ biến gồm Random Forest (RF), Support Vector Machine (SVM) và Gradient Tree Boosting (GTB) được triển khai nhằm lựa chọn mô hình phù hợp nhất với điều kiện thực tế của khu vực nghiên cứu. Tập dữ liệu đầu vào bao gồm các chỉ số hạn hán tính từ ảnh vệ tinh Landsat và Sentinel-2 trong giai đoạn 2016-2025. Kết quả dự báo được đánh giá hiệu năng mô hình thông qua các chỉ số độ chính xác bao gồm độ chính xác tổng thể và chỉ số Kappa. Kết quả nghiên cứu góp phần cung cấp công cụ hỗ trợ giám sát hạn hán, phục vụ quy hoạch sản xuất nông nghiệp và quản lý tài nguyên nước bền vững trong bối cảnh biến đổi khí hậu ngày càng gia tăng.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Dữ liệu và khu vực nghiên cứu

Trong nghiên cứu sử dụng dữ liệu ảnh viễn thám từ hai hệ thống vệ tinh là Landsat và Sentinel-2, trong khoảng thời gian từ năm 2016 đến 2025. Mỗi năm xây dựng 01 cặp ảnh đại diện. Việc sử dụng đa nguồn dữ liệu cho phép tận dụng ưu điểm độ phân giải không gian cao của ảnh Sentinel 2 và các kênh hồng ngoại nhiệt của ảnh Landsat. Các ảnh vệ tinh được xử lý và phân tích trên nền tảng Google Earth Engine (GEE), một hệ thống điện toán đám mây mạnh mẽ, cho phép truy cập và xử lý khối lượng lớn dữ liệu viễn thám một cách hiệu quả.

Dữ liệu Landsat (bao gồm Landsat 7 ETM+, Landsat 8 OLI/TIRS và Landsat 9 OLI-2/TIRS-2) cung cấp ảnh với độ phân giải không gian 30 mét

và chu kỳ lặp lại 16 ngày, phù hợp cho việc theo dõi các biến động về thảm thực vật và độ ẩm đất. Trong khi đó, dữ liệu Sentinel-2 (MSI) cung cấp ảnh với độ phân giải cao hơn ($10 \div 20$ m) và chu kỳ lặp lại 5 ngày, giúp tăng cường khả năng giám sát chi tiết các khu vực bị ảnh hưởng bởi hạn hán.

Tỉnh Quảng Trị, nằm ở khu vực Bắc Trung Bộ Việt Nam, là một trong những địa phương chịu ảnh hưởng nặng nề của biến đổi khí hậu, đặc biệt là hiện tượng hạn hán kéo dài (Hình 1). Với địa hình đa dạng, bao gồm vùng ven biển, đồng bằng, trung du và miền núi, tạo nên sự phong phú về sinh thái nhưng cũng đồng thời làm tăng tính phức tạp trong việc quản lý tài nguyên nước (Cổng thông tin điện tử tỉnh Quảng Trị, 2025a).

Khí hậu Quảng Trị mang đặc trưng nhiệt đới gió mùa, phân thành hai mùa rõ rệt: mùa khô từ tháng 3 đến tháng 9 và mùa mưa từ tháng 10 đến tháng 2 năm sau. Trong mùa khô, gió Tây Nam khô nóng thường xuyên thổi mạnh, làm tăng nhiệt độ

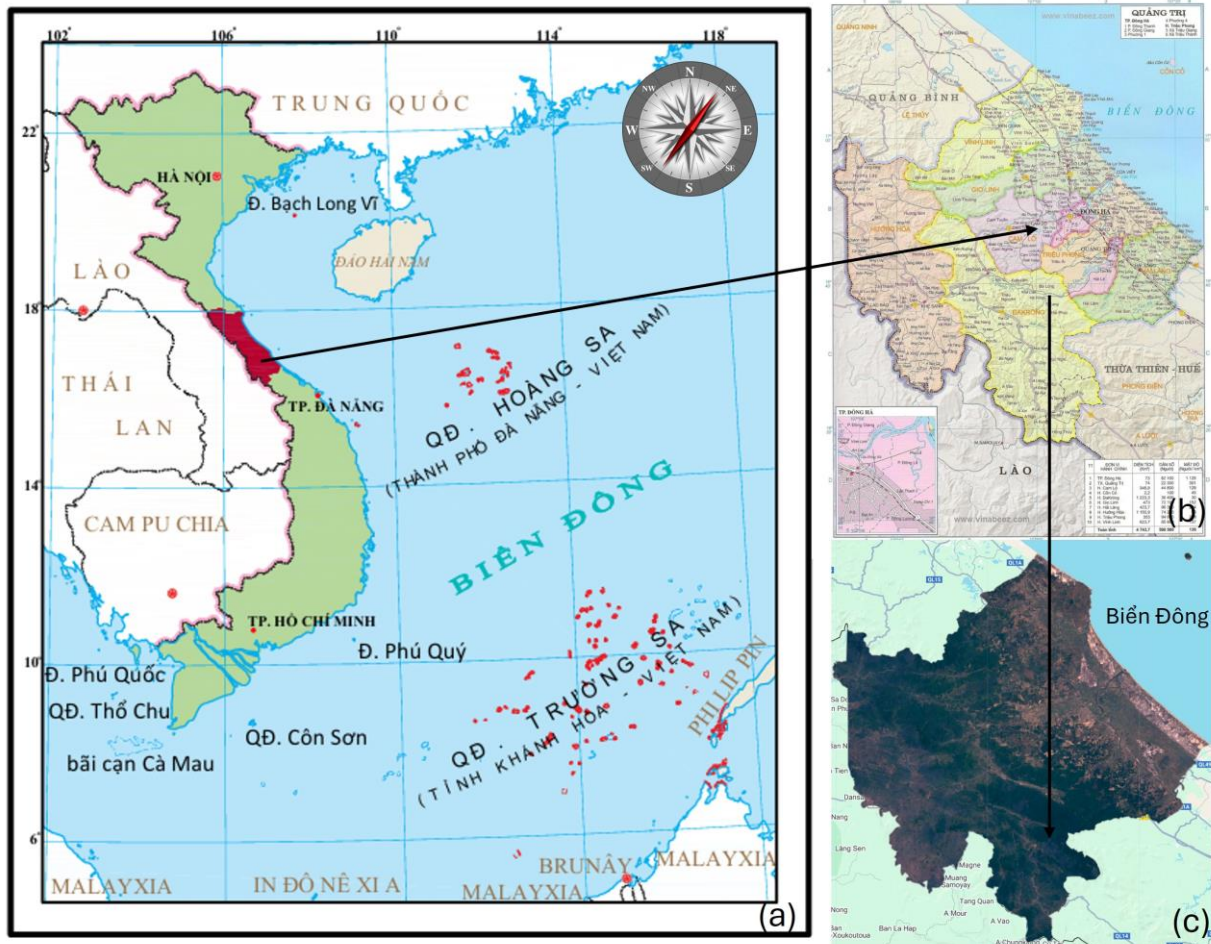
và giảm độ ẩm, dẫn đến tình trạng hạn hán nghiêm trọng.

2.2. Phương pháp nghiên cứu

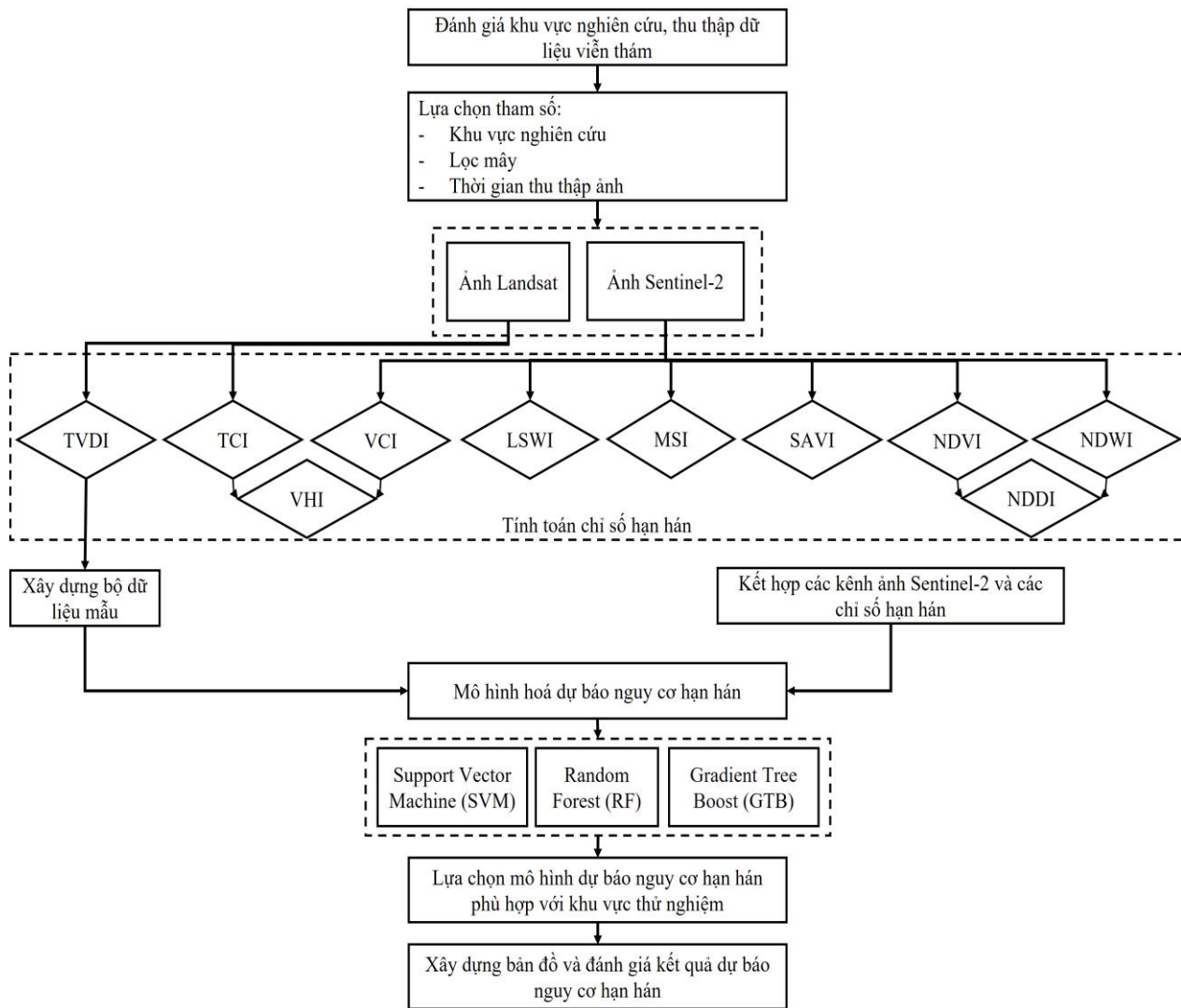
Nghiên cứu được triển khai với mục tiêu xây dựng mô hình dự báo nguy cơ hạn hán tại tỉnh Quảng Trị bằng cách kết hợp dữ liệu viễn thám đa nguồn và các kỹ thuật học máy. Toàn bộ quy trình được triển khai trên nền tảng điện toán đám mây Google Earth Engine (GEE) nhằm tận dụng khả năng xử lý dữ liệu lớn và tự động hóa. Quy trình xử lý dữ liệu được trình bày cụ thể trong Hình 2, bao gồm các bước như sau:

Bước 1: Đánh giá, thu thập dữ liệu

Bước đầu tiên của quy trình nghiên cứu là đánh giá tổng quan khu vực và tiến hành thu thập dữ liệu viễn thám phục vụ phân tích hạn hán. Tỉnh Quảng Trị được lựa chọn làm khu vực nghiên cứu do thường xuyên chịu ảnh hưởng bởi tình trạng



Hình 1. Mô tả vị trí khu vực nghiên cứu.



Hình 2. Sơ đồ mô hình dự báo nguy cơ hạn hán.

khô hạn kéo dài vào mùa khô, đặc biệt tại khu vực phía Đông và trung du. Nghiên cứu sử dụng dữ liệu ảnh vệ tinh từ hai nguồn chính: Landsat và Sentinel-2 với thời gian thu thập giai đoạn từ năm 2016 đến năm 2025.

Thời gian lựa chọn ảnh được giới hạn trong các tháng mùa khô, cụ thể từ tháng 3 đến tháng 9 hằng năm. Đây là giai đoạn dễ xảy ra tình trạng thiếu hụt độ ẩm bề mặt và giảm mạnh sức khỏe thực vật nên có khả năng phản ánh rõ ràng các dấu hiệu hạn hán qua phổ phản xạ.

Sau khi thu thập, ảnh vệ tinh được tiền xử lý bằng các bước: lọc mây và bóng mây; cắt ảnh theo ranh giới hành chính tỉnh Quảng Trị bằng shapefile khu vực nghiên cứu; chuẩn hóa phổ phản xạ nhằm đảm bảo đồng nhất giữa các năm và

các nguồn ảnh. Dữ liệu sau xử lý được lưu trữ thành bộ ảnh tổ hợp để tính toán các chỉ số hạn hán và phục vụ xây dựng bộ dữ liệu huấn luyện cho mô hình.

Bước 2: Tính toán chỉ số hạn hán

Tiếp theo, các chỉ số hạn hán được tính toán từ ảnh viễn thám đã tiền xử lý. Các chỉ số này bao gồm nhóm phản ánh tình trạng độ ẩm đất và nước bề mặt như NDWI (Normalized Difference Water Index), LSWI (Land Surface Water Index), MSI (Moisture Stress Index), NDDI (Normalized Drought Difference Index); nhóm phản ánh trạng thái thảm thực vật như NDVI (Normalized Difference Vegetation Index), SAVI (Soil-Adjusted Vegetation Index), VCI (Vegetation Condition

Index); và nhóm chỉ số tổng hợp như TVDI (Temperature Vegetation Dryness Index), VHI (Vegetation Health Index), trong đó VHI là chỉ số trung bình giữa VCI và TCI (Temperature Condition Index). Cụ thể, TVDI và TCI được tính toán từ ảnh Landsat, các chỉ số còn lại được tính toán từ ảnh Sentinel 2. Các chỉ số được tính theo công thức chuẩn từ phổ dữ liệu ảnh và được tổ chức thành bảng dữ liệu để đưa vào mô hình học máy. Công thức của từng chỉ số ảnh được tính toán như sau:

NDWI (Gao, 1996):

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (1)$$

LSWI (Xiang và nnk., 2020):

$$LSWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2)$$

MSI (Hunt và Rock, 1989):

$$MSI = \frac{SWIR}{NIR} \quad (3)$$

NDVI (Tucker, 1979):

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (4)$$

NDDI (Gu và nnk., 2007):

$$NDDI = \frac{NDVI - NDWI}{NDVI + NDWI} \quad (5)$$

SAVI (Huete, 1988):

$$SAVI = \frac{(1 + L)(NIR - Red)}{NIR + Red + L}, L = 0,5 \quad (6)$$

VCI (Kogan, 1995):

$$VCI = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \times 100 \quad (7)$$

TCI (Kogan, 1995):

$$TCI = \frac{T_{max} - T}{T_{max} - T_{min}} \times 100 \quad (8)$$

VHI (Kogan, 1997):

$$VHI = \alpha \cdot VCI + (1 - \alpha) \cdot TCI, \alpha = 0,5 \quad (9)$$

TVDI (Sandholt và nnk., 2002):

$$TVDI = \frac{T_s - T_{min}(NDVI)}{T_{smax}(NDVI) - T_{min}(NDVI)} \quad (10)$$

Bước 3: Khởi tạo bộ dữ liệu mẫu

Dựa trên các chỉ số hạn hán đã tính toán, bước tiếp theo trong quy trình nghiên cứu là xây dựng bộ dữ liệu huấn luyện cho các mô hình học máy. Trong giai đoạn này, chỉ số TVDI được sử dụng làm cơ sở để gán nhãn mức độ hạn hán cho các điểm ảnh. TVDI là chỉ số phản ánh trực tiếp mức độ khô hạn bề mặt, được tính toán từ mối quan hệ giữa nhiệt độ bề mặt đất (Land Surface Temperature - LST) và chỉ số NDVI, do đó rất phù hợp để phân vùng hạn hán (Trịnh, 2014).

Giá trị TVDI được chia thành năm khoảng ngưỡng, tương ứng với năm mức độ hạn hán như sau (Chen và nnk., 2023):

- 0,0÷0,2: Không hạn,
- 0,2÷0,4: Hạn nhẹ,
- 0,4÷0,6: Hạn trung bình,
- 0,6÷0,8: Hạn nặng,
- 0,8÷1,0: Hạn rất nặng.

Từ phân vùng theo chỉ số TVDI, nghiên cứu tiến hành xây dựng bộ dữ liệu huấn luyện theo kịch bản phân loại đa lớp, với mục tiêu nhận diện chính xác các mức độ khác nhau của hạn hán theo không gian và thời gian. Việc phân loại theo nhiều mức độ hạn cho phép mô hình học máy không chỉ phân biệt vùng có và không có hạn hán, mà còn định lượng được mức độ nghiêm trọng của hiện tượng hạn hán tại mỗi thời điểm (Danandeh, 2021).

Từ mỗi nhóm hạn hán theo phân loại TVDI, một tập hợp điểm ảnh đại diện được chọn ngẫu nhiên, đảm bảo số lượng mẫu tương đương và phân bố đều theo không gian. Cụ thể, nghiên cứu lựa chọn 200 điểm ảnh từ mỗi nhóm (tổng cộng 1000 điểm ảnh) để tạo thành tập dữ liệu huấn luyện và kiểm tra mô hình. Mỗi điểm ảnh mang nhãn tương ứng với mức độ hạn hán từ 1 đến 5, và đi kèm với bộ đặc trưng bao gồm các giá trị của 10 chỉ số hạn hán đã tính toán.

Cách tiếp cận này cho phép áp dụng các thuật toán học máy trong bài toán phân loại đa lớp (multi-class classification), qua đó nâng cao khả năng nhận diện chính xác và chi tiết bản đồ nguy cơ hạn hán trên toàn khu vực nghiên cứu. Dữ liệu huấn luyện được xây dựng đảm bảo tính đại diện cả về không gian và mức độ hạn, là yếu tố nền tảng

quan trọng cho các bước huấn luyện và đánh giá mô hình ở các giai đoạn tiếp theo.

Bước 4: Mô hình hóa dự báo nguy cơ hạn hán bằng học máy

Sau khi hoàn tất quá trình xây dựng bộ dữ liệu huấn luyện với các nhãn hạn hán phân cấp theo chỉ số TVDI, nghiên cứu tiến hành bước mô hình hóa nhằm xây dựng công cụ dự báo nguy cơ hạn hán tại khu vực nghiên cứu. Trong bước này, các thuật toán học máy được áp dụng để học từ mối quan hệ giữa các đặc trưng đầu vào và mức độ hạn hán, từ đó tạo ra mô hình có khả năng dự báo mức độ hạn cho những điểm ảnh chưa biết nhãn.

Ba mô hình học máy phổ biến và hiệu quả được lựa chọn để so sánh và đánh giá hiệu suất trong nghiên cứu bao gồm: Random Forest (RF), Support Vector Machine (SVM) và Gradient Tree Boosting (GTB). Đây là các thuật toán có khả năng xử lý tốt dữ liệu phi tuyến tính, đa chiều và thường được sử dụng trong các bài toán phân loại ảnh viễn thám (Maxwell và nnk., 2018).

Random Forest (RF) là một phương pháp học máy dựa trên tổ hợp nhiều cây quyết định (Breiman, 2001). Mỗi cây học từ một mẫu ngẫu nhiên của dữ liệu và kết quả cuối cùng được lấy theo nguyên tắc bỏ phiếu đa số. Ưu điểm chính của RF là khả năng chống overfitting tốt, hoạt động hiệu quả với dữ liệu có nhiều chiều và không yêu cầu giả định phân phối của dữ liệu.

Support Vector Machine (SVM) là một thuật toán phân loại mạnh, hoạt động trên nguyên lý tìm siêu phẳng tối ưu để phân tách các lớp dữ liệu (Suthaharan, 2016). Trong trường hợp dữ liệu không tuyến tính, SVM sử dụng các hàm kernel để ánh xạ dữ liệu sang không gian đặc trưng cao hơn, giúp tăng khả năng phân biệt giữa các lớp. Tuy nhiên, SVM thường yêu cầu tinh chỉnh tham số kỹ lưỡng và không linh hoạt bằng RF khi số lớp phân loại tăng cao.

Gradient Tree Boosting (GTB) là một thuật toán boosting kết hợp nhiều cây quyết định theo chuỗi, trong đó mỗi cây mới được huấn luyện nhằm giảm sai số còn lại từ cây trước (Friedman, 2002). GTB thường cho độ chính xác cao nhưng yêu cầu thời gian huấn luyện lâu hơn và có nguy cơ overfitting nếu không kiểm soát tốt các siêu tham số.

Để đánh giá hiệu năng mô hình, dữ liệu được chia thành hai tập: tập huấn luyện (70%) và tập

kiểm tra (30%). Các mô hình được huấn luyện trên tập huấn luyện và sau đó được áp dụng để phân loại các điểm trong tập kiểm tra. Các chỉ số đánh giá bao gồm độ chính xác tổng thể (tỷ lệ giữa số điểm phân loại đúng và tổng số điểm kiểm tra) và hệ số Kappa được sử dụng để so sánh hiệu quả giữa các mô hình.

Bước 5: Dự báo và xây dựng bản đồ nguy cơ hạn hán

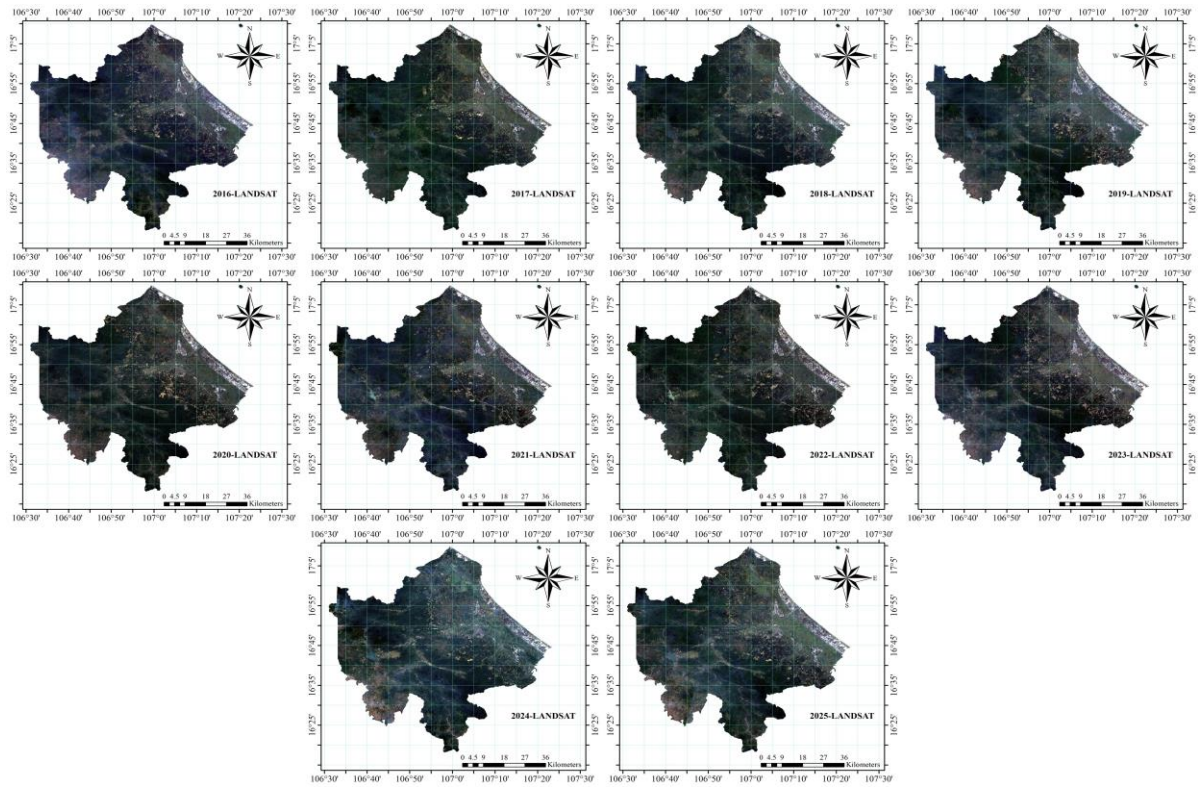
Sau khi chọn được mô hình có hiệu suất tốt nhất, mô hình này được áp dụng lên toàn bộ tập dữ liệu ảnh viễn thám tại tỉnh Quảng Trị để dự báo bản đồ nguy cơ hạn hán. Mỗi điểm ảnh được gán một giá trị dự báo tương ứng với một trong năm mức độ hạn hán. Kết quả được trình bày dưới dạng bản đồ phân vùng nguy cơ hạn hán theo từng năm, thể hiện sự thay đổi không gian của hạn hán trong giai đoạn từ 2016 đến 2025.

Đồng thời, nghiên cứu tiến hành so sánh sự thay đổi diện tích các vùng hạn qua các năm, đánh giá sự thay đổi của hạn hán theo thời gian.

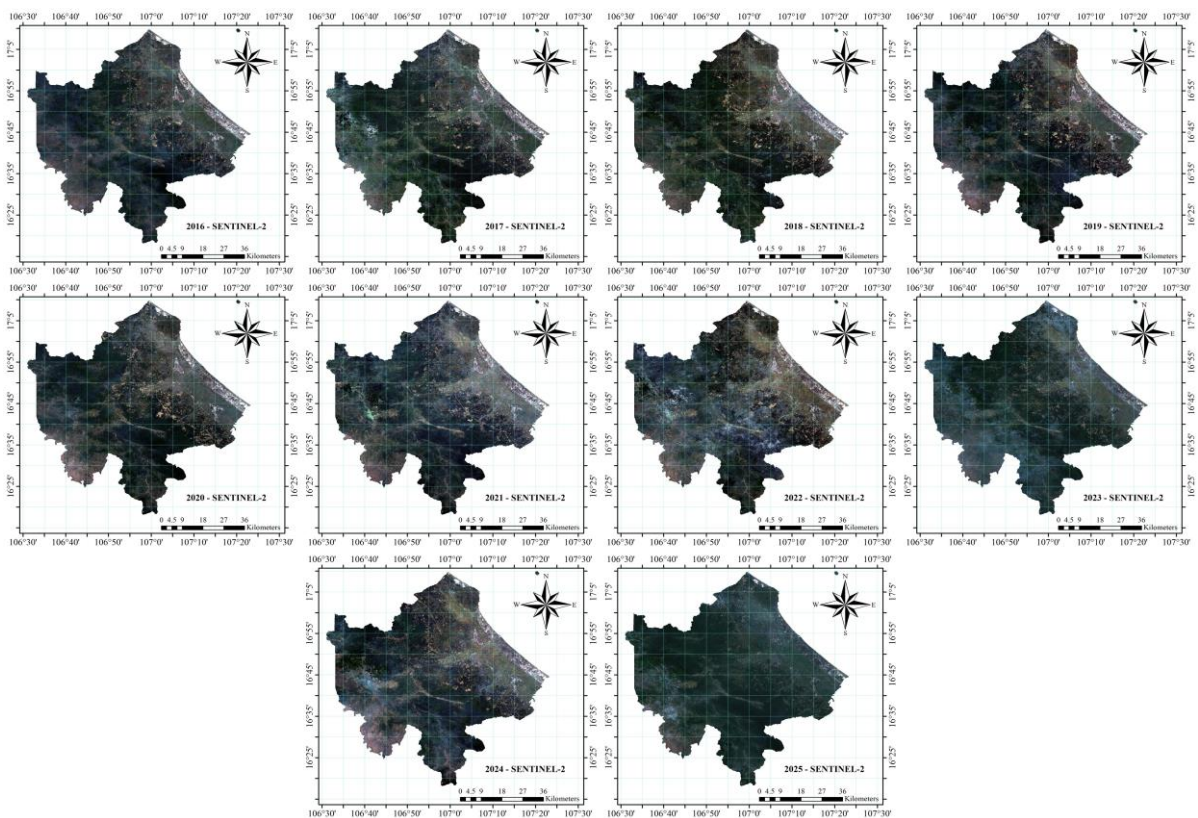
3. Kết quả và thảo luận

Hình 3 và Hình 4 lần lượt trình bày chuỗi hình ảnh Landsat và Sentinel-2 theo năm khu vực tỉnh Quảng Trị giai đoạn 2016 - 2025. Các hình ảnh được biên tập theo tổ hợp màu R-G-B (đỏ, xanh lục, xanh lam) nhằm nổi bật các đặc điểm bề mặt như thực vật, đất và nước. Chuỗi hình ảnh cung cấp cái nhìn trực quan về tốc độ thay đổi cảnh quan tại Quảng Trị. Đồng thời, bộ dữ liệu là hình ảnh đầu vào cho các bước tính toán tiếp theo.

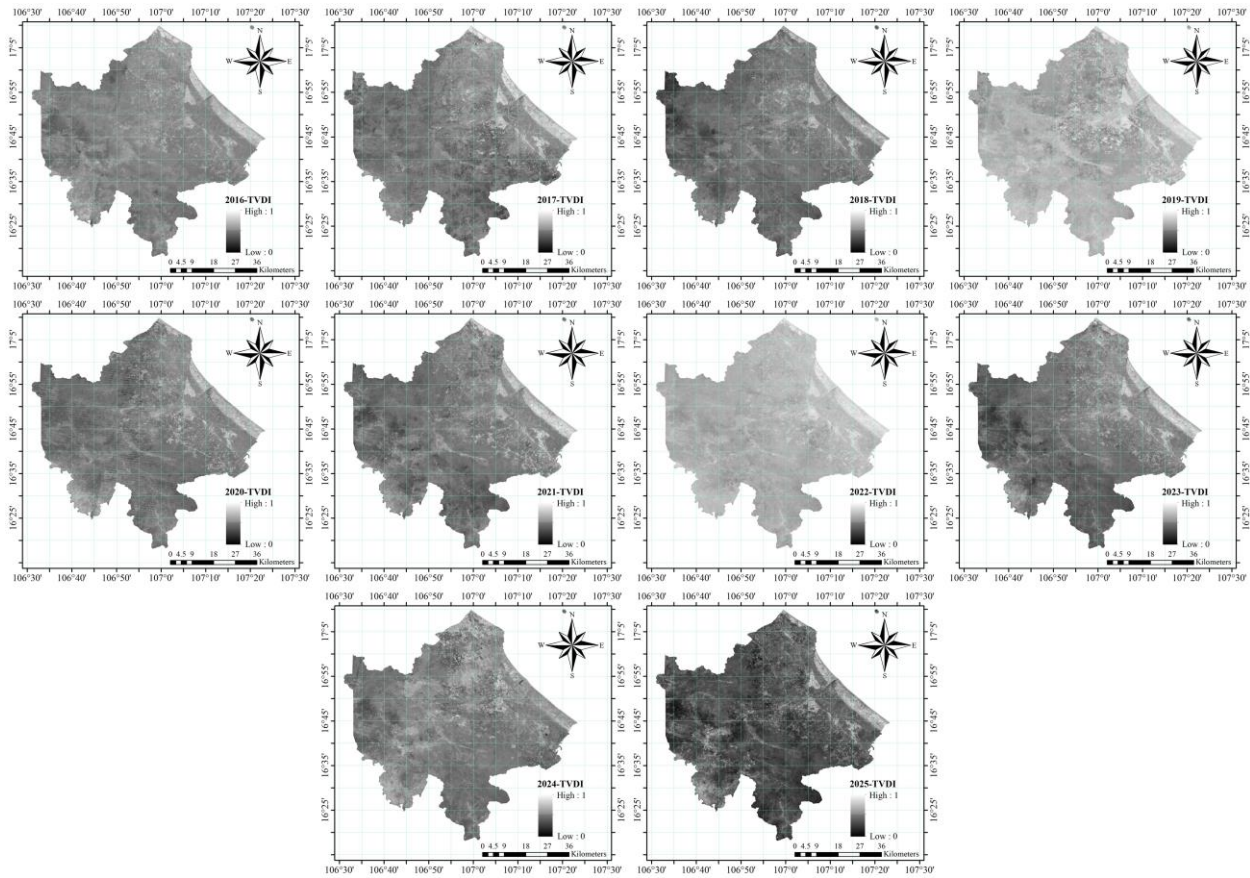
Hình 5 hiển thị hình ảnh chỉ số TVDI khu vực tỉnh Quảng Trị từ năm 2016 đến 2025, phản ánh mức độ khô hạn dựa trên mối quan hệ giữa nhiệt độ bề mặt và độ che phủ thực vật. Chỉ số TVDI dao động từ 0 (ẩm ướt) đến 1 (khô hạn), với thang màu từ đen tới trắng. Trong giai đoạn từ năm 2016 đến năm 2018, phần lớn khu vực có màu xám và xám đậm, đặc biệt ở vùng miền núi phía Tây, cho thấy điều kiện ẩm ướt, giá trị TVDI thấp. Khu vực phía Đông có màu sáng hơn, báo hiệu khô hạn nhẹ (giá trị TVDI tăng). Tiếp theo, ảnh TVDI năm 2019 và năm 2022 có màu sáng với nhiều khu vực chuyển sang màu xám nhạt và trắng đặc biệt ở vùng đồng bằng phía Đông, cho thấy khô hạn gia tăng (TVDI tiến gần 1). Bên cạnh đó, từ năm 2023 đến năm 2025, xu hướng khô hạn giảm dần với các hình ảnh có thang màu tối hơn.



Hình 3. Hình ảnh Landsat khu vực tỉnh Quảng Trị giai đoạn 2016-2025.



Hình 4. Hình ảnh Sentinel-2 khu vực tỉnh Quảng Trị giai đoạn 2016-2025.



Hình 5. Hình ảnh chỉ số TVDI khu vực tỉnh Quảng Trị giai đoạn 2016-2025.

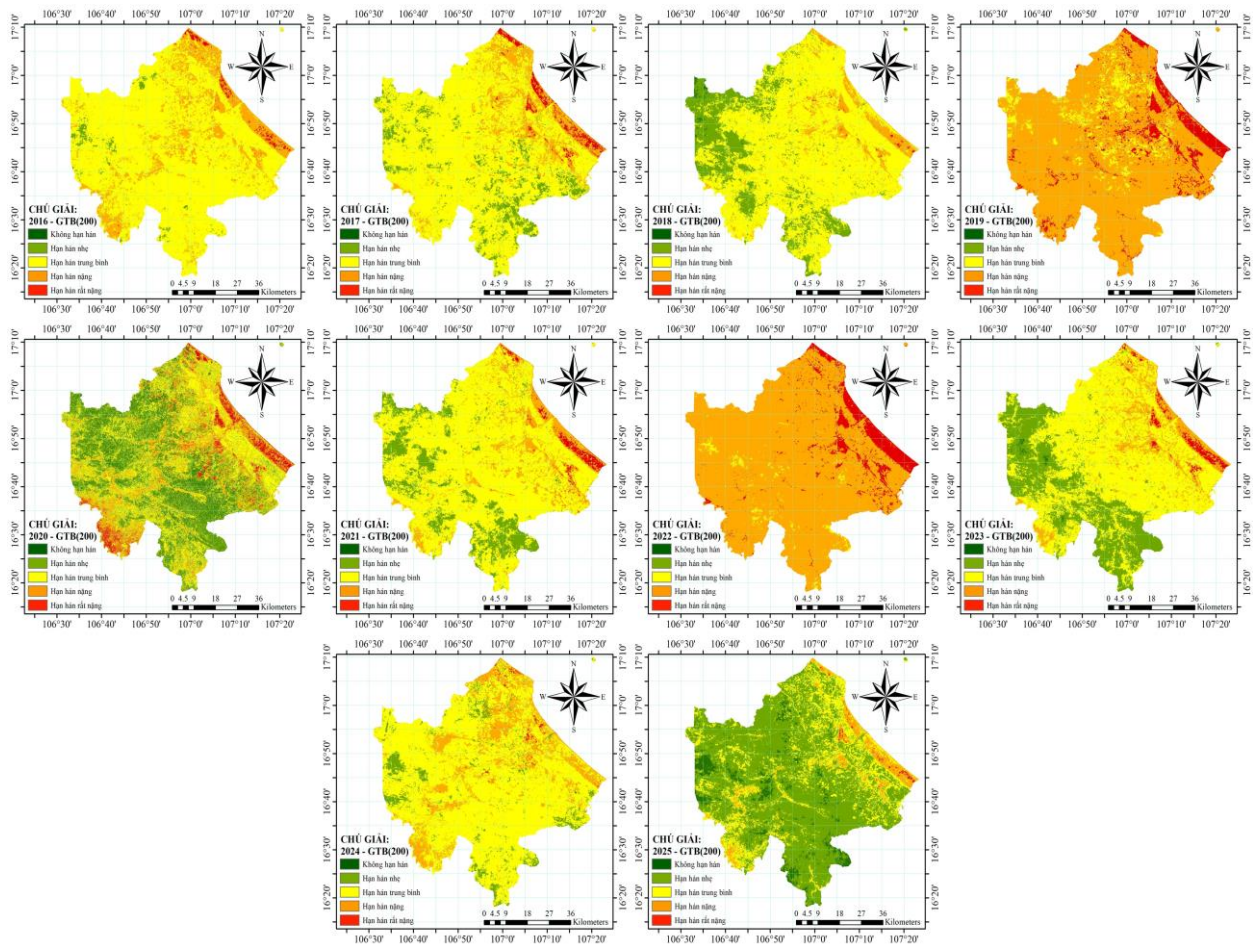
Tiếp theo, các tham số của thuật toán GTB và RF được tối ưu với số lượng cây (numberOfTrees) là 200. Đối với thuật toán SVM, tham số cost được đặt là 100. Các tham số còn lại của các thuật toán được đặt theo giá trị mặc định. Kết quả phân loại hạn hán năm 2025 cho thấy thuật toán GTB cho kết quả tốt nhất, độ chính xác cao nhất với chỉ số độ chính xác tổng thể và hệ số Kappa lần lượt là 91,67% và 0,89, hiệu quả hơn RF và SVM (lần lượt có độ chính xác tổng thể là 90,33% và 89,00%, hệ số Kappa là 0,87 và 0,86).

Do đạt hiệu quả cao nhất, thuật toán GTB đã được lựa chọn để thành lập bản đồ dự báo nguy cơ hạn hán giai đoạn 2016-2025 tại khu vực nghiên cứu tỉnh Quảng Trị. Hình 6 trình bày kết quả phân loại hạn hán khu vực tỉnh Quảng Trị giai đoạn 2016-2025 theo thuật toán GTB

Bảng 1 và Bảng 2 trình bày diện tích và tỉ lệ diện tích hạn hán khu vực tỉnh Quảng Trị giai đoạn 2016-2025. Trong suốt giai đoạn nghiên cứu, diện tích không hạn tại tỉnh Quảng Trị luôn chiếm tỷ lệ rất thấp. Giai đoạn 2016-2019, diện tích không

hạn gần như bằng 0, phản ánh tình trạng khô hạn phổ biến và kéo dài trên toàn tỉnh. Năm 2020 ghi nhận sự cải thiện rõ rệt với diện tích không hạn đạt 809,63 km² (16,71%), tuy nhiên, các năm sau đó, tỷ lệ này lại giảm mạnh, dao động từ 0,00% đến 5,45%. Điều này cho thấy, mặc dù có những năm điều kiện khí hậu thuận lợi hơn, nhưng xu hướng chung vẫn là nguy cơ hạn hán bao trùm hầu hết toàn khu vực.

Bên cạnh đó, diện tích hạn nhẹ giai đoạn 2016-2017, chiếm tỷ lệ thấp, từ 1,60% đến 9,09%. Tuy nhiên, năm 2018 tỉ lệ này tăng tỷ lệ này tăng vọt lên tới 19,26%. Năm 2019, diện tích hạn nhẹ chỉ đạt 1,32 km² (0,03%) phản ánh điều kiện khí hậu khô hạn đang mở rộng. Giai đoạn 2020-2025, diện tích hạn nhẹ dao động mạnh, tăng cao năm 2023 (26,04%), sau đó giảm dần về mức thấp hơn 6,83% rồi tăng lên 66,51%. Xu hướng này cho thấy sự thay đổi thất thường, có thể do biến động về mưa mùa khô hoặc tác động của các biện pháp điều tiết nước.



Hình 6. Kết quả phân loại hạn hán khu vực tỉnh Quảng Trị giai đoạn 2016-2025 theo thuật toán GTB.

Bảng 1. Diện tích và tỉ lệ diện tích hạn hán khu vực tỉnh Quảng Trị giai đoạn 2016-2020.

Diện tích	2016		2017		2018		2019		2020	
	km ²	%	km ²	%	km ²	%	km ²	%	km ²	%
Không hạn	0,09	0,00	2,10	0,04	17,40	0,36	0,23	0,00	809,63	16,71
Hạn nhẹ	77,69	1,60	440,25	9,09	933,11	19,26	1,32	0,03	1083,93	22,37
Hạn trung bình	3880,25	80,08	3635,53	75,03	3512,94	72,50	389,67	8,04	1892,98	39,06
Hạn nặng	846,98	17,48	663,80	13,70	373,94	7,72	4026,70	83,10	825,95	17,04
Hạn rất nặng	40,75	0,84	104,09	2,15	8,36	0,17	427,84	8,83	233,27	4,81

Bảng 2. Diện tích và tỉ lệ diện tích hạn hán khu vực tỉnh Quảng Trị giai đoạn 2021-2025.

Diện tích	2021		2022		2023		2024		2025	
	km ²	%	km ²	%	km ²	%	km ²	%	km ²	%
Không hạn	2,34	0,05	0,16	0,00	14,77	0,30	1,99	0,04	264,22	5,45
Hạn nhẹ	655,46	13,53	0,47	0,01	1261,96	26,04	331,08	6,83	3223,08	66,51
Hạn trung bình	3521,98	72,68	94,10	1,94	2899,41	59,83	3640,01	75,12	1070,80	22,10
Hạn nặng	582,31	12,02	4359,77	89,97	583,33	12,04	852,58	17,59	272,38	5,62
Hạn rất nặng	83,67	1,73	391,25	8,07	86,29	1,78	20,10	0,41	15,27	0,32

Hạn trung bình là mức độ phổ biến nhất trong giai đoạn 2016-2018, chiếm từ 72÷80% diện tích tỉnh. Tuy nhiên, từ năm 2019 trở đi, diện tích hạn trung bình giảm dần, đạt 39,06% vào năm 2020. Năm 2021, tỷ lệ này tăng trở lại lên 72,68%, nhưng sau đó tiếp tục suy giảm liên tục, xuống mức 22,10% vào năm 2025. Sự giảm sút này phản ánh sự chuyển dịch từ mức độ hạn trung bình sang hạn nặng hoặc rất nặng, cho thấy tình trạng khô hạn ngày càng nghiêm trọng.

Ngoài ra, năm 2016, diện tích hạn nặng chiếm 17,48%, nhưng tăng mạnh từ năm 2018 trở đi, đạt đỉnh điểm vào năm 2019 với 83,10% - năm có diện tích hạn nặng lớn nhất trong giai đoạn 2016-2020. Từ 2021-2022, tỷ lệ hạn nặng có sự biến động mạnh, tăng từ 12,02% lên đến 89,97%. Xu hướng này cho thấy hạn nặng đang trở thành mức độ chiếm ưu thế, là dấu hiệu đáng lo ngại cho môi trường và sản xuất nông nghiệp khu vực. Giai đoạn 2023-2025 ghi nhận diện tích hạn nặng có sự giảm dần (Cổng thông tin điện tử tỉnh Quảng Trị, 2025b; Sở Tài nguyên và Môi trường tỉnh Quảng Trị, 2025).

Trong giai đoạn 2016-2018, diện tích hạn rất nặng luôn chiếm tỷ lệ thấp (~2%). Tuy nhiên, từ năm 2019, diện tích này bắt đầu tăng lên, đạt 8,83% vào năm 2019 và 8,07% vào năm 2022. Năm 2025, diện tích hạn rất nặng giảm xuống còn 0,32%, có thể do chuyển dịch về hạn nặng hoặc sự cải thiện cục bộ. Mặc dù diện tích nhỏ, nhưng sự gia tăng của hạn rất nặng là dấu hiệu cảnh báo về mức độ tổn thương sinh thái và sản xuất cao nhất.

4. Kết luận

Nghiên cứu đề xuất mô hình dự báo nguy cơ hạn hán cho tỉnh Quảng Trị bằng cách tích hợp dữ liệu viễn thám và thuật toán học máy. Sử dụng dữ liệu từ vệ tinh Landsat và Sentinel-2 trong giai đoạn 2016-2025, các chỉ số hạn hán bao gồm NDVI, LSWI, NDWI, MSI, NDDI, SAVI, VCI, TCI, TVDI và VHI đã được tính toán để phản ánh tình trạng độ ẩm đất và sức khô thảm thực vật. Dựa trên chỉ số TVDI, dữ liệu được phân loại thành năm mức độ hạn hán: không hạn, hạn nhẹ, hạn trung bình, hạn nặng và hạn rất nặng. Tập dữ liệu huấn luyện được xây dựng bằng cách chọn ngẫu nhiên các điểm ảnh đại diện cho từng mức độ hạn hán. Các mô hình học máy gồm Random Forest (RF), Support Vector Machine (SVM) và Gradient Tree Boosting (GTB) được áp dụng để phân loại và

dự báo nguy cơ hạn hán. Kết quả cho thấy mô hình GTB đạt độ chính xác cao nhất với độ chính xác tổng thể và chỉ số Kappa lần lượt là 91,67% và 0,89, vượt trội so với SVM và RF. Bản đồ phân bố nguy cơ hạn hán được xây dựng từ mô hình GTB phản ánh rõ ràng các khu vực có nguy cơ hạn hán cao, đặc biệt là ở phía đông và trung tâm tỉnh Quảng Trị. Nghiên cứu khẳng định hiệu quả của việc kết hợp dữ liệu viễn thám và kỹ thuật học máy trong việc dự báo hạn hán, cung cấp công cụ hữu ích cho các nhà quản lý trong việc lập kế hoạch ứng phó và giảm thiểu tác động của hạn hán. Mặc dù nghiên cứu đã đạt được kết quả khả quan, nhưng hạn chế về dữ liệu viễn thám như độ phân giải không gian, ảnh hưởng của mây và thiếu dữ liệu thực tế từ các trạm quan trắc có thể ảnh hưởng đến độ chính xác. Các nghiên cứu tiếp theo cần sử dụng dữ liệu đa nguồn, áp dụng phương pháp học sâu và thực hiện khảo sát hiện trường để cải thiện độ chính xác và tính khả thi của mô hình.

Lời cảm ơn

Bài báo có sử dụng một phần số liệu và kết quả của đề tài Cơ sở mã số T25-39.

Đóng góp của các tác giả

Trần Thị Thu Trang - lên ý tưởng, viết phần phương pháp và hoàn thiện bản thảo; Trịnh Lê Hùng - xây dựng chương trình; Đỗ Thị Phương Thảo - đánh giá các kết quả thực nghiệm và biên tập bản thảo; Lê Văn Phú - lọc dữ liệu; Nguyễn Mai Hà - biên tập và phân tích số liệu.

Tài liệu tham khảo

- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Chen, A., Jiang, J., Luo, Y., Zhang, G., Hu, B., Wang, X., & Zhang, S. (2023). Temperature vegetation dryness index (TVDI) for drought monitoring in the Guangdong Province from 2000 to 2019. *PeerJ*, 11, e16337.
- Cổng thông tin điện tử tỉnh Quảng Trị (2025a). Điều kiện tự nhiên. <https://www.quangtri.gov.vn/new-chi-ti%E1%BA%BFT-tin-t%E1%BB%95ng-quan/-/view-article/1/1573433122481/1717639121658>. Truy cập ngày 22-4-2025.

- Cổng thông tin điện tử tỉnh Quảng Trị (2025b). Kiểm tra tình hình ứng phó với hạn hán, nắng nóng tại Vĩnh Linh, Gio Linh. https://www.quangtri.gov.vn/chi-tiet-tin/-/view-article/1/138482_41113627/1577456441574. Truy cập ngày 22-4-2025.
- Danandeh Mehr, A. (2021). Drought classification using gradient boosting decision tree. *Acta Geophysica*, 69(3), 909-918.
- Dũng, L. H., Lê Giang, N., & Phương, T. T. (2016). Hạn hán năm 2015-2016 trên hai lưu vực sông chính ở Tây Nguyên. *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*, 58(6).
- Elbeltagi, A., Pande, C.B., Kumar, M. *et al.* Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environ Sci Pollut Res* 30, 43183-43202 (2023). <https://doi.org/10,1007/s11356-023-25221-3>.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4), 367-378.
- Gao, B. C. (1996). NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3), 257-266.
- Gu, Y., Brown, J. F., Verdin, J. P., & Wardlow, B. D. (2007). A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophysical Research Letters*, 34(6).
- Ha, D. H., Duc, P. N., Luong, T. H., Duc, T. T., Ngoc, T. T., Minh, T. N., & Minh, T. N. (2024). Application of Artificial Intelligence to Forecast Drought Index for the Mekong Delta. *Applied Sciences*, 14(15), 6763. <https://doi.org/10,3390/app14156763>.
- He, Z., Gu, X., Wang, M., & Xu, M. (2025). The coupled effect of climate change and LUCC on meteorological drought in a karst drainage basin, Southwest China. *Scientific Reports*, 15(1), 9134. <https://doi.org/10,1038/s41598-025-93035-z>.
- Huete, A. R., (1988). A Soil-adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25(3): 295-309. DOI: 10,1016/0034-4257(88)90106-X.
- Hunt, E. R., & Rock, B. N. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*, 30(1), 43-54.
- Khảm, D. V., Khánh, Đ. Q. (2022). Đánh giá đặc điểm khí hậu và điều kiện khí hậu nông nghiệp huyện Kỳ Anh tỉnh Hà Tĩnh. *Tạp chí Khí tượng Thủy văn* 2022, 733, 15-28; doi:10,36335/VNJHM.2022(733).15-28.
- Kogan, F.N. (1995). Application of vegetation index and brightness temperature for drought detection. *Advances in Space Research*, 15(11), 91-100. [https://doi.org/10,1016/0273-1177\(95\)00079-T](https://doi.org/10,1016/0273-1177(95)00079-T).
- Kogan, F. N. (1997). Global drought watch from space. *Bulletin of the American meteorological society*, 78(4), 621-636.
- Mallick, J., Alqadhi, S., Alsubih, M., & Alkahtanis, M. (2025). Integrating traditional and advanced technologies for drought monitoring and management: a systematic review of global methodologies and applications. *Theoretical and Applied Climatology*, 156(5), 1-18. <https://doi.org/10,1007/s00704-025-05469-0>.
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9), 2784-2817.
- Mukherjee, S., Mishra, A. & Trenberth, K.E. (2018) Climate Change and Drought: a Perspective on Drought Indices. *Curr Clim Change Rep* 4, 145-163 (2018). <https://doi.org/10,1007/s40641-018-0098-x>.
- Liu, Z. N., Li, Q. F., Nguyen, L. B., Xu, G. H. (2018). Comparing Machine-Learning Models for Drought Forecasting in Vietnam's Cai River Basin. *Polish Journal of Environmental Studies*, 27(6), 2633-2646. <https://doi.org/10.15244/pjoes/80866>.

- Oyarzabal, R. S., Santos, L. B., Cunningham, C., Broedel, E., de Lima, G. R., Cunha-Zeri, G., ... & Cunha, A. P. (2025). Forecasting drought using machine learning: a systematic literature review. *Natural Hazards*, 1-29. <https://doi.org/10,1007/s11069-025-07195-2>.
- Phùng, M. L., & Nguyễn, Á. T. (2023). Tác động của thiên tai và biến đổi khí hậu tới sản xuất nông nghiệp và thủy sản của hộ nông dân khu vực miền Trung và Tây Nguyên Việt Nam. *Tạp chí Khoa học Thăng Long: Khoa học Ứng dụng*, 1, 47-60. <https://science.thanglong.edu.vn/index.php/vola/article/view/9>.
- Phan, V. T., & Ngô, Đ. T. (2013). Biến đổi khí hậu ở Việt Nam: Một số kết quả nghiên cứu, thách thức và cơ hội trong hội nhập quốc tế. *VNU Journal of Science: Earth and Environmental Sciences*, 29(2).
- Quiring, S. M. & Ganesh, S. (2010). Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas. *Agricultural and Forest Meteorology*, 150(3), 330-339. <https://doi.org/10,1016/j.agrformet.2009,11,015>.
- Sandholt, I., Rasmussen, K., & Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79(2-3), 213-224.
- Serban, C., & Maftai, C. (2025). Remote Sensing Evaluation of Drought Effects on Crop Yields Across Dobrogea, Romania, Using Vegetation Health Index (VHI). *Agriculture*, 15(7), 668.
- Sở Tài nguyên và Môi trường tỉnh Quảng Trị. (2022). Báo cáo 2022. <https://quangtriclimate.vn/2022/10/page/2/>. Truy cập ngày 22-4-2025.
- Suthaharan, S. (2016). *Support vector machine. In Machine learning models and algorithms for big data classification: thinking with examples for effective learning* (pp. 207-235). Boston, MA: Springer US.
- Tamrakar, Y., Das, I.C. & Sharma, S. Machine learning for improved drought forecasting in Chhattisgarh India: a statistical evaluation. *Discov Geosci* 2, 84 (2024). <https://doi.org/10,1007/s44288-024-00089-z>.
- Trịnh, L.H. (2014). Ứng dụng dữ liệu viễn thám hồng ngoại nhiệt Landsat nghiên cứu độ ẩm đất trên cơ sở chỉ số khô hạn nhiệt độ thực vật. *Tạp chí Các khoa học về Trái Đất*, 36(3), 262-270.
- Trịnh L. H., & Đào K. H. (2019). Ứng dụng viễn thám đánh giá nguy cơ hạn hán khu vực huyện Bắc Bình, tỉnh Bình Thuận. *HCMUE Journal of Science*, 5(70), 128. [https://doi.org/10.54607/hcmue.js.0,5\(70\).539\(2015\)](https://doi.org/10.54607/hcmue.js.0,5(70).539(2015)).
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
- Wu, J., & Chen, Y. F. (2017), August). Research of using RF model to drought forecast on Huaihe River. In *IOP Conference Series: Earth and Environmental Science* (Vol. 82, No. 1, p. 0120 16). IOP Publishing.
- Xiang, K., Yuan, W., Wang, L., & Deng, Y. (2020). An LSWI-Based Method for Mapping Irrigated Areas in China Using Moderate-Resolution Satellite Data. *Remote Sensing*, 12(24), 4181. <https://doi.org/10,3390/rs12244181>
- Zamani, H., Pakdaman, Z., Shakari, M. et al. Enhancing drought monitoring with a multivariate hydrometeorological index and machine learning-based prediction in the south of Iran. *Environ Sci Pollut Res* 32, 5605-5627 (2025). <https://doi.org/10,1007/s11356-025-36049-4>.