

TỰ ĐỘNG HÓA

Vietnam
AUTOMATION

NGÀY NAY

HỘI TỰ ĐỘNG HÓA VIỆT NAM



TODAY

Số 267+268 | Tháng 5+6/2023

ISSN 1859-0551

Chào mừng 98 năm
NGÀY BÁO CHÍ CÁCH MẠNG VIỆT NAM

- **Lịch sử báo chí số**
- **Khai thác "mỏ vàng" dữ liệu số**
Thúc đẩy phát triển kinh tế số

Trong số này

Số 267+268 | Tháng 5+6/2023

Diễn đàn/Forum

- 14** Khai thác "mỏ vàng" dữ liệu số thúc đẩy tăng trưởng kinh tế số
Explore the "gold mine" of digital data to boost digital economy growth
- 20** Lịch sử hình thành báo chí số
History of Digital Journalism
- 23** Tự động hóa sáng tạo nội dung truyền hình nhìn từ phóng sự viết bằng AI của Đài Truyền hình TP.HCM (HTV)
Automating TV content creation seen from the HTV's reportage written by AI

Công nghệ ứng dụng/Application technology

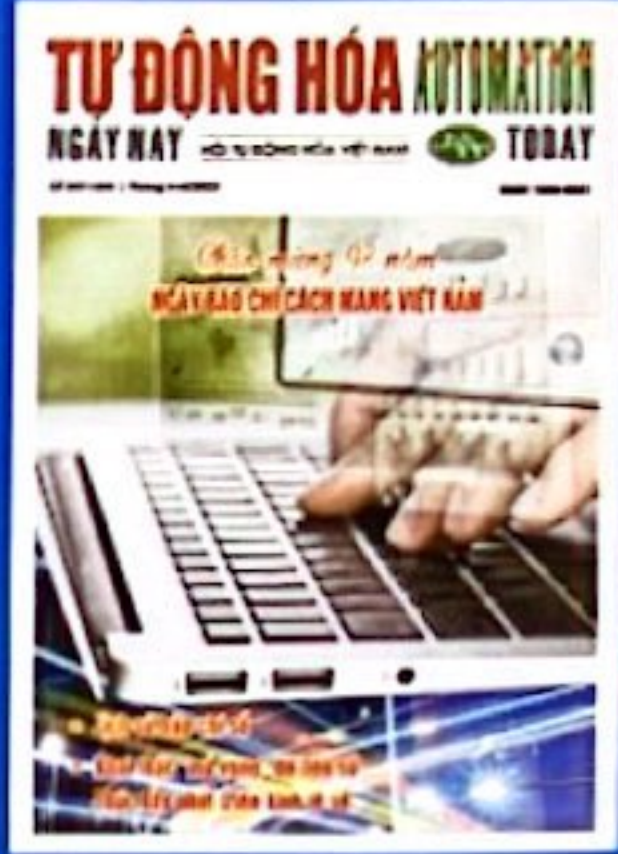
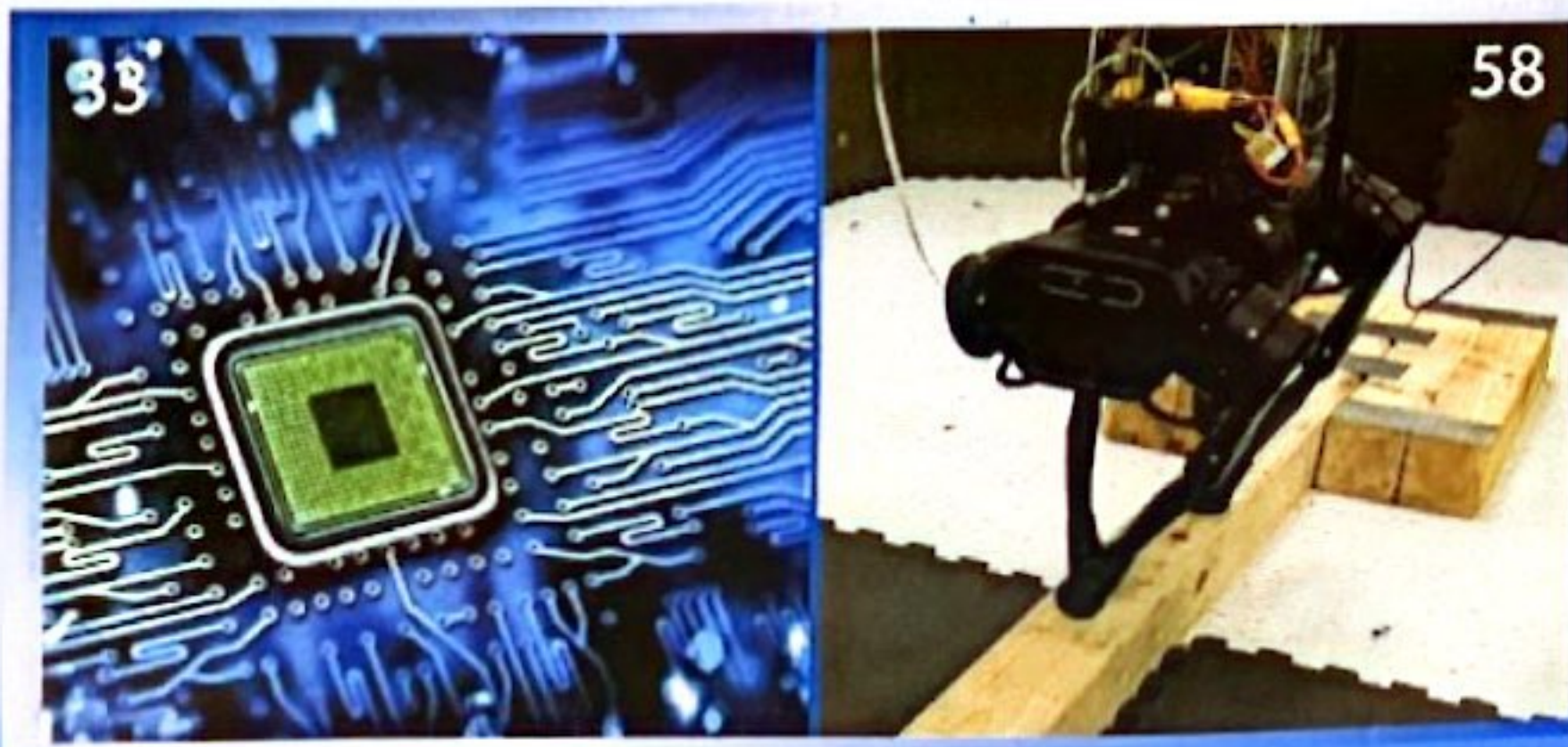
- 29** Fortinet cải tiến mạnh mẽ các giải pháp giúp CIO quản lý mạng hiệu quả hơn
Fortinet drastically improves solutions to help CIOs manage networks more efficiently
- 33** Chip bán dẫn: linh kiện làm thay đổi thế giới
Semiconductor chip: the component which changes the world
- 38** Ứng dụng học máy trong dự báo nguy cơ xảy ra thiên tai
Application of machine learning in assessing the susceptibility of natural disasters
- 41** Nghiên cứu và phát triển Robot dạng người thông minh IVASTBot ứng dụng trong giao tiếp và phục vụ con người
Research and Develop IVASTBot intelligent humanoid robot applied in communication and serving people

Rô bốt/Robot

- 55** Đại học Công nghiệp Hà Nội giành ngôi vô địch Robocon Việt Nam 2023
Hanoi University of Industry has won the championship of Robocon Vietnam 2023
- 57** Khám phá năng lực mới của những chú chó robot
Discover the new power of robot dogs

Doanh nghiệp sản phẩm/Enterprises and products

- 59** Khát vọng xây dựng một Rạng Đông mới bằng con đường chuyển đổi số
Aspiration to build a new Rang Dong by digital transformation
- 64** Ứng dụng nhận diện khuôn mặt thông minh trong quản lý học sinh - sinh viên tự động
Application of smart face recognition in automatic student management



HỘI ĐỒNG BIÊN TẬP

TS. Nguyễn Quân
PGS.TS. Hoàng Minh Sơn
PGS.TS. Nguyễn Hữu Công
PGS.TS. Tạ Cao Minh
TS. Trương Đình Châu
PGS.TS. Lê Văn Doanh
PGS.TS. Trần Anh Dũng
GS.TSKH. Nguyễn Tiến Dũng
PGS.TS. Trịnh Lương Miên
TS. Đinh Văn Hiến
TS. Trần Văn Lê
Ths. Lê Việt Cường

CỐ VẤN

TS. Dương Nguyên Bình
Nhà báo Phạm Việt Tiến

TỔNG BIÊN TẬP

TS. Trần Bảo Khánh

PHÓ TỔNG BIÊN TẬP

PGS.TS. Thái Quang Vinh
Ths. Trần Thị Giang

THƯ KÝ TÒA SOẠN

Ths. Trần Thị Giang

TÒA SOẠN TẠP CHÍ

P.406 - Tòa nhà Trung tâm Công nghệ cao,
Trường ĐHBK Hà Nội, Số 1, Đại Cồ Việt
Q. Hai Bà Trưng, Hà Nội
Tel: 84.24.35665327 / 35665328 | Fax: 84.24.35665327
Hotline quảng cáo: 0934519698 - 0985725187 - 0984585839
E-mail: tapchitudonghoa@gmail.com
E-mail phát hành: phathanh.tdh@gmail.com
Hotline phát hành phía Bắc: 090 219 0307
VĂN PHÒNG ĐẠI DIỆN PHÍA NAM: Lầu 8, số 224
Điện Biên Phủ, phường Võ Thị Sáu, quận 3, Tp.HCM.
Tel: 028.22118588
Email: vnautomate.hcm@gmail.com / Hotline : 0908855888
Tạp chí điện tử:
www.vnautomate.net

Giấy phép xuất bản số:

276/GP-BTTTT ngày 05/09/2014
của Bộ Thông Tin và Truyền thông
In tại Công ty TNHH MTV - In quân đội 1

Giá: 50.000 VND

ỨNG DỤNG HỌC MÁY TRONG DỰ BÁO NGUY CƠ XẢY RA THIÊN TAI

LÊ VĂN HƯNG

Khoa Công nghệ Thông tin, Trường Đại học Mở - Địa chất

Bài báo trình bày tổng quan về ứng dụng học máy dự báo nguy cơ xảy ra các thiên tai như trượt lở đất, cháy rừng và lũ quét, nhằm giảm thiểu thiệt hại do thiên tai gây ra. Dự báo thiên tai là một thách thức phức tạp, yêu cầu xử lý và phân tích một lượng lớn dữ liệu đa nguồn và đa chiều. Các nghiên cứu gần đây cho thấy sự vượt trội của các mô hình học máy so với các phương pháp truyền thống. Mặc dù còn nhiều thách thức, nhưng với sự phát triển của các phương pháp học máy và thu thập dữ liệu kỹ thuật số, chúng ta có thể xây dựng các mô hình dự báo nguy cơ xảy ra thiên tai ngày càng chính xác hơn, đóng góp vào công tác phòng chống thiên tai.

Giới thiệu

Dự báo nguy cơ xảy ra thiên tai là một công việc quan trọng để phòng tránh và giảm thiểu thiệt hại do thiên tai gây ra. Học máy là kỹ thuật giúp máy tính học từ dữ liệu và có thể áp dụng để phân tích mối quan hệ phức tạp giữa các yếu tố ảnh hưởng với thiên tai. Nhiều nghiên cứu về ứng dụng học máy trong dự báo nguy cơ xảy ra thiên tai đã đạt được những kết quả đáng chú ý. Các mô hình học máy đã giúp nâng cao độ chính xác và tin cậy của việc dự báo ở nhiều nơi trên thế giới. Tuy nhiên, việc này cũng đặt ra nhiều thách thức, bao gồm việc lựa chọn các yếu tố ảnh hưởng phù hợp cho khu vực nghiên cứu, chất lượng của dữ liệu thu thập, chất lượng của mô hình học máy và việc ứng dụng trong thực tế khi điều kiện ở khu vực nghiên cứu thay đổi.

Một số ứng dụng học máy trong dự báo nguy cơ xảy ra thiên tai

Trên thế giới, học máy đã và đang được sử dụng rộng rãi trong dự báo nguy cơ xảy ra nhiều loại thiên tai khác nhau, trong đó có trượt lở đất, cháy rừng và lũ quét. Các nghiên cứu đã cho thấy sự vượt trội của các mô hình này so với các phương pháp truyền thống.

Học máy đã được ứng dụng nhiều trong dự báo nguy cơ xảy ra trượt lở đất [1-3]. Để xây dựng các mô hình dự báo này, dữ liệu huấn luyện chứa các thông tin về địa hình, địa chất, đất đai, thực vật, thời tiết và hoạt động của con người tại khu vực nghiên cứu được sử dụng. Những thông tin này bao gồm độ dốc, độ cao, độ cong bề mặt, sông suối, đường giao thông, đứt gãy, thạch học, thổ nhưỡng, loại

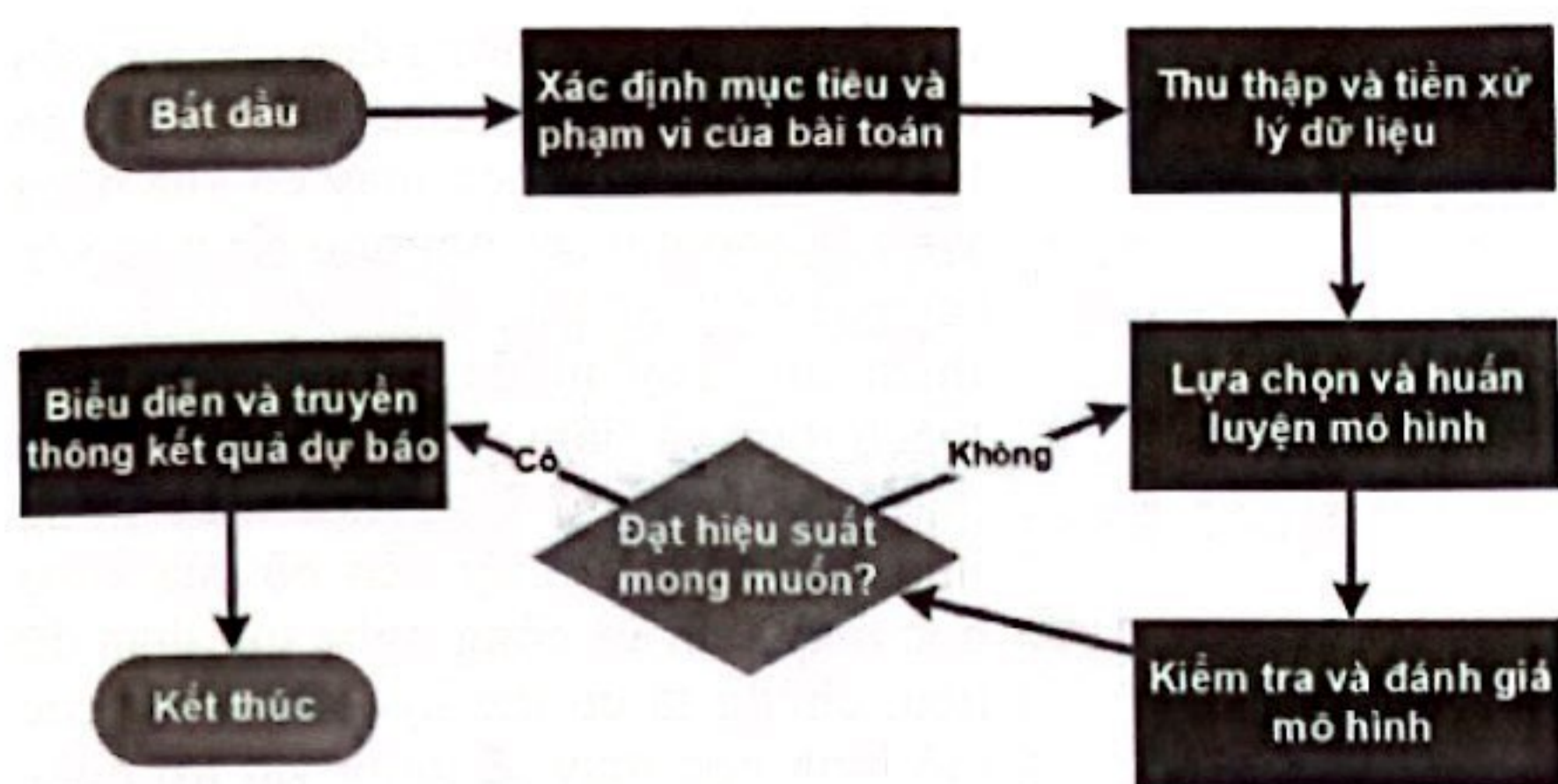
cây trồng, sử dụng đất, lượng mưa, độ ẩm, nhiệt độ và các hoạt động khai thác mỏ, xây dựng đường,... tại các vị trí đã xảy ra trượt lở đất trong quá khứ.

Một ứng dụng thông dụng khác của học máy là dự báo nguy cơ cháy rừng [4-6]. Việc phát triển các mô hình học máy có thể dự đoán nguy cơ cháy rừng dựa trên dữ liệu huấn luyện chứa các thông tin về thời tiết, đặc điểm địa lý, thực vật và hoạt động con người trong khu vực nghiên cứu. Các yếu tố quan trọng bao gồm nhiệt độ, độ ẩm, tốc độ gió, lượng mưa, độ cao, độ dốc, các chỉ số thực vật (NDVI, NDWI,...), các hoạt động xây dựng công trình, đường giao thông và khai thác tài nguyên. Tất cả các thông tin này được thu thập từ các vị trí đã xảy ra cháy rừng trong quá khứ để xây dựng các mô hình dự báo nguy cơ cháy rừng.

Học máy cũng có thể được sử dụng để dự đoán khả năng xảy ra lũ quét [7,8]. Để xây dựng các mô hình này, thông tin về đặc điểm địa lý, khí tượng thủy văn, thực vật và hoạt động của con người tại khu vực nghiên cứu được sử dụng trong dữ liệu huấn luyện. Những thông tin này bao gồm độ cao, độ dốc, hình thái mặt đất, lượng mưa, dòng chảy của sông suối, loại cây trồng, sử dụng đất, các hoạt động khai thác tài nguyên, xây dựng công trình, làm đường,... tại các vị trí đã xảy ra lũ quét trong quá khứ.

Các bước xây dựng mô hình học máy trong dự báo nguy cơ xảy ra thiên tai

Bài toán dự báo nguy cơ xảy ra thiên tai thường được mô hình hóa như một bài toán phân hai lớp cho các điểm trên bản đồ raster của khu vực nghiên cứu: âm (không xảy ra) và dương (đã xảy ra).



Hình 1. Các bước ứng thực hiện

Nhìn chung, bài toán ứng dụng học máy trong dự báo nguy cơ xảy ra thiên tai bao gồm các bước như (hình 1):

Bước 1: Xác định mục tiêu và phạm vi của dự báo. Đây là bước quan trọng để xác định loại thiên tai cần dự báo, khu vực nghiên cứu, thời gian và độ chính xác mong muốn của dự báo.

Bước 2: Thu thập và tiền xử lý dữ liệu. Đây là bước cần thiết để có được các dữ liệu cho mô hình dự báo, bao gồm dữ liệu về thiên tai đã xảy ra trong quá khứ và các yếu tố ảnh hưởng đến thiên tai. Dữ liệu có thể được thu thập từ nhiều nguồn khác nhau như ảnh vệ tinh, cảm biến, khảo sát, báo cáo, thống kê,... Dữ liệu cần được tiền xử lý để loại bỏ nhiễu, thiếu sót, trùng lặp và định dạng lại cho phù hợp với mô hình dự báo. Đối với dữ liệu về thiên tai đã xảy ra, cần phân chia dữ liệu thành hai tập: tập huấn luyện (chiếm đa số) và tập kiểm tra. Tập huấn luyện sẽ được sử dụng để huấn luyện mô hình, còn tập kiểm tra sẽ được sử dụng để đánh giá hiệu năng của mô hình. Tập huấn luyện và tập kiểm tra còn được bổ sung một số lượng tương ứng dữ liệu của các điểm không xảy ra thiên tai.

Bước 3: Lựa chọn và huấn luyện mô hình dự báo. Đây là bước áp dụng các thuật toán học máy để tìm ra mối quan hệ giữa các dữ liệu đầu vào và đầu ra. Có nhiều loại mô hình dự báo khác nhau có thể sử dụng trong dự báo nguy cơ của các thiên tai nói trên. Các thuật toán thường được sử dụng là các thuật toán phân lớp (classification) bao gồm: hồi quy logistic (Logistic Regression); cây quyết định (Decision Tree); các mô hình học tập thể (Ensemble Learning) dựa trên cây quyết định như Random Forest và Gradient Boosting; Support Vector Machines; mạng nơ ron nhân tạo (Neural Networks) và học sâu (Deep Learning);... Cũng có thể sử dụng các thuật toán tối ưu để tối ưu các siêu tham số (hyperparameters) của mô hình học máy để đạt được kết quả tốt hơn.

Bước 4: Kiểm tra và đánh giá mô hình dự báo. Đây là bước kiểm tra hiệu suất của mô hình dự báo

trên tập dữ liệu kiểm tra chưa được sử dụng trong quá trình huấn luyện. Các phương pháp kiểm tra và đánh giá có thể bao gồm kiểm tra chéo k-lần (k-fold cross-validation), giữ lại (hold-out), kiểm tra nhóm (bootstrap),... Các chỉ số đánh giá có thể bao gồm độ chính xác (accuracy), diện tích dưới đường cong ROC (AUC),...

Bước 5: Biểu diễn và truyền thông kết quả dự báo. Đây là bước sử dụng mô hình có hiệu suất như mong muốn để thành lập bản đồ phân vùng nguy cơ thiên tai cho khu vực nghiên cứu. Áp

dụng mô hình học máy cho từng điểm trên bản đồ của khu vực nghiên cứu sử dụng dữ liệu của các yếu tố tại điểm đó làm đầu vào, đầu ra của mô hình là giá trị chỉ số nguy cơ xảy ra thiên tai của điểm. Sau đó, sử dụng một phương pháp phân loại các giá trị nguy cơ của toàn bộ khu vực nghiên cứu thành nhiều cấp (ví dụ: rất cao, cao, trung bình, thấp và rất thấp) và biểu diễn bằng các màu sắc khác nhau. Bản đồ phân vùng nguy cơ này cần được truyền thông đến các bên liên quan để có biện pháp phòng ngừa và giảm thiểu thiệt hại do thiên tai gây ra.

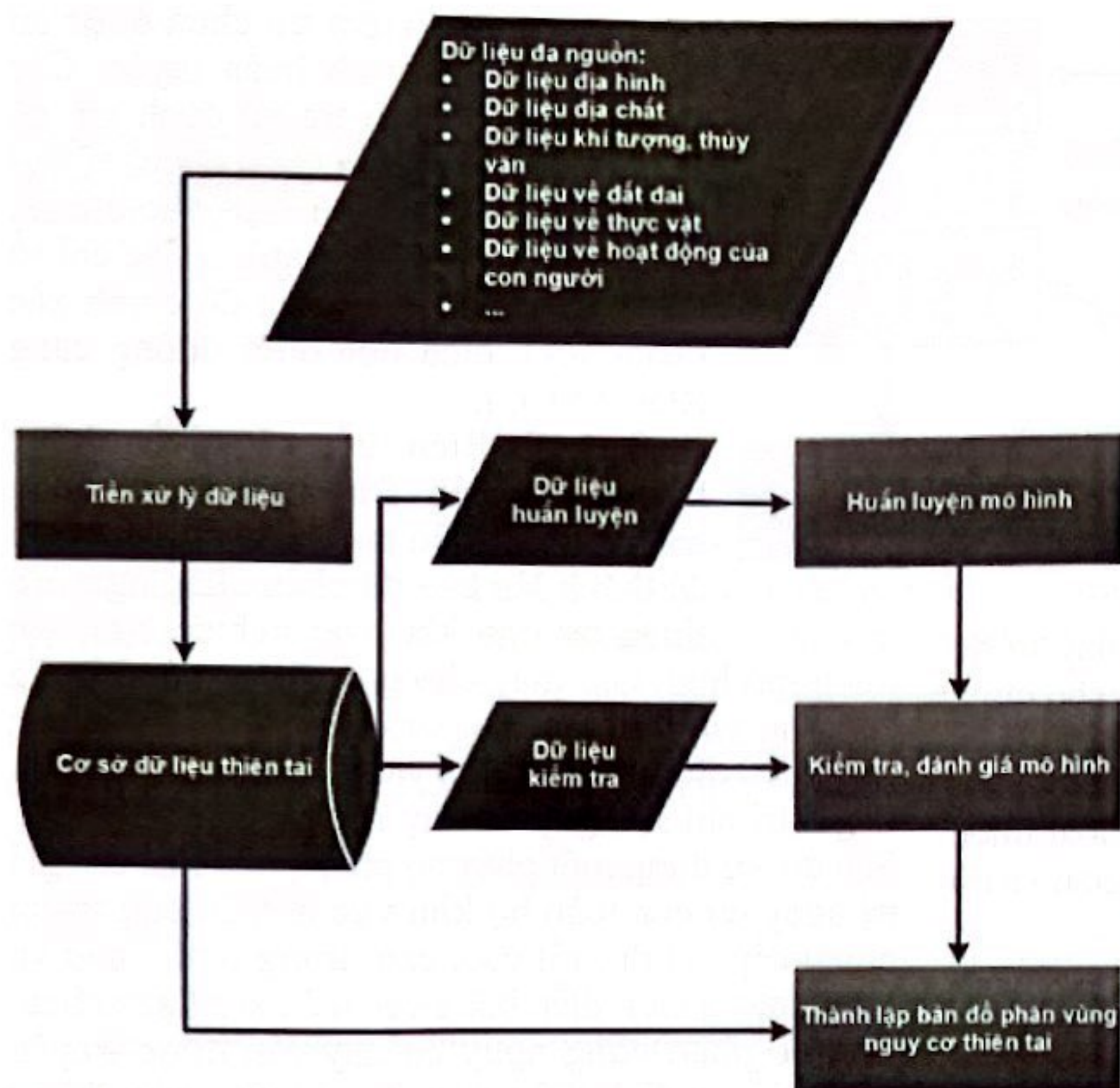
Cụ thể hơn, quy trình kỹ thuật xây dựng mô hình học máy và thành lập bản đồ phân vùng nguy cơ thiên tai như (hình 2):

Các thách thức trong việc xây dựng và sử dụng các mô hình học máy trong dự báo nguy cơ xảy ra thiên tai

Một số thách thức trong việc xây dựng và sử dụng các mô hình học máy dự báo nguy cơ xảy ra thiên tai như sau:

Về dữ liệu của các yếu tố ảnh hưởng: Đây là thách thức liên quan đến việc thu thập, tiền xử lý, lựa chọn và phân tích các dữ liệu liên quan đến thiên tai. Các yếu tố gây ra thiên tai là rất đa dạng, từ địa hình, địa chất, thời tiết đến sự tác động của con người và cần được lựa chọn phù hợp với khu vực nghiên cứu. Các dữ liệu này có thể bị thiếu sót, không chính xác hoặc không đồng nhất. Các dữ liệu cần được tiền xử lý để loại bỏ nhiễu, thiếu sót, trùng lặp và định dạng lại cho phù hợp với mô hình học máy.

Về mô hình: Đây là thách thức liên quan đến việc lựa chọn, huấn luyện, kiểm tra và đánh giá các mô hình học máy. Có nhiều loại mô hình học máy khác nhau có thể được sử dụng cho dự báo nguy cơ xảy ra thiên tai nhưng không có mô hình nào là tối ưu cho mọi trường hợp. Vì vậy, cần phải xây dựng nhiều loại mô hình khác nhau và lựa chọn mô hình tốt nhất. Các mô hình học máy có thể bị ảnh hưởng



Hình 2. Quy trình kỹ thuật

bởi các vấn đề như quá khớp (overfitting), thiếu ổn định (instability) hoặc thiếu khả năng giải thích (interpretability).

Về triển khai và theo dõi trong tương lai: Đây là thách thức liên quan đến việc áp dụng và duy trì các mô hình học máy dự báo nguy cơ xảy ra thiên tai cho khu vực nghiên cứu trong tương lai khi dữ liệu về các yếu tố ảnh hưởng có thể thay đổi. Các mô hình học máy có thể không còn chính xác khi gặp các điều kiện mới hoặc khác biệt so với điều kiện huấn luyện. Ví dụ, các yếu tố gây ra thiên tai có thể thay đổi theo mùa, theo năm hoặc do các sự kiện khí hậu bất thường. Các mô hình học máy cần được cập nhật và điều chỉnh theo thời gian để phản ánh sự thay đổi của các yếu tố gây thiên tai. Ví dụ, các mô hình học máy có thể được huấn luyện lại trên các dữ liệu mới hoặc sử dụng các kỹ thuật học tăng cường (reinforcement learning) để tự động cải thiện hiệu suất của mình. Các kết quả dự báo cũng cần được kiểm tra và đánh giá liên tục để đảm bảo tính chính xác và tin cậy.

Kết luận

Dự báo nguy cơ xảy ra các thiên tai như trượt lở đất, cháy rừng, lũ quét là một nhiệm vụ quan trọng và cấp thiết để giảm thiểu thiệt hại về người và tài sản, cũng như bảo vệ môi trường và đa dạng sinh học. Các mô hình học máy đã được áp dụng rộng rãi và hiệu quả trong dự báo thiên tai, sử dụng các dữ liệu đa nguồn và đa chiều liên quan đến các yếu

tổ gây ra và ảnh hưởng đến các sự kiện thiên tai. Các nghiên cứu gần đây đã cho thấy các mô hình học máy có khả năng vượt trội so với các phương pháp truyền thống trong dự báo nguy cơ xảy ra các thiên tai. Tuy nhiên, cũng còn nhiều thách thức và tiềm năng để cải thiện và phát triển các mô hình học máy trong lĩnh vực này. Với sự tiến bộ của khoa học máy tính và công nghệ thu thập dữ liệu, chúng ta có thể kỳ vọng rằng các mô hình học máy sẽ đóng vai trò ngày càng quan trọng trong dự báo và ứng phó với các thiên tai trong tương lai.

Tài liệu tham khảo

1. Merghadi, A. et al. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207: 103225, 2020.
2. Chang, Z. et al. Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. *Journal of Rock Mechanics and Geotechnical Engineering*, 15(5): 1127-1143, 2023.
3. Hoang, D.A. et al. Hybrid BBO-DE Optimized SPAARCTree Ensemble for Landslide Susceptibility Mapping. *Remote Sensing*, 15(8):2187, 2023.
4. Sayad, Y.O. et al. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, 104: 130-146, 2019.
5. Le, H.V. et al. A Hybrid Intelligence System Based on Relevance Vector Machines and Imperialist Competitive Optimization for Modelling Forest Fire Danger Using GIS. *Journal of Environmental Informatics*, 36: 43-57, 2020.
6. Le, H.V. et al. A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. *Ecological Informatics*, 63: 101300, 2021.
7. Alipour, A. et al. Leveraging machine learning for predicting flash flood damage in the Southeast US. *Environmental Research Letters*, 15(2): 024011, 2020.
8. El-Magd S.A.A., et al. Hybrid-based Bayesian algorithm and hydrologic indices for flash flood vulnerability assessment in coastal regions: machine learning, risk prediction, and environmental impact. *Environmental Science and Pollution Research*, 29: 57345-57356, 2022.