

Bragana, chứa một số quan sát thời tiết (ví dụ: Tốc độ gió) được ghi lại trong khoảng thời gian 30 phút bởi một trạm khí tượng nằm ở trung tâm của công viên Montesinho. Hai cơ sở dữ liệu được lưu trữ trong hàng chục bảng tính riêng lẻ, dưới các định dạng riêng biệt và tích hợp một cách thủ công vào một tập dữ liệu duy nhất với tổng số 517 mục. Dữ liệu có thể được tải từ đường link sau: <http://www.dsi.uminho.pt/pcortez/forestfires/>.

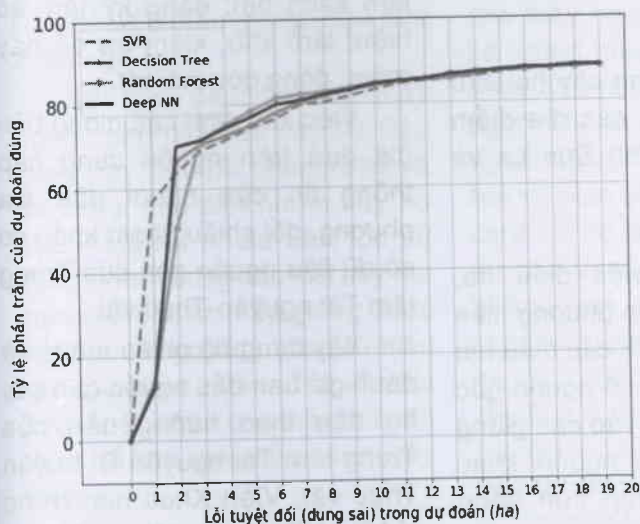
### Phương pháp sử dụng

Trong nghiên cứu này, các thuật toán khác nhau đã được sử dụng, các thuật toán bao gồm: Vector hồi quy hỗ trợ - 'SVR', Cây quyết định - 'Decision Tree', Rừng ngẫu nhiên - 'Random Forest', và Mạng nơ-ron sâu - 'Deep NN'. Sau đó, kết quả dự đoán nguy cơ cháy rừng với cùng bộ dữ liệu được đem so sánh để thấy được thuật toán nào có độ chính xác cao nhất. Các bước so sánh bao gồm: Tìm diện tích thực sự đã bị cháy; Dự đoán lỗi (số ha cháy); Tìm sai số toàn phương trung bình; Tìm lỗi tuyệt đối (dung sai) trong dự đoán; Tỷ lệ dự đoán đúng.

### Một số kết quả

**Hình 3:** Kết quả cuối cùng khi so sánh độ chính xác dự đoán của 4 thuật toán

Đường đặc tính lỗi cho các mô hình khác nhau



Hình 3 thể hiện kết quả cuối cùng cho thấy quan hệ giữa dung sai trong dự đoán, tỷ lệ phần trăm dự đoán đúng của 4 thuật toán. Khi nhận kết quả tính toán Sai số toàn phương trung bình (RMSE), SVR cho kết quả xấu nhất với RMSE là 0.671031657615, Cây quyết định là 0.624177314982, Random Forest là 0.614885707514 và Deep NN là 0.610289206294.

### Kết luận

Với bộ dữ liệu [5], thực nghiệm cho thấy thuật toán Deep NN có sai số thấp nhất so với các thuật toán khác, hứa hẹn sẽ cho kết quả tốt hơn khi áp dụng ở các bài toán tương tự khác, sử dụng học sâu (Deep learning). Ngoài ra, chất lượng của mô hình cháy rừng có thể được cải thiện khi các thuật toán tối ưu hóa mới khác được chọn để tìm kiếm các tham số tốt nhất cho mô hình Deep NN. Những cải tiến khác có thể có của nghiên cứu này bao gồm điều tra các mô hình máy học mới để cải thiện khả năng dự đoán của mô hình dự báo cháy rừng. Kết luận cuối cùng, kết quả trong nghiên cứu này rất hữu ích cho việc lập kế hoạch và quản lý rừng ở các khu vực dễ bị cháy rừng.

### Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Mở - Địa chất trong khuôn khổ đề tài NCKH số T19-01 của ThS. Hoàng Anh Đức chủ nhiệm: "Nghiên cứu ứng dụng phương pháp học sâu trong đánh giá nguy cơ cháy rừng, áp dụng thí điểm cho một tỉnh miền núi Việt Nam"

### TÀI LIỆU THAM KHẢO

- [1] Eastaugh, C.S., Hasenauer, H., 2014. Deriving forest fire ignition risk with biogeochemical process modelling. Environ. Model. Softw. 55, 132–142.
- [2] Ganteaume, A. et al., 2013. A review of the main driving factors of forest fire ignition over Europe. Environ. Manage., 51(3), p. 651–662.
- [3] Hùng, P. N., 2004. Quản lý cháy rừng ở Việt Nam. s.l.:Nhà xuất bản Nghệ An.
- [4] McGrattan, K. et al., 2013. Fire Dynamics Simulator Technical Reference Guide, Volume 1: Mathematical Model. s.l.:National Institute of Standards and Technology, U.S. Department of Commerce.
- [5] P. Cortez and A. Morais. "A Data Mining Approach to Predict Forest Fires using Meteorological Data", In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal, pp. 512-523, 2007. APPIA, ISBN-13 978-989-95618-0-9.
- [6] Verde, J., Zêzere, J., 2010. Assessment and validation of wildfire susceptibility and hazard in Portugal. Nat. Hazards Earth Syst. Sci. 10 (3), 485–497. ■