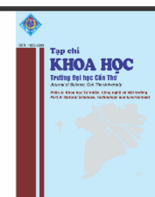




Tạp chí Khoa học Trường Đại học Cần Thơ
Phần A: Khoa học Tự nhiên, Công nghệ và Môi trường

website: sj.ctu.edu.vn



DOI:10.22144/ctu.jsi.2017.022

ỨNG DỤNG GIẢI THUẬT GỢI Ý DỰA TRÊN NỘI DUNG HỖ TRỢ NÔNG DÂN PHÒNG TRỪ BỆNH ĐẠO ÔN

Trần Nguyễn Minh Thư, Nguyễn Thị Thanh Lan và Nguyễn Hoàng Mẫn

Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận bài: 15/09/2017

Ngày nhận bài sửa: 10/10/2017

Ngày duyệt đăng: 20/10/2017

Title:

Content-based
recommendation system to
support farmers in blast
prevention

Từ khóa:

Bệnh đạo ôn, Bayes thơ ngây,
cây quyết định, hệ thống gợi ý
dựa trên nội dung

Keywords:

Blast disease, content-based
recommendation system, naïve
bayes, decision tree

ABSTRACT

Blast disease is caused by the *Pyricularia oryzae* fungus and has been recorded in more than 80 rice-producing countries around the world, and the disease is progressively more complex, causing many difficulties for farmers. From previous studies on rice blast disease, six important factors including rice variety, density, temperature, moisture, leaf color (protein), and lesion status have been found to have significant influence on the pathogenesis of diseases. Today, with the rapid development of internet networks, mobile devices, etc., most of the farmers own mobile phones. In this study, the content-based of recommender method is used to build the mobile application "BLASTREC" that supports farmers in blast prevention. The software BLASTREC functions Android operating system based on two Naive Bayes and Decision Tree classification algorithms. Experimental results show that the accuracy of two algorithms is more than 90%. The experiment data on blast in Trung An area, Thot Not district, Can Tho city combines with agricultural experts' opinion to provide farmers with appropriate treatment.

TÓM TẮT

Bệnh đạo ôn do nấm *Pyricularia oryzae* gây ra và đã được ghi nhận ở 80 quốc gia có trồng lúa trên thế giới và bệnh đang có sự chuyển biến ngày càng phức tạp hơn, gây ra nhiều khó khăn cho người nông dân. Từ những nghiên cứu về bệnh đạo ôn trước đây, người ta đã phát hiện được 6 yếu tố quan trọng ảnh hưởng đến việc phòng trị bệnh: giống lúa, mật độ sạ, nhiệt độ, độ ẩm, màu lá (độ đạm), tình trạng vết bệnh. Ngày nay, với sự phát triển nhanh chóng của mạng lưới internet, các thiết bị di động,... đa phần người nông dân đều sở hữu cho điện thoại di động. Trong nghiên cứu này, để hỗ trợ cho người nông dân chọn lựa phương pháp phòng trừ bệnh đạo ôn, phương pháp gợi ý dựa trên nội dung được sử dụng để xây dựng ứng dụng trên điện thoại di động "BLASTRec". Ứng dụng BLASTRec hoạt động trên hệ điều hành android dựa trên hai giải thuật phân lớp Naive bayes và cây quyết định (Decision Tree). Kết quả thực nghiệm cho thấy việc phân lớp của hai giải thuật có độ chính là trên 90%. Thực nghiệm được thực hiện dựa trên tập dữ liệu thu thập về bệnh đạo ôn tại địa bàn Trung An, quận Thốt Nốt, thành phố Cần Thơ kết hợp với ý kiến chuyên gia nông nghiệp để đưa ra cho người nông dân các giải pháp điều trị phù hợp.

Trích dẫn: Trần Nguyễn Minh Thư, Nguyễn Thị Thanh Lan và Nguyễn Hoàng Mẫn, 2017. Ứng dụng giải thuật gợi ý dựa trên nội dung hỗ trợ nông dân phòng trừ bệnh đạo ôn. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Công nghệ thông tin: 164-170.

1 GIỚI THIỆU

Bệnh đạo ôn đã xuất hiện ở tất cả các quốc gia trồng lúa như: Trung Quốc, Nhật Bản, Philippines, Ấn Độ, Thái Lan, Việt Nam (Nguyễn Văn Viên và *ctv.*, 2013)... Bệnh xảy ra ở khoảng 85 quốc gia trên thế giới. Sự xuất hiện đầu tiên được biết đến của nó là vào năm 1637 ở Trung Quốc (G-L.Wang and B. Valent, 2009). Bệnh đạo ôn có tính hủy diệt cao đối với lúa gạo ở vùng đất ôn đới và á nhiệt đới Châu Á, và lúa nương ở vùng Á Châu nhiệt đới, Mỹ La Tinh và Châu Phi. Bệnh đạo ôn có mặt ở các vùng trồng lúa khắp Việt Nam và gây nhiều thiệt hại cũng như tổn kém cho người nông dân, đây là một trong những bệnh nghiêm trọng ở Đồng bằng sông Cửu Long. Nếu bà con nông dân không phát hiện sớm và phòng trị kịp thời thì dẫn đến bệnh nặng làm mất trắng năng suất. Theo ước tính của FAO, thiệt hại do bệnh đạo ôn gây ra làm giảm năng suất lúa trung bình từ 0,7 – 17,5%, những nơi thiệt hại nặng có thể làm giảm đến 80% (Nguyễn Thị Phong Lan và *ctv.*, 2012; Bonman *et al.*, 1986). Theo nghiên cứu của tác giả Alvin R. Malicdem *et al.* (2015), thiệt hại từ 5-10%, 8% và 14% được ghi nhận lần lượt ở Ấn Độ (1960-1961), Hàn Quốc (giữa thập niên 1970), và ở Trung Quốc (1980-1981). Đặc biệt, tại Philippines thiệt hại năng suất dao động từ 50-85% đã được báo cáo (Alvin R. Malicdem *et al.*, 2015). Tiến trình của bệnh thay đổi ở các vị trí và thời điểm khác nhau, phụ thuộc chủ yếu vào điều kiện thời tiết. Do đó, dự báo dịch bệnh là cần thiết nếu người trồng phải ngăn ngừa tổn thất năng suất nghiêm trọng do bệnh. Các nghiên cứu trên thế giới hiện tại tập trung vào việc dự đoán sự bùng phát của bệnh đạo ôn thông qua việc sử dụng mô hình máy học vector hỗ trợ (SVM), mạng nơ ron nhân tạo (ANN),... (Alvin R. Malicdem *et al.*, 2015).

Tại Việt Nam, các phương pháp phòng trừ bệnh đạo ôn hiện nay được triển khai đến người nông dân thông qua các chi cục bảo vệ thực vật, các trang web của Sở Nông nghiệp hay các Chi Cục Bảo vệ Thực vật, các chương trình khuyến nông. Tuy nhiên, thực tế thì các chương trình khuyến nông cũng chỉ có thể hướng dẫn người nông dân một vài trường hợp cụ thể. Người nông dân chủ yếu chọn theo sự giới thiệu của người bán thuốc bảo vệ thực vật mà không theo đúng tình trạng bệnh thực tế gây tổn nhiều tiền bạc và công sức.

Vì vậy, việc áp dụng công nghệ nói chung và công nghệ thông tin nói riêng nhằm giảm bớt gánh nặng cho bà con nông dân trong sản xuất lúa và giúp họ lựa chọn những giải pháp phù hợp cho việc phòng trừ bệnh đạo ôn là một việc làm rất cần thiết. Hiện nay, điện thoại thông minh được bán rộng rãi trên thị trường với giá khá rẻ. Mỗi người

nông dân đa phần đều trang bị một chiếc điện thoại. Với sự tiện lợi đó cùng với sự cần thiết của việc giúp người nông dân phòng trừ bệnh đạo ôn hiệu quả, hệ thống gợi ý phòng trừ bệnh Đạo ôn trên thiết bị di động- BLASTRec được đề xuất.

Theo Adomavicius và Tuzhilin (2005), hệ thống gợi ý có ứng dụng rất đa dạng trong thực tế, giúp người dùng giải quyết vấn đề quá tải thông tin và cung cấp cho họ những gợi ý được cá nhân hóa về các sản phẩm và dịch vụ. Do đó, việc ứng dụng hệ thống gợi ý trong các hệ thống thông tin, đặc biệt là hệ thống thông tin với cơ sở dữ liệu lớn, mang lại nhiều lợi ích cho cả người dùng lẫn hệ thống. Theo Balabanovic và Shoham (1997), hệ thống gợi ý thường được chia ra thành ba loại dựa trên cách mà các gợi ý được tạo ra: (1) gợi ý dựa trên nội dung: người dùng được gợi ý những mục dữ liệu tương tự với những mục dữ liệu mà chính họ thích trong quá khứ; (2) gợi ý dựa trên lọc cộng tác: người dùng được gợi ý những mục dữ liệu mà những người dùng khác (có cùng thị hiếu và sở thích) thích trong quá khứ; (3) gợi ý hỗn hợp: kết hợp hai phương pháp trên. Ứng dụng BLASTRec được xây dựng dựa trên phương pháp gợi ý theo nội dung.

Bài báo này bao gồm 5 phần. Phần một, sự cần thiết của bệnh đạo ôn và hệ thống gợi ý dựa trên nội dung được giới thiệu. Phần hai, các ứng dụng công nghệ thông tin đối với bệnh đạo ôn được phân tích và đánh giá. Các phương pháp xây dựng hệ thống gợi ý dựa trên nội dung cho ứng dụng BLASTRec được giới thiệu chi tiết. Phần bốn, ứng dụng BLASTRec được trình bày và phần cuối chúng tôi sẽ trình bày kết luận của nghiên cứu.

2 ỨNG DỤNG CÔNG NGHỆ THÔNG TIN ĐỐI VỚI BỆNH ĐẠO ÔN LÚA

Các nhà nghiên cứu trên thế giới ở miền Bắc Philippines hiện tại tập trung vào việc dự đoán sự bùng nổ của bệnh đạo ôn thông qua việc sử dụng mô hình máy học vecto hỗ trợ (SVM) và mạng nơ ron nhân tạo (ANN) với độ chính xác trên 90%. Mô hình này sử dụng 4 thuộc tính quan trọng là lượng mưa, nhiệt độ tối thiểu, nhiệt độ tối đa và độ ẩm. Kết quả dự đoán mô hình SVM cung cấp dự đoán chính xác hơn ANN với MSE trung bình của 0,2374 và R^2 (hệ số tương quan) của 0,7758 nào tốt hơn nhiều so với mô hình ANN (Alvin R. Malicdem *et al.*, 2015). Nghiên cứu này chỉ là mô hình dự báo chưa có phương pháp hỗ trợ phòng trừ bệnh đạo ôn.

Ngoài ra, theo nghiên cứu về mô hình máy học như mô hình đa hồi quy (REG), máy học vector (SVM), mạng nơ ron lan truyền ngược (BPNN), mạng lưới nơ ron hồi quy tổng quát (GRNN) ở một

số trường hợp nghiên cứu dự đoán về bệnh đạo ôn. Với 6 thuộc tính quan trọng như: nhiệt độ (Min – Max), độ ẩm tương đối (Min – Max), lượng mưa và ngày mưa mỗi tuần được ghi lại hàng ngày. Kết quả dự đoán mô hình REG, BPNN, GRNN: % MAE tương ứng là 77,54; 66,11; 58,26 và hệ số trung bình tương quan $r = 0,48; 0,56; 0,66$. Phương pháp dựa vào SVM thực hiện tốt hơn cả ba cách tiếp cận trên khi sử dụng chỉ số % MAE đến 44,12 và $r = 0,74$ (Rakesh Kaundal *et al.*, 2007). Tuy nhiên, không có công cụ dự báo bệnh trực tuyến cũng như không có tiềm năng hướng dẫn nông dân và các nhà nghiên cứu xây dựng các chiến lược khác nhau để kiểm soát bệnh thực vật một cách hiệu quả bằng cách dự báo kịp thời sự xuất hiện của bệnh. Nghiên cứu này chỉ là mô hình dự báo chưa có phương pháp hỗ trợ phòng trừ bệnh đạo ôn.

Hiện nay, có những trang web hỗ trợ phòng trừ bệnh trên lúa có thể kể đến như: Tin tức nông nghiệp¹, Cục Bảo vệ Thực vật², hỗ trợ nông nghiệp³,.... Những trang web này cung cấp những tin tức thông tin về bệnh dịch hại hỗ trợ bà con phòng và điều trị dịch bệnh một cách chung chung chưa có phương pháp cụ thể cho bà con nông dân.

Về mặt nghiên cứu hiện nay đã một số đề tài nghiên cứu ứng dụng công nghệ thông tin vào nông nghiệp như đề tài: Hệ thống so màu lá lúa trên thiết bị di động, sử dụng giải thuật K láng giềng cho độ chính xác tương ứng là 92% (Huỳnh Xuân Hiệp và *ctv.*, 2014); Tương quan giữa các yếu tố ảnh hưởng đến bệnh đạo ôn lúa nhằm phục vụ công tác dự báo dịch hại bằng hệ thống thông tin địa lý (GIS) tại huyện Châu Thành, tỉnh An Giang (Đỗ Thanh Nghị và Lê Thanh Vân, 2012); Giải pháp nhận dạng bệnh trên lúa từ thiết bị di động thông minh (Nguyễn Hữu Hòa và *ctv.*, 2016). Tuy nhiên, chưa có đề tài nào hỗ trợ người dân phòng trừ bệnh đạo ôn trên lúa. Do đó, cần xây dựng một hệ thống gợi ý dựa trên nội dung nhằm giúp đỡ người nông dân tìm được phương pháp phù hợp với bệnh hại lúa.

3 XÂY DỰNG HỆ THỐNG GỢI Ý DỰA TRÊN NỘI DUNG CHO ỨNG DỤNG BLASTREC

3.1 Giới thiệu về bệnh đạo ôn

Bệnh đạo ôn là một trong những bệnh gây hại nhiều nhất trên cây lúa. Khi dịch cháy lá xảy ra trên diện rộng thì sự thiệt hại đến năng suất và sản lượng thể hiện rõ rệt và thiệt hại nghiêm trọng đến kinh tế. Bệnh có thể gây hại trên cổ lá nên gọi là

thối cổ lá, hoặc gây hại trên cổ bông nên được gọi là thối cổ bông làm lép hạt; đôi khi bệnh có thể gây lem vỏ hạt lúa (Trần Thị Thu Thủy và *ctv.*, 2005; Bùi Cách Tuyến và Lê Cao Lượng, 2013; Prem Bahadur Magar *et al.*, 2015).



Hình 1: Dấu hiệu của bệnh đạo ôn trên các bộ phận của cây lúa

Trên mạ: vết bệnh trên lá mạ lúc đầu hình bầu dục nhỏ sau thành hình thoi nhỏ hoặc dạng tương tự hình thoi (hình mắt én), có màu nâu hoặc màu nâu vàng. Khi bệnh nặng, từng đám bệnh kế tiếp nhau làm cho cây mạ khô héo hoặc chết. Trên lá: đặc điểm của vết bệnh có thể thay đổi theo tuổi cây, điều kiện thời tiết và tính nhiễm của giống:

Vết bệnh ban đầu chỉ là vết chấm kim nhỏ màu nâu hay chỉ là những đốm úng nước nhỏ, màu xám xanh. Vết bệnh sau đó lan ra, tạo thành vết bệnh to, hình thoi (mắt én), màu nâu nhạt, có khi xung quanh xuất hiện quầng màu vàng nâu, phần giữa vết bệnh có màu tro xám. Kích thước vết bệnh (1-1,5 cm x 0,3-0,5 cm). Trong điều kiện ẩm ướt, được bón đạm nhiều, các giống nhiễm sẽ xuất hiện vết bệnh cấp tính hình tròn hay hình bầu dục, nâu xanh tái do đài và bào tử nấm phát triển mạnh trên đó, dạng thấm nước về sau cũng chuyển thành dạng mãn tính điển hình.

Vết bệnh là những đốm nâu nhỏ có kích thước bằng đầu kim đến 1-2 cm. Ở các giống kháng vừa, vết bệnh có hình tròn hay trứng, tâm sáng trắng viền nâu, kích thước khoảng 2-3 mm. Tùy thuộc vào mức độ kháng bệnh của từng giống lúa mà vết bệnh có các kích thước khác nhau. Nhiễm nặng và sớm lúa có thể bị lùn, nhiều vết bệnh trên lá liên kết lại gây cháy lá. Khi bị nhiễm bệnh, các bộ phận trên cây lúa như đốt thân, cổ gie, nhánh gie sẽ có màu nâu sậm đến đen. Khi gặp khí hậu ẩm ướt, vết bệnh ướt và có mốc xám xanh. Khi trời khô, vết bệnh bị nhăn lại. Bệnh làm gãy thân, gãy gie, lép hạt hay giảm trọng lượng hạt.



Hình 2: Minh họa triệu chứng của bệnh đạo ôn trên cây lúa

¹ <http://www.tintucnongnghiep.com>

² <http://ppd.gov.vn>

³ <http://hotronongnghiep.com/vn/>

– Bệnh xuất hiện trên hạt hình đốm tròn, viền nâu, tâm màu xám trắng, đường kính 1-2 mm (Huỳnh Nguyệt Ánh và *ctv.*, 2008).

Trong đó, thang đánh giá cấp bệnh đạo ôn theo nguồn IRRI- 2014 được mô tả như sau:

Bảng 1: Thang đánh giá cấp bệnh đạo ôn

Cấp bệnh	Mô tả vết bệnh	Đánh giá
0	Không có vết bệnh trên lá	Rất kháng
1	Có chấm màu nâu li ti như đầu kim	Kháng
2	Vết bệnh nhỏ, hơi kéo dài 1-2 mm, có viền nâu	
3	Vết bệnh giống như cấp 2, nhưng với số lượng đáng kể trên các lá	Hơi kháng
4	Vết bệnh điển hình, hình elip, dài 3 cm hoặc dài hơn, tác hại <4% diện tích lá	
5	Vết bệnh điển hình, tác hại 4-10% diện tích lá	Hơi nhiễm
6	Vết bệnh điển hình, tác hại 11-25% diện tích lá	Nhiễm
7	Vết bệnh điển hình, tác hại 26-50% diện tích lá	
8	Vết bệnh điển hình, tác hại 51-75% diện tích lá	Rất nhiễm

3.2 Giới thiệu các phương pháp phòng trừ bệnh

Hiện nay, biện pháp hóa học được sử dụng phổ biến trong việc phòng trừ bệnh đạo ôn vì cho hiệu quả nhanh, dễ sử dụng. Tuy nhiên, nếu sử dụng không đúng, không hợp lý, sử dụng lâu dài sẽ gây ảnh hưởng xấu tới môi trường, tăng khả năng hình thành tính kháng thuốc của sâu bệnh, tiêu diệt thiên địch, phá vỡ cân bằng sinh thái, gây nhiều dịch hại mới. Biện pháp phòng trừ dịch hại tổng hợp IPM (Integrated Pest Management) được xem là biện pháp chính trong kiểm soát dịch hại. Chọn giống tốt sẽ kháng bệnh khừ lẫn tạp hạt cỏ trước khi gieo. Dùng biện pháp sạ hàng, bón phân cân đối N-P-K, không bón thừa phân đạm, nên bón phân đạm theo nhu cầu cây lúa. Sau mùa thu hoạch, cày vùi rơm rạ để trả lại nguồn hữu cơ cho đất đồng thời diệt được mầm bệnh, hạn chế đốt rơm vì biện pháp này chỉ trả lại một số chất khoáng có trong tro, đất dần dần kém màu mỡ, mau suy kiệt. Vệ sinh đồng ruộng bằng cách thu gom, tiêu diệt lúa rầy, cỏ dại mọc ven bờ là nơi lưu tồn và lây lan mầm bệnh sau này. Giữ mực nước đầy đủ thường xuyên trên mặt ruộng tùy theo nhu cầu nước theo từng giai đoạn của cây lúa, tránh để ruộng khô nước khi bệnh cháy lá xảy ra. Nông dân cần thăm đồng thường xuyên, phát hiện kịp thời khi bệnh mới xuất hiện

(Nguyễn Văn Đĩnh, 2004). Vì vậy, để phòng trừ hiệu quả phải điều tra, phát hiện theo dõi phân tích các điều kiện liên quan tới sự phát sinh của bệnh như thời tiết, giống lúa,... Áp dụng biện pháp phòng trừ bệnh đạo ôn trên cây lúa phụ thuộc vào loại giống (giống đó có kháng bệnh – nhiễm bệnh), mật độ sạ (sạ dày: khi sạ $\geq 20\text{kg}/1000\text{m}^2$ hay sạ thưa), nhiệt độ môi trường (cao – thấp), độ ẩm không khí, màu lá (độ đậm) và tình trạng vết bệnh (Alvin R. Malicdem *et al.*, 2015; Trần Cao Đệ, 2016). Các trường hợp về bệnh đạo ôn để đưa ra phương pháp phòng trừ bệnh trên cây lúa bao gồm:

– Các giống lúa có tính nhiễm với mật độ sạ khác nhau như: sạ dày và sạ thưa. Trong quá trình trồng trọt thì có xuất hiện sự chênh lệch giữa các tình trạng như: nhiệt độ (cao-thấp), độ ẩm (cao-thấp), màu lá (độ đậm thừa-thiếu) và các tình trạng vết bệnh đều có ảnh hưởng đến việc đưa ra những phương pháp phù hợp phòng trị cho cây lúa khác nhau.

– Các giống kháng đạo ôn cũng có các thuộc tính tương tự giống nhiễm như: mật độ sạ, nhiệt độ, độ ẩm, màu lá và các tình trạng vết bệnh đều có tầm ảnh hưởng quan trọng đến việc đưa ra dự đoán chính xác cho phương pháp phòng trừ phù hợp.

3.3 Giới thiệu về tập dữ liệu

Dữ liệu thu thập tại địa bàn Trung An, quận Thốt Nốt, thành phố Cần Thơ. Tập dữ liệu ban đầu được thu thập với nhiều thuộc tính khác nhau như: kinh độ, vĩ độ, giống lúa, mật độ sạ, nhiệt độ, độ ẩm,... Sau khi phân tích dữ liệu, loại bỏ những thuộc tính dư thừa, thuộc tính nhiễu, thu thập ý kiến từ các chuyên gia, các giáo trình và bài báo khoa học có liên quan đến bệnh đạo ôn thì tổng hợp được 6 thuộc tính dự báo quan trọng:

- Giống lúa (giống nhiễm và giống kháng)
- Mật độ sạ (sạ dày và sạ thưa) (kg/ha)
- Nhiệt độ (cao hay thấp) (độ C)
- Độ ẩm (cao hay thấp) (%)
- Màu lá (độ đậm của lúa - thừa đạm hay không thừa đạm).
- Tình trạng vết bệnh

Dựa vào giá trị của các thuộc tính này, nghiên cứu được hỗ trợ trong việc gán nhãn các phương pháp phòng và trị bệnh đạo ôn trên tập dữ liệu bởi chuyên gia nông nghiệp. Dữ liệu có 4.735 dòng với các thuộc tính cần thiết để xác định phương pháp phòng trừ: giống lúa, mật độ sạ, nhiệt độ, độ ẩm, màu lá (độ đậm của lúa), tình trạng vết bệnh. Sau đây là một ví dụ về 8 trường hợp cụ thể trong tập dữ liệu được sử dụng nghiên cứu.

Bảng 2: Thông tin dữ liệu nghiên cứu

Giong lúa	Mat đo sạ	Nhiệt độ	Độ ẩm	Màu lá	Tình trạng bệnh	PP
OM1490	18	30	83	5	1	7
VD20	18	28	82	5	3	9
VD95-20	23	27	91	4	1	12
OM2395	16	25	90	4	4	24
IR64	20	29	83	4	5	27
IR64	19	30	85	4	2	29
OM3240	12	25	90	4	3	30
OM2517	18	29	91	4	4	31

3.4 Gợi ý dựa trên nội dung cho ứng dụng BLASTRec

Hệ thống gợi ý dựa trên nội dung chủ yếu dựa vào đặc điểm của các mục dữ liệu, ở đây mục dữ liệu là các phương pháp phòng trừ, và các thuộc tính của một mục dữ liệu chính là 6 thuộc tính ảnh hưởng đến việc xác định phương pháp phòng trừ: giống lúa, mật độ sạ, nhiệt độ, độ ẩm, màu lá, tình trạng vết bệnh. Trong nghiên cứu này, chúng tôi sử dụng phương pháp phân lớp Bayes thơ ngây và cây quyết định để cài đặt cho ứng dụng BLASTRec cũng như đánh giá hiệu quả của ứng dụng (Hình 3). Người dùng thông qua điện thoại của mình sẽ cung cấp các thông tin cần thiết, ứng dụng sẽ tự động đưa ra phương pháp phòng trừ bệnh đạo ôn phù hợp với thông tin người nông dân đã cung cấp.



Hình 3: Kiến trúc của hệ thống

3.4.1 Xây dựng ứng dụng BLASTRec sử dụng phương pháp Bayes thơ ngây (Naive Bayes)

Để xác định phương pháp phòng trừ bệnh đạo ôn trên cây lúa. Khi phân tích tập dữ liệu thu được về bệnh được quan sát như sau (Nguyễn Văn Viên và ctv., 2013; Phạm Văn Kim, 2015):

Dữ liệu liệt kê gồm các thuộc tính (giống lúa, màu lá, tình trạng bệnh).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Dữ liệu liên tục gồm các thuộc tính (mật độ sạ, nhiệt độ, độ ẩm) dùng hàm mật độ xác suất nên tính theo công thức sau:

+ Hàm mật độ xác suất $f(x)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

+ Standard deviation σ

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$$

+ Mean μ

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

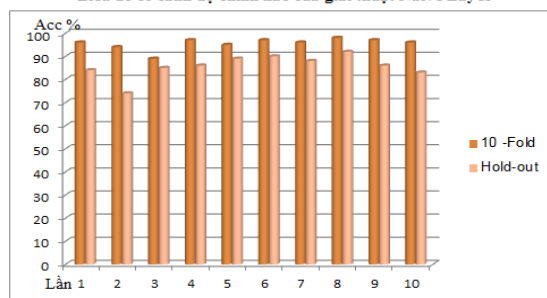
Tuy nhiên, miền giá trị của các thuộc tính có độ chênh lệch khá cao vì để đảm bảo sự chênh lệch trong mức cho phép tiến hành chuẩn hóa dữ liệu về khoảng $[0,1]$. Sau đó tiến hành tính xác suất của từng thuộc tính như: giống lúa, mật độ sạ, nhiệt độ, độ ẩm, màu lá, tình trạng vết bệnh. Lưu ý, các dữ liệu dạng liên tục thì sử dụng hàm mật độ xác suất.

$$\text{new_val} = (\text{val} - \text{min}) / (\text{max} - \text{min})$$

Kết quả kiểm tra của giải thuật Naive Bayes

Tiến hành kiểm tra độ chính xác của giải thuật theo nghi thức k-fold (10- Fold) và Hold-out, độ chính xác trung bình sau 10 lần lặp của 10-fold là 95,5% và hold-out là 85,7%. Hình 4 là biểu đồ thể hiện độ chính xác của giải thuật Bayes thơ ngây sau 10 lần lặp.

Biểu đồ so sánh độ chính xác của giải thuật Naive Bayes



Hình 3: Biểu đồ so sánh độ chính xác của giải thuật Naive Bayes

3.4.2 Xây dựng ứng dụng BLASTRec sử dụng phương pháp cây quyết định (Decision Tree)

Để xác định phương pháp phòng trừ bệnh đạo ôn trên cây lúa. Khi phân tích tập dữ liệu thu được về bệnh được quan sát như sau:

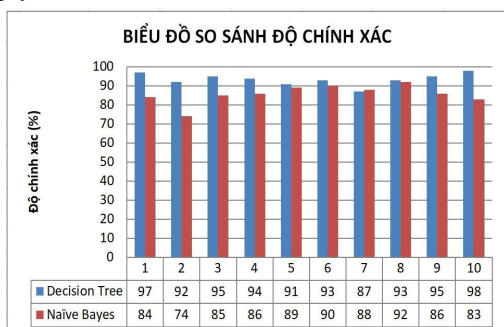
Dữ liệu liệt kê gồm các thuộc tính (giống lúa, màu lá, tình trạng bệnh).

Dữ liệu liên tục gồm các thuộc tính (mật độ sạ, nhiệt độ, độ ẩm): có nhiều nhánh để phân hoạch.

Giải thuật học cây quyết định gồm 2 bước lớn: Xây dựng cây (Top – down), cắt nhánh (Bottom-up) để tránh học vẹt.

Kết quả kiểm tra của giải thuật Decision Tree:

Tiến hành kiểm tra độ chính xác của giải thuật theo nghi thức k-fold (10- Fold) và Hold-out, độ chính xác trung bình sau 10 lần lặp của 10-fold là 97% và hold-out là 93,5%. Hình 5 là biểu đồ thể hiện độ chính xác của giải thuật cây quyết định so sánh với giải thuật Bayes thơ ngây sau 10 lần lặp. Trong hầu như các lần lặp thì độ chính xác của phương pháp cây quyết định tốt hơn Bayes thơ ngây.



Hình 5: Biểu đồ so sánh độ chính xác của giải thuật Decision Tree và bayes thơ ngây

4 ỨNG DỤNG BLASTREC

Ứng dụng BLASTRec được xây dựng trên hệ điều hành Android, người nông dân chỉ cần cung cấp 6 thông tin về tình trạng bệnh và điều kiện môi trường: nhiệt độ, độ ẩm, màu lá, tình trạng vết bệnh, giống lúa (Hình 6), hệ thống sẽ đưa ra phương pháp phòng trừ bệnh đạo ôn một cách cụ thể và tức thời (Hình 7).

Hình 6: Thông tin tình trạng bệnh và điều kiện môi trường

Hình 7: Phương pháp phòng trừ tư vấn cho nông dân

Để đánh giá hiệu quả của hệ thống, chúng tôi tiến hành đánh giá độ chính xác trên cả hai giải thuật Bayes thơ ngây và cây quyết định. Nghi thức hold-out và 10-fold được sử dụng để đánh giá đều cho độ chính xác cao. Tuy nhiên, độ chính xác của nghi thức Hold-out đối với giải thuật Bayes thơ ngây cũng như cây quyết định tương đối thấp so với phương pháp 10-fold. Điều này cho thấy ảnh hưởng của độ lớn tập dữ liệu để xây dựng mô hình lên độ chính xác của hệ thống.

Giải thuật	Độ chính xác	
	10 Fold	Holdout
Bayes thơ ngây	95,50%	85,70%
Cây quyết định	97,0%	93,5%

5 KẾT LUẬN

Trong nghiên cứu này, hệ thống gợi ý dựa trên nội dung đã áp dụng thành công để xây dựng hệ thống hỗ trợ người nông dân phòng trừ bệnh đạo ôn - BLASTRec. BLASTRec giúp người nông dân cung cấp thông tin về tình trạng bệnh của ruộng lúa và đưa ra phương pháp phòng trừ gợi ý trong khoảng thời gian ngắn và phù hợp với tình huống cụ thể của người nông dân dựa trên những dữ liệu thu thập được cũng như kinh nghiệm của chuyên gia nông nghiệp.

Tuy nhiên, tập dữ liệu chúng tôi sử dụng chưa được đa dạng vùng miền mà chỉ tập trung tại một đơn vị hành chính là huyện Trung An, quận Thốt Nốt, thành phố Cần Thơ. Trong tương lai, chức năng chụp ảnh lá lúa để dự đoán màu lá (độ đậm)

được sử dụng mang đến sự tiện lợi và chính xác hơn dành cho người sử dụng.

TÀI LIỆU THAM KHẢO

- Adomavicius G and Tuzhilin A. Towards the next generation of recommender systems, 2005: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17 (6): 734-749.
- Alvin R. Malicdem and Proceso L. Fernandez, 2015. Rice Blast Disease Forecasting for Northern Philippines. *Wseas transaction on Information science and applications*, E-ISSN: 2224-3402, volume 12.
- Balabanovic, M. and Y. Shoham, 1997. Fab: Content-based, collaborative recommendation. *Communications of the ACM* 40 (3): 66-72.
- Bonman, J. M, T. I. Vergel De Dios, và M. M. Khin, 1986. Physiologic specialization of *Pyricularia oryzae* in the Philipnies. *Plant Disease* 70: 167-169.
- Bùi Cách Tuyến và Lê Cao Lượng, 2013. Đánh giá hiệu quả kiểm soát bệnh đạo ôn lá và cổ bông của hỗn hợp một số hoạt chất hóa học trên vụ lúa xuân hè tại huyện Cai Lậy, tỉnh Tiền Giang. *Hội thảo Khoa học Công nghệ: Nông nghiệp và phát triển nông thôn - kỳ 2 - tháng 3*.
- Đỗ Thanh Nghị và Lê Thanh Vân, 2012. Các hệ tri thức và khai thác dữ liệu. *Đại học Cần Thơ*.
- G.-L. Wang and B. Valent, 2009. *Advances in Genetics, Genomics and Control of Rice Blast Disease*, ISBN 978-1-4020-9499-6, DOI 10.1007/978-1-4020-9500-9, Library of Congress Control Number: 2008940153. Springer Science+Business Media B.V.
- Huỳnh Nguyệt Ánh, Lê Thùy Nương, Võ Hiền Đức, Nguyễn Thành Phước, và Phạm Thị Liên, 2008. Bệnh cháy lá lúa và bố trí giống chống chịu bệnh ở Đồng bằng sông Cửu Long. *Vol 10*, tr 82–91.
- Huỳnh Xuân Hiệp, Nguyễn Thái Nghe và Trần Nguyễn Minh Thư, 2014. *Giáo trình mô hình hóa quyết định*. Nhà xuất bản Đại học Cần Thơ.
- Michael J. Pazzani and Daniel Billsus, 2007. Content-based recommendation systems. In *The adaptive web*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). *Lecture Notes In Computer Science*, Vol. 4321. Springer-Verlag, Berlin, Heidelberg 325-341.
- Nguyễn Hữu Hòa, Lâm Tấn Phương và Nguyễn Thái Nghe, 2016. Giải pháp nhận dạng bệnh trên lúa từ thiết bị di động thông minh. *Kỷ yếu Hội nghị khoa học công nghệ quốc gia lần thứ IX – Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR 2016)*, trang 159-164. Nhà xuất bản Khoa học tự nhiên và Công nghệ. ISBN 978-604-913-472-2.
- Nguyễn Thị Phong Lan, Võ Thị Thu Ngân, Lương Hữu Tâm, Nguyễn Đức Cường, Trần Hà Anh, Trần Phước Lộc, Trần Thị Nam Lý, Trần Thị Kiều, Nguyễn Thị Xuân Mai và Võ Thị Dạ Thảo, 2012. Nghiên cứu các giải pháp quản lý bền vững bệnh đạo ôn hại lúa vùng Đồng bằng sông Cửu Long. *Hội thảo Quốc gia về Khoa học cây trồng lần thứ hai*.
- Nguyễn Văn Viên, Hồ Viết Cường, Đỗ Tấn Dũng và Nguyễn Đức Huy, 2013. *Bệnh đạo ôn hại lúa và biện pháp phòng trừ*. Nhà xuất bản Nông nghiệp Hà Nội.
- Phạm Văn Kim, 2015. *Các bệnh hại lúa quan trọng ở đồng bằng sông Cửu Long*. Nhà xuất bản nông nghiệp.
- Prem Bahadur Magar, Basistha Acharya, và Bishnu Pandey, 2015. Use of Chemical Fungicides for the Management of Rice Blast (*Pyricularia Grisea*) Disease at Jyotinagar, Chitwan, Nepal. P.B. Magar et al., *Int J Appl Sci Biotechnol*, Vol 3(3): 474-478.
- Rakesh Kaundal, Amar S Kapoor, and Gajendra PS Raghava, 2006. Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics*.
- Trần Cao Đệ. (2016). *Nghiên cứu và ứng dụng công nghệ thông tin ở Đồng bằng sông Cửu Long*. Nhà xuất bản Đại học Cần Thơ.
- Trần Thị Thu Thủy, Nguyễn Thị Lùng, và Hans Jorgen Lyngs Jorgensen, 2005. Khảo sát khả năng kích kháng bệnh cháy lá lúa do nấm *Pyricularia grisea* (Cooke) Sacc. của dịch hại trích thực vật trên khía cạnh sinh học và mô học, vol 36, tr 57–62.
- Vũ Triệu Mẫn, 2017. *Bệnh cây chuyên khoa*. Đại Học Nông Nghiệp | Hà Nội - Khoa Nông Học.