

Cải thiện việc dự đoán năng suất lúa mì dựa trên các đặc điểm phụ bằng mẫu dữ liệu thực nghiệm trong môi trường nhiệt độ khắc nghiệt

Nguyễn Minh Tiến^{1,2,3}, Tạ Nhật Minh^{1,2,3}, Nguyễn Hoàng Minh^{1,2,3}, Nguyễn Thiện Thuật^{1,2,3}, and Đỗ Trọng Hợp^{1,2,4}

¹ University of Information Technology, Ho Chi Minh City, Vietnam

² Vietnam National University, Ho Chi Minh City, Vietnam

³ {20522010, 20521614, 20521609, 20521998}@uit.edu.vn

⁴ hopdt@uit.edu.vn

Ngày 12 tháng 7 năm 2022

Tóm tắt nội dung

Mục tiêu chính để cải thiện lúa mì (*Triticum aestivum*) là làm thay đổi năng suất. Tuy nhiên, có nhiều nguyên nhân đã làm ảnh hưởng đến năng suất lúa mì như mức độ khảo nghiệm thực tế, biến động môi trường và lượng thời gian cần thiết để có được các đánh giá trong nhiều năm. Các đặc điểm thứ cấp như độ phản xạ quang phổ và nhiệt độ tán cây (CT), có thể được đo nhanh nhiều lần trong suốt mùa sinh trưởng, thường nó sẽ tương quan với năng suất hạt và có thể được sử dụng để chọn lọc gián tiếp trong các quần thể lớn, đặc biệt là ở các thể hệ trước trong chu kỳ nhân giống để kiểm tra năng suất nhân rộng. Trong khi đó, việc thu thập dữ liệu cảm biến càng ngày càng được triển khai với các nền tảng đa dạng cung cấp thông tin mạnh mẽ và giá cả phải chăng, việc sử dụng hiệu quả những dữ liệu này đang là một thách thức. Mục tiêu của nghiên cứu là theo dõi sự phát triển của lúa mì và dự đoán năng suất hạt trong các lần thử nghiệm nhân giống lúa mì bằng cách sử dụng các phép đo cảm biến với mật độ cao trong điều kiện nhiệt độ khắc nghiệt vào giai đoạn cuối cực phổ biến ở Bangladesh. Qua năm mùa trồng trọt, chúng tôi đã phân tích chỉ số chuẩn hóa ở thực vật (NDVI) và các phép đo CT được thu thập trong các loại giống siêu trội từ Trung tâm cải tiến ngô và lúa mì quốc tế tại Trạm nghiên cứu nông nghiệp khu vực, Jamalpur, Bangladesh. Chúng tôi đã khám phá một số kỹ thuật khử biến đổi và điều hòa, sau đó sử dụng các đặc điểm thứ cấp kết hợp để dự đoán năng suất ngũ cốc. Qua các năm, sản lượng ngũ cốc có sự khác biệt rõ rệt, mối tương quan giữa năng suất hạt và các tính trạng thứ cấp dao động từ -0,55 đến 0,53. Độ chính xác của dự đoán được tính toán bằng độ đo R^2 và được xây dựng các dự đoán bằng cách sử dụng các mô hình nhóm đơn biến và đa biến. Chúng tôi nhận thấy rằng các mô hình kết hợp các yếu tố phụ cho độ chính xác dự đoán về năng suất ngũ cốc cao hơn so với các mô hình nhóm đơn biến (tốt hơn khi chỉ cho dự đoán dựa trên các nhóm nông học). Trong đó, kết quả cho thấy mô hình Ridge regression xây dựng trên tất cả thuộc tính (CT, NDVI, các yếu tố nông học) cho kết quả thực hiện tốt hơn tương đối so với các mô hình khác trong việc dự đoán năng suất ngũ cốc. Khi kết hợp tất cả các đặc điểm phụ vào các mô hình, chúng tôi thu được độ chính xác dự đoán dao động từ 0,42–0,8 trong năm mùa trồng trọt. Kết quả cho thấy không có sự khác biệt giữa các dự đoán. Các mô hình dự đoán có thể tận dụng các đặc điểm thứ cấp để đưa ra các dự đoán chính xác về năng suất hạt lúa mì, cho phép các chương trình nhân giống thực hiện các lựa chọn nhanh chóng và mạnh mẽ hơn.

Từ khóa: Lúa mì, phân tích thống kê, dự đoán năng suất ngũ cốc, chỉ số thực vật khác biệt chuẩn hóa, năng suất ngũ cốc, nông học, nhiệt độ khắc nghiệt, ridge, lasso, decision tree, random forest, support vector regression.

1 Giới thiệu

Lúa mì chiếm 26% sản lượng ngũ cốc thế giới và 44% tổng lượng ngũ cốc tiêu thụ [6]. Kinh tế nhanh chóng và tăng trưởng thu nhập, đô thị hóa và toàn cầu hóa đang dẫn đầu đến những thay đổi mạnh mẽ về chế độ ăn uống, đặc biệt là ở Châu Á nói chung và ở Việt Nam nói riêng khi người tiêu dùng tăng tiêu thụ các sản phẩm lúa mì của họ. Ở thời điểm hiện tại, sản lượng lúa mì tăng còn chậm so với nhu cầu tiêu thụ của thế giới và có thể trong tương lai sản lượng lúa mì còn có thể bị giảm do sức ảnh hưởng của biến đổi khí hậu. Trong khi lúa mì được phân bố trên toàn cầu và phải đối mặt với nhiều thách thức sinh học và phi sinh học, thì ở Nam Á, nhiệt độ là một trong những yếu tố quan trọng và nó làm hạn chế năng suất lúa mì ở các vùng này và nó cũng xảy ra ở các vùng khác có nhiệt độ khắc nghiệt [4]. Môi trường nhiệt độ khắc nghiệt có thể tác động rất nhiều đến kiểu hình của thực vật (ví dụ như nhiệt độ của tán cây - CT) làm cho cây sinh trưởng không bình thường và sẽ làm suy thoái khả năng sinh trưởng nếu nhiệt độ quá cao.

Nhiều khu vực sản xuất lúa mì trên thế giới đã có nhiệt độ được dự đoán là tiếp tục tăng từ 1,7 đến 4,8°C vào cuối thế kỷ này [8]. Do đó, tăng năng suất hạt dưới nhiệt độ cao là một mục tiêu được đặt lên hàng đầu kết hợp với các phương pháp và công nghệ để tăng tỷ lệ di truyền đạt được trong môi trường nhiệt độ khắc nghiệt. Vì sự phức tạp của nhiệt độ nên ta không thể dùng một loại chiến lược tăng năng suất nhất định nào để cải thiện sản lượng lúa mì mà phải sử dụng một cách biến hóa và linh hoạt các biện pháp cho phù hợp với điều kiện hoàn cảnh cụ thể, nghiên cứu để tìm ra và tác động tới các yếu tố có ảnh hưởng đến năng suất của cây giúp cho cây có được năng suất tốt hơn (ví dụ như cho cây ra hoa sớm, đóng khí khổng ở cây để cho cây hạn chế mất nước,...) [1]. Sự sắp xếp của những cặp tính trạng tạo cơ hội để cải thiện lúa mì nếu biến thể di truyền có lợi được đo lường chính xác và sử dụng các phương pháp phù hợp để tác động lên các gen di truyền hợp lý. Theo truyền thống, trước khi phát hiện ra DNA và chỉ thị phân tử, các nhà chọn tạo giống cây trồng chỉ lựa chọn các dòng có triển vọng trên cơ sở kiểu hình. Bằng cách tạo ra số lượng lớn phép lai và đánh giá các thể hệ kế tiếp trong nhiều môi trường từ đó xác định được các cá thể ưu việt và dùng nó trong các kỳ tăng trưởng tiếp theo để tăng năng suất. Tuy nhiên cách này phải gặp một thách thức lớn trong việc định hình chính xác các quần thể lớn trong thời gian ngắn để xác định thể hệ con cháu tốt nhất. Với sự ra đời của các loại cây có dạng giống (gen) vượt trội đã giúp cho việc nghiên cứu trở nên dễ dàng hơn trong việc xác định các tính trạng trội có lợi một cách chính xác và nhanh chóng hơn. [3]

Việc nghiên cứu để tăng năng suất lúa trong điều kiện nhiệt độ khắc nghiệt được Bangladesh rất quan tâm trong chương trình nhân giống quốc gia nhằm cải thiện sản lượng lúa mì ở nơi này giúp cân bằng nguồn tiêu thụ trong nước cũng như xuất khẩu. Năng suất ngũ cốc là đặc điểm cơ bản được quan tâm và tất cả các yếu tố thực nghiệm được xem là đặt điểm phụ nhằm nghiên cứu các yếu tố ảnh hưởng đến năng suất lúa mì. Nếu các tính trạng thứ cấp có thể được định hình chính xác trong chương trình nhân giống, thì các tính trạng thứ cấp này có thể được sử dụng để dự đoán tính trạng chính và để cải thiện khả năng di truyền đặc biệt sớm hơn trong chu kỳ nhân giống trước khi tiến tới các thử nghiệm năng suất nhân rộng. Hai đặc điểm phụ tiềm năng có thể sửa đổi đối với các phép đo thông lượng cao bao gồm chỉ số thực vật khác biệt chuẩn hóa (NDVI)[2] và nhiệt độ tán cây (CT).[5]

Mục tiêu của nghiên cứu của chúng tôi là để đánh giá các phép đo kiểu hình theo thời gian, mật độ từ cảm nhận gần của NDVI và CT cũng như các đặc điểm nông học khác có thể được sử dụng trong các chương trình nhân giống cây trồng quốc gia của Bangladesh để đánh giá năng suất dòng trong môi trường áp lực nhiệt [7]. Ngoài ra, chúng tôi còn xây dựng các mô hình máy học thống kê trên các phép đo tương quan cao của đặc điểm phụ để từ đó có thể đánh giá và dự đoán các loại giống cho được năng suất cao trước khi nhân rộng và phục vụ cho việc đánh giá các gen di truyền cho năng suất tốt.

2 Dữ liệu

2.1 Nguồn dữ liệu

Bộ dữ liệu được chúng tôi thu thập từ zenodo.org - kho dữ liệu nghiên cứu được OpenAIRE và CERN tạo ra để cung cấp địa điểm cho các nhà nghiên cứu gửi các tập hợp dữ liệu với kích thước lớn. Dữ liệu được thực hiện cho nghiên cứu cung cấp khoảng thời gian từ năm 2016 đến năm 2020 với số lượng cột từ 32 đến 44 tùy từng giai đoạn. Dữ liệu được thu thập tại RARS (Regional Agricultural Research Station - Trạm Nghiên cứu Nông nghiệp), BARI (Bangladesh Agricultural Research Institute - Viện

Nghiên cứu Nông nghiệp Bangladesh), Jamalpur, Bangladesh. Quỹ Bill Melinda Gates (BMGF) cùng với Văn phòng Phát triển và Thịnh vượng chung (FCDO) của Vương quốc Anh đã hỗ trợ các hoạt động nhân giống lúa mì của CIMMYT thông qua dự án cung cấp tính trạng trội được di truyền trong Lúa mì do Đại học Cornell quản lý.

Bộ dữ liệu đánh giá các bộ gồm 540 dây chuyền tiên tiến của International Maize and Wheat Improvement Center (CIMMYT) (tạm dịch Trung tâm cải tiến ngô và lúa mì thế giới) qua năm mùa trồng trọt (2015–2016, 2016–2017, 2017–2018, 2018–2019 và 2019–2020) tại Bangladesh. Mỗi năm, từng bộ của 540 dây chuyền từ CIMMYT được đánh giá khi nguyên vật liệu chịu nhiệt mới xuất hiện, đồng thời, cùng vào lúc đó cũng thực hiện các bài kiểm tra cục bộ bao gồm BARI Gom 26 hoặc BARI Gom 30, kết quả từ bài kiểm tra được dùng để làm tham số điểm chuẩn cho bài kiểm tra ở Bangladesh. Tất cả các dây chuyền được thử nghiệm dưới môi trường nhiệt độ cao tại Regional Agricultural Research Station (RARS), Bangladesh Agricultural Research Institute (BARI), Jamalpur, Bangladesh (N 24.93, E 89.93, 23 masl). Khí hậu của khu vực này thường nóng và ẩm, là nơi thích hợp để tái tạo môi trường nhiệt độ cao, được phân loại là ME5A dựa trên tiêu chuẩn của hệ thống phân loại môi trường cho lúa của CIMMYT (Rajaram et al., 1993).

Để đảm bảo về sự biến đổi của không gian, các dây chuyền được chạy qua nhiều lần thử của từng mùa trồng trọt khác nhau. Mỗi cuộc thử nghiệm gồm có 60 dây chuyền, trong đó có 53 dây để nuôi trồng và 7 dây để kiểm tra độ đa dạng. Các lần thử nghiệm hoàn thiện được đem đi trồng vào các ngày phù hợp cho từng mùa như 4–8/12/2015; 25–28/11/2016; 29–30/11/2017; 28/11/2018 và 05/12/2019. Các lần thử nghiệm được sắp xếp dưới dạng thiết kế mạng alpha lặp lại 2 lần cho ta tổng cộng 120 lô đất cho mỗi lần thử nghiệm. Mỗi bên gồm 6 hàng dài 4.17 mét và cách nhau 20 cm với diện tích của lô đất là 5m². Mỗi lô đất cách nhau bằng một rãnh sâu 40 cm. Mùa năm 2015 – 2016 có tổng cộng 10 lần thử. Các năm sau đó có 11 lần thử, với lần thử thứ 11 là lần thử đạt kết quả cao nhất của mùa trước.

Các hoạt động nông học của Bangladesh Wheat Research Center được thực hiện trong các mùa trồng trọt. Phân bón được áp dụng gồm N:P:K:S:Zn:B với tỉ lệ tương ứng là 100:26:50:20:5:1 kg/ha mỗi năm. Việc tưới tiêu được thực hiện khi có biểu hiện của sự thiếu nước. Vào mùa trồng trọt năm 2015–2016, 3 lần tưới tiêu được thực hiện khi xới đất, tủa ngọn và độn hạt trong khi năm 2016–17 và 2019–20, 2 lần tưới tiêu được thực hiện khi xới đất và làm đồng (Zadoks et al., 1974). Nhổ cỏ bằng tay được thực hiện mỗi mùa để giữ cho nơi trồng trọt không bị mọc cỏ dại. Thuốc diệt côn trùng không được sử dụng trong mùa trồng trọt.

2.2 Bộ dữ liệu - các thuộc tính có trong bộ dữ liệu

Bộ dữ liệu cho thấy sản lượng lúa là thuộc tính chính, CT (Cycle threshold - ngưỡng chu kỳ) và NDVI (Normalized Difference Vegetation Index - mức độ tập trung của thực vật trên mặt đất) được hiểu như thuộc tính thứ cấp dựa trên cảm biến, và tất cả thuộc tính khác là đặc điểm nông học. Tổng sản lượng lúa của mỗi ô được thu hoạch, phơi khô, cân và chia cho kích thước ô (5m²) để có được sản lượng (kg/m²) và sau đó quy đổi sang đơn vị tấn trên hecta. Trong suốt mùa sinh trưởng, dữ liệu kiểu hình được ghi lại cho các đặc điểm nông học như độ che phủ mặt đất (GrndCov), DTHD, số ngày đến khi chín (DAYSMT), PH, lượng hạt trên cành (GRNSPK), bệnh bạc lá do nhiễm *Helminthosporium* nghiêm trọng (HELSPSEV), số lượng bông lúa trên mỗi đơn vị diện tích (SN), số lượng bông con trên mỗi bông lúa (SPLN), độ dài của bông (SPKLNG) và trọng lượng nghìn hạt (TGW). GrndCov là một đơn vị ước tính trực quan của mặt đất được bao phủ bởi sinh khối của cây trồng bắt đầu từ 30 ngày sau khi gieo và tiếp tục trong 15 ngày. DTHD được ghi nhận là số ngày mà 50% tổng số cây trong một ô đã kéo dài bông lúa ra khỏi bẹ lá. DAYSMT được ghi nhận khi 80% số cây trong lô có phần thân chuyển từ xanh sang vàng. Chiều cao của cây được đo bằng chiều dài từ mặt đất đến đỉnh của bông lúa không bao gồm ngọn bông lúa. Tổng số hạt lúa từ 5 bông lúa được đếm và chia cho 5 để có được thông số GRNSPK. HELSPSEV được tính theo thang điểm để đánh giá cường độ bệnh trên lá lúa mì (Saari và Prescott, 1975). Tổng số bông lúa trên một mét vuông (tức là SN) được đo bằng cách đo số lượng bông lúa đếm được từ lô dài 3,5 m, rộng 20 cm (0,7 m²) và được chuyển đổi thành số lượng bông lúa trên một mét vuông. SPKLNG được đo trên một bông lúa miêu tả trong biểu đồ là chiều dài từ gốc đến đỉnh của một bông lúa không bao gồm ngọn bông lúa.

Các thuộc tính thứ cấp của dữ liệu CT (Cycle threshold - ngưỡng chu kỳ) và NDVI (Normalized Difference Vegetation Index - mức độ tập trung của thực vật trên mặt đất) được thu thập từ 8 đến 15 lần trong các mùa trồng trọt (8, 14, 12, 13 và 15 các mốc thời gian cho các mùa lần lượt là 2016–2017, 2017–2018, 2018–2019 và 2019–2020). Các phép đo miêu tả cho sự phát triển của thực vật từ khi xới

đất đến khi già đi (Zadoks et al., 1974) với các phép đo lường mặt trời vào buổi trưa được thực hiện từ 11 giờ đến 14 giờ của mỗi ngày quan sát được. CT được đo bằng nhiệt kế hồng ngoại cầm tay (IRT) (Apogee, Logan, UT, USA), cung cấp độ chính xác cao, đo nhiệt độ bề mặt không tiếp xúc từ 30 đến 65 ° C với độ chính xác $\pm 0,124^{\circ}\text{C}$. Các nghiên cứu IRT được thực hiện ở góc 30° đo từ đường chân trời và 70 cm trên tán cây (Pask et al., 2012). IRT hoạt động ở 0,6 hertz, nhưng chỉ có CT trung bình được ghi lại cho mỗi lần đo. NDVI được thu thập bằng cảm biến cầm tay GreenSeeker (Trimble Inc. Sunnyvale, CA, USA). GreenSeeker được sử dụng bằng cách đưa cảm biến 75 cm qua tán cây. Đội gồm hai người được sử dụng để thu thập CT và NDVI, với một người vận hành thiết bị và người kia ghi lại dữ liệu. Mất 3 giờ với hai đội (tức là bốn người) để đo CT và NDVI của tất cả các lô. Dữ liệu được ghi lại trong kế hoạch Field Book (Rife và Ba Lan, 2014).

3 Phân tích thống kê

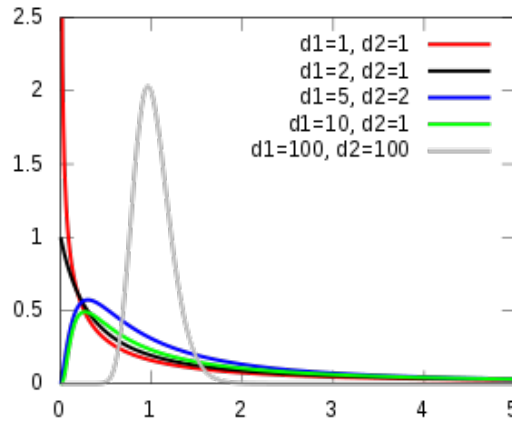
3.1 Cơ sở lý thuyết kiểm định sự khác nhau giữa các yếu tố

3.1.1 One – way ANOVA

One-way ANOVA là một loại thử nghiệm thống kê nhằm đánh giá tác động của một biến độc lập duy nhất lên một biến phản hồi duy nhất. Nó là một công cụ rất hữu ích để xác định liệu có bất kỳ sự khác biệt có ý nghĩa nào giữa các giá trị trung bình của các đối tượng của ba (hay nhiều nhóm) mẫu độc lập với nhau hay không với khả năng phạm sai lầm chấp nhận cho phép (thông thường là 5%).

$$F = \frac{s_w^2}{s_a^2}(1)$$

Với s_w^2 là ước lượng của σ^2 được tính từ phương sai si bên trong các nhóm (Within group variance), s_a^2 là ước lượng của phương sai chung (Among group variance).



Hình 1: F - Distribution

3.1.2 Kiểm định hậu định Post-hoc anova - Tukey HSD

Kiểm định HSD (Honestly significant difference) là kiểm định trung bình khác nhau của sự khác biệt từng cặp trong phân tích thống kê. Được tính bằng công thức dưới đây:

$$q = \frac{|ybar_i - ybar_j|\sqrt{2}}{\sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}(2)$$

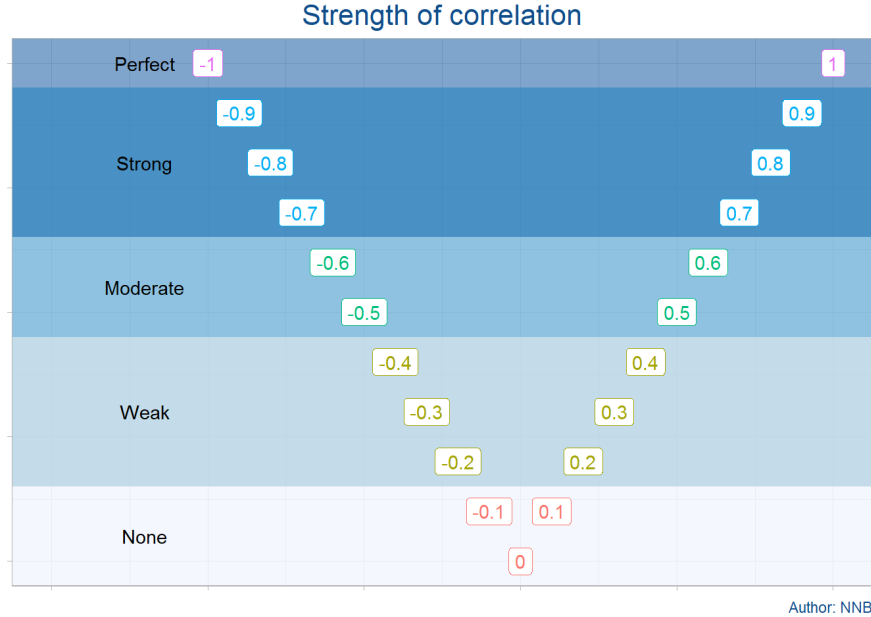
3.1.3 Hệ số tương quan Pearson (Correlation Coefficient)

Trong phân tích thống kê, hệ số tương quan (Kí hiệu là r) đóng một vai trò quan trọng, vì nó làm rõ mối quan hệ giữa các yếu tố khác nhau. Hệ số tương quan là chỉ số thống kê đo lường mức độ mạnh yếu

của mối quan hệ giữa hai biến số, được biểu diễn bằng công thức toán học sau:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (3)$$

Các giá trị thể hiện mức độ tương quan nằm trong khoảng $[-1,1]$, mức độ tương quan giữa hai thuộc tính mạnh hay yếu được quy ước như hình:



Hình 2: Strength of correlation

3.2 Kết quả phân tích

3.2.1 Kiểm định sự khác nhau giữa các nhóm thuộc tính qua các năm

Bảng 3.2.1: Bảng kết quả kiểm định One-way ANOVA các yếu tố nông học (Agronomic Traits) qua các năm

Agronomic Traits	Df	Sum Square	Mean Sq	F value	Pr(>F)
Day to heading growth	4	57377	14344	806.1	<2e-16 ***
Day to maturity	4	77476	19369	1762	<2e-16 ***
Plant height	4	104012	26003	595	<2e-16 ***
Number of spikes per m2	4	21146909	5286727	1462	<2e-16 ***
Length of spike*	4	232	77.34	50.59	<2e-16 ***
Spikelet number per spike	4	1960	490.0	96.15	<2e-16 ***
Grains per spike	4	16085	4021	56.62	<2e-16 ***
Thousand kernel weight	4	114697	28674	796.2	<2e-16 ***

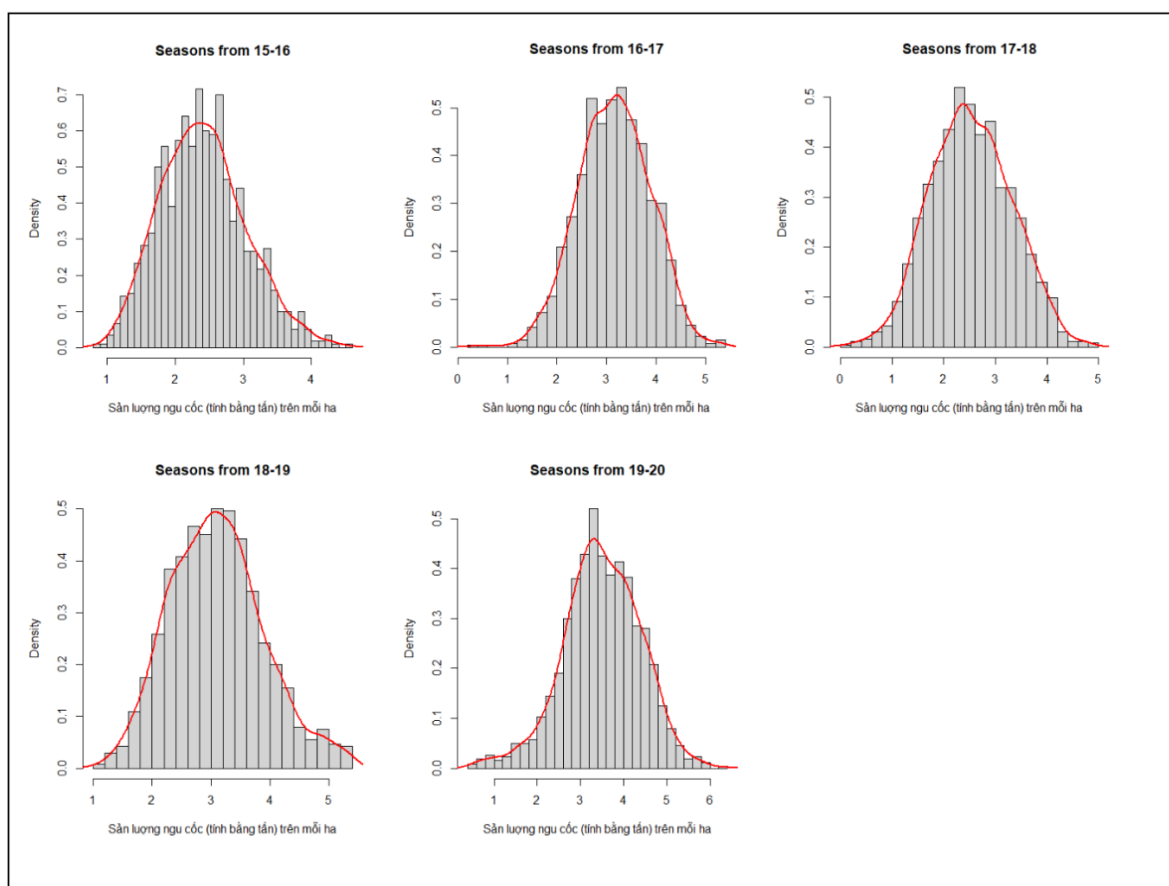
Ghi chú: H0: Không có sự khác nhau giữa các nhóm theo từng thuộc tính nông học qua các năm , *: Không có dữ liệu vào mùa vụ 2020. ***: Có ý nghĩa thống kê với mức ý nghĩa alpha < 0.001.

Từ bảng trên (table) cho thấy tất cả các yếu tố nông học (Agronomic Traits) có sự khác nhau rõ rệt qua các mùa vụ với mức ý nghĩa rất thấp. Điều này chứng tỏ có sự khác nhau trong từng cặp mùa là rất lớn. Sự chênh lệch giữa sản lượng ngũ cốc (xem table dưới) có thể do môi trường nhiệt độ ở bangladesh tương đối khắc nghiệt và thất thường qua từng năm đã làm ảnh hưởng đến sản lượng ngũ cốc qua các mùa làm cho chênh lệch sản lượng ngũ cốc có sự khác biệt giữa các nhóm (xem thêm hình).

Bảng 3.2.2: Bảng kết quả Kiểm định Tukey HSD giữa sản lượng ngũ cốc (Grain yield) qua các năm

	Diff	P(>q)
2017-2016	0.745	0.0 ***
2018-2016	0.116	0.001 ***
2019-2016	0.687	0.0 ***
2020-2016	1.088	0.0 ***
2018-2017	-0.629	0.0 ***
2019-2017	-0.058	0.323 ns
2020-2017	0.342	0.0 ***
2019-2018	0.571	0.0 ***
2020-2018	0.972	0.0 ***
2020-2019	0.4	0.0 ***

Ghi chú: H0: Không có sự khác nhau giữa năng suất ngũ cốc qua các năm, ns: Không có ý nghĩa thống kê với mức ý nghĩa $\alpha = 0.05$. ***: Có ý nghĩa thống kê với mức ý nghĩa $\alpha < 0.001$



Hình 3: Bảng kết quả mối tương quan giữa các yếu tố được đo lường với năng suất theo từng năm

3.2.2 Mối tương quan giữa các đặt điểm đo lường

Bảng 3.2.3: Bảng kết quả mối tương quan giữa các yếu tố được đo lường với năng suất theo từng năm

Traits	Correlation Coefficient (r)				
	2015–16	2016–17	2017–18	2018–19	2019–20
CT Average	-0.06	-0.45*	-0.55**	-0.45*	-0.43*
NDVI Average	0.48*	0.32*	0.39*	0.53**	0.52**
Day to heading growth	-0.06	-0.24	-0.31*	-0.18	-0.17
Day to maturity	0.23	0.05	0.04	-0.01	0.15
Plant height	0.37*	0.26	0.32*	0.35*	0.43*
Number of spikes per m ²	0.33*	0.23	0.46*	0.14	0.37*
Length of spike	0.09	0.09	0.2	0.05	-
Spikelet number per spike	0.12	0.06	0.04	-0.003	0.01
Grains per spike	0.14	0.16	0.06	0.14	0.12
Thousand kernel weight	0.18	0.24	0.38*	0.09	0.24

Ghi chú: * Tương quan yếu, ** Tương quan vừa phải

Chúng tôi tiến hành tính toán các tương quan kiểu hình cho tất cả các thuộc tính CT, NDVI và nông học được đo lường để xem xét tất cả các thử nghiệm cùng nhau để xác định mối quan hệ giữa chúng và GRYLD (table). Trong đó, các thuộc tính CT, NDVI được chúng tôi tính các giá trị trung bình các lần lấy kết quả trong từng năm trước khi tiến hành tính toán các mối tương quan.

Kết quả cho thấy, HTHD cho thấy dường như có mối tương quan tỉ lệ nghịch với năng suất nhưng yếu (có mùa rất yếu $r = -0.06$). DAYSMT cũng cho thấy mối tương quan tỉ lệ thuận trong bốn trong số năm mùa tăng trưởng nhưng yếu. Mối tương quan cao nhất được quan sát thấy đối với các yếu tố nông học là giữa SN (Number of spikes per m²) và GRYLD là 0.46, tiếp theo là giữa PH (Plant height) và GRYLD ($r = 0.43$) trong mùa vụ 2019 – 20. Các yếu tố CT, NDVI, HTHD, PH, SN, SPKLN (Length of spike), GRNSPK (Grains per spike), TGW (Thousand kernel weight) có tính nhất quán đối với GRYLD (năng suất ngũ cốc) qua các mùa trồng trọt.

Cả 2 thuộc tính CT và NDVI đa phần có tương quan với năng suất qua từng mùa, nằm ở mức tương quan khá yếu và vừa phải (mùa 2017 – 18 có hệ số tương quan với CT lên tới 0.55 và mùa 2018 – 19 hệ số tương quan với NDVI là 0.53), năng suất ngũ cốc (GRYLD) có tương quan tỉ lệ nghịch với CT và tỉ lệ thuận với NDVI. Tương quan giữa NDVI và GRYLD được giữ ổn định qua các mùa, tương tự, CT và GRYLD cũng tương quan tương đối ổn định trừ mùa vụ 2015–16. Nhìn chung, mối tương quan giữa các đặt điểm phụ so với năng suất ngũ cốc là khả quan và có sự ảnh hưởng nhất định.

4 Xây dựng mô hình dự đoán năng suất

4.1 Mô hình dự đoán cho sản lượng lúa mì

Chúng tôi sử dụng các thuật toán máy học cho mỗi đặc điểm, năm mô hình máy học khác nhau đã được sử dụng để dự đoán sản lượng lúa mì bằng cách sử dụng nhiều phép đo NDVI, CT và các đặc điểm nông học. Các mô hình được áp dụng bao gồm Ridge Lasso Regression, Support Vector Regression, Random Forest Regression, Decision Tree. Ridge và Lasso Regression là hai trong số các mô hình Regression (Hồi quy). Hai mô hình này là biến thể của Linear Regression (Hồi quy tuyến tính) được sử dụng để xử lý vấn đề về overfitting với tập dữ liệu lớn. Ridge Regression (L2 regularization) sẽ được thêm hệ số tổng bình phương của các hệ số nhằm mục tiêu tối ưu hóa. Còn Lasso Regression (L1 Regularization) sẽ thêm hệ số tổng giá trị tuyệt đối của các hệ số. Support Vector Regression (SVR) là mô hình cung cấp một sự linh hoạt để xác định mức độ lỗi có thể chấp nhận được trong mô hình và sẽ tìm một đường thích hợp (hoặc là một siêu phẳng ở các kích thước cao hơn) để phù hợp với dữ liệu. SVR sử dụng nguyên tắc tương tự như SVM, nhưng chủ yếu với các vấn đề hồi quy. Có nghĩa là tìm một hàm gần đúng ánh xạ từ miền đầu vào tạo thành các số thực trên cơ sở một mẫu huấn luyện. Random Forest là một thuật toán học máy có giám sát sử dụng phương pháp học máy theo nhóm để phân loại và hồi quy. Random là ngẫu nhiên, Forest là rừng, nên ở thuật toán Random Forest mình sẽ xây dựng nhiều cây quyết định bằng thuật toán Decision Tree, tuy nhiên mỗi cây quyết định sẽ khác

nhau (có yếu tố random). Sau đó kết quả dự đoán được tổng hợp từ các cây quyết định. Decision Tree là một công cụ ra quyết định sử dụng cấu trúc giống cây lưu đồ hoặc là một mô hình về các quyết định và tất cả các kết quả có thể có của chúng, bao gồm cả kết quả, chi phí đầu vào và tiện ích. Các mô hình được xây dựng trong một quy trình lặp đi lặp lại; cho mỗi năm, chúng tôi dự đoán các mô hình thông qua NDVI, CT, và nhóm thuộc tính nông học với nhau.

4.2 Sử dụng mô hình dự đoán năng suất lúa mì bằng một nhóm thuộc tính đơn lẻ

Yield predictors	Models	2015-2016		2016-2017		2017-2018		2018-2019		2019-2020	
		r	R ²	r	R ²	r	R ²	r	R ²	r	R ²
Canopy temperature (CT)	SVR	0.48	0.2	0.56	0.31	0.7	0.49	0.55	0.3	0.67	0.44
	RFR	0.47	0.22	0.54	0.27	0.71	0.51	0.57	0.33	0.68	0.43
	Ridge	0.32	0.1	0.55	0.29	0.67	0.45	0.53	0.28	0.64	0.39
	Lasso	0.32	0.09	0.54	0.28	0.67	0.45	0.52	0.27	0.64	0.39
	Decision tree	0.24	0.03	0.28	0.07	0.64	0.41	0.47	0.22	0.57	0.31
Normalized difference vegetation index (NDVI)	SVR	0.53	0.28	0.47	0.22	0.71	0.5	0.68	0.45	0.69	0.47
	RFR	0.53	0.29	0.51	0.25	0.71	0.49	0.69	0.48	0.69	0.47
	Ridge	0.54	0.29	0.46	0.2	0.71	0.51	0.69	0.47	0.69	0.47
	Lasso	0.54	0.28	0.46	0.2	0.71	0.51	0.69	0.47	0.69	0.47
	Decision tree	0.48	0.23	0.38	0.14	0.59	0.35	0.6	0.36	0.54	0.29
Nhóm các đặc điểm nông học	SVR	0.63	0.4	0.6	0.35	0.71	0.5	0.53	0.26	0.68	0.45
	RFR	0.62	0.35	0.6	0.35	0.66	0.42	0.53	0.27	0.62	0.37
	Ridge	0.62	0.38	0.57	0.31	0.7	0.49	0.51	0.26	0.66	0.43
	Lasso	0.62	0.38	0.57	0.31	0.7	0.49	0.51	0.26	0.66	0.43
	Decision tree	0.4	0.16	0.41	0.16	0.41	0.16	0.32	0.1	0.5	0.25

Bảng 2.1 Kết quả dự đoán khi sử dụng mô hình dự đoán bằng một nhóm thuộc tính đơn lẻ

Các mô hình cho kết quả với mức giải thích của các biến là R^2 từ 0.03 đến 0.51, tương quan giữa kết quả dự đoán và kết quả thực tế là r từ 0.24 đến 0.71 (Bảng 2.1). Khi dự đoán với các biến của nhóm thuộc tính CT, NDVI hay các đặc điểm nông học thì kết quả đầu ra không ổn định, sự chênh lệch giữa các mô hình và các năm là tương đối cao. Ngoài ra, mô hình Decision tree cho kết quả thấp. Hai mô hình Support Vector Regression (SVR) và Random Forest Regression (RFR) cho kết quả cao hơn so với phần còn lại. Mặt khác khi sử dụng nhóm thuộc tính NDVI, kết quả cũng cao hơn so với 2 nhóm thuộc tính còn lại.

Nhìn chung, sử dụng một nhóm đặc điểm thứ cấp hoặc nông học duy nhất, kết quả không phù hợp với độ chính xác dự đoán từ 0.03 đến 0.51. Độ chính xác dự đoán của các nhóm đặc điểm thứ cấp riêng lẻ rất khác nhau tùy thuộc vào đặc điểm và thời gian đo lường, với các nhóm đặc điểm CT và NDVI được đo xung quanh giai đoạn sinh trưởng cung cấp các giá trị cao nhất, trong khi tại thời điểm điểm đầu hoặc cuối mùa sinh trưởng có các giá trị không nhất quán, khiến việc sử dụng đơn lẻ các nhóm thuộc tính này cho kết quả không cao so với việc sử dụng kết hợp nhóm các thuộc tính nông học.

4.3 Sử dụng mô hình dự đoán năng suất lúa mì bằng nhóm thuộc tính kết hợp

Yield predictors	Models	2015-2016		2016-2017		2017-2018		2018-2019		2019-2020	
		r	R ²	r	R ²	r	R ²	r	R ²	r	R ²
NDVI và CT	SVR	0.58	0.33	0.57	0.32	0.74	0.54	0.67	0.45	0.75	0.54
	RFR	0.63	0.38	0.57	0.32	0.75	0.56	0.71	0.49	0.74	0.5
	Ridge	0.56	0.32	0.56	0.3	0.74	0.55	0.69	0.47	0.73	0.53
	Lasso	0.56	0.32	0.56	0.3	0.74	0.54	0.69	0.48	0.73	0.53
	Decision tree	0.41	0.16	0.4	0.16	0.66	0.44	0.62	0.38	0.62	0.39
Tất cả thuộc tính	SVR	0.68	0.46	0.71	0.49	0.78	0.61	0.7	0.47	0.76	0.57
	RFR	0.71	0.48	0.7	0.46	0.79	0.61	0.71	0.51	0.78	0.55
	Ridge	0.7	0.49	0.71	0.5	0.8	0.63	0.73	0.53	0.8	0.63
	Lasso	0.71	0.5	0.71	0.5	0.79	0.63	0.73	0.53	0.79	0.62
	Decision tree	0.42	0.17	0.39	0.15	0.62	0.38	0.55	0.3	0.56	0.31

Bảng 2.2 Kết quả dự đoán năng suất lúa mì khi sử dụng mô hình đa biến

Độ chính xác dự đoán năng suất của các mô hình R² từ 0.16 đến 0.63 và sự tương quan giữa giá trị dự đoán với giá trị thực là r từ 0.41 đến 0.8 (Bảng 2.2). Khi sử dụng các biến nhóm CT kết hợp với NDVI làm biến dự đoán, các mô hình cho kết quả gần như là giống nhau, nhưng mức độ giải thích bởi mô hình SVR cao hơn nhưng không đáng kể so với các mô hình khác. Khi sử dụng tất cả các biến để dự đoán, kết quả dự đoán tăng rõ rệt so với các mô hình trước đó, trong đó mô hình Ridge và Lasso cho kết quả cao hơn so với các mô hình còn lại. Mặt khác, mô hình Decision tree tiếp tục cho kết quả thấp, không phù hợp với bộ dữ liệu này.

Nhìn chung, với việc kết hợp các nhóm thuộc tính với nhau, kết quả dự đoán của các mô hình tăng lên đáng kể. Đặc biệt khi kết hợp nhóm các đặc điểm nông học với nhóm biến CT và NDVI, khả năng dự đoán được cải thiện hơn và sự chênh lệch dự đoán giữa các mô hình được giảm thiểu.

4.4 Thảo luận và đánh giá

Khi sử dụng 5 mô hình để đánh giá, chúng tôi nhận thấy mô hình Decision tree cho kết quả thấp trong khi 4 mô hình còn lại cho kết quả tương đương nhau. Có thể mô hình này không thực sự phù hợp với xu hướng của bộ dữ liệu này. Vì vậy chúng tôi quyết định không xét kết quả của mô hình Decision tree trong quá trình thảo luận và đánh giá.

Các mô hình dự đoán năng suất hạt được phát triển lặp đi lặp lại với hai đặc điểm thứ cấp riêng biệt, đó là NDVI và CT, và các đặc điểm nông học khác với độ chính xác dự đoán trong phạm vi 0.09–0.51 chỉ sử dụng CT (Bảng 2.1). Sử dụng NDVI, độ chính xác dự đoán thường cao hơn so với chỉ sử dụng CT trong khoảng từ 0.2 đến 0.51. Khi chúng tôi kết hợp cả NDVI và CT vào mô hình, độ chính xác của dự đoán tiếp tục tăng từ 0.3 đến 0.56 (Bảng 2.2). Việc kết hợp tất cả các đặc điểm lại với nhau dẫn đến độ chính xác dự đoán tổng thể cao nhất từ 0.46 đến 0.63 trong những năm thử nghiệm.

Chúng tôi đã đánh giá cách các đặc điểm đo được có thể được sử dụng để dự đoán năng suất hạt thông qua các mô hình thống kê khác nhau. Chúng tôi đã sử dụng một số mô hình đơn biến để dự đoán năng suất hạt bằng cách sử dụng dữ liệu kiểu hình vì chúng tôi dự định so sánh mô hình đơn biến với các mô hình dự đoán đa biến phức tạp hơn. Chúng tôi quan sát thấy rằng các mô hình đơn biến có độ chính xác dự đoán thấp hơn bất kỳ mô hình đa biến nào được thử nghiệm trong nghiên cứu này. Các mô hình cho kết quả gần như nhau ngoại trừ Decision tree cho kết quả thấp. Chúng tôi nhận thấy rằng SVR và RFR là tốt nhất trong số bốn mô hình được triển khai trong việc dự đoán năng suất ngũ cốc bằng cách sử dụng nhóm đặc điểm riêng lẻ (CT, NDVI, hay các đặc điểm nông học) hay thuộc tính kết hợp CT và NDVI trong lúa mì. Mô hình hồi quy Ridge và Lasso lại kết quả tốt nhất khi sử dụng tất cả các đặc điểm để dự đoán.

Nhìn chung, các mô hình cho kết quả dự đoán với mức giải thích ý nghĩa của biến phụ thuộc (năng suất) là gần như nhau (trừ Decision Tree). Với giả thiết không được đặt ra là H0: “Không có sự khác nhau giữa kết quả dự đoán của các mô hình và với kết quả thực tế”. Khi thực hiện phân tích One-way ANOVA của các nhóm kết quả dự đoán theo từng năm và từng nhóm thuộc tính được dùng để dự

đoán, tất cả đều cho giá trị P-value > 0.05 (Bảng 3). Từ đó, chúng tôi kết luận không có sự khác biệt giữa kết quả dự đoán của 4 mô hình là SVR, RFR, Ridge và Lasso. Khi đánh giá kết quả dự đoán với giá trị thực tế thì kết quả cho thấy giữa kết quả được dự đoán và kết quả thực tế không có sự khác biệt.

Yield predictors	Year	P-value
Canopy temperature (CT)	2015-2016	0.96
	2016-2017	0.48
	2017-2018	0.91
	2018-2019	0.66
	2019-2020	0.99
Normalized difference vegetation index (NDVI)	2015-2016	0.84
	2016-2017	0.39
	2017-2018	0.99
	2018-2019	0.99
	2019-2020	0.98
NDVI và CT	2015-2016	0.95
	2016-2017	0.55
	2017-2018	0.99
	2018-2019	0.94
	2019-2020	0.99
Nhóm các đặc điểm nông học	2015-2016	0.99
	2016-2017	0.45
	2017-2018	0.85
	2018-2019	0.19
	2019-2020	0.75
Tất cả thuộc tính	2015-2016	0.97
	2016-2017	0.46
	2017-2018	0.96
	2018-2019	0.40
	2019-2020	0.96

Bảng 3. Giá trị P-value khi thực hiện One-way ANOVA

Tóm lại, so với việc sử dụng các nhóm thuộc tính riêng lẻ để dự đoán thì việc kết hợp các nhóm thuộc tính lại với nhau sẽ giúp cải thiện hơn về độ chính xác của mô hình. Tuy nhiên, việc sử dụng các nhóm thuộc tính riêng biệt cũng giúp phát hiện các nhóm thuộc tính có ảnh hưởng lớn đến năng suất lúa mì (NDVI).

5 Ứng dụng mô hình dự đoán vào thực tế nâng cao năng suất trên diện rộng

Ở một nước đang phát triển như Việt Nam, các cơ sở phân tích kiểu gen vẫn chưa có sẵn. Do đó, các mô hình kiểu hình này có thể áp dụng trực tiếp để thực hiện việc nhân giống được ứng dụng để dự đoán năng suất và dễ kiểm soát hơn là chọn lọc chỉ dựa trên cấu trúc bộ gen. Ngoài ra, nghiên cứu của chúng tôi cung cấp một lượng lớn dữ liệu bằng các công cụ định hình kiểu hình với chi phí thấp. Theo chúng tôi, đây là nghiên cứu được thực hiện bởi các công cụ phân tích với chi phí thấp có thể được hoàn thành bằng máy tính cá nhân (tức là các nguồn sẵn có cho nhiều chương trình nhân giống của quốc gia) mà không cần tiêu tốn lượng lớn tài nguyên cho cả kiểu hình và tính toán. Mô hình này của chúng tôi có thể áp dụng được cho bất kỳ chương trình nhân giống nào, cung cấp lượng lớn dữ liệu về các tính trạng thứ cấp để dự đoán tính trạng chính cần được quan tâm và tăng độ chính xác của quá trình chọn lọc. Bên cạnh các kết quả hiện tại đang được mong đợi thì các nguồn lực như kỹ thuật viên lành nghề, phần cứng và phần mềm hiện tại vẫn chưa ở mức có thể thực hiện được trong

nhiều chương trình nhân giống quốc gia. Hy vọng trong tương lai, các nguồn tài nguyên này sẽ trở nên hợp lý và thân thiện hơn với người dùng.

Trong các thử nghiệm nhân giống này, chúng tôi đã đánh giá sự đa dạng của các mầm nhân giống ưu tú cho thấy nhiều hứa hẹn trong việc xác định các giống được ứng cử có năng suất cao tại Việt Nam. Nhìn chung, có một tỷ lệ cao (24–57%) các dòng được đánh giá vượt trội hơn các giống thông thường như BARI Gom 26 và BARI Gom 30 (Hình bổ sung S2). Ngoài ra, sản lượng trung bình của các mục đã chọn (tức là 10% dòng được đánh giá hàng đầu) mỗi năm cao hơn năng suất của các giống thông thường (Bảng bổ sung S12).

Những quan sát và kết quả chọn lọc này hỗ trợ triển vọng đi lên của việc tiếp tục lựa chọn vật liệu nhân giống chịu nhiệt và phát triển các giống ứng cử mới, ưu việt cho nhiệt độ siêu tối ưu được tìm thấy ở Việt Nam. Việc sử dụng kết hợp các lựa chọn với các công cụ định hình kiểu hình và phương pháp chọn lọc được đề xuất có thể đẩy nhanh hơn nữa việc xác định các giống ứng viên ưu việt này. Mục tiêu của chúng tôi là cải thiện dự đoán năng suất lúa mì bằng cách sử dụng các đặc điểm thứ cấp và mô hình thống kê có thể điều chỉnh các biến tương quan cao (Bảng bổ sung S11). Trong khi chúng tôi tìm hiểu về các mô hình với dữ liệu nông nghiệp và thứ cấp thì chúng tôi thấy được rằng chúng dựa trên cảm biến của NDVI và CT có thể được đo lường dễ dàng hơn so với các đặc điểm nông nghiệp, cái mà đòi hỏi nhiều thời gian hơn và thường không thể đo được cho đến cuối vụ.

Hơn nữa, nghiên cứu của chúng tôi có thể hỗ trợ giá trị của các phép đo cảm biến sinh lý này trong chăn nuôi, dự đoán năng suất chỉ với dữ liệu dựa trên cảm biến cho thấy khả năng cao gần như dự đoán sử dụng được tất cả các đặc điểm tựa nhau.

Các đặc điểm dựa trên cảm biến này dễ dàng đo lường lặp lại trong mùa. Chúng cho phép các nhà lai tạo sử dụng các đặc điểm dựa trên cảm biến để dự đoán năng suất lúa mạch một cách linh hoạt tùy thuộc vào thiết bị hiện có và thực hiện dự đoán năng suất trên các ô quan sát nhỏ. Nếu cơ sở vật chất hạn chế, NDVI có thể được sử dụng thay cho CT để dự đoán năng suất. Bất kể loại chính xác của phép đo dựa trên cảm biến là gì, các nhà lai tạo sẽ có thể tăng khả năng dự đoán bằng cách kết hợp các đặc điểm thứ cấp. Người chăn nuôi có thể sử dụng các phép đo tính trạng thứ cấp, thu được trong mùa sinh trưởng, để tăng độ chính xác của việc lựa chọn trước khi thu hoạch các ô nhằm đảm bảo rằng các ô đó có năng suất cao. Mục đích cuối cùng là cho máy dự đoán ở các mẫu nhỏ sau đó nhân rộng giống cho dự đoán năng suất tốt nhất ra những mẫu đất lớn hơn để đạt sản lượng nhiều hơn.

6 Kết luận

6.1 Kết quả đạt được và đóng góp

Sau khi thực hiện những phương pháp thực nghiệm với những mô hình khác nhau, chúng tôi thấy rằng độ chính xác trong việc dự đoán của các nhóm đặc thứ cấp riêng lẻ rất khác nhau. Ở giai đoạn sinh trưởng sẽ cung cấp các giá trị có sự nhất quán cao nhất, trong khi tại thời điểm điểm đầu hoặc cuối mùa sinh trưởng sẽ có các giá trị không nhất quán, khiến việc sử dụng các nhóm đặc thứ cấp riêng lẻ cho kết quả không cao so với việc sử dụng kết hợp nhóm các thuộc tính nông học. Khi kết hợp các nhóm thuộc tính với nhau, kết quả dự đoán của các mô hình tăng lên đáng kể. Đặc biệt khi kết hợp nhóm các đặc điểm nông học với nhóm biến CT và NDVI, khả năng dự đoán được cải thiện hơn và sự chênh lệch dự đoán giữa các mô hình được giảm thiểu.

6.2 Khó khăn

Trong quá trình tìm kiếm nguồn dữ liệu uy tín và đầy đủ thông tin cần thiết để phát triển dự án gặp nhiều khó khăn, tốn nhiều công sức và thời gian. Bên cạnh đó với vốn kiến thức đòi hỏi nâng cao nên gặp một số khó khăn nhất định trong việc lựa chọn mô hình phù hợp để giải quyết các vấn đề trong thực tiễn. Thời gian nghiên cứu và thực nghiệm cũng bị hạn chế nên chưa thể tìm ra được giải pháp tối ưu để giải quyết những vấn đề vẫn còn tồn tại trong dự án này.

6.3 Hướng phát triển

Thông qua những quy trình phân tích và thiết kế thực nghiệm đã được tiến hành. Chúng tôi hướng đến mục tiêu cải thiện dự đoán năng suất lúa mì tại Việt Nam thông qua những mô hình được ứng dụng tại Bangladesh. Từ những quy trình trên chúng tôi sẽ mở rộng hướng phát triển lên nhiều đối

tượng khác nhau kể cả trong chăn nuôi. Qua đó, giúp cho việc chọn lựa những nguồn giống tốt được cải thiện và đạt hiệu quả cao hơn, góp phần hướng tới một nền nông nghiệp tiên tiến, hiện đại, cho ra nhiều sản phẩm đạt chất lượng cao và mang đến những giá trị nhất định trong xã hội hiện nay.

Tài liệu

- [1] Senthold Asseng, IAN Foster, and Neil C Turner. The impact of temperature variability on wheat yields. *Global Change Biology*, 17(2):997–1012, 2011.
- [2] MA Babar, M Van Ginkel, AR Klatt, B Prasad, and MP Reynolds. The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation. *Euphytica*, 150(1):155–172, 2006.
- [3] M. P. Reynolds I. Amani, R. A. Fischer. Canopy Temperature Depression Association with Yield of Irrigated Spring Wheat Cultivars in a Hot Climate. *J. Agronomy Crop Sci*, 176(2):119–129, 1996.
- [4] Tsutomu Ishimaru, Hideyuki Hirabayashi, Masashi Ida, Toshiyuki Takai, Yumiko A San-Oh, Satoshi Yoshinaga, Ikuo Ando, Tsugufumi Ogawa, and Motohiko Kondo. A genetic resource for early-morning flowering trait of wild rice *oryza officinalis* to mitigate high temperature-induced spikelet sterility at anthesis. *Annals of botany*, 106(3):515–520, 2010.
- [5] MS Lopes, MP Reynolds, MR Jalal-Kamali, M Moussa, Y Feltaous, ISA Tahir, N Barma, Mannes Vargas, Y Mannes, and M Baum. The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. *Field Crops Research*, 128:129–136, 2012.
- [6] Shelley McGuire. Fao, ifad, and wfp. the state of food insecurity in the world 2015: meeting the 2015 international hunger targets: taking stock of uneven progress. rome: Fao, 2015. *Advances in Nutrition*, 6(5):623–624, 2015.
- [7] Mark Tester and Peter Langridge. Breeding technologies to increase crop production in a changing world. *Science*, 327(5967):818–822, 2010.
- [8] Zhu Zhongming, Lu Linong, Yao Xiaona, Zhang Wangqiang, Liu Wei, et al. Climate change 2014: Synthesis report. 2015.