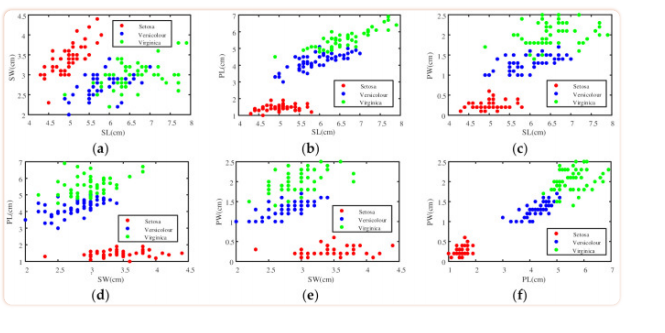
(Vi). Mỗi lớp chứa 50 mẫu và có bốn thuộc tính: chiều dài đài hoa (SL), chiều rộng đài hoa (SW), chiều dài cánh hoa (PL) và chiều rộng cánh hoa (PW). Theo phương pháp đề xuất trong bài báo này, bốn thuộc tính có thể được sử dụng để xác định sáu BPA của một mẫu thử nghiệm. Phân phối mẫu dựa trên hai thuộc tính được thể hiện trong **[Hình 4](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f004)** .



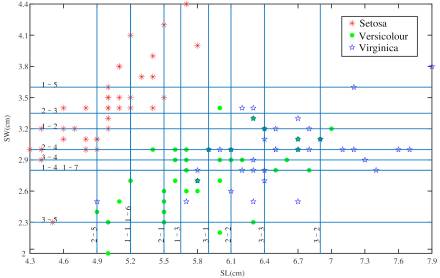
**Hình 4.**

Sáu số liệu phân phối dựa trên hai thuộc tính bất kỳ trong bộ dữ liệu Iris. ( **a** ) Phân phối mẫu dựa trên các thuộc tính SL và SW; ( **b** ) Phân phối mẫu dựa trên các thuộc tính SL và PL; ( **c** ) Phân phối mẫu dựa trên thuộc tính SL và PW; ( **d** ) Phân phối mẫu dựa trên các thuộc tính SW và PL; ( **e** ) Phân phối mẫu dựa trên các thuộc tính SW và PW; ( **f** ) Phân phối mẫu dựa trên các thuộc tính PL và PW.

4.1.1. Xác định BPA của Singleton Proposition

Trong thử nghiệm này, 40 nhóm mẫu được chọn ngẫu nhiên từ mỗi lớp của tập dữ liệu Iris, tổng cộng 120 mẫu được chọn làm tập dữ liệu huấn luyện và 30 mẫu còn lại được sử dụng làm tập dữ liệu kiểm tra. Theo dữ liệu của hai thuộc tính trong tập dữ liệu huấn luyện, một bộ phân loại mạnh được tạo ra, được sử dụng để bỏ phiếu cho các mẫu thử nghiệm nhằm xác định BPA. Các chi tiết của thí nghiệm được hiển thị dưới đây.

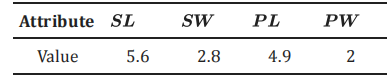
Một mẫu từ bộ thử nghiệm của Virginica được lấy làm ví dụ và dữ liệu được đưa ra trong **[Bảng 2](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t002)** . Vì tập dữ liệu huấn luyện chứa bốn thuộc tính nên chúng ta có thể nhận được sáu bộ phân loại mạnh. Dựa trên các mẫu đào tạo của SL và SW, **[Hình 5](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f005)** cho thấy các quy trình phân loại của phương pháp được đề xuất. Mỗi dòng trong biểu đồ đại diện cho một bộ phân loại yếu và ý nghĩa của số phía trên dòng có thể được mô tả là i−j, trong đó *i* đại diện cho quy trình phân loại hai lần thứ *i* và *j* đại diện cho trình phân loại yếu thứ *j được đào tạo bởi quy trình phân loại hai lần thứ i* .



**Hình 5.** Quá trình sử dụng bộ phân loại yếu để phân loại mẫu dựa trên SL và SW.

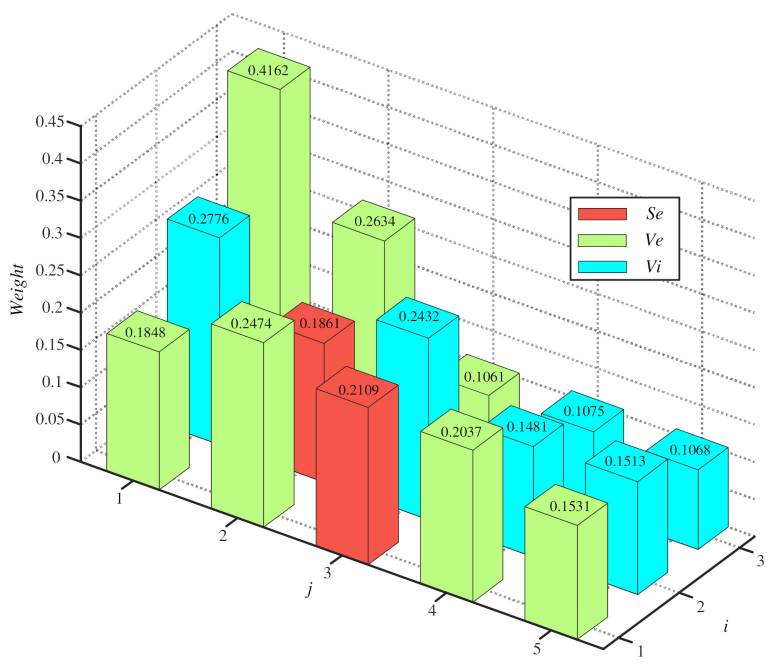
**Bảng 2.**

 Giá trị thuộc tính của mẫu.



Trọng số của các bộ phân loại yếu trong hai cách phân loại khác nhau được thể hiện trong **[Hình 6](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f006)** . Màu sắc khác nhau đại diện cho các lớp khác nhau và chiều cao đại diện cho giá trị của phiếu bầu. Bằng cách sử dụng Công thức (13)–(15) và phiếu bầu của tất cả các bộ phân loại yếu, khối lượng của mẫu này trong SL và SW được đưa ra như sau:

m(Se) = 0.1323, m(V e) = 0.5249, m(V i) = 0.3427

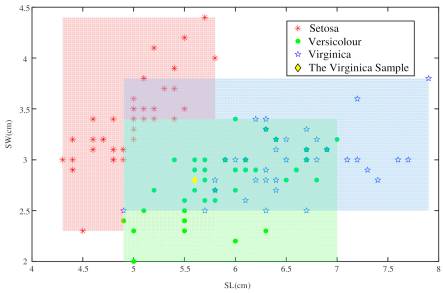


**Hình 6.** Trọng số của tất cả các bộ phân loại yếu dựa trên **[Hình 5](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f005)** .

Theo cách tương tự, chúng ta cũng có thể thu được kết quả biểu quyết của mẫu này ở hai thuộc tính bất kỳ khác. Tuy nhiên, mẫu Virginica đã cho nằm trong các vùng giao nhau của một số phân phối mẫu và nên xem xét tính không chắc chắn của mệnh đề tổng hợp. Do đó, chúng tôi đã sử dụng phương pháp tỷ lệ diện tích trong ví dụ này.

4.1.2. Xác định BPA của Composite Proposition

Như trong **[Hình 7](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f007)** , chúng tôi đã sử dụng các vùng hình chữ nhật với các màu khác nhau để biểu thị phạm vi phân phối của các phân phối mẫu cho các loại SL và SW khác nhau. Mẫu Virginica đã cho nằm trong khu vực giao nhau của ba khu vực phân phối. Do đó, chúng tôi đã tính toán tỷ lệ diện tích của các vùng giao nhau để chuẩn bị cho việc phân bổ lại BPA. Phạm vi và diện tích của tất cả các vùng trong thí nghiệm này được đưa ra trong **[Bảng 3](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t003)** .

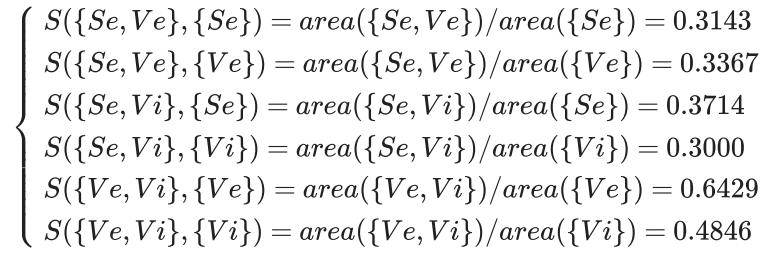


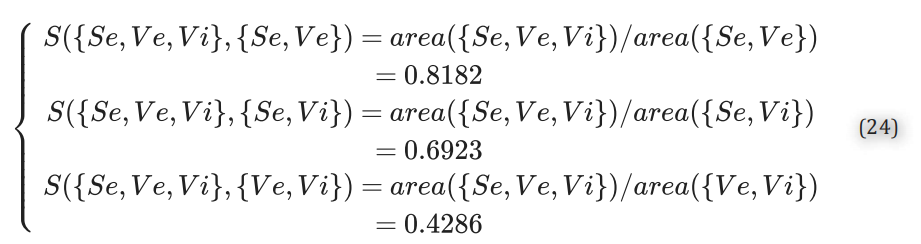
**Hình 7.** Các vùng phân phối mẫu của tập dữ liệu huấn luyện dựa trên SL và SW.

**Bảng 3.** Phạm vi và diện tích của tất cả các vùng (cm).



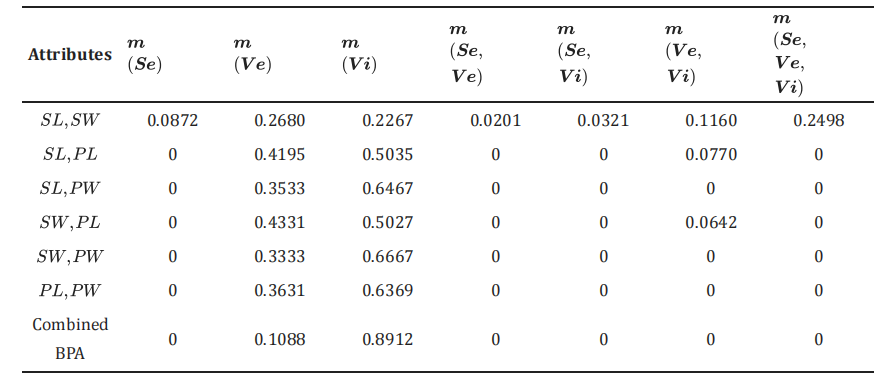
Bằng cách sử dụng các phương trình (17) và (18), chúng tôi nhận được:





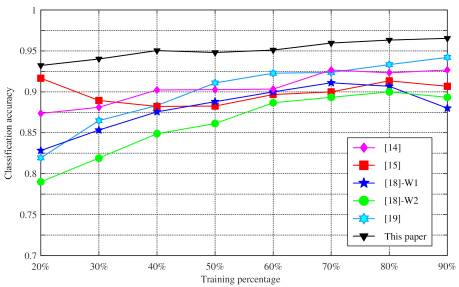
Từ phương trình (23) và (24), chúng ta có thể phân bổ lại kết quả biểu quyết của mẫu này với hai thuộc tính bất kỳ. Tất cả các kết quả biểu quyết được phân bổ lại và kết quả hợp nhất theo quy tắc kết hợp của Dempster được đưa ra trong **[Bảng 4](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t004)** .

**Bảng 4.** Tất cả các BPA được xác định và BPA kết hợp.



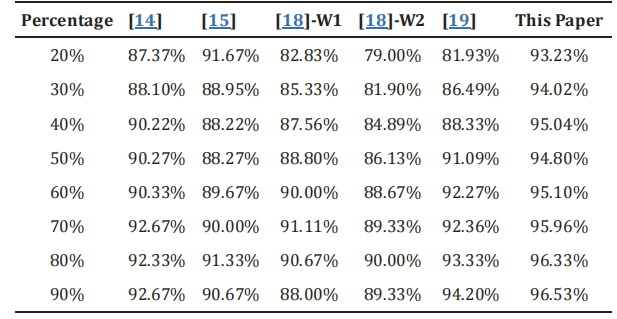
Từ các giá trị của BPA kết hợp trong **[Bảng 4](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t004)** , chúng ta có thể kết luận rằng loại của mẫu thử nghiệm là Virginica, phù hợp với kết quả của bộ dữ liệu Iris.

Để chứng minh tính ưu việt của phương pháp đề xuất, chúng tôi vẫn lấy tập dữ liệu Iris làm ví dụ để so sánh phương pháp đề xuất với phương pháp số khoảng [ **[15](https://www.mdpi.com/1099-4300/23/7/812" \l "B15-entropy-23-00812" \o ")** ] và phương pháp số mờ tam giác tổng quát [ **[14](https://www.mdpi.com/1099-4300/23/7/812" \l "B14-entropy-23-00812" \o ")** , **[18](https://www.mdpi.com/1099-4300/23/7/812" \l "B18-entropy-23-00812" \o ")** , **[19](https://www.mdpi.com/1099-4300/23/7/812" \l "B19-entropy-23-00812" \o ")** ]. Trong thí nghiệm này, số mẫu huấn luyện được chọn ngẫu nhiên từ mỗi lớp là 10, 15, 20, 25, 30, 35, 40 và 45. Các mẫu còn lại được dùng làm tập kiểm tra. Thí nghiệm được lặp lại 100 lần bằng phương pháp Monte Carlo và ghi lại giá trị trung bình của các kết quả thí nghiệm. Như trong **[Hình 8](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f008)** và **[Bảng 5](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t005)** , phương pháp được đề xuất trong bài báo này có độ chính xác phân loại cao hơn.



**Hình 8.** So sánh độ chính xác của bốn phương pháp khác nhau với phương pháp trong bài báo này.

**Bảng 5.** Độ chính xác so với phonon trăm đào tạo đối với các phương pháp khác nhau.

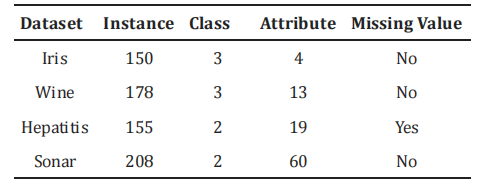


4.2. Các thử nghiệm về việc thay đổi tỷ lệ phonon trăm đào tạo của bốn bộ dữ liệu UCI

Trong phonon này, chúng tôi so sánh phương pháp được đề xuất với sáu bộ phân loại nổi tiếng sau: máy vector hỗ trợ (SVM), SVM với hàm cơ sở xuyên tâm (RBF), mạng RBF (RBFN), perceptron đa lớp (MP), naive Bayesian (NB) ) và Trình học cây quyết định (REPTree).Chúng tôi cũng xem xét Adaboost được đề cập trong **[Phần 2.2](https://www.mdpi.com/1099-4300/23/7/812" \l "sec2dot2-entropy-23-00812)** để minh họa tính hiệu quả của phương pháp được đề xuất. Ngoài bộ dữ liệu Iris, các thử nghiệm trong phonon này đã sử dụng ba bộ dữ liệu khác: Rượu, Viêm gan và Sonar, cũng từ kho lưu trữ máy học của UC Irvine.

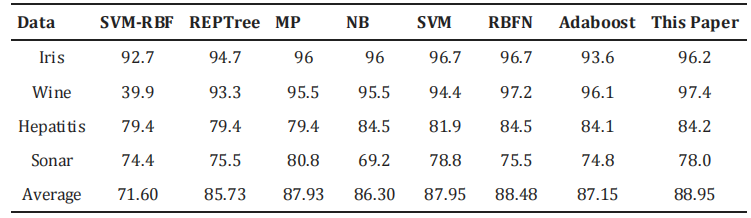
Bộ dữ liệu Rượu bao gồm 13 loại dữ liệu, là kết quả phân tích hóa học của ba loại rượu vang khác nhau được sản xuất trong cùng một vùng của Ý. Bộ dữ liệu Viêm gan chứa 19 thuộc tính, bao gồm thông tin bệnh nhân và kết quả xét nghiệm chức năng gan và dữ liệu của các thuộc tính này được sử dụng để dự đoán liệu bệnh nhân còn sống hay không. Bộ dữ liệu Sonar được sử dụng để dự đoán xem đối tượng mục tiêu là đá hay mỏ dựa trên dữ liệu cường độ do một sonar nhất định trả về từ các góc khác nhau. Thông tin cơ bản về các bộ dữ liệu này được đưa ra trong **[Bảng 6](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t006)** , bao gồm số lượng phiên bản, số lượng lớp, số lượng thuộc tính và tình trạng thiếu giá trị.

**Bảng 6.** Thông tin của các bộ dữ liệu khác nhau.

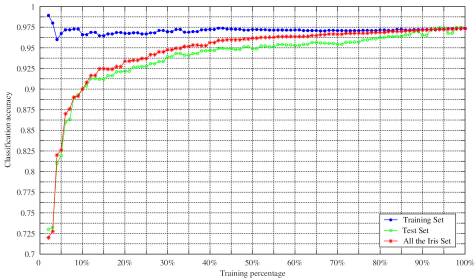


**[Bảng 7](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t007)** cho thấy dữ liệu về độ chính xác phân loại của các phương pháp phân loại khác nhau sử dụng bốn bộ dữ liệu trên. Trong thực nghiệm của mỗi phương pháp, 80% mẫu được chọn ngẫu nhiên làm tập dữ liệu huấn luyện và các mẫu còn lại làm tập dữ liệu kiểm tra. Sau đó, chúng tôi lặp lại thí nghiệm 100 lần và sử dụng độ chính xác trung bình của các thí nghiệm này làm độ chính xác cuối cùng. Bằng cách so sánh độ chính xác trung bình của từng phương pháp, có thể thấy phương pháp được đề xuất trong bài báo này hiệu quả hơn.

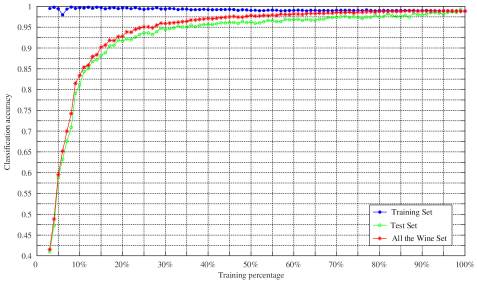
**Bảng 7.** Độ chính xác của các cách phân loại khác nhau (%).



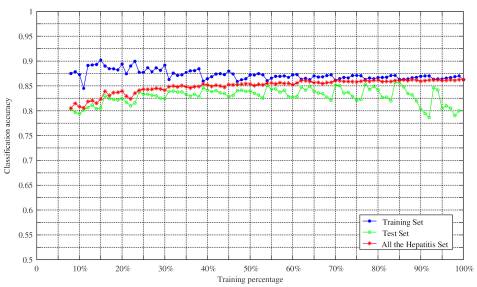
Để xác minh tính hiệu quả của phương pháp đề xuất trong phân loại, phương pháp đề xuất đã được thử nghiệm thêm bằng cách thay đổi tỷ lệ phần trăm đào tạo. N phần trăm mẫu của tập dữ liệu được chọn ngẫu nhiên làm tập huấn luyện và các mẫu còn lại được sử dụng làm tập kiểm tra. Chúng tôi đặt tỷ lệ phần trăm đào tạo của bộ dữ liệu Viêm gan từ 8% đến 98% vì nó chứa các giá trị bị thiếu, trong khi tỷ lệ phần trăm đào tạo của các bộ dữ liệu khác thay đổi từ 2% thành 98% trong quá trình đào tạo. Sau đó, phương pháp Monte Carlo được sử dụng để lặp lại thí nghiệm 100 lần để thu được độ chính xác phân loại trung bình của tập huấn luyện, độ chính xác phân loại trung bình của tập kiểm tra và độ chính xác phân loại trung bình của toàn bộ tập dữ liệu. Kết quả thí nghiệm được thể hiện trong **[Hình 9](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f009)** , **[Hình 10](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f010)** , **[Hình 11](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f011)** và **[Hình 12](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f012)** .



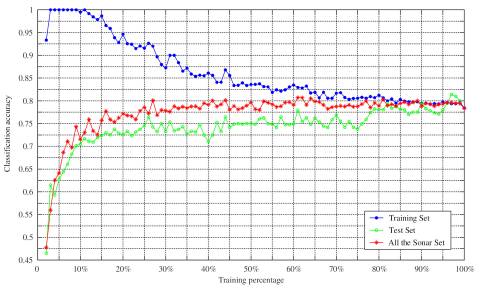
**Hình 9.** Độ chính xác phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Iris.



**Hình 10.** Độ chính xác của phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Rượu.



**Hình 11.** Độ chính xác của phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Viêm gan.

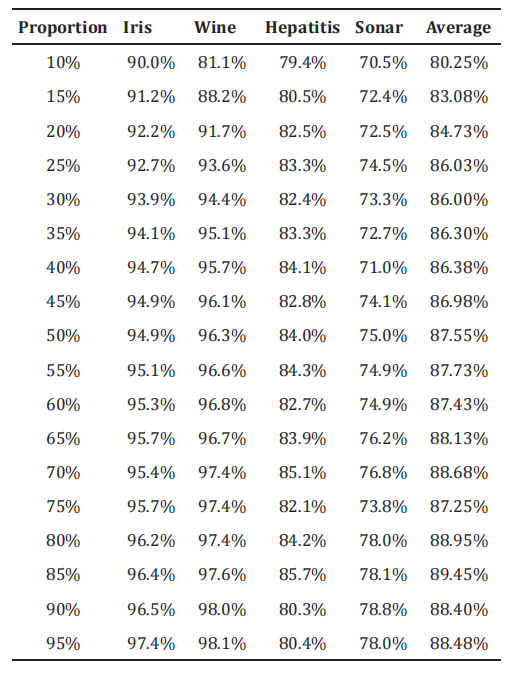


**Hình 12.** Độ chính xác phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Sonar.

Như có thể thấy từ **[Hình 9](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f009)** , **[Hình 10](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f010)** , **[Hình 11](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f011)** và **[Hình 12](https://www.mdpi.com/1099-4300/23/7/812" \l "fig_body_display_entropy-23-00812-f012)**, độ chính xác phân loại trung bình cho bộ dữ liệu Iris, bộ dữ liệu Rượu và bộ dữ liệu Sonar được cải thiện với tỷ lệ phần trăm đào tạo ngày càng tăng. Tuy nhiên, đối với bộ dữ liệu Viêm gan, xu hướng của độ chính xác phân loại trung bình không giống với các bộ dữ liệu khác và giảm khi số lượng mẫu thử nghiệm tăng lên. Điều này là do có 60 thuộc tính trong bộ dữ liệu Viêm gan và diện tích của vùng giao nhau giữa các thuộc tính khác nhau là lớn. Điều này gây khó khăn trong việc phân loại, đó là lý do tại sao hầu hết các thuật toán có độ chính xác phân loại tương tự nhau trong bộ dữ liệu Sonar. Tuy nhiên, độ chính xác phân loại trung bình của bộ dữ liệu Viêm gan vẫn tương đối cao.

Ngoài ra, trong lĩnh vực ứng dụng thực tế, có thể không thu được một số lượng lớn các mẫu đào tạo. Vì vậy, trong trường hợp này, tính khả thi của phương pháp xác định BPA là đặc biệt quan trọng. Như có thể thấy trong **[Bảng 8](https://www.mdpi.com/1099-4300/23/7/812" \l "table_body_display_entropy-23-00812-t008)** , độ chính xác của tập dữ liệu Iris và tập dữ liệu Sonar lần lượt đạt 81,28% và 90,26%, với tập huấn luyện là 10%. Khi tỷ lệ đào tạo là 15%, độ chính xác của bộ dữ liệu Rượu và bộ dữ liệu Viêm gan lần lượt là 88,2% và 80,5%. Điều đáng chú ý là độ chính xác phân loại trung bình của bốn bộ dữ liệu là 80,25% khi tỷ lệ đào tạo là 10%. Kết quả này cho thấy phương pháp trong bài báo này vẫn hợp lý và hiệu quả trong trường hợp số lượng mẫu huấn luyện nhỏ.

**Bảng 8.** Độ chính xác so với phần trăm đào tạo cho các bộ dữ liệu khác nhau.



1. Kết Luận

Trong thuyết Dempster-Shafer (DSET), làm thế nào để xác định một phép gán xác suất cơ bản hợp lý (BPA), một bước quan trọng và đầu tiên, vẫn còn là một vấn đề mở. Trong bài báo này, một phương pháp mới để xác định BPA dựa trên Adaboost được đề xuất. Trong phương pháp được đề xuất này, nhiều bộ phân loại mạnh đã được xây dựng bằng cách sử dụng các mẫu huấn luyện và các trọng số tương ứng đã được ghi lại, được sử dụng để xác định BPA của mệnh đề đơn lẻ. BPA của mệnh đề tổng hợp được xác định bởi tỷ lệ diện tích vùng giao nhau của mệnh đề đơn lẻ. Ưu điểm của phương pháp đề xuất như sau:

1. Phương pháp được đề xuất trong bài báo này là dựa trên dữ liệu để giảm độ không chắc chắn do tính chủ quan gây ra.
2. Không có giả định nào được đưa ra về phân phối dữ liệu huấn luyện, điều này cho thấy có thể áp dụng phương pháp này trong nhiều lĩnh vực khác nhau.
3. Phương pháp tỷ lệ diện tích được đưa ra để cải thiện khả năng xử lý thông tin không chắc chắn của BPA ,đồng thời tăng độ chính xác của phân loại.
4. Đây là một phương pháp đơn giản, thiết thực và nó có thể xác định BPA trong trường hợp có số lượng mẫu đào tạo nhỏ.Sử dụng phương pháp được đề xuất để phân loại bộ dữ liệu Iris, thử nghiệm kết luận rằng tổng tỷ lệ nhận dạng là 96,53% và độ chính xác phân loại trung bình là 90% có thể đạt được khi tỷ lệ đào tạo là 10%.

Khi có quá nhiều thuộc tính của mẫu huấn luyện sẽ gây ra gánh nặng tính toán lớn hơn, đây chính là hạn chế của bài báo này. Là một phần mở rộng của các kết quả của bài báo này, các phương pháp xác định BPA dựa trên phân loại đa thuộc tính sẽ được xem xét trong công việc tương lai của chúng tôi.

## **Sự nhìn nhận**

Các tác giả đánh giá rất cao những gợi ý của các bài đánh giá và sự khuyến khích của biên tập viên.

## **Sự đóng góp của tác giả**

Khái niệm hóa, XW; phương pháp, WF; xác nhận, SY; phân tích chính thức, WF; quản lý dữ liệu, WF; viết—chuẩn bị bản thảo gốc, WF; viết—đánh giá và chỉnh sửa, WF, SY và XW; giám sát, SY; quản lý dự án, XW; mua lại tài trợ, XW Tất cả các tác giả đã đọc và đồng ý với phiên bản đã xuất bản của bản thảo.

## **Kinh phí**

Công trình này được hỗ trợ bởi Quỹ khoa học tự nhiên quốc gia Trung Quốc theo Khoản tài trợ số 61573132, Quỹ nghiên cứu cơ bản tại các tổ chức giáo dục đại học của tỉnh Hắc Long Giang theo Khoản tài trợ số KJCX201809 và RCYJTD201806.

## **Xung đột lợi ích**

Các tác giả tuyên bố không có xung đột lợi ích.