

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ



VƯƠNG CÔNG ĐỊNH

NGHIÊN CỨU THUẬT TOÁN CÂY QUYẾT ĐỊNH SỬ
DỤNG PHÂN TÍCH NGƯỠNG KÉP CHO ỨNG DỤNG
PHÂN LOẠI HÀNH VI CỦA BÒ

LUẬN VĂN THẠC SĨ
CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG

HÀ NỘI – 2017

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ



VƯƠNG CÔNG ĐỊNH

**NGHIÊN CỨU THUẬT TOÁN CÂY QUYẾT ĐỊNH SỬ
DỤNG PHÂN TÍCH NGƯỠNG KÉP CHO ỨNG DỤNG
PHÂN LOẠI HÀNH VI CỦA BÒ**

Ngành: Công Nghệ Kỹ thuật Điện tử, Truyền thông

Chuyên ngành: Kỹ thuật Viễn thông

Mã số: 60520208

**LUẬN VĂN THẠC SĨ
CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG**

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. TRẦN ĐỨC TÂN

HÀ NỘI – 2017

LỜI NÓI ĐẦU

Trong những năm gần đây, việc giám sát hoạt động của gia súc ở trang trại lớn trở nên quan trọng và rất phổ biến. Để nâng cao nguồn lợi kinh tế, người ta càng ngày càng mở rộng chăn nuôi gia súc với số lượng lớn. Nếu sử dụng cách truyền thống là giám sát sức khỏe của từng cá thể gia súc bằng cách quan sát trực tiếp đem áp dụng vào việc quản lý sức khỏe tại nông trại lớn thì đó là việc khó khăn và không hiệu quả. Do vậy việc đề xuất các đề tài nghiên cứu khoa học phục vụ cho phát triển chăn nuôi, kiểm soát gia súc, nhằm nâng cao năng suất, chất lượng con giống; cải tiến kỹ thuật về giống, chăm sóc, nuôi dưỡng, thú y ... là điều hết sức cấp thiết. Việc giám sát hành vi hàng ngày của bò sữa giúp chủ trang trại biết được tình trạng sức khỏe của chúng. Nó giúp nông dân có cái nhìn toàn diện về sức khỏe trong suốt quá trình phát triển và có biện pháp điều chỉnh chế độ chăm sóc phù hợp. Nghiên cứu này tập trung vào dự đoán hành vi của bò bằng việc sử dụng và phân tích dữ liệu từ thiết bị giám sát được đeo ở cổ cho từng cá thể. Thiết bị giám sát này sử dụng cảm biến gia tốc 3 chiều, dữ liệu từ cảm biến này được sử dụng để phân loại các hành vi đơn giản của bò như: ăn, nằm và đứng.

Trên thực tế có nhiều thuật toán được dùng để phân loại hành vi của bò. Trong luận văn này, sử dụng thuật toán cây quyết định để phân loại hành vi đứng, nằm và ăn của bò. Thuật toán tìm ra 2 ngưỡng quyết định một cách đồng thời. Việc tìm ngưỡng đồng thời này giúp nâng cao độ chính xác so với phương pháp [1,9,10] tìm ngưỡng lần lượt. Ngoài ra thuật toán được thực hiện và so sánh trên các bộ dữ liệu lấy cảm biến gia tốc [7] được lấy mẫu với thời gian khác nhau.

LỜI CẢM ƠN

Xuất phát từ những ý nghĩa thực tế của việc quản lý, phân loại hành vi trên số lượng lớn của bò, luận văn là kết quả của quá trình nghiên cứu lý luận và thực tiễn của cá nhân tác giả dựa trên sự chỉ bảo, hướng dẫn tận tình của PGS.TS. Trần Đức Tân. Thầy đã không quản khó khăn, thời gian, công sức để giúp tôi hoàn thành luận văn này, nhân đây, tôi xin gửi lời cảm ơn sâu sắc tới PGS.TS Trần Đức Tân. Được thầy hướng dẫn là một niềm hạnh phúc đối với cá nhân tác giả, bởi lẽ thầy là một nhà giáo trẻ, mẫu mực, say mê nghiên cứu khoa học, là người có phương pháp nghiên cứu, có nhiều đóng góp cho sự nghiệp nghiên cứu khoa học – là hình mẫu cho chúng tôi noi theo.

Tôi cũng xin gửi lời cảm ơn đến các thầy, cô giáo và bạn bè trong lớp K21 Kỹ thuật viễn thông, Khoa Điện Tử – Viễn Thông, Trường Đại Học Công Nghệ, Đại Học Quốc Gia Hà Nội đã có những nhận xét, góp ý cho luận văn này của tôi.

Cuối cùng tôi xin gửi lời cảm ơn đến gia đình tôi, cơ quan tôi đang công tác, những người đã tạo điều kiện cho tôi học tập và nghiên cứu. Gia đình là động lực cho tôi vượt qua những thử thách, luôn luôn ủng hộ và động viên tôi hoàn thành luận văn này.

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này là sản phẩm của quá trình nghiên cứu, tìm hiểu của cá nhân dưới sự hướng dẫn và chỉ bảo của các thầy hướng dẫn, thầy cô trong bộ môn, trong khoa và các bạn bè. Tôi không sao chép các tài liệu hay các công trình nghiên cứu của người khác để làm luận văn này.

Nếu vi phạm, tôi xin chịu mọi trách nhiệm.

Vương Công Định

MỤC LỤC

LỜI NÓI ĐẦU.....	i
LỜI CAM ĐOAN	iii
MỤC LỤC	iv
DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT.....	v
DANH MỤC CÁC BẢNG.....	vii
DANH MỤC CÁC HÌNH VẼ.....	viii
CHƯƠNG I. GIỚI THIỆU TỔNG QUAN	1
1.1. Đặt vấn đề.....	1
1.2. Cấu trúc chung của hệ thống.....	1
1.3. Nội dung thực hiện	3
1.4. Tổ chức luận văn.....	3
CHƯƠNG II. THỰC HIỆN THUẬT TOÁN.....	4
2.1. Các đặc tính đặc trưng.....	4
2.2. Lưu đồ thuật toán.....	5
2.3. Hiệu năng hệ thống.....	7
2.4. Thực hiện thuật toán.....	8
2.5. Kết quả khi thực hiện thuật toán.....	9
2.5.1. Kịch bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 10 phút/lần	9
2.5.2. Kịch bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 05 phút/lần	13
2.5.3. Kịch bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 01 phút/lần	17
2.6. Nhận xét.....	21
CHƯƠNG III. ĐÁNH GIÁ THUẬT TOÁN.....	22
3.1. Hiệu năng của thuật toán ở các tần số lấy mẫu khác nhau	22
3.1.1. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 10 phút/lần	22
3.1.2. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 05 phút/lần	24
3.1.3. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 01 phút/lần	27
3.2. Hiệu năng của thuật toán so với phương pháp ROC	31
KẾT LUẬN	35
TÀI LIỆU THAM KHẢO.....	36

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Ký hiệu	Tiếng Anh	Tiếng Việt	Đơn vị
g	Gravity of Earth	Gia tốc trọng trường ($1g = 9,8 \text{ m/s}^2$)	m/s^2
DBA	Dynamic Body Acceleration	Gia tốc chuyển động	m/s^2
ODBA	Overall Dynamic Body Acceleration	Tổng gia tốc chuyển động toàn thân	m/s^2
VeDBA	Vectorial Dynamic Body Acceleration	Véc tơ tổng hợp gia tốc chuyển động toàn thân	m/s^2
SCAY	Static Component of the Acceleration in the Y-axis	Gia tốc tĩnh theo phương Y, được chuẩn hóa theo g.	m/s^2
ROC	Receiver Operating Characteristic	Vẽ đường cong đặc trưng	
Cont	Contour plot	Vẽ đường viền	
SVM	supported vector machine	Máy véc tơ hỗ trợ	
k-mean		K phân cụm (là một thuật toán)	
TN	True negative	Âm tính thật	
TP	True positive	Dương tính thật	
FN	False negative	Âm tính giả	
FP	False positive	Dương tính giả	
Sen	Sensitivity	Độ nhạy	%

Pre	Precision	Độ chính xác	%
Spe	Specificity	Độ chỉ rõ	%
Max	Maximum	Giá trị lớn nhất	
TPR	True positive rate	Tỉ lệ độ nhạy	%
FPR	False positive rate	Tỉ lệ báo động giả	%

DANH MỤC CÁC BẢNG

Bảng 2.1. Các giá trị ngưỡng khi dữ liệu lấy mẫu 10 phút/lần	12
Bảng 2.2. Các giá trị ngưỡng khi dữ liệu lấy mẫu 05 phút/lần	16
Bảng 2.3. Các giá trị ngưỡng khi dữ liệu lấy mẫu 01 phút/lần	20
Bảng 3.1. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 10 phút/lần	22
Bảng 3.2. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 10 phút/lần	23
Bảng 3.3. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 10 phút/lần	24
Bảng 3.4. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 05 phút/lần	25
Bảng 3.5. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 05 phút/lần	26
Bảng 3.6. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 05 phút/lần	26
Bảng 3.7. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 01 phút/lần	27
Bảng 3.8. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 10 phút/lần	28
Bảng 3.9. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 01 phút/lần	29
Bảng 3.10. Hiệu năng của hệ thống với các chu kỳ lấy mẫu khác nhau	30
Bảng 3.11. Ví dụ so sánh giá trị ngưỡng khi thực hiện 2 thuật toán	32
Bảng 3.12. Hiệu năng của hệ thống khi so sánh 2 thuật toán	33

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Hệ thống quản lý chăn nuôi bò.....	2
Hình 1.2. Vị trí gắn cảm biến trên cổ bò [1]	2
Hình 2.1. Định hướng của cảm biến gắn trên cổ bò, thay đổi khi đứng (a) và khi nằm (b) [1].....	5
Hình 2.2. Lưu đồ thuật toán xác định hành vi của bò	6
Hình 2.3. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 10 phút/lần	9
Hình 2.4. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 10 phút/lần.....	10
Hình 2.5. Sự thay đổi giá trị ngưỡng theo độ nhảy, dữ liệu lấy mẫu 10 phút/lần	10
Hình 2.6. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 10 phút/lần .	11
Hình 2.7. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 10 phút/lần.....	11
Hình 2.8. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 10 phút/lần	12
Hình 2.9. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 05 phút/lần	13
Hình 2.10. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 05 phút/lần.....	14
Hình 2.11. Sự thay đổi giá trị ngưỡng theo độ nhảy, dữ liệu lấy mẫu 05 phút/lần	14
Hình 2.12. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 05 phút/lần	15
Hình 2.13. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 05 phút/lần.....	15
Hình 2.14. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 05 phút/lần	16
Hình 2.15. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 01 phút/lần	17
Hình 2.16. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 01 phút/lần.....	18
Hình 2.17. Sự thay đổi giá trị ngưỡng theo độ nhảy, dữ liệu lấy mẫu 01 phút/lần	18
Hình 2.18. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 01 phút/lần	19
Hình 2.19. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 01 phút/lần.....	19
Hình 2.20. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 01 phút/lần	20
Hình 3.1. Đường cong ROC xác định ngưỡng A theo độ nhảy tốt nhất, dữ liệu lấy mẫu 10 phút/lần [1]	31
Hình 3.2. Đường cong ROC xác định ngưỡng B theo độ nhảy tốt nhất, dữ liệu lấy mẫu 10 phút/lần [1]	32

CHƯƠNG I. GIỚI THIỆU TỔNG QUAN

1.1. Đặt vấn đề

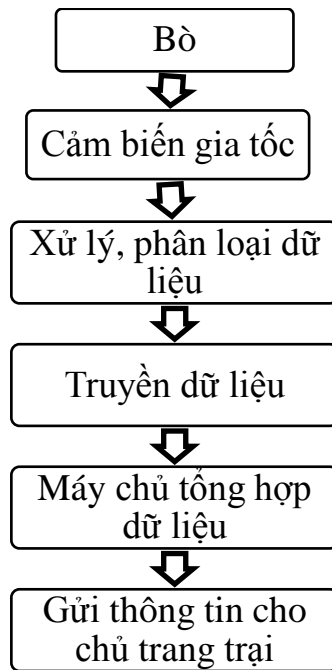
Việc nâng cao chất lượng chăn nuôi bò sữa ở các nông trại lớn là điều quan trọng và cần thiết. Hơn nữa ở Việt nam, có điều kiện thuận lợi về khí hậu để phát triển chăn nuôi bò sữa. Nên hiện nay, có một số công ty chế biến sữa trong nước đã tiên phong áp dụng các công nghệ tiên tiến để nâng cao năng suất và chất lượng sữa. Có thể kể đến như: TH True milk, Vinamilk ... Vì thế, nếu sử dụng việc quản lý từng cá thể gia súc ở quy mô nhỏ để áp dụng trên quy mô trang trại lớn sẽ gặp nhiều khó khăn. Do người chăn nuôi luôn có nhu cầu kiểm soát hành vi hoạt động của từng con bò càng thường xuyên càng tốt để có những hành động chăm sóc kịp thời. Trong khi ở những trang trại chăn nuôi có diện tích lớn, số lượng bò rất nhiều và bò là di chuyển, vì vậy không thể quan sát từng cá thể bò sữa bằng mắt thường. Do vậy cần có các hệ thống tự động được thiết kế để theo dõi giám sát và phân loại hành vi của bò. Cảm biến quán tính có rất nhiều trong thực tế [12-15]. Trong ứng dụng giám sát bò, cảm biến gia tốc được gắn trên cổ bò [2] và sử dụng dữ liệu từ cảm biến để phân loại hành vi. Có nhiều phương pháp được sử dụng để phân loại hành vi của bò một cách tự động, chủ yếu dựa trên các thuật toán học máy như: cây quyết định, k-mean, máy vector hỗ trợ (SVM – supported vector machine).

Tuy nhiên, trong thực tế có nhiều hệ thống thích hợp cho việc phân loại một hoặc hai hành vi của bò cùng một lúc. Trong các hệ thống này, loại cảm biến gia tốc được sử dụng khá phổ biến để theo dõi hành vi và sức khỏe của động vật. Như trong báo cáo [1], tác giả đã sử dụng máy đo gia tốc 3 chiều để tự động theo dõi và phân biệt các hành vi của nhiều động vật, đặc biệt đối với gia súc thông qua các chuyển động của cổ bò, mà trên cổ bò có gắn cảm biến gia tốc 3 chiều.

Trong nội dung luận văn này, tập trung nghiên cứu thuật toán phân loại hành vi của bò dựa trên bộ dữ liệu từ cảm biến gia tốc đã có [7]. Thuật toán sẽ đưa ra cách tìm ngưỡng cho VeDBA (Vectorial Dynamic Body Acceleration) và SCAY (Static Component of the Acceleration in the Y-axis) sử dụng đồ thị Contour, qua đó để tìm 2 ngưỡng tốt nhất một cách đồng thời, mà không sử dụng thuật toán ROC (Receiver Operating Characteristic) để tìm lần lượt 2 ngưỡng này. Tiếp theo là đưa ra đánh giá chất lượng khi sử dụng đồ thị Contour và khi sử dụng ROC. Ngoài ra, còn đánh giá việc sử dụng đồ thị Contour với bộ 3 dữ liệu với tần số cập nhật dữ liệu khác nhau, qua đó lựa chọn bộ dữ liệu phù hợp áp dụng trong thực tế.

1.2. Cấu trúc chung của hệ thống

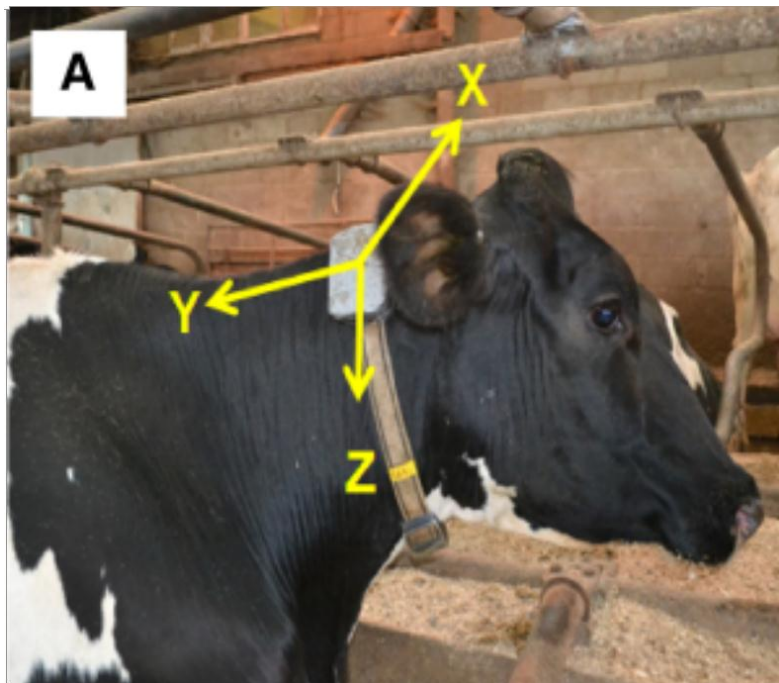
Với mục tiêu là giám sát các hoạt động của bò. Để từ đó thông báo khi có sự kiện và phát hiện một cách kịp thời, từ đó giúp tăng năng suất chăn nuôi. Ta có sơ đồ một hệ thống quản lý bò như sau:



Hình 1.1. Hệ thống quản lý chăn nuôi bò

Cấu trúc của hệ thống bao gồm:

- Đối tượng cần theo dõi là bò.
- Cảm biến gia tốc được gắn trên cơ thể bò [11]. Cụ thể trong nghiên cứu này sử dụng cảm biến gia tốc 3 chiều và được gắn trên cổ bò. Việc gắn cảm biến ở cổ, giúp cho hướng của cảm biến gia tốc cố định và chắc chắn.



Hình 1.2. Vị trí gắn cảm biến trên cổ bò [1]

- Khối xử lý, phân loại dữ liệu: lấy giữ liệu từ cảm biến gia tốc, sau đó xử lý để xác định hoạt động của bò là: ăn, nằm, đứng ...
- Bộ thu phát không dây: giúp cho truyền dữ liệu hành vi về máy chủ.

- Máy tính chủ tổng hợp dữ liệu của nhiều cá thể bò, để thấy được thời gian phân bố các trạng thái trong ngày cũng như tình hình sức khỏe của bò. Sau đó gửi các thông tin này đến chủ trang trại để biết được tình trạng của gia súc và có các hành động phù hợp. Phát hiện con nào bị thương, có dấu hiệu khác thường để kịp thời chữa trị, giảm tiêu hao năng suất.

Trong luận văn này sẽ tập trung vào khối xử lý dữ liệu thu được từ cảm biến để phân loại hành vi, hoạt động của bò.

1.3. Nội dung thực hiện

Việc phân loại hành vi của bò sử dụng thuật toán cây quyết định dựa trên bộ dữ liệu đã có [7]. Thuật toán sử dụng 2 tham số ngưỡng được sử dụng để phân loại đó là: VeDBA, SCAY. Do vậy nội dung thực hiện của luận văn như sau:

- Tìm ngưỡng cho VeDBA và SCAY sử dụng đồ thị Contour (không sử dụng ROC - tìm lần lượt 2 ngưỡng này) để tìm 2 ngưỡng tốt nhất một cách đồng thời.
- Đưa ra đánh giá chất lượng khi sử dụng Contour và khi sử dụng ROC
- Đánh giá việc sử dụng đồ thị Contour với bộ 3 dữ liệu với tần số cập nhật dữ liệu khác nhau.

1.4. Tổ chức luận văn

Phần còn lại của luận văn này được tổ chức như sau: Chương 2 trình bày về thực hiện thuật toán, chỉ ra được thuật toán thực hiện và kết quả tương ứng. Chương 3 nói về đánh giá hiệu năng của thuật toán với các bộ dữ liệu lấy mẫu khác nhau và so sánh với phương pháp trước ROC [1,9,10]. Cuối cùng là kết luận và hướng nghiên cứu tiếp theo.

CHƯƠNG II. THỰC HIỆN THUẬT TOÁN

2.1. Các đặc tính đặc trưng

Trong bài toán này, để phân loại hành vi của bò, ta sử dụng dữ liệu gia tốc thu được từ cảm biến gia tốc 3 chiều. Sau đó tính toán ra hai thành phần của dữ liệu gia tốc: thành phần tĩnh và thành phần động.

Thành phần động được gây ra bởi sự chuyển động của vật mang cảm biến. Tổng gia tốc chuyển động toàn thân – ODBA (Overall Dynamic Body Acceleration) [3,4] và vector tổng hợp gia tốc chuyển động toàn thân (VeDBA) có thể đại diện phần động cho tập giá trị gia tốc của đối tượng [5,6]. Do đó người ta sử dụng vector gia tốc chuyển động toàn thân (VeDBA) để phân biệt giữa hành vi với hoạt động cao (như ăn) và hoạt động thấp (như đứng hoặc nằm) [2].

Để tính toán được ra giá trị VeDBA, trước hết ta tính gia tốc động DBA (Dynamic Body Acceleration) theo từng trục X, Y, Z. DBA được tính như sau:

$$DBA_i = A_{it} = |A_{it}^* - \mu_{it}| \quad (2.1)$$

A_i : là gia tốc động tương ứng là A_x, A_y, A_z

A_i^* : là gia tốc thu được tại thời điểm lấy mẫu

μ_{it} : là gia tốc tĩnh (tính được bởi lấy trung bình một số mẫu)

μ_{it} được tính theo công thức sau:

$$\mu_{it} = \sum_{t-\frac{win_size}{2}}^{t+\frac{win_size}{2}} \frac{A_{it}^*}{win_size} \quad \text{với } i \text{ tương ứng với trục X, Y và Z.} \quad (2.2)$$

win_size là độ rộng cửa sổ thời gian lấy mẫu.

Giá trị của DBA được sử dụng để tính toán giá trị ODBA và giá trị VeDBA như công thức dưới đây:

$$ODBA = |A_x + A_y + A_z| \quad (2.3)$$

$$VeDBA = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (2.4)$$

VeDBA là giá trị đại diện cho tổng hợp gia tốc chuyển động toàn thân mà trong các cảm biến ngày nay người ta hay dùng, đơn vị của VeDBA là g (gia tốc trọng trường).

Thành phần tĩnh của gia tốc được gây ra bởi sự định hướng các trục của cảm biến so với trường hấp dẫn của trái đất và có thể được tính như trung bình động để xác định tư thế cơ thể [1,2]. Cụ thể trong bài toán này, SCAY được sử dụng để xác định sự thay đổi gia tốc trọng trường Y.

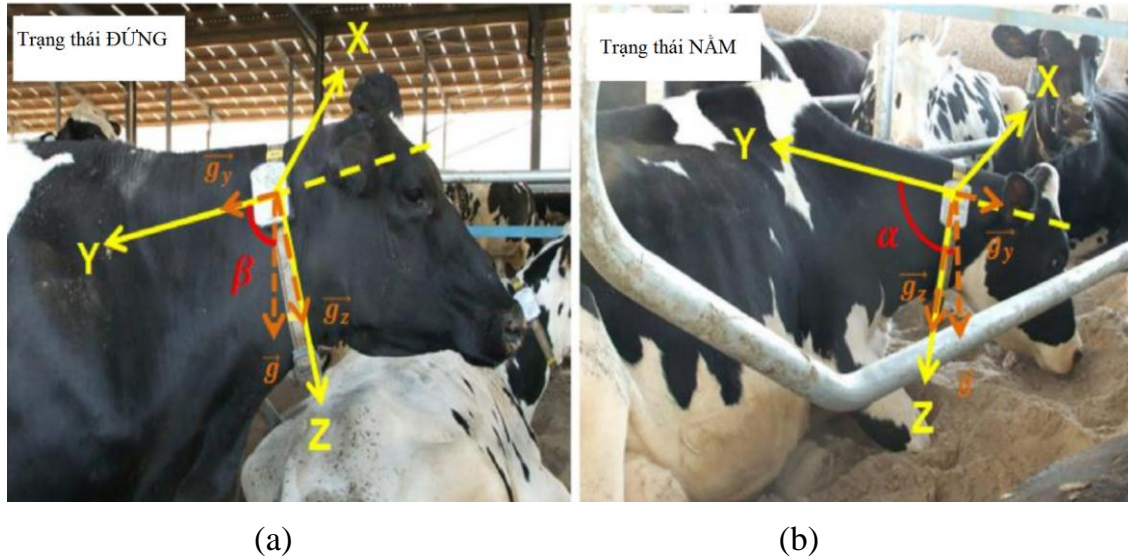
$$SCAY: \vec{g}_y = g \times \cos(180 - \beta) \quad (2.5)$$

Trong đó: β là góc thể hiện sự tương đối giữa trục Y với phương ngang.

\vec{g}_y : véc tơ gia tốc theo trục Y

g : gia tốc trọng trường

Hình 2.1 minh họa cho việc sử dụng SCAY để phân loại trạng thái đứng và nằm của bò.



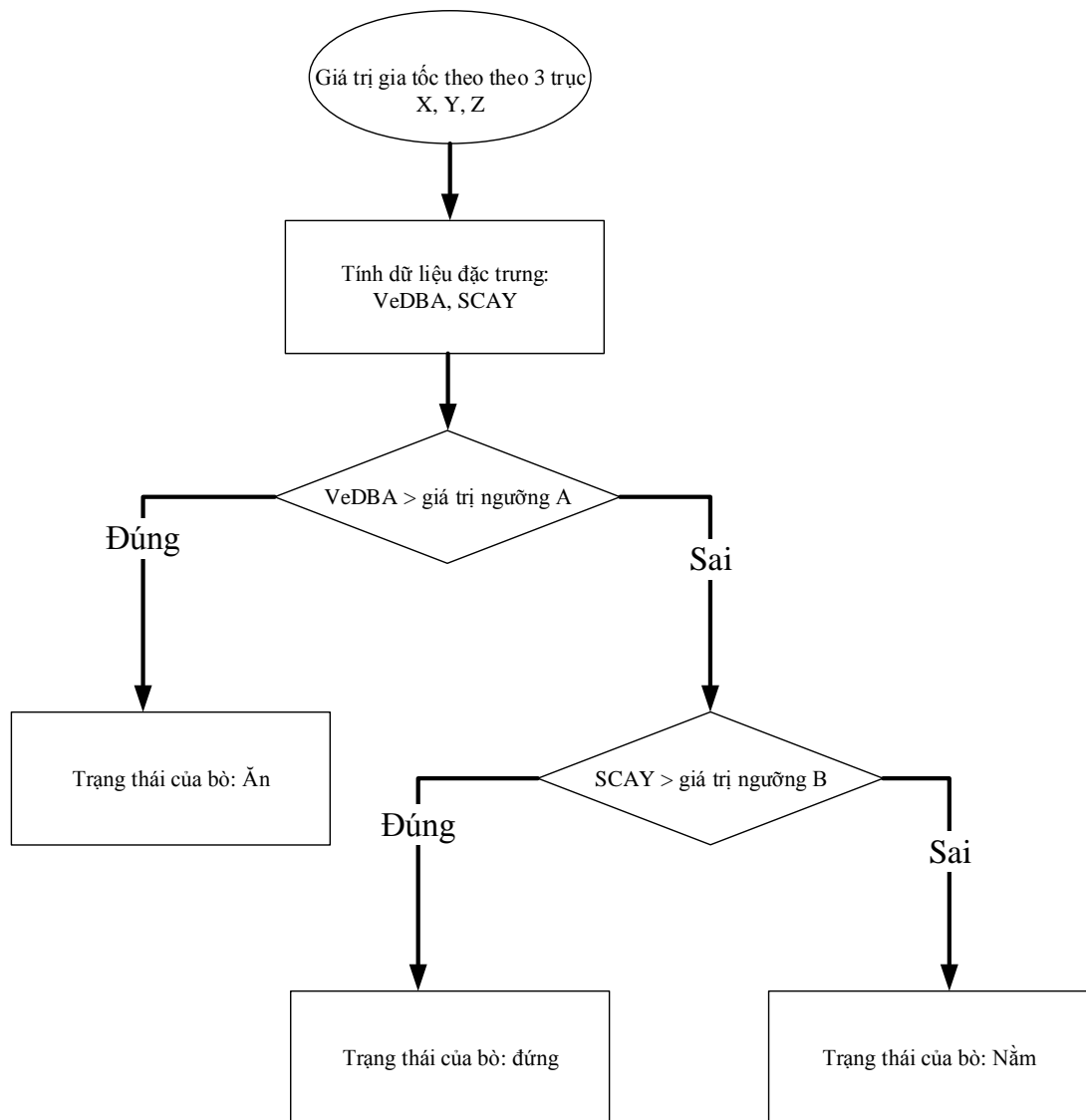
Hình 2.1. Định hướng của cảm biến gắn trên cổ bò, thay đổi khi đứng (a) và khi nằm (b) [1]

Thành phần tĩnh này được sử dụng để phân loại hoạt động thấp của bò, là hành vi nằm hoặc đứng.

Do vậy để phân loại trạng thái của bò, chúng ta cần tính toán ra các tham số đặc trưng như VeDBA và SCAY tại mỗi thời điểm, sau đó sử dụng chúng so sánh với hai giá trị ngưỡng trong thuật toán cây quyết định.

2.2. Lưu đồ thuật toán

Lưu đồ của thuật toán thực hiện việc xác định hành vi của bò như dưới đây.



Hình 2.2. Lưu đồ thuật toán xác định hành vi của bò

Ban đầu, để nhận biết được hành vi của bò, các hoạt động của bò được ghi lại bằng cách quan sát trực tiếp: bằng mắt thường hoặc qua camera, cho thấy:

- Trạng thái ăn: Con bò phải nằm ở khu ăn và con bò đang ăn thực phẩm. Con bò thường lắc nhẹ và cúi đầu.
- Trạng thái nằm: Con bò đang nằm trong chuồng.
- Trạng thái đứng: Con bò đứng trên 4 chân của nó.

Dựa trên bộ dữ liệu thực tế này, thuật toán sẽ phân loại, tính toán ra được các giá trị ngưỡng A cho việc so sánh với VeDBA, ngưỡng B cho việc so sánh với SCAY.

Sau khi có dữ liệu ngưỡng A, ngưỡng B. Thuật toán phân loại hành vi của bò được thực hiện tự động như sau:

- Từ cảm biến ta thu được dữ liệu của gia tốc theo 3 trục A_x , A_y , A_z .
- Sau đó ta tính toán được 2 tham số đặc trưng là VeDBA theo công thức (2.4) và SCAY theo công thức (2.5) từ dữ liệu cảm biến gia tốc 3 chiều.

- Tiếp theo đem so sánh VeDBA vừa tìm được với giá trị ngưỡng A. Nếu giá trị VeDBA lớn hơn ngưỡng A, thì trạng thái của bò là ăn. Ngược lại thì trạng thái của bò là nằm hoặc đứng.

- Để phân loại trạng thái nằm hoặc đứng, ta đem so sánh SCAY vừa tìm được với giá trị ngưỡng B. Nếu giá trị SCAY lớn hơn ngưỡng B, thì trạng thái của bò là đứng. Ngược lại thì trạng thái của bò là nằm.

2.3. Hiệu năng hệ thống

Có nhiều tham số có thể sử dụng để đánh giá hiệu năng của một thuật toán. Tùy vào mục đích khác nhau mà có những tham số khác nhau. Hiệu năng của hệ thống bị ảnh hưởng rất nhiều bởi sự lựa chọn các giá trị ngưỡng này. Trong luận văn này đưa ra 3 tham số về hiệu năng hệ thống: độ nhạy, độ chính xác và độ chỉ rõ [8]. Giá trị ngưỡng tìm được phụ thuộc vào tham số hiệu năng mà ta lựa chọn.

$$\text{Độ nhạy:} \quad \text{Sen} = \frac{TP}{TP+FN} \quad (2.6)$$

$$\text{Độ chính xác:} \quad \text{Pre} = \frac{TP}{TP+FP} \quad (2.7)$$

$$\text{Độ chỉ rõ:} \quad \text{Spe} = \frac{TN}{TN+FP} \quad (2.7)$$

TP: (true positive) những trường hợp mà trạng thái thực tế quan sát được và phân loại đúng theo thuật toán.

FP: (False positive) những trường hợp mà trạng thái được phân loại bởi thuật toán nhưng không được quan sát trong thực tế.

FN: (False negative) những trường hợp mà trạng thái được quan sát trong thực tế nhưng không phân loại theo thuật toán.

TN: (True negative) những trường hợp mà trạng thái không được phân loại theo thuật toán và cũng không quan sát được thấy trong thực tế.

Độ nhạy: Sen (sensitivity) được tính theo công thức (2.6), đặc trưng cho khả năng có thể phân loại được của thuật toán. Độ nhạy càng cao, chứng tỏ khả năng phân loại các trường hợp của thuật toán là tốt.

Độ chính xác: Pre (precision) được tính theo công thức (2.7), đặc trưng cho khả năng phân loại đúng của thuật toán. Độ chính xác càng lớn, thì khả năng phân loại của thuật toán càng chính xác.

Độ chỉ rõ: Spe (specificity) được tính theo công thức (2.8), đặc trưng cho khả năng chỉ ra chính xác bao nhiêu % khả năng không phải sự kiện cần phân loại. Độ chính xác càng cao, chứng tỏ khả năng chỉ ra sự kiện không cần phân loại càng lớn. Ví dụ: có 100 sự kiện thực tế ko phải là X (với X là sự kiện cần phân loại), nhưng giải thuật chỉ chỉ được chính xác 90 sự kiện trong đó không phải là X=> Độ chỉ rõ=90%.

Tùy thuộc bài toán yêu cầu, mà trong thực tế người ta sẽ tìm ngưỡng theo tham số hiệu năng cụ thể. Trong thuật toán này, sẽ xét tính toán ngưỡng theo lần lượt sao

cho hiệu năng là lớn nhất theo độ nhạy, độ chính xác, độ chỉ rõ và tính lớn nhất cho cả trung bình 3 tham số hiệu năng này.

2.4. Thực hiện thuật toán

Thuật toán được đề xuất, xác định được 2 ngưỡng A và B một cách đồng thời. Trong đó, ngưỡng A là giá trị VeDBA được dùng để phân loại giữa trạng thái có hoạt động cao (ăn) và trạng thái có hoạt động thấp (nằm và đứng). Ngưỡng B là giá trị SCAY được dùng để phân loại giữa đứng và nằm. Như phần trên, có 3 tham số về hiệu năng của hệ thống được sử dụng khi lựa chọn các giá trị ngưỡng này, đó là: độ nhạy, độ chính xác, độ chỉ rõ.

Thuật toán sử dụng đồ thị Contour để tìm 2 ngưỡng A và B một cách đồng thời. Bộ dữ liệu VeDBA và SCAY được sử dụng để tìm ngưỡng có đặc điểm sau [7]:

- Được sử dụng từ nguồn chia sẻ trực tuyến:

https://static-content.springer.com/esm/art%3A10.1186%2Fs40317-015-0045-8/MediaObjects/40317_2015_45_MOESM2_ESM.txt

- Bộ dữ liệu gồm 3 tham số: VeDBA, SCAY và 3 trạng thái hành vi thực tế của bò: ăn, nằm và đứng.

- Có 3 bộ dữ liệu được lấy mẫu lần lượt: 1 phút, 5 phút, 10 phút. Cụ thể ở bộ dữ liệu lấy mẫu 1 phút, khoảng cách thời gian giữa 2 mẫu dữ liệu là 1 phút. Ở bộ dữ liệu lấy mẫu 5 phút, khoảng cách thời gian giữa 2 mẫu dữ liệu là 5 phút. Ở bộ dữ liệu lấy mẫu 10 phút, khoảng cách thời gian giữa 2 mẫu dữ liệu là 10 phút.

Thuật toán được thực hiện như sau:

Thuật toán: Contour Threshold

- 1: Nhập bộ dữ liệu với tần số lấy mẫu là: 1 phút, 5 phút, hoặc 10 phút.
- 2: Phân loại từng dữ liệu về VeDBA, bộ dữ liệu về SCAY và bộ dữ liệu về trạng thái thực tế của bò.
- 3: Tìm giá trị lớn nhất, nhỏ nhất của dữ liệu VeDBA.
- 4: Khởi tạo n giá trị ngưỡng A liên tục, cách đều trong khoảng giá trị lớn nhất và nhỏ nhất của VeDBA.
- 5: Tìm giá trị lớn nhất, nhỏ nhất của tham số SCAY.
- 6: Khởi tạo n giá trị ngưỡng B liên tục, cách đều trong khoảng giá trị lớn nhất và nhỏ nhất của SCAY.
- 7: Khởi tạo bộ giá trị TP, TN, FP, FN cho từng trạng thái của bò: ăn, nằm và đứng.
- 7: **for** $i = 1$ đến giá trị n, **do**
- 8: Tại mỗi vòng lặp của i, so sánh giá trị của VeDBA với ngưỡng A, và so sánh SCAY với ngưỡng B, để đưa ra kết luận của thuật toán.

9: Mỗi kết luận của thuật toán ta đem so sánh với kết quả thực tế trạng thái bò, sau đó tăng các biến TP, TN, FP, FN phù hợp.

10: **end for**

11: Khi có bộ dữ liệu TP, TN, FP, FN ta sẽ tìm ra được độ nhạy, độ chính xác, độ chỉ rõ.

12: Dựa vào tiêu chí cần đạt được, ta sẽ tìm được cặp giá trị ngưỡng A và B sao cho độ nhạy lớn nhất hoặc độ chính xác lớn nhất hoặc độ chỉ rõ lớn nhất hoặc cân bằng được cả 3 tiêu chí này.

2.5. Kết quả khi thực hiện thuật toán

2.5.1. Kịch bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 10 phút/lần

Tham số đầu vào:

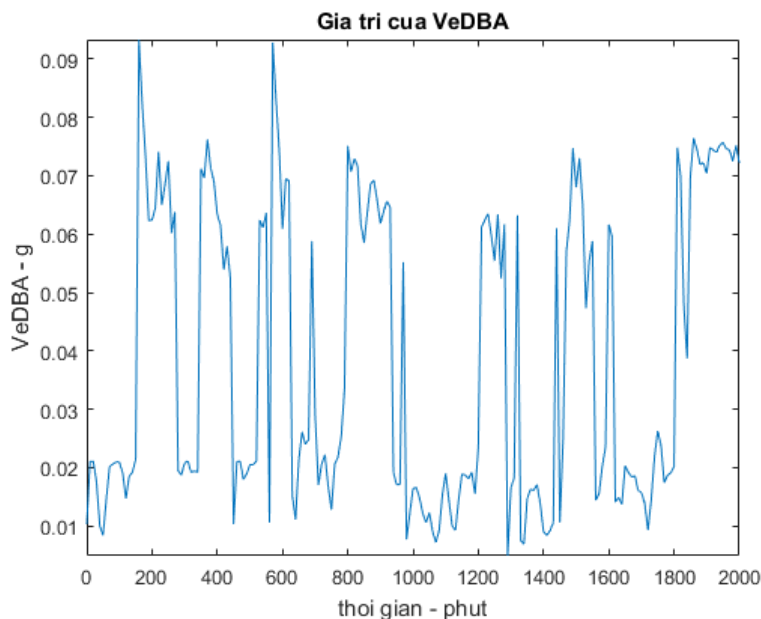
Dữ liệu đầu vào [7] có khoảng thời gian giữa các mẫu là 10 phút. Thời gian để thu thập bộ dữ liệu là khoảng 2000 phút.

Dữ liệu về VeDBA, SCAY và trạng thái của bò có 201 mẫu.

Khởi tạo dữ liệu ngưỡng A và B là 500 mẫu.

Kết quả mô phỏng thuật toán:

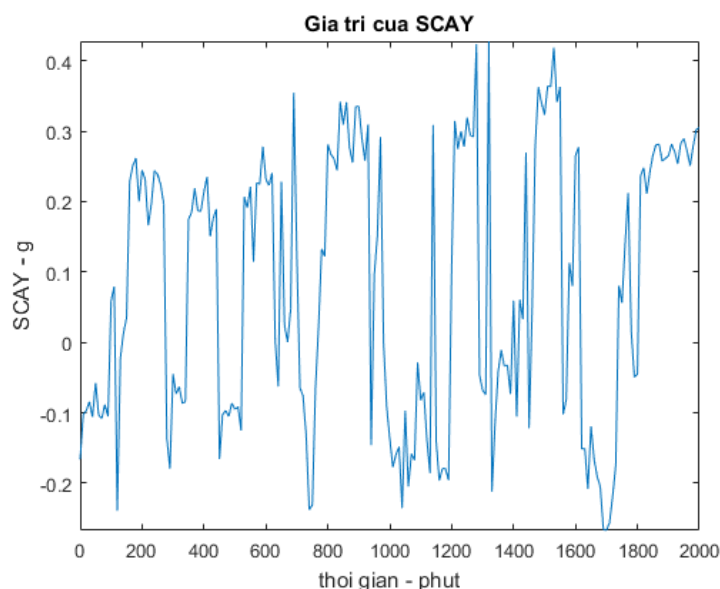
Ban đầu ta khảo sát dữ liệu VeDBA, để từ đó tính toán được khoảng giá trị cho ngưỡng A.



Hình 2.3. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 10 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng A là [0,0049 ; 0,0933].

Tiếp theo ta khảo sát dữ liệu SCAY, để từ đó tính toán được khoảng giá trị cho ngưỡng B.



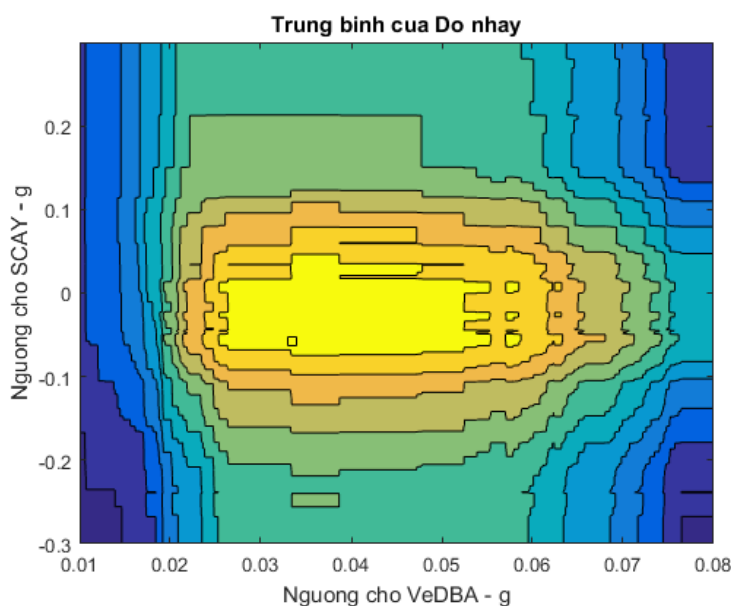
Hình 2.4. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 10 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng B là $[-0,2672 ; 0,4280]$.

Khi hệ thống cần đạt độ nhạy tốt nhất.

$$\text{Max Sen} = \frac{TP}{TP+FN}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



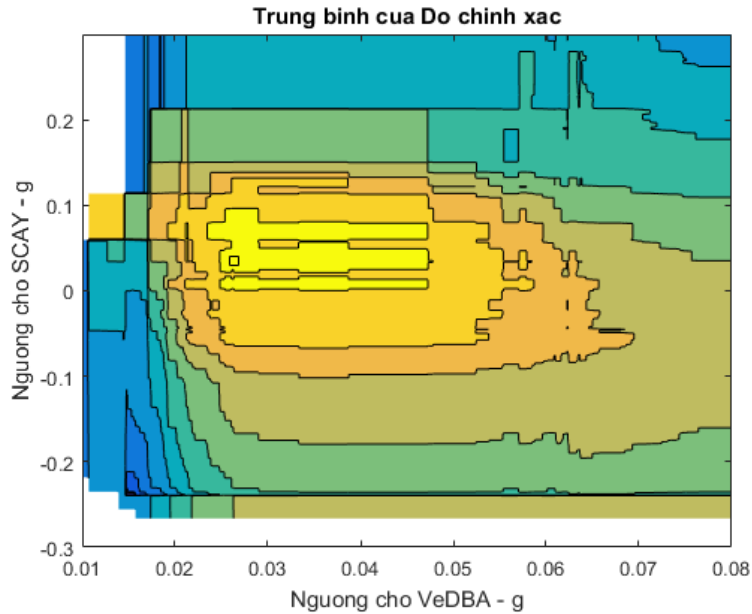
Hình 2.5. Sự thay đổi giá trị ngưỡng theo độ nhạy, dữ liệu lấy mẫu 10 phút/lần

Như trên đồ thị kết quả, giá trị độ nhạy thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ nhạy lớn cho ta kết quả tốt nhất về hiệu năng độ nhạy. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = $0,0334g$, ngưỡng B = $-0,0571g$, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chính xác tốt nhất.

$$\text{Max Pre} = \frac{TP}{TP+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



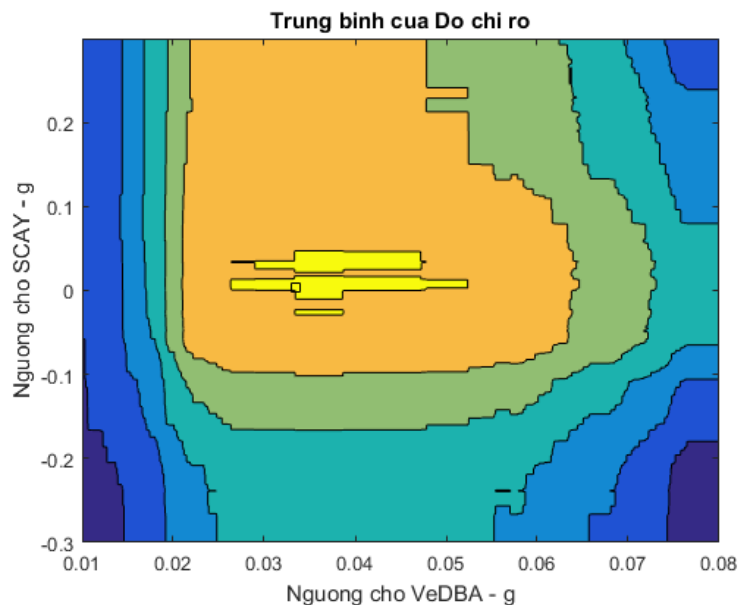
Hình 2.6. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 10 phút/lần

Như trên đồ thị kết quả, giá trị độ chính xác thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chính xác lớn cho ta kết quả tốt nhất về hiệu năng độ chính xác. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0264g, ngưỡng B = 0,0343g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chỉ rõ tốt nhất.

$$\text{Max Spe} = \frac{TN}{TN+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



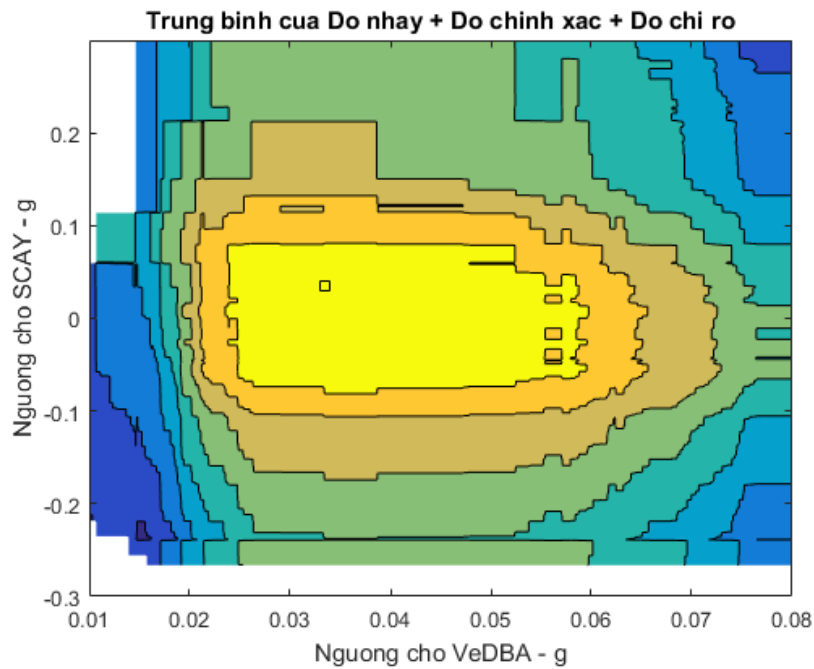
Hình 2.7. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 10 phút/lần

Như trên đồ thị kết quả, giá trị độ chỉ rõ thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chỉ rõ lớn cho ta kết quả tốt nhất về hiệu năng độ chỉ rõ. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0224g, ngưỡng B = 0,003g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt tốt nhất và đồng đều cho cả độ nhạy, độ chính xác và độ chỉ rõ.

$$\text{Max } \frac{\text{Sen} + \text{Pre} + \text{Spe}}{3}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



Hình 2.8. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 10 phút/lần

Như trên đồ thị kết quả, giá trị trung bình các tham số thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị trung bình các tham số lớn cho ta kết quả tốt và đồng đều trung bình các tham số. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0334g, ngưỡng B = 0,0343g, ($1g = 9,8 \text{ m/s}^2$).

Ta có kết quả bảng tổng kết các giá trị ngưỡng vừa tìm được ở trên như sau:

Bảng 2.1. Các giá trị ngưỡng khi dữ liệu lấy mẫu 10 phút/lần

	Ngưỡng VeDBA, A Đơn vị là g	Ngưỡng SCAY, B Đơn vị là g
Đạt độ nhạy lớn nhất	0,0334	-0,0571
Đạt độ chính xác lớn nhất	0,0264	0,0343

Đạt độ chỉ rõ lớn nhất	0,0334	0,0030
Trung bình cả Độ nhạy, Độ chính xác, Độ chỉ rõ	0,0334	0,0343

Từ bảng trên ta thấy các giá trị ngưỡng A và ngưỡng B tính toán cho bộ dữ liệu lấy mẫu 10 phút/lần, thay đổi tùy thuộc vào tham số hiệu năng hệ thống cần đạt được. Cụ thể đối với bộ dữ liệu lấy mẫu 10 phút/ lần, khi cần đạt độ nhạy lớn nhất, ta chọn ngưỡng A là 0,0334 và ngưỡng B là -0,0571. Trong khi muốn đạt độ chỉ rõ lớn nhất thì ta chọn ngưỡng A là 0,0334 và ngưỡng B là 0,0030.

2.5.2. Kích bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 05 phút/lần

Tham số đầu vào:

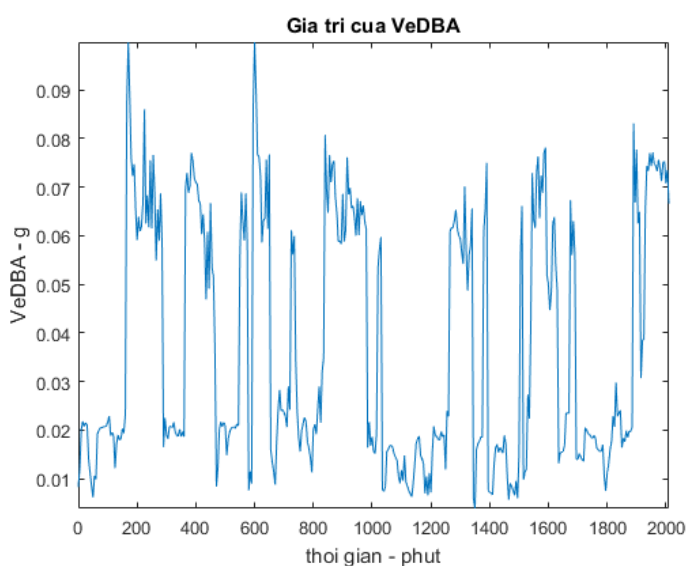
Dữ liệu đầu vào [7] có khoảng thời gian giữa các mẫu là 05 phút. Thời gian để thu thập bộ dữ liệu là khoảng 2000 phút.

Dữ liệu về VeDBA, SCAY và trạng thái của bò có 403 mẫu.

Khởi tạo dữ liệu ngưỡng A và B là 500 mẫu.

Kết quả mô phỏng thuật toán:

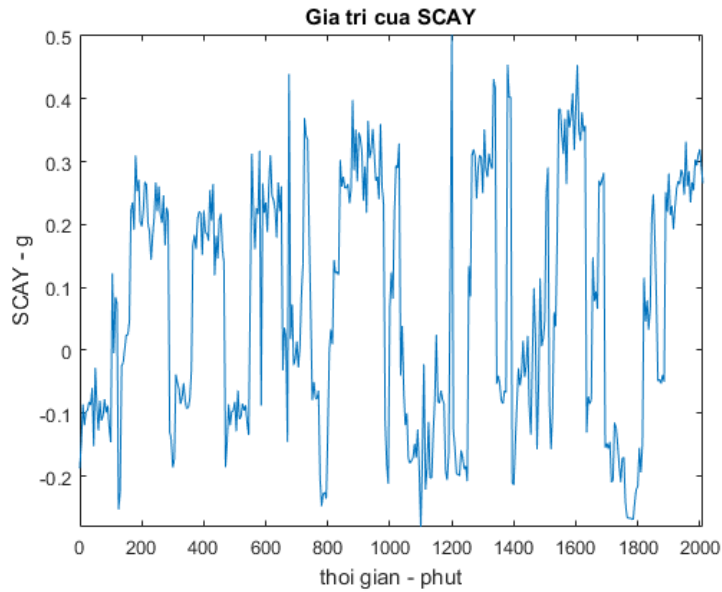
Ban đầu ta khảo sát dữ liệu VeDBA, để từ đó tính toán được khoảng giá trị cho ngưỡng A.



Hình 2.9. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 05 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng A là [0,0039 ; 0,0998]

Tiếp theo ta khảo sát dữ liệu SCAY, để từ đó tính toán được khoảng giá trị cho ngưỡng B.



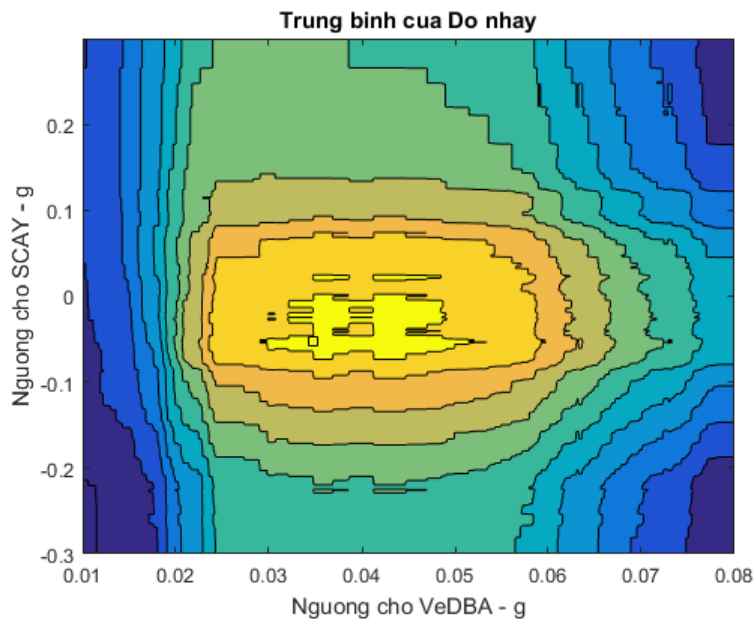
Hình 2.10. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 05 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng B là $[-0,2808 ; 0,5020]$

Khi hệ thống cần đạt độ nhạy tốt nhất.

$$\text{Max Sen} = \frac{TP}{TP+FN}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



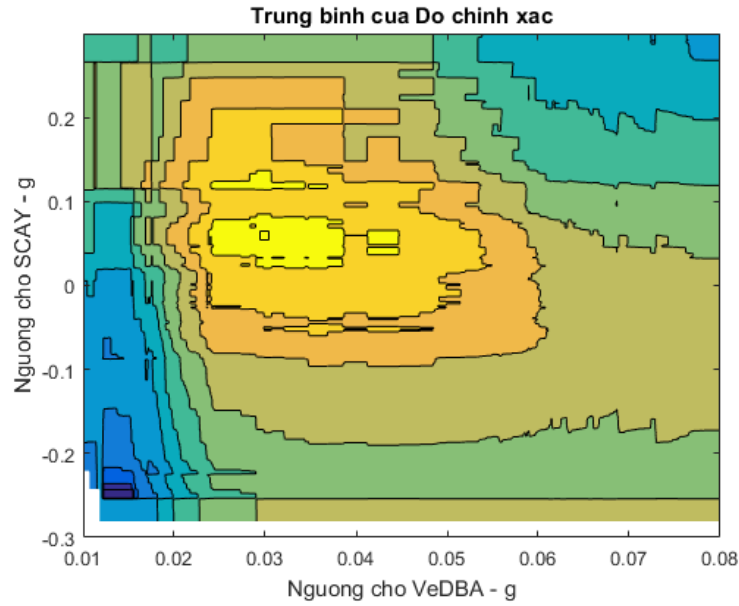
Hình 2.11. Sự thay đổi giá trị ngưỡng theo độ nhạy, dữ liệu lấy mẫu 05 phút/lần

Như trên đồ thị kết quả, giá trị độ nhạy thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ nhạy lớn cho ta kết quả tốt nhất về hiệu năng độ nhạy. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0348g, ngưỡng B = -0,0535g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chính xác tốt nhất.

$$\text{Max Pre} = \frac{TP}{TP+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



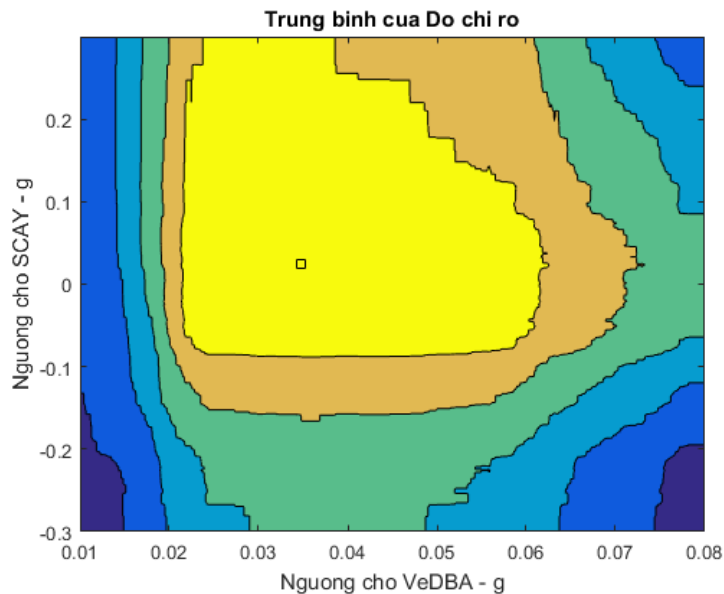
Hình 2.12. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 05 phút/lần

Như trên đồ thị kết quả, giá trị độ chính xác thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chính xác lớn cho ta kết quả tốt nhất về hiệu năng độ chính xác. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0299g, ngưỡng B = 0,0595g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chỉ rõ tốt nhất.

$$\text{Max Spe} = \frac{TN}{TN+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



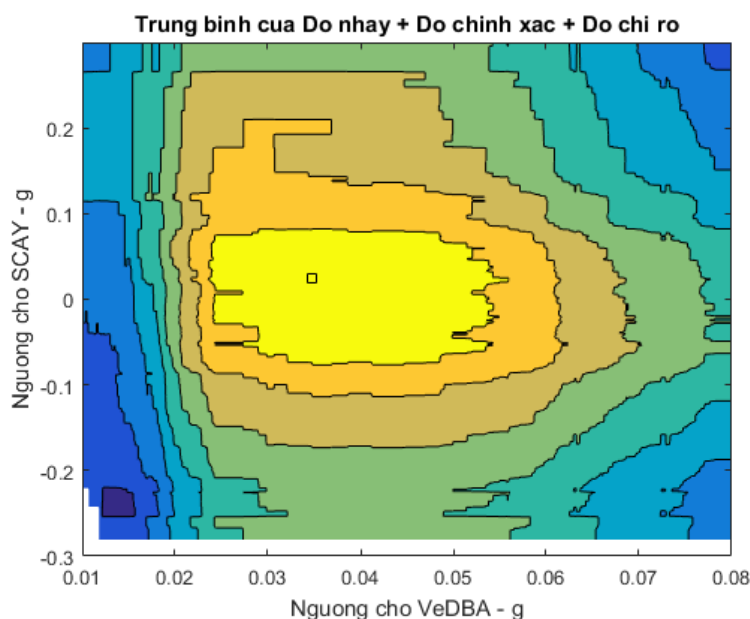
Hình 2.13. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 05 phút/lần

Như trên đồ thị kết quả, giá trị độ chỉ rõ thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chỉ rõ lớn cho ta kết quả tốt nhất về hiệu năng độ chỉ rõ. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0348g, ngưỡng B = 0,0234g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt tốt nhất và đồng đều cho cả độ nhạy, độ chính xác và độ chỉ rõ.

$$\text{Max } \frac{\text{Sen} + \text{Pre} + \text{Spe}}{3}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



Hình 2.14. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 05 phút/lần

Như trên đồ thị kết quả, giá trị trung bình các tham số thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị trung bình các tham số lớn cho ta kết quả tốt và đồng đều trung bình các tham số. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0348g, ngưỡng B = 0,0234g, ($1g = 9,8 \text{ m/s}^2$).

Ta có kết quả bảng tổng kết các giá trị ngưỡng vừa tìm được ở trên như sau:

Bảng 2.2. Các giá trị ngưỡng khi dữ liệu lấy mẫu 05 phút/lần

	Ngưỡng VeDBA, A Đơn vị là g	Ngưỡng SCAY, B Đơn vị là g
Đạt độ nhạy lớn nhất	0,0348	-0,0535
Đạt độ chính xác lớn nhất	0,0299	0,0595
Đạt độ chỉ rõ lớn nhất	0,0348	0,0234

Trung bình cả Độ nhạy, Độ chính xác, Độ chỉ rõ	0,0348	0,0234
---	--------	--------

Từ bảng trên ta thấy các giá trị ngưỡng A và ngưỡng B tính toán cho bộ dữ liệu lấy mẫu 05 phút/lần, thay đổi tùy thuộc vào tham số hiệu năng hệ thống cần đạt được. Cụ thể đối với bộ dữ liệu lấy mẫu 05 phút/ lần, khi cần đạt độ nhạy lớn nhất, ta chọn ngưỡng A là 0,0348 và ngưỡng B là -0,0535. Trong khi muốn đạt độ chính xác lớn nhất thì ta chọn ngưỡng A là 0,0299 và ngưỡng B là 0,0595.

2.5.3. Kịch bản mô phỏng thuật toán với bộ dữ liệu lấy mẫu 01 phút/lần

Tham số đầu vào:

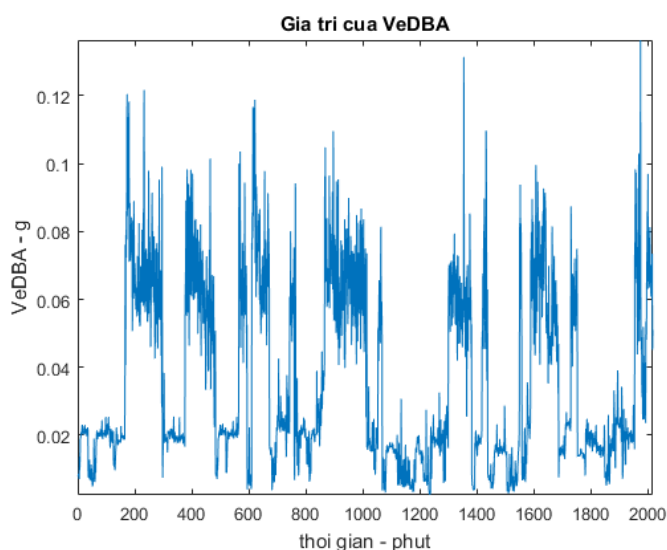
Dữ liệu đầu vào [7] có khoảng thời gian giữa các mẫu là 01 phút. Thời gian để thu thập bộ dữ liệu là khoảng 2000 phút.

Dữ liệu về VeDBA, SCAY và trạng thái của bò có 2019 mẫu.

Khởi tạo dữ liệu ngưỡng A và B là 500 mẫu.

Kết quả mô phỏng thuật toán:

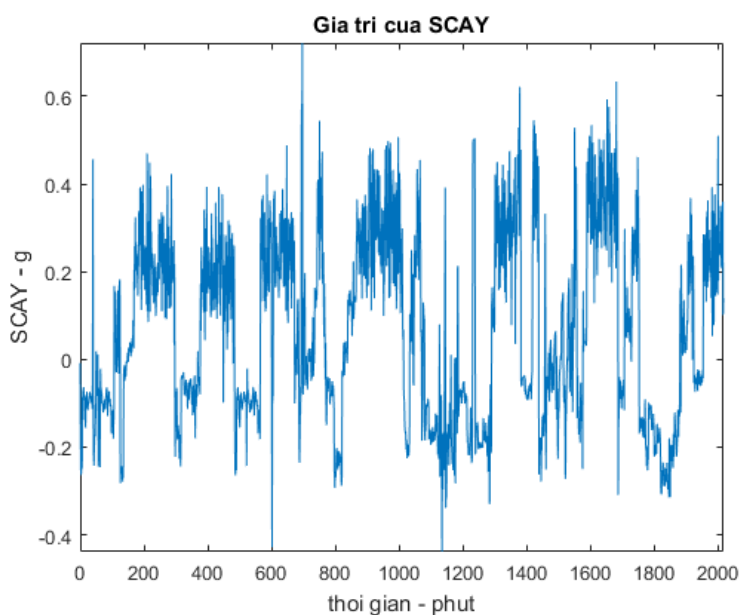
Ban đầu ta khảo sát dữ liệu VeDBA, để từ đó tính toán được khoảng giá trị cho ngưỡng A.



Hình 2.15. Sự thay đổi của giá trị VeDBA với dữ liệu lấy mẫu 01 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng A là [0,0024 ; 0,1364]

Tiếp theo ta khảo sát dữ liệu SCAY, để từ đó tính toán được khoảng giá trị cho ngưỡng B.



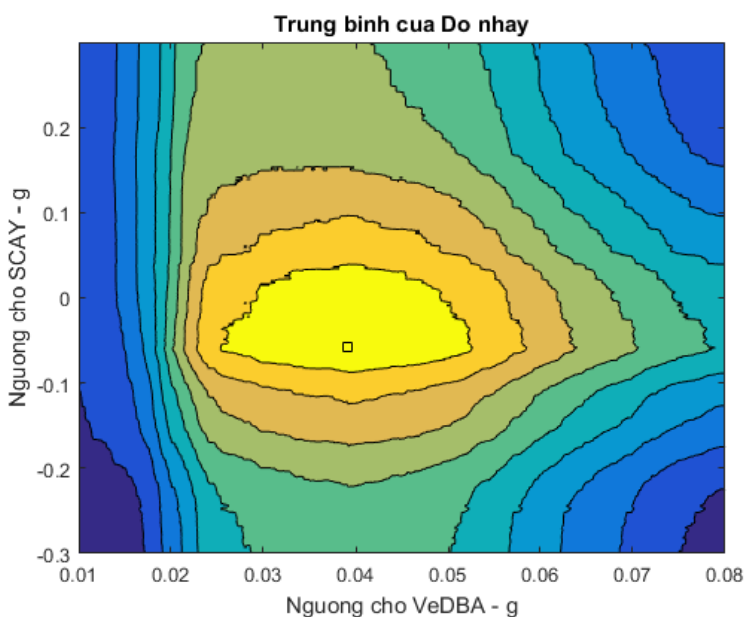
Hình 2.16. Sự thay đổi của giá trị SCAY với dữ liệu lấy mẫu 01 phút/lần

Từ đồ thị ta có giới hạn của ngưỡng B là $[-0,4389 ; 0,7216]$

Khi hệ thống cần đạt độ nhạy tốt nhất.

$$\text{Max Sen} = \frac{TP}{TP+FN}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



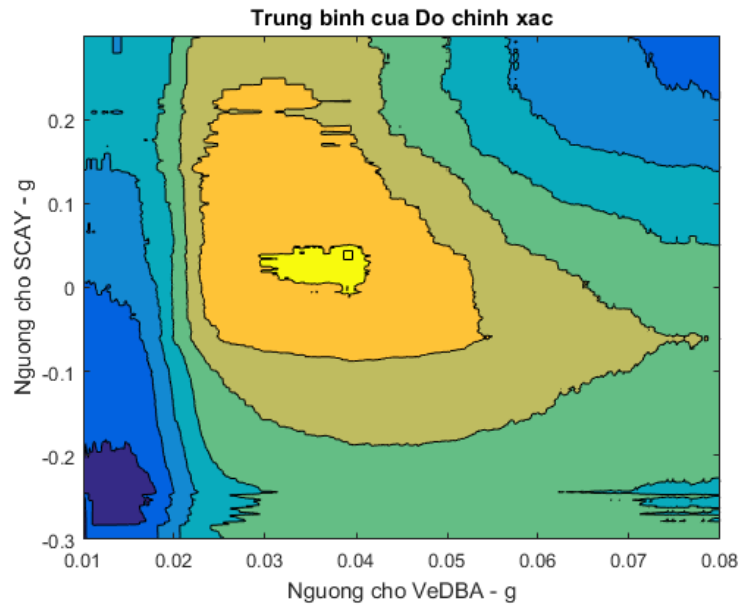
Hình 2.17. Sự thay đổi giá trị ngưỡng theo độ nhạy, dữ liệu lấy mẫu 01 phút/lần

Như trên đồ thị kết quả, giá trị độ nhạy thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ nhạy lớn cho ta kết quả tốt nhất về hiệu năng độ nhạy. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0392g, ngưỡng B = -0,0583g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chính xác tốt nhất.

$$\text{Max Pre} = \frac{TP}{TP+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



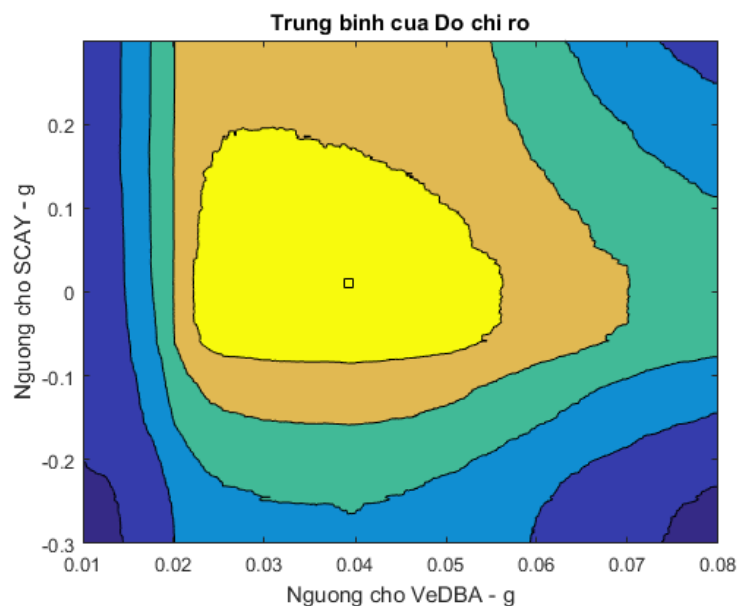
Hình 2.18. Sự thay đổi giá trị ngưỡng theo độ chính xác, dữ liệu lấy mẫu 01 phút/lần

Như trên đồ thị kết quả, giá trị độ chính xác thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chính xác lớn cho ta kết quả tốt nhất về hiệu năng độ chính xác. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0392g, ngưỡng B = 0,0379g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt độ chỉ rõ tốt nhất.

$$\text{Max Spe} = \frac{TN}{TN+FP}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



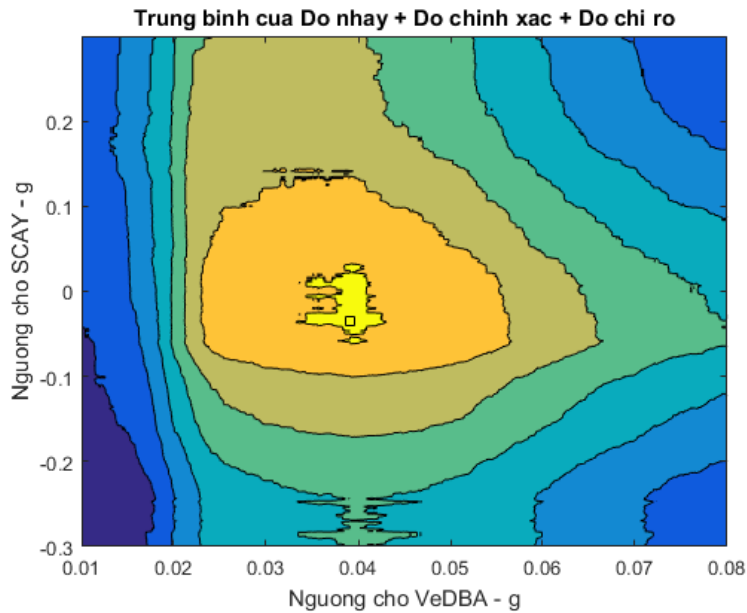
Hình 2.19. Sự thay đổi giá trị ngưỡng theo độ chỉ rõ, dữ liệu lấy mẫu 01 phút/lần

Như trên đồ thị kết quả, giá trị độ chỉ rõ thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị độ chỉ rõ lớn cho ta kết quả tốt nhất về hiệu năng độ chỉ rõ. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0393g, ngưỡng B = 0,0102g, ($1g = 9,8 \text{ m/s}^2$).

Khi hệ thống cần đạt tốt nhất và đồng đều cho cả độ nhạy, độ chính xác và độ chỉ rõ.

$$\text{Max } \frac{\text{Sen} + \text{Pre} + \text{Spe}}{3}$$

Ta có đồ thị Contour xác định ngưỡng A và B đồng thời như sau:



Hình 2.20. Sự thay đổi giá trị ngưỡng theo trung bình các tham số, dữ liệu lấy mẫu 01 phút/lần

Như trên đồ thị kết quả, giá trị trung bình các tham số thay đổi theo ngưỡng A và ngưỡng B. Các đường đồng mức bên trong cho giá trị lớn hơn các đường đồng mức bên ngoài. Điểm ô vuông đánh dấu nằm bên trong vùng giá trị trung bình các tham số lớn cho ta kết quả tốt và đồng đều trung bình các tham số. Từ đó ta tìm được giá trị cho 2 ngưỡng là: Ngưỡng A = 0,0392g, ngưỡng B = -0,0355g, ($1g = 9,8 \text{ m/s}^2$).

Ta có kết quả bảng tổng kết các giá trị ngưỡng vừa tìm được ở trên như sau:

Bảng 2.3. Các giá trị ngưỡng khi dữ liệu lấy mẫu 01 phút/lần

	Ngưỡng VeDBA, A Đơn vị là g	Ngưỡng SCAY, B Đơn vị là g
Đạt độ nhạy lớn nhất	0,0392	-0,0583
Đạt độ chính xác lớn nhất	0,0392	0,0379
Đạt độ chỉ rõ lớn nhất	0,0393	0,0102

Trung bình cả Độ nhạy, Độ chính xác, Độ chỉ rõ	0,0392	-0,0355
---	--------	---------

Từ bảng trên ta thấy các giá trị ngưỡng A và ngưỡng B tính toán cho bộ dữ liệu lấy mẫu 01 phút/lần, thay đổi tùy thuộc vào tham số hiệu năng hệ thống cần đạt được. Cụ thể đối với bộ dữ liệu lấy mẫu 01 phút/ lần, khi cần đạt độ nhạy lớn nhất, ta chọn ngưỡng A là 0,0392 và ngưỡng B là -0,0583. Trong khi muốn đạt độ chỉ rõ lớn nhất thì ta chọn ngưỡng A là 0,0393 và ngưỡng B là 0,0102.

2.6. Nhận xét

Chương II đã thành công trong việc tìm ngưỡng A và ngưỡng B của thuật toán. Việc xác định được giá trị của 2 ngưỡng này phụ thuộc vào tiêu chí tham số hiệu năng cần đạt được, phụ thuộc vào bộ dữ liệu lấy mẫu cũng như thuật toán đề ra.

Như kết quả thấy, việc lựa chọn tham số hiệu năng nào để tìm ngưỡng và bộ dữ liệu lấy mẫu, sẽ cho kết quả là giá trị các ngưỡng thay đổi khác nhau.

CHƯƠNG III. ĐÁNH GIÁ THUẬT TOÁN

3.1. Hiệu năng của thuật toán ở các tần số lấy mẫu khác nhau

Việc so sánh hiệu năng của thuật toán ở các bộ dữ liệu [7] ở tần số lấy mẫu khác nhau: 1 phút/ lần, 5 phút/ lần và 10 phút/ lần, sẽ cho ta thấy được sự khác nhau của hiệu năng khi tần số lấy mẫu dữ liệu thay đổi, để từ đó có thể lựa chọn tần số lấy mẫu dữ liệu trong thực tế sao cho phù hợp.

Để có thể so sánh hiệu năng của thuật toán ở các tần số lấy mẫu khác nhau. Trước hết đánh giá hiệu năng thuật toán dựa vào tham số về độ nhạy, độ chính xác, độ chỉ rõ ở từng bộ dữ liệu lấy mẫu. Sau khi có kết quả các tiêu chí đánh giá hiệu năng của từng bộ tần số lấy mẫu dữ liệu, ta sẽ tiến hành so sánh với các bộ dữ liệu tần số lấy mẫu với nhau.

3.1.1. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 10 phút/lần

Như phần trước đã tìm ra được các giá trị ngưỡng A và B tương ứng với các tiêu chí hiệu năng của thuật toán cần đạt được, của bộ dữ liệu 10 phút/lần. Ở phần này, để có thể đánh giá hiệu năng của thuật toán, ta sẽ chọn giá trị ngưỡng A và B tương ứng ở phần trước, để tính các tham số phù hợp tương ứng.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ nhạy tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 10 phút/lần và giá trị ngưỡng $A = 0,0334g$, ngưỡng $B = -0,0571g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 10 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.1. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 10 phút/lần

	Độ nhạy lớn (%)	Độ chính xác (%)	Độ chỉ rõ (%)
Ăn	100,00	93,18	94,90
Nằm	80,90	98,63	95,65
Đứng	95,45	53,85	80,00
Tổng hợp	92,12	81,90	90,20

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ nhạy của thuật toán tốt nhất thì giá trị độ nhạy lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ nhạy của việc phân loại trạng thái ăn là 100%.
- Độ nhạy của việc phân loại trạng thái nằm là 80,9%.
- Độ nhạy của việc phân loại trạng thái đứng là 95,45%.
- Độ nhạy trung bình của việc phân loại trạng thái trên là 92,12%.

Do chỉ chọn giá trị ngưỡng để độ nhạy của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ nhạy và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 53,85%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chính xác tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 10 phút/lần và giá trị ngưỡng $A = 0,0264g$, ngưỡng $B = 0,0343g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 10 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.2. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 10 phút/lần

	Độ nhạy (%)	Độ chính xác lớn (%)	Độ chỉ rõ (%)
Ăn	100,00	91,11	93,27
Nằm	95,50	93,41	72,27
Đứng	72,72	80,00	95,50
Tổng hợp	89,40	88,17	87,17

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chính xác của thuật toán tốt nhất thì giá trị độ chính xác lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chính xác của việc phân loại trạng thái ăn là 91,11%.
- Độ chính xác của việc phân loại trạng thái nằm là 93,41%.
- Độ chính xác của việc phân loại trạng thái đứng là 80,00%.
- Độ chính xác trung bình của việc phân loại trạng thái trên là 88,17%.

Do chỉ chọn giá trị ngưỡng để độ chính xác của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chính xác và không đồng đều. Cụ thể, độ chỉ rõ của việc phân loại nằm chỉ đạt 72,27%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chỉ rõ tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 10 phút/lần và giá trị ngưỡng $A = 0,0334g$, ngưỡng $B = 0,0030g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 10 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.3. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 10 phút/lần

	Độ nhạy (%)	Độ chính xác (%)	Độ chỉ rõ lớn (%)
Ăn	100,00	93,18	94,90
Nằm	92,13	95,35	82,60
Đứng	81,81	69,23	91,10
Tổng hợp	91,32	85,90	89,50

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chỉ rõ của thuật toán tốt nhất thì giá trị độ chỉ rõ lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chỉ rõ của việc phân loại trạng thái ăn là 94,90%.
- Độ chỉ rõ của việc phân loại trạng thái nằm là 82,60%.
- Độ chỉ rõ của việc phân loại trạng thái đứng là 91,10%.
- Độ chỉ rõ trung bình của việc phân loại trạng thái trên là 89,50%.

Do chỉ chọn giá trị ngưỡng để độ chỉ rõ của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chỉ rõ và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 69,23%.

3.1.2. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 05 phút/lần

Như phần trước đã tìm ra được các giá trị ngưỡng A và B tương ứng với các tiêu chí hiệu năng của thuật toán cần đạt được, của bộ dữ liệu 05 phút/lần. Ở phần này, để có thể đánh giá hiệu năng của thuật toán, ta sẽ chọn giá trị ngưỡng A và B tương ứng ở phần trước, để tính các tham số phù hợp tương ứng.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ nhạy tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 05 phút/lần và giá trị ngưỡng $A = 0,0348g$, ngưỡng $B = -0,0535g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 05 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.4. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 05 phút/lần

	Độ nhạy lớn (%)	Độ chính xác (%)	Độ chỉ rõ (%)
Ăn	99,40	92,80	95,10
Nằm	78,40	97,97	94,00
Đứng	93,60	50,57	77,13
Tổng hợp	90,40	80,45	88,75

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ nhạy của thuật toán tốt nhất thì giá trị độ nhạy lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ nhạy của việc phân loại trạng thái ăn là 99,40%.
- Độ nhạy của việc phân loại trạng thái nằm là 78,40%.
- Độ nhạy của việc phân loại trạng thái đứng là 93,60%.
- Độ nhạy trung bình của việc phân loại trạng thái trên là 90,40%.

Do chỉ chọn giá trị ngưỡng để độ nhạy của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ nhạy và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 50,57%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chính xác tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 05 phút/lần và giá trị ngưỡng $A = 0,0299g$, ngưỡng $B = 0,0595g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 05 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.5. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 05 phút/lần

	Độ nhạy (%)	Độ chính xác lớn (%)	Độ chỉ rõ (%)
Ăn	100,00	91,20	93,90
Nằm	96,70	89,90	57,40
Đứng	58,70	81,80	96,80
Tổng hợp	85,20	87,70	82,70

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chính xác của thuật toán tốt nhất thì giá trị độ chính xác lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chính xác của việc phân loại trạng thái ăn là 91,20%.
- Độ chính xác của việc phân loại trạng thái nằm là 89,90%.
- Độ chính xác của việc phân loại trạng thái đứng là 81,80%.
- Độ chính xác trung bình của việc phân loại trạng thái trên là 87,70%.

Do chỉ chọn giá trị ngưỡng để độ chính xác của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chính xác và không đồng đều. Cụ thể, độ chỉ rõ của việc phân loại nằm chỉ đạt 57,40%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chỉ rõ tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 05 phút/lần và giá trị ngưỡng A = 0,0348g, ngưỡng B = 0,0234g. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 05 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.6. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 05 phút/lần

	Độ nhạy (%)	Độ chính xác (%)	Độ chỉ rõ lớn (%)
Ăn	99,30	92,80	95,10
Nằm	94,00	93,00	74,00
Đứng	72,30	70,80	92,60
Tổng hợp	88,60	85,60	87,20

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chỉ rõ của thuật toán tốt nhất thì giá trị độ chỉ rõ lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chỉ rõ của việc phân loại trạng thái ăn là 95,10%.
- Độ chỉ rõ của việc phân loại trạng thái nằm là 74,00%.
- Độ chỉ rõ của việc phân loại trạng thái đứng là 92,60%.
- Độ chỉ rõ trung bình của việc phân loại trạng thái trên là 87,20%.

Do chỉ chọn giá trị ngưỡng để độ chỉ rõ của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chỉ rõ và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 70,80%.

3.1.3. Hiệu năng thuật toán với bộ dữ liệu lấy mẫu 01 phút/lần

Như phần trước đã tìm ra được các giá trị ngưỡng A và B tương ứng với các tiêu chí hiệu năng của thuật toán cần đạt được, của bộ dữ liệu 01 phút/lần. Ở phần này, để có thể đánh giá hiệu năng của thuật toán, ta sẽ chọn giá trị ngưỡng A và B tương ứng ở phần trước, để tính các tham số phù hợp tương ứng.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ nhạy tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 01 phút/lần và giá trị ngưỡng A = 0,0392g, ngưỡng B = - 0,0583g. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 01 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.7. Hiệu năng của thuật toán khi quan tâm đến độ nhạy với dữ liệu lấy mẫu 01 phút/lần

	Độ nhạy lớn (%)	Độ chính xác (%)	Độ chỉ rõ (%)
Ăn	97,28	91,20	94,60
Nằm	76,40	97,10	92,20
Đứng	91,70	47,80	74,20
Tổng hợp	88,46	78,70	87,00

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ nhạy của thuật toán tốt nhất thì giá trị độ nhạy lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ nhạy của việc phân loại trạng thái ăn là 97,28%.
- Độ nhạy của việc phân loại trạng thái nằm là 76,40%.
- Độ nhạy của việc phân loại trạng thái đứng là 91,70%.
- Độ nhạy trung bình của việc phân loại trạng thái trên là 88,46%.

Do chỉ chọn giá trị ngưỡng để độ nhạy của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ nhạy và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 47,80%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chính xác tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 01 phút/lần và giá trị ngưỡng $A = 0,0392g$, ngưỡng $B = 0,0379g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 01 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.8. Hiệu năng của thuật toán khi quan tâm đến độ chính xác với dữ liệu lấy mẫu 10 phút/lần

	Độ nhạy (%)	Độ chính xác lớn (%)	Độ chỉ rõ (%)
Ăn	97,30	91,20	94,60
Nằm	93,30	89,10	61,80
Đứng	60,00	63,60	91,10
Tổng hợp	83,54	81,30	82,50

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chính xác của thuật toán tốt nhất thì giá trị độ chính xác lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chính xác của việc phân loại trạng thái ăn là 91,20%.
- Độ chính xác của việc phân loại trạng thái nằm là 89,10%.
- Độ chính xác của việc phân loại trạng thái đứng là 63,60%.
- Độ chính xác trung bình của việc phân loại trạng thái trên là 81,30%.

Do chỉ chọn giá trị ngưỡng để độ chính xác của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chính xác và không đồng đều. Cụ thể, độ nhạy của việc phân loại đứng chỉ đạt 60,00%.

Khi việc đánh giá hiệu năng thuật toán chỉ cần đạt độ chỉ rõ tốt nhất.

Ta sử dụng bộ dữ liệu lấy mẫu 01 phút/lần và giá trị ngưỡng $A = 0,0393g$, ngưỡng $B = 0,0102g$. Tiếp theo ta thực hiện phân loại hành vi của bò theo lưu đồ thuật toán Hình 2.2. Với kết quả phân loại hành vi của thuật toán, kết hợp với kết quả hành vi của bộ dữ liệu lấy mẫu 01 phút/lần. Ta sẽ tính ra được các giá trị TP, TN, FP và FN từ đó sẽ tìm ra được tham số hiệu năng độ nhạy, độ chính xác và độ chỉ rõ.

Như bảng dưới đây cho ta kết quả hiệu năng về độ nhạy, độ chính xác và độ chỉ rõ.

Bảng 3.9. Hiệu năng của thuật toán khi quan tâm đến độ chỉ rõ với dữ liệu lấy mẫu 01 phút/lần

	Độ nhạy (%)	Độ chính xác (%)	Độ chỉ rõ lớn (%)
Ăn	97,30	91,20	94,60
Nằm	90,30	91,30	71,47
Đứng	69,56	59,70	87,87
Tổng hợp	85,70	80,70	84,60

Nhìn vào bảng ta thấy, khi lựa chọn giá trị ngưỡng A và B sao cho độ chỉ rõ của thuật toán tốt nhất thì giá trị độ chỉ rõ lớn và đồng đều cho việc phân loại các trạng thái của bò: ăn, nằm và đứng. Cụ thể:

- Độ chỉ rõ của việc phân loại trạng thái ăn là 94,60%.
- Độ chỉ rõ của việc phân loại trạng thái nằm là 71,47%.
- Độ chỉ rõ của việc phân loại trạng thái đứng là 87,87%.
- Độ chỉ rõ trung bình của việc phân loại trạng thái trên là 84,60%.

Do chỉ chọn giá trị ngưỡng để độ chỉ rõ của thuật toán tốt nên giá trị hiệu năng của các tiêu chí khác của thuật toán sẽ thấp hơn của độ chỉ rõ và không đồng đều. Cụ thể, độ chính xác của việc phân loại đứng chỉ đạt 59,70%.

3.1.4. So sánh hiệu năng thuật toán với bộ dữ liệu lấy mẫu khác nhau

Trước khi đưa ra bảng so sánh, ta thấy giá trị của chu kỳ lấy mẫu ảnh hưởng tới hiệu năng của thuật toán cây quyết định. Bởi vì, nếu thời gian lấy mẫu nhanh dẫn sẽ rất khó có thể thu được sự chuyển động lên xuống thường xuyên của cổ bò khi chúng đang ăn, vì vậy dẫn đến giá trị VeDBA thấp và gây ra phân loại nhầm. Nhưng nếu thời gian lấy mẫu quá lâu thì sẽ có rất ít điểm giá trị có thể so sánh với hiệu năng của thuật

toán. Để tìm được thời gian lấy mẫu phù hợp, hiệu năng của hệ thống được tính toán với các điều kiện cùng dữ liệu đầu vào, giá trị ngưỡng cố định và thời gian lấy mẫu là 1 phút/lần, 5 phút/lần và 10 phút/lần.

Từ những kết quả trên, ta có bảng so sánh hiệu năng hệ thống giữa các chu kỳ lấy mẫu dữ liệu khác nhau: 1 phút/lần, 5 phút/lần và 10 phút/lần.

Bảng 3.10. Hiệu năng của hệ thống với các chu kỳ lấy mẫu khác nhau

		1 phút/lần	5 phút/lần	10 phút/lần
Độ nhạy (%)	Ăn	97,28	99,40	100,00
	Nằm	76,40	78,40	80,90
	Đứng	91,70	93,60	95,45
	Tổng hợp	88,46	90,40	92,12
Độ chính xác (%)	Ăn	91,20	91,20	91,11
	Nằm	89,10	89,90	93,41
	Đứng	63,60	81,80	80,00
	Tổng hợp	81,30	87,70	88,17
Độ chỉ rõ (%)	Ăn	94,60	95,10	94,90
	Nằm	71,47	74,00	82,60
	Đứng	87,87	92,60	91,10
	Tổng hợp	84,60	87,20	89,50

Bảng trên được thống kê lại từ kết quả tính toán hiệu năng thuật toán với bộ dữ liệu lấy mẫu 10 phút/lần, 05 phút/lần, 01 phút/lần, và xét ở các điều kiện thuật toán đạt độ nhạy tốt nhất, độ chính xác tốt nhất và độ chỉ rõ tốt nhất. Cụ thể, giá trị về độ nhạy ở cột 01 phút/lần, được lấy từ dữ liệu độ nhạy lớn nhất ở bảng 3.7.

Nhìn vào bảng 3.10, ta thấy nhìn chung tổng thể hiệu năng của thuật toán đạt cao nhất khi giá trị lấy mẫu là 10 phút/lần và giảm dần khi giá trị lấy mẫu là 05 phút/lần và 01 phút/lần. Cụ thể:

-Độ nhạy của phát hiện hành vi ăn là 100,00% ở bộ dữ liệu lấy mẫu 10 phút/lần, và giảm dần là 99,40% ở bộ dữ liệu lấy mẫu 05 phút/lần, tiếp theo là 97,70% ở bộ dữ liệu lấy mẫu 01 phút/lần.

- Độ chính xác của phát hiện hành vi nằm là 93,41% ở bộ dữ liệu lấy mẫu 10 phút/lần, và giảm dần là 89,90% ở bộ dữ liệu lấy mẫu 05 phút/lần, tiếp theo là 89,10% ở bộ dữ liệu lấy mẫu 01 phút/lần.

3.2. Hiệu năng của thuật toán so với phương pháp ROC

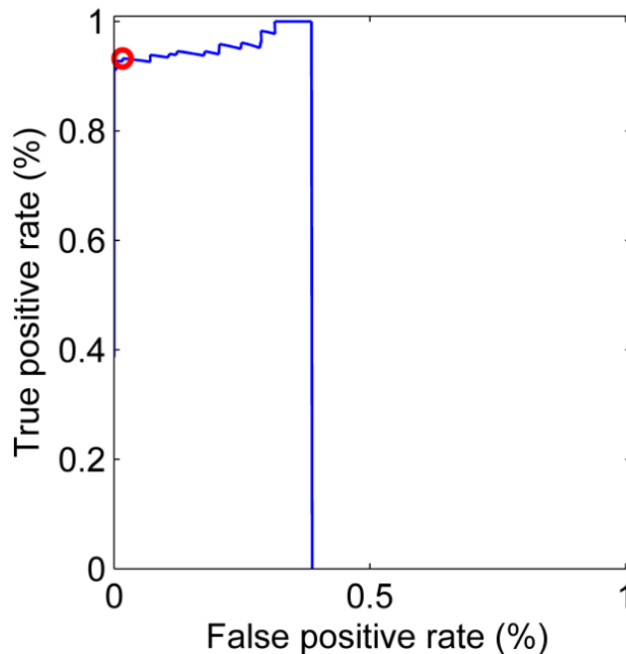
Phương pháp tìm ngưỡng ROC [1,9,10] là phương pháp sử dụng đường cong ROC (đường cong đặc trưng hoạt động) tìm ngưỡng A và B một cách độc lập. Tức là giá trị ngưỡng A được xác định trước và dựa vào tiêu chí hiệu năng thuật toán tốt nhất về độ nhạy hoặc về độ chính xác và bộ dữ liệu cho trước [7]. Điều này cũng tương tự với ngưỡng B.

Phương pháp tìm ngưỡng thực hiện trong luận văn là phương pháp tìm ngưỡng sử dụng đồ thị Contour để tìm 2 ngưỡng A và B một cách đồng thời.

Phần này sẽ tập trung vào việc so sánh các kết quả đạt được giữa 2 phương pháp ROC [1] và phương pháp đồ thị Contour, dựa trên cùng một bộ dữ liệu [7], để đưa ra kết luận so sánh.

Cả 2 phương pháp đều được thực hiện trên cùng bộ dữ liệu đầu vào [7], cùng phương pháp đánh giá hiệu năng. Trong tài liệu [1,9,10], phương pháp tìm ngưỡng ROC dựa vào 2 tiêu chí của hiệu năng hệ thống là độ nhạy tốt nhất hoặc độ chính xác tốt nhất. Kết quả của phương pháp ROC được lấy từ tài liệu [1]. Trong khi phương pháp tìm ngưỡng Contour dựa vào 3 tiêu chí của hiệu năng hệ thống là độ nhạy tốt nhất hoặc độ chính xác tốt nhất hoặc độ chỉ rõ tốt nhất.

Cụ thể ở phương pháp ROC trong tài liệu [1], dữ liệu đầu vào được lấy mẫu 10 phút/lần. Các giá trị ngưỡng A và B được xác định dựa vào tiêu chí tốt nhất về độ nhạy. Việc xác định ngưỡng A và B một cách lần lượt như hình vẽ dưới. Ban đầu xác định giá trị ngưỡng A, sau khi tìm được ngưỡng A tiến hành xác định ngưỡng B.

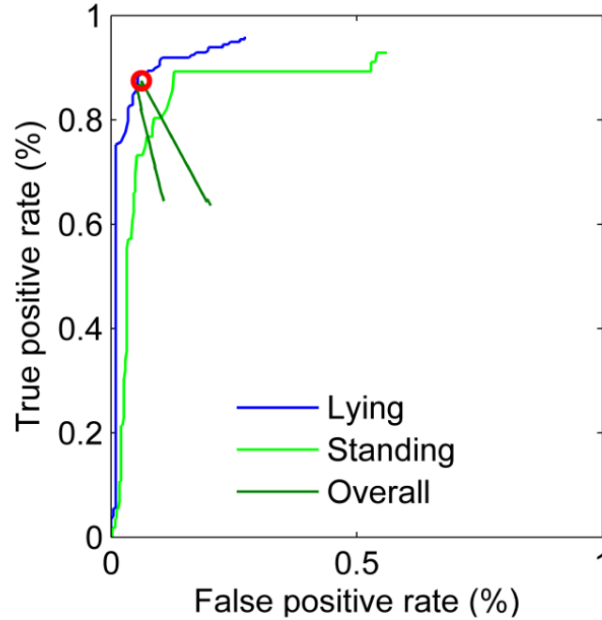


Hình 3.1. Đường cong ROC xác định ngưỡng A theo độ nhạy tốt nhất, dữ liệu lấy mẫu 10 phút/lần [1]

Hình vẽ trên biểu diễn đường cong ROC khi ta biến đổi giá trị ngưỡng A trong khoảng $-0,1g$ đến $0,9g$. Giá trị ngưỡng tốt nhất được lựa chọn sao cho TPR (true

positive rate) là lớn nhất và FPR (false positive rate) là nhỏ nhất. Như hình vẽ trên, điểm tốt nhất sẽ nằm ở góc trên bên trái, được đánh dấu bằng vòng tròn và giá trị ngưỡng A này là 0,0413g [1].

Sau khi tìm được ngưỡng A, tiến hành xác định ngưỡng B cho phân loại hành vi nằm và đứng của bò.



Hình 3.2. Đường cong ROC xác định ngưỡng B theo độ nhạy tốt nhất, dữ liệu lấy mẫu 10 phút/lần [1]

Hình vẽ trên biểu diễn đường cong ROC khi ta biến đổi giá trị ngưỡng B trong khoảng -0,9g đến 0,9g. Giá trị ngưỡng tốt nhất được lựa chọn sao cho TPR là lớn nhất và FPR là nhỏ nhất. Như hình vẽ trên, điểm tốt nhất sẽ nằm ở góc trên bên trái, được đánh dấu bằng vòng tròn và giá trị ngưỡng B là -0,055g [1].

Tương tự với phương pháp Contour, dữ liệu đầu vào được lấy mẫu với chu kỳ 10 phút/lần. Như phần trước đã tính toán, giá trị ngưỡng A và B được xác định dựa vào tiêu chí tốt nhất về độ nhạy, ta tính được ngưỡng A là 0,0334g và ngưỡng B là -0,0571g.

Bảng 3.11. Ví dụ so sánh giá trị ngưỡng khi thực hiện 2 thuật toán

	Ngưỡng V_{eDBA} , A		Ngưỡng $SCAY$, B	
	Đơn vị là g		Đơn vị là g	
	ROC	CONT	ROC	CONT
Đạt độ nhạy lớn nhất Với bộ dữ liệu lấy mẫu 10 phút/lần	0,041	0,0334	-0,055	-0,0571

Từ 2 giá trị ngưỡng khác nhau, dẫn đến kết quả đánh giá hiệu năng của 2 thuật toán cũng khác nhau.

Tương tự như ví dụ trên về việc so sánh giá trị ngưỡng A và B khi thực hiện bởi 2 thuật toán. Các giá trị ngưỡng phù hợp với các tiêu chí hiệu năng thuật toán (độ nhạy, độ chính xác, độ chỉ rõ) và với các bộ dữ liệu lấy mẫu thời gian khác nhau (10 phút/lần, 05 phút/lần, 01 phút/lần) đã được tính toán ở [1] đối với ROC và ở phần trước đối với đồ thị Contour. Từ các giá trị ngưỡng này, sẽ có kết quả đánh giá hiệu năng thuật toán.

Ta có bảng so sánh 2 thuật toán được thống kê lại như sau:

Bảng 3.12. Hiệu năng của hệ thống khi so sánh 2 thuật toán

		1 phút/lần		5 phút/lần		10 phút/lần	
		ROC	CONT	ROC	CONT	ROC	CONT
Độ nhạy lớn (%)	Ăn	95,65	97,28	97,44	99,40	98,78	100,00
	Nằm	74,09	76,40	74,09	78,40	77,42	80,90
	Đứng	82,08	91,70	88,46	93,60	88,00	95,45
	Tổng hợp	83,94	88,46	86,66	90,40	88,06	92,12
Độ chính xác lớn (%)	Ăn	92,03	91,20	93,25	91,20	93,10	91,11
	Nằm	96,57	89,10	97,95	89,90	98,63	93,41
	Đứng	47,01	63,60	47,92	81,80	55,00	80,00
	Tổng hợp	78,53	81,30	79,71	87,70	82,24	88,17
Độ chỉ rõ lớn (%)	Ăn	x	94,60	x	95,10	x	94,90
	Nằm	x	71,47	x	74,00	x	82,60
	Đứng	x	87,87	x	92,60	x	91,10
	Tổng hợp	x	84,60	x	87,20	x	89,50

Bảng trên thống kê lại dữ liệu đã được tính toán ở phần trên đối với phương pháp Contour và lấy từ tài liệu [1] đối với phương pháp ROC. Cụ thể, với phần độ

nhảy lớn ở 1 phút/lần của phần Contour được lấy từ cột độ nhảy lớn (%) trong bảng 3.7.

Trong phương pháp Contour được tính toán trong luận văn này, có sử dụng 3 tham số hiệu năng của hệ thống là độ nhảy, độ chính xác và độ chỉ rõ. Trong khi tài liệu [1] cho phương pháp ROC chỉ sử dụng 2 tham số hiệu năng hệ thống là độ nhảy và độ chính xác.

Như dữ liệu trong bảng cho ta thấy, ở mỗi tham số đánh giá trong cùng điều kiện về thời gian lấy mẫu và tham số hiệu năng, phương pháp Contour đều cho giá trị lớn và đồng đều hơn phương pháp ROC. Cụ thể như:

- Với tham số hiệu năng là độ nhảy lớn, ở bộ dữ liệu có chu kỳ lấy mẫu là 10 phút/lần, phương pháp Contour cho độ nhảy ở hành vi ăn là 100,00% trong khi phương pháp ROC cho độ nhảy chỉ là 98,78%.

- Với tham số hiệu năng là độ chính xác lớn, ở bộ dữ liệu có chu kỳ lấy mẫu là 10 phút/lần, phương pháp Contour cho độ chính xác ở hành vi đứng là 80,00% trong khi phương pháp ROC cho độ chính xác chỉ là 55,00%.

- Với tham số hiệu năng là độ chính xác lớn, ở bộ dữ liệu có chu kỳ lấy mẫu là 10 phút/lần, phương pháp ROC cho bộ dữ liệu là: 93,10%, 98,63%, 55%, 82,24%, trong khi phương pháp Contour cho bộ dữ liệu là: 91,11%, 93,41%, 80,00%, 88,17%. Ta thấy được sự đồng đều và tin cậy hơn của phương pháp Contour so với ROC.

KẾT LUẬN

Trong chương này chúng ta sẽ đi vào đánh giá về thuật toán và hướng nghiên cứu tiếp theo.

Thuật toán cây quyết định đơn giản và hiệu quả, sử dụng dữ liệu từ cảm biến gia tốc để phân loại được một số hành vi quan trọng: nằm, đứng, ăn... Cụ thể là đã khảo sát thuật toán sử dụng đồ thị Contour.

Thuật toán sử dụng đồ thị Contour này đã được khảo sát tính toán trên bộ dữ liệu đã có [7], được lấy mẫu ở các tần số khác nhau 10 phút/lần, 05 phút/lần và 01 phút/lần. Qua việc tính toán này, đã có bảng so sánh hiệu năng thuật toán ở các bộ dữ liệu lấy mẫu khác nhau. Việc đánh giá sử dụng đồ thị Contour với bộ 3 dữ liệu với tần số cập nhật dữ liệu khác nhau sẽ giúp ta lựa chọn được bộ dữ liệu tốt và phù hợp. Kết quả cho thấy dữ liệu được lấy mẫu ở 10 phút/lần cho kết quả tốt nhất.

Thuật toán tìm ngưỡng cho VeDBA và SCAY sử dụng đồ thị Contour (bằng cách tìm 2 ngưỡng tốt nhất một cách đồng thời) được so sánh với thuật toán ROC trên cùng bộ dữ liệu. Do mỗi thuật toán cho ta các giá trị ngưỡng VeDBA và ngưỡng SCAY khác nhau, dẫn đến hiệu năng của mỗi thuật toán cũng khác nhau. Việc tính toán so sánh hiệu năng của 2 thuật toán được cụ thể hóa trong bảng so sánh. Kết quả cho thấy thuật toán sử dụng đồ thị Contour cho kết quả tốt hơn thuật toán ROC.

Bước tiếp theo là thu thập thêm nhiều dữ liệu. Kết hợp với nhiều cảm biến khác gắn lên chân bò, lên thân, để có thể cho kết quả chính xác hơn, cũng như phân loại được nhiều hành vi hơn.

TÀI LIỆU THAM KHẢO

- [1] Diosdado, Jorge A. Vázquez, et al. "*Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system.*" *Animal Biotelemetry* 3.1, vol.1, 2015.
- [2] Venkatraman, Subramaniam, et al. "*Wireless inertial sensors for monitoring animal behavior.*" *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE. IEEE, 2007.*
- [3] Guo, Ying, et al. "*Animal behaviour understanding using wireless sensor networks.*" *Local Computer Networks, Proceedings 2006 31st IEEE Conference on. IEEE, 2006.*
- [4] Roelofs, Judith B., et al. "*Pedometer readings for estrous detection and as predictor for time of ovulation in dairy cattle.*" *Theriogenology* 64.8, 2005, pp.1690-1703.
- [5] Qasem, Lama, et al. "*Tri-axial dynamic acceleration as a proxy for animal energy expenditure; should we be summing values or calculating the vector?.*" *PLoS One* 7.2 (2012): e31187.
- [6] Gleiss, Adrian C., Rory P. Wilson, and Emily LC Shepard. "*Making overall dynamic body acceleration work: on the theory of acceleration as a proxy for energy expenditure.*" *Methods in Ecology and Evolution* Vol. 2, No.1, (2011), pp.23-33.
- [7]https://static-content.springer.com/esm/art%3A10.1186%2Fs40317-015-0045-8/MediaObjects/40317_2015_45_MOESM2_ESM.txt
- [8]<https://uberpython.wordpress.com/2012/01/01/precision-recall-sensitivity-and-specificity>
- [9] Nguyen Thi Huyen Nga, Le Thi Thu Ha, Doan Ba Cuong. "*Giám sát và phân loại hoạt động của bò sử dụng cảm biến gia tốc ba chiều*" Công trình NCKH của sinh viên DHQGHN năm 2015-2016, pp.13-26.
- [10] Nguyen Thi Huyen Nga. "*Classification of behavior of cows using acceleration data and decision tree algorithm*" Thesis of Electronics and communications, 2016, pp.13-50.
- [11] Chinh Nguyen Dinh, Khanh Phung Cong Phi, Tan Tran Duc and Ha Le Vu, "*Nghiên cứu và thiết kế mô hình hệ thống giám sát hành vi trên bò*", The 2016 National Conference on Electronics, Communications and Information Technology, REV, 12/2016, pp. 6:19-6.22.
- [12] Tran, D. T., Huynh, H. T., Nguyen, T. L., Nguyen, P. T., & Nguyen, V. C. (2006), Designing Kalman filters for integration of inertial navigation system and

global positioning system, The 10th biennial Vietnam Conference on Radio & Electronics, REV-2006. Hanoi, pp. 6-10.

[13] Tran, D. T., Luu, M. H., Nguyen, T. L., Nguyen, P. T., & Huynh, H. T. (2007). Performance Improvement of MEMS-Based Sensor Applying in Inertial Navigation Systems. *Posts, Telematics & Information Technology Journal*, 2, 19-24.

[14] Tran, D. T., Luu, M. H., Nguyen, T. L., Nguyen, D. D., & Nguyen, P. T. (2007). Land-vehicle mems INS/GPS positioning during GPS signal blockage periods. *Journal of Science, Vietnam National University, Hanoi*, 23(4), 243-251.

[15] Van Thanh, P., Nguyen, T. A., Duc, N. T., Anh, N. D., & Duc-Tan, T. (2017). Development of a Real Time Supported Program for Motorbike Drivers Using Smartphone Built-in Sensors. *International Journal of Engineering and Technology (IJET)*, 9(2).