

Phân loại hành vi bò: Bộ tham số tối ưu cho thuật toán Rừng ngẫu nhiên

Đỗ Viết Mạnh¹, Trần Đức Tân², Nguyễn Thị Thanh Huyền³, Trần Đức Nghĩa^{1*}

¹Phòng Tin học Quản lý, Viện Công nghệ Thông tin (IOIT-VAST), Hà Nội, Việt Nam;

²Khoa Điện - Điện tử, Trường Đại học Phenikaa, Hà Nội, Việt Nam;

³Khoa Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội, Hà Nội, Việt Nam.

*Email: nghiatd@ioit.ac.vn

Nhận bài: 30/12/2022; Hoàn thiện: 20/02/2023; Chấp nhận đăng: 12/6/2023; Xuất bản: 25/6/2023.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.88.2023.34-41>

TÓM TẮT

Hành vi phản ánh tình trạng sức khỏe động vật, phân tích dữ liệu hành vi có thể hỗ trợ phát hiện sớm một số bệnh của vật nuôi. Theo dõi sức khỏe của bò thông qua việc giám sát hành vi là công cụ hỗ trợ đắc lực cho trang trại lớn, giúp nâng cao hiệu quả chăn nuôi. Dữ liệu gia tốc là một trong những giải pháp hữu hiệu cho hệ thống phân loại hành vi của bò sử dụng gia tốc gắn trên vật nuôi. Chúng tôi đã đề xuất các bộ đặc trưng, cửa sổ dữ liệu và sử dụng thuật toán Rừng ngẫu nhiên nhằm phân loại bốn hành vi cơ bản của bò, bao gồm: ăn, nằm, đứng và đi. Trong bài báo này, nhằm mục đích nâng cao hiệu suất của hệ thống phân loại, chúng tôi khảo sát một số tham số quan trọng của thuật toán Rừng ngẫu nhiên trên bộ dữ liệu thực nghiệm để tìm ra bộ tham số tối ưu. Kết quả cho thấy với giá trị của bộ tham số: số lượng cây = 25 và độ sâu = 15 cho hiệu suất phân loại tốt với độ chính xác là 95,9%.

Từ khoá: Phân loại; Giám sát; Cảm biến gia tốc; Hành vi; Bò; Random forest.

1. MỞ ĐẦU

Đối với ngành chăn nuôi bò để cho chất lượng và sản lượng sữa tốt thì việc đảm bảo sức khỏe cho bò là rất quan trọng [1]. Nhiều nghiên cứu đã chỉ ra, thông qua việc giám sát hành vi bò giúp theo dõi sức khỏe của chúng một cách hữu ích [2-4]. Thông qua dữ liệu hoạt động và hành vi ăn ở bò giúp dự đoán một số bệnh [5-7]. Những bệnh này tác động trực tiếp đến hành vi của bò, làm giảm hoạt động và dẫn đến biếng ăn [8-9].

Đối với chăn nuôi thâm canh thì việc theo dõi thủ công rất tốn kém chi phí và nhân lực, đây là một thách thức lớn cho các nhà quản lý [10]. Giải pháp sử dụng các thuật toán học máy [11] dựa trên dữ liệu thu được từ các cảm biến gắn trên chân và cổ bò giúp phân loại các hành vi của chúng. Cảm biến gia tốc có chi phí rẻ, lắp đặt gọn nhẹ, phù hợp với việc thu thập dữ liệu [12].

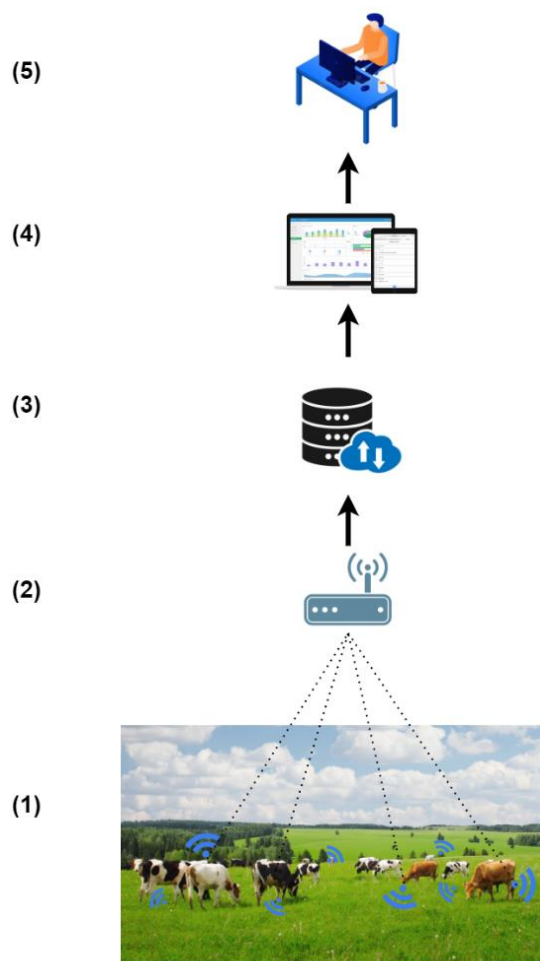
Trong nghiên cứu trước [13], chúng tôi đã đề xuất các bộ đặc trưng, cửa sổ dữ liệu cho thuật toán Random Forest (RF) giúp phân loại bốn hành vi quan trọng của bò bao gồm: ăn, nằm, đi và đứng [14]. Nghiên cứu này có mục đích kiểm chứng và tìm ra bộ tham số tốt nhất của thuật toán RF đối với bộ dữ liệu thực nghiệm của chúng tôi, điều này giúp nâng cao chất lượng hệ thống giám sát, phân loại hành vi của bò. Kết quả của nghiên cứu được thảo luận và so sánh với Wang và cộng sự [15].

2. PHƯƠNG PHÁP

2.1. Tổng quan các thành phần

Mô hình tổng quan hệ thống giám sát hành vi của bò bằng thuật toán Rừng ngẫu nhiên thể hiện ở hình 1.

Xem hình 1, ở phần (1) tất cả con bò của trang trại đều được đeo hai thiết bị là gia tốc bậc tự do 3DoF (Degrees of Freedom) gắn ở cổ và chân bò để đo chuyển động quay xung quanh các trục tọa độ: trục X, trục Y, trục Z. Để có hiệu quả hơn trong phân loại thì hai thiết bị 3DoF phải được đồng bộ hoá dữ liệu với nhau.



Hình 1. Sơ đồ tổng quan các thành phần chính.

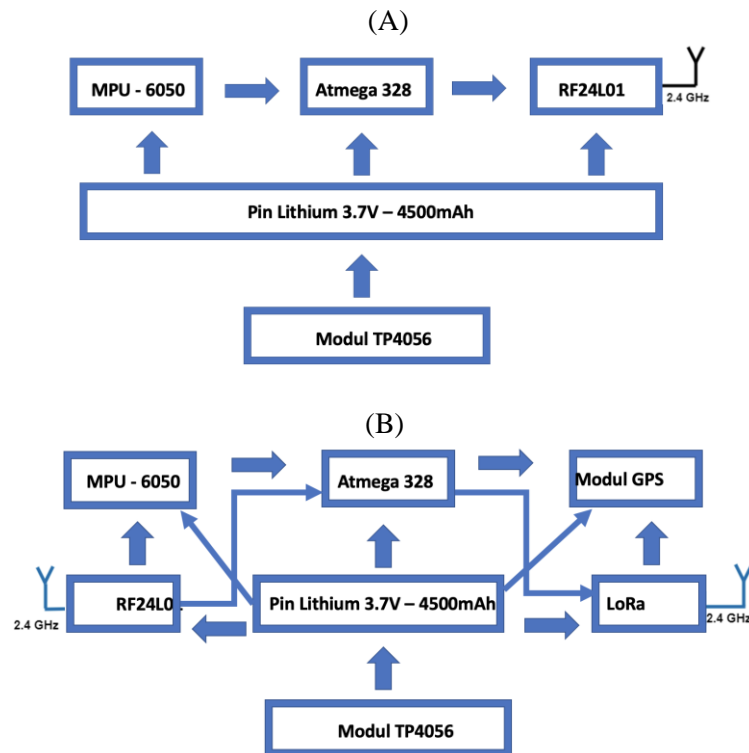
Các dữ liệu từ hai gia tốc 3DoF được gắn ở các con bò sữa trong trang trại thu được sẽ được gửi sang phần (2) để các dữ liệu này được đưa vào mô hình học máy để xác định các hành vi của bò sữa. Sau đó, các kết quả phân loại hành vi của bò tại phần (2) sẽ gửi lên phần (3) để sao lưu, lưu trữ dữ liệu của toàn bộ kết quả phân loại thu được vào cơ sở dữ liệu. Để kiểm tra kết quả phân loại từng con bò, chủ trang trại sẽ sử dụng các thiết bị đầu cuối như laptop, tablet, máy tính để bàn để theo dõi hành vi của các con bò ở trong phần (4) và phần (5).

2.2. Thiết bị thu thập dữ liệu gia tốc ba trục

2.2.1. Sơ đồ khối thiết bị gia tốc kế gắn ở chân và cổ bò

Để có cái nhìn tổng quan hơn về sơ đồ khối của thiết bị chúng tôi đang sử dụng để đeo cho bò được mô tả trong hình 2.

Trên hình 2, phía trên (A) là sơ đồ thiết bị được gắn ở chân bò để đo chuyển động để đo chuyển động xung quanh các trục X,Y,Z. Thiết bị này gồm có các thành phần sau: Cảm biến MPU-6050 là loại 3DoF, Nhân vi xử lý Atmega 328, Modul truyền thông RF24L01, modul có tần số 2.4 GHz, Pin Lithium 3.7 V-4500 mAh để cấp nguồn năng lượng điện để thiết bị hoạt động. Phía dưới (B) là sơ đồ thiết bị gắn ở cổ bò dùng để đo các chuyển động tại phần cổ của bò. Những thành phần của loại thiết bị đeo ở cổ của bò gồm: Cảm biến gia tốc kế MPU-6050 là loại 3DoF, nhân vi xử lý Atmega 328, modul truyền thông RF24L01, modul LoRa, pin Lithium 3.7 V – 4500 mAh cung cấp nguồn điện, cuối cùng là modul LoRa sẽ tạo thành một mạng cảm biến không dây.



Hình 2. Sơ đồ khối thiết bị đo chuyển động gắn ở chân và cổ bò.

2.2.2. Đồng bộ dữ liệu hai cảm biến gia tốc gắn trên chân và cổ

Điểm mấu chốt của nghiên cứu là đồng bộ hóa dữ liệu giữa hai thiết bị (A) và (B) (hình 2). Bộ xử lý sẽ thực hiện những việc sau:

- + Lấy số mẫu gia tốc chân bò nhận được trong một giây gọi là N , $N \leq 10$, tính trung bình cộng giá trị gia tốc của N mẫu thu được trên mỗi trục X , Y , Z .
- + Lấy số mẫu gia tốc nhận được từ cổ bò trong một giây gọi là M , $M \leq 10$, tính trung bình cộng giá trị gia tốc của M mẫu thu được trên mỗi trục X , Y , Z .
- + Nếu các mẫu chỉ thu được dữ liệu ở chân hoặc ở cổ do đường truyền thì các mẫu này sẽ bị loại bỏ. Chúng tôi lưu các giá trị trung bình mà thông tin thời gian (tính bằng giây) tồn tại ở cả chân và cổ

Ở bộ vi điều khiển, dữ liệu gia tốc cổ và chân được đồng bộ hóa theo đơn vị là giây và được đóng gói thành một siêu khung bao gồm các thành phần sau:

- + Mã định danh (ID) của bò
- + Thời gian tính bằng giây
- + Phần trăm pin còn lại trong thiết bị cổ và chân
- + Gia tốc ba trục X , Y , Z đo ở chân bò.
- + Gia tốc ba trục X , Y , Z đo ở cổ con bò.

Dữ liệu đóng gói được gửi đến nút trung tâm thông qua mô-đun giao tiếp LoRa. Chúng tôi đã sử dụng mô-đun LoRa từ AI-Thinker. Con chip RA-02 có lõi SX1278 của SEMTech có tần số hoạt động 433 MHz và khoảng cách giao tiếp 10 km.

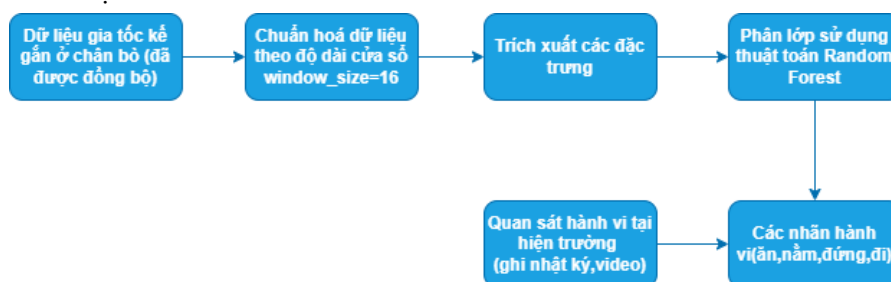
Chúng tôi tiến hành thử nghiệm trên 5 con bò. Để đánh giá sự thích nghi của con bò khi đeo thiết bị, những con bò được chia thành hai nhóm. Các cảm biến được gắn vào bò của nhóm đầu

tiên (bao gồm ID1 và ID4) hai tuần trước khi thử nghiệm, và cho bò của nhóm còn lại (bao gồm ID2, ID3 và ID5) vào ngày thử nghiệm.

20 mẫu/giây là tốc độ lấy mẫu của tín hiệu thô; chúng tôi lấy trung bình ở một mẫu/giây để lưu trữ tín hiệu lấy mẫu xuống. Không có hạn chế hoặc ảnh hưởng đến các hoạt động của bò được công nhận trong quá trình thử nghiệm.

2.3. Phân loại hành vi của bò sử dụng thuật toán Random Forest

2.3.1. Mô hình thuật toán



Hình 3. Mô hình thuật toán phân loại hành vi của bò.

Hình 3 mô tả mô hình thuật toán phân loại hành vi của bò được sử dụng trong nghiên cứu này. Mở đầu là “Dữ liệu gia tốc kế gắn ở chân và cổ bò (đã được đồng bộ hóa)” với “Chuẩn hoá dữ liệu theo độ dài cửa sổ $window_size = 16$ ”. “Trích xuất các đặc trưng” để tính toán cho các bản ghi gồm các: giá trị hiệu dụng (RMS), độ lệch chuẩn (SD) và giá trị trung bình được (mean). “Các nhãn hành vi” gồm bốn lớp: ăn, nằm, đứng, đi.

Dữ liệu sử dụng cho thuật toán RF sẽ được chia thành dữ liệu đào tạo 60% (train data) dựa trên dữ liệu đề xuất, dữ liệu thử nghiệm 40% (test data) dựa trên mô hình phân loại. Các dữ liệu của kết quả về hành vi của bò được đánh giá từ quan sát tại hiện trường.

2.3.2. Các hành vi được phân loại

Nhóm nghiên cứu của chúng tôi tập trung vào bốn hành vi phân loại chính gồm: *Ăn* - Bò đang ăn (nhai/cấn và lấy thức ăn dùng miệng), *Nằm* - Bò đang ở tư thế nằm nghỉ ngơi, *Đứng* - Bò đang đứng trên tất cả 4 chân, *Đi* - Bò đang bước trong vòng 1 giây.

2.3.3. Trích chọn đặc trưng

Đây là bước quan trọng nhất trong mô hình phân loại giúp huấn luyện mô hình tốt hơn, các đặc trưng được trích chọn được mô tả như sau:

$$m(X_j) = \frac{1}{N} \sum_{j=1}^N x_j \tag{1}$$

$$\sigma(X_j) = \sqrt{\frac{1}{N} \sum_{j=1}^k (x_j - m)^2} \tag{2}$$

$$RMS_{X_j} = \sqrt{\frac{1}{N} \sum_{j=1}^N x_j^2} \tag{3}$$

Các công thức ở trên là của trục X, với X là dữ liệu của trục X; X_j là bản ghi j ; N là số lượng mẫu của một bản ghi (độ dài cửa sổ); x_i là mẫu i của bản ghi X_j ; $m(X_j)$ là trung bình của X_j ; $\sigma(X_j)$ là độ lệch chuẩn của X_j ; RMS_{X_j} là giá trị hiệu dụng của X_j . Các công thức (1),(2),(3) dành cho trục X; các trục khác có công thức tương tự như trục X.

2.3.4. Đánh giá phân loại

Công thức để tính độ chính xác trong phân loại như sau:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (4)$$

Dương tính thật (True Positive - TP) là các giá trị dự đoán khớp với giá trị thực tế, giá trị thực tế là hành vi nào đó và mô hình dự đoán đúng hình vi đó. Âm tính thật (True Negative - TN) là giá trị dự đoán khớp với giá trị thực tế, giá trị thực tế là không phải hành vi nào đó và mô hình dự đoán đúng không phải hành vi đó. Dương tính giả (False Positive - FP) - là giá trị dự đoán đã được dự đoán sai, giá trị thực tế là không phải hành vi nào đó nhưng mô hình dự đoán đó là hành vi đó. Âm tính giả (False Negative - FN) - là giá trị dự đoán đã được dự đoán sai, giá trị thực tế là hành vi nào đó nhưng mô hình dự đoán không phải hành vi đó.

2.3.5. Khảo sát các tham số của RF

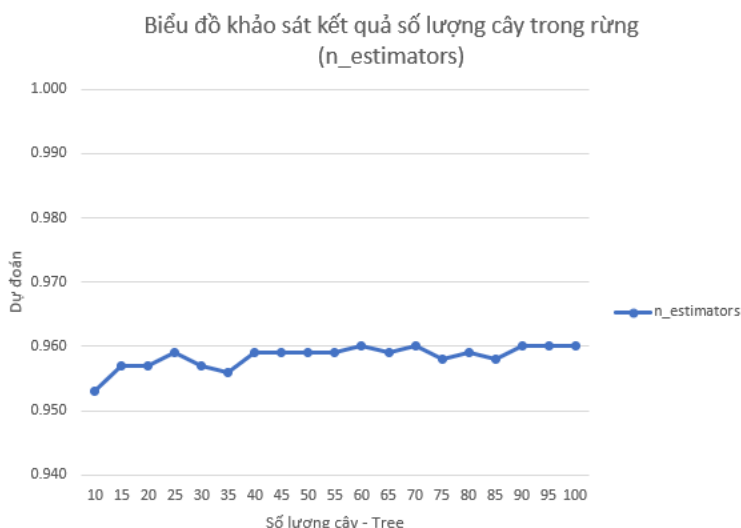
Chúng tôi lựa chọn thuật toán Rừng ngẫu nhiên với ưu điểm tránh được vấn đề overfitting và thực hiện tốt ngay cả khi dữ liệu bị thiếu (trong quá trình thu thập dữ liệu hoặc truyền dữ liệu, dữ liệu có thể bị mất). Tuy nhiên vì nó dựa trên kết quả dự đoán của các cây quyết định con nên mỗi lần thực hiện phân loại hành vi sẽ tốn thời gian hơn, điều này cũng là vấn đề hạn chế khi triển khai trên vi mạch.

Trong nghiên cứu này hai tham số chính để khảo sát của thuật toán Rừng ngẫu nhiên gồm: số lượng cây ($n_estimators$) và độ sâu (max_depth). Với số lượng cây **chúng tôi khảo sát các giá trị trong khoảng từ 10 đến 100 cây. Đối với độ sâu chúng tôi khảo sát các giá trị từ 3 đến 50.**

3. KẾT QUẢ VÀ THẢO LUẬN

Như đã trình bày trong phần 2.2.5, nghiên cứu này khảo sát giá trị của tham số số lượng cây và độ sâu nào là phù hợp nhất cho mô hình phân loại do chúng tôi đề xuất. Với mỗi giá trị của hai tham số trong quá trình huấn luyện sẽ cho ra các mô hình tương ứng, tiêu chí độ chính xác được sử dụng để đánh giá các mô hình này, từ đó giúp lựa chọn được các giá trị tham số phù hợp nhất. Độ chính xác thể hiện trong kết quả là độ chính xác bình quân của việc phân loại 4 hành vi gồm: ăn, đứng, đi, nằm.

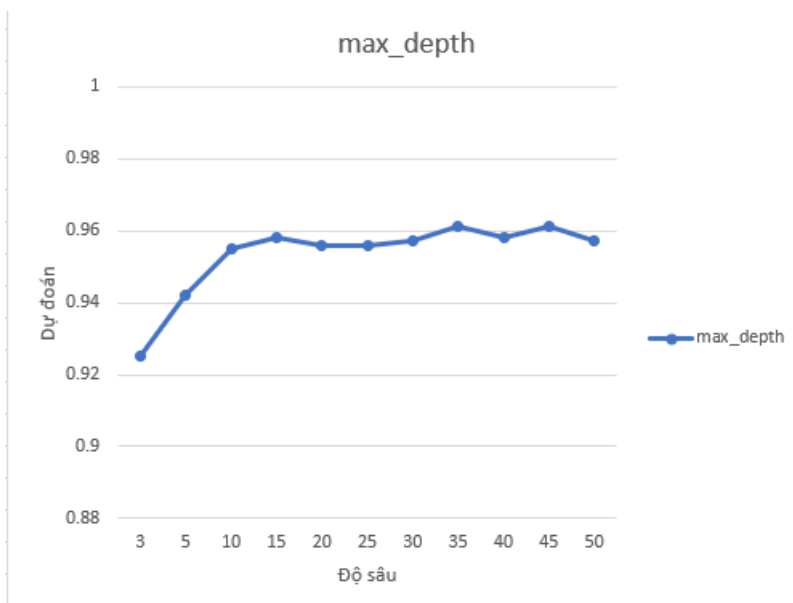
A. Kết quả khảo sát số lượng cây



Hình 4. Kết quả khảo sát giá trị số lượng cây trong RF.

Biểu đồ đường thể hiện kết quả trên hình 4, cho thấy với giá trị số lượng cây $n_estimators = 25$ cho độ chính xác tốt hơn các giá trị từ 10 đến 20 và tương đương với các giá trị cây lớn hơn 40 (đều là 0.96). Với giá trị số lượng cây càng lớn thì độ phức tạp tính toán của mô hình cao hơn. Do đó, chúng tôi xác định được với số lượng cây $n_estimators = 25$ là phù hợp nhất với mô hình.

B. Kết quả khảo sát độ sâu



Hình 5. Kết quả khảo sát giá trị độ sâu trong RF.

Từ hình 5, với các giá trị độ sâu của cây từ max_depth = 3 đến max_depth = 5 cho kết quả phân loại thấp hơn dưới 0.95. Các giá trị độ sâu từ 15 trở lên đều cho kết quả phân loại tốt với giá trị đều đạt trên 0.95. Trong đó với giá trị độ sâu max_depth = 35 và max_depth = 45 cho kết quả tốt nhất đều là 0.961. Với giá trị max_depth = 15, độ chính xác của mô hình đạt kết quả là gần 0.958, đây là giá trị tham số phù hợp với mô hình. Với giá trị độ sâu là 15 nhỏ hơn sẽ mang lại lợi ích giúp mô hình tính toán nhanh hơn, độ phức tạp thấp hơn mà vẫn đảm bảo được độ chính xác trong mô hình phân loại. So sánh công trình liên quan.

C. Thảo luận

Bảng 1. So sánh công trình liên quan.

Nghiên cứu	Hành vi	Công nghệ sử dụng	Tỷ lệ lấy mẫu	Cửa sổ dữ liệu	Thuật toán	Độ chính xác
Wang và các cộng sự[15]	đứng, nằm, ăn, đi bộ, nằm xuống, đứng lên, đi bộ tích cực	kết hợp gia tốc kế gắn trên chân và cảm biến vị trí	1 Hz	6 giây	AdaBoost	86%
Chúng tôi	đứng, nằm, ăn, đi bộ	kết hợp gia tốc kế gắn trên cổ áo và gia tốc kế gắn trên chân	20 Hz	16 giây	Random Forest	95%

Wang và cộng sự [15] đề xuất một phương pháp dựa trên việc kết hợp cảm biến gia tốc và vị trí gắn trên chân và thuật toán AdaBoost để phân loại bảy hành vi của bò. Trong thí nghiệm của họ, tốc độ lấy mẫu là 1 Hz, thời lượng cửa sổ là 6 giây. Trong đó có bốn hành vi như nghiên cứu của chúng tôi: ăn, nằm, đứng, đi. Độ chính xác của hệ thống được đề xuất trong nghiên cứu này tốt hơn của Wang và cộng sự [15] (0,95 so với 0,86). Để phân biệt giữa cho ăn và đứng, kết quả phân loại và vị trí của bò đã được phân tích. Kết quả hành vi ăn của Wang và cộng sự [15] độ

chính xác 0,75, kết quả đúng là độ chính xác 0,75. Những kết quả này thấp hơn trường hợp xấu nhất của chúng tôi. Nó chứng minh rằng giải pháp kết hợp cảm biến chân và cảm biến cổ của chúng tôi hoạt động tốt hơn so với giải pháp cảm biến chân và cảm biến vị trí của Wang và cộng sự [15]. Hơn nữa, giải pháp của Wang và cộng sự [15] chỉ phù hợp với bò nuôi ở chuồng vì vị trí cho ăn đã cố định. Trong trường hợp này, sự nhầm lẫn có thể xuất hiện nếu bò đang đứng ở vị trí cho ăn cố định.

4. KẾT LUẬN

Nhằm mục đích nâng cao hiệu suất của hệ thống phân loại hành vi bò từ dữ liệu gia tốc đồng bộ thu được từ gia tốc kế gắn trên chân và cổ, chúng tôi đề xuất sử dụng thuật toán Rừng ngẫu nhiên/Random Forest với bộ đặc trưng giá trị trung bình, độ lệch chuẩn, giá trị hiệu dụng và lựa chọn giá trị tối ưu cho hai tham số: $n_estimators = 25$ (số lượng cây) và $max_depth = 15$ (độ sâu cây) giúp mô hình đạt độ chính xác 0.96. Điều này giúp cải tiến hiệu suất phân loại tốt hơn, phù hợp triển khai trên vi mạch phần cứng. Trong tương lai, chúng tôi tiếp tục nghiên cứu các vấn đề ảnh hưởng tới kết quả phân loại của mô hình, ví dụ như vấn đề cảm biến bị đeo lệch để giúp cải tiến tốt mô hình hơn nữa và triển khai mô hình phân loại được nhiều hành vi hơn.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số: **02/2022/TN**. Trần Đức Nghĩa được tài trợ bởi Chương trình học bổng sau tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), mã số **VINIF.2022.STS.38**.

TÀI LIỆU THAM KHẢO

- [1]. Rutten, C.J., Steeneveld, W., Vernooij, J.C.M., Huijps, K., Nielen, M., Hogeveen, H., "A prognostic model to predict the success of artificial insemination in dairy cows based on readily available data", *J. Dairy Sci.*, Vol. 99, No. 8, pp. 6764–6779, (2016). <https://doi.org/10.3168/jds.2016-10935>
- [2]. Benaissa, S., Tuytens, F.A.M., Plets, D., Catrysse, H., Martens, L., Vandaele, L., Joseph, W., Sonck, B., "Classification of ingestive-related cow behaviours using RumiWatch halter and neck-mounted accelerometers", *Appl. Animal Behaviour Sci.*, Vol. 211, pp. 9–16, (2018). <https://doi.org/10.1016/j.applanim.2018.12.003>
- [3]. Borchers, M.R., Chang, Y.M., Tsai, I.C., Wadsworth, B.A., Bewley, J.M., "A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors", *J. Dairy Sci.*, Vol. 99, No. 9, pp. 7458–7466, (2016). <https://doi.org/10.3168/jds.2015-10843>.
- [4]. Khanh, P.C.P., Dinh Chinh, N., Cham, T.T., Vui, P.T., Tan, T.D., "Classification of cow behavior using 3-DOF accelerometer and decision tree algorithm", In: 2016 International Conference on Biomedical Engineering (BME-HUST), IEEE, Hanoi, Vietnam, pp. 45–50, (2016). <https://doi.org/10.1109/BME-HUST.2016.7782100>.
- [5]. I. Halachmi, "Precision livestock farming applications", Wageningen Academic Publishers, vol. 10:9, pp. 1482–1483, (2016).
- [6]. K. Fogsgaard, C. Røntved, P. Sørensen, M. Herskin, "Sickness behavior in dairy cows during *Escherichia coli* mastitis", *Int. J. Dairy Sci.*, vol. 95, pp. 630–638, (2012).
- [7]. J. Siivonen, S. Taponen, M. Hovinen, M. Pastell, B. J. Lensink, S. Pyörälä, L. Hänninen, "Impact of acute clinical mastitis on cow behaviour", *Appl. Anim. Behav. Sci.*, vol. 132, pp. 101–106, (2011).
- [8]. T. Halasa, K. Huijps, O. Østerås, H. Hogeveen, "Economic effects of bovine mastitis and mastitis management: a review", *Veterinary Quarterly*, Vol. 29, No. 1, pp. 18–31, (2011).
- [9]. P. Sepulveda-Varas, K. L. Proudfoot, D. M. Weary, M. A.G. von Keyserlingk, "Changes in behaviour of dairy cows with clinical mastitis", *Appl. Anim. Behav. Sci.*, Vol. 175, pp. 8–13, (2016).
- [10]. Watanabe, N., Sakanoue, S., Kawamura, K., Kozakai, T., "Development of an automatic classification system for eating, ruminating and resting behavior of cattle using an accelerometer", *Grassland Sci.*, Vol. 54, No. 4, 231–237, (2008).
- [11]. Cong Phi Khanh, P., Tran, D. T., Van Duong, T., Hong Thinh, N., and Tran, D. N., "The new design of cows' behavior classifier based on acceleration data and proposed feature set", *Mathematical Biosciences and Engineering*, Vol. 17, No. 4, pp. 2760-2780, (2020).
- [12]. Duc-Nghia Tran, Phung Cong Phi Khanh, Vijender Kumar Solanki, and Duc-Tan Tran, "A robust

- classification system for Southern Yellow cow behavior using 3-DoF accelerometers*", Journal of Intelligent & Fuzzy Systems, Vol. 43, No. 2, pp. 2211-2218, (2022), DOI: 10.3233/JIFS-219319
- [13]. Vazquez ´ Diosdado, J.A., Barker, Z.E., Hodges, H.R., Amory, J.R., Croft, D.P., Bell, N.J., Codling, E.A., "Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system", Anim. Biotelem. Vol. 3, No. 1, (2015), <https://doi.org/10.1186/s40317-015-0045-8>.
- [14]. Nguyen Thi Thu, To-Hieu Dao, Bo Quoc Bao, Duc-Nghia Tran, Pham Van Thanh and Duc-Tan Tran, "Real-Time Wearable-Device Based Activity recognition Using Machine Learning Methods", International Journal of Computing and Digital Systems, Vol. 5, No.3, pp. 189–201, (2022).
- [15]. J. Wang, Z. He, "Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data", PLoS One, vol. 13, (2018).

ABSTRACT

Cow behavior classification: The optimal set of parameters for the random Forest algorithm

Accelerometer data are key in animal behavior classification systems using pet-mounted accelerometers. Behavioral data reflecting the health status, early detection of some diseases of cows, monitoring the health of cows through behavior are effective support tools for ranchers to help improve performance and save money cost. In the previous study, we proposed feature sets, data windows and used a random forest algorithm to classify four important cow behaviors, including: eating, lying, standing and walking. In this study, in order to improve the performance of the classification system, we propose to use suitable values for the important parameter set of the random forest algorithm on the experimental data set. The experimental results show that with the value of the parameter set: number of trees = 25 and depth = 15, the classification performance is good with an accuracy of 95,9%.

Keywords: Classification; Monitoring; Accelerometer; Behavior; Cow; Random forest.