



## NHẬN DẠNG HÀNH VI CỦA BÒ SỬ DỤNG GIA TỐC KẾ VÀ GIẢI THUẬT MÁY HỌC RỪNG NGẪU NHIÊN

Lê Đình Chiến<sup>1</sup>, Lê Văn Lâm<sup>2</sup> và Trần Công Án<sup>1</sup>

<sup>1</sup>Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

<sup>2</sup>Phòng Hợp tác Quốc tế, Trường Đại học Cần Thơ

### Thông tin chung:

Ngày nhận bài: 15/09/2017

Ngày nhận bài sửa: 10/10/2017

Ngày duyệt đăng: 20/10/2017

### Title:

Cow behaviour recognition using accelerometer and random forest algorithm

### Từ khóa:

Bò, gia tốc kế, nhận dạng hành vi, rừng ngẫu nhiên

### Keywords:

Accelerometer, behaviour, cow, random forests, recognition

### ABSTRACT

Cattle behaviour patterns provide significant information about cattle health. Therefore, early behaviour recognition may help breeders be aware of cattle health problems promptly to have appropriate treatment to reduce negative impact. In this paper, an approach to cow behaviour recognition based on accelerated data will be proposed. The behaviour recognition model is built using random forest algorithm. This study focuses on four popular behaviours, i.e. walking, standing, eating (grazing), and lying. The model is validated using a real cow activity dataset. The overall classification result of the model is about 95% of accuracy. The comparison on the classification result with other recent approaches is provided. It is shown that the proposed approach in this paper is promising, and it can be used for developing cow behavior recognition application.

### TÓM TẮT

Mẫu hành vi của bò cung cấp thông tin rất quan trọng về sức khỏe, tình trạng của bò. Do đó, phát hiện sớm các hành vi của bò có thể giúp người chăn nuôi nhận biết các vấn đề đối với sức khỏe của bò một cách kịp thời; qua đó, có thể sử dụng các biện pháp điều trị thích hợp để giảm thiểu các thiệt hại. Bài báo này sẽ đề xuất một phương pháp phát hiện hành vi của bò dựa vào dữ liệu gia tốc kế. Mô hình nhận dạng hành vi được xây dựng dựa trên giải thuật máy học rừng ngẫu nhiên (random forest). Nghiên cứu này sẽ tập trung nhận dạng 4 hành vi: đi lại, đứng, ăn (gặm cỏ) và nằm. Mô hình nhận dạng được đánh giá trên 1 bộ dữ liệu thật do chúng tôi thu thập. Kết quả phân loại hành vi của mô hình là 95%. Chúng tôi cũng so sánh kết quả thực nghiệm với một số phương pháp đã được đề xuất trong các nghiên cứu khác. Kết quả so sánh cho thấy phương pháp được đề xuất là khả thi và có thể áp dụng để xây dựng các ứng dụng nhận dạng hành vi của bò.

Trích dẫn: Lê Đình Chiến, Lê Văn Lâm và Trần Công Án, 2017. Nhận dạng hành vi của bò sử dụng gia tốc kế và giải thuật máy học rừng ngẫu nhiên. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Công nghệ thông tin: 80-87.

## 1 GIỚI THIỆU

Sức khỏe của bò ảnh hưởng rất lớn đến năng suất (sữa, thịt) của chúng. Có nhiều nghiên cứu về

mối liên hệ giữa các mẫu hành vi (behaviour pattern) với sức khỏe của bò. Các nghiên cứu này cho rằng, các mẫu hành vi của bò, hay động vật nói chung, cung cấp các thông tin rất quan trọng về sức

khỏe của chúng (Doyle *et al.*, 2015, Mench, 1998). Do đó, đối với những người chăn nuôi bò thì việc theo dõi mẫu hành vi của bò được đặc biệt quan tâm vì chúng có thể giúp họ phát hiện sớm các vấn đề về sức khỏe của bò; từ đó, đưa ra các biện pháp phòng ngừa hay điều trị sớm để giảm thiểu các thiệt hại (Sterrett *et al.*, 2014).

Có nhiều phương pháp được đề xuất hay các hệ thống được quan tâm xây dựng nhằm giám sát hoạt động của bò. Phương pháp sơ khai nhất là quan sát trực tiếp hành vi của bò một cách thủ công. Đối với khả năng của con người thì có thể nhận dạng hành vi của chúng một cách chính xác. Tuy nhiên, phương pháp này trên thực tế là không khả thi vì sự hiện diện của con người có thể ảnh hưởng đến hành vi của gia súc. Ngoài ra, việc quan sát một cách liên tục hay với số lượng lớn các con bò hoặc trong một không gian lớn (trên 1 cánh đồng hay trang trại) thì phương pháp này rõ ràng là không khả thi.

Do đó, có 1 phương pháp khác được sử dụng rộng rãi hơn là dùng các cảm biến đặt trên các con bò cần theo dõi hành vi. Một trong những loại cảm biến thông dụng nhất được đề xuất sử dụng trong nhiều nghiên cứu về theo dõi hành vi của bò là gia tốc kế.

Bonk Stephanie *et al.* (2013) tập trung vào hành động nằm của bò. Các tác giả muốn kiểm tra độ chính xác của việc sử dụng gia tốc kế Hobo Pendant G data logger (Hobbo, 2017) để tính tổng thời gian nằm của bò và số lần nằm của bò trong ngày. Gia tốc kế được gắn vào chân sau của bò để thu thập dữ liệu. Dữ liệu này sẽ được xử lý một cách thủ công bằng phần mềm thống kê SPSS để tính phát hiện hành vi nằm của bò; từ đó, tính ra thời gian nằm và số lần nằm của bò trong một khoảng thời gian nào đó. Để tính độ chính xác của phương pháp này, các tác giả đã sử dụng một camera để quay các hoạt động của bò. Dữ liệu từ camera sẽ được so sánh với kết quả phân tích để đánh giá độ chính xác. Kết quả thực nghiệm cho thấy độ chính xác của phương pháp sử dụng gia tốc kế Hobo Pendant G data logger kết hợp với sự phân tích dữ liệu một cách thủ công như trong nghiên cứu có độ chính xác rất cao với predictability, sensitivity, và specificity đều đạt trên 97%.

Martiskainen *et al.* (2009) có tính tổng quát hơn, có thể phát hiện nhiều hành vi của bò hơn. Các tác giả đề xuất phương pháp phân lớp hành vi của bò sử dụng một gia tốc kế 3 chiều đeo trên cổ bò và giải thuật máy học vector hỗ trợ (SVM). Có tất cả 8 hành vi được quan tâm trong nghiên cứu bao gồm đứng, nằm, chạy, ăn, đi bình thường, đi

khập khiễng, nằm xuống và đứng lên. Độ chính xác trung bình của phương pháp này cho 8 hành vi là khoảng 78%, trong đó hành vi có độ chính xác cao nhất là chạy (86%) và thấp nhất là nằm xuống (0%).

Trong nghiên cứu của Diosdado Jorge A. Vázquez *et al.* (2015), các tác giả đề xuất sử dụng một gia tốc kế đặt ở cổ của bò. Tần số thu thập dữ liệu là 50Hz. Giải thuật dùng để xây dựng mô hình phân loại hành vi là giải thuật cây quyết định (decision tree). Các hành vi được phân loại trong mô hình này bao gồm nằm (77,42% sensitivity, 98,63% precision), ăn (98,78% sensitivity, 93,10% precision), và đứng (88,00% sensitivity, 55,00% precision). Ngoài ra, mô hình nhận dạng trong nghiên cứu này còn cho phép phát hiện sự thay đổi hành vi từ đứng sang nằm với 96,45% sensitivity và 87,5% precision. Các tác giả cũng so sánh kết quả của giải thuật cây quyết định với một số giải thuật khác như máy vector hỗ trợ (SVM), Hidden Markov Model (HMM) và K-Mean.

Kết quả thực nghiệm cho thấy giải thuật cây quyết định tốt hơn các giải thuật được so sánh trong hầu hết các trường hợp.

Robert *et al.* (2009) sử dụng gia tốc kế để phân biệt các hành vi nằm, đứng và đi. Tần suất lấy mẫu được thử nghiệm với 3 trường hợp là 3, 5 và 10 giây. Độ chính xác tốt nhất cho 3 hành vi trên lần lượt là 99,2%, 98% và 67,8%.

Tại Việt Nam, nghiên cứu của Nguyen Thi Huyen Nga *et al.* (2016) đã phát triển một thiết bị để giám sát và phát hiện hành vi của bò. Thiết bị này bao gồm một mạch Arduino (Uno), trên đó có gắn cảm biến gia tốc (ADXL345) và một SD card. Thiết bị này được gắn vào cổ của bò để thu thập dữ liệu gia tốc và lưu vào SD card. Các dữ liệu này sẽ được phân tích một cách thủ công để phân biệt các hành vi đứng, nằm và ăn của bò. Tuy nhiên, độ chính xác của mô hình chưa được kiểm tra.

Ngoài ra, còn nhiều nghiên cứu khác liên quan đến nhận dạng hành vi của bò như Laca *et al.*, 2000 sử dụng một thiết bị quay video và một microphone không dây để nhận dạng 2 hành vi cắn và nhai với độ chính xác 94%. Hay Gibb *et al.*, 1998 sử dụng một thiết bị gọi là IGER để thu lại tiếng nhai cỏ nhằm tính toán số lần gặm cỏ, thời gian mỗi lần gặm cỏ và tổng thời gian gặm cỏ trong ngày của bò.

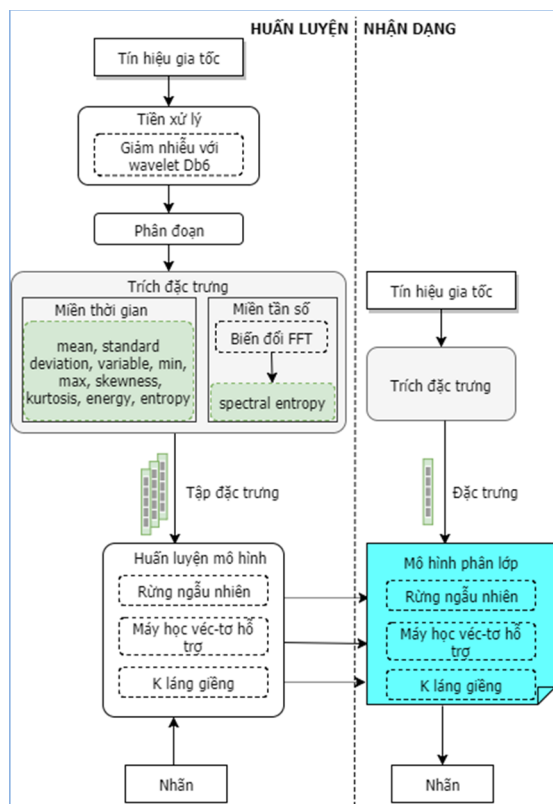
Trong nghiên cứu này, chúng tôi đề xuất một phương pháp phát hiện hành vi của bò dựa trên gia tốc kế sử dụng giải thuật rừng ngẫu nhiên (random forest). Lý do chúng tôi đề xuất giải thuật này là vì giải thuật này đã được sử dụng để phân loại hành

vi của người khá thành công. Tuy nhiên, trong phân loại hành vi của bò thì vẫn chưa có nghiên cứu nào sử dụng giải thuật này. Nghiên cứu này sử dụng 1 gia tốc kế gắn trên chân bò để phát hiện các hành vi đi, đứng, ăn (gặm cỏ) và nằm. Chúng tôi cũng sẽ so sánh kết quả của phương pháp này với các nghiên cứu trước đó nhằm đánh giá phương pháp được đề xuất.

Phần tiếp theo của bài báo được tổ chức như sau: Phần 2 sẽ mô tả chi tiết phương pháp nhận dạng hành vi của bò bằng gia tốc kế và giải thuật rừng ngẫu nhiên; Phần 3 trình bày kết quả thực nghiệm; cuối cùng, chúng tôi sẽ rút ra các kết luận cũng như đề xuất các định hướng phát triển của nghiên cứu trong phần 4.

## 2 NHẬN DẠNG HÀNH VI CỦA BÒ BẰNG GIA TỐC KẾ VÀ GIẢI THUẬT RỪNG NGẪU NHIÊN

Mô hình tổng quát của phương pháp nhận dạng hành vi của bò bằng gia tốc kế được mô tả trong Hình 1.



Hình 1: Mô hình nhận dạng hành vi của bò

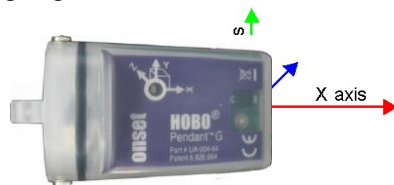
Từ dữ liệu của gia tốc kế (dữ liệu huấn luyện), hệ thống sẽ thực hiện các bước tiền xử lý để loại bỏ các dữ liệu nhiễu. Sau đó, dữ liệu huấn luyện sẽ được phân thành các đoạn có kích thước bằng nhau. Để tăng độ chính xác phân lớp thì chúng tôi sẽ bổ sung một số đặc trưng vào tập dữ liệu. Tập

dữ liệu đã được phân đoạn cùng với các đặc trưng mới sẽ được huấn luyện bởi giải thuật máy học rừng ngẫu nhiên để xây dựng mô hình phân lớp. Mô hình phân lớp này sẽ được sử dụng để nhận dạng hành vi của bò từ các dữ liệu gia tốc kế.

Chi tiết của các bước trong mô hình sẽ được mô tả trong phần sau.

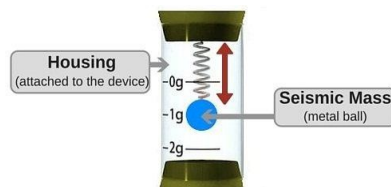
### 2.1 Gia tốc kế

Gia tốc kế là một thiết bị dùng để đo sự biến đổi gia tốc của đối tượng mang thiết bị. Các loại gia tốc kế thông dụng hiện tại đều có khả năng đo sự biến đổi gia tốc theo cả 3 chiều: x, y, và z. Các cảm biến gia tốc vi cơ được chế tạo theo công nghệ MEMS có nhiều ưu điểm so với các gia tốc kế chế tạo bằng công nghệ điện tử trước đây. Loại cảm biến này ngày càng nhanh hơn, nhạy hơn, nhẹ hơn, rẻ hơn và quan trọng hơn cả là có độ tin cậy rất cao. Hình 2 minh họa một gia tốc kế và chiều của các trục tương ứng.



Hình 2: Gia tốc kế và chiều của các trục

Nguyên lý để đo gia tốc theo trục y được mô tả trong Hình 3.



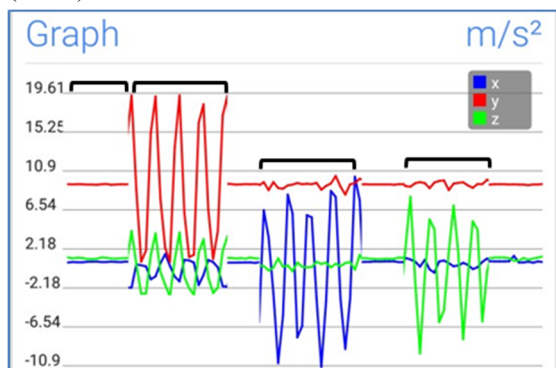
Hình 3: Nguyên lý đo gia tốc theo trục y

(Source: <https://goo.gl/gB74Kj>)

Khi đặt gia tốc kế thẳng đứng theo trục y, do tác động của trọng lực thì khối lượng chuyển động (seismic mass) sẽ bị kéo xuống và giá trị chuẩn của trạng thái này là +1G. Khi di chuyển khoang chứa (housing) lên xuống theo phương thẳng đứng thì khối lượng chuyển động sẽ di chuyển, dẫn đến giá trị của y sẽ thay đổi. Độ biến thiên của y phụ thuộc vào việc gia tốc chuyển động của khoang chứa theo chiều thẳng đứng. Một gia tốc nhiều chiều sẽ bao gồm nhiều đơn vị đo gia tốc trong Hình 3 được đặt theo các hướng tương ứng.

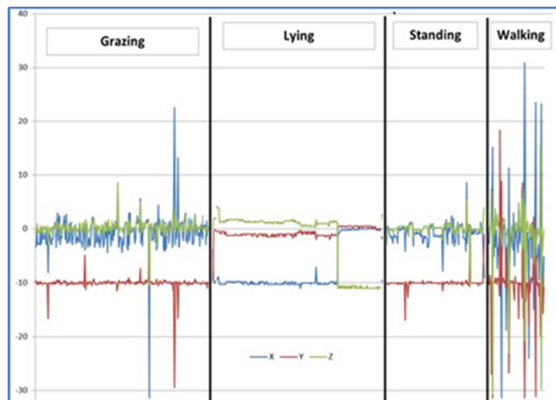
Hình 4 minh họa giá trị của x, y, z trong một số trường hợp khác nhau: (1) là trường hợp đặt gia tốc kế đứng yên theo phương thẳng đứng của trục y; (2) là trường hợp di chuyển gia tốc kế lên xuống theo phương thẳng đứng (trục y); (3) là trường hợp

di chuyển gia tốc kể qua lại theo chiều ngang (trục x); và (4) là trường hợp di chuyển gia tốc kể tới lui (trục z).



Hình 4: Sự biến thiên giá trị x, y và z

Trong nghiên cứu này, gia tốc kế sẽ được gắn vào chân sau của bò với chiều của các trục như được mô tả trong Hình 2. Giá trị x, y, z của một số hành vi được minh họa trong Hình 5. Dữ liệu minh họa cho thấy việc phân biệt hành vi dựa vào gia tốc kế là hoàn toàn khả thi. Tuy nhiên, có một số trường hợp dễ nhầm lẫn là đi và gặm cỏ.



Hình 5: Giá trị x, y, z của một số hành vi

## 2.2 Các hành vi của bò

Trong nghiên cứu này sẽ tập trung vào các hành vi gồm đi, đứng, ăn (gặm cỏ) và nằm được mô tả như sau:

**Bảng 1: Hành vi sử dụng cho nhận dạng**

Hành vi	Mô tả
Đi	Di chuyển về phía trước, ít nhất là 2 chân trên mặt đất
Đứng	Tất cả bốn chân trên mặt đất
Ăn	Có chuyển động, nhưng rất chậm
Nằm	Thân tiếp xúc đầy đủ với mặt đất, bên trái hoặc bên phải

## 2.3 Xây dựng mô hình nhận dạng hành vi

Đầu vào của việc xây dựng mô hình nhận dạng là dữ liệu gia tốc kế 3 chiều x, y, z nằm trong miền

giá trị  $[-3G, 3G]$  của các hành vi thu thập từ các gia tốc kế gắn trên chân bò. Các bước thực hiện huấn luyện mô hình bao gồm.

### 2.3.1 Tiền xử lý

Dữ liệu thu thập được có thể có nhiễu, mất hoặc sai giá trị ở một số đoạn, hoặc sự trôi (drift) (Doyle *et al.*, 2015). Vì vậy, cần phải khử nhiễu ra khỏi dữ liệu.

Có nhiều giải thuật khử nhiễu như các bộ lọc đơn giản: trung bình dịch chuyển, trung vị, thông cao, thông thấp, dải thông. Hoặc các bộ lọc phức tạp hơn như bộ lọc Butterworth thông cao hoặc thấp, savitzky-golay, kalman và wavelet. Các bộ lọc này được coi là các bộ lọc tối ưu (Wang *et al.*, 2011). Trong nghiên cứu này, chúng tôi sử dụng bộ lọc db6 thuộc họ mẹ wavelet vì cho kết quả khử nhiễu tốt so với các bộ lọc khác (Parsons *et al.*, 1999).

Các bước khử nhiễu của phương pháp wavelet được mô tả như sau:

1) Phân tích tín hiệu, lựa chọn wavelet, chọn mức N và tính wavelet thuận

$$C(a, b) = \int_R s(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt$$

$$a = 2^j, b = k2^j, (j, k) \in \mathbb{Z}^2$$

2) Lấy ngưỡng các hệ số chi tiết, với mọi mức từ 1 tới N, chọn ngưỡng và áp dụng lấy ngưỡng mềm hay cứng vào các hệ số chi tiết.

3) Khôi phục, tính wavelet ngược.

$$s(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C(j, k) \psi_{j,k}(t)$$

Chi tiết của các công thức trên bước 1) và 3) có thể được tham khảo trong nghiên cứu (Wang *et al.*, 2011). Cách chọn ngưỡng tại bước 2) có thể tính theo công thức của Donoho và Jonhstone (Donoho *et al.*, 1994).

Kết quả khử nhiễu phụ thuộc nhiều vào hệ số lấy ngưỡng. Trong nghiên cứu này, chúng tôi áp dụng họ wavelet là Daubechies (db6) ở mức 2 và chọn ngưỡng mềm vì sự đơn giản và phổ biến trong nhiều nghiên cứu (Breiman *et al.*, 1984).

### 2.3.2 Phân đoạn

Chuỗi dữ liệu cảm biến được phân đoạn thành các khung có kích thước bằng nhau. Trong đó, chiều dài của chúng có thời lượng dài hơn khoảng thời gian cho một hoạt động bất kỳ. Chiều dài một đoạn thường là số chẵn: 8, 16 ..., 128, 512. Nhiều nghiên cứu trước đó đã cho thấy độ dài của cửa sổ trượt có ảnh hưởng quan trọng tới hiệu năng của

thuật toán nhận dạng hoạt động. Vì thế, chiều dài của phân đoạn sẽ được lựa chọn từ thực nghiệm bằng cách thay đổi từ nhỏ đến lớn và chọn giá trị cho kết quả tốt nhất.

### 2.3.3 Rút trích đặc trưng

Rút trích đặc trưng nhằm rút gọn các tín hiệu thành các đặc trưng để phân biệt các hoạt động hiệu quả hơn. Trong nghiên cứu này, chúng tôi sử dụng 8 đặc trưng cộng với các đặc trưng x, y và z từ dữ liệu thô. Các đặc trưng này có thể được chia làm 2 loại: các đặc trưng miền thời gian và các đặc trưng miền tần số.

#### Đặc trưng miền thời gian

Các thống kê đơn giản nhằm phân biệt dễ dàng các hoạt động gồm: giá trị trung bình (mean), giá trị lớn nhất (max), giá trị nhỏ nhất (min), độ lệch chuẩn (standard deviation), độ bất đối xứng (skewness), độ nhọn (kurtosis), năng lượng (energy), độ tương quan (correlation).

#### Đặc trưng miền tần số

Tín hiệu miền thời gian được chuyển về miền tần số bằng chuyển đổi Fourier rồi lấy, sau đó tính năng lượng (energy) và spectral entropy (Daubechies, 1992, Lee *et al.*, 2008). Các đặc trưng này được tính qua 5 bước:

Bước 1: Fast Fourier Transform

$$X(i, n) = \sum_{m=1}^M x(m, n) e^{-j \frac{2\pi}{N} im}$$

Bước 2: Phổ năng lượng (Power spectral)

$$S(i, n) = |X(i, n)|^2$$

Bước 3: Mật độ phổ năng lượng (Power spectral density)

$$P[S(i, n)] = \frac{S(i, n)}{\sum_{m=1}^{M/2} S(m, n)}$$

Bước 4: Spectral entropy theo công thức Shannon's entropy

$$H_{FFT}(n) = - \sum_{i=1}^{M/2} P[S(i, n)] \log_2 P[S(i, n)]$$

Bước 5: Chuẩn hóa spectral entropy

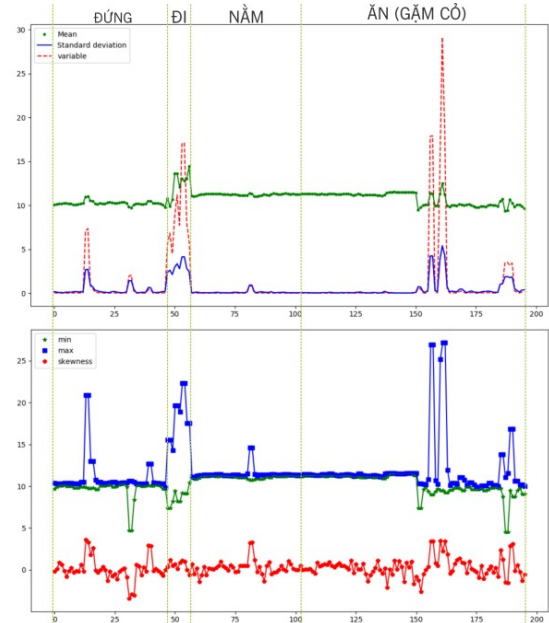
$$H_X(n) = \frac{H_{FFT}}{\log_2(n)}$$

Chi tiết của các công thức trong các bước trên có thể được tham khảo trong (Daubechies, 1992). Hình 6 minh họa một số đặc trưng tương ứng với các hành vi của bò.

### 2.3.4 Phân loại hành vi với thuật toán rừng ngẫu nhiên

Trong các nghiên cứu nhận dạng hoạt động của con người cho thấy thuật toán rừng ngẫu nhiên cho phân lớp có kết quả chính xác khá cao. Tuy nhiên, trong phân lớp hành vi của bò thì chưa có nghiên cứu nào áp dụng. Do đó, chúng tôi đề xuất sử dụng thuật toán này để xây dựng mô hình phân lớp trong nghiên cứu này. Chúng tôi sẽ so sánh kết quả của thuật toán này với một số thuật toán được sử dụng trong các nghiên cứu đã thực hiện về nhận dạng hành vi của bò để đánh giá hiệu quả của giải thuật.

Brieman (2001) đề xuất giải thuật Random Forest (RF) thực hiện phân lớp và hồi quy, RF dựa trên việc kết hợp kết quả dự đoán của một số lượng lớn các cây quyết định. Trong mô hình RF truyền thống mỗi cây quyết định được xây dựng từ tập dữ liệu được lấy ngẫu nhiên từ tập dữ liệu ban đầu và việc phát triển các nút con từ một nút cha dựa trên thông tin trong một không gian con các thuộc tính được chọn ngẫu nhiên từ không gian thuộc tính ban đầu. Do đó, RF xây dựng các cây quyết định từ một tập con những thuộc tính được lựa chọn ngẫu nhiên và tổng hợp kết quả dự đoán của các cây để tạo ra kết quả dự đoán cuối cùng.



Hình 6: Minh họa một số đặc trưng

Các cây quyết định được xây dựng sử dụng thuật toán CART (Brieman, 1984) mà không thực hiện việc cắt tỉa do đó thu được những cây với độ lệch thấp. Bên cạnh đó, mối quan hệ tương quan giữa các cây quyết định cũng được giảm thiểu nhờ việc xây dựng các không gian con thuộc tính một cách ngẫu nhiên. Do đó, việc kết hợp kết quả của một số lượng lớn những cây quyết định độc lập có

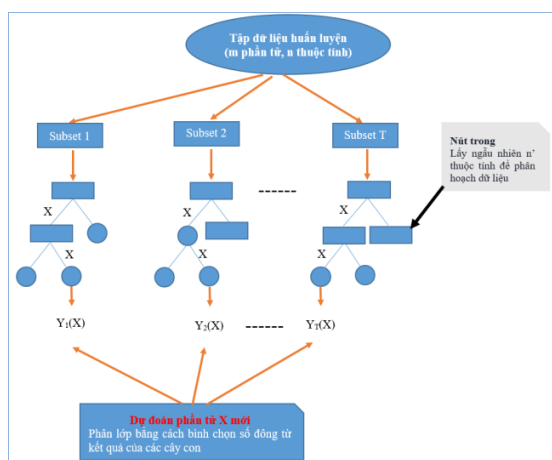
độ lệch thấp, phương sai thấp sẽ giúp RF đạt được cả độ lệch thấp và phương sai thấp.

Quá trình xây dựng rừng ngẫu nhiên được trình bày tại Hình 7 và gồm các bước chính sau:

1) Lấy ra T mẫu Subset (bootstrap) từ tập huấn luyện.

2) Đối với mỗi mẫu Subset xây dựng một cây phân lớp không được tỉa (unpruned tree) theo hướng dẫn sau: Tại mỗi nút thay vì chọn một phân chia tốt nhất trong tất cả các biến dự đoán, ta chọn ngẫu nhiên một mẫu m của các biến dự đoán sau đó chọn một phân chia tốt nhất trong các biến này.

3) Đưa ra các dự đoán bằng cách tổng hợp số đông các dự đoán của T cây.



Hình 7: Giải thuật xây dựng rừng ngẫu nhiên

### 3 KẾT QUẢ THỰC NGHIỆM

#### 3.1 Tập dữ liệu thực nghiệm

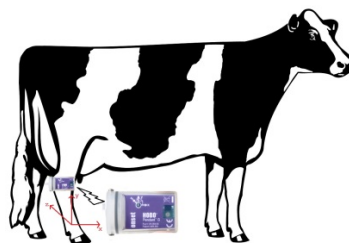
Trong nghiên cứu này, chúng tôi thu thập dữ liệu thực nghiệm bằng gia tốc kế HOB0 (Hobo, 2017). Chi tiết kỹ thuật chính của gia tốc kế này và cấu hình cài đặt để thu thập dữ liệu được mô tả trong Bảng 2.

Đối tượng dùng để thu thập dữ liệu là một con bò đã trưởng thành. Gia tốc kế được gắn vào chân phải của bò với trục y hướng lên trên, trục x song song với mặt đất và trục z vuông góc với 2 trục còn lại và hướng về bên trong của bò như được mô tả trong Hình 8.

Sau khi thiết lập thông số thì chúng bắt đầu lấy mẫu các trường hợp đi, đứng, ăn (gặm cỏ) và nằm của bò. Bò được gắn gia tốc kế sẽ được thả ra đồng để thực hiện các thao tác thường ngày của chúng.

Bảng 2: Thông số của gia tốc kế

Chi tiết	Giá trị
Độ nhạy	[-3G, 3G]
Tần suất lấy mẫu	0.1Hz (10 giây/lần)
Sai số về thời gian	±1 phút/tháng
Thời lượng pin	7 ngày đến 1 năm
Bộ nhớ	64Kb
Đơn vị gia tốc	m/s



Hình 8: Vị trí đặt và hướng của gia tốc kế

Một quan sát viên trực tiếp ghi nhận hoạt động của bò theo thời gian. Một quan sát viên khác sẽ quay lại các hoạt động của bò. Ngoài ra, còn một quan sát viên theo dõi hoạt động và có thể điều khiển để con bò thực hiện các hành vi mong muốn. Tập dữ liệu được ghi nhận trong tổng khoảng thời gian là 4 giờ 20 phút, bao gồm 1562 mẫu hành vi (đứng: 370; đi: 80; nằm: 752 và ăn: 360). Mỗi mẫu dữ liệu có 4 thuộc tính: thời gian, giá trị x, y, z.

Tập dữ liệu này sẽ được gán nhãn dựa vào các ghi chép của quan sát viên và video ghi lại các hoạt động của bò. Sau khi bước tiền xử lý, khử nhiễu dữ liệu bằng phương pháp wavelet, chúng tôi tiến hành tạo đặc trưng với kích thước của số là 16 điểm dữ liệu và nạp chồng ½ cửa sổ. Chúng tôi tạo 3 tập dữ liệu thực nghiệm với số đặc trưng khác nhau như mô tả trong Bảng 3.

Bảng 3: Mô tả các tập dữ liệu thực nghiệm

Tên tập dữ liệu	Mô tả
XYZ	27 đặc trưng: gồm 3 đặc trưng thô x, y, z và 8 đặc trưng mở rộng (mean, std, var, min, max, skew, kurtosis, energy, entropy) cho mỗi đặc trưng thô
V	9 đặc trưng: gồm đặc trưng V và 8 đặc trưng mở rộng của V
XYZV	36 đặc trưng: tổng hợp tất cả các đặc trưng của tập XYZ và V

### 3.2 Kết quả thực nghiệm

Chúng tôi sử dụng kỹ thuật kiểm tra chéo 5 lần (5-fold cross validation) để đánh giá độ chính xác của các giải thuật phân lớp. Độ chính xác (Accuracy) được tính theo công thức:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Với

TP: số hành vi được phân loại lớp hành vi

FP: số hành vi không phải hành vi được phân loại vào lớp hành vi

FN: số hành vi được phân loại vào lớp không phải hành vi

TN: số hành vi không phải hành vi được phân vào lớp không phải hành vi

Kết quả trung bình cộng của 5 fold được trình bày trong Bảng 4.

**Bảng 4: Kết quả thực nghiệm**

Tập các đặc trưng	ĐỘ CHÍNH XÁC (%)		
	SVM (nhân RBF)	K láng giềng	Rừng ngẫu nhiên
XYZ (27 chiều)	94,90	92,83	<b>96,94</b>
V (9 chiều)	96,94	91,83	<b>95,38</b>
XYZV (36 chiều)	97,45	92,33	<b>97,95</b>

*Các tham số chính (bằng cách thử sai, do tìm trong khoảng nào đó, tìm giá trị tham số tối ưu)*

**SVM (nhân RBF):** tập V chọn  $c = 1000$ ,  $\Gamma = 10$  hoặc  $c = 10000$ ,  $\gamma = 0.01$ ; XYZ và XYZV cùng chọn  $c = 100$ ,  $\gamma = 0.01$ .

**K láng giềng:** tập V chọn  $k = 5$ ; XYZ chọn  $k = 12$  và XYZV chọn  $k = 11$ ;  $d = \text{Euclidean}$  và có trong số.

**Rừng ngẫu nhiên:** Số cây = 300; số thuộc tính để tính toán phân hoạch =  $\sqrt{\text{số đặc trưng}}$ ; đường phân hoạch = entropy.

Các kết quả thực nghiệm ở Bảng 4 cho thấy mô hình nhận dạng hành vi của con Bò dựa trên dữ liệu gia tốc được trích đặc trưng trong miền thời gian và miền tần số, kết hợp với rừng ngẫu nhiên cho độ chính xác cao và đáng tin cậy.

## 4 KẾT LUẬN

Bài báo đã đề xuất một mô hình nhận dạng hành vi của bò bao gồm: đi, đứng, ăn (gặm cỏ) và nằm sử dụng một thiết bị đo gia tốc và mô hình máy học rừng ngẫu nhiên. Mô hình được kiểm thử

trên một tập dữ liệu gồm 1.562 mẫu với độ chính xác kiểm thử cao nhất với 97,95%. Kết quả này tốt hơn 2 giải thuật máy học véc-tơ hỗ trợ và k-láng giềng trên cùng tập dữ liệu. Với độ chính xác này thì việc áp dụng mô hình này vào các ứng dụng nhận dạng hành vi của bò là hoàn toàn khả thi.

Trong tương lai, chúng tôi tiếp tục nghiên cứu cải thiện độ chính xác phân lớp, tăng cường đặc trưng có lợi cho mô hình phân lớp thông qua học đặc trưng tự động, thông qua ý kiến chuyên gia và đánh giá cụ thể bộ lọc tín hiệu phù hợp. Ngoài ra, chúng tôi cũng sẽ tiếp cận tăng số loại hành vi hơn có thể nhận dạng và thử nghiệm trên dữ liệu lớn, đi đến việc phát triển ứng dụng nhận dạng hành vi của bò.

## TÀI LIỆU THAM KHẢO

- Bonk, Stephanie, Onno Burfeind, V. S. Suthar, and Wolfgang Heuwieser, 2013. Evaluation of data loggers for measuring lying behavior in dairy calves. *Journal of dairy science* 96, No.5: 3265-3271.
- Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen, 1984. Classification and regression trees. CRC press.
- Breiman, Leo, 2011. Random forests. *Machine learning* 45. No.1: 5-32.
- Daubechies, Ingrid, 1992. Ten lectures on wavelets. Society for industrial and applied mathematics.
- Diosdado, Jorge A. Vázquez, et al, 2015. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*. Vol3, No.1.
- Donoho, David L., and Jain M. Johnstone, 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81. No.3: 425-455.
- Doyle, Rebecca, and John Moran, 2015. Cow Talk: Understanding Dairy Cow Behaviour to Improve Their Welfare on Asian Farms. Csiro Publishing
- Gibb, M. J., C. A. Huckle, and R. Nuthall, 1998. Effect of time of day on grazing behaviour by lactating dairy cows. *Grass and Forage Science* 53. No.1: 41-46.
- HOBOPendant G Acceleration Data Logger. [http://www.onsetcomp.com/files/manual\\_pdfs/10872-B-MAN-UA-004.pdf](http://www.onsetcomp.com/files/manual_pdfs/10872-B-MAN-UA-004.pdf) (truy xuất lần cuối: 06/09/2017).
- Jiaying Du, 2016. Signal processing for mems sensor based motion analysis system. Mälardalen University Press Licentiate Theses. 13-14
- Laca, E. A., and M.F. WallisDeVries, 2000. Acoustic measurement of intake and grazing behaviour of cattle." *Grass and Forage Science* 55. No.2: 97-104.
- Lee, W. S., Roh, Y. W., Kim, D. J., Kim, J. H., & Hong, K. S., 2008. Speech emotion recognition using spectral entropy. In *International*

- Conference on Intelligent Robotics and Applications. Springer, Berlin, Heidelberg. 45-54
- Martiskainen P, Jarvinen M, Skon JP, Tiirikainen J, Kolehmainen M, et al., 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science* 119. 32–38.
- Mench, Joy, 1998. Why it is important to understand animal behavior. *ILAR journal* 39. No.1: 20-26.
- Nguyen Thi Huyen Nga and Le Thi Thu Ha and Doan Ba Cuong and Tran Duc Tan, 2016. Monitoring and classification of cow activities using three-dimensional accelerometers. In: *SSC: the 2016 Student Scientific Contest*. Hanoi.
- Parsons, Antony C., W. Mack Grady, and Edward J. Powers, 1999. A wavelet-based procedure for automatically determining the beginning and end of transmission system voltage sags. In *Power Engineering Society 1999 Winter Meeting*. IEEE, vol.2: 1310-1315.
- Quinlan, J. Ross. C4. 5: programs for machine learning. Elsevier, 2014.
- Robert B, White BJ, Renter DG, Larson RL, 2009. Evaluation of three-dimensional accelerometers to monitor and classify behaviour patterns in cattle. *Compu Electron Agric.* 67:80–4.
- S.Bonk, O.Burfeind, V.S.Suthar, W.Heuwieser, 2013. Technical note: Evaluation of data loggers for measuring lying behavior in dairy calves. *Journal of Dairy Science.* 96:5,3265-3271.
- Sterrett, Amanda E., Donna M. Amaral- Phillips, Jeffrey M. Bewley, and Michelle Arnold, 2014. A Fresh Cow Health Monitoring System.
- Wang, W. Z., Guo, Y. W., Huang, B. Y., Zhao, G. R., Liu, B. Q., & Wang, L., 2011-November. Analysis of filtering methods for 3D acceleration signals in body sensor network. In *Bioelectronics and Bioinformatics (ISBB)*. IEEE, 2011 International Symposium: 263-266.