# A 216 µW, 87% Accurate Cow Behavior Classifying Decision Tree on FPGA With Interpolated Arctan2

Jim Bartels[1], Korkut Kaan Tokgoz[2], Masamoto Fukawa[1], Shohei Otsubo[1]
Li Chao[1], Ikumi Rachi[1], Kenichi Takeda[3], Hiroyuki Ito[1]

[1]Department of Electrical and Electronic Engineering, Tokyo Institute of Technology
[2]Research Institute for the Earth Inclusive Sensing (EISESiV), Tokyo Institute of Technology
[3]Faculty of Agriculture, Shinshu University
Email: bartels.j.aa@m.titech.ac.jp, paper@lsi.pi.titech.ac.jp

2021 IEEE International Symposium on Circuits and Systems (ISCAS) | 978-1-7281-9201-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCAS51556.2021.9401342

***Abstract*—This work presents five different 8-bit fixed point low power hardware implementations of cow behavior classifying axis-parallel decision trees (DT) for a range of five feature sets (FS). Each DT is trained and tested with 3-axis accelerometer data labeled beforehand containing four main behaviors, resting, eating, rumination and moving. Investigation shows a 10% F1-score decrement from software (77.4%) to 8-bit fixed point hardware (67.4%) for the conventional feature set, i.e. average mean and variation of the 3D accelerometer vector magnitude (FS4) due to quantization. The horizontal-longitudinal angle of acceleration of the cow (arctan2(y, x)) is proposed as a new FS together with the vertical $z$-axis (FS1&FS2). The results of an experimental setup, which simulates a stratified 20% split of the entire dataset from PC to FPGA, shows an average accuracy of 86.81% respectively for FS2. This FS's hardware implementation programmed on a Lattice ICE40UP5K FPGA has a power consumption of 216 µW with a latency of 2.58 s at the frequency of 2.9 kHz resulting in an energy of 557 µJ per classification, which is competitive with state of the art.**

***Index Terms*—low power hardware, decision tree, cow behavior, FPGA, interpolated arctangent**

## I. Introduction

According to a United Nations report in 2013, 14.5% of all human induced greenhouse gasses originate from livestock, with 65% of these emissions arising from cattle farming [1]. Next to emissions, the livestock sector contributes to waste problems worldwide and has strong indicators of growth in the future [2]. On the other hand, breakthroughs in signal processing, Artificial Intelligence (AI) and sensory systems have enabled an industry of intelligent farming called Precision Agriculture (PA). Precision Livestock Farming (PLF) applies this intelligent farming to animal husbandry, which other than economic reasons, provides ways to mitigate the environmental impact of livestock [3].

A fundamental part of PLF is the monitoring of animal wellness. In the case of applying AI to monitoring, 3D accelerometers have been used extensively with software based Machine Learning (ML) algorithms achieving classification of animal behavior with accuracies of over 85% for at least 3 behaviors [4]–[8]. However, sensory systems with classifiers in the cloud consume a lot of power because of wireless raw sensor data transmission. As such, Low Power Wide Area (LPWA) technology shows an exponential decline of device lifetime for a daily throughput of over 10 bytes [9]. To prevent such frequent transmission, moving average reduces data dimensions but this acts as a low pass filter (LPF) [10]. On the other hand, Edge-AI, which proposes to implement AI algorithms on a hardware device reduces the classification latency [11] and also performs this dimension reduction, enabling lower power consumption [12].

In this work, a comparison is revealed for hardware and software based implementations of Decision Trees (DTs) considering five different Feature Sets (FS). The DTs are trained in software and tested in software and hardware using a prelabeled 3-axis accelerometer cow behavior dataset containing the following behaviors: resting, eating, rumination and moving. This investigation shows a 10% F1-score decrement for the conventional FS (represented as FS4). To increase the accuracy of the system and overcome the large hardware-software implementation difference, the horizontal-longitudinal angle of acceleration of the cow (arctan2($y$, $x$)) is proposed as a new FS together with the vertical $z$ axis (FS1&FS2). Furthermore, a novel approximation of arctan2($y$, $x$) is described with nearest-neighbor interpolation using binary search, reducing power consumption while maintaining DT performance.

More about the dataset and the selection of features and ML algorithms in hardware are explored in Section II accompanied by the approximation of arctan2($y$, $x$). Proceeding to Section III, a setup to measure the classification accuracy of the different FS implementations on an FPGA is shown. Subsequently, in Section IV, the hardware architecture programmed on a Lattice ICE40UP5K FPGA is illustrated followed by a comparison with software in Section V. Finally, the best performing hardware implementation (FS2) is compared with previous works.

## II. ML Algorithm and Feature Set Considerations

The selection of features and ML algorithm has a large influence on the power consumption and performance of a classification system in hardware. As regards the complexity of ML algorithms, Neural Networks (NN) in hardware extensively use multiply-accumulate operations and therefore have a high

TABLE I
3-AXIS ACCELEROMETER BASED FEATURE SETS

| Feature set | Data representation: Mean magnitude ($a_n$) | Mean variation ($a_{var}$) |
|---|---|---|
| FS1 | arctan2($y$, $x$), $z$ | arctan2($y$, $x$), $z$ |
| FS2 | arctan2($y$, $x$), $z$ | $z$ |
| FS3 | arctan($y/x$), $z$ | $z$ |
| FS4 | $\sqrt{x^2+y^2+z^2}$ | $\sqrt{x^2+y^2+z^2}$ |
| FS5 | $\lvert x\rvert + \lvert y\rvert + \lvert z\rvert$ | $\lvert x\rvert + \lvert y\rvert + \lvert z\rvert$ |

power consumption [13]. Moreover, Naive Bayes algorithms need a large amount of Look Up Tables (LUT) to determine exponential and logarithmic operations [14]. Besides that, Hidden Markov Models (HMM) require a high amount of states and resources for state calculation [15]. Regarding cow behavior classification, [4] showed the Decision Tree (DT) to be the most accurate among HMM and 2 other ML algorithms.

Concerning DTs, decision rules are realizable with functions of any order $n$. In this case, the 0-th order axis-parallel DT (only constant decision rules) is selected. According to [16], this usually results in bigger trees but requires no multiplications. Also, the Universal Node hardware architecture, which they show to be the lowest power DT architecture is adapted.

Regarding DT creation, sensor data from six different cows is split up in a stratified way with 80% for training using Gini index as splitting criterion and 20% for testing whereafter the trained DTs' decision rules are implemented in hardware. This dataset contains 295,674 labeled samples of fourteen behaviors obtained beforehand with a 3-axis accelerometer at a sampling frequency of 25 Hz attached to the cow's neck. The sensor's axes are the longitudinal, horizontal and vertical direction for $x$, $y$ and $z$ respectively. Since classification of fourteen behaviors for a decision tree is not realizable, the dataset is transformed to the primary behaviors of cattle i.e. eating, rumination and resting [17]. In addition, moving that indicates when the cow is running or walking is also considered.

Conventional features in cow behavior monitoring with 3D accelerometers are such as the static horizontal axis, the vector magnitude and area, the spectral energy and entropy [4]–[7]. However, none of these consider the horizontal-longitudinal (y-x) acceleration angle with the two argument arctangent. Table I shows details of the five feature sets that are considered in this work where FS1-FS3 include novelty of the arctangent and FS4 and FS5 represent the conventional. Each feature's magnitude and variation (absolute difference for two timesteps) is averaged for every 64 samples, which acts as an LPF and decreases the data size by 98.44%.

Other than data reduction, 87.5% memory size decrease is achieved by reducing 64 floating point (software) to 8-bit fixed point. This results in a quantization error of $\epsilon \approx 0.01563$ G for an acceleration range of $\pm$ 2 G. Furthermore, Table II, which shows the proposed approximation of the two argument arctangent, arctan2($y$, $x$), reduces the function to a binary search lookup which runs in logarithmic time. In the case of **i**, this takes 57.78% less time compared to linear search. This

TABLE II
APPROXIMATION OF THE TWO ARGUMENT ARCTANGENT CONSIDERING PERCENTAGE OF DATASET AND MACHINE LEARNING APPLICATION

| Boundaries | %Dataset | Theoretical | Approximation |
|---|---|---|---|
| $0 < \frac{y}{x} < 1$ | 1.80 | arctan($\frac{y}{x}$), $\frac{\pi y}{4x}$ | $\frac{\pi}{8}$ |
| $\frac{y}{x} > 1, x > 0$ | 19.69 | arctan($\frac{y}{x}$) | $k_n(n=1.....15)^*$ |
| $\frac{y}{x} < 0, x > 0$ | 0.41 | arctan($\frac{y}{x}$) | $-\frac{\pi}{8}$ |
| $x < 0, y \geq 0$ | 73.48 | arctan($\frac{y}{x}$) + $\pi$ | $-k_n(n=1.....15)^*$+ $\pi$ |
| $x < 0, y < 0$ | 0.29 | arctan($\frac{y}{x}$) - $\pi$ | $-\frac{7\pi}{8}$ |
| $x = 0, y > 0$ | 4.27 | $\frac{\pi}{2}$ | $\frac{\pi}{2}$ |
| $x = 0, y < 0$ | 0.01 | $-\frac{\pi}{2}$ | $-\frac{\pi}{2}$ |
| $x = 0, y = 0$ | 0 | undefined | undefined |

*$\mathbf{k} = [0.78,\ 1.11,\ 1.25,\ 1.328,\ 1.375,\ 1.406,\ 1.422,\ 1.453,\ 1.469,\ 1.484,\ 1.500,\ 1.516,\ 1.531,\ 1.547,\ 1.563]$, $n$ is calculated by argmin($\pm\mathbf{i} - \frac{y}{x}$), $\mathbf{i} = [1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 10,\ 11,\ 14,\ 19,\ 30,\ 55,\ 70]$

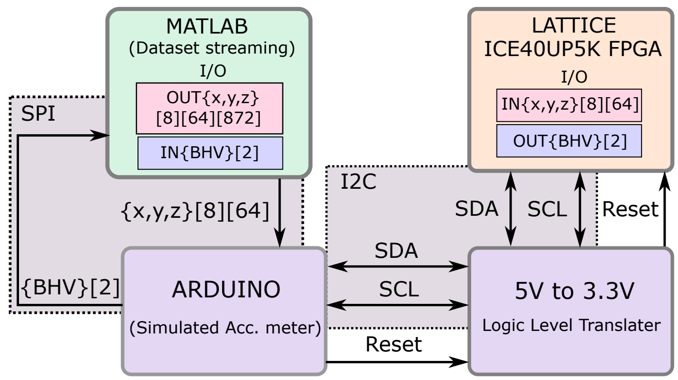



Fig. 1. Setup used to send accelerometer data and receive classified behavior from Matlab on PC to the implementation on FPGA

approximation was carefully tuned by considering the DT's performance while assigning constants to data regions.

## III. FPGA MEASUREMENT SETUP

Figure 1 shows the setup constructed to compare the DT hardware implementation on the Lattice ICE40UP5K FPGA, that is used for this work, for each FS with software. After Arduino stores 64 $\{x,\ y,\ z\}$ samples from Matlab, the I2C master module in the FPGA is trigerred by the reset, requesting all this stored data. Upon reception, the FPGA responds with a 2-bit classified cow behavior. All blocks other than the FPGA are not considered for the power considerations in this work.

## IV. PROPOSED DECISION TREE ARCHITECTURE

Figure 2 shows the VLSI implementable architecture of each feature set described in Table I. A custom implementation is developed for each FS, which changes the amount of addition and memory blocks for the data representations as well as the algorithm of the main features. The previously discussed 64 sample average and approximated arctan2($y$, $x$) are applied. Depending on the FS, four addition-accumulate (AA) blocks for magnitude and variation indicated by yellow, blue, green and red are used and mapped on FPGA. After accumulation

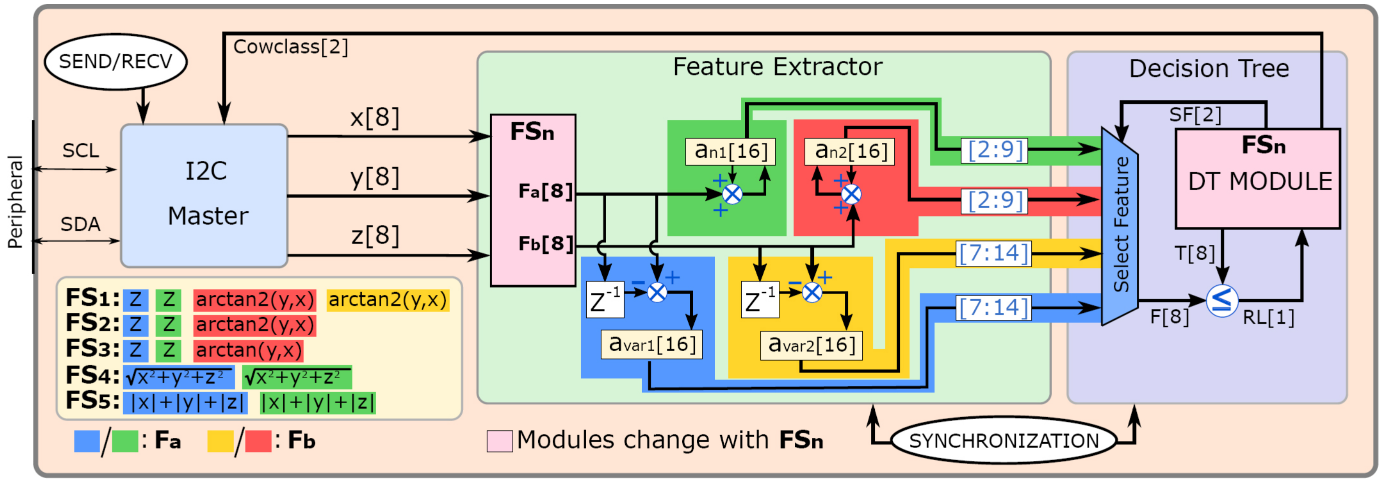



Fig. 2. Top level view of the in FPGA implemented hardware architecture for each feature set, addition-accumulate block usage is indicated with colors and bit size with brackets, the blue brackets represent a shift operation and reduction to 8-bit where index 0 indicates the MSB.

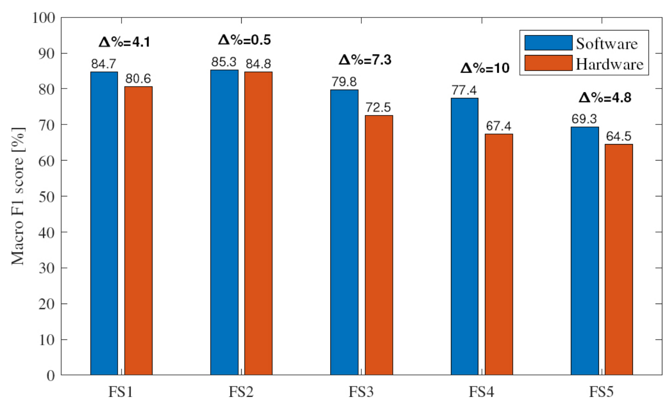



Fig. 3. The macro average F1 score of five different feature sets applied to four cow behaviors on software and 8-bit fixed point hardware.

TABLE III
RESOURCE ALLOCATION FOR FS2 ON LATTICE ICE40UP5K FPGA

| Module | #LUT4 | %FPGA* | #REG | %FPGA † |
|---|---|---|---|---|
| Control | 270 | 5.11 | 70 | 1.33 |
| Decision Tree | 119 | 2.25 | 35 | 0.66 |
| Feature Extractor | 597 | 11.31 | 145 | 2.75 |
| I2C | 247 | 4.68 | 114 | 2.16 |
| Total | 1222 | 23.14 | 364 | 6.89 |

*Total LUT4: 5280
†Total Registers: 5280

of 64 samples, the magnitude and variation are divided by 64 and 2 respectively using shift operations. This division by 2 prevents further rounding error since the variation has a relatively small range. The DT module, which contains all the tree nodes in memory, selects an 8-bit feature $F$ with a 2-bit wire, called $SF$. Repeatedly, a 1-bit wire, $RL = F \leq T$, selects a new node in the tree, which contains $SF$ and decision rule $T$ until a leaf is reached. The behavior stored in this leaf is then forwarded to the I2C module and sent to the peripheral. In consideration of the 25 Hz accelerometer, the clock frequency for the system is set to 2.9 kHz for the I2C to communicate on time. Because the cycle count of the combinational logic for each FS is smaller than an I2C message, its frequency is also set to 2.9 kHz, causing dynamic power to be small relative to static power. The resource use of the FPGA implementation of FS2 are shown in Table III. Using 3 AA blocks and 22 tree nodes amounts to 11.31% and 2.25% of the logic elements (LUT4) and 2.75% and 0.66% of the registers for the Feature Extractor and Decision Tree modules respectively.

## V. HARDWARE-SOFTWARE COMPARISON

Figure 3 shows the macro F1 score (the harmonic mean of the precision and recall for each behavior) of four behavior classification obtained from software and the 8-bit fixed point implementations in hardware with the previously stated architecture, setup and conditions. The percentage difference ($\Delta\%$) is the hardware-software error originating from the quantization of numbers and algorithms. Similarly, Table IV shows the precision and recall in hardware (HW), including resources such as the amount of LUTs, registers and static power, and software (SW) for each behavior and FS.

While FS1 requires the most hardware resources, its F1 score is 4.2% lower than FS2, which considers only the magnitude of arctan2($y$, $x$). Inspection of the FS1 DT in software shows that the variation of arctan2($y$, $x$) does not improve accuracy. Figure 4 supports this by showing less evident axis-parallel boundaries for Fig. 4a and 4b which are used in FS1, compared to the boundaries in Fig. 4c, also used in FS2. Furthermore, F1 score decrement for hardware-software is 0.5% for FS2 while it is 4.1% for FS1 since variation of arctan2($y$, $x$) results in additional error in hardware. Also, while arctan2($y$, $x$) uses slightly more resources than arctan($y/x$), the F1 score increases by 12.3% in hardware. Data analysis shows a more distinct division of cow behavior for 4 quadrants of arctan2($y$, $x$) compared to 2 quadrants of

TABLE IV
A RESOURCE-PERFORMANCE COMPARISON FOR THE HARDWARE IMPLEMENTATIONS ON FPGA AND SOFTWARE APPLYING 5 DIFFERENT FEATURE SETS

| | Hardware Resources | | | Classification Performance (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Resting | | Eating | | Rumination | | Moving | |
| Feature set | #LUT4 | #REG | Static power (µW) | Recall | Precision | Recall | Precision | Recall | Precision | Recall | Precision |
| FS1-HW | 1349 | 473 | 223 | 84.68 | 92.69 | 84.52 | 88.75 | 97.94 | 57.23 | 70.88 | 80.12 |
| FS1-SW | - | - | - | 90.74 | 92.69 | 86.25 | 86.25 | 98.25 | 67.47 | 72.41 | 91.30 |
| **FS2-HW** | **1222** | **364** | **211** | **89.79** | **92.69** | **83.72** | **90.00** | **98.43** | **75.30** | **72.07** | **80.12** |
| FS2-SW | - | - | - | 91.9 | 92.69 | 84.15 | 86.25 | 98.33 | 71.08 | 72.14 | 90.06 |
| FS3-HW | 1160 | 358 | 208 | 89.44 | 95.36 | 46.81 | 83.54 | 100 | 68.87 | 67.92 | 47.68 |
| FS3-SW | - | - | - | 91.90 | 95.14 | 66.28 | 72.15 | 100 | 69.54 | 69.54 | 80.13 |
| FS4-HW | 724 | 319 | 185 | 94.69 | 93.14 | 37.14 | 32.50 | 88.81 | 71.69 | 54.17 | 73.58 |
| FS4-SW | - | - | - | 95.91 | 90.36 | 57.58 | 71.25 | 89.21 | 74.70 | 64.95 | 79.25 |
| FS5-HW | 673 | 285 | 179 | 82.95 | 92.04 | 40.21 | 48.75 | 97.44 | 68.67 | 59.32 | 65.22 |
| FS5-SW | - | - | - | 88.98 | 92.04 | 40.21 | 48.75 | 97.44 | 68.67 | 49.30 | 48.75 |

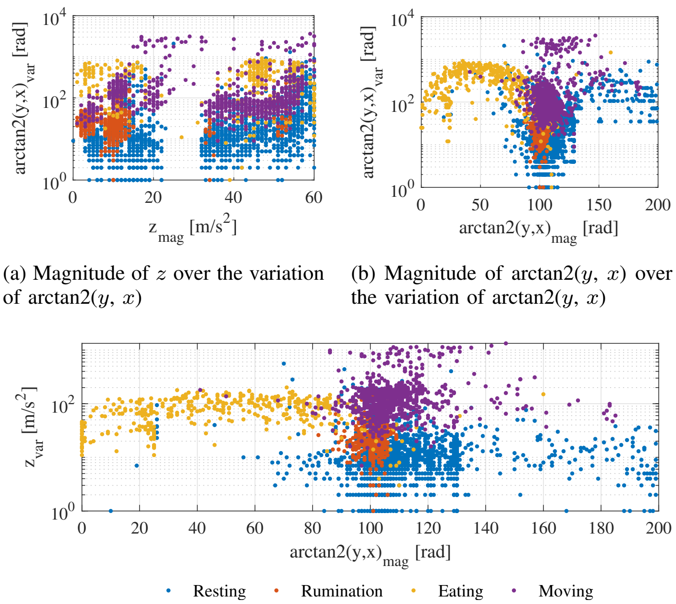
TABLE V
COMPARISON OF FS2 ON FPGA AND STATE OF THE ART

| Reference | [18]* 2018 | [19] 2018 | This Work |
|---|---|---|---|
| Application | Horse | Sheep | Cow |
| Sensor | 3D Accl. | 3D Accl. | 3D Accl. |
| Sensor position | Jaw | Neck | Neck |
| Sampling Frequency (Hz) | 30 | 100 | 25 |
| Windowsize (s) | 0.033 | 6.4 | 2.56 |
| Algorithm | NN | DT | DT |
| #Behaviors | 3 | 5 | 4 |
| Implementation | MCU | MCU | FPGA |
| Power (µW) | N/R | 4380† | 216 |
| Energy (µJ) | N/R | 1259 | **557** |
| Frequency (kHz) | N/R | 23722 | 2.9 |
| Latency (cycle) | N/R | 1865† | 7469 |
| Mean Accuracy (%) | 80.77 | 87.97 | 86.81 |

N/R: Not Reported

* The best result [18] reports applies a kalman filter to all 3-axis accelerometer, gyrometer and magnetometer data to obtain a mean accuracy of 95.39%. This is a high power application and outside the scope of this work.

† This result is only for DT operation without feature extraction and time from windowsize.




(a) Magnitude of $z$ over the variation of arctan2($y$, $x$)

(b) Magnitude of arctan2($y$, $x$) over the variation of arctan2($y$, $x$)

(c) Magnitude of arctan2($y$, $x$) over the variation of $z$

Fig. 4. Scatterplots of quantized FS1 where decision rules occur in the DT for the processed cow behavior dataset containing 4359 samples, (a) and (b) show a less separable scatterplot than (c) which is also part of FS2.

arctan($y/x$). Finally, FS4, similarly used in software in [4], [6] shows an F1 score decrement of 10 % for hardware-software and along with FS5 are inadequate for the classification of moving and eating.

FS2 is selected since the classification accuracy of four behaviors increases significantly compared to FS3 and conventional FS4 and FS5 while only sacrificing a small amount of static power consumption. Table V shows a comparison of the hardware implementation of FS2 with other animal behavior classifiers. Our implementation performs close to or better than state of the art with a classification latency of 7469 cycles or 2.58 s at the frequency of 2.9 kHz, suitable for cow behavior monitoring, and an average power of 216 µW resulting in a total energy consumption of 557 µJ for one classification.

## VI. CONCLUSION

In this paper, a cow behavior classifying hardware implementation competitive with state of the art was presented. For this result, a comparison of software and hardware and the methodology for different feature sets using an axis-parallel decision tree to optimize power consumption and accuracy was described. It was shown that the error involved in using 8-bit representation for low power use cases has to be taken into account for a reliable selection of a feature set. Investigating a new feature and its nearest-neighbor approximation using binary search, the arctan2($y$, $x$), brought about a significant increase of accuracy for a small increase of power. The investigation of classification accuracy for multiple feature sets in hardware proves to be important for optimal low power and small bit size implementations.

## ACKNOWLEDGEMENT

This research was partially supported by the Center of Innovation Program from the Japan Science and Technology Agency (JPMJCE1309) on Development of Innovative Technology. The authors would like to thank Ludovico Minati, Ph.D., of WRHI, Tokyo Institute of Technology for his advice on FPGA implementation and the measurement setup.

## REFERENCES


[1] P. J. Gerber, H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, G. Tempio *et al.*, *Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities.* Food and Agriculture Organization of the United Nations (FAO), 2013.

[2] J. Martinez, P. Dabert, S. Barrington, and C. Burton, "Livestock waste treatment systems for environmental quality, food safety, and sustainability," *Bioresource technology*, vol. 100, no. 22, pp. 5527–5536, 2009.

[3] E. Tullo, A. Finzi, and M. Guarino, "Environmental impact of livestock farming and precision livestock farming as a mitigation strategy," *Science of the total environment*, vol. 650, pp. 2751–2760, 2019.

[4] J. A. Vázquez Diosdado, Z. E. Barker, H. R. Hodges, J. R. Amory, D. P. Croft, N. J. Bell, and E. A. Codling, "Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system," *Animal Biotelemetry*, vol. 3, no. 1, pp. 1–14, 2015. [Online]. Available: http://dx.doi.org/10.1186/s40317-015-0045-8

[5] S. Benaissa, F. A. Tuyttens, D. Plets, T. de Pessemier, J. Trogh, E. Tanghe, L. Martens, L. Vandaele, A. Van Nuffel, W. Joseph, and B. Sonck, "On the use of on-cow accelerometers for the classification of behaviours in dairy barns," *Research in Veterinary Science*, vol. 125, no. August 2017, pp. 425–433, 2019.

[6] T. Tamura, Y. Okubo, Y. Deguchi, S. Koshikawa, M. Takahashi, Y. Chida, and K. Okada, "Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers," *Animal Science Journal*, vol. 90, no. 4, pp. 589–596, 2019.

[7] L. A. González, G. J. Bishop-Hurley, R. N. Handcock, and C. Crossman, "Behavioral classification of data from collars containing motion sensors in grazing cattle," *Computers and Electronics in Agriculture*, vol. 110, pp. 91–102, 2015.

[8] F. A. Alvarenga, I. Borges, V. H. Oddy, and R. C. Dobos, "Discrimination of biting and chewing behaviour in sheep using a tri-axial accelerometer," *Computers and Electronics in Agriculture*, vol. 168, no. August 2019, p. 105051, 2020. [Online]. Available: https://doi.org/10.1016/j.compag.2019.105051

[9] J. Finnegan and S. Brown, "An analysis of the energy consumption of lpwa-based iot devices," in *2018 International Symposium on Networks, Computers and Communications (ISNCC).* IEEE, 2018, pp. 1–6.

[10] S. Golestan, M. Ramezani, J. M. Guerrero, F. D. Freijedo, and M. Monfared, "Moving average filter based phase-locked loops: Performance analysis and design guidelines," *IEEE Transactions on Power Electronics*, vol. 29, no. 6, pp. 2750–2763, 2013.

[11] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge ai: On-demand accelerating deep neural network inference via edge computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, 2019.

[12] J. P. Dominguez-Morales, A. Rios-Navarro, M. Dominguez-Morales, R. Tapiador-Morales, D. Gutierrez-Galan, D. Cascado-Caballero, A. Jimenez-Fernandez, and A. Linares-Barranco, "Wireless sensor network for wildlife tracking and behavior classification of animals in doñana," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2534–2537, 2016.

[13] A. Page and T. Mohsenin, "Fpga-based reduction techniques for efficient deep neural network deployment," in *2016 IEEE 24th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM).* IEEE, 2016, pp. 200–200.

[14] M. N. Marsono, M. W. El-Kharashi, and F. Gebali, "Binary lns-based naïve bayes inference engine for spam control: noise analysis and fpga implementation," *IET Computers & Digital Techniques*, vol. 2, no. 1, pp. 56–62, 2008.

[15] Y. Sun, P. Li, G. Gu, Y. Wen, Y. Liu, and D. Liu, "Accelerating hmmer on fpgas using systolic array based architecture," in *2009 Ieee International Symposium on Parallel & Distributed Processing*. IEEE, 2009, pp. 1–8.

[16] J. R. Struharik, "Implementing decision trees in hardware," *SISY 2011 - 9th International Symposium on Intelligent Systems and Informatics, Proceedings*, pp. 41–46, 2011.

[17] J. Hancock, "Studies of grazing behaviour in relation to grassland management i. variations in grazing habits of dairy cattle," *The Journal of Agricultural Science*, vol. 44, no. 4, pp. 420–433, 1954.

[18] D. Gutierrez-Galan, J. P. Dominguez-Morales, E. Cerezuela-Escudero, A. Rios-Navarro, R. Tapiador-Morales, M. Rivas-Perez, M. Dominguez-Morales, A. Jimenez-Fernandez, and A. Linares-Barranco, "Embedded neural network for real-time animal behavior classification," *Neurocomputing*, vol. 272, pp. 17–26, 2018.

[19] S. P. le Roux, R. Wolhuter, and T. Niesler, "Energy-aware feature and model selection for onboard behavior classification in low-power animal borne sensor applications," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2722–2734, 2018.