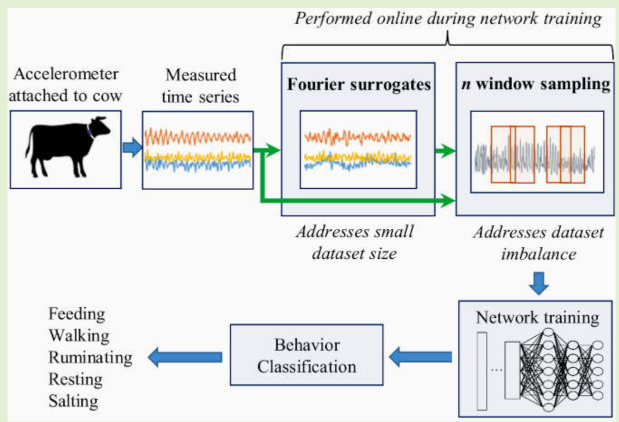
Integrated Data Augmentation for Accelerometer Time Series in Behavior Recognition: Roles of Sampling, Balancing, and Fourier Surrogates

Chao Li, *Graduate Student Member, IEEE*, Ludovico Minati, *Senior Member, IEEE*, Korkut Kaan Tokgoz, *Member, IEEE*, Masamoto Fukawa, Jim Bartels, *Graduate Student Member, IEEE*, Sihan A, *Graduate Student Member, IEEE*, Ken-ichi Takeda, and Hiroyuki Ito, *Member, IEEE*

***Abstract*—The behavioral monitoring of farmed animals such as cattle is a fundamental element of precision farming in which it enables unobtrusive ongoing health monitoring. This application presents two ubiquitous challenges typical of sensing applications of the Internet of Things: limited dataset size and dataset imbalance. Recently, data augmentation has emerged as a way of addressing their negative influences on the training process without overburdening the data acquisition phase. However, there remains no consensus regarding which methods should be applied to time series and in what combination. Here, we present the first comprehensive analysis that synergistically combines multiple approaches. These approaches are benchmarked on a dataset of triaxial accelerometer time series, which were acquired from six freely roaming cows through a collar-mounted sensor and labeled by experienced human observers according to five behaviors. Our results indicate that integrating data augmentation with the training process can substantially improve the time-series classification performance while retaining a fixed convolutional neural network architecture. The improvement is maximized when the dataset is balanced by applying a suitable sampling scheme and the negative influence of data duplication is reduced via generating synthetic time series with Fourier surrogates. With the proposed approach, the overall accuracy is improved from 90% to 96%, and the classification accuracy of an under-represented behavior, namely, grazing, is elevated from 45% to 91%. This work provides a direction toward a general methodology, motivating research on other datasets and applications.**



***Index Terms*—Accelerometer, animal behavior, data augmentation, Fourier surrogates, imbalanced dataset, sensor data processing, time series.**

Manuscript received 20 September 2022; accepted 31 October 2022. Date of publication 9 November 2022; date of current version 14 December 2022. This work was supported in part by the Center of Innovation Program from Japan Science and Technology Agency and Research Program on Development of Innovative Technology [funding agency: Bio-oriented Technology Research Advancement Institution, National Agriculture and Food Research Organization (NARO)] under Grant JPMJCE1309 and in part by the Japan Science and Technology Agency-Support for Pioneering Research Initiated by the Next Generation (JST-SPRING) under Grant JPMJSP2106. The associate editor coordinating the review of this article and approving it for publication was Dr. Pedro Oliveira Conceição Junior. *(Corresponding author: Chao Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Animal Care and Use Committee of Shinshu University.

Chao Li, Ludovico Minati, Jim Bartels, Sihan A, and Hiroyuki Ito are with the Nano Sensing Unit, Tokyo Institute of Technology, Yokohama 226-8503, Japan (e-mail: li.c.ag@m.titech.ac.jp; paper@lsi.pi.titech.ac.jp).

Korkut Kaan Tokgoz was with the Research Institute for the Earth Inclusive Sensing (EISESiV), Tokyo Institute of Technology, Tokyo 152-8552, Japan. He is now with the Faculty of Engineering and Natural Sciences, Sabancı University, 34956 Istanbul, Turkey.

Masamoto Fukawa is with the Nano Sensing Unit, Tokyo Institute of Technology, Yokohama 226-8503, Japan, and also with TechnoPro Design, Tokyo 106-6135, Japan.

Ken-ichi Takeda is with the Faculty of Agriculture, Shinshu University, Nagano 399-4598, Japan.

Digital Object Identifier 10.1109/JSEN.2022.3219594

I. Introduction

CATTLE behavior is a valuable and easily observable indicator of livestock health and well-being, which influence the amount and quality of the end products that can be obtained [1]. Therefore, monitoring cattle behavior is a necessary aspect of effective livestock welfare support systems aiming to improve the yield and reduce the environmental impact of farming [2], [3]. However, it is impractical to personally monitor livestock behaviors, especially in large herds, at all times. Identifying all individuals in a herd and accurately determining their health statuses via direct observation requires excessive effort and is prone to error since it depends heavily on expertise. Automatic behavior monitoring systems address these limitations by recognizing and quantifying the prevalence of a predetermined set of behaviors through instrumental measurements, often consisting of triaxial accelerometer time series [4], [5].

However, due to the diversity of behaviors, interindividual variability, and the presence of various external noise sources, drawing inferences from these time series is not trivial. Recently, neural networks have become a powerful behavior and activity recognition tool, but the success of neural network-based classifiers across diverse fields rests primarily upon the availability of large amounts of high-quality training data [6]. Accordingly, data quantity is universally associated with model robustness and generalization performance [7]. The process of training suitable classifiers for cattle behavior classification, in particular, encounters two practical issues that tend to be ubiquitous in edge-based sensing applications of the Internet of Things, namely, limited dataset size and dataset imbalance [8], [9]. The first problem, namely, limited dataset size, arises because of the human and financial costs of data gathering. In the present case, long-term video monitoring requires dedicated personnel for footage acquisition and subsequent behavioral labeling; it also encounters practical difficulties associated with filming while attempting not to disturb the natural cattle behavior. The second problem, namely, dataset imbalance, arises because it is rare to observe a balanced prevalence across behaviors. On the contrary, natural and artificial processes exhibit some activity patterns more frequently than others because their daily functioning requires it. Similar situations are encountered when considering wearable systems for livestock activity monitoring [10], human behavior classification [11], gesture recognition [12], as well as room occupancy and activity detection [13].

One way of addressing these situations is by applying data augmentation techniques, that is, suitable algorithmic manipulations that enlarge the amount of data entered into the training process without overburdening the data acquisition stage. In essence, augmenting data can be considered a preprocessing step that leverages prior information regarding the expected invariant features of the time series concerning certain transforms, such as sensor axis rotations, scalings, and temporal manipulations. When appropriately applied, data augmentation generates synthetic patterns that expand the classifier's decision boundaries [14], improving network generalization performance at a minimal cost.

Data augmentation has recently become a common practice in the image domain. For example, many well-established deep learning architectures for image classification, such as AlexNet [15], residual networks (ResNets) [16], and very deep convolutional networks (VGGs) [17], include data augmentation approaches as a standard practice in the process of model training. Considerably fewer time-series-based classifiers benefit from data augmentation [18]. Most existing time-series augmentation approaches for sensor signals are based on random transforms, such as window slicing [19], warping [20], and scaling [21]. These transforms are, in fact, mainly inspired by image augmentation, suggesting that data augmentation has not yet been thoroughly investigated in the time-series domain.

This article proposes the integrated and synergistic use of multiple approaches to better tackle classifying cattle behaviors while facing limited data availability. Namely, it combines three key ideas resulting in substantially increased classification performance and having possibly general usefulness.

1) We propose using a random selection of time-series snippets during each training epoch to provide a built-in source of variability that aids the determination of classifier boundaries, yielding a high generalization ability.
2) We propose applying a form of biased sampling, which aims to offset the dataset imbalance problem, further aiding the training process in determining the boundaries between the most and least represented behaviors.
3) We propose combining the above with the generation of Fourier transform-based surrogates to alleviate the issue of data duplication encountered when repeatedly sampling over the least represented behaviors.

In particular, we apply a multivariate version of the Iterative amplitude-adjusted Fourier transform (IAAFT) method, preserving the autocorrelations and cross correlations alongside the value distributions [22]. Finally, we perform a step-by-step reduction to reveal the time-series properties supporting the generalization performance.

Throughout the remainder of this article, in Section II, we first discuss related works on time-series data augmentation. Section III introduces the dataset and describes the methods for data processing, model training, data augmentation, and performance evaluation. Section IV presents the classification performance results across the sampling and surrogate schemes, followed by an analysis of the relevance of surrogate time series based on a deductive approach. Finally, Section V offers general conclusions regarding the benefits of the proposed method, its drawbacks, and the possible directions for future work.

II. Related Works

The majority of existing augmentation approaches used in time-series-based research are based on random transformations originally inspired by image data augmentation. Some examples include scaling (global magnitude changes), window slicing (equivalent to image cropping), magnitude warping (modulating signal magnitude by a smooth curve), rotating (flipping for univariate cases and rotating for multivariate cases), and jittering (e.g., adding Gaussian noise). These augmentation operations have been widely applied across diverse data sources and time-series classification tasks. Three recent reviews, namely, Iwana and Uchida [14], Wen et al. [18], and Ge et al. [23], offer a comprehensive survey of this field.

For instance, Le Guennec et al. [19] proposed a window slicing and warping method, which generates new samples by randomly slicing the original time series and speeding up or slowing down the extracted small-size slices. Um et al. [24] applied various data augmentation methods, including permutating, cropping, rotating, scaling, jittering, time-warping, and magnitude-warping methods to wearable sensor data, with a focus on the application of convolutional neural networks (CNNs) to Parkinson's disease monitoring. Frequency warping is also a prominent approach in augmenting time-series data, although it is mainly used for audio and speech recognition. The issue with random transformation-based augmentation is that such methods do not rigorously consider which signal properties should be invariant; instead, they rely on simple heuristics.

The method proposed in this work attempts to be more principled. It involves extracting a fixed-length snippet from each segment of the time series, in which a segment is defined as the time interval between two behavior transitions. Given that such transitions are irregular, some segments are longer than others; however, a fixed snippet is always extracted, starting from a random time. Within the framework of Iwana and Uchida [14], this resembles an implementation of the slicing method. According to Wen et al. [18], this method, while not explicitly mentioned, would be an instance of cropping. However, it should be underlined that neither cropping nor slicing, unlike the method that we have implemented, provides a homogeneous snippet length.

Another critical aspect of the proposed sampling method is that, as clarified in the following, it is performed online and fully integrated with the training process. While online data augmentation is commonplace in computer vision applications, it remains almost unexplored in time-series analysis, as confirmed by three recent reviews [14], [18], [23]. In line with the results reported next, the present work shows that even a relatively straightforward method such as segment sampling can significantly improve the performance when implemented in an online form, that is, running during training rather than only once before training.

For completeness, it should also be mentioned that an alternate approach to random transformations is pattern mixing, which combines multiple samples of intraclass data to generate new ones. An example application is the one reported by Takahashi et al. [25]. A problem with this approach is that out-of-phase overlap can occur, and nonperiodic time-series data may lead to malformed patterns. Therefore, this approach is not well-suited for applications such as animal behavior recognition.

As regards the issue of dataset imbalance, broadly put, three approaches are possible: undersampling, that is, artificially reducing the prevalence of the most frequent behaviors, oversampling, that is, increasing the prevalence of the least frequent behaviors either by repetitive presentation or by interpolation, and generating new data based on heuristic rules. Comprehensive reviews of these approaches can be found by Kaur et al. [26], Patel et al. [27], and Tanha et al. [28], with additional considerations about the impact on the learning process given by Krawczyk [29] and He and Garcia [30].

Our approach, further detailed in the following, is distinguished by two aspects. First, it explicitly leverages the temporal adjacency of samples so that, rather than rejecting or inserting new data points as done customarily, it effectively biases how densely the time-series segments covering a particular behavior get covered by the snippets that are extracted and submitted to the training process. Second, it is performed online, which represents a fundamental difference with respect to most existing literature in this field. This aspect implies that, no matter the relative under- or over-representation of specific classes, a high level of variability is retained in the data input to the training process across epochs. It is a crucial advantage and contribution of this work because the most severe drawback of performing under- or over-sampling prior to training is the reduction or magnification of the variance in each class as observed by the training algorithm.

In other words, an integrated approach to dataset imbalance mitigation appears to alleviate the difficulties encountered in learning the boundaries between differently represented classes while ensuring that the majority of the available input data variance is still made available to the training process. It appears noteworthy that, even in the specific survey on resampling provided by Moniz et al. [31], these aspects are not considered; this indicates that the focus remains firmly on precalculated dataset adjustments and that the beneficial impact of integrated preprocessing remains largely to be clarified. An exception is the work of Cao et al. [32], which, however, is different from the present one in that it relies on oversampling by interpolation between neighbors in the feature space rather than sampling in the time domain.

Finally, as regards the generation of Fourier surrogates, it is worth pointing out that the IAAFT method was initially devised for a purpose different from data augmentation, namely, the generation of null-hypothesis datasets against which to compare experimental data [33]. Because phase randomization destroys any nonlinear structure, this method can detect nonlinear features over noise, for example, in brain activity time series [34]. In the course of previous research [35], we noted that time reversal could serve as a helpful data augmentation method. Because this operation inherently alters the nonlinear properties of the data, we concluded that the features necessary for behavior classification would plausibly be found in linear properties of the kind preserved by the IAAFT method. Therefore, in this article, we consider it a data augmentation technique.

As reviewed by Ge et al. [23], this approach has been gradually emerging as a potential candidate for data augmentation. Unlike simple operations such as cropping, it ensures that no time-domain data duplication can occur; importantly, it does not suffer from the issues associated with potentially class-altering operations such as scaling. It is also more desirable compared to pattern mixing, as it avoids data corruption in the presence of nonstationarity [36], and compared to generative approaches such as adversarial networks, which have considerable potential but require resource-intensive training to become reliable data generators [37], [38].

The exact modality of using surrogates for data augmentation varies across studies, and the literature remains scant. For example, Lee et al. [39] used surrogate data intermixed with original accelerometric and neurophysiological recordings, drawing the training and test data from the resulting pool. Schwabedal et al. [40] adopted a similar approach, but applied surrogates only to the training data, while performing cross validation. On the other hand, in a later study, Lee et al. [41] proposed using surrogates as the exclusive basis for training and validation, reserving the entirety of the experimental recordings for testing. In this article, we consider three possibilities more systematically: training only on the original data, training only on the surrogates, and training on an evenly mixed pool. In all cases, validation and testing are performed exclusively on the original data and remain unchanged throughout the entire training process. As a result,

Fig. 1. Representative video frames for the behavioral classes and sensor positioning, taken on the farm at Shinshu University, Nagano, Japan.

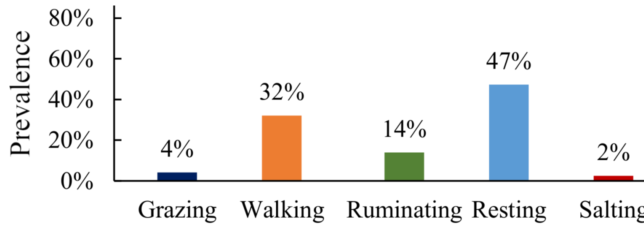


Fig. 2. Relative behavior prevalence (normalized).

TABLE I
DESCRIPTION OF THE OBSERVED (LABELED) BEHAVIORS AND THE CORRESPONDING NUMBER OF SAMPLES

Label	N. of samples	Description
Grazing	19,632	Eating grass
Walking	150,682	Moving on the grassy field or in the farm pen
Ruminating	65,509	Regurgitating, re-chewing and re-swallowing feed
Resting	221,853	Standing or lying still
Salting	11,419	Lick mineral salt blocks

it is possible to explicitly address the effect of surrogate data inclusion.

Three more aspects of this work stand out with respect to the existing literature. One advancement is that, similar to sampling, surrogate generation is integrated, ensuring that each training epoch can receive uncorrelated input vectors. In the field of time-series analysis, data augmentation tends to be performed just once on the entire dataset prior to model training. In addition to the obvious drawback of requiring additional storage size, this approach suffers from the fundamental limitation that the surrogates are, on par with experimental data, static, i.e., identical across epochs. As our results have shown, generating the surrogates during training and integrating this approach with the random subsampling of data segments can confer a substantial performance boost that would not be available when precalculating the surrogates. Importantly, this also counters the risk of overfitting. Another innovation is that, unlike the existing studies such as the one by Lee et al. [41], we explicitly consider the multivariate nature of the triaxial vector accelerometer data and consequently employ an iterative method that preserves not only the autocorrelations but also the cross correlations, more appropriately representing the real-world kinematics. Finally, the authors are unaware of any studies that explicitly attempt to demonstrate why surrogate data help for data augmentation. In this work, we rigorously deploy a step-by-step, deductive approach under which the retained features are gradually removed, observing how the classification performance declines and therefore inferring which statistical aspects support the use of surrogate data.

III. DATA AND PROPOSED PROCESSING METHODS

A. Data Acquisition

Data acquisition was conducted on a sample of six Japanese black beef cows at a cattle farm located at Shinshu University, Nagano, Japan. The animals were allowed to roam freely over a grassy field and a farm pen. The presence of disease conditions was excluded via veterinary monitoring. All animal handling procedures were reviewed and approved by the Institutional Animal Care and Use Committee of Shinshu University.

The acceleration data were acquired utilizing KX122-1037 silicon micromachined accelerometers (ROHM Company Ltd., Kyoto, Japan). The sampling rate was set to 25 Hz, the conversion precision was set to 16 bit, and the range was set to ± 4 g, providing a noise floor of 0.75 mg. The data were recorded using a microcontroller and a local storage device.

The sensor device was strapped to the cow's neck during the experiments using a dedicated nylon collar belt. Throughout the existing studies, the majority of sensors used for cattle behavior monitoring are positioned within collars because this reduces the risk of chewing while allowing the installation of sufficiently sized batteries [42]. The tightness was adjusted to provide good acceleration coupling in response to natural movement while minimizing discomfort.

Table I lists the five prevalent behaviors alongside their descriptions and number of samples available, and example video frames are shown in Fig. 1. These represent the most prevalent behaviors in the cattle's daily activity, and subtle changes in their distribution are diagnostic of subclinical or "hidden" disease [43]. Other behaviors were less prevalent and therefore excluded from the analyses.

To label the data, synchronized video recordings were acquired while maintaining the time axes aligned as accurately as possible through GPS time stamping. Experienced human observers manually annotated the data frame-by-frame based on the videos and later reviewed it. Subsequently, the time series were cut into segments according to the annotated behaviors, with a median length ranging between 200 and 1200 points. Fig. 2 shows the relative prevalence of data samples across the activities. The data logger and underlying raw time series have been made publicly available [44], [45], and the representative examples are shown in Fig. 3.

B. Machine Learning Model

The machine learning model consisted of a CNN receiving three-axial time-series snippets as inputs and outputting the

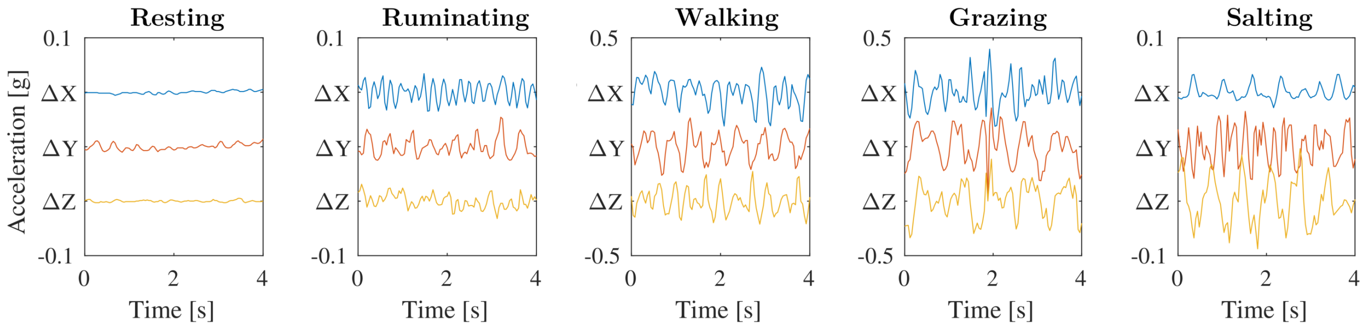


Fig. 3. Representative time-series excerpts for the behavior classes. Mean subtracted for visualization purposes (unit: gram).

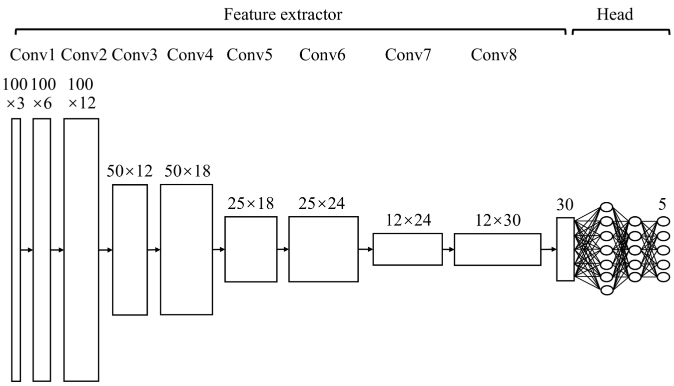


Fig. 4. Temporal CNN architecture, showing the input time-series segment length (100), the feature size (30), and the number of output classes (5).

predicted classes. CNNs are widely used for detecting local features through filters sliding across temporal sequences [46], [47] and automatically extract features while keeping the computational complexity down to a minimum [48]. They are, therefore, particularly desirable in wearable edge Internet-of-Things devices, for which the additional complexity of feature extraction (such as Fourier transform calculation) would complicate low-power deployment, and feeding raw sensor signals is preferable [4].

A multilayer CNN having eight convolutional layers was used in this work. The basic architecture comprised two logical entities: a cascade of convolutional blocks and a multilayer "head" of the network (see Fig. 4). Convolutional and max-pooling layers were alternatively instantiated to extract a series of feature maps from the original time series, and each convolutional layer comprised convolution and batch normalization operations alongside a rectified linear unit (ReLU) activation function [49].

The input layer had a size of 100×3, where 3 denotes the axes and 100 indicates the input data length (corresponding to a fixed snippet of 4 s). The convolutional layers had the sizes of 100×6, 100×12, 50×12, 50×18, 25×18, 25×24, 12×24, and 12×30, and the convolution kernels had the sizes of 6×3, 12×3, 12×3, 18×3, 18×3, 24×3, 24×3, and 30×3. Global average pooling (GAP) [50] was used in the final step of the feature extractor, aggregating the deep features along the spatial dimensions. Finally, the features were entered in the network "head," which consisted of two fully connected layers having 30 inputs and five outputs. A five-way softmax layer was instantiated to obtain the predicted labels, i.e., resting, ruminating, walking, grazing, and salting.

Model training and evaluation were performed in Python using Keras with the backend of TensorFlow (version 2.4.0) [52]. The model was trained using the adaptive moment estimation (Adam) optimizer [51], setting the initial learning rate to 0.02, the number of epochs to 1500, and the batch size to 256. The optimal convolutional blocks, kernel sizes, initial learning rate, and number of epochs were determined during preliminary work, omitted for brevity, aiming to maximize the overall validation performance in the baseline case, that is, in the absence of sampling and augmentation (case OA).

C. Model Evaluation Metrics

After learning was completed, the effectiveness of the classifier model was examined based on an independent test set. The overall accuracy was calculated as the agreement between the behaviors predicted by the classifier and the human annotations from video analysis and was defined as overall accuracy $= C/(C+I)$, where C and I denote the number of correct and incorrect classifications, respectively.

As regards the accuracy separately for each class (i.e., behavior), the following metrics were calculated based on the number of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) derived from the confusion matrix. The precision and recall were first obtained as $\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$ and $\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$. Then, the F_1 score was calculated as the harmonic mean of them, namely, $F_1 = 2 \times (\mathrm{Precision} \times \mathrm{Recall})/(\mathrm{Precision} + \mathrm{Recall})$ [53]. Finally, the Matthews correlation coefficient (MCC) was obtained with $\mathrm{MCC} = (\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN})/(((\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})/(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})))^{1/2}$ [54].

D. Data Augmentation

Fig. 5 shows the data flow supporting the experimental design. Stratified splitting was performed following data segmentation according to behavior labels, retaining 72% of the overall data for training and the remaining 20% and 8% for test and validation, respectively. After that, the analysis split into three homologous branches: a first one involving only the original data, a second one involving only the surrogates, and a third one involving an even mixture of the two (identified, respectively, with "O," "S," and "M"). This branching pertained to the training data only, and only the original data were

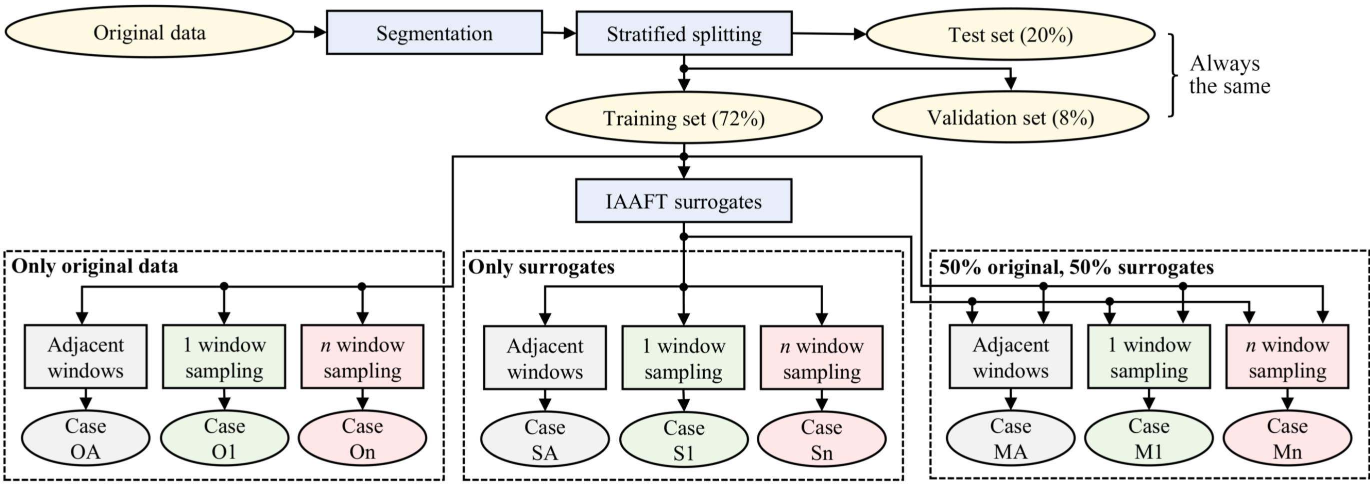


Fig. 5. Data processing flow, showing the 3 × 3 split according to surrogate usage (O, S, M) and sampling scheme (A, 1, *n*).

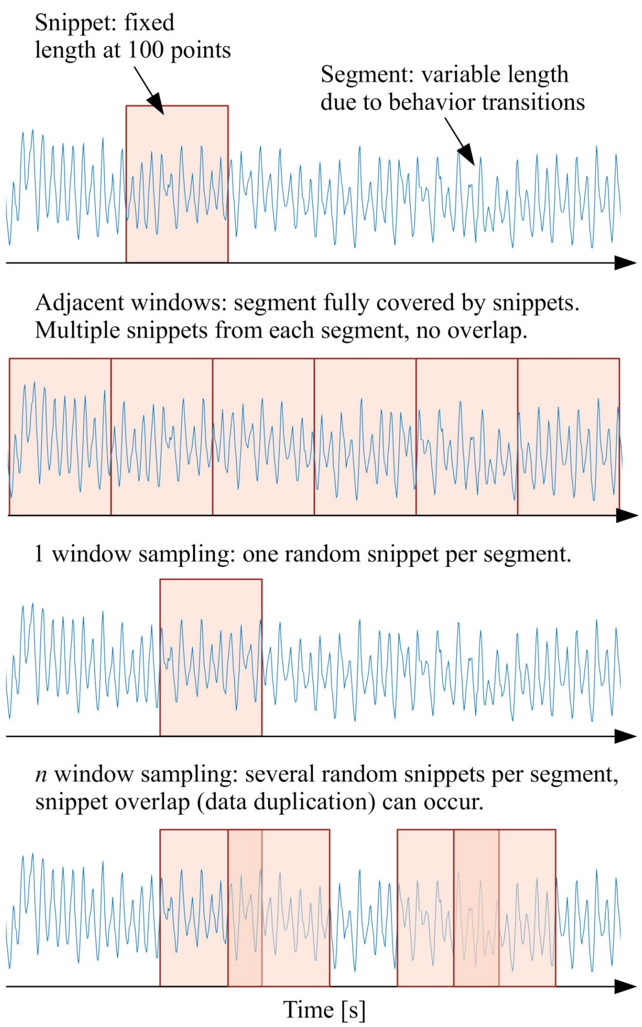


Fig. 6. Sampling schemes (A, 1, *n*) used in deriving snippets (fixed length of 4-s time intervals submitted to the CNN) from segments (variable length of 8–48 s time intervals of homogeneous behavior).

used for validation and testing. Within each branch, a further split into three sampling approaches, detailed in Fig. 6, was present. These consisted of considering adjacent windows (all data entered during each training epoch, retaining the original class distribution), considering one window per data segment (extracting one snippet per segment starting from a different random location for each epoch and retaining the original class distribution), and considering n windows per data segment to approximately balance the distribution (identified, respectively, with "A," "1," and "n"). In other words, the study followed a 3 × 3 design according to surrogate usage and windowing/sampling approach.

All aspects of the data augmentation methods were seamlessly integrated with the training process; in other words, no precalculation was performed, and the surrogate generation and sampling were performed independently for each training epoch. Surrogate time series were generated through the IAAFT method, separately for each segment of the original data, ensuring that no labels were mixed. The method is described in detail elsewhere [33] but, in brief, consists of the iterative adjustment of value distributions and Fourier spectra, starting from a shuffle of the initial data points. The Fourier transform is calculated at each iteration, retaining the phases but replacing the amplitudes with those of the original time series. Then, the values of the iterated time series are replaced with those from the original time series, according to their ranks. This process is iterated until the initial signal's value distribution and autocorrelation are both sufficiently preserved. The obtained time series is entirely uncorrelated in the time domain, which prevents data duplication. In addition, for triaxial data, useful information may be contained in the cross correlations; therefore, we applied a multivariate extension of this method, which also preserves cross correlations. The entire calculation details and a source code are available [55].

Finally, previous works have used surrogate time series solely based on empirical evidence that they can aid the training process. Here, as shown in Fig. 7, we adopted a deductive approach to understanding precisely, which retained features render the surrogate time series usable for training. Starting from the original data, the nonlinear structure was first destroyed by the IAAFT method itself. After that, we switched from a multivariate approach to a univariate one, thereby ceasing to retain the cross correlations. Next, we relinquished the

TABLE II
PERFORMANCE OF THE CLASSIFICATION RESULTS ON A TEST DATASET ACROSS THE SAMPLING AND SURROGATE SCHEMES. SEE FIG. 5 FOR DEFINITION OF THE CASES

	Case	OA	O1	On	SA	S1	Sn	MA	M1	Mn
F_1-scores	Grazing	45%	83%	86%	52%	85%	89%	68%	87%	91%
	Walking	89%	95%	96%	88%	95%	98%	85%	96%	97%
	Ruminating	88%	92%	95%	89%	91%	93%	83%	91%	92%
	Resting	95%	96%	97%	94%	95%	97%	90%	96%	97%
	Salting	83%	98%	100%	79%	93%	100%	78%	93%	100%
Overall accuracy		90%	95%	96%	90%	94%	96%	86%	95%	96%

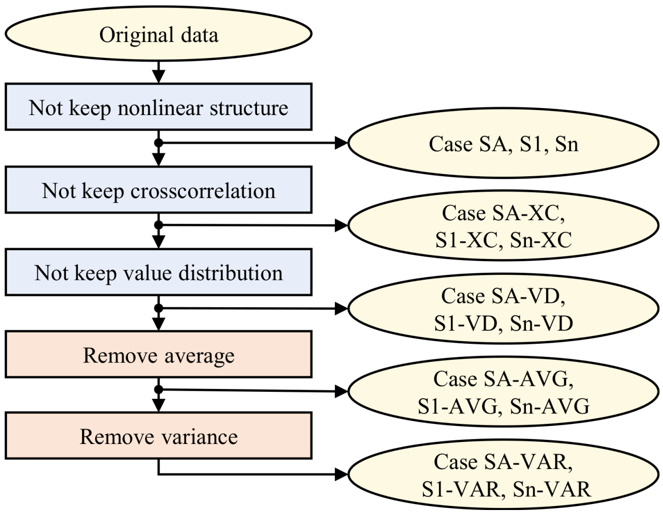


Fig. 7. Deductive steps used to determine the elements of surrogate data supporting high training performance.

IAAFT iterative approach and randomized the Fourier phases, leading to time series that retains the autocorrelation but not the value distribution. Finally, we subtracted the average and normalized the variance to unity: this aspect is essential since useful information about the behavior can be conveyed by the average and variance, which are in part represented within the Fourier amplitudes, even when the value distribution is not adjusted.

IV. RESULTS

A. Classification Performance Across the Sampling and Surrogate Schemes

The classification accuracy values, aggregated and separate for each behavior class, are given in Table II, while the corresponding confusion matrices are shown in Fig. 8. Considering the original data only and dividing up all segments into adjacent windows (case OA), an overall accuracy of 90% was observed. With respect to this, the most significant improvement, to 95%, was obtained by introducing random sampling (case O1), according to which one window is extracted starting from a different random time point for each training epoch. The improvement was particularly notable for the least-represented behaviors, namely, grazing and salting, whose scores increased, respectively, from 45% to 83% and from 83% to 98%.

When introducing biased sampling, which rendered the number of windows entered in the training process approximately even across behaviors (case On), a further improvement to 96% was recorded; even though the additional score increase was quantitatively smaller, it was consistent across all classes.

TABLE III
ADDITIONAL EVALUATION METRICS ON A TEST DATASET. SEE FIG. 5 FOR DEFINITION OF THE CASES

	Precision		Recall		MCC	
Case	OA	Mn	OA	Mn	OA	Mn
Grazing	67%	96%	34%	86%	46%	91%
Walking	89%	98%	89%	97%	84%	96%
Ruminating	85%	91%	91%	93%	86%	90%
Resting	95%	97%	95%	97%	91%	94%
Salting	76%	100%	90%	100%	82%	100%

Considering that next, the training performed exclusively on the surrogates, the principal finding was that the accuracy levels across the three sampling schemes were closely comparable to those obtained when using the original data, namely, 90% for both cases OA and SA, 95% and 94% for cases O1 and S1, respectively, and 96% for both cases On and Sn. When considering the score for the grazing class, the performance was better when training using the surrogates rather than the original data, namely, with 45% versus 52% for cases OA and SA, 83% versus 85% for cases O1 and S1, and 86% versus 89% for cases On and Sn. This could plausibly be ascribed to statistical aspects such as improved stationarity. It should be noted that the online generated surrogates, being uncorrelated across epochs, strongly counter overfitting; therefore, even when the performance is the same as the original data, they help rule out overfitting.

The most favorable performance was obtained via combining the original data and the surrogates, randomly intermixing them for each epoch. At the overall classification accuracy level, the case Mn scores were comparable to those obtained using only the original data or only the surrogates, namely, with 96% (case On) and 96% (case Sn). However, considering the individual behaviors, the score for Grazing was maximized, reaching 91% for case Mn as opposed to 86% for case On. Due to a saturation effect, this was not evident for salting. The advantage of mixing the surrogates and original data plausibly stems from their different statistical properties as regards which features are retained and from the data variability.

Further insight into the performance of the proposed method could be obtained by considering separately the precision and recall, as well as the MCC, which are given in Table III for cases OA and Mn. It can be seen that these metrics were consistently higher for case Mn across all behaviors, with the most marked improvement being observed, again, for grazing

Fig. 8. Confusion matrices for a selection of sampling and surrogate schemes (test data).

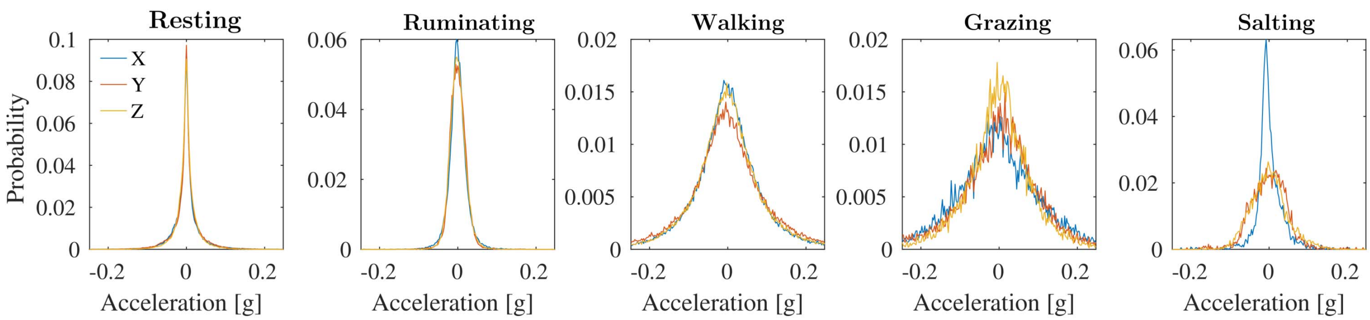


Fig. 9. Value distributions across the behavior classes.

followed by salting. Overall, the improvement was similar across precision, recall, and MCC.

B. Analysis of the Relevance of Surrogate Time Series

Considering the ability of the surrogates to sustain high training performance, the following results were noted (see Table IV); for brevity, they are presented concerning case Sn, but similar considerations hold for cases SA and S1. Starting from the attained score of 96%, switching to a univariate approach not retaining the cross correlations had a negligible effect on the overall accuracy, which remained high at 95%. Removing the iterative process in the IAAFT algorithm, thus retaining the Fourier amplitudes without the value distribution, also did not reduce the performance.

On the other hand, removing the average outright by subtracting it from the data had a more complex effect. The overall accuracy for case Sn remained similar, down to 94%; however, for case SA, the accuracy for the grazing class collapsed to 4%. Overall, removing the variance information by normalization had a stronger and more generalized detrimental effect, reducing the overall accuracy down to 80% for case Sn.

Altogether, these results suggest that the bulk of relevant information was contained in the autocorrelation, the only feature retained after this cascade, and in the variance. Therefore, to further characterize the dynamical features retained by the IAAFT surrogates that support the use of surrogate generation as a data augmentation technique, the value distribution, autocorrelation, and cross correlation were considered in detail across the behavior classes.

First, we note that, as shown in Fig. 9, the five behaviors were characterized by markedly different distributions of the acceleration values; for convenience, they are illustrated after detrending. For resting, a narrow Lorentz-like distribution was observed, with near-complete overlap between the three axes (standard deviations of 0.03, 0.03, and 0.02 g for X, Y, and Z, respectively). For ruminating, the distributions were marginally broader and less peaked around zero, albeit retaining a comparable standard deviation (0.02 for all three axes). By contrast, for walking, the acceleration distributions for all three axes were markedly broader and more Gaussian-like (standard deviations of 0.10, 0.13, and 0.11 g). For grazing, the situation was comparable, albeit with larger noise due to the smaller amount of data (standard deviations of 0.11, 0.10, and 0.08 g). For salting, the variability was intermediate, and there was an evident difference between a Lorentz-like distribution for the X-axis and Gaussian-like distributions for the Y- and Z-axes (standard deviations of 0.03, 0.04, and 0.05 g). In summary, the resting and ruminating behaviors appeared closely comparable and well-separated from walking and grazing, which were similar to each other, whereas salting represented an intermediate condition. Therefore, the relevant information contained in the value distribution consisted mainly of different variances, supporting the distinction between these classes.

Second, we note that, as shown in Fig. 10, across the three axes, the five behaviors were characterized by visibly different autocorrelation and cross correlation profiles, in which the latter tended to be smaller. On the whole, resting was associated with a relatively slow and monotonic autocorrelation decay, which was comparable for the three axes; its cross correlation dwelled around zero, except for the YZ combination, which was strongly anticorrelated. These features plausibly stem from the absence of regular movements alongside occasional head rotations during resting. By comparison, the autocorrelation envelope for ruminating showed a faster decay, which was additionally associated with a prominent periodic oscillation

TABLE IV
PERFORMANCE OF THE CLASSIFICATION RESULTS ON A TEST DATASET

	Case	SA-XC	S1-XC	Sn-XC	SA-VD	S1-VD	Sn-VD	SA-AVG	S1-AVG	Sn-AVG	SA-VAR	S1-VAR	Sn-VAR
F_1-scores	Grazing	57%	80%	84%	50%	87%	84%	4%	78%	85%	5%	49%	64%
	Walking	87%	95%	96%	88%	96%	96%	84%	94%	94%	63%	78%	70%
	Ruminating	85%	89%	91%	81%	89%	91%	78%	94%	93%	60%	82%	82%
	Resting	93%	96%	96%	90%	95%	97%	92%	96%	96%	84%	86%	85%
	Salting	79%	93%	100%	95%	95%	100%	67%	95%	98%	83%	92%	98%
Overall accuracy		89%	94%	95%	87%	94%	95%	84%	95%	94%	73%	82%	80%

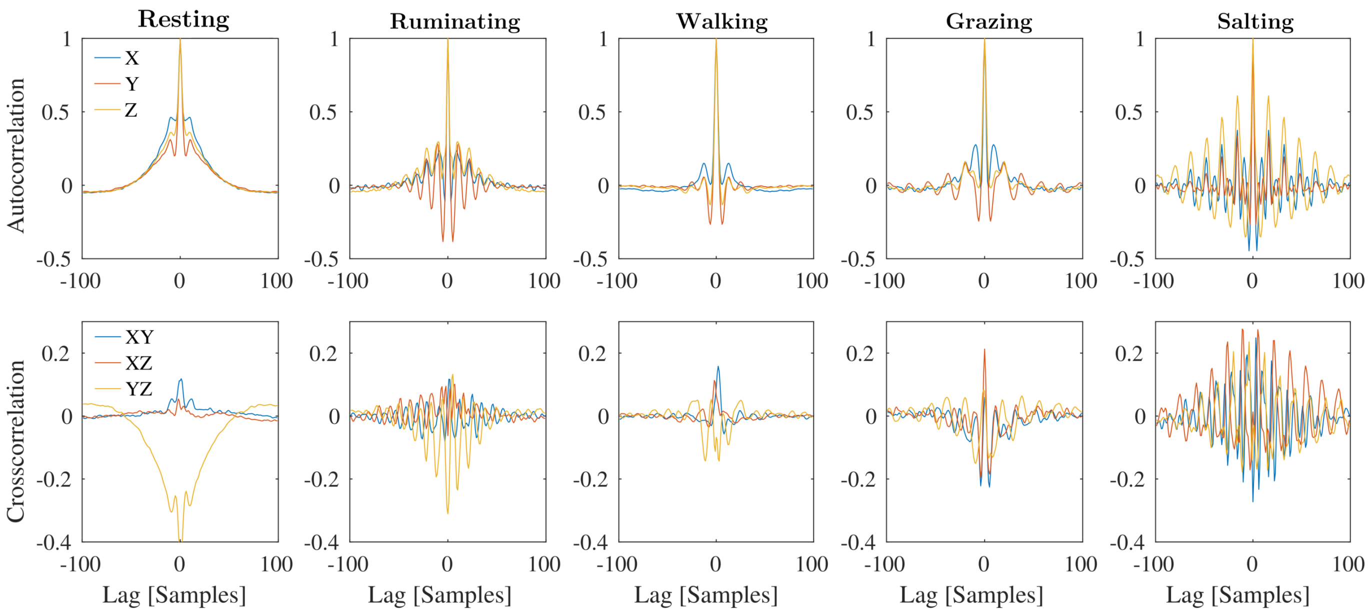


Fig. 10. Autocorrelation and cross correlation across the behavior classes.

peaking at a lag of around ten samples, particularly for the Y-axis; this periodicity could also be appreciated for the cross correlation. Ruminating knowingly involves prolonged chewing and associated rhythmic neck movements, which may explain the observed pattern. Conversely, walking was associated with the fastest autocorrelation decay, minimal periodicity, and weakest cross correlation. These features plausibly reflect the fact that, regardless of leg movement, the head movements are minimized in this condition due to staring forward. Grazing exhibited properties intermediate between ruminating and walking, as regards both the autocorrelation decay and the strength of cross correlation; compared to ruminating, this behavior was associated with a somewhat slower periodicity, peaking at a lag of around 15 samples, again in line with behavioral expectations of slower chewing in this condition. Finally, salting was markedly different from all other behaviors in which it was hallmarked by a very strong periodicity at an intermediate frequency, again peaking at a lag of around 15 samples, which was visible on all three axes for autocorrelation and all three axis combinations for cross correlation. Since salting involves large repetitive "sliding" movements associated with licking, this pattern was expected. Therefore, the most crucial distinguishing feature appeared to be autocorrelation, followed by the value distribution. Notably, the separability of the behaviors was different and complementary between them since resting and ruminating, walking, and grazing had similar value distributions but markedly different autocorrelation profiles.

V. CONCLUSION

This article introduced and systematically evaluated a set of time-series augmentation methods, especially suitable for short, low-dimensional sensor time series, aiming to address the challenges stemming from small dataset size and dataset imbalance. The case of cattle behavior recognition using a CNN-based classifier was considered; however, the results are expected to be relevant beyond the specific example under consideration. The key finding is that the performance is maximized when combining a suitable random sampling scheme with surrogate data and integrating it with the training process to realize online augmentation. In particular, our results demonstrate that dynamically sampling time-series snippets during each epoch can facilitate the training process by expanding the classifier boundaries. Introducing a biased coverage that compensates for the imbalanced original distribution further enhances the performance, not only for the least-represented behavior classes. The issue of limited dataset size is effectively addressed through Fourier surrogates, which support high training performance while avoiding any duplication in the time-domain data submitted to the training algorithm. Extending previous works in which this technique was used empirically, we show how a deductive approach can be used to dissect and identify explicitly which properties retained from the original time series are most relevant. Overall, the proposed method improved the average performance from 90% to 96%, and the classification accuracy of grazing from 45% to 91%, without modifying the classifier architecture.

The potential downsides of the proposed method with respect to other time-series augmentation approaches should be considered. First, since the process of surrogate generation by construction destroys all nonlinear content in the time series, the proposed method entails the assumption that the linear features are sufficient to support high classification performance. The extent to which this assumption holds also for other datasets and applications remains to be clarified. Second, the processes of surrogate generation and sampling do not include any physical assumptions about the specific application, such as, in this case, the influence of collar rotation. Given that such assumptions may be a powerful basis for data augmentation, the proposed method and other approaches should not be viewed as adversarial but eventually combined.

In light of the fact that the proposed method is in principle generally applicable, future work should be directed at systematically investigating other possible scenarios. First, since triaxial accelerometer time series are widely used in monitoring the behavior of other animals as well as human activity, it is expected that the proposed method will lead to improved performance across a broad range of possible applications. Second, since no aspect of the proposed method is specific to accelerometer time series, its application to other multivariate datasets, for example, derived from multiple sensors, should be considered. Third, as the issues of dataset imbalance and size affect network training in general, the usefulness of the proposed method should also be evaluated on other types of networks, such as recurrent neural networks.

Acknowledgment

The authors would like to thank the members of the Research Institute for the Earth Inclusive Sensing (EISESiV), Tokyo, Japan, for their support.

References

[1] D. Fraser, "Understanding animal welfare," *Acta Veterinaria Scandinavica*, vol. 50, no. 1, pp. 1–7, 2008.

[2] E. Tullo, A. Finzi, and M. Guarino, "Review: Environmental impact of livestock farming and precision livestock farming as a mitigation strategy," *Sci. Total Environ.*, vol. 650, pp. 2751–2760, Feb. 2019.

[3] D. Lovarelli, J. Bacenetti, and M. Guarino, "A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production?" *J. Cleaner Prod.*, vol. 262, Jul. 2020, Art. no. 121409.

[4] J. Bartels et al., "TinyCowNet: Memory-and power-minimized RNNs implementable on tiny edge devices for lifelong cow behavior distribution estimation," *IEEE Access*, vol. 10, pp. 32706–32727, 2022.

[5] T. Hasegawa, "Smartphone sensor-based human activity recognition robust to different sampling rates," *IEEE Sensors J.*, vol. 21, no. 5, pp. 6930–6941, Mar. 2021.

[6] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[7] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Dec. 2015.

[8] D. She, X. Lou, and W. Ye, "RadarSpecAugment: A simple data augmentation method for radar-based human activity recognition," *IEEE Sensors Lett.*, vol. 5, no. 4, pp. 1–4, Apr. 2021.

[9] N. Tufek, M. Yalcin, M. Altintas, F. Kalaoglu, Y. Li, and S. K. Bahadir, "Human action recognition using deep learning methods on limited sensory data," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3101–3112, Mar. 2020.

[10] L. Schmeling et al., "Training and validating a machine learning model for the sensor-based monitoring of lying behavior in dairy cows on pasture and in the barn," *Animals*, vol. 11, no. 9, p. 2660, 2021.

[11] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–40, May 2022.

[12] Y. Song, L.-P. Morency, and R. Davis, "Distribution-sensitive learning for imbalanced datasets," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Apr. 2013, pp. 1–6.

[13] N. S. Sani, I. I. S. Shamsuddin, S. Sahran, A. Rahman, and E. N. Muzaffar, "Redefining selection of features and classification algorithms for room occupancy detection," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 8, nos. 2–4, pp. 1486–1493, 2018.

[14] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLoS ONE*, vol. 16, no. 7, Jul. 2021, Art. no. e0254841.

[15] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[18] Q. Wen et al., "Time series data augmentation for deep learning: A survey," in *Proc. Int. Joint Conf. Artif. Intell.*, 2021, pp. 4653–4660.

[19] A. Le Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," in *Proc. ECML/PKDD Workshop Adv. Anal. Learn. Temporal Data*, 2016, pp. 1–9.

[20] K. M. Rashid and J. Louis, "Time-warping: A time series data augmentation of IMU data for construction equipment activity identification," in *Proc. Int. Symp. Autom. Robot. Construct. (IAARC)*, May 2019, pp. 651–657.

[21] M. Kim and C. Y. Jeong, "Label-preserving data augmentation for mobile sensor data," *Multidimensional Syst. Signal Process.*, vol. 32, no. 1, pp. 115–129, Jan. 2021.

[22] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.*, vol. 77, no. 4, pp. 635–638, Jul. 1996.

[23] Y. Ge, X. Xu, S. Yang, Q. Zhou, and F. Shen, "Survey on sequence data augmentation," *J. Frontiers Comput. Sci. Technol.*, vol. 15, no. 7, pp. 1207–1219, 2021.

[24] T. T. Um et al., "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, 2017, pp. 216–220.

[25] N. Takahashi, M. Gygli, B. Pfister, and L. Van Gool, "Deep convolutional neural networks and data augmentation for acoustic event detection," 2016, *arXiv:1604.07160*.

[26] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, Jul. 2020.

[27] H. Patel, D. S. Rajput, G. T. Reddy, C. Iwendi, A. K. Bashir, and O. Jo, "A review on classification of imbalanced data for wireless sensor networks," *Int. J. Distrib. Sens. Netw.*, vol. 16, no. 4, 2020, Art. no. 1550147720916404.

[28] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: An experimental review," *J. Big Data*, vol. 7, no. 1, pp. 1–47, Dec. 2020.

[29] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.

[30] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[31] N. Moniz, P. Branco, and L. Torgo, "Resampling strategies for imbalanced time series," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2016, pp. 282–291.

[32] H. Cao, V. Y. F. Tan, and J. Z. F. Pang, "A parsimonious mixture of Gaussian trees model for oversampling in imbalanced and multimodal time-series classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2226–2239, Dec. 2014.

[33] T. Schreiber and A. Schmitz, "Surrogate time series," *Phys. D, Nonlinear Phenomena*, vol. 142, nos. 3–4, pp. 346–382, 2000.

[34] L. Minati, P. Chiesa, D. Tabarelli, L. D'Incerti, and J. Jovicich, "Synchronization, non-linear dynamics and low-frequency fluctuations: Analogy between spontaneous brain activity and networked single-transistor chaotic oscillators," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 3, Mar. 2015, Art. no. 033107.

[35] C. Li et al., "Data augmentation for inertial sensor data in CNNs for cattle behavior classification," *IEEE Sens. Lett.*, vol. 5, no. 11, pp. 1–4, Nov. 2021.

[36] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Phys. D*, vol. 58, nos. 1–4, pp. 77–94, 1992.
[37] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," 2020, *arXiv:2001.06937*.
[38] F. N. Hatamian, N. Ravikumar, S. Vesal, F. P. Kemeth, M. Struck, and A. Maier, "The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1264–1268.
[39] T. E. K. M. Lee, Y. L. Kuah, K.-H. Leo, S. Sanei, E. Chew, and L. Zhao, "Surrogate rehabilitative time series data for image-based deep learning," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.
[40] J. T. C. Schwabedal, J. C. Snyder, A. Cakmak, S. Nemati, and G. D. Clifford, "Addressing class imbalance in classification problems of noisy signals by using Fourier transform surrogates," 2018, *arXiv:1806.08675*.
[41] T. K.-M. Lee, H. Chan, K. Leo, E. Chew, L. Zhao, and S. Sanei, "Surrogate data for deep learning architectures in rehabilitative edge systems," in *Proc. Signal Process., Algorithms, Archit.*, 2020, pp. 30–35, doi: 10.23919/SPA50552.2020.9241275.
[42] A. da Silva Santos, V. W. C. de Medeiros, and G. E. Gonçalves, "Monitoring and classification of cattle behavior: A survey," *Smart Agricult. Technol.*, vol. 3, Feb. 2023, Art. no. 100091.
[43] A. Stone, "Precision dairy farming technology solutions for detecting dairy cow disease to improve dairy cow well-being," Ph.D. dissertation, Univ. Kentucky, Lexington, KY, USA, 2017, doi: 10.13023/ETD.2017.028.
[44] M. Fukawa and Y. Oko. (2022). *Cattleuser/Spresense_EVK-701_RECORDER*. Zenodo. [Online]. Available: https://zenodo.org/record/5848608#.Y2VFuHZBzIU
[45] H. Ito. *Japanese Black Beef Cow Behavior Classification Dataset*. Zenodo. [Online]. Available: https://zenodo.org/record/5848608#.Y2VFuHZBzIU
[46] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017.
[47] T. Azuroğlu and H. Oğul, "A deep learning approach for Parkinson's disease severity assessment," *Health Technol.*, vol. 12, no. 5, pp. 943–953, Sep. 2022.
[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[49] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Jun. 2011, pp. 315–323.
[50] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[52] M. Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: http://tensorflow.org
[53] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.*, vol. 2, pp. 37–63, Dec. 2011.
[54] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–6, Dec. 2020.
[55] Mathworks Inc. (2022). *Surrogate Data*. [Online]. Available: https://it.mathworks.com/matlabcentral/fileexchange/4612-surrogate-data

Chao Li (Graduate Student Member, IEEE) received the B.E. degree in electronic information and electrical engineering from the Dalian University of Technology, Dalian, China, in 2017, and the M.E. degree in electrical and electronic engineering from the Tokyo Institute of Technology, Yokohama, Japan in 2020, where she is currently pursuing the Ph.D. degree.

She has been the Chair of the IEEE Student Branch, Tokyo Institute of Technology, since 2021. Her research interests include data collection for machine learning, activity classification at the edge, and low-power Internet-of-Things devices.

Ms. Li is a Student Member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Ludovico Minati (Senior Member, IEEE) received the B.S. degree in information technology and computing and the M.S. degree in science from The Open University, Milton Keynes, U.K., in 2004 and 2006, respectively, the M.S. degree in applied cognitive neuroscience from the University of Westminster, London, U.K., in 2008, the B.S. degree in physical science and the M.S. degree in medical physics from The Open University in 2009, the Ph.D. degree in neuroscience from the Brighton and Sussex Medical School, Falmer, U.K., in 2012, and the D.Sc. (doktor habilitowany) degree in physics from the Institute of Nuclear Physics, Polish Academy of Sciences, Kraków, Poland, in 2017.

He has held research and consulting roles across public institutions and private companies, including the Institute of Nuclear Physics, Polish Academy of Sciences, the Carlo Besta Neurological Institute, Milan, Italy, and the Brighton and Sussex Medical School. He is currently a Specially Appointed Associate Professor with the Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan, an Affiliate Research Fellow with the Center for Mind/Brain Sciences, University of Trento, Trento, Italy, and a Freelance Research and Development Consultant. He has authored more than 150 articles and several patents. His research interests include nonlinear dynamical systems, chaotic oscillators, reconfigurable analog and digital computing, functional magnetic resonance imaging, advanced techniques for biosignal analysis, brain–machine/computer interfaces, and robotics.

Dr. Minati is a European Engineer (Eur. Ing.), a Chartered Engineer, and a member of the Institution of Engineering and Technology, U.K; a Chartered Physicist and a member of the Institute of Physics, U.K.; and a Chartered Scientist and a member of the Institute of Physics and Engineering in Medicine, U.K. He is also a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), the Institute of Electrical Engineers of Japan, and the Japan Neuroscience Society.

Korkut Kaan Tokgoz (Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, in 2009 and 2012, respectively, and the M.Eng. and Ph.D. degrees from the Department of Physical Electronics, Tokyo Institute of Technology, Tokyo, Japan, in 2014 and 2018, respectively.

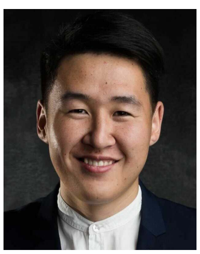
From 2018 to 2019, he worked as a Senior Researcher/Assistant Manager at NEC Corporation, Kawasaki, Japan, where he was involved in 5G systems and fixed point-to-point wireless links. From 2019 to 2022, he worked as an Assistant Professor at the Tokyo Institute of Technology. He is currently working as a Faculty Member with the Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Turkey. He is also the Co-Founder and CTO of Evrim Company Ltd., Yokohama, Japan. His research interests include analog RF, millimeter-wave and subterahertz transceivers for wireless communications, low-power edge artificial intelligence (AI) for monitoring systems, the Internet of Things (IoT), sensors and systems, deembedding, device characterization, and high-power, high-efficiency power amplifiers (Pas) for wireless systems.

Dr. Tokgoz was a recipient of several awards, scholarships, and grants, including the TUBITAK 2232-B International Fellowship for Early Stage Researchers in 2022, the Marie Skłodowska-Curie Actions Post-doctoral Fellowship in 2022, the SSCS Predoctoral Achievement Award in 2018, the IEEE MTT-S Graduate Student Fellowship in 2017, the IEICE Student Encouragement Prize in 2017, the Seiichi Tejima Overseas Student Research Award, and the IEEE/ACM ASP-DAC University LSI Design Contest 2017 Best Design Award.

Masamoto Fukawa joined TechnoPro Design, Tokyo, Japan, in 2017. He is currently working on a joint research project with the Tokyo Institute of Technology, Yokohama, Japan, where his tasks involve machine learning and data science.

Jim Bartels (Graduate Student Member, IEEE) received the B.Sc. (Hons.) degree in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2019, and the M.E. degree from the Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, Japan, in 2021. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with the Tokyo Institute of Technology.

He is also the Vice-Chair of the IEEE Student Branch, Tokyo Institute of Technology. His current research interests include embedded machine learning, spiking neural networks, quantization, and neural architecture search.



Sihan A (Graduate Student Member, IEEE) received the B.Sc. (Agr.) degree in animal science from China Agricultural University, Beijing, China, in 2016. He is currently pursuing the M.E. degree in electrical and electronic engineering with the Tokyo Institute of Technology, Yokohama, Japan.

He is interested in farm applications of deep learning on edge devices to improve both animal welfare and production.

Ken-ichi Takeda received the B.Agr. degree from Nippon Veterinary and Animal Science University, Tokyo, Japan, in 1995, and the M.Agr. and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1997 and 2000, respectively.

He is currently an Associate Professor at the Institute of Agriculture, Academic Assembly, Shinshu University, Nagano, Japan. His research interests are farm animal welfare and herbivores management in grassland.

Hiroyuki Ito (Member, IEEE) received the B.E. degree from the Department of Electronics and Mechanical Engineering, Chiba University, Chiba, Japan, in 2002, and the M.E. and Ph.D. degrees from the Department of Advanced Applied Electronics, Tokyo Institute of Technology, Yokohama, Japan, in 2004 and 2006, respectively.

From 2004 to 2007, he was a Research Fellow of the Japan Society for the Promotion of Science. He was a temporary Visiting Researcher and a Visiting Professor with the Communications Technology Laboratory, Intel Corporation, Hillsboro, OR, USA, in 2006 and 2007, respectively. He was an Assistant Professor at the Precision and Intelligence Laboratory, Tokyo Institute of Technology, from 2007 to 2013, where he was an Associate Professor from 2013 to 2015. From 2008 to 2010, he was with Fujitsu Laboratories Ltd., Yokohama, where he developed an RF CMOS transceiver and digital calibration techniques for mobile-WiMAX applications. Since 2016, he has been an Associate Professor with the Institute of Innovative Research, Tokyo Institute of Technology. He has been the Co-Funder/CEO of Evrim Company Ltd., Yokohama, since 2020. His research interests include ultralow-power RF circuits, a high-sensitivity MEMS accelerometer, a low-noise and crystal-less RF synthesizer, a cow management system exploiting the Internet of Things (IoT), and deep/machine learning.

Dr. Ito is a member of the IEEE Solid-State Circuits Society, the Institute of Electronics, Information and Communication Engineers (IEICE), the Japan Society of Applied Physics (JSAP), and the Japanese Society of Agricultural Machinery and Food Engineers. He is also an International Technical Program Committee Member of the International Solid-State Circuits Conference. He was a recipient of the European Solid-State Circuits Conference (ESSCIRC) 2008 Best Paper Award.