Journal of Agriculture and Food Research 15 (2024) 101021

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Multi-algorithmic approach for detecting outliers in cattle intake data

Jae-Min Jung [a], Dong-Hyeon Kim [a], Hyunjin Cho [b], Mingyung Lee [b], Jinhui Jeong [a], Dae-Hyun Lee [a], Seongwon Seo [b,**], Wang-Hee Lee [a,c,*]

[a] Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon, 34134, South Korea
[b] Division of Animal and Dairy Sciences, Chungnam National University, Daejeon, 34134, South Korea
[c] Department of Smart Agriculture Systems, Chungnam National University, Daejeon, 34134, South Korea

A R T I C L E I N F O

Keywords:
Anomaly detection
Cattle intake
Machine learning
Statistics
Voting process

A B S T R A C T

Monitoring cattle feed intake is crucial for evaluating animal health, productivity, and farm profitability. In particular, an abnormal intake is related to the cattle activity. Therefore, outlier detection forms the basis for intake monitoring. This study employed multiple algorithms ranging from statistics to deep learning to detect outliers in time-series data of cattle intake. We used five models implementing mean + standard deviation, moving average, box plot, time series decomposition, and autoencoder, and attempted to enhance the detection performance by a voting system to combine more than one model. Both box plot and time-series decomposition demonstrated high accuracy (over 95 %) and F1-score (harmonic mean of precision and recall). Thus, it reliably distinguished normal values from outliers. Moving average exhibited a high true-skill statistic (TSS), thereby rendering it suitable for outlier detection. The voting system gave F1 and TSS scores of 0.49 and 0.65, respectively. Thus, it enhanced the detection performance compared with the individual model. These results demonstrate that the performance metrics vary depending on the type of algorithm. This, in turn, highlights the need to select algorithms adapted to the monitoring objectives. The algorithmic selection can be complemented by a voting system. This demonstrates its potential for generating a reliable database with accurate outlier detection and aiding decision-making by livestock producers.

1. Introduction

Feed intake is closely related to the productivity. It is an important indicator of the diet preference, health, and physiological state of animals. Measurements of feed intake are necessary to monitor the status of the animals; measure the feed efficiency, which directly affects the profitability of the farm [1,2]; and provide feeding that satisfies the animal's nutrient requirements without excessive or insufficient nutrients [3]. Notwithstanding its importance, it has conventionally been highly difficult to measure feed intake on farms. This is because it is labor-intensive and time-consuming. However, recent applications of technologies such as radio-frequency identification (RFID) and wearable biosensors in the livestock industry have become the means to solve these problems.

Although technological advancements have alleviated the difficulty of collecting data on feed intake, analyzing large amounts of collected data remains a challenge. Because of the emphasis on precision and smart livestock farming, technologies capable of analyzing real-time, time-series, and large-sized data acquired through sensor technology are required [4,5]. Moreover, with advances in monitoring and sensing techniques, large amounts of data have been generated in the form of time series to effectively manage livestock farming [6–10]. However, livestock data are highly likely to vary significantly owing to the harsh and variable environments of livestock farming. This generally impedes decision-making in livestock management [11,12]. Time-series data are generally influenced by unusual observations that may be caused by equipment malfunctions, errors in data recording, and an abrupt interference in the measurement target [6,13,14]. These outliers reduce the quality of the time-series data and yield biased results. This poses a significant challenge in extracting latent information through data analysis [9,10,15]. Therefore, outlier detection is important to identify and remove large deviations in time-series data [6,13,15,16]. Accordingly, efforts have been undertaken to enhance the quality of livestock monitoring data and process data abnormalities [17–19]. In particular,

* Corresponding author. Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon, 34134, South Korea.
** Corresponding author.
E-mail addresses: swseo@cnu.ac.kr (S. Seo), wanghee@cnu.ac.kr (W.-H. Lee).

https://doi.org/10.1016/j.jafr.2024.101021
Received 14 October 2023; Received in revised form 14 December 2023; Accepted 26 January 2024
Available online 28 January 2024

neural network was applied to detect outliers in time series data under different scenarios with reliable performance by advancing previously developed algorithms [9,10,20,21]. However, their algorithms are too complicated for practical users working in a livestock farm. Practically, simple and easy real-time monitoring is required for livestock farmers on a daily basis, necessitating a practical method for destecting outliers without large computational cost and effort.

This study aimed to develop different statistical and machine learning algorithms for detecting outliers in the time-series feed intake data of cattle. A simple statistical approach to an unsupervised deep learning algorithm was applied to identify outliers. Moreover, the detection results were compared to outliers manually curated by experts. Based on this comparison, we ultimately attempted to identify the optimal algorithm available for feed intake data of cattle. It could be used to diagnose abnormalities in system monitoring as well as in cattle physiology.

2. Materials and methods

2.1. Experimental animals

The experiments were conducted at the Center for Animal Science Research, Chungnam National University, Korea. The use of animals and protocols for this experiment were reviewed and pre-approved by the Chungnam National University Animal Research Ethics Committee (CNU- 01021).

Fifty-nine-month-old Hanwoo steers participated in this experiment (average body weight: 256 ± 22.7 kg). The steers were housed in one of the four pens. Each pen (10 × 10 m) was equipped with an automatic concentrate feeding station (ACFS) and four forage feed bunks. Each ACFS (1.9 × 2 × 0.8 m) and forage feed bunk automatically measured the individual feed intake. It identified each steer by detecting an RFID neck tag attached to it (Dawoon Co., Incheon, Korea). Forage was fed *ad libitum* two times a day at 0800 and 1800 h. The steers were provided with the amount of daily permissible concentrate mix divided into four or six times through the ACFS. Within a pen, the steers consumed an identical diet (i.e., forage and concentrate mix).

2.2. Data acquisition and processing

Among the 50 steers, 40 were selected based on the body weight and birth time to minimize the external factors that could influence their intake (Fig. 1). The intake data were collected over a period ranging from 283 days to 496 days. A record of the intake (kg) of each of the 40 individuals over time (in days) was retrieved from the raw data. Consequently, a time-series database that recorded the daily intake of each steer was constructed. In the database, four specific cases were identified to classify outliers: 1) intake values of zero, 2) sensor detachment from a measurement target, 3) intake measurements captured on the day of blood and gastric juice collection, and 4)

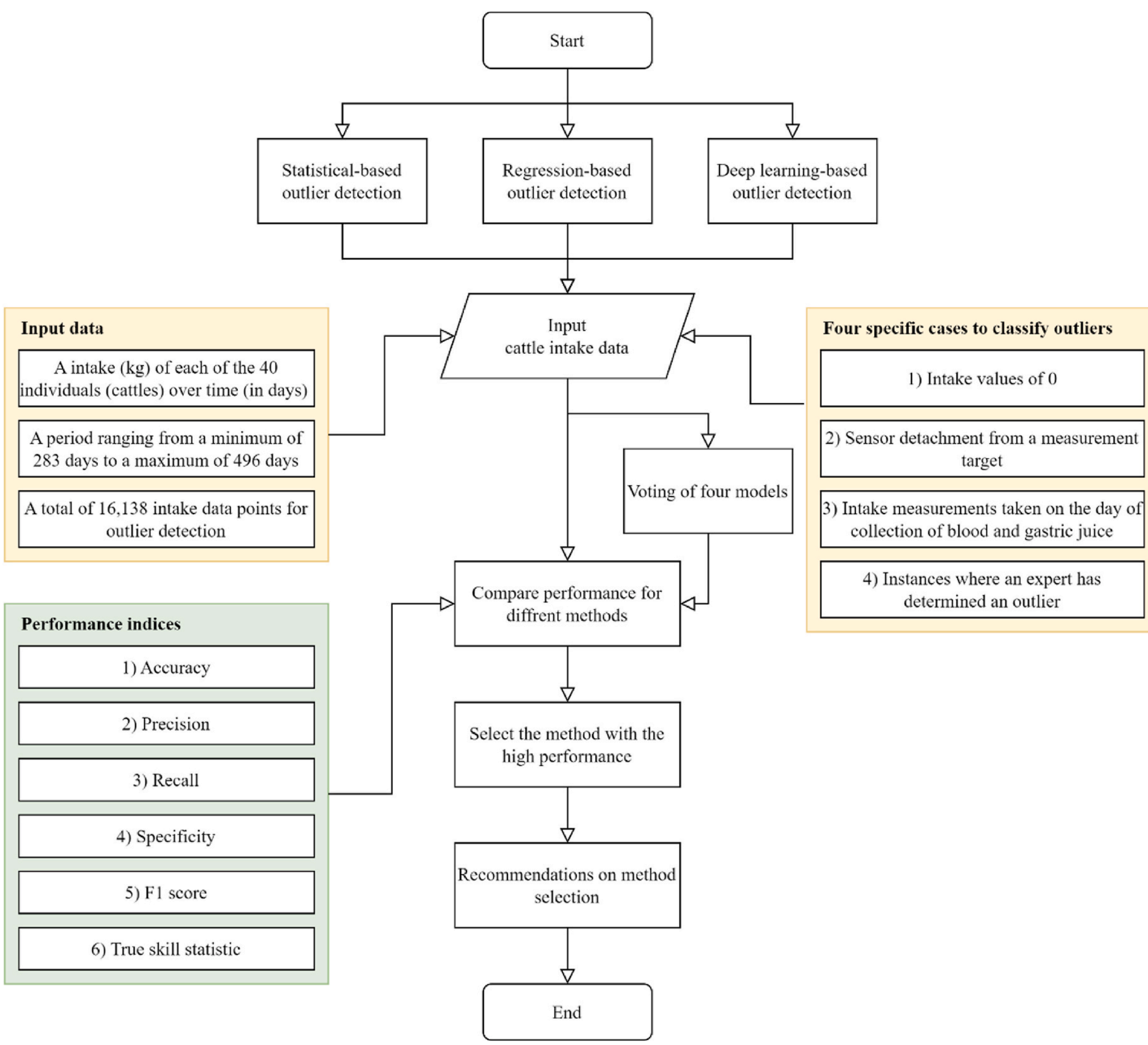

Fig. 1. Process of outlier detection methods.

instances in which an expert determined a rapid increase or decrease in intake (Fig. 1). Finally, 16,138 intake data points with 754 labeled outliers were verified using the format required by the software coding algorithms. Because the first and final data points could not be used, 741 labeled outliers were used for the moving average. The autoencoder used 11,360 data points with 497 labeled outliers that were truncated to match the shortest data length (283 days) because it required uniform data lengths for all the steers.

2.3. Developing outlier detection algorithms

2.3.1. Simple descriptive statistical methods

Simple statistical methods were initially employed to identify the outliers. Outliers in cattle intake were identified using three statistical methods (box plot, average ± standard deviation, and moving average ± average variation). These three statistical methods are employed because they are easily and quickly applicable in practical fields, devoid of large efforts for interpretation of results, and computational load. First, we used Tukey's fences based on a box plot of the intake data [22, 23]. The box plot identifies outliers based on the distribution (quartile) of the data, allowing it to capture the distribution characteristics of individual cattle intake data. This method employs quartiles, specifically utilizing the Interquartile Range (IQR), which encompasses the 50 % portion between the 25th (1st quartile) and 75th (3rd quartile) percentiles. To identify outliers, the IQR is multiplied by a specified value, determining the lower fence (1st quartile – k × IQR) and upper fence (3rd quartile + k × IQR). Any data falling outside this range is considered an outlier. In our study, we adopted the default value of 1.5 for multiplying the IQR. Second, the outliers were screened using a value deviating from the mean by more than the standard deviation (Equation (1)). This method is a classic approach, akin to a box diagram, but easy to apply for detecting outliers, making popular in livestock farms as users. Initially, this method established the outlier detection threshold by multiplying the standard deviation by different factors: 1 (32 % of significance level 32 %), 1.96 (5 % of significance level 5 %), and 2.58 (1 % of significance level 1 %). However, when using the factors 1.96 and 2.58, outlier detection proved nearly impossible. Consequently, a single standard deviation was employed for this purpose. Third, we developed an algorithm based on the moving algorithm and overall average variation (Equation (2)). Moving average can detect outliers by analyzing trends with consideration of preceding and following values. In addition, moving average is especially valuable for identifying sudden changes in momentary data, enabling the detection of anomalies such as sensor failure, machine issues, or cattle diseases. We identified that the intake to be classified as an outlier was most strongly associated with the values measured on the previous day and the next day. In addition, as this method was conceived for real-time monitoring, we utilized three time points (t = 3). Accordingly, we calculated the three-day moving average at three time points (measurement day - 1 day, measurement day, and measurement day + 1 day). Meanwhile, for all intakes, individual variation was computed by subtracting the intake measured the previous day, and subsequently, the total average of these variation was calculated. The algorithm then detected an outlier that deviated from the criteria at all the three time points.

$$\frac{\sum_{i=1}^{n} x_i}{n} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \le x_i \le \frac{\sum_{i=1}^{n} x_i}{n} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \quad \text{Equation 1}$$

where n is the total number of individual data.

$$\frac{1}{t}\sum_{j=n-k+1}^{n} x_j - \frac{\sum_{j=2}^{n}|x_{j-1} - x_j|}{n-1} \le x_j \le \frac{1}{t}\sum_{j=n-k+1}^{n} x_i + \frac{\sum_{j=2}^{n}|x_{j-1} - x_j|}{n-1} \quad \text{Equaton 2}$$

where k is the window size (t = 3).

2.3.2. Times series decomposition

Time-series decomposition is a filtering procedure that classifies time-series data into latent components of trends, seasonality, and residuals [24]. Time series decomposition breaks down time series data into components like trends, seasonality, and residuals, facilitating the identification of structural changes in the time series. This method is particularly suitable for growing cattle, considering that intake often increases as cattle grow. Therefore, it was chosen as a method adept at determining outliers in cattle intake. Residual is the remainder component of the time-series data after eliminating the trend and seasonal variations. It is assumed to be a random variable with a constant variance [25–27]. Thus, a data point with a large residual is likely an outlier that deviates from the general trend or seasonality [13,15].

In this study, we decomposed the intake data of individual cows to identify the residuals and their distribution. The additive model that was converted from the multiplicative model by determining the logarithm of the variables was employed with a weekly seasonality (Equation (3)). Because the residuals conformed to a normal distribution with zero mean and a constant variation, the intake data with a residual outside the 95 % confidence interval (i.e., more than two times the standard deviation from the average of the residuals) were regarded as outliers. This algorithm was coded by Python version 3.7 with the "seasonal_decompose" library.

$$\mathrm{x(t) = tr(t) \times s(t) \times r(t)}$$

$$\mathrm{\log[x(t)] = \log[tr(t) \times s(t) \times r(t)]}$$

$$\mathrm{\log[x(t)] = \log[tr(t)] + \log[s(t)] + \log[r(t)]} \quad \text{Equaton 3}$$

where t, x(t), tr(t), s(t), and r(t) are the time, observed data, trend, seasonality, and residual, respectively, at t.

2.3.3. Autoencoder

An autoencoder is an unsupervised neural network used to generate output data by learning the encoded characteristics of the input [16]. It generally consists of encoding and decoding layers with hidden layers between these. The encoder extracts information from the input data, and the decoder regenerates an output close to the input data based on the information through the hidden layers. The autoencoder is trained to learn the compressed pattern of the input data. Therefore, this technique has been used to detect abnormalities, i.e., outputs that do not conform to the information in the input data [28].

We developed a long short-term memory (LSTM) anti-encoder composed of three layers of encoder and decoder with hyperparameters of 14 and 30 for the window size and batch size, respectively, for standardized data conforming to N (0,1). The epoc was set to 50, and the model with the best learning results was obtained. The discrepancy between the input and output data reconstructed by the autoencoder was used to detect the outliers in the cow intake data. The threshold discrepancy was set to 0.9. We varied the window size and batch size from 2 to 7, respectively, up to a maximum of 30. Additionally, the threshold and epoch were adjusted from 0.5 to 1.5 and 50 to 200, respectively, aligning with the previous hyperparameter settings to select the values that most appropriately distinguish between outliers and normal values. The input data that exceeded this threshold were regarded as outliers. This algorithm was coded by Python version 3.7 with the "TimeseriesGenerator," "Sequential Dense," "Dropout," "LSTM," "RepeatVector" and "TimeDistributed" libraries.

2.4. Voting of four models for optimal outlier detection

To improve the performance of outlier detection, a voting process that utilizes two or more algorithms consecutively in combination is required. However, because of the relatively small dataset in the autoencoder and its low performance compared with the other models, we excluded it from this method.

In the voting process, the number of positive votes among the four algorithms was counted, and the outliers were identified based on the counted algorithms, i.e., the number of votes. We evaluated the outlier detection performance by varying the number of votes determining outliers from one (at least one algorithm collectively detected an outlier) to four (the four algorithms collectively detected an outlier), and determined the optimal voting process at the highest detection performance.

2.5. Performance evaluation

The actual outliers of 754 owing to record errors or eating disorders were manually assessed and marked by experts so that these could be compared with the systemic outliers detected by the outlier detection algorithms (Fig. 1). The actual and predicted outliers were compared to calculate the confusion matrix by determining the number of true negatives, true positives, false negatives, and false positives for each outlier detection algorithm. To minimize the uncertainty of the evaluation using an individual performance metric, we measured the predictive performance using multiple metrics. In this study, we used accuracy, precision, recall (sensitivity), specificity, F1-score (harmonic mean of precision and recall), and true skill statistics (TSS) to evaluate the

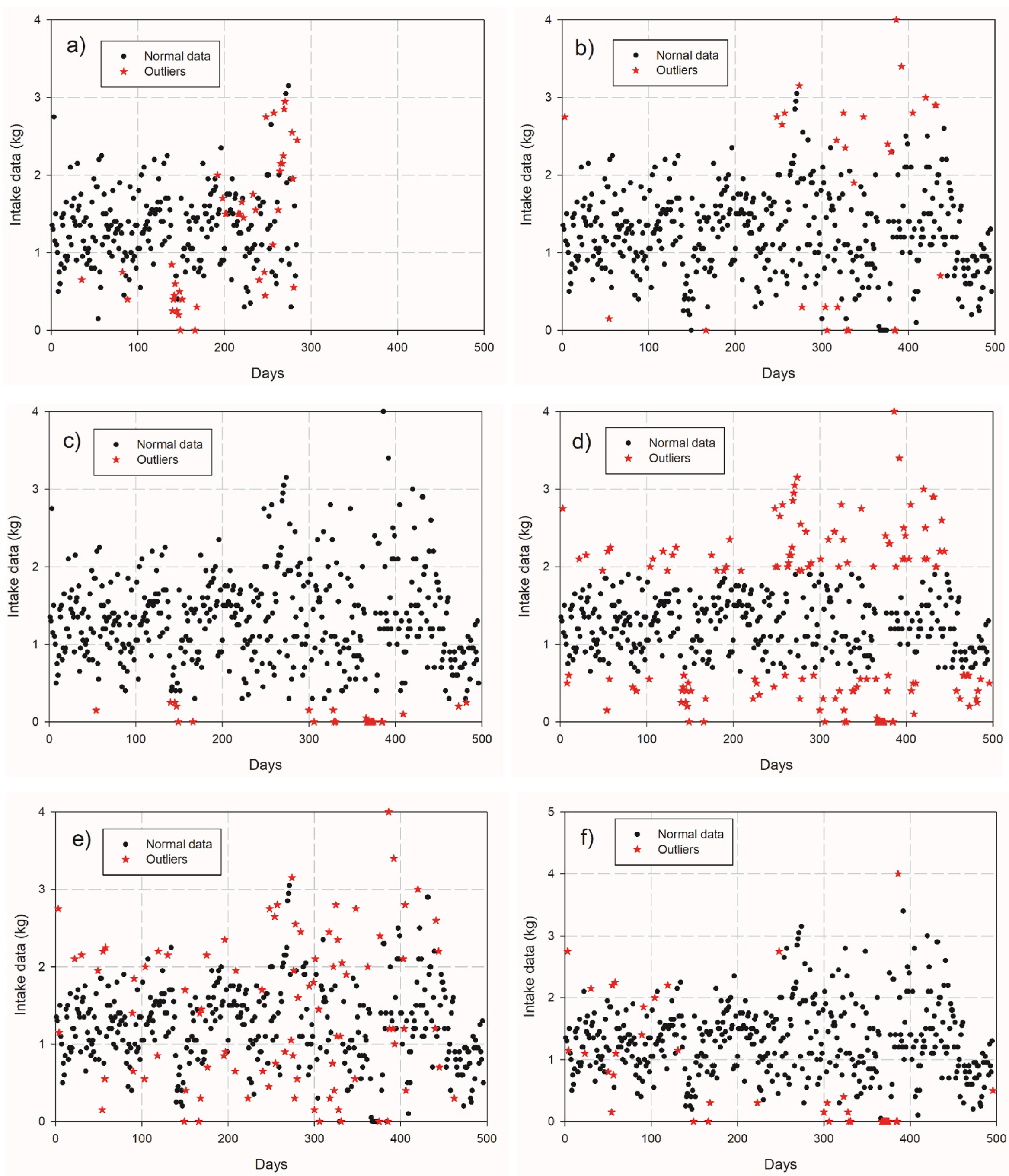

Fig. 2. Outliers depicted in representative cattle intake data by a) autoencoder, b) time series decomposition, c) box plot, d) average ± standard deviation e) moving average, and f) marking actual outliers.

performance of each developed algorithm [29–31].

3. Results

3.1. Result of outlier identification by algorithms

The outliers predicted by the mean and standard deviation accounted for 30.4 % of the total data, with 4673 outliers identified (Fig. 2 and Table 1). This method detected all the outliers recording zero value. However, there was a tendency to predict data assessed to be normal intake, as outliers. On an average, 117 outliers were detected for each cow. A maximum of 161 were detected in the longest data length of 496 days, and a minimum of 65 outliers were detected in the shortest data length of 218 days. The ratio of outliers to the number of data points ranged from 21 % to 33.6 %. Thus, it classified a larger number of data points as outliers than the other algorithms in the first study.

The moving average identified 16.9 % of the data as outliers, which corresponded to 2592 instances (Fig. 2 and Table 1). An advantage of the moving average method is that it can dynamically detect outliers and predict values that show rapid variations in a short period. However, the method is limited to detecting outliers when data values continuously approach zero. This resulted in the outlier detection criterion converging to zero and thereby, failure to detect outliers. On an average, 65 outliers were identified for each steer, and the ratio of outliers to the number of data points ranged from 13.1 % to 19.4 %. Similar to the previous method, the largest outlier (89 data points) had the longest data length of 496 days, whereas the smallest outlier (47 data points) had the shortest data length of 218 days. The second-largest amount of data was identified as outliers using this approach.

The box plot identified 2.8 % of the data as outliers, which corresponded to 428 identifications (Fig. 2 and Table 1). The box plot detected all the outliers with a value of zero and predicted the smallest number of outliers among all the methods. The largest and smallest numbers of outliers by this method were 51 and 2, respectively, thereby averaging 11 outliers for each cow. The total length of the data was unrelated to the number of outliers identified. The ratio of outliers to the number of data points ranged from 0.3 % to 10 %, which was the lowest number of outliers identified by any of the algorithms.

Time series decomposition identified 513 data points (3.3 % of the total data). It displayed an outlier identification similar to that of the box plot-based method (Fig. 2 and Table 1). On an average, 13 outliers were detected for each cow. A maximum of 40 and minimum of 1 were predicted as outliers. These corresponded to ratios of 10 % and 0.3 %, respectively, in terms of the ratio of outliers to the number of data points. Notwithstanding the strength of the time series decomposition in the pattern analysis of time-series data, the current cattle intake data did not exhibit any specific pattern over time. This limited the detection performance of this approach.

Unlike the other models, the autoencoder could not identify outliers in the data from individual cattle. It rather used the entire set of cattle data by date to identify abnormal intake. All the steers had outliers marked on the same dates. This resulted in a total of 1600 outliers, with 42 outliers per entity for 40 entities detected from a dataset spanning 283 days (Fig. 2 and Table 1). These outliers accounted for 14.8 % of the total data. In general, an autoencoder is used to detect outliers occurring on a specific date for data of an equal length. This implies that this method is likely to be unsuitable for detecting outliers on an individual basis.

3.2. Performance evaluation by manually curated outliers

The accuracies of the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average were 0.84, 0.95, 0.96, 0.74, and 0.86, respectively. These generally displayed a reasonable accuracy for outlier detection (Table 2). The box plot and time series decomposition exhibited accuracies of over 95 % and a true negative rate of 94 %. Thus, these demonstrated high reliability in not misclassifying normal values as outliers. Conversely, the mean ± deviation method showed the lowest accuracy. It displayed a false positive rate of 25 %, thereby indicating a higher likelihood of incorrectly detecting a normal value as an outlier. The high accuracy of over 80 % except for the method using mean ± deviation indicated its potential capability of outlier detection [32].

The precision values for the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average were 0.11, 0.49, 0.57, 0.12, and 0.2, respectively. These were relatively low except for the box plot and time series decomposition (Table 2). The reason for the higher precision in the time series decomposition and box plot-based approaches is the relatively small number of data points classified as

Table 2
The performance of five methods based on accuracy, precison, recall, specificity, true skill statistic, and F1-score.

	Autoencoder	Time series decomposition	Box plot	Average ± standard deviation	Moving average
Accuracy[a]	0.84	0.95	0.96	0.74	0.86
Precison[b]	0.11	0.49	0.57	0.12	0.2
Recall[c]	0.36	0.34	0.33	0.77	0.69
Specificity[d]	0.86	0.98	0.99	0.73	0.86
True skill statistic[e]	0.22	0.32	0.31	0.5	0.55
F1-score[f]	0.16	0.4	0.42	0.21	0.31

[a] Accuracy was calculated as the true positive/(true positive + false positive + false negative + true negative).
[b] Precision was calculated as the true positive/(true positive + false positive).
[c] Recall was calculated as the true positive/(true positive + false negative), and.
[d] Specificity was calculated as the true negative/(false positive + true negative).
[e] The true skill statistic (TSS) was calculated as the recall + specificity - 1.
[f] The F1-score was computed as 2 × {(precision × recall)/(precision + recall)}.

Table 1
Confusion matrix of five methods for detecting outliers.

		Predicted class									
		Positive (outlier)					Negative (normal)				
		LSTM[a]	TSD[b]	BP[c]	A ± S[4]	MA[e]	LSTM	TSD	BP	A ± S	MA
Actural class	Positive (outlier)	178	253	246	581	510	319	501	508	173	231
	Negative (normal)	1502	261	182	4093	2082	9361	15,123	15,202	11,291	13,235

[a] LSTM is the Long Short Term Memory antoencoder.
[b] TSD is the time series decompostion.
[c] BP is the box plot.
[d] A ± S is the averger ± standard deviation, and.
[e] MA is the moving average.

outliers, which resulted from their low false positives (1.6 % and 1.1 %, respectively) compared with the other methods. Consequently, both time series decomposition and box plot methods could be used to prevent false identification of normal intake levels as outliers. This is essential for the continuous monitoring of cattle intake.

The recall values for the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average were 0.36, 0.34, 0.33, 0.77, and 0.69, respectively (Table 2). All the methods showed high specificities of 0.86, 0.98, 0.99, 0.73, and 0.86 for the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average, respectively. In all the methods except for that using mean ± standard deviation, the specificity values were higher than the recall values. The methods using mean ± standard deviation and moving average had high recall values because their true positive values were approximately two times higher than those of the other models. The high true positive value was owing to the large number of data points classified as outliers, which included a significant portion of the actual outliers. In contrast, the false-negative values were half those of the other models, which resulted in relatively high recall values. Although the autoencoder, time-series decomposition, and box plot demonstrated significantly high specificities, their ratios to the recall value were significantly skewed. Meanwhile, mean ± deviation and moving average displayed specificities marginally lower than those of the other methods. However, their ratio with the recall value was close to 1:1. This indicated a good performance [29].

The TSS and F1-score are comprehensive evaluation methods. These exhibit relatively low performances in all models [29,30]. The moving average method achieved the highest TSS, whereas the box plot method obtained the best F1-score. The TSS values for the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average were 0.22, 0.32, 0.31, 0.5, and 0.55, respectively (Table 2). Notably, the method using mean ± deviation and moving average showed relatively high TSS values. This is reasonable given their high recall and specificity. The F1-scores for the autoencoder, time series decomposition, box plot, mean ± deviation, and moving average were 0.16, 0.4, 0.42, 0.21, and 0.31, respectively (Table 2). The low performance of the autoencoder in all the indicators resulted from its failure to detect individual outliers for each steer, whereas higher values were obtained in the time series decomposition and box plot-based methods because of the true negative rate. Considering both TSS and the F1-score, the moving average was determined to be the best method for detecting outliers, followed by the box plot.

3.3. Result and performance of outlier identification by voting process

A total of 89 outliers were collectively detected by the four algorithms (received four votes). Of these 69 were actual outliers. This demonstrated a precision of 77.5 %. However, this approach has a significant limitation in that it identifies only a small number of outliers. When the outliers were determined by only one model, the voting process identified the most outliers, totaling 663 (Table 3). However, this resulted in a trade-off because the number of false positives was high at 5208. This indicates a problem in determining the optimal number of votes owing to the trade-off between precision and recall (Table 4). Therefore, it was evident that the confusion matrix and the performance of outlier detection were significantly affected by the number of votes and amount of data labeled as outliers (Table 4).

The highest accuracy (0.96) was achieved using more than two models. However, it was lower than the accuracy of the box plot method (Fig. 3). Precision attained the highest value of 0.78 with the use of all the four models, whereas the recall was significantly low at 0.09 (Table 4). However, the precision decreased to 0.11, whereas the recall increased to 0.88 with the use of at least one model for outlier detection (Table 4). The F1-score showed the highest value (0.49) with voting of three models, thereby surpassing the F1-scores of the box plot and autoencoder methods by 0.07, and 0.33, respectively (Table 4). Meanwhile, the highest TSS (0.65) was achieved with voting of two models. This outperformed the moving average and autoencoder methods by 0.1 and by 0.43, respectively (Table 4).

4. Discussion

Detecting outliers in cattle intake facilitates early disease diagnosis and intervention. It also prevents potential economic losses by early decisions on appropriate actions, such as the optimal slaughter times [20]. Moreover, outlier detection technology can be applied to smart livestock monitoring systems. This would enable the real-time management of cattle [20,33].

Because of its simplicity and comprehensibility, this statistical approach serves as a straightforwardly accessible method for livestock producers to identify outliers [32]. The methods utilizing averages (mean ± standard deviation and moving average) identified a relatively larger amount of data as outliers. This indicated that their criteria for detecting outliers were lenient compared with the other models. Thus,

Table 4
The accuracy, precision, recall, specificity, true skill statistic, and F1-score of voting methods were evaluated for detection in at least one model (vote 1), more than one model (vote 2), more than two models (vote 3), and by all the models (vote 4), respectively.

	Vote 1	Vote 2	Vote 3	Vote 4
Accuracy[a]	0.67	0.91	0.96	0.95
Precison[b]	0.11	0.31	0.62	0.78
Recall[c]	0.88	0.73	0.41	0.09
Specificity[d]	0.66	0.92	0.99	≒1
True skill statistic[e]	0.54	0.65	0.4	0.09
F1-score[f]	0.2	0.44	0.49	0.16

[a] Accuracy was calculated as (true positive + true negative)/(true positive + false positive + false negative + true negative).
[b] Precision was calculated as true positive/(true positive + false positive).
[c] Recall was calculated as true positive/(true positive + false negative), and.
[d] Specificity was calculated as true negative/(false positive + true negative).
[e] The true skill statistic (TSS) was calculated as recall + specificity - 1.
[f] The F1-score was computed as 2 × {(precision × recall)/(precision + recall)}.

Table 3
Confusion matrix of voting methods for detecting outliers.

		Predicted class							
		Positive (outlier)				Negative (normal)			
		Vote 1[a]	Vote 2[b]	Vote 3[c]	Vote 4[d]	Vote 1	Vote 2	Vote 3	Vote 4
Actural class	Positive (outlier)	663	547	310	69	91	207	444	685
	Negative (normal)	5208	1196	194	20	10,176	14,188	15,190	15,364

[a] Vote 1 indicates that an outlier was detected in at least one model.
[b] Vote 2 indicates that an outlier was detected by more than one models.
[c] Vote 3 indicates that an outlier was detected by more than two models, and.
[d] Vote 4 indicates that an outlier was detected by all the models.

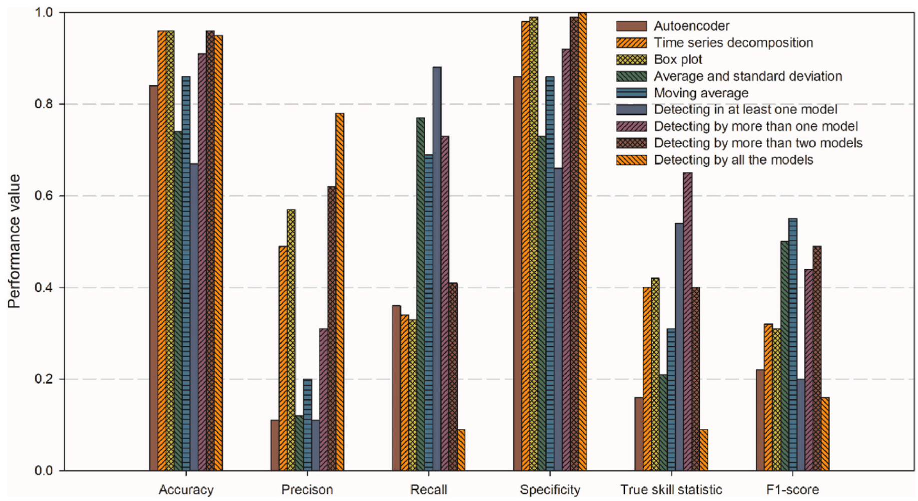

Fig. 3. Performance value of autoencoder, time series decomposition, box plot, average ± standard deviation, moving average, and voting method wherein an outlier is detected by more than one model.

this approach is more efficient when a livestock producer seeks to sensitively identify abnormal intake. In the box plot method, the criteria for assessing outliers were not more stringent than those in the other methods. Therefore, it detected fewer outliers than the actual labeled outliers. This method would be suitable in cases where continuous monitoring of normal intake is more important than the sensitive detection of outliers. However, this could result in a high likelihood of failure to identify abnormal intake.

Machine learning is a highly effective tool for data classification. This enhances its application to outlier detection [32]. However, it does not exhibit a high performance in all the cases. Moreover, its application is occasionally constrained by the data characteristics. Time-series decomposition is a highly effective method for analyzing data with identifiable trends, seasonality, and residuals [25]. Nonetheless, the application of this method to data without periodic patterns is challenging. Furthermore, the patterns exhibited varying forms across individuals. This limited the application of time series decomposition to the detection of outliers. The autoencoder was demonstrated to be suitable for detecting outliers or errors occurring on specific dates. However, it was not suitable for different individual intake patterns. In addition, the autoencoder method requires an equal data length for each steer [34].

The outlier detection methods exhibited distinct characteristics based on the algorithms. When the recall value was high, the precision tended to be low, and *vice versa*. An attempt to identify the maximum number of labeled outliers could result in the erroneous inclusion of a large amount of normal data. For example, the approaches employing mean ± standard deviation and moving average effectively identified over 500 outliers. However, these involved a substantial amount of normal data (2000–4000). In contrast, methods such as time series decomposition and box plot classified approximately 250 outliers, thus capturing only half of the labeled outliers. However, these showed fewer misclassifications with a significantly low error rate. Consequently, the number of detected outliers would decrease with stringent criteria, whereas the accuracy of classification of normal data would increase proportionally [32]. This situation underlines the challenge of accurately labeling outliers owing to the high variance in cattle data, the presence of measurement errors in the environment, and data collection by non expert farmers [35].

To overcome low-reliability outlier data owing to the high variability and characteristics of livestock environments, it is necessary to determine optimal data analysis algorithms in conjunction with high-precision data collection. In this study, a voting-based approach was adopted to enhance the outlier detection performance by leveraging the strengths of each model. It is noteworthy that votes 1 and 4 did not enhance the detection performance compared with the individual algorithm. The optimal outlier detection performance was observed when at least two models voted. Although this approach tended to show a lower accuracy, the F1-score and TSS (which consider both outliers and normalities) were relatively higher than those of the other methods. Therefore, we consider that a consensus from at least two models is more reliable for monitoring cattle intake data than the use of an individual model. It has the advantage of correcting the erroneous classification of outliers by an individual model. This advantage is likely to expand the identification of health issues based on the intake levels, thereby enabling their recognition and assessment based on the number of votes. Furthermore, it presents an opportunity to establish an additional database for outlier detection, provides precise values to livestock producers, and aids in accurate decision making.

5. Conclusion

Harsh environments in ranches and non-expert data recordings decrease the reliability of livestock data collection. This limits the performance of the data analysis. Therefore, the livestock industry is transforming into smart livestock management using sensor-based measurements and database technology. As a feasible approach for smart livestock, we developed multiple algorithms that can be utilized for monitoring collected data and outlier detection. Depending on the algorithm, the performance metrics varied. This indicated that a suitable selection of the optimal algorithm is necessary in conjunction with the data characteristics and the target to be monitored. Using a voting process with multiple algorithms is a potential solution for overcoming performance variations and achieving a high performance. Although there are further requirements such as methods for enhancing model performance and selecting optimal combinations of multiple algorithms, this study would contribute to improved decision-making in livestock farming and provide valuable insights for producers to ensure animal health and farm profitability.

Funding

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) and Korea Smart Farm R&D Foundation (KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) and Ministry of Science and ICT (MSIT), Rural Development Administration (RDA) (Project

No.421022–04).

CRediT authorship contribution statement

Jae-Min Jung: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Dong-Hyeon Kim:** Formal analysis, Software, Validation. **Hyunjin Cho:** Investigation, Resources, Writing – review & editing. **Mingyung Lee:** Investigation, Resources, Writing – review & editing. **Jinhui Jeong:** Formal analysis, Software, Validation. **Dae-Hyun Lee:** Formal analysis, Software, Writing – review & editing. **Seongwon Seo:** Funding acquisition, Investigation, Project administration, Resources, Supervision. **Wang-Hee Lee:** Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M.M. Culbertson, S.E. Speidel, R.K. Peel, R.R. Cockrum, M.G. Thomas, R.M. Enns, Optimum measurement period for evaluating feed intake traits in beef cattle, J. Anim. Sci. 93 (5) (2015) 2482–2487, https://doi.org/10.2527/jas.2014-8364.
[2] P. Martin, V. Ducrocq, D.G.M. Gordo, N.C. Friggens, A new method to estimate residual feed intake in dairy cattle using time series data, Animal 15 (2) (2021) 100101, https://doi.org/10.1016/j.animal.2020.100101.
[3] M. Lee, S. Seo, Wearable wireless biosensor technology for monitoring cattle: a review, Animals 11 (10) (2021) 2779, https://doi.org/10.3390/ani11102779.
[4] I. Andonovic, C. Michie, P. Cousin, A. Janati, C. Pham, M. Diop, Precision livestock farming technologies, in: Proceedings of the 2018 Global Internet of Things Summit (GIoTS), 2018, pp. 1–6, https://doi.org/10.1016/j.animal.2021.100429. Bilbao, Spain.
[5] Y. Qiao, H. Kong, C. Clark, S. Lomax, D. Su, S. Eiffert, S. Sukkarieh, Intelligent perception for cattle monitoring: a review for cattle identification, body condition score evaluation, and weight estimation, Comput. Electron. Agric. 185 (2021) 106143, https://doi.org/10.1016/j.compag.2021.106143.
[6] S. Basu, M. Meckesheimer, Automatic outlier detection for time series: an application to sensor data, Knowl. Inf. Syst. 11 (2007) 137–154, https://doi.org/10.1007/s10115-006-0026-6.
[7] E.A. Laca, Precision livestock production: tools and concepts, Rev. Bras. Zootec. 38 (2009) 123–132, https://doi.org/10.1590/S1516-35982009001300014, 2009.
[8] M. Taneja, J. Byabazaire, N. Jalodia, A. Davy, C. Olariu, P. Malone, Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle, Comput. Electron. Agric. 171 (2020) 105286, https://doi.org/10.1016/j.compag.2020.105286.
[9] G.K. Vishwakarma, C. Paul, A.M. Elsawah, An algorithm for outlier detection in a time series model using backpropagation neural network, J. King Saud Univ. Sci. 32 (8) (2020) 3328–3336, https://doi.org/10.1016/j.jksus.2020.09.018.
[10] G.K. Vishwakarma, C. Paul, A.M. Elsawah, A hybrid feedforward neural network algorithm for detecting outliers in non-stationary multivariate time series, Expert Syst. Appl. 184 (2021) 115545, https://doi.org/10.1016/j.eswa.2021.115545.
[11] Y. Chung, J. Lee, S. Oh, D. Park, H.H. Chang, S. Kim, Automatic detection of cow's oestrus in audio surveillance system, Asian-Australas. J. Anim. Sci. 26 (7) (2013) 1030, https://doi.org/10.5713/ajas.2012.12628.
[12] V. Bloch, H. Levit, I. Halachmi, Design a system for measuring individual cow feed intake in commercial dairies, Animal 15 (7) (2021) 100277, https://doi.org/10.1016/j.animal.2021.100277.
[13] J. Xiong, Q. Wang, J. Wan, B. Ye, W. Xu, J. Liu, Detection of outliers in sensor data based on adaptive moving average fitting, Sens. Lett. 11 (5) (2013) 877–882, https://doi.org/10.1166/sl.2013.2657.
[14] R. Gauthier, C. Largouët, L. Rozé, J.Y. Dourmad, Online forecasting of daily feed intake in lactating sows supported by offline time-series clustering, for precision livestock farming, Comput. Electron. Agric. For. 188 (2021) 106329, https://doi.org/10.1016/j.compag.2021.106329.
[15] M. Zhang, X. Li, L. Wang, An adaptive outlier detection and processing approach towards time series sensor data, IEEE Access 7 (2019) 175192–175212, https://doi.org/10.1109/ACCESS.2019.2957602.
[16] H.D. Nguyen, K.P. Tran, S. Thomassey, M. Hamad, Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management, Int. J. Inf. Manag. 57 (2021) 102282, https://doi.org/10.1016/j.ijinfomgt.2020.102282.
[17] R.M. De Mol, G. André, E.J.B. Bleumer, J.T.N. Van der Werf, Y. De Haas, C.G. Van Reenen, Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows, J. Dairy Sci. 96 (6) (2013) 3703–3712, https://doi.org/10.3168/jds.2012-6305.
[18] Z.H. Ismail, A.K.N.K. Chun, M.I.S. Razak, Efficient herd–outlier detection in livestock monitoring system based on density–based spatial clustering, IEEE Access 7 (2019) 175062–175070, https://doi.org/10.1109/ACCESS.2019.2952912.
[19] J. Navarro, I.M. de Diego, P.C. Pérez, F. Ortega, Outlier detection in animal multivariate trajectories, Comput. Electron. Agric. 190 (2021) 106401, https://doi.org/10.1016/j.compag.2021.106401, 2021.
[20] A. Blázquez-García, A. Conde, U. Mori, J.A. Lozano, A review on outlier/anomaly detection in time series data, ACM Comput. Surv. 54 (3) (2021) 1–33, https://doi.org/10.1145/3444690.
[21] T. Kieu, B. Yang, C.S. Jensen, Outlier detection for multidimensional time series using deep neural networks, in: 19th IEEE International Conference on Mobile Data Management (MDM), 2018, pp. 125–134, https://doi.org/10.1109/MDM.2018.00029. Aalborg, Denmark, 2018.
[22] S. Walfish, A review of statistical outlier methods, Pharmaceut. Technol. 30 (11) (2006) 82. https://www.statisticaloutsourcingservices.com/Outlier2.pdf.
[23] H. Vinutha, B. Poornima, B. Sagar, Detection of outliers using interquartile range technique from intrusion dataset, in: Information and Decision Sciences: Proceedings of the 6th International Conference on FICTA, Springer, 2018, pp. 511–518, https://doi.org/10.1007/978-981-10-7563-6_53.
[24] R.B. Cleveland, W.S. Cleveland, J.E. McRae, I. Terpenning, STL: a seasonal-trend decomposition, J. Off. Stat 6 (1) (1990) 3–73. https://www.nniiem.ru/file/news/2016/stl-statistical-model.pdf.
[25] M. West, Time series decomposition, Biometrika 84 (2) (1997) 489–494, https://doi.org/10.1093/biomet/84.2.489.
[26] M. Theodosiou, Forecasting monthly and quarterly time series using STL decomposition, Int. J. Forecast. 27 (4) (2011) 1178–1195, https://doi.org/10.1016/j.ijforecast.2010.11.002.
[27] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, OTexts, 2018. https://otexts.com/fpp2/.
[28] M. Said Elsayed, N.A. Le-Khac, S. Dev, A.D. Jurcut, Network anomaly detection using LSTM based autoencoder, in: Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks, 2020, pp. 37–45, https://doi.org/10.1145/3416013.3426457.
[29] O. Allouche, A. Tsoar, R. Kadmon, Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS), J. Appl. Ecol. 43 (6) (2006) 1223–1232, https://doi.org/10.1111/j.1365-2664.2006.01214.x, 2006.
[30] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, F. score and Roc: a family of discriminant measures for performance evaluation, in: A. Sattar, B. H. Kang (Eds.), AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science, vol. 4304, 2006, pp. 1015–1021, https://doi.org/10.1007/11941439_114.
[31] V.M. Patro, M.R. Patra, Augmenting weighted average with confusion matrix to enhance classification accuracy, Trans. mach. learn. artif. intell. 2 (4) (2014) 77–91, https://doi.org/10.14738/tmlai.24.328.
[32] H. Alimohammadi, S.N. Chen, Performance evaluation of outlier detection techniques in production timeseries: a systematic review and meta-analysis, Expert Syst. Appl. 191 (2022) 116371, https://doi.org/10.1016/j.eswa.2021.116371.
[33] R.K. Dwivedi, A.K. Rai, R. Kumar, A study on machine learning based anomaly detection approaches in wireless sensor network, in: 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Amity University Noida, 2020, pp. 194–199, https://doi.org/10.1109/Confluence47617.2020.9058311.
[34] B. Zong, Q. Song, M.R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, Deep autoencoding Gaussian mixture model for unsupervised anomaly detection, in: The 6th International Conference on Learning Representations, 2018. https://www.researchgate.net/profile/Song-Qi-6/publication/328437435_TGNet_Learning_to_Rank_Nodes_in_Temporal_Graphs/links/5bd608004585150b2b8c7d62/TGNet-Learning-to-Rank-Nodes-in-Temporal-Graphs.pdf.
[35] A. Huuskonen, P. Huhtanen, E. Joki-Tokola, The development of a model to predict feed intake by growing cattle, Livest. Sci. 158 (1–3) (2013) 74–83, https://doi.org/10.1016/j.livsci.2013.10.005.