IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received October 2, 2018, accepted October 16, 2018, date of publication October 25, 2018,
date of current version November 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2877970

An Early Detection System for Proactive Management of Raw Milk Quality: An Australian Case Study

ATEFE ZAKERI, MORTEZA SABERI, OMAR KHADEER HUSSAIN, AND ELIZABETH CHANG
School of Business, UNSW Canberra, Canberra, ACT 2612, Australia

Corresponding author: Atefe Zakeri (a.zakeri@student.unsw.edu.au)

ABSTRACT Milk is a highly perishable product, whose quality degrades while moving downstream in an imperfect cold dairy supply chain. Existing literature adopts a reactive approach for evaluating and preventing milk with a high microbial index from moving further downstream in a dairy supply chain. In this paper, we argue that such an approach is not the best response if the intention is to maximize milk life in terms of quality. We propose a proactive approach that monitors the metrics of the temperature and the level that are the building blocks of microorganisms in milk. This information is then used to determine the status at which the storage tank should hold the milk in accordance with standards. This status is then compared with the tank's actual status, and if they are different from one another, it will prompt the farmers to take the required preventive actions to manage the quality of milk. The developed proactive management of raw milk quality approach is modeled by using a rule-based system and machine learning techniques with a high level of accuracy. To test the validity of our approach and demonstrate its applicability, we apply it to a milk farm in Queensland, Australia.

INDEX TERMS Smart farming, milk quality, early detection system, early warning system, machine learning.

I. INTRODUCTION

With the continuing application of big data and the Internet of Things (IoT) in different domains, "Smart Farms" are now a reality in the agri-sector. In the dairy sector, such farms are enhanced versions of traditional farms as they use real-time data to provide farmers with value added benefits. This has led farmers to gain valuable insights in a timely manner, leading them to efficiently exploit their available resources and increase both the quantity and quality of milk produced. Some other examples of such smart initiatives in dairy farms are:

- Automated monitoring of a milking process with the use of RFID tags to improve the yield and quality of milk [1].
- Using robotic milkers to harvest milk without human involvement in the milking process. This has not only improved the quality of milk but also bolstered cows' health [2].
- Determination of cows' health by monitoring their mobility when they are grazing, being milked and resting, by using IoT platforms and cyber-physical systems [2], [3] as well as Wireless Sensor Networks (WSN) [4].
- Real-time monitoring, early detection and prediction of disease in cattle using wireless sensor networks [5], nano-sensors [6] and biosensors [7], [8].
- Feed management of cattle by improving the nutrient profile of the feedstock, monitoring hormone levels for improved fertility and also by adding of nano-scale digestive-aids for further feeding efficiency [6].
- The use of calving alert and birth monitoring using temperature and pH sensors [9].
- Monitoring fields' condition such as air and water quality in a dairy farm using the Internet of Nano Things (IoNT) [6].
- Classifying cattle behaviour based on sensor data and machine learning algorithms to enhance genetic selection programs emphasizing more on the individual well-being and performance rather than a more traditional herd based approach [10].
- Developing a milk yield prediction and analysis tool to help dairy farmers accurately predict future milk yield [11].

Such efforts and initiatives on a dairy farm have, overall, helped the farmers to ensure that the produced milk is of high quality, disease free, and free from microorganisms. However, like any perishable product, it is crucial that the initial quality of the milk is precisely controlled and maintained prior to moving downstream in the dairy supply chain.

In the next section, we describe a dairy supply chain and the issues raised by the current reactive management of milk quality and then we explain why a proactive approach in managing the quality of milk is needed.

II. PROBLEM DEFINITION

Fig. 1, shows an overview of the dairy supply chain which consists of the following stakeholders: 1) a dairy producer (*farmer*), 2) logistics transport provider (*transporter*), 3) processing unit where the milk is pasteurized, homogenized, converted and packed in accordance with the customers package shape requirements (*processor*), 4) the *retailer*, and 5) the *end customer*.

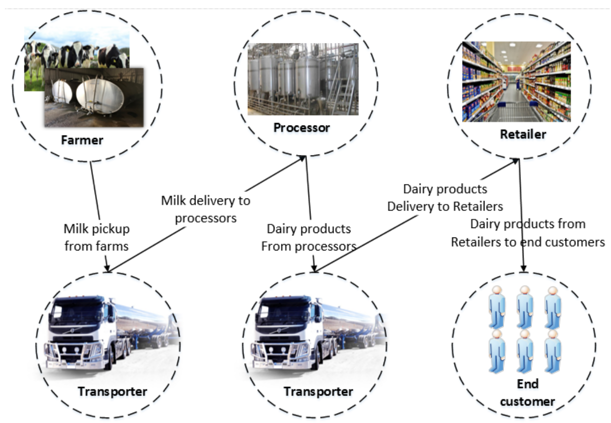

FIGURE 1. Dairy supply chain.

Every stakeholder in a dairy supply chain has to satisfy the quality requirements in order for the whole chain to achieve its highest value. Our focus in this paper is on the initial stakeholders of farmers and transporters.

In the dairy chain, different metrics are used in the literature to quantify and represent milk quality. For example, some approaches measure milk quality in terms of its nutritional value (such as fat and protein) [12] while some measure milk quality as the presence or absence of diseases in the milk [13]. A majority of works, [14]–[20] including this paper, quantify milk quality according to its *bacterial index*.

Irrespective of the quality metrics used, processors are increasingly demanding to receive milk of high quality from the farmers. This is due to rapid changes in the food supply patterns from small stores to large supermarkets, less frequent shopping cycles, and also export markets, dairy products with high quality standards and extended shelf life have recently become in demand [20]–[22]. Thus, for the processors to satisfy this need, they need high quality milk from the farmers.

The literature attempts to address this by using the notion of "cold chain" which is an uninterrupted temperature controlled transport and storage system for perishable goods [23]. While such a cold chain is beneficial, it reaches maximum efficiency only when the farmer manages the milk in such a way that its quality does not degrade beyond prescribed limits during the storage time (after it is milked and before it is collected by the transporter). The current literature does not consider this and, in most cases, the farmer manages the quality of milk in a reactive way after it is milked. We explain this further in the next sub-section.

A. EXISTING REACTIVE APPROACH TO MANAGE THE QUALITY OF RAW MILK

As the initial point of the dairy supply chain, farmers, having extracting fresh milk, use a temperature-controlled storage tank, to cool down the extracted raw milk. This is because cooling is the main means of slowing down the bacteria growth in the milk [24]. The temperature at which raw milk should be stored differs according to each continent's food safety legislations. For example, the European Union (EU) requires on-farm raw milk to be cooled down immediately (within 2 hours from the end of milking) so that its temperature stays below 8 °C, if the transporter collects the milk daily from farms. In the case of raw milk being collected on an alternate day basis, the temperature should not rise above 6 °C [25]. Food Standards Australia New Zealand (FSANZ)'s guidelines state that the milk should be cooled to 5 °C within 3.5 hours from the start of milking [25]. In the current structure of a dairy supply chain, transporters (milk tanker drivers) upon arriving at the farm, conduct a temperature test to check if that tank of milk is at the temperature indicated in the standard. In addition, the driver conducts senses tests before transferring the milk from the storage tank to the truck. The driver rejects the milk should it fail these tests [26], [27]. After the milk is picked up, it is taken to the processor who performs a comprehensive analysis of the quality of the milk in terms of its bacterial count to see if it is acceptable for used. If the milk consignment has a high bacterial index, the processor will reject it. While the existing approach compels farmers to maintain milk quality in terms of temperature, it does so in a reactive way which leads to the following challenges:

- While conducting temperature and senses tests by the transporter at the farmer's pickup point may prevent incompatible milk of low quality from moving further downstream in a dairy supply chain, these tests are very subjective and prone to failure. In real world conditions, this has led to scenarios where the milk passed the temperature and senses tests, but was subsequently rejected by the processor. This has led to substantial financial loss for both the farmer and the transporter. According to a survey [4], the majority of farmers (93.8%) had their milk rejected either once or twice per month by the processor.
- Relying on the testing of milk temperature at the pickup point does not guarantee that the milk was actually maintained at that temperature by the farmer after milking.

In other words, these tests do not guarantee that the milk is cooled down to the required temperature within the required time (3.5 hours after milking). It may be that the stored milk was brought to the accepted temperature range just before the scheduled pickup. Hence, relying solely on the tests at pickup does not guarantee acceptance by downstream stakeholders. To avoid this, a functionality is required by the transporter to check the whole cooling profile of the milk on a farm after it is milked. This is to ensure that the milk stored in the tank is not only cooled down according to the required temperature, but has not considerably fluctuated. Fluctuations in the milk temperature may occur for reasons such as electricity outage or the farmer neglecting to switch on the electricity while milk is entering to the tank. These scenarios may cause an increase in the milk's bacterial growth which cannot be identified at pickup point if the reactive approach of temperature and senses tests is taken directly before pickup.

- Undertaking the first comprehensive quality check at the processor is too late in the dairy supply chain to take any preventive actions to prolong the milk quality. In addition, deciding on the milk quality based on the microbial tests at the processor imposes a reactive approach to managing the quality of the milk since it is based on a symptom-driven approach rather than a root-cause-driven approach. For example, as shown in Fig. 2, bacterial growth in milk can have many different causes. In the current reactive milk quality management approach, the focus is more on the symptoms (bacterial index) which is a representation of what has already happened to the milk from different causes. However, if the intention is to prolong the milk life and slow down the decay in its quality, apart from just measuring milk's bacterial count as its quality parameter, there should be a mechanism to also measure, monitor and control parameters such as the temperature and the duration of the time milk is stored at that temperature resulting in that bacteria.

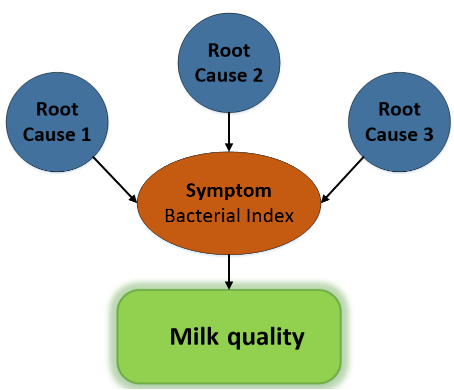

FIGURE 2. A hierarchical representation of factors impacting the milk quality.

B. NEED FOR A PROACTIVE APPROACH TO MANAGE THE QUALITY OF RAW MILK

We argue that a reactive approach is not the best possible response if the intention is to prolong the milk life according to acceptable standards. We need a quality management process that assists farmers measure and manage the quality of milk as soon as it is milked. Furthermore, rather than measuring the bacterial index which represents the quality of the milk as a composite measure, we need to monitor and manage the building blocks which lead to that bacterial index. To this end, we propose online monitoring of different key factors that impact bacterial quality; which are the temperature and the time the milk is stored at that temperature. This will present a comprehensive picture of the milk quality from the first point of entering the tank to the end point of leaving the tank on a farm.

By having such online monitoring, if there are any inconsistencies between the temperature of the milk in the tank and the standard temperature, then, preventive actions can be taken immediately by the farmer to rectify it. We term such an approach "proactive management of milk quality" which will be of great benefit for the farmer, transporter and processor due to their following unique requirements as follows:

- Farmers: The collection schedule of the transporter varies from being daily to alternate days depending on the optimal milk collection routes between farmer and processor [28], [29]. Between any two pickups, farmers store the extracted raw milk in tanks as shown in Fig. 3. As there is no segregation in a tank's milk, raw milk, which is milked at different times, is mixed and stored. This implies that the rate of change in bacterial quality of the tank's milk is not in accordance with the freshness of each batch of milk but according to the overall quality of the milk in the tank. If a reactive approach is taken to manage a tank's milk, then there is a strong possibility of either the whole quantity of milk being rejected or its life reduced [30], if actions to manage it are taken after the milk has deteriorated to a certain level. On the other hand, if a proactive approach to manage milk quality is used, it will then assist farmers to make smart predictive decisions towards maximizing the milk quality and achieving the maximum economic benefit from it.
- Transporter: As the collection schedule of milk varies and its quality is managed in a reactive way, it is possible that when the transporter delivers it to the processor, that may lead to milk rejection due to the high bacterial index which was not identified at the time of pickup. As shown in Fig. 3, in such scenarios, the rejected milk is sent back to the farmer leading to the transporter and farmer experiencing an economic loss [31], [32]. Being proactive in managing the milk quality will not only assist the farmer but also the transporter in avoiding such scenarios and ensuring that further bull whip effects such as milk shortages and loss in revenue downstream in the dairy supply chain are avoided.

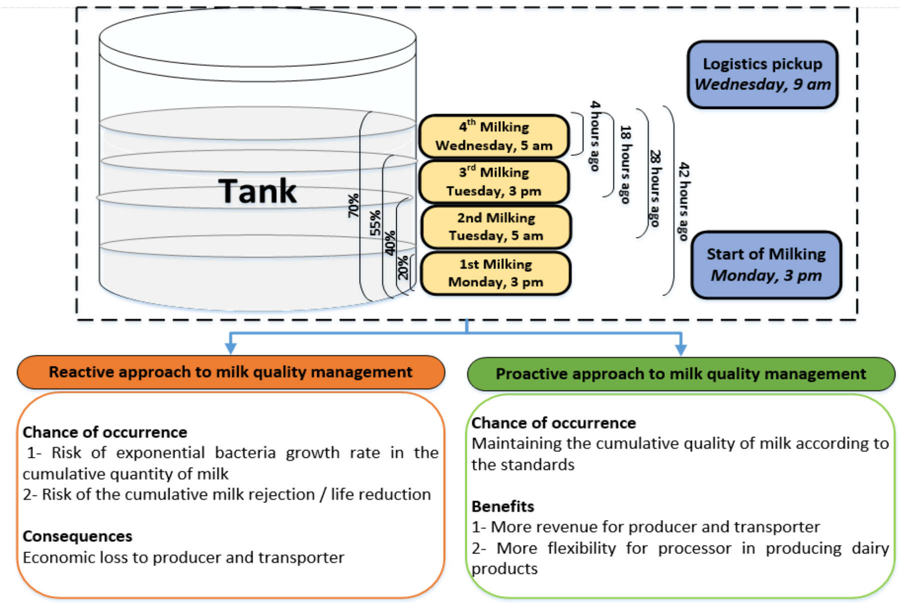

FIGURE 3. Reactive vs. proactive approaches in managing the raw milk quality in a tank.

- Processors: If the processor is provided with high-quality raw milk, it not only increases the processor's flexibility and efficiency in producing high-quality alternative dairy products, but also reduces the milk rejection in the processing plant. This is demonstrated by the fact that, processors are offering incentives to the farmer/transporter for providing higher-quality raw milk with microbial levels well below the regular health-based limits [20], [33]. This can be best achieved if a proactive approach to managing milk quality is adopted by the farmers that will benefit the processors.

We propose in this paper, that such benefits can be realized through a proactive approach for managing milk quality. To the best of our knowledge, monitoring the impacting factors on the quality of the milk when it is stored on a farm and provide farmers with an opportunity to take preventive actions if required has not been addressed in the literature.

The paper is organized as follows. Section 3 discusses the related work from the literature. In Section 4, a conceptual framework for proactive management of milk quality using the sensor data is proposed. Section 5 discusses the methodology in detail. For validating the proposed approach, in Section 6, real world data from a farm located in Queensland, Australia is used as our case study. Section 7 concludes the paper with a discussion on future work.

III. RELATED WORK

There is a vast amount of research in the literature that focuses on evaluating and monitoring the quality of raw milk at different levels of a dairy supply chain. Such approaches range from early sanitation practices prior to and during the milking process in a dairy farm, to monitoring the temperature and bacterial count in the last stages of the dairy chain where milk and other dairy products are sold by retailers to customers. To justify the need and importance of a proactive approach in managing milk quality and also to highlight the gap that such an approach addresses, we discuss the existing literature in two broad categories:

A. BEFORE THE STORAGE OF RAW MILK ON A FARM

This category includes papers that focus on improving milk quality by taking preventive actions and applying sanitation practices *before and during the milking process* until it is stored on a farm. These papers achieved this by implementing on-farm management practices [16], [18], various hygiene practices such as sanitation prior to milking [34] or Hazard Analysis and Critical Control Points (HACCP) [35]. Implementation of modern dairy practices, as mentioned earlier can also be categorized in this group. Some examples include automated monitoring of a milking process [1] or using robotic milkers instead of labor [2] to harvest the milk.

However, our focus in this paper is to maintain and manage milk quality as soon as it is extracted and stored in a tank on the farm. Therefore, such researches are not considered within the scope of our work since they are more focused on quality management practices *before* the milk is stored. The initial assumption in our research is that, due to the implementation of high hygiene standards on the farms, the extracted warm milk is totally fresh with an unavoidable minimal bacterial contamination. However, if not properly managed, then the milk starts to decay [24].

B. AFTER THE STORAGE OF RAW MILK ON A FARM

Some existing approaches from the literature represented in Table 1, monitor and evaluate the quality of the milk right after it is picked up by the transporter until it reaches the

TABLE 1. A classification of the literature according to the stages of a dairy chain.

Reference	Dairy Chain Stage							Measured Variable for monitoring	
	1	2	3	4	5	6	7	Temperature	Bacterial count
(Arati et al.) [15]			*				*		*
(Ruangwittayanusorn et al.) [17]		*			*				*
(Nada et al.) [19]			*	*					*
(Ndraha et al.) [23]						*		*	
(O'Connell et al.) [33]			*						*
(El Zubeir et al.) [36]	*								*
(Hill et al.) [37]	*								*
(Laguna et al.) [38]		*						*	
(De Las Morenas et al.) [39]		*						*	
(Kadam et al.) [40]			*					*	
(Van Schaik et al.) [41]			*						*
(Pantoja et al.) [42]			*						*
(Mhone et al.) [43]			*	*					*
(Boopathi et al.) [44]			*					*	
(Carullo et al.) [45]					*			*	
(Koutsoumanis et al.) [46]					*	*	*		*
(Morelli et al.) [47]						*		*	
(Nainggolan et al.) [48]							*		*
(Wang et al.) [49]							*	*	

1: At the pickup point 2: During transportation to processors 3: At the processors
4: After pasteurization 5: During transportation to retailers 6: At the retailer
7: End customer

retailers' store. As there are different stages in a dairy chain, for a better representation, we divide them as follows.

- Stage 1: At the pickup point
- Stage 2: During transportation to processors
- Stage 3: At the processors
- Stage 4: After pasteurization
- Stage 5: During transportation to retailers
- Stage 6: At the retailer
- Stage 7: End customer

The second column of Table 1 shows the particular stage/s which existing approaches address in managing milk quality. Column 3 in Table 1 shows the metric used by the approaches to measure and monitor milk quality (and dairy products). The following observations can be made from Table 1.

- First, as seen in Column 3 of Table 1, most of the papers in this category focus on determining the milk quality in terms of its bacterial count. While it gives a representation of its quality, as shown in Fig. 2, such an approach is symptom-driven rather than root-cause driven. Thus, we need to monitor and manage parameters such as the temperature and the time (duration after milking) [14], [50], [51] for proactive management of milk quality. It should be noted that while temperature as a parameter is measured by some existing approaches in Table 1, this measurement is done from the pickup point and not as soon as the milk is extracted and stored in the tank.
- Moreover, none of the papers in Table 1 indicate continual monitoring of the temperature of the milk while it is stored in a tank. As mentioned earlier, such a random measurement of temperature does not guarantee that the quality of the milk in the tank is perfectly maintained until it is picked up. Considering the cooling as the main means of slowing down the bacteria growth in the milk [24], if the cooling is not properly done, the bacteria will start to grow in the stored milk. While monitoring and preventing the milk (or other dairy products) with a high bacterial index moving downstream at each stage of the supply chain is of great importance, there is also a need for meticulously monitoring the whole cooling performance of the milk in the farmer's tank and taking proactive actions to ensure that it is cooled down according to the required standard.

These observations will lead to the reactive management of milk quality. In order to overcome these challenges, the contribution of this paper is to offer a proactive approach for managing the quality of raw milk stored in the tank. By adopting a data-driven approach and using online monitoring to analyze the impacts of key factors such as storage temperature and time on bacterial quality of tank milk this is achieved.

FIGURE 4. **Tank installation kit: sensors and controller.**

This guarantees that milk of high quality reaches the next steps in the supply chain and consequently provides the processor with a maximum milk life not only allowing more flexibility for handling raw milk and increasing efficiencies but reduces the risk of raw milk reaching bacterial levels of concern [20]. In the next section, we propose an approach for proactive management of milk quality.

IV. CONCEPTUAL FRAMEWORK FOR PROACTIVE MANAGEMENT OF MILK QUALITY: TOWARDS AN EARLY WARNING SYSTEM

The use of IoT devices in high risk industries such as the food supply chain has recently attracted much attention [52]. To monitor the cooling of milk which is stored in a tank, tanks are equipped with IoT sensors as shown in Fig. 4, which can be accessed by the farmer as well as the transporter.

This enables both the farmer and the transporter to detect instances where the milk temperature is exceeding the relative standards (in this case the cooling curve) that could negatively impact milk quality and potentially lead towards the risk of milk rejection. To avoid this, preventive actions need to be taken by the farmer within a certain time frame after the detection of such an event. This can be done by developing an "Early-warning system" that is capable of notifying the farmer when the temperature values increase beyond the acceptable limits based on the milk cooling curve along with the recommended action to be taken. Fig. 5 shows the milk temperature and level along with the standard cooling curve.

Milking is defined as a process in which fresh milk from cows is extracted and stored in the tank. As mentioned in Section 2 and shown in Fig. 3, a farmer can have more than one milking per day. As shown in Fig. 5, there is an increase in the milk's level at the time of each milking. Thus two milkings can be identified in Fig. 5. According to the milking pattern, the cooling curve is determined showing the temperature that the milk in the tank should not exceed. After the transporter picks up the milk, the level reaches zero. Recorded temperature in the tank by the sensors show occasions in which the registered temperature exceeds the cooling curve that requires action to be taken by the farmer as soon as possible to avoid the degradation of the milk quality. This is where the "Early-warning system" comes into play.

As shown in Fig. 6, the pre-requisites for developing an early warning system requires capturing data and knowledge from two sources:

- Data in the form of critical information related to milk quality captured from the installed sensors in the tank.
- Knowledge captured from the domain expert relating to the milk's status in the tank and converted to a form required for modeling the system.

Once the system is modeled, it can be used for proactive management of the milk quality. This first requires the early detection of events that may negatively impact milk quality

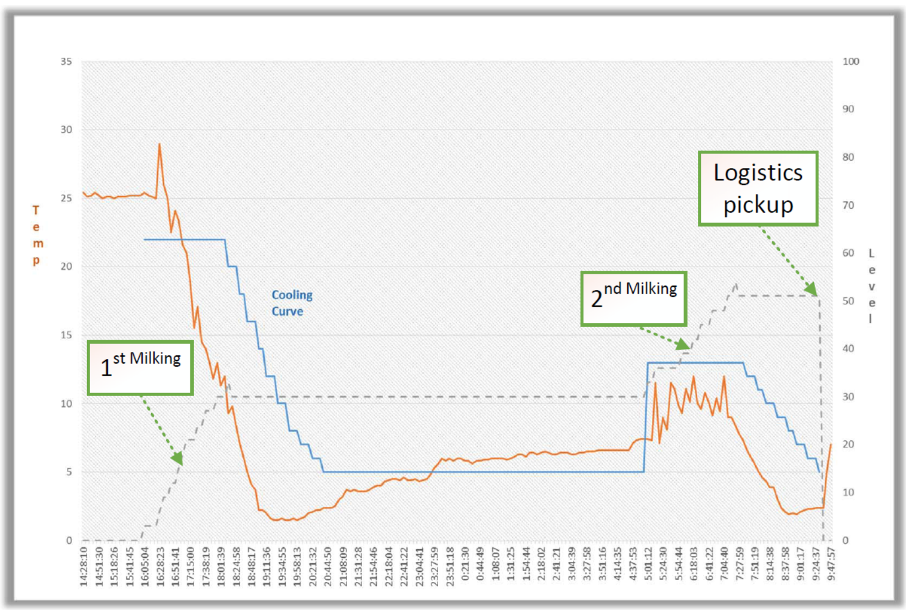

FIGURE 5. **Monitoring the milk temperature based on the milk cooling curve.**

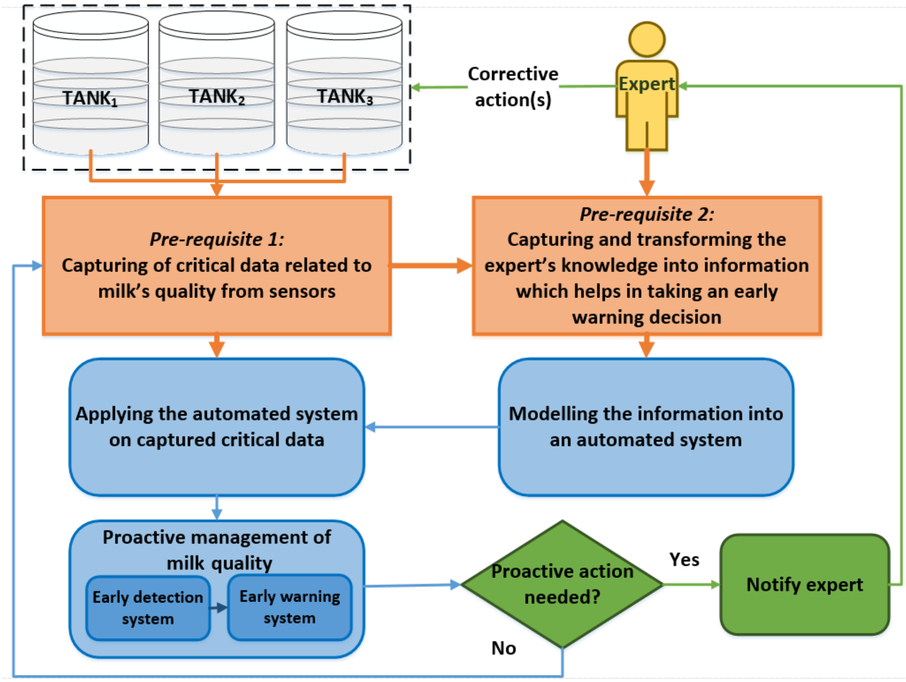

FIGURE 6. Conceptual framework for an early warning system.

followed by an early warning system, which recommends preventive actions to be taken.

In other words, an "Early detection system" allows the user to monitor and detect the milk temperature exceeding the standards of the cooling curve which increases the growth rate of bacteria beyond acceptable limits on the detection of such events. The "Early warning system" generates alarms for the farmer of risky situations and recommends proactive actions to be taken.

In this paper, the focus is towards developing an early detection system that is a pre-requisite for the early-warning system. To this end, the objectives are to first capture and transform the experts' knowledge and model it into an automated system. Thus, in the next section, the proposed steps for developing such an early detection system are explained.

V. METHODOLOGY: DESIGNING AN EARLY DETECTION APPROACH FOR MANAGING THE QUALITY OF RAW MILK USING DATA MINING TECHNIQUES

The steps for the proposed early detection approach for managing the quality of raw milk are explained and shown as Fig. 7. Our proposed approach is composed of two phases:

Phase I, involves having an interactive relationship between a domain expert who understands the nature of the problem (e.g. a farmer) with a user who is going to model the domain expert's mind for further analysis. During Phase I, the user is trained in the basic knowledge required to determine what the current state of the tank should be in terms of the milking status, which dictates if early detection needs to be activated. In other words, the objective is to understand the relationships between the input parameters and the output variable and model it in the form of a rule-based system. Thus, this phase is called "user training and knowledge modeling".

The output of Phase I is used as a training set for a machine learning algorithm in Phase I where the objective is to automatically process the extracted information and classify it into associated classes and deal with instances where the input data is missing, which leads the rule-based approach to not work in such scenarios.

A. PHASE I: USER TRAINING AND KNOWLEDGE MODELING

The objective of this phase is to model the knowledge in an expert's mind who has a general, clear understanding of the status to happen in the tank throughout a whole milking cycle. Modeling such information involves forming an interactive relationship between the expert and the user. This enables the user to capture the required assumptions to model the problem. As shown in Fig. 7, this phase is composed of three stages described as follows.

1) STAGE 1. GETTING TO KNOW THE SYSTEM UNDER STUDY

The focus of the user at this stage is to become familiar with the milking cycle and the resulting events. A "milking cycle" is defined as a sequence of multiple recordable "events" starting from an empty, clean tank ready for the first milking to the time the tank is next cleaned. Between these two instances of tank cleaning, there are multiple other represented events in a milking cycle. These events with their semantics description need to be defined by the experts. Semantics defined should be clear and able to distinguish when an event ends and the next starts.

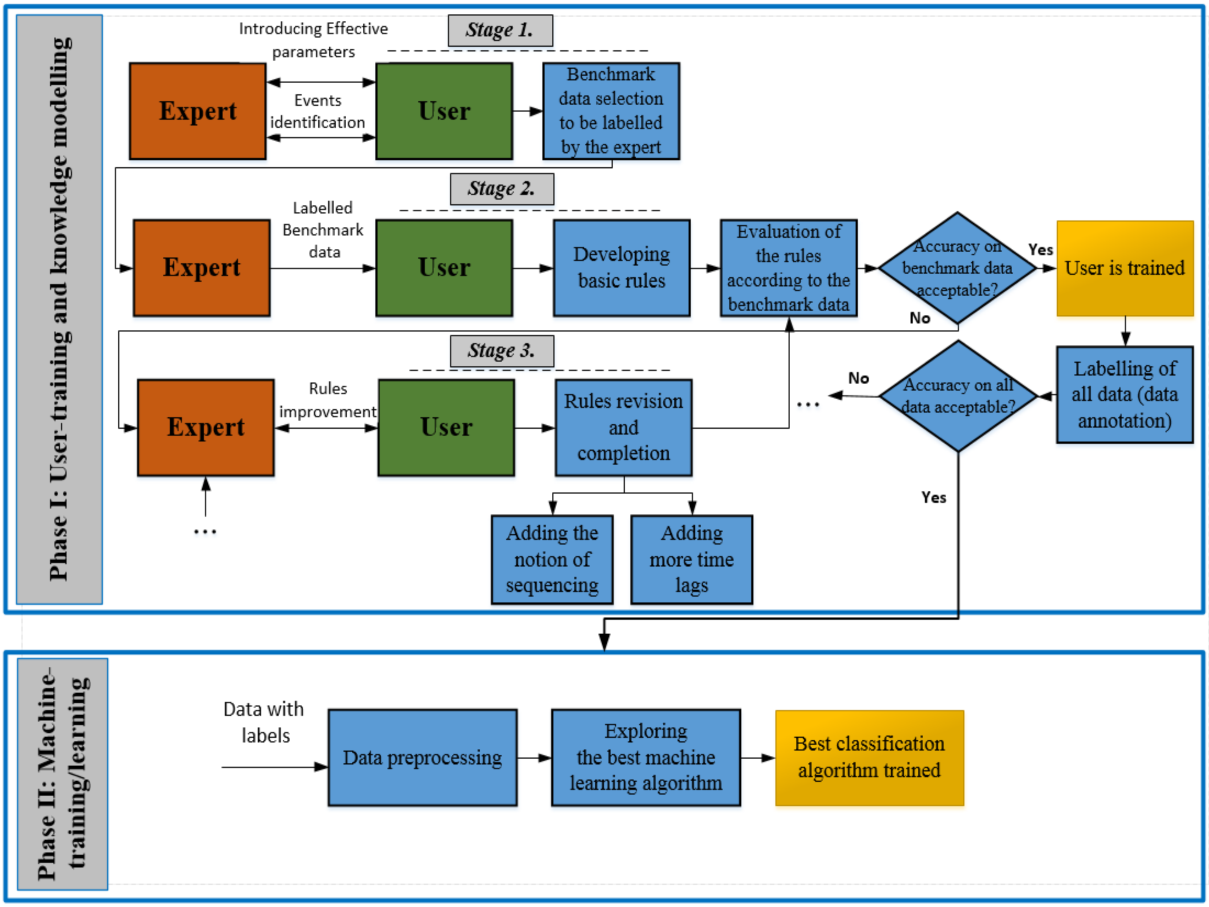

FIGURE 7. A schematic representation of the proposed methodology for early detection of events impacting raw milk quality.

Fig. 8 and Table 2 represent a set of events, and associated semantics for each event along with their assigned labels, respectively in a milking cycle.

2) STAGE 2. DEVELOPING AND EVALUATING THE RULES

Once the different events in a milking cycle are defined, the expert is then asked to label a selected sample of critical data captured according to the events they satisfy as shown in Table 3. This set of data is formed as a benchmark data. The last two columns of Table 3 represent the corresponding events and labels determined by an expert to the benchmark data.

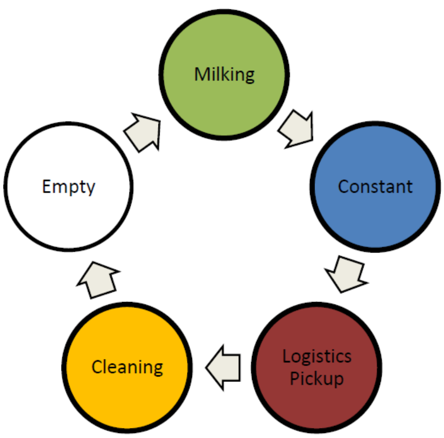

FIGURE 8. An example of events in a milking cycle.

To explain the labeled data in Table 3, this milking cycle starts when the tank is "Empty" followed by a "Milking" event in which the level rises to 16% of the tank capacity. Then, there is no increase in the level called "Constant" until the second "Milking" happens in which the level increases to 23%. Again, the level is "Constant" until the transporter picks up the milk from the tank which is determined as "Logistics Pickup". The last event before "Empty" is "Cleaning" of the tank with hot water in which the temperature rises for cleaning purposes.

Having such benchmark data with their corresponding labels enables the user to model the expert's knowledge in the form of If-Then rules. An example of these rules is as follows:

(1) If *Level* is "increasing", Then *label* is "M".
(2) If *Level* is "fixed" and *Temp* is "fixed", Then *label* is "C".

As can be seen from the above rules, the meaning of statements such as "increasing", "fixed" or "sudden drops"

TABLE 2. Semantics for identified events in a given milking cycle.

Event	Description/ Semantics	Label
Constant	The status when the level of milk in the tank is constant (non-zero), thus, the temperature should be controlled based on the standards.	C
Milking	The status when the milking process is in progress resulting in an increase of the milk's level during a certain number of data points.	M
Pickup	Starts from the first time when the level of milk starts to decrease significantly until it remains constant for some data points.	P
Cleaning	Those points in time when there is no milk in the tank and the temperature starts to have a significant increase until the first point when the temperature gradually starts to decrease.	Cl
Empty	A cleaned tank with no milk in it ready to be filled with the fresh milk.	E

TABLE 3. An example of a benchmark data from the sensors.

Row	Date	Time	Temp (°C)	Level (%)	Event (Label)
1	1/07/2017	15:00:34	15.3	0	Empty (E)
2	1/07/2017	15:06:35	15.2	0	Empty
3	1/07/2017	15:12:35	19.6	3	Milking (M)
4	1/07/2017	15:18:35	15.4	6	Milking
5	1/07/2017	15:24:36	14.2	6	Milking (M)
6	1/07/2017	15:30:36	16.3	8	Milking
7	1/07/2017	15:36:36	13.2	9	Milking (M)
8	1/07/2017	15:42:37	12.9	12	Milking
9	1/07/2017	15:48:37	10.2	15	Milking (M)
10	1/07/2017	15:54:38	8.3	15	Milking
11	1/07/2017	16:00:38	6.3	16	Milking (M)
12	1/07/2017	16:06:39	4.2	16	Constant
13	1/07/2017	16:12:39	3.9	15	Constant (C)
14	1/07/2017	16:18:39	3.6	15	Constant
15	1/07/2017	16:24:39	3.7	15	Constant (C)
16	1/07/2017	16:30:40	3.6	14	Constant
17	1/07/2017	16:36:40	9.9	19	Milking (M)
18	1/07/2017	16:42:40	8.1	22	Milking
19	1/07/2017	16:48:41	6.5	23	Milking (M)
20	1/07/2017	16:54:41	3.8	23	Constant
21	1/07/2017	17:00:41	3.7	22	Constant (C)
22	1/07/2017	17:06:42	3.5	22	Constant
23	1/07/2017	17:12:42	3.5	0	Logistics Pickup (P)
24	1/07/2017	17:18:42	16.6	0	Cleaning
25	1/07/2017	17:24:42	55	0	Cleaning (Cl)
26	1/07/2017	17:30:43	54.5	0	Cleaning
27	1/07/2017	17:36:43	61.4	0	Cleaning (Cl)
28	1/07/2017	17:42:43	46.6	0	Empty
29	1/07/2017	17:48:44	44.6	0	Empty (E)

cannot be determined by just looking at an individual data point, but by considering a sequence of data points. As a result, the above rules are revised to consider such a sequence of data points:

(1) If $Level_{t-2} > Level_{t-1} > Level_t$ Then *label* is "M".
(2) If $Level_{t-2} = Level_{t-1} = Level_t$ & $Temp_{t-2} = Temp_{t-1} = Temp_t$, Then *label* is "C".

After forming the If-Then rules, they need to be evaluated to determine accuracy in the labeling of input data. For this purpose, the rules are hard-coded and tested on the benchmark data already labeled by the expert. If the labeling accuracy on benchmark data was greater than a given threshold (λ), it indicates that the rules are well designed for the conversion of the knowledge in the expert's mind to a structured format of If-Then rules. It also implies that, if required, the user is now well trained and qualified to annotate a partial dataset.

After the user completed the labeling, the accuracy of the rule-based model on the whole data set under study also needs to satisfy a given threshold (γ), as shown in Fig. 7.

However, in case the accuracy of the whole data set is below the threshold (γ), the user had to re-label the data and remake the rules and subsequently, the accuracy of the rules on the whole data needs to be tested again to determine if they satisfy the given threshold. If it does not, then the domain expert needs to be consulted to retrain the user.

It should be noted that while labeling a big set of temporal data is a labor-expensive, time-consuming task, it is required at this step to have a complete set of labeled data which is going to be used for both the evaluation of the developed rule-based model on the whole data set as well as for training machine learning algorithms as explained in Phase I.

3) STAGE 3: IMPROVING THE RULES

In case the accuracy of the rule-based model in Stage 2 was not satisfying, it is required that a meeting with the expert be held to discuss mis-labeled data instances. The reason being that there might be some rules or even some new events that have not been considered in the model. To capture that during the retraining process, the strategy of "sequencing" can be used to improve the accuracy. The approach here is to first consider another strategy that suggests "using more time lags" in developing the rules before considering sequencing. These two strategies are discussed below.

4) STRATEGY 1: IMPROVING THE RULES USING MORE TIME LAGS

Since data recorded by the sensors are time-based and sequential, a combined sequence of data points is required to make a firm decision about the current event. The number of sequential data points considered in developing each rule may vary depending on how the values of input features (level and temperature) are changing during each event. However, while more data points make the rules more complex, they can usually improve the rules and help in getting more accurate results.

Therefore, to improve the accuracy of rules, we make use of more data points (time lags) in developing the rules. This can especially improve the determination of labels in those periods when the changes in input features are not logical. For instance, when a milking event is in progress, there might be an occasion where the level of milk stops increasing for a few points due to a break in the milking process. To capture such events and in order to avoid confusion between a Milking event with a Constant one, more data points should be considered in developing the associated rules for the Milking.

5) STRATEGY 2: IMPROVING THE RULES USING THE NOTION OF SEQUENCING

As shown in Table 3, the relationship between data recorded by the temperature and level sensors and the corresponding label are not logical, leading to an inconsistency between what measurements are shown and what the assigned label is. In other words, there can be noise in the recorded data. For instance, from Table 3, while the percentage of milk level in the tank is unchanged from reading # 4 to 5 which is (6%), the label is still "M". The same thing happens for readings #9 and 10. This is determined by seeing an increased level in the next few readings that indicate the milking is still in progress. Such difference in the time lags should be considered while developing the milking rules. Inversely, the same thing can be observed for readings #15-16 and #20-21. While the labels indicate the milk level is Constant, there are fluctuations in the level of the milk. Therefore, to capture such scenarios, rules should be revised in such a way that ignores these fluctuations due to noise. Consequently, two types of changes can be observed in the data:

- Changes in inputs that indicate a transition from one label to another. For example, to move from "Constant" to "Milking", (1) can be revised as the following for which the variation in the level, between the current point and either one, two or three previous lags is considered to be at least 3 units.
  (1) If $Level_t - Level_{t-1} \geq 3$ & $Level_t - Level_{t-2} \geq 3$ & $Level_t - Level_{t-3} \geq 3$, Then *label* is "M".
- Changes in inputs that do not lead to a transition from one label to another. Given the noise happening in data coming from the IoT sensors and devices, the rules should be modified to consider the allowed range of changes for the values of features under which the label should not change. Since these changes do not indicate a transition from one label to another, they are called "allowed range of changes". Considering these allowed range of changes in developing the rules, (2) can be revised as follows:
  (2) If$|Level_t - Level_{t-1}| \leq 1$ & $|Level_{t-1} - Level_{t-2}| \leq 1$& $|Level_t - Level_{t-2}| \leq 1$, Then *label* is "C".

The user should determine an allowable range of changes by having a look at the data in consultation with the expert.

6) SEQUENCING MATRIX

To revise the rules by taking time changes and the allowed range of changes, a notion of sequencing between events comes into play and a binary $n * n$ sequencing matrix can be accordingly arranged. Such a matrix is used to represent a binary relationship between two events in a given milking cycle with the total of n events. The entries of S are defined as follows:

$$S_{ij} = \begin{cases} 0 & \text{if the move from i to j is permitted,} \\ 1 & \text{if the move from i to j is not permitted.} \end{cases}$$

For example, sequencing matrix for the data in Table 3 is as follows:

$$S_{5\times5} = \begin{array}{c} \\ M \\ C \\ P \\ Cl \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} M & C & P & Cl & E \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

According to matrix S, there are ten elements for which $a_{ij} = 1$ indicates the permitted movements. For example, if the current event is "Milking", movements from "Milking" to either "Milking" or "Constant" are permitted.

Sequencing of events should consider all permitted movements from one event to another in a milking cycle. Thus, the rules should be extended and revised with respect to the sequencing matrix in such a way that every element of $a_{ij} = 1$ in matrix S is covered by associated rules.

After revising the rules, they are again evaluated based on the benchmark data as represented in Fig. 7 to see if the resultant accuracy is acceptable.

At the end of Phase I, a set of labeled data enables us to train a machine learning algorithm which is going to be used to complement our developed rule-based model.

B. PHASE I: DEVELOPING A MACHINE LEARNING-BASED CLASSIFIER FOR AUTOMATIC LABELING OF DATA

At the completion of Phase I, a historical unlabeled set of data from a tank can be automatically labeled to an event in a milking cycle with high accuracy. However, as the objective in proactive management of milk is to assign a label in a real-time mode, the rule-based system has some drawbacks when applied to an online real situation:

1) To determine the label of a current data point, some previous points (lags) are used to acquire a more accurate label. This results in some delay in making the current label available which can further result in delays in taking proactive action towards a risky occasion if needed.
2) In many applications, for data coming from the IoT sensors and devices, scenarios such as mis-recording and false measurements [53] are possible. This causes the rule-based system to be ineffective as it cannot deal with such scenarios.

In order to address these drawbacks, by using several machine learning techniques in Phase I, a classifier is developed that gives labels based on the input. Therefore, using our developed rule-based model, we automatically determine the labels for a set of data which is going to be used as a training set for finding the corresponding class (labels) for online unseen data (testing set) in the classification problem. This will help suppliers in monitoring the milk cooling performance on the farm at any point in time.

1) STEP 1: DATA PREPROCESSING AND FEATURE SELECTION

As a preprocessing step in machine learning problems, feature selection is very crucial in reducing dimensionality, removing irrelevant data and finally improving the classification accuracy [54]. Since If-Then rules are developed based on the features measured by the installed sensors, our selected features in this phase are the milk temperature and the level. Thus, the objective of this phase is to classify the data in one of the identified milking events determined in the previous steps.

Other preprocessing steps may include removing missing values for which the sensors fail to record measurements and removing out-of-range values such as negative measurements for the "level (%)" in order to avoid producing misleading results.

After the preprocessing step, data is divided into two sets; training and testing. Training data is used for model training while the testing data (without labels) is used as a measure of how well the classification algorithm performs on unseen data.

2) STEP 2: EXPLORING THE BEST MACHINE LEARNING ALGORITHM

In this step we compare the performance of typical machine learning algorithms on our milking data set. The objective is to find the best algorithm with the highest accuracy (lowest error rate) on testing data. To nullify the effect of different scales of input parameters used in the model, Mean Absolute Percentage Error (MAPE) is used to measure the accuracy. MAPE is measured by:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \widehat{Y}_i|}{Y_i} \tag{1}$$

Ensemble methods can also be considered and compared with single classifiers. This is due in some cases to ensemble methods such as bagging or boosting can perform better than single classifiers due to their committee structure which enables ensemble methods to utilize an agreement between several individual classifiers and vote on the classification of testing data [55].

In the next section, we discuss a real world case study to evaluate the practicality of our proposed early detection methodology.

VI. CASE STUDY: MONITORING THE STATUS OF ON-FARM MILK USING REAL WORLD DATA

The Queensland, Australia, farm under study currently has one active tank equipped with sensors as shown in Fig. 4. This device is a cloud-based system composed of a controller and a sensor. Every six minutes, a record is registered which contains the following input parameters:

1) Temp: measures the temperature of milk currently in the tank.
2) Level: measures the percentage of the tank that is filled with milk.
3) Since the data being registered is sequential, date and time are also recorded.

To monitor the status of milk stored in the tank and proactively manage it to comply with regulations, the following events as outputs need to be accurately recognized from the input stream of data being captured:

- The commencement and termination of the milking process (Milking)
- Logistics pickup
- Tank cleaning

For the particular farm under study, milking into the tank happens two times a day; morning and afternoon. Logistics

TABLE 4. Semantics for identified events in a milking cycle for the case under study.

Event	Description/ Semantics	Label
Constant	There is milk in the tank, so the temperature should be controlled (under °5C).	C
Milking	Starts from the first time when there is a jump in the temperature and subsequently a continuous increase in the level, until the level reaches to its local maximum (no more increase at the level after that).	M
Pickup	Starts from the first time when the level of milk starts to decrease (a significant decrease) until it reaches to zero.	P
Cleaning	Those points of time when there is no milk in the tank and the temperature starts to have a significant increase until the first point when the temperature gradually starts to decrease.	Cl
Empty	A cleaned tank with no milk in it which is ready to be filled with the new milk.	E

TABLE 5. Basic If-Then rules according to the labeled benchmark data.

#	level(t-1)	Temp(t-1)	level(t)	Temp(t)	Event(Label)
1	x	y	x	y	Constant (C)/Milking (M) (Temporarily stopped)
2	$x+$	y	x	y	Pickup (P)
3	$x-$	y	x	y	Milking (M)
4	x	$y-$	x	y	Constant (C)/ Milking (M) (Temporarily stopped)
5	x	$y+$	x	y	Constant (C)/ Milking (M) (Temporarily stopped)
6	$x-$	$y+$	x	y	Milking (M)
7	$x+$	$y+$	x	y	Not possible
8	$x-$	$y-$	x	y	Milking (M)
9	$x-$	$y-$	x	y	Milking (M)
10	0		x		Milking (M)
11	0	y	0	y	Empty (E)
12	0	$y-$	0	y	Cleaning (Cl)/ Empty (E)
13	0	$y+$	0	y	Empty (E)
14	x	y	0	y	Empty (E)
15	x	$y-$	0	y	Not possible
16	x	$y+$	0	y	Not possible

pickup happens either on a daily or alternative basis depending on the plan scheduled by the logistics company, which depends on the average milk production per farm. Therefore, the number of milkings before the logistics pickup is either two or four times. Thus, by monitoring the input parameters of the level and temperature, we aim to determine the output parameter which is the status of milk in the tank with high accuracy.

A. PHASE I: USER TRAINING AND KNOWLEDGE MODELING

1) STAGE 1. GETTING TO KNOW THE SYSTEM UNDER STUDY

In order to get to know the following domain knowledge, several meetings were held:

- General requirements for the on-farm milk that should be met by the farm,
- Knowing the processes and procedures happening throughout a milking cycle in order to identify the events,
- Working of sensors

During this project, both face-to-face and online meetings were held with one of the experts from the logistics provider. This created a friendly environment for the expert to share his knowledge and also for the researchers (users) to ask their questions and share their findings as the project proceeded.

According to the discussion with the expert, five events were identified throughout a milking cycle as shown in Table 4, with a description of their semantics.

216 benchmark data were selected out of almost 16000 data instances from the whole data set under study. Benchmark data to be labeled by the expert includes all the events defined in Table 4 and were selected from a whole milking cycle.

2) STAGE 2. DEVELOPING AND EVALUATING THE RULES

Benchmark data was sent to the expert to be labeled according to the defined labels in the previous stage. Due to space limitations, sample benchmark data with its associated labels from the farm under study can be found in the electronic version of the paper.

Having benchmark data with their corresponding label, the basic If-Then rules were developed which are shown in Table 5. For this study, accuracy thresholds i.e. (λ) and (γ) were selected as 90%.

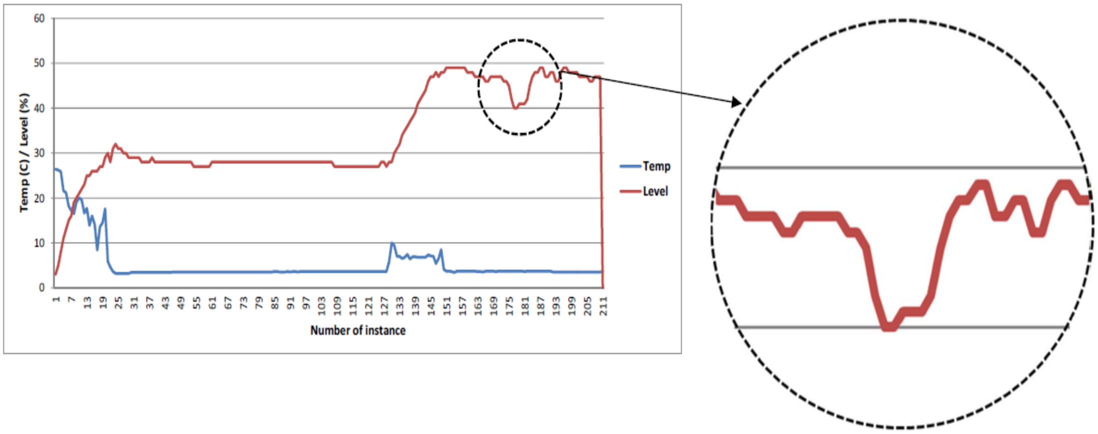

FIGURE 9. An example from a milking cycle to show how Strategy 1 can improve the rule-based accuracy.

In developing the rules at this stage, only one previous lag (reading happened at time $t-1$) was considered to determine the label at time (t):

In Table 5, x and y are variables used to respectively show the changes in the level and the temperature of milk. According to rule 2, if $Level_t < Level_{t-1}$ and $Temp_t = Temp_{t-1}$then the event is pickup. It can also be observed from the table that some of the rules, for example, rule #1 events/ labels are not determined conclusively indicating that the current information used is not sufficient to make a firm decision. It indicates that the knowledge obtained from the process is still not sufficient to make a final decision about the labels of some data instances. Also, as can be seen from Table 5, some scenarios, (for example, #7) are not expected to happen according to the processes happening in a milking cycle in this study.

In order to see how accurate the primary rule-based model is performing in labeling the benchmark data, the rules are implemented in *Matlab* software. The accuracy of the rules in Table 5 is 76.5% which is less than $\lambda = 90\%$. The accuracy is determined by the total percentage of the labels, which are correctly determined by the rules in comparison with the labels determined by the expert. However, the accuracy is not promising and indicates the weakness of the rules in translating expert knowledge. As also shown in Fig. 7, to improve the rules in order to achieve greater accuracy, we move to Stage 3.

3) STAGE 3: IMPROVING THE RULES

A meeting was conducted with the domain expert where the misclassified instances were discussed to see how the rule-based system could be revised to avoid these instances being incorrectly classified. It was observed that a new event needed to be defined to consider points in time when the logistics pickup is complete but cleaning has not yet started. This event is called 'Delay in cleaning' which is different from the 'Cleaning' event where in the 'Cleaning' the temperature has an increasing trend while during the 'Delay in cleaning' event, the temperature is almost unchanged (below 5 C). Therefore, another event was added to the previous defined events. Although at this stage, there is no milk whose quality may be affected, the overall accuracy of the system is being impacted due to the incorrect classification of the event. Hence, a new event was designed.

Also, improvement was made in terms of both adding extra time lags to the developed rules as well as involving the notion of sequencing between events as suggested in improving strategies in the methodology. These are discussed in greater detail.

4) STRATEGY 1: IMPROVING THE RULES USING MORE TIME LAGS

It is beneficial to use more time lags in developing the rules as suggested in Strategy 1 due to the following issues seen in the data:

- During a ''Milking'', sometimes there is a break that results in the level of milk to halt for a few data points. This description is similar to the semantics of a ''Constant'' event and the two might be confused.
- When moving from the event of ''Constant'' to ''Milking'', there is usually a time lag of one reading before the level increases which results in the delay of recognizing the ''Milking'' event. This is due to the technical design of the sensors that are based on a change in the pressure of liquid in the tank.
- When the milk status is ''Constant'', there are some occasions in which the level goes down significantly. Fig. 9 represents a sample from a milking cycle with two ''Milking'' events and a ''logistics pickup'' happening at the end of the curve. As can be seen after the second milking, the level suddenly decreases. Initially, it might be confused with the ''Logistics pickup'' but as the next data points come through, it turns out to be related to the cooling system since the milk goes out from the tank to a

refrigerated tank to be cooled down and then comes back again to the tank. This results in the volume of milk in the tank to be underestimated by the sensor.

Therefore, we decided to include a maximum of three time lags, whenever needed, in our rule-based model.

5) STRATEGY 2: IMPROVING THE RULES USING THE NOTION OF SEQUENCING

Using the notion of sequencing as suggested by the second improvement strategy is also helpful for improving the rules due to the following reasons:

As previously discussed, there is noise in the recorded data. For example, as shown in Table 3, when the milk is "constant", there are still some fluctuations in the level of milk. These "allowed range of changes" as discussed in the methodology, need to be differentiated with the changes occurring due to a transition from one event to another. This can be done using the notion of sequencing as discussed in improving Strategy 2 in the methodology. Thus, the sequencing flowchart can be represented as shown in Fig 10.

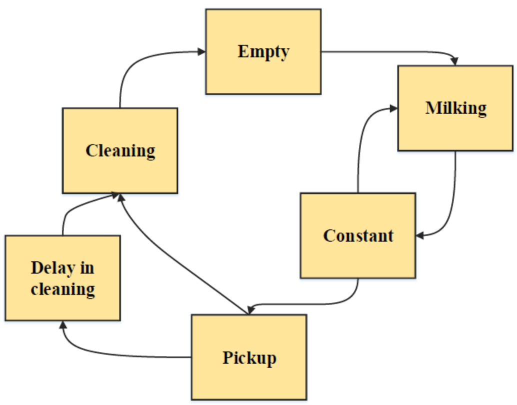

FIGURE 10. Sequencing flowchart.

Below is the sequencing matrix corresponding to the above sequencing flowchart. In this matrix, 1 indicates the sequencing that is allowed while 0 indicates the sequencing that is not permitted.

$$
S_{6\times 6} = \begin{array}{c} \\ M \\ C \\ P \\ D \\ Cl \\ E \end{array}
\begin{array}{c} \begin{array}{cccccc} M & C & P & D & Cl & E \end{array} \\
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \end{array}
$$

Consequently, 14 different transitions between events are possible as shown in Fig. 11. Thus, each of these transitions between events should be separately considered in developing the rules.

By revising the rules based on both strategies, 14 final rules resulted and are listed in Table 6.

The revised rule-based model was able to correctly determine the labels for all benchmark data (Accuracy = 100%). Thus, the model is ready to be used for determining the labels for the whole data set by the user as they are qualified to label the rest of the data manually.

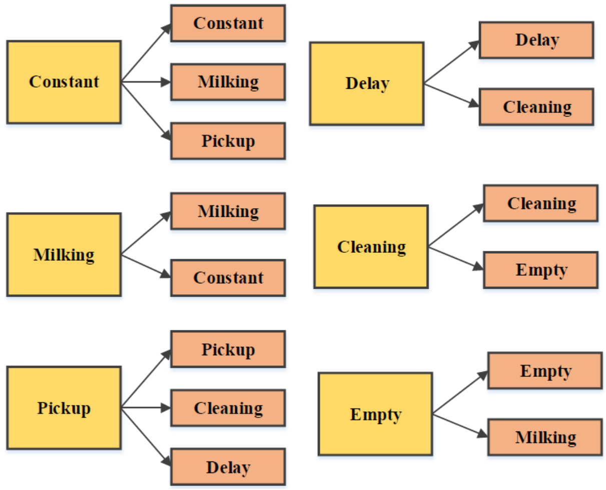

FIGURE 11. Permitted transitions between the events according to the sequencing matrix.

The accuracy of the rule-based model on the whole data set was 99%, which is very promising. This indicates that the model is capable of labeling a previously unlabeled set of data with high accuracy. The labeled data set is used in the next step.

The developed rule-based model would assist the farmer in the early detection of the events when both input variables are available. However, in scenarios where input is either missing or viewed as an outlier, the rule-based system may not give an accurate label. To overcome that in the next phase, we propose a machine learning based classifier by using different algorithms.

B. PHASE I: DEVELOPING A MACHINE LEARNING-BASED CLASSIFIER FOR AUTOMATICALLY LABELING DATA

We use the training data labeled by the rule-based model as a training set to find the best machine learning algorithm by which the online IoT data coming from the sensor could be classified accurately in one of the identified events in a milking cycle.

1) STEP 1: DATA PREPROCESSING AND FEATURE SELECTION

For some data points, when the tank is empty, the value for the "level" was recorded as a small negative value, which was changed to zero due to out-of-range values. 80% of all data was used for training and the remainder used for testing.

2) STEP 2: EXPLORING THE BEST MACHINE LEARNING ALGORITHM

For classification purposes, *Weka* machine learning workbench was used. Twenty-one typical machine learning algorithms from six different categories were trained based on the training set and evaluated on the testing set. Results are summarized in Table 7.

TABLE 6. Final rules for conversion of the expert knowledge.

#	Transition	Rule
1	**Constant to Constant**	If $(Level_{t-2} > 0)\&(Level_{t-1} > 0)\&(Level_t > 0)\&\vert(Level_{t-1} - Level_t\vert \leq 1\&\vert(Level_{t-1} - Level_{t-2}\vert \leq 1\&\vert(Level_t - Level_{t-3}\vert \leq 1$ Then Label is Constant
2	**Constant to Milking**	If $(Level_{t-1} > 0)\&(Level_t > 0)\&(Temp_t - Temp_{t-1} \geq 2.5)$ Then Label is Milking
3	**Milking to Milking**	If $(Level_{t-2} > 0)\&(Level_{t-1} > 0)\&(Level_t > 0)\&Level_t \geq (Level_{t-1}\&Level_t \geq (Level_{t-2}\&Level_t \geq (Level_{t-3}\&Level_t - Level_{t-2} \geq 1$ Then Label is Milking
4	**Constant to pickup**	If $(Level_{t-3} > 0)\&(Level_{t-2} > 0)\&(Level_{t-1} > 0)\&(Level_t > 0)\&Level_{t-1} - Levelt \geq 2\&Level_{t-2} - Levelt \geq 2\&Level_{t-3} - Levelt \geq 2$ Then Label is Logistics Pickup
5	**Milking to Constant**	If $(Level_{t-2} > 0)\&(Level_{t-1} > 0)\&(Level_t > 0)\&\vert(Level_{t-1} - Level_t\vert \leq 1\&\vert(Level_{t-1} - Level_{t-2}\vert \leq 1\&\vert(Level_t - Level_{t-3}\vert \leq 1$ Then Label is Constant
6	**Pickup to Pickup**	If $(Level_{t-3} > 0)\&(Level_{t-2} > 0)\&(Level_{t-1} > 0)\&Level_t - Levelt - 1 \leq 0\&Level_t - Levelt - 3 \leq 0\&Level_{t-2} - Levelt \geq 2$ Then Label is Logistics Pickup
7	**Pickup to Cleaning**	If $(Level_t = 0)\&(Temp_t - Temp_{t-1} \geq 4)$ Then Label is Cleaning
8	**Pickup to Delay**	If $(Level_t = 0)\&(Level_{t-1} = 0)\&(Temp_t \leq 5.5)$ Then Label is Delay in cleaning
9	**Delay to Delay**	If $(Level_t = 0)\&(Level_{t-1} = 0)\&(Temp_t \leq 5.5)$ Then Label is Delay in cleaning
10	**Delay to Cleaning**	If $(Level_t = 0)\&(Level_{t-1} = 0)\&(Temp_t - Temp_{t-1} \geq 10)$ Then Label is Cleaning
11	**Cleaning to Cleaning**	If $(Level_t = 0)\&(Level_{t-1} = 0)\&(Temp_t - Temp_{t-1} \geq 4)$ Then Label is Cleaning
12	**Cleaning to Empty**	If $(Level_t = 0)\&(Level_{t-1} = 0)\&(Temp_{t-1} - Temp_t > 1)$ Then Label is Cleaning
13	**Empty to Empty**	If $(Level_t \leq 2)\&(Level_{t-1} \leq 2)\&(Level_{t-2} \leq 2)\&(Level_{t-3} \leq 2)$ Then Label is Empty
14	**Empty to Milking**	If $(Level_{t-1} = 0)\&(Level_t - Level_{t-1} > 1)$Then Label is Milking

TABLE 7. Classification algorithms with their corresponding accuracy on the testing set.

Category	Classification Algorithm	Accuracy
Decision Trees	**C4.5 (J48)**	**0.911**
	REPTree	0.909
	RandomForests	0.87
	RandomTree	0.85
	Hoeffding	**0.915**
	Decision Stumps	0.738
Functions	**Artificial Neural Network (ANN)**	**0.912**
	Support Vector Machine (SVM)	0.906
	Logistics	0.906
	SMO	0.89
Instance-Based Methods	LWL	0.881
	Kstar	0.897
	KNN(N=5)	0.908
Bays	NaiveBays	0.904
	NaiveBaysUpdatable	0.904
Rules	OneR	0.857
	ZeroR	0.62
Ensemble Methods	AdaBoostM1	0.903
	Bagging	0.908
	Stacking	0.908
	Voting	0.62

In this table, bold numbers represent the best results. As can be seen, the greatest accuracy is achieved by Hoeffding (91.5%) and c4.5 (91.1%) from "Decision trees" category and also ANN (91.2%) from "Functions" category. Ensemble methods which utilize an agreement between several individual classifiers have generally produced good results (except for "Voting") though their accuracy is not the best achieved.

Consequently, using the proposed early detection system, we are able to automatically determine each event in a milking cycle with a high accuracy of 91.5%. This will help farmers monitor the milk cooling performance on a farm at any point in time as well as transporters in monitoring the milk quality stored on a farm and deciding in advance if it is likely to be unacceptable for pickup. However, instead of taking a reactive approach as utilized in earlier literature, the privilege of the proposed methodology is that it takes a proactive approach by designing an early detection system. In fact, this methodology is mutually beneficial for both farmers and transporters.

VII. CONCLUSION AND FUTURE WORK

Existing literature on milk quality is quite vast. The focus is more on a reactive approach of evaluating the quality based on the bacterial index and preventing milk with a high bacterial index from moving further downstream in a dairy supply chain. In this paper, we argue that it is not the best possible response if the intention is to maximize the milk life and slow down the decay in its quality. To this end, we have developed an early detection system by using IoT data to automatically determine the events in a milking cycle with high accuracy.

As discussed in the conceptual framework, the goal of our future research is to propose an early warning system using the present early detection system for proactive management of milk quality. As shown in Fig. 6, an early warning system enables the farmer to receive highly accurate alarms in case a particular action is required.

REFERENCES

[1] S. Stankovski, G. Ostojic, I. Senk, M. Rakic-Skokovic, S. Trivunovic, and D. Kucevic, "Dairy cow monitoring by RFID," *Scientia Agricola*, vol. 69, no. 1, pp. 75–80, 2012.

[2] M. Stuber, G. Kazmer, S. Gupta, and S. Zinn. *Milking Cows for Data Not Just Dairy Products*. Accessed: Sep. 14, 2018. [Online]. Available: https://phys.org/news/2018-07-cows-dairy-products.html

[3] D. Vasisht *et al.*, "FarmBeats: An IoT platform for data-driven agriculture," in *Proc. NSDI*, 2017, pp. 515–529.

[4] K. Bhargava, S. Ivanov, C. Kulatunga, and W. Donnelly, "Fog-enabled WSN system for animal behavior analysis in precision dairy," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Jan. 2017, pp. 504–510.

[5] B. Sharma and D. Koundal, "Cattle health monitoring system using wireless sensor network: A survey from innovation perspective," *IET Wireless Sensor Syst.*, vol. 8, no. 4, pp. 143–151, 2018.

[6] K. Bhargava, S. Ivanov, and W. Donnelly, "Internet of nano things for dairy farming," in *Proc. 2nd Annu. Int. Conf. Nanosc. Comput. Commun.*, 2015, Art. no. 24.

[7] S. Neethirajan, "Recent advances in wearable sensors for animal health management," *Sens. Bio-Sens. Res.*, vol. 12, pp. 15–29, Feb. 2017.

[8] S. Neethirajan, S. K. Tuteja, S.-T. Huang, and D. Kelton, "Recent advancement in biosensors technology for animal and livestock health management," *Biosensors Bioelectron.*, vol. 98, pp. 398–407, Dec. 2017.

[9] G. Caja, A. Castro-Costa, and C. H. Knight, "Engineering to support wellbeing of dairy animals," *J. Dairy Res.*, vol. 83, no. 2, pp. 136–147, 2016.

[10] A. Rahman, D. V. Smith, B. Little, A. B. Ingham, P. L. Greenwood, and G. Bishop-Hurley, "Cattle behaviour classification from collar, halter, and ear tag sensors," *Inf. Process. Agriculture*, vol. 5, no. 1, pp. 124–133, 2018.

[11] W. J. Yan, X. Chen, O. Akcan, J. Lim, and D. Yang, "Big data analytics for empowering milk yield prediction in dairy supply chains," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2132–2137.

[12] C. E. Eskildsen, "Prediction of phenotypic milk parameters using FT-IR spectroscopy, variable selection and pls regression-the overfitting issue," in *Proc. 9th Int. Symp. Milk Genomics Hum. Health*, 2012.

[13] Nidhi. (2017). *Hatsun Agro Milking the IoT Opportunity in Indian Dairy*. [Online]. Available: https://rctom.hbs.org/submission/hatsun-agro-milking-theiot-opportunity-in-indian-dairy/

[14] J. Niza-Ribeiro, A. C. Louz a, P. Santos, and M. Lima, "Monitoring the microbiological quality of raw milk through the use of an ATP bioluminescence method," *Food Control*, vol. 11, no. 3, pp. 209–216, 2000.

[15] P. S. Kajal, K. Y. Shivashkti, R. B. Shivangi, B. G. Arati, and R. K. Pooja, "IoT based detection of microbial activity in raw milk by using Intel Galileo Gen II," *Int. Res. J. Eng. Technol.*, vol. 4, no. 1, pp. 487–491, 2017.

[16] S. De Silva, K. Kanugala, and N. Weerakkody, "Microbiological quality of raw milk and effect on quality by implementing good management practices," *Procedia Food Sci.*, vol. 6, pp. 92–96, 2016.

[17] K. Ruangwittayanusorn, D. Promket, and A. Chantiratikul, "Monitoring the hygiene of raw milk from farms to milk retailers," *Agricult. Agricult. Sci. Procedia*, vol. 11, pp. 95–99, 2016.

[18] A. M. Elmoslemany, G. P. Keefe, I. R. Dohoo, J. J. Wichtel, H. Stryhn, and R. T. Dingwell, "The association between bulk tank milk analysis for raw milk quality and on-farm management practices," *Preventive Vet. Med.*, vol. 95, nos. 1–2, pp. 32–40, 2010.

[19] S. Nada, D. Ilija, T. Igor, M. Jelena, and G. Ruzica, "Implication of food safety measures on microbiological quality of raw and pasteurized milk," *Food Control*, vol. 25, no. 2, pp. 728–731, 2012.

[20] S. C. Murphy, N. H. Martin, D. M. Barbano, and M. Wiedmann, "Influence of raw milk quality on processed dairy products: How do raw milk quality test results relate to product quality and yield?" *J. dairy Sci.*, vol. 99, no. 12, pp. 10128–10149, 2016.

[21] L. Valík, F. Görner, and D. Lauková, "Growth dynamics of bacillus cereus and shelf-life of pasteurised milk," *Czech J. Food Sci.*, vol. 21, no. 6, pp. 195–202, 2003.

[22] G. Rysstad and J. Kolstad, "Extended shelf life milk—Advances in technology," *Int. J. Dairy Technol.*, vol. 59, no. 2, pp. 85–96, 2006.

[23] N. Ndraha, H.-I. Hsiao, J. Vlajic, M.-F. Yang, and H.-T. V. Lin, "Time-temperature abuse in the food cold chain: Review of issues, challenges, and recommendations," *Food Control*, vol. 89, pp. 12–21, Jul. 2018.

[24] M. S. Rahman, *Handbook Food Preservation*. Boca Raton, FL, USA: CRC Press, 2007.

[25] Australian Government Department of Agriculture and Water Resources. (2015). *Understanding and Complying With European Union Requirements for Raw Milk Temperatures*. [Online]. Available: http://www.agriculture.gov.au/SiteCollectionDocuments/aqis/exporting/dairy/goods-for-eu/raw-milk-temps.pdf

[26] M. S. Rahman, "Devondale murray goulburn supplier handbook," in *Supplier Handbook Southern Milk Region*, Murray Goulburn Co-operative Co., 2017, pp. 27–33. [Online]. Available: http://www.mgc.com.au/media/24629/Supplier-Handbook-Southern-Milk-Region.pdf

[27] *Milk Policy*, Dairy Farmers Milk Cooperative, Melbourne, VIC, USA, 2016, pp. 24–30.

[28] E. W. T. Ngai, T. K. P. Leung, Y. H. Wong, M. C. M. Lee, P. Y. F. Chai, and Y. S. Choi, "Design and development of a context-aware decision support system for real-time accident handling in logistics," *Decis. Support Syst.*, vol. 52, no. 4, pp. 816–827, 2012.

[29] J. E. Mendoza, A. L. Medaglia, and N. Velasco, "An evolutionary-based decision support system for vehicle routing: The case of a public utility," *Decis. Support Syst.*, vol. 46, no. 3, pp. 730–742, 2009.

[30] L. C. Q. Pimenta, "The logistics of milk collection: An exploratory case study between New Zealand and Brazil," M.S. thesis, Logistics Supply Chain Manage., Massey Univ., Palmerston North, New Zealand, 2010.

[31] K. Gwezuva, "An analysis of raw milk transportation and possible strategies to maintain milk quality in matabeleland region, Zimbabwe," M.S. thesis, Larenstein Univ. Appl. Sci., Leeuwarden, Netherlands, 2011.

[32] K. V. Katuku, "Farm factors associated with milk rejection at dairy cooperatives peri-urban nairobi," Ph.D. dissertation, Univ. Nairobi, Nairobi, Kenya, 2009.

[33] A. O'Connell, S. McParland, P. L. Ruegg, B. O'Brien, and D. Gleeson, "Seasonal trends in milk quality in Ireland between 2007 and 2011," *J. Dairy Sci.*, vol. 98, no. 6, pp. 3778–3790, 2015.

[34] Y. Suranindyah, E. Wahyuni, S. Bintara, and G. Purbaya, ''The effect of improving sanitation prior to milking on milk quality of dairy cow in farmer group,'' *Procedia Food Sci.*, vol. 3, pp. 150–155, 2015.

[35] M. J. Vilar, J. L. Rodríguez-Otero, M. L. Sanjuán, F. J. Diéguez, M. Varela, and E. Yus, ''Implementation of HACCP to control the influence of milking equipment and cooling tank on the milk quality,'' *Trends Food Sci. Technol.*, vol. 23, no. 1, pp. 4–12, 2012.

[36] I. E. M. El Zubeir and M. I. A. Ahmed, ''The hygienic quality of raw milk produced by some dairy farms in Khartoum State, Sudan,'' *Res. J. Microbiol.*, vol. 2, no. 12, pp. 988–991, 2007.

[37] B. Hill, B. Smythe, D. Lindsay, and J. Shepherd, ''Microbiology of raw milk in New Zealand,'' *Int. J. Food Microbiol.*, vol. 157, no. 2, pp. 305–308, 2012.

[38] J. C. Laguna, A. G. Higuera, R. Z. Cantabrana, and J. de las Morenas, ''Comprehensive traceability system of milk samples using RFID,'' in *Proc. Eur. Conf. Smart Objects, Syst. Technol. (SmartSysTech)*, Jun. 2012, pp. 1–8.

[39] J. de Las Morenas, A. García, and J. Blanco, ''Prototype traceability system for the dairy industry,'' *Comput. Electron. Agriculture*, vol. 101, pp. 34–41, Feb. 2014.

[40] P. Kadam *et al.*, ''IoT based detection of microbial activities in raw milk,'' *Int. J. Softw. Eng. Comput. Archit.*, vol. 1, no. 1, 2017.

[41] G. Van Schaik, L. E. Green, D. Guzmán, H. Esparza, and N. Tadich, ''Risk factors for bulk milk somatic cell counts and total bacterial counts in smallholder dairy farms in the 10th region of Chile,'' *Preventive Vet. Med.*, vol. 67, no. 1, pp. 1–17, 2005.

[42] J. C. F. Pantoja, D. J. Reinemann, and P. L. Ruegg, ''Associations among milk quality indicators in raw bulk milk,'' *J. Dairy Sci.*, vol. 92, no. 10, pp. 4978–4987, 2009.

[43] T. A. Mhone, G. Matope, and P. T. Saidi, ''Aerobic bacterial, coliform, *Escherichia coli and Staphylococcus aureus* counts of raw and processed milk from selected smallholder dairy farms of Zimbabwe,'' *Int. J. Food Microbiol.*, vol. 151, no. 2, pp. 223–228, 2011.

[44] S. Boopathi, F. P. Ahmed, S. T. Ram, T. Devika, and N. Suthanthira-Vanitha, ''Arm based temperature monitoring and control for milk pasteurization,'' *Int. J. Eng. Trends Technol.*, vol. 16, no. 5, 2014.

[45] A. Carullo, S. Corbellini, M. Parvis, and A. Vallan, ''A wireless sensor network for cold-chain monitoring,'' *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1405–1411, May 2009.

[46] K. Koutsoumanis, A. Pavlis, G.-J. E. Nychas, and K. Xanthiakos, ''Probabilistic model for *listeria monocytogenes* growth during distribution, retail storage, and domestic storage of pasteurized milk,'' *Appl. Environ. Microbiol.*, vol. 76, no. 7, pp. 2181–2191, 2010.

[47] E. Morelli, V. Noel, P. Rosset, and G. Poumeyrol, ''Performance and conditions of use of refrigerated display cabinets among producer/vendors of foodstuffs,'' *Food Control*, vol. 26, no. 2, pp. 363–368, 2012.

[48] I. Nainggolan, T. I. Nasution, Z. A. Ahmad, K. R. Ahmad, and A. S. Arif, ''Detection of milk quality degradation using biopolymer chitosan sensor,'' *Indonesian J. Mater. Sci.*, vol. 5, pp. 5–8, 2011.

[49] L. Wang, J. S. L. Ting, and W. H. Ip, ''Design of supply-chain pedigree interactive dynamic explore (SPIDER) for food safety and implementation of hazard analysis and critical control points (HACCPs),'' *Comput. Electron. Agriculture*, vol. 90, pp. 14–23, Jan. 2013.

[50] A. O'Connell, P. L. Ruegg, K. Jordan, B. O'Brien, and D. Gleeson, ''The effect of storage temperature and duration on the microbial quality of bulk tank milk,'' *J. Dairy Sci.*, vol. 99, no. 5, pp. 3367–3374, 2016.

[51] M. W. Griffiths, J. D. Phillips, and D. D. Muir, ''Effect of low-temperature storage on the bacteriological quality of raw milk,'' *Food Microbiol.*, vol. 4, no. 4, pp. 285–291, 1987.

[52] M. Thibaud, H. Chi, W. Zhou, and S. Piramuthu, ''Internet of Things (IoT) in high-risk environment, health and safety (EHS) industries: A comprehensive review,'' *Decis. Support Syst.*, vol. 108, pp. 79–95, Apr. 2018.

[53] A. Zakeri, M. Saberi, O. K. Hussain, and E. Chang, ''Addressing missing data and data competitiveness issues: Transforming tacit knowledge into explicit form by fuzzy inference learning system,'' *Int. J. Fuzzy Syst.*, vol. 20, no. 4, pp. 1224–1239, 2018.

[54] L. Yu and H. Liu, ''Feature selection for high-dimensional data: A fast correlation-based filter solution,'' in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 856–863.

[55] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

Authors' photographs and biographies not available at the time of publication.