Distributed Inference Condition Monitoring System for Rural Infrastructure in the Developing World

Heloise Greeff, Achut Manandhar, Patrick Thomson, Robert Hope, and David A. Clifton

***Abstract*— Remote condition monitoring systems for rural infrastructure lack "intelligent" analysis and advanced insights offered by recent Internet of Things devices. This is because the extreme and inaccessible operating locations necessitate the conservative use of limited resources, such as battery life and data transmission. Present implementations are often limited to usage data loggers, which are informative of general usage but post-processed advanced insights lag real-time system changes. A light-weight novelty filter is implemented onboard rural handpumps to identify subsets of data as potential infrastructure failure. The "intelligent" summaries of these data subsets are sent to a cloud-based system, where more advanced machine learning approaches are applied to increase the fidelity of potential failure predictions. The proposed method was tested on three independent data sets and found that the on-pump novelty filter could predict failure with up to 61.6% *in situ*. Incorporating more advanced machine learning methods on the cloud-based platform increased the classifiers' positive predictive value by at least an additional 10%–73%. This novel method has proven that it is possible for rural operating, resource-constrained devices to use lightweight, onboard machine learning approaches to perform anomaly detection in the embedded system. Distributed inference between the embedded system at the rural node and powerful cloud-based machine learning algorithms offers robust information without the need for expensive hardware or sensors embedded *in situ*—making the possibility of a large-scale (and perhaps even continent-wide) monitoring system feasible.**

***Index Terms*— Distributed inference, embedded systems, remote condition monitoring, machine learning, rural development.**

I. Introduction

RURAL infrastructure, such as manual handpumps, play an important role in improving quality of life as well as driving economic growth, particularly in developing countries where rural areas account for the larger part of the geographical regions [1]. However, sustainable provision of reliable infrastructure hinges on the high standard of both installation and maintenance, the latter of which is often under invested in or completely neglected [2]. Downtime due to system failure in rural settings is often even greater than in urban settings due to the practical challenges in the supply of spare parts combined with a lack of local skills. In sub-Saharan Africa, it is estimated around one third of the one million handpumps used daily by nearly 200 million people are not working at any time and often remain broken for up to 30 days [3].

Predictive health monitoring is widely used in engineering applications to detect damage to infrastructure as early as possible. Forecasting failure rather than merely detecting failure once it occurs helps to reduce the downtime of systems, and, ideally, performing predictive maintenance can avoid downtime completely. With this approach already widely used in many fields from commercial and military jet engines [4], through to patient monitoring in health systems [5], it can be directly extended to monitoring the condition and use of rural infrastructure, such as off-grid solar home systems [6] or, in the case discussed in this paper, handpumps [7]–[9].

Despite the importance of rural infrastructure and the potential impact of predictive monitoring, the implementation of remote condition monitoring systems in these extreme rural settings has historically been limited to data loggers [10], [11]. This is largely due to the technical and logistical challenges, such as battery life, data-transmission bandwidth limitations and long or expensive maintenance cycles, associated with operating in such remote locations. This necessitates the use of sensors that are robust, reliable, low-power and low-cost. These considerations may compromise performance leading to data that is lower frequency, more coarsely quantised or with a poor signal-to-noise ratio, in turn making advanced data analysis challenging but increasingly important [12]. This is in stark contrast to, sophisticated in-home patient monitoring systems using advanced machine learning techniques operating in a benign environment with ready access to grid power and wireless broadband that satisfy intense processes and data transfers [13], [14].

Wilson *et al.* [15] demonstrated that the use of ensemble machine learning in remote monitoring of rural handpumps, when combined with a preventive maintenance service model, could increase the uptime in rural handpumps to 99 per cent. Although such performance improvements can translate directly into positive health impacts for local communities [16], the proposed model sacrifices prediction sensitivity (51.0%) over specificity (99.3%) when identifying independent failure events. However, when considering the failures as a series of "failure days", the proposed method correctly

Manuscript received October 8, 2018; revised November 7, 2018; accepted November 7, 2018. Date of publication November 22, 2018; date of current version February 6, 2019. This work was supported in part by UNICEF and in part by NERC through the UPGro Consortium Grant. The associate editor coordinating the review of this paper and approving it for publication was Dr. Marco J. Da Silva. *(Corresponding author: Heloise Greeff.)*

H. Greeff, A. Manandhar, and D. A. Clifton are with the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford OX3 7DQ, U.K. (e-mail: heloise.greeff@eng.ox.ac.uk; achut.manandhar@eng.ox.ac.uk; david.clifton@eng.ox.ac.uk).

P. Thomson and R. Hope are with the School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, U.K. (e-mail: patrick.thomson@ouce.ox.ac.uk; robert.hope@ouce.ox.ac.uk).

Digital Object Identifier 10.1109/JSEN.2018.2882866

identifies 24 out of 25 events. Since failures are novel events, a low sensitivity would lead to an extremely high number of false alerts that have a direct cost associated to it when performing rural monitoring, unlike when a supervisor needs to walk for only a few minutes to inspect the alert on a production line or at a patient bed. The advantage of the approach proposed in this paper is that the in situ novelty filter greatly reduces the data from statistically healthy handpumps, which in turn increases the sensitivity of the classifier and thereby reduces the number of costly false alarms.

The importance of rural infrastructure and the lack of adequate advanced remote monitoring systems suggests that the implementation of a robust, automated and continuous condition monitoring system for rural infrastructure could be of great significance for rural development. While usage monitoring is a crucial first step to increase transparency of tracking metrics for the UN Sustainable Development Goals [17], it does not address the local users' need for a sustainable and reliable water, transport, or energy supply network.

To bridge this gap, we propose a system designed for the needs of rural infrastructure. This uses lightweight machine learning techniques, in the form of a logistic regressor (LR) novelty filter, applied at the nodes to perform initial processing of the raw observations prior to transmission to a cloud-based platform. The on-board novelty threshold will then determine the proportion of the original data to be transmitted based on the on-board assessment of the condition of the node. More advanced machine learning methods can then be applied at the cloud-based platform, where power and processing requirements are not limited by the factors mentioned earlier. We will demonstrate how this can be implemented using rural handpumps as nodes.

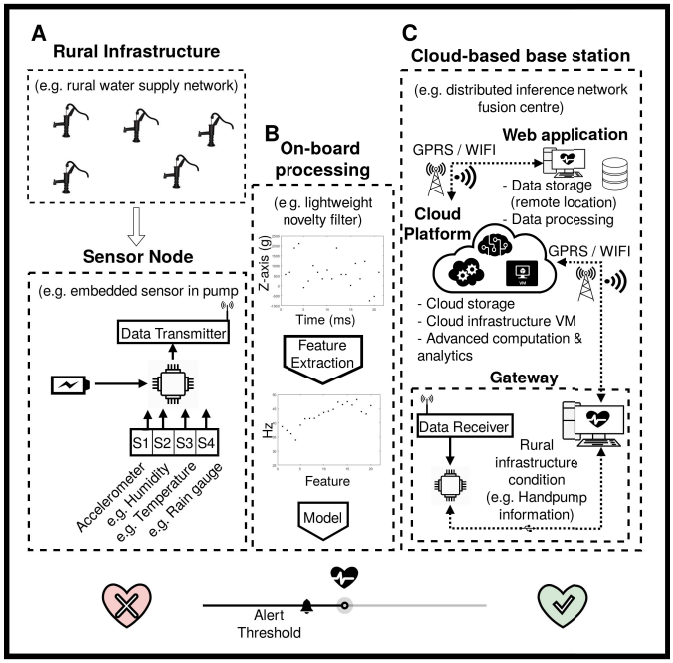

Fig. 1. Example of a distributed inference condition monitoring system: (A) Node with related embedded sensors, (B) On-board lightweight novelty filter producing data summaries, and (C) Gateway node and heavyweight cloud-based computing and analytics.

TABLE I
ESTIMATED POWER CONSUMPTION OF ESSENTIAL COMPONENTS IN PROPOSED EMBEDDED SYSTEM OPERATING AT 25°C

Component	Run Mode	Sleep Mode
Microprocessor (8-bit at 8 MHz)	20 mA	5 mA
Accelerometer at 96 Hz	200 μA	0.5 μA
GSM Module (for SMS)	250 mA	40 μA

*Approximation only. Actual consumption will depend on the components selected and operating conditions.

II. METHODS

The processing sequence for distributed inference in predicting handpump failure is shown in Figure 1. The algorithm consist of four steps: accelerometry feature extraction at the sensor node on-board the handpump, lightweight novelty filtering on-board, transmission of smaller data packages, and the heavyweight processing in the cloud. The following sections will describe the proposed labeling protocols, the data sets used to train and validate the algorithm, the on-board data preprocessing procedure, and each step of the proposed distributed inference condition monitoring system.

A. System Design Requirements

The distributed condition monitoring system should aim to: (i) use low-cost, embedded sensors to acquire accelerometry data from the routine daily use of handpumps, (ii) perform on-pump condition monitoring and pre-processing of data, (iii) analyze the resulting data in situ to produce informative assessments of the equipment condition [18], (iv) communicate with a cloud-based platform via the local mobile telecommunications network, and (v) perform more advanced machine learning techniques on the cloud-based system to increase the accuracy and certainty of the classified condition of the monitored equipment [19], [20]. An overview of the proposed system can be seen in Figure 1 with three main sections:

A. *Sensor node* contains the sensor, battery, and data transmitter. In this study, the embedded system attached to the handpump handle to record vibrations consists of three essential elements: (a) an IC-based, 96 Hz accelerometer; (b) an 8-bit microprocessor; and (c) a GSM modem. The estimated power consumption of this node in run and sleep mode is compared in Table I. Alternatively, nodes could make provision for other sensors depending on the infrastructure type that is being monitored, like power meters in home solar units, heart rate monitors in patients or temperature and humidity sensors in agriculture. This network could contain hundreds of nodes in a small geographical region covered by the local telecommunications network and managed by the local maintenance delivery team.

B. *On-board novelty filter* performs real-time preprocessing of the data acquired during use, like pumping, and analyses the data using a LR novelty filter to produce "intelligent" data summaries that aim to flag potential infrastructure failure.

TABLE II
DESCRIPTION OF THE MECHANICAL CONDITION AND SHORT-TERM WATER QUANTITY CLASSIFICATION LABELS ASSIGNED TO EACH RECORDING

Condition Description	Water Flow	Label
Excellent working condition	High flow	**1**
Noisy but working	Average to Low flow	**1**
Dry borehole	No flow	**0**
Rising main leak	No flow	**0**
Broken U-seal	No flow	**0**
Worn U-seal	Average to low flow	**1**
Handpump body leaking	Average to low flow	**1**
Worn bush bearing	Average Flow	**1**
Stiff handle	Average Flow	**1**
Other	Average to low flow	**1**

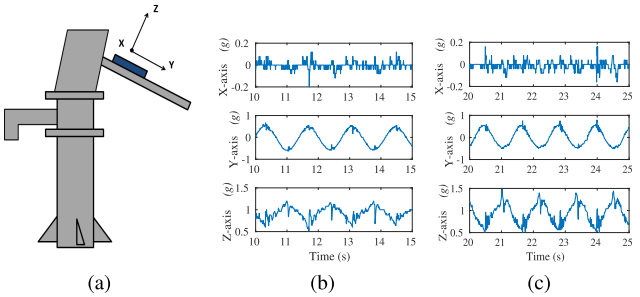

Fig. 2. (a) A diagram showing the experimental set-up of a vibration sensor attached to a handpump and the relevant orthogonal axes of accelerometry with a 5 s interval of unprocessed gravitational force (g) accelerometer data of the same handpump in a (b) normal and (c) abnormal condition in the X, Y, and Z dimensions (upper to lower plots, respectively) for the same user.

C. *Cloud-based platform as base station* performs more complex processing of the data summaries using advanced machine learning methods to increase prediction fidelity. A web application with a bespoke user interface could be used to access and setup data transmission and user alert protocols.

B. Defining Functionality and Labeling Data

Unlike patient-monitoring, there are no standardized labeling protocols for rural infrastructure conditions. Following interviews with domain experts and an in-depth literature review of handpump functionality in West- and East-Africa [21]–[24], we introduced two main attributes to classify handpump functionality according to their conditions:

1) *Short-Term Water Quantity:* a handpump is either classed as normal (C1) or abnormal (C0). A handpump is considered normal when water flows from the spout while pumping and abnormal when no water flows from the spout while pumping.
2) *Mechanical Performance:* ten sub-categories, shown in Table II, are used to identify the mechanical attributes that describe the functionality and physical condition of the handpump.

The data was labeled using notes collected during in-person, contemporaneous observations. This level of labeling is limited in that it allows for only two classes. It is clear that certain conditions, like those with average or low flow, are not entirely normal nor entirely abnormal. However, for this initial proof of concept, it is assumed that the proposed labels are adequately descriptive.

C. Sensor Nodes and Data Collection

To determine the condition of a node, we measured the vibrations of an operating *Afridev* handpump [25] by retrofitting a sensor, containing a consumer grade accelerometer with a sampling frequency of 96 Hz. Each sensor was housed in a waterproof casing and mounted with tamper-proof bolts inside the handle at a position closest to the handpump body, as shown in Fig. 2a, without interfering with the range of motion of the handle. The accelerometer operates in three orthogonal axes, the Y-axis being along the line of the handpump handle. A 5 s interval of data from a handpump in a normal and abnormal condition is shown in Fig. 2b and 2c, respectively. Following 5 min of inactivity, the loggers switch to a low power state to preserve battery life, restarting after 10 s of continuous motion. For a regularly used handpump, operating nearly constantly for 8 to 12 hours per day, this translates to about 1 gigabyte of data per handpump per month. All of the handpumps in the original study region are located in areas with sufficient network coverage to transmit the data via the telecommunications network. However, for this initial study, to preserve battery and cost of data transmission, the data was stored locally on a micro-SD card and downloaded manually.

D. Data Sets

We consider three data sets collected from pumps in our study site in Kwale, Kenya. Data sets contain high-frequency (96 Hz) three axes accelerometery readings from a logger mounted inside the handle of handpumps. For this study we report results on data from the Y-axis, which has shown to be the most informative for this initial analysis. While we observed a significant difference in the spectra of deep and shallow handpumps, as illustrated in Fig. 3, we focused mainly on data collected from deep wells, operating at depths greater than 25 m. In our study area, deep handpumps are typically located at greater elevations where other groundwater sources tend to be sparse, often making them the primary source of drinking water for the nearby communities and households. However, the greater weight of water and handpump rods being lifted, combined with the increased level of use, lead to more frequent breakdowns of these handpumps compared to those located at shallower wells. Failures at deeper wells are more labor intensive and time consuming to repair. This, together with their more inaccessible locations, result in longer downtimes at deeper wells, making remote condition monitoring and timely repair even more important.

The first data set, $\mathcal{D}_m$, represents a general inter-handpump system consisting of twelve different handpumps of varying operating depths ranging between 6 m to 53 m, and was included to establish the baseline performance of a general classifier. The second data set, $\mathcal{D}_{d,1}$, represents a deep-operating inter-handpump system consisting of eight different

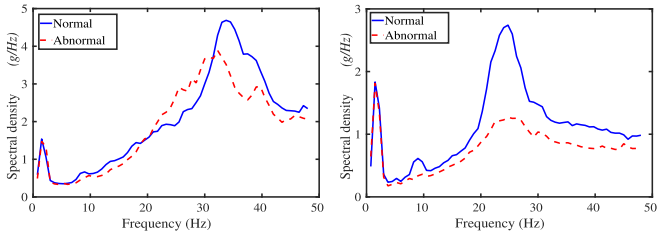

Fig. 3. Median amplitude of the spectral data for (a) deep and (b) shallow operating handpumps.

handpumps operating at depths between 33 m to 54 m. The third data set, $\mathcal{D}_{d,2}$, represents a deep-operating intra-handpump system of one handpump operating at 54 m. Although the implementation of a region-wide intra-handpump system is unfeasible, this data set was selected to investigate the influence of different failure types, while controlling for the handpump.

The data set contained recordings from eight different common handpump failure types. All the data sets were balanced and randomly divided into a training-and-validation set (80%) and a test set (20%).

E. On-Board Pre-Processing Procedure

Data pre-processing consisted of: peak and trough detection, high-pass filtering, windowing in the time-series domain, and transforming to the frequency domain.

First, the peaks and troughs from the pumping waveforms were determined. For waveforms on the Y-axis, we considered periods with a difference greater than 0.25 s between the peaks and troughs. This was done to eliminate measurements from children playing with the handpumps.

Next, we applied high-pass filtering to the signal. Our hypothesis is that changes in handpump conditions are not affected by the component of the signal caused by the relative low-frequency motion of the user moving the handpump handle. As such, we use a high-pass filter (HPF) to remove the low-frequency components associated with the pumping tempo that are not indicative of handpump failure, and which retains an estimate of the fast-moving trend components of the noisy vibrations.

Due to the resource constraints, such as an 8-bit microprocessor and limited battery, we use a phase-corrected 4-point moving average (MA) finite impulse response (FIR) filter to represent the shape of the recording, which is then removed from the original signal. The filter calculates the average of a number of points from the input signal such that each point of the output signal, y, is calculated as [26]:

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \tag{1}$$

where x is the input signal and M is the number of points used in the moving average.

Finally, Fast-Fourier transforms [27] are used to decompose the signal into a sum of sinusoidal basis functions used to describe the frequency content within the time-series waveform. The recordings were partitioned into 1.3 s windows with 50% overlap. This creates 128 samples per window, equivalent to 64 frequency components with a resolution of 0.75 Hz per component for a sampling frequency of 96 Hz. To account for truncated waveforms with discontinuous endpoints resulting from the finite windows [28], a 128-point Hamming window function is applied [29]. The final result after FFT is a feature vector with 64 frequency components per window, up to the Nyquist rate.

We then selected a subset of 20 features by uniformly sampling across frequency bins 3 to 60, discarding low frequency components, equivalent to 0 to 2.25 Hz, which represent the pumping motion of the user, where a full handpump stroke has a median period of 1.1 s. Artifacts of this pumping motion can be seen in Fig. 3.

F. Lightweight On-Board Novelty Filter

As a first layer of condition monitoring, the on-board classifier should be quick to train and fast to classify unknown records, such that it is suitable for applications with limited processing power and bandwidth. In machine learning, logistic regression can be used to model the posterior probability of input variables, $\mathbf{X}$, being associated with a class by fitting a linear model to the feature space. As a linear classification method it is used to categorize the dichotomous dependent variable and predict the probability (0,1) of membership of one class (e.g., True/False) in a two-class setting, making it suitable for this lightweight approach.

1) Logistic Regression (LR): The LR model was formulated using the sigmoidal hypothesis function, $h(\mathbf{x}_n)$, with a probability that a given example is of class 1:

$$P(\ell_n = 1|\mathbf{x}_n, \mathbf{w}) = h(\mathbf{x}_n) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_n}} \tag{2}$$

where $\mathbf{w}$ is a set of weights assigned to each input feature, $\mathbf{x}_n$. The decision threshold, T, is used to assign a given example to class 1 based on whether the hypothesis function is greater than or less than T. This threshold was varied to change the size of the data subsets that was subsequently transmitted to the offline classifier. As the value of T is decreased, the size of the subset s increases, as more of the novelty scores are deemed abnormal.

The LR model was trained using 5-fold cross-validation (CV), where each training set, $\mathcal{D}_t$, was randomly subdivided into 5 equal subsets to construct 5 independent training-and-validation sets. The LR regularization parameter, λ, for each independent LR model was optimized by maximizing the area under receiver operator curve (AUROC) on the held-out validation sets.

G. Heavyweight Cloud-Based Classifier

The next stage of condition monitoring involved performing heavyweight machine learning processing on the subsets of data flagged by the lightweight on-board novelty filter. In addition to LR, we considered support vector machine (SVM) [30] and random forest (RF) [31] classifiers for the cloud-based

processing. The on-pump novelty filter is used to ensure that under normal operating conditions the vast majority of data is not transmitted and only as the on-pump model suspects the condition is degrading will data be transmitted to the cloud, which means that in most cases the cloud-based system will only be receiving data for novelty-filtered data related to abnormal conditions. To simulate this operating scenario, the cloud-based classifiers are trained using data which contain both normal and abnormal examples and tested using data which only contain examples flagged as abnormal by the on-pump novelty filter. However, in reality, these test examples may contain both normal and abnormal examples given that the on-board novelty filter is likely to misclassify some proportion of data.

1) Logistic Regression (LR): For comparison, we repeated the LR model method on the novelty filtered data, using the same method described in section II-F.1.

2) Support Vector Machines (SVM): The SVM classifier model was trained using the radial basis function, $\exp(-\gamma \|\mathbf{x}-\mathbf{x}'\|^2)$, to project the individual scores from the novelty filter where two classes may be linearly separable. The SVM model was also trained using the 5-fold CV method, using different training and validation sets. The SVM hyperparameters: the kernel bandwidth, γ, and penalty cost factor, C, were optimized using grid search, where $\gamma = 2^a$ for $a \in [-10, -9, \ldots 5]$ and $C = 2^b$ for $b \in [-5, -4, \ldots 10]$, by maximizing the sum of all AUCs over all CV folds. The grid search was done independently for each CV fold. Once this was completed, we repeated the process to perform a fine grid search, where $a_{\text{opt}} \in [a_{\text{opt}} - 1, a_{\text{opt}} - 0.75, \ldots a_{\text{opt}} + 1]$ and $b_{\text{opt}} \in [b_{\text{opt}} - 1, b_{\text{opt}} - 0.75, \ldots b_{\text{opt}} + 1]$. The refined hyperparameters, γ^* and C^*, from the fine grid search was used to train the SVM models from the training sets from each of the data sets.

3) Random Forest (RF): The RF classifier model was trained using a random selection of a subset of features, Θ_k, and a random subset of the training data, $\mathcal{D}(t)$, to grow each decision tree, T. At each node, t, of the tree, the split $s_t = s^*$ to separate the input vector, $\mathbf{X}$, was chosen to minimize the impurity, $i(t)$, in class labels [32] by minimizing the misclassification such that $i_E(t) = 1 - \max\{p_c\}$, where p_c is the probability of a class C. The importance of the variable input feature $\mathbf{X}$ for predicting the output is based on their weighted impact on decreasing the impurity of that node for all N_T trees in the forest:

$$\text{Imp}(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_m} \frac{N_t}{N} \Delta i(s_t, t) \qquad (3)$$

where v_s is the variable used in split s_t.

The RF hyperparameters: the number of threes, N_T, the number of feature vectors in each decision tree, and the proportion of training data to be bootstrapped were again optimized using a grid search.

H. Condition Estimation

Following the analysis above, we produced a condition score in situ at the remote node, $Q_{n,i}$, on-board operating handpumps. Due to the lightweight processing requirement of the on-board classifier used during this feasibility study, the temporal dependence of the accelerometer observations were not considered at the node. This was done during post-processing by aggregating the classifier scores over consecutive examples to varying degrees by applying a moving average (MA) window and increasing the size of the window from 7 s to 27 s. This produced three lightweight condition scores, $Q_{n,i}$, per data set with $i = 1 \ldots 3$ equivalent to [Raw on-board score, 7 s MA window score, 27 s MA window score].

The in situ condition score, $Q_{n,1}$, was then used to filter the transmitted data such that data summaries sent to the cloud-based classifier contain only abnormal examples, as labeled by the on-board novelty filter. Finally, we produced condition scores for each of the three offline classifier methods (LR, SVM, RF) using the novelty filtered data.

I. Evaluation Method

The ability of our on-pump novelty filter and subsequent cloud-based classifier to verify CM reliability was assessed using the receiver operating characteristic (ROC) to compare the performance. This metric compares the actual and predicted outputs for each class. The true positive rate (TPR), or *sensitivity*, of a classifier is defined to be the probability of detection such that, $\text{TPR} = \frac{\sum \text{True Positive}}{\sum \text{Condition Positive}}$ and the false positive rate (FPR), or *fall-out*, is defined to be the probability of a false alarm such that, $\text{FPR} = \frac{\sum \text{False Positive}}{\sum \text{Condition Negative}}$. Optimizing the area under the ROC (AUC) will maximize handpump failure detection while simultaneously minimizing false alarms, which can be costly in real-life. In the ideal case, the classifier would be very sensitive (TPR = 1) with no false alarms (FPR = 0).

For the on-pump condition classification, we compared the the performance of $Q_{n,i}$, to a baseline control score, $Q_{n,\text{lab}}$, generated in the lab using the same original data but assuming no processing or power constraints as would be experienced on-board the local novelty filter.

III. Results

A. Performance of On-Pump Novelty Filter

Table III shows that the intra-handpump classifier, $\{Q_{n,i}, \mathcal{D}_{d,2}\}$ pairs, performs substantially better than the inter-handpump classifiers, $\{Q_{n,i}, \mathcal{D}_m \,/\, \mathcal{D}_{d,1}\}$ pairs, achieving up to 86.2 per cent AUROC compared to 65.7 per cent. However, the performance of the general inter-handpump classifier is sufficient to use as a lightweight novelty filter since the large-scale implementation of pump-specific classifiers would be too costly and unrealistic to roll-out across entire region-wide rural water supply networks.

In all three cases, the lab generated scores, $Q_{n,\text{lab}}$, outperforms those generated by the on-pump classifier, $Q_{n,i}$, by 7.5 to 12.1 per cent. Given the limitations of the embedded system, it was expected that the accuracy of the on-pump classifiers would suffer compared to the lab-simulated results. Due to the lightweight processing requirement of the on-board classifier, the temporal dependence of the accelerometer observations

TABLE III

RESULTS FOR FIELD-BASED, $Q_{n,i}$, AND LAB-SIMULATED, $Q_{n,\text{lab}}$, ON-BOARD CONDITION CLASSIFICATION SCORES, GIVEN THE MEAN AUC OF 20 ITERATIONS (ONE STANDARD DEVIATION)

	$\mathcal{D}_m$	$\mathcal{D}_{d,1}$	$\mathcal{D}_{d,2}$
$Q_{n,1}$	61.32±0.7	61.55±0.4	68.87±0.3
$Q_{n,2}$	64.76±0.4	65.08±0.2	77.31±0.2
$Q_{n,3}$	65.26±0.1	65.67±0.0	86.21±0.0
$Q_{n,lab}$	68.89±0.3	69.80±0.2	88.01±0.2

have not been considered. However, post-processing of the ROC scores indicate that the classifier performance improves when temporal correlation is incorporated by aggregating the classifier scores over consecutive examples (to varying degrees as the moving average window size is increased 7 s to 27 s). This type of post-processing is fairly lightweight and can be easily implemented on-board the handpump to improve on-pump novelty scores, which will bring it nearly on par with the lab-simulated results.

Fig. 4 compares the receiver operator curve (ROC) scores for (a) a general classifier, (b) a depth-specific inter-handpump classifier, and a depth-specific intra-handpump classifier with two different failure types, (c) a broken handpump rod and (d) a leaking rising main. The curves for the general and depth-specific inter-handpump classifiers look almost identical. This is likely a result of the underrepresentation of data from shallow handpump failures in the training and test sets of the general classier. Given that deep handpumps are likely to break more frequently and repairs are more time- and labor-intensive, the need for such classifiers are more important for deep operating handpumps.

The case studies shown in Fig. 4 demonstrate two key findings: (i) general on-board classifiers may perform sufficiently well as not to necessitate the need for depth-specific classifiers, as shown in Fig. 4a and 4b; and (ii) it may be possible to identify specific failure types if the system has *a priori* knowledge of the handpump operating depth. However, certain extreme failure types that are physically located closer to the sensor, like a broken handpump rod shown in Fig. 4c, are easier to detect than less severe failures located further away from the sensor, like a leak in the rising main 43 m down the borehole shown in Fig. 4d. It may also confirm the limitations in the classification labels of our data, such as labeling low water flow caused by a leak in the rising main as abnormal when in reality it is neither entirely normal nor entirely abnormal but rather indicative of an imminent failure event than a failure in itself.

B. Performance of Offline Novelty Filtered Classifier

The next stage of condition monitoring involved performing more advanced machine learning methods on the data flagged by the lightweight on-pump algorithm.

Table IV shows that in all three cases the LR classifier is sufficiently lightweight in that it reaches the optimum classification accuracy by using only 6 to 15 per cent of the flagged data from the on-pump novelty filter, compared to 89 to 98 per cent required by the RF classifier and 97 to 100

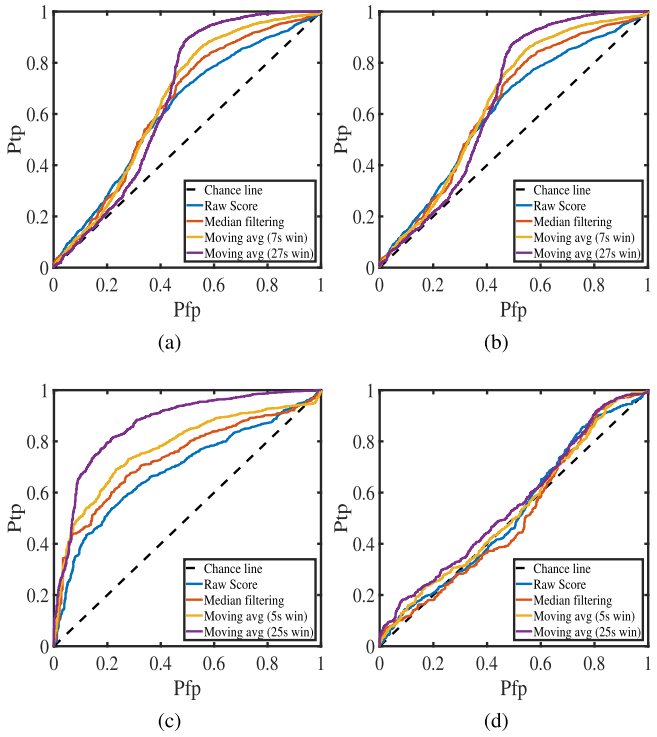

Fig. 4. ROC comparing the on-pump novelty filter classifier performance trained using different data subsets for: (a) a general classifier trained using $\mathcal{D}_m$, (b) a depth-specific inter-handpump classifier trained using $\mathcal{D}_{d,1}$, (c) a broken handpump rod in a depth-specific intra-handpump classifier trained using $\mathcal{D}_{d,2}$, and (d) a rising main leak in a depth-specific intra-handpump classifier trained using $\mathcal{D}_{d,2}$.

TABLE IV

RESULTS FOR LAB SIMULATED CLOUD-BASED CONDITION CLASSIFICATION OF ON-PUMP PROCESSED DATA, GIVEN THE MEAN AUC OF 20 ITERATIONS (ONE STANDARD DEVIATION)

		$\mathcal{D}_m$	$\mathcal{D}_{d,1}$	$\mathcal{D}_{d,2}$
Offline scores at 100% data	LR	59.6 ± 0.1	60.0 ± 0.0	79.4 ± 0.0
	RF	$\mathbf{71.8 \pm 2.5}$	$\mathbf{70.9 \pm 2.0}$	$\mathbf{86.8 \pm 1.4}$
	SVM	60.8 ± 0.2	65.2 ± 0.0	75.0 ± 0.0
Max offline scores	LR	60.9 ± 1.8 at 8%	60.7 ± 2.3 at 6%	80.8 ± 2.7 at 15%
	RF	$\mathbf{73.0 \pm 1.7}$ at 89%	$\mathbf{72.1 \pm 2.3}$ at 98%	$\mathbf{87.5 \pm 1.5}$ at 91%
	SVM	60.8 ± 0.6 at 99%	65.3 ± 0.5 at 97%	75.0 ± 0 at 100%

per cent for the SVM. In all three cases, the RF classifier outperforms the LR and SVM classifiers.

As before, the LR classifier shows little difference in performance between a general, $\mathcal{D}_m$, or depth-specific, $\mathcal{D}_{d,1}$, inter-handpump data set. The RF classifier does marginally better for depth-specific,$\mathcal{D}_{d,1}$, data set. Both the LR and RF classifier performance benefit from processing in the cloud.

Unlike, the RF and LR classifiers, the SVM classifier performance benefits more from the depth-specific data set, $\mathcal{D}_{d,1}$, than the general data set, $\mathcal{D}_m$. The SVM classifier performance is likely to increase as we continue to collect more depth-specific data. This is suggested by the significant reduction in variance for the SVM classifier as the proportion of data is increased.

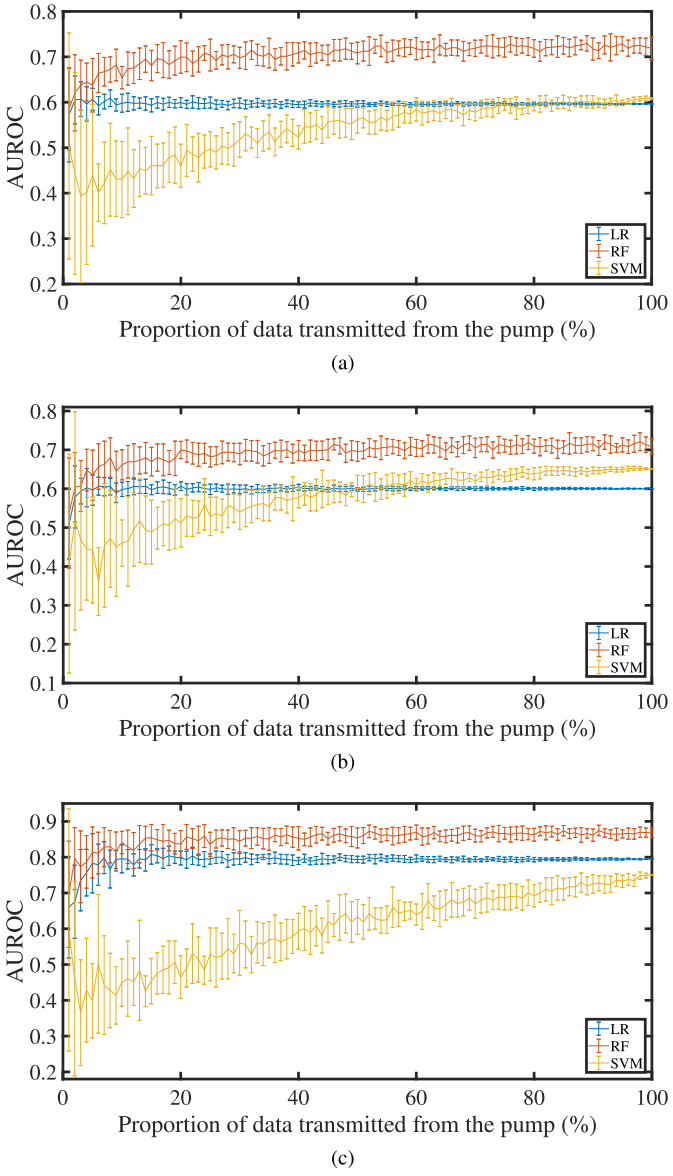

Fig. 5. AUROC comparison of the offline classifier using novelty filtered data subsets for: (a) a general classifier trained using $\mathcal{D}_m$, (b) a depth-specific inter-handpump classifier trained using $\mathcal{D}_{d,1}$, and (c) a depth-specific intra-handpump classifier trained using $\mathcal{D}_{d,2}$.

Fig. 5 compares the AUROC scores for the three types of classifiers trained using the three different proposed data sets. In all three sets, the offline LR classifiers benefits the least from the addition of an increase in the subset used for testing. However, the standard deviation in the predictive accuracy of the LR model does reduce as the test set size increased. Conversely, however, in all three cases the RF classifier achieves the highest overall accuracy, with relatively little data, and benefits minimally from more data both in improving prediction accuracy or decreasing prediction variance.

Overall, more advanced cloud-based machine learning methods offer a 10 per cent improvement from the raw on-pump generated condition scores, $Q_{n,i}$. The three cases show that there is a trade-off between accuracy and specificity. Whilst the RF classifier may offer a higher overall prediction accuracy, both LR and SVM can dramatically reduce the variability in predictions as the proportion of data supplied is increased. This is an important trade-off to note for the final system design since it may have a direct impact on operational cost of the final condition monitoring system.

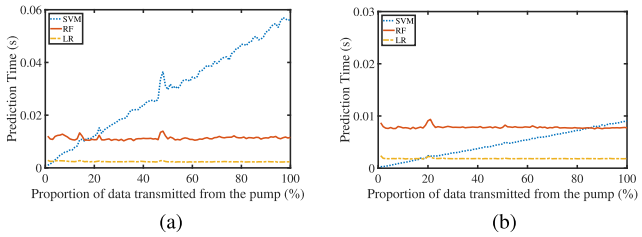

Fig. 6. Comparison of the classifier prediction run times as the proportion of data transmitted from the handpump varies for: (a) inter-handpump and (b) intra-handpump condition monitoring systems.

C. Improving Real-Time Performance

To ensure the system is suitable for real-time implementation within the constraints of the limited resources, such as battery life and data transmission, two additional design factors were considered that would lead to make the final model leaner without affecting the overall system performance.

For optimizing the run time cost model of operating such a large-scale distributed system, we considered the impact of potentially distributed run time plans and the machine learning characteristics of each classifier as a direct trade-off of its overall prediction accuracy.

1) Prediction Run Time: Time implementation of complex, region-wide monitoring systems should aim to optimize machine learning approaches by being sensitive to memory use and parallelism. Fig. 6 shows a comparison of the prediction run times for each classifier for the intra- and inter-node systems. In both cases, the prediction time for the LR and RF classifier remains constant as the proportion of data increases while SVM prediction time increases linearly. As the most lightweight method, the LR classier shows the fastest run time, irrespective of the system type.

2) Number of Features: As another method of reducing the transmission of data, we considered the additional gain of increasing the number of features on the misclassification rate of the different classifiers, shown in Fig 7. In both cases, the intra- and inter-handpump condition monitoring systems gain very little predictive accuracy from using more than 8 or 10 features, respectively. This suggests that the cost of data transmission during implementation can be reduced by nearly half by reducing the size of the data packages required by the cloud-based system.

While these factors may seem trivial when considered independently, the trade-offs between the different classifiers and respective prediction times along with the number of features transmitted as part of the "intelligent" data summaries transmitted from the rural node are important design considerations for the implementation of the final distributed system, without sacrificing the performance of the system.

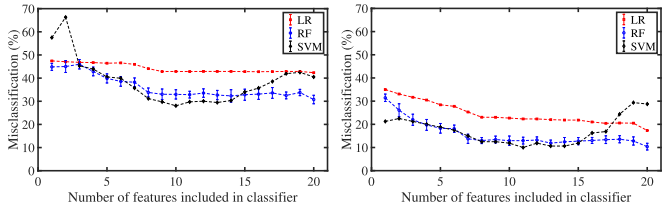

Fig. 7. Comparison of the classifier performance for varying number of features on two deep well data sets for (a) inter-handpump and (b) intra-handpump condition monitoring systems.

IV. Discussion

The current work demonstrates that a distributed inference condition monitoring system for rural infrastructure offers a number of advantages over existing condition monitoring system that are both energy and bandwidth heavy; however, further exploration of dynamic threshold variation of the on-board novelty filter is still required.

A. Summary of Main Points

Existing condition monitoring systems are not suitable for monitoring rural infrastructure that often operate in harsh environments and with constraints on data-transmission and battery life. We have described an appropriate set of labels that can be used for handpump condition monitoring (Table II) based on the recent definition of functionality [23]. We have shown that low-cost, lightweight machine learning methods can be implemented on-board with minimum bandwidth and battery requirements to apply novelty filtering (Fig. 4 and Table III). Furthermore, incorporating more advanced condition monitoring methods on a cloud-based platform have been shown to increase the system's overall positive predictive value by at least 10 per cent when "intelligent" subsets of flagged data from the rural node is transmitted (Fig. 5 and Table IV).

We defined a novel remote condition monitoring system for rural infrastructure that optimizes the limited resources in these operating conditions while taking advantage of the predictive power of advanced machine learning methods. Our algorithm resulted in a trade-off between increasing prediction accuracy and reducing prediction specificity. However, our method greatly improves the remote monitoring of rural infrastructure that is currently limited to usage.

We concluded that distributed inference using logistic regression on-board the rural node followed by random forests at the cloud-based server provided the best performing condition monitoring system for rural infrastructure while optimizing limited resources. We found that the combination of LR and RF provides the optimal prediction run time and can both be successfully implemented with less than half the number of features to be transmitted to the cloud-based platform.

B. Advantage of the Proposed Method

Dynamically combining a lightweight on-board and heavy-weight cloud-based system not only increases the specificity of the novelty detection, but more importantly, optimizes the allocation of limited computational resources by strategically scheduling new measurements and data transmission. This is important because the need for low-cost sensing in these rural applications can result in low signal-to-noise ratio of the processing hardware which severely compromises the performance of machine learning methods fitted to this type of data.

The system design is proposed in a way that can easily make it generalizable to other rural infrastructure types, such as off-grid home solar systems or agriculture monitoring systems. Furthermore, the proposed system provides a framework to tune both the type and the size of data to be transmitted from the node to the cloud. The cloud has the ability to proactively request more data, but also more detailed data, such as requesting feature vectors or raw accelerometer data rather than novelty scores. Our current experiments are limited to one-way communication; however, since the proposed system can perform interactive two-way communication this can easily be implemented in real-time.

C. Challenges and Future Work

There are two critical components required to perform predictive maintenance: (1) a health monitoring system that flags imminent failure, and (2) a service delivery model that performs the timely maintenance once the alert has been triggered. We have presented a condition monitoring system without clear guidelines for the alert protocols, since these should be defined by the integrated service delivery model rather than the technology itself.

Furthermore, a "plug-and-play" cloud-based platform would enable the translation of the system to other handpump or infrastructure types as well as other geographical regions. However, expanding this system to other infrastructure would require large amounts of time and data to train the on-board novelty filter. Future work should investigate methods that can reduce training time and allow transfer learning.

V. Conclusion

The present work describes the development of a distributed inference condition monitoring system for rural infrastructure. We use the example of failure detection of rural handpumps, the large-scale implementation of which could have significant health impacts for rural communities who rely solely on handpumps for their daily water supply [16].

The interaction between the on-board LR novelty filter and cloud-based RF processing is a novel research field, which is applicable to other cognate fields. The model was trained with a subset of general and depth-specific data and verified using another independent subset of data from the same study region. The results showed that it is possible to use lightweight machine learning methods to classify the condition of a rural node in situ with limited resources. It also showed that subsets of flagged data summaries from the node can be processed at a cloud-based platform to further improve prediction fidelity. This work marks an important milestone in the application of machine learning in rural condition monitoring as a tool for tracking rural development. However, further validation of the

proposed method will follow the collection and analysis of a long-term large-scale data set, as well as further exploring other time- and frequency-domain features.

Data Access Statement

This manuscript is in compliance with the UK Research Councils Common Principles on Research Data Policy, and some of the data collected and used in this research has been made available to the public. Remote collection of the data means that it is completely anonymous.

Acknowledgment

The authors would like to thank Suleiman Mwakurya, Idd Mwaropia and James Okoti of the RFL/FundiFix team in Bomani, Kenya for their support during fieldwork. As well as Chris Cormency, Peter Harvey, and Jessica Tribe from UNICEF.

References

[1] W. F. Fox and S. Porca, "Investing in rural infrastructure," *Int. Regional Sci. Rev.*, vol. 24, no. 1, pp. 103–133, 2001.
[2] S. G. Furey, "RWSN handpump survey 2013: Summary of findings," SKAT, St. Gallen, Switzerland, Tech. Rep., 2013.
[3] E. Baumann, "May-day! May-day! Our handpumps are not working!" RWSN Perspect., St. Gallen, Switzerland, Tech. Rep. RWSN/A/200, 2009, vol. 1.
[4] D. A. Clifton and L. Tarassenko, "Novelty detection in jet engine vibration spectra," *Int. J. Condition Monit.*, vol. 5, no. 2, pp. 2–7, 2015.
[5] L. A. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson, and L. Tarassenko, "Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 722–730, May 2014.
[6] V. Mehra, R. Ram, and C. Vergara, "A novel application of machine learning techniques for activity-based load disaggregation in rural off-grid, isolated solar systems," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2016, pp. 372–378.
[7] P. Thomson, R. Hope, and T. Foster, "Is silence golden? Of mobiles, monitoring, and rural water supplies," *Waterlines*, vol. 31, no. 4, pp. 280–292, 2012.
[8] F. E. Colchester, H. Greeff, P. Thomson, R. Hope, and D. A. Clifton, "Smart handpumps: A preliminary data analysis," in *Proc. Appropriate Healthcare Technol. Low Resource Settings (AHT)*, Sep. 2014, pp. 1–4.
[9] C. Nagel, J. Beach, C. Iribagiza, and E. A. Thomas, "Evaluating cellular instrumentation on rural handpumps to improve service delivery—A longitudinal study in rural rwanda," *Environ. Sci. Technol.*, vol. 49, no. 24, pp. 14292–14300, 2015.
[10] E. A. Thomas *et al.*, "Remotely accessible instrumented monitoring of global development programs: Technology development and validation," *Sustainability*, vol. 5, no. 8, pp. 3288–3301, 2013.
[11] P. Thomson, R. Hope, and T. Foster, "GSM-enabled remote monitoring of rural handpumps: A proof-of-concept study," *J. Hydroinform.*, vol. 14, no. 4, pp. 829–839, Oct. 2012.
[12] H. H. W. J. Bosman, A. Liotta, G. Iacca, and H. J. Wörtche, "Anomaly detection in sensor systems using lightweight machine learning," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 7–13.
[13] A. G. Logan *et al.*, "Mobile phone–based remote patient monitoring system for management of hypertension in diabetic patients," *Amer. J. Hypertension*, vol. 20, no. 9, pp. 942–948, 2007.
[14] K. Hung and Y.-T. Zhang, "Implementation of a WAP-based telemedicine system for patient monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 2, pp. 101–107, Jun. 2003.
[15] D. L. Wilson, J. R. Coyle, and E. A. Thomas, "Ensemble machine learning and forecasting can achieve 99% uptime for rural handpumps," *PLoS ONE*, vol. 12, no. 11, p. e0188808, 2017.
[16] E. A. Thomas, "Improving global impact: How the integration of remotely reporting sensors in water projects may demonstrate and enhance positive change," Global Water Forum, Sydney, NSW, Australia, Tech. Rep. 12, 2013, GWF Discussion Paper 1349.
[17] K. Fogelberg, "Filling the knowledge gap: Monitoring post-construction water and sanitation sustainability," *Waterlines*, vol. 29, no. 3, pp. 220–235, 2010.
[18] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 370, no. 1958, pp. 100–117, 2012.
[19] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.
[20] S. Funiak, C. Guestrin, M. Paskin, and R. Sukthankar, "Distributed inference in dynamical systems," in *Proc. NIPS*, 2006, pp. 1–8.
[21] M. van den Broek and J. Brown, "Blueprint for breakdown? Community based management of rural groundwater in uganda," *Geoforum*, vol. 67, pp. 51–63, Dec. 2015.
[22] M. Fisher *et al.*, "Understanding handpump sustainability: Determinants of rural water source functionality in the Greater Afram Plains region of Ghana," *Water Resour. Res.*, vol. 51, no. 10, pp. 8431–8449, 2015.
[23] R. C. Carter and I. Ross, "Beyond 'functionality' of handpump-supplied rural water services in developing countries," *Waterlines*, vol. 35, no. 1, pp. 94–110, 2016.
[24] T. Papastylianou *et al.*, "Smart handpumps: Technical aspects of a one-year field trial in rural kenya," in *Proc. Appropriate Healthcare Technol. Low Resource Settings (AHT)*, no. 1, Sep. 2014, pp. 1–4.
[25] E. Baumann and J. Keen, "Afridev handpump specification," SKAT-RWSN, RWSN/SKAT, St. Gallen, Switzerland, Tech. Rep., 2007.
[26] S. W. Smith, "Moving average filters," in *The Scientist & Engineer's Guide to Digital Signal Processing*, 2nd ed. Poway, CA, USA: California Technical Publishing, 1999, pp. 277–284.
[27] R. G. Lyons, *Understanding Digital Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2011.
[28] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.
[29] A. Testa, D. Gallo, and R. Langella, "On the processing of harmonics and interharmonics: Using Hanning window in standard framework," *IEEE Trans. Power Del.*, vol. 19, no. 1, pp. 28–34, Jan. 2004.
[30] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
[31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[32] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 431–439.

Heloise Greeff, photograph and biography not available at the time of publication.

Achut Manandhar, photograph and biography not available at the time of publication.

Patrick Thomson, photograph and biography not available at the time of publication.

Robert Hope, photograph and biography not available at the time of publication.

David A. Clifton, photograph and biography not available at the time of publication.