Digital Monitoring for Health Facility Emission Prediction and Climate Finance: A Case Study from the Democratic Republic of the Congo

Daniel Steven Doxazo Sitompul[1], Samuel B. Miles[1], Rashmi Varma[1], Sylvia Imanirakiza[2] , Sanuratu Koroma[2], Jackson Mughuma[3], Ryan McCord[4] Olufolahan Osunmuyiwa[5], Janvier Kamundala[6], Ngalula Sandrine Mubenga[6,7], Jay Taneja[2], Daniel M. Kammen[1,8], Laura H Kwong[1]
[1]University of California, Berkeley, [2]University of Massachusetts, Amherst, [3]Research Center for Humanitarian Aid, [4]Duke University, [5]nLine, Inc., [6]DRC Electricity Regulatory Authority (ARE), [7]University of Toledo, [8]Johns Hopkins University,
daniel_sitompul@berkeley.edu, samuel.b.miles@berkeley.edu, rashmi.varma@berkeley.edu, simanirakiza@umass.edu, skoroma@umass.edu, jacksonmugh@rcha-rdc.org, ryan.mccord@duke.edu, olu@nline.io, j.kamundala@are.gouv.cd, ngalula.mubenga@utoledo.edu, jtaneja@umass.edu, kammen@jhu.edu, lakwong@berkeley.edu

2025 IEEE Global Humanitarian Technology Conference (GHTC) | 979-8-3315-9502-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/GHTC66843.2025.11265999

Abstract*— Reliable electricity access remains a fundamental challenge for health facilities in the Democratic Republic of the Congo (DRC), where many depend on diesel generators to sustain clinical operations. These generators emit substantial quantities of carbon dioxide (CO_2), nitrogen oxides (NO_x), and fine particulate matter smaller than 2.5 microns ($PM_{2.5}$), posing significant risks to both climate and health. This paper presents a digital monitoring and auditing methodology that integrates voltage and frequency sensing, machine learning (ML), and Internet of Things (IoT) technologies to quantify generator runtime, emissions, and carbon finance potential. Empirical data from one hospital is used to develop and validate generator detection models based on random forest (RF) and extreme gradient boosting (XGBoost), achieving over 99% accuracy through feature engineering and ML. Estimated annual emissions for 2023 totaled 113 tonnes of carbon dioxide equivalent (tCO_2e), corresponding to potential carbon credit revenues of $900–$7,900, depending on market pricing. The approach is further demonstrated across three additional facilities, illustrating scalability and generalizability. These findings outline a data-driven solution for supporting cleaner, more resilient health facility energy systems in resource-constrained settings.

***Keywords*— Digital monitoring; machine learning (ML); emissions modeling; Internet of Things (IoT); diesel generators; climate finance**

I. Introduction

The Democratic Republic of Congo (DRC) faces persistent energy challenges, particularly in health facilities, where inadequate grid power necessitates the widespread use of distributed power systems [1]. As in many low- and middle-income country (LMIC) settings, the DRC's national electricity grid suffers from low coverage, frequent outages, and voltage instability, prompting health facilities to rely on diesel generators to maintain critical operations [2].

While generators provide critical energy resilience, their environmental and health impacts are significant, emitting pollutants such as carbon dioxide (CO_2), nitrogen oxides (NO_x), and fine particulate matter smaller than 2.5 microns ($PM_{2.5}$) at much higher rates than the grid. In the DRC, generators release 55 times more CO_2 and 9,000 times more $PM_{2.5}$ than the grid [3]. These emissions contribute directly to air pollution, the fifth-leading risk factor for mortality worldwide, responsible for over one million deaths annually across Africa, nearly two billion lost intelligence quotient (IQ) points in African children annually, and economic costs totaling 1–2% of national gross domestic product (GDP) per African country [4], [5]. Many of these emissions also drive climate change. Despite the well-established link between diesel generators, greenhouse gas emissions, and the negative impacts of climate change on health, less than 5% of climate finance—capital allocations dedicated to adaptation to or mitigation of climate impacts—was targeted toward health-related activities between 2009 and 2019 [6]. Voluntary carbon markets (VCMs) are emerging as critical mechanisms for filling gaps left by public commitments, acting as a bridge between grant funding and sustainable finance. By traded volume, VCMs grew by 30% from 2016 to 2021, reaching a current market valuation near $2 billion. Projections estimate a market size of $5–50 billion by 2030, with present carbon credit prices ranging from $12–35 per tonne [7]. In these markets, digital monitoring, reporting, and verification (dMRV) methodologies have become the gold standard in response to scrutiny over the accuracy of claimed, verified, and sold climate targets, particularly in forest conservation and improved cookstove projects that dominate global mitigation and adaptation efforts [8], [9], [10].

Building on recent advances in artificial intelligence (AI), big data, and the Internet of Things (IoT) at the energy–climate–health nexus, we develop a digital monitoring methodology that employs machine learning (ML) to quantify generator power and usage hours at healthcare facilities using voltage and frequency data. ML predictions are then used to estimate annual generator emissions of CO_2, $PM_{2.5}$, NO_x, and hydrocarbons (HC) and calculate the climate finance potential based on historical and plausible voluntary carbon market tonne-of-CO_2 (tCO_2) valuations. The monitoring methodology is primarily designed for healthcare facilities but is extensible to tracking carbon emissions benefits for carbon market and health efficacy studies. These methods could further assist projects aiming to reduce hospitals' reliance on polluting and costly diesel generators by linking emissions to disability-adjusted life years (DALYs) and other health impact metrics in future work. Specifically, our research questions are as follows: 1) With what accuracy, precision, and recall can ML detect generator usage based on voltage and frequency data? 2) Based on estimated generator use, what are the annual greenhouse gas emissions from generator use at selected healthcare facilities in the DRC? 3) How much could a solarization project that displaces generator emissions at such hospitals earn through voluntary carbon markets?

II. Methodology

We installed five sensors in the facility, two power quality sensors (nLine "PowerWatch"[1] sensors) at outlets in the laboratory and three single-phase power energy sensor (A2EI "HopMeters"[2]) in several facilities as shown in Figure 1. PowerWatch sensors sample voltage and frequency data, every two minutes, calculating outages in tandem, while HOPmeter sensors are used to measure power draw, at the equipment, ward, or whole facility level. Both types of sensors relay data via integrated SIM cards to cloud data warehouses accessible through commercially available data platforms with a real-time lag of approximately 5-10 minutes.

A. Study Sites

This study uses data from four health facilities in North Kivu Province, DRC, each with different energy setups. Our main focus is Centre Hospitalier (CH) Bethesda Ndosho—the only site with complete records on generator fuel use, power, and emissions. The other three sites—HGR Katwa, HGR Kitatumba, and HGR Walikale—are used to test how well our machine learning (ML) model works in different settings.

At CH Bethesda Ndosho, two backup generators (80 and 30 KVA) are used alternately (Figure 1). Technicians keep a logbook of generator use, which we accessed for January to April 2023. These logs help validate our ML detection method and assess generator efficiency.

Although installing a power or current meter on the generator may seem like a simple way to measure usage, it doesn't work well in places like Ndosho where the power system includes multiple sources—like grid, solar, batteries, and generators—all connected to the same circuit. In such setups, the generator might be running but not actually supplying power. That's why we use our ML approach to more accurately detect when the generator is truly in use and estimate its emissions.

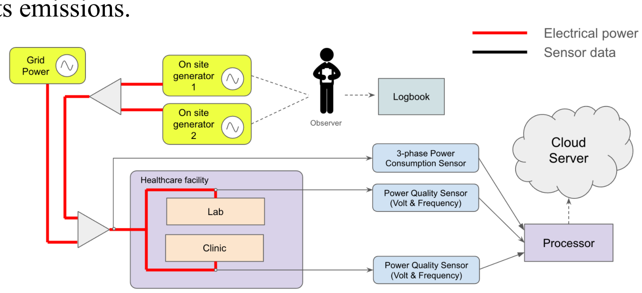
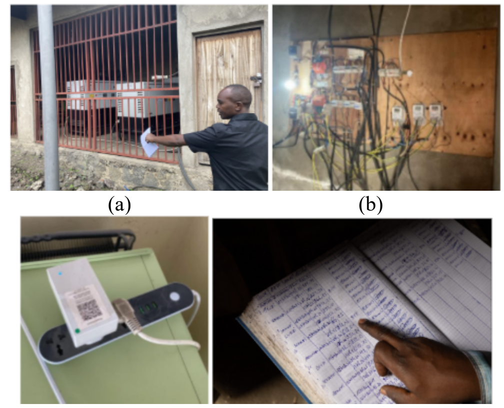


Figure 1. Electrical System Monitoring and Data Flow Diagram

To illustrate the field implementation, Figure 2 shows the generator, power sensors, voltage and frequency sensors, and the technician's logbook. The images were taken at CH Bethesda Ndosho in Goma, which serves as our primary observation site.



(a) (b) (c) (d)

Figure 2. a) Generators assessed ; b) A2EI power consumption sensors; c) nLine power quality sensors; d) generator and grid expenditure log books at a large hospital in the DRC.

B. Machine Learning Detection Model

To estimate generator emissions, we collected voltage and frequency data from the power source and developed a ML detection method to determine generator usage during consecutive 2-minute intervals. We iteratively refined the model's accuracy through feature engineering and validation. Using the detection results, we then calculated the generator's energy consumption, corresponding emissions, and associated financial value over time (see Figure 3). The model for CH Bethesda Ndosho was trained on data collected during two distinct periods—1 February to 28 February 2023 and 12 August to 9 September 2024—to capture both operational and seasonal variations.

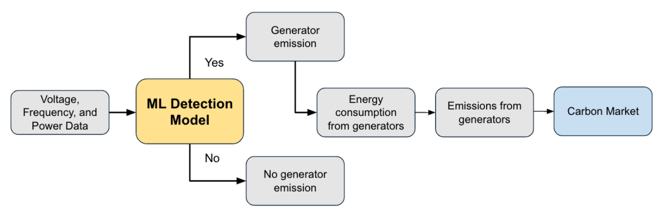


Figure 3. Emission and Carbon Market Calculation

1) Labeling the Data

We manually labeled generator usage during the training period based on visual inspection of voltage and frequency signals, where stable and high voltage indicate generator operation. This was supplemented with contextual knowledge obtained through site visits, facility expenditure logbooks, surveys, and correspondence with facility technicians and medical staff.

2) Feature Engineering

We trained and evaluated our model using four distinct feature groups. Feature group 1 consists of the raw voltage and frequency data collected from sensors in both the clinic and laboratory, yielding four features. Feature group 2 builds upon the first by adding indicator features that flag whether each

[1] https://nline.io/docs/powerwatch_spec_sheet.pdf

[2] https://a2ei.org/resources/uploads/2021/03/A2EI-Smart-Meter.pdf

voltage or frequency value exceeds the mean computed over the 5.5-week training period, expanding the feature set to eight. Feature group 3 further extends the features by incorporating the hour of the day and the first-order differences of voltage and frequency between consecutive timestamps, resulting in 50 features. Feature group 4 further extends the feature by adding temporal patterns using Hidden Markov Model (HMM) with ten states to the raw voltage and frequency data, as illustrated in Figure 4.

HMMs are well-established tools for modeling time series data characterized by state-driven processes, effectively capturing temporal dependencies and identifying distinct regimes or patterns within the signal [13], [14]. In this application, the HMM detects underlying behaviors in voltage and frequency signals, such as stable operating periods, transitions, and anomalies. Each timestamp is consequently assigned a state based on the HMM predictions, thereby categorizing the signal into discrete patterns that represent system behavior.

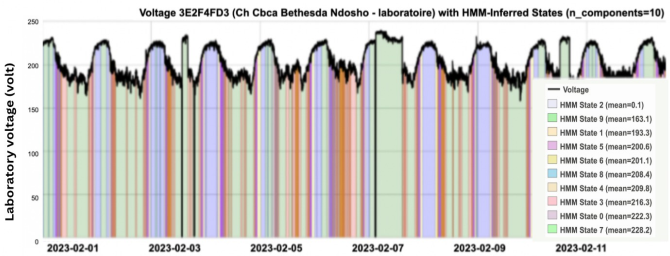


Figure 4. Feature generation by HMM using ten states

3) Modeling

We applied the baseline model and three different ML models (logistic regression, random forest (RF), and XGBoost) for each feature group. The baseline model was a non-ML technique, classifying generator use based solely on the most common state observed in the training data. This allowed us to compare the performance of ML models against a simple heuristic approach. The three ML models were chosen for their unique strengths in handling classification problems. Logistic regression is a simple, interpretable model well-suited for understanding linear relationships in the data. It serves as a benchmark for comparing more complex models. RF is an ensemble method that handles nonlinear relationships and is less prone to overfitting due to its use of multiple decision trees. XGBoost is a powerful gradient boosting method that improves accuracy by learning from previous mistakes, and it performs well on structured data. Furthermore, we performed hyperparameter tuning to increase prediction accuracy, precision, and recall. These models differ in their algorithmic approaches and computational complexity, which allows us to evaluate their relative strengths and trade-offs in predicting generator usage.

4) Model Assessment

For each model and feature group, we evaluated performance using standard classification metrics: accuracy, precision, and recall. The model with the highest scores across these metrics was selected for generator usage prediction. To validate model outputs, we compared the predicted generator operating hours for January, March, and April 2023 against technician logbook records and calculated the absolute difference between predicted and logged generator usage. Model generalizability was assessed by applying the trained models to additional facilities beyond CH Bethesda Ndosho. For each site, we used one month of data, with a 60/40 train-test split, to evaluate how each model and feature group performed across different settings in terms of accuracy, precision, and recall.

C. Model Generalization: Leave-One-Site-Out Cross-Validation (LOSO-CV)

To evaluate the robustness and generalizability of our ML models, we performed a comprehensive leave-one-site-out cross-validation (LOSO-CV) analysis. This shows how well models transfer across unseen sites. We systematically train the RF and XGBoost classifiers on data from all sites except one, then evaluate performance on the held-out site. This process is repeated iteratively, with each facility serving as the test set exactly once. For each training iteration, we perform independent hyperparameter tuning using cross-validation on the training sites to ensure optimal model configuration for each fold. We also compared LOSO-CV results against site-specific models to quantify the trade-offs between generalizability and site-specific performance. Recall the facility-specific models are trained and validated solely on data from individual facilities using a 60/40 train-test split. The cross-validation also ensures that the test data is 40% of the test facility in a particular fold for a fair comparison. Additionally, we compared the model performance using two particular metrics for this analysis: F1 Score (the harmonic mean of precision and recall) and the generator operating time detection error (the percentage absolute difference between predicted and actual generator operating time in minutes).

D. Power Consumption and Energy Usage Calculations

To calculate total energy usage, we utilized data from nLine "PowerWatch" sensors. Due to limitations in real-time data availability throughout 2023, power data is available for only one facility: CH Bethesda Ndosho. Using the average hourly power consumption, we applied our workflow to calculate the total generator runtime and estimate the corresponding fuel consumption, continuous power output in kilowatts ($kW_{continuous}$), and total energy in kilowatt-hours (kWh) during the identified generator runtime. The total energy consumption is estimated by multiplying the average power load at each minute *i* (M_i) of generator usage by the duration of generator operation in minutes, as shown in equation (1).

$$T_{total} = \Sigma_{i=1}^{n} \left(\frac{M_i}{60} \times P_{avg,i} \right) \quad (1)$$

E. Estimated Emissions and Potential Climate Financing

We seek to examine the potential amount of climate financing that could be available to health facilities from eliminating generator use. To estimate the CH Bethesda Ndosho hospital's generator CO_2, NO_x, and $PM_{2.5}$ emissions in 2023, we multiply the generator-produced energy consumption (T_{total}) by the U.S. Environmental Protection Agency's gaseous pollutant-specific emission factors (EF_i) for large stationary diesel equipment [15]. To estimate the potential climate financing associated with eliminating these emissions, we first calculated the CO_2 equivalent (CO_2e) of the generator emissions by multiplying the mass of each emitted pollutant by its global warming potential (GWP) over a 100-year timeframe (in accordance with IPCC AR5) [16]. Thus, the total CO_2e for the CH Bethesda Ndosho generator was calculated as the sum of CO_2e from CO_2 emissions and CO_2e from NOx emissions. Finally, we multiplied the resulting CO_2e for two carbon price scenarios (8 and 70 USD/tonne) to estimate the potential climate financing the project could generate.

III. Results and Discussion

A. Detection Model

1) Detection Model for CH Bethesda Ndosho

As shown in Table 1, the RF and XGBoost models performed the best when we used the mean threshold feature and time-series (feature group 3). The time series feature increases the accuracy for all RF and XGBoost models to more than 99% with precision above 99% and recall above 96% compared to feature group 2. Adding the HMM increased the logistic regression model accuracy, precision and recall.

Table 1. Accuracy, precision, and recall of various models detecting whether a diesel generator is on or off, based on different features on the test set. Bold indicates the models with the best performance.

Model	Feature Group	Accuracy (%)	Precision (%)	Recall (%)
Baseline	1	81.66	0	0
Logistic Regression		96.80	95.28	86.86
Random Forest		96.37	94.95	84.73
XGBoost		96.64	95.27	85.96
Logistic Regression	2	78.4	38.59	99.88
Random Forest		91.18	94.95	54.81
XGBoost		96.64	95.27	85.96
Logistic Regression	3	96.98	98.53	84.81
Random Forest		99.33	99.69	96.69
XGBoost		99.16	99.43	95.94
XGBoost (hyperparameter tuned)		99.39	99.51	97.20
Logistic Regression	4	98.37	99.86	91.22
Random Forest		98.7	99.87	93.14
XGBoost (tuned hyperparameter)		99.19	99.68	95.91

Our model predicts generator usage when voltage is relatively high and stable. The RF model with feature group 3, estimated that the generator was used for 113,200 minutes (~1,886.7 hours) during 2023. In this model, voltage in the clinic emerges as the most significant contributor to the model's performance followed by voltage in the laboratory and frequency in the laboratory and clinic. In accordance with the training data, the model predicted that the generator was on when the voltage was relatively high and stable, as shown in Figure 5.

2) Validation with Logged Data from CH Bethesda Ndosho

We compared the model's predictions to the technician's logbook entries (Table 2) to evaluate accuracy, using the logbook as the ground truth for generator usage during the analysis period. The model underestimated generator usage by only 2.43%, demonstrating high predictive accuracy. Although the degree of over- or underestimation varied slightly across months, the model's error remained within ±6.05% throughout the evaluation period.

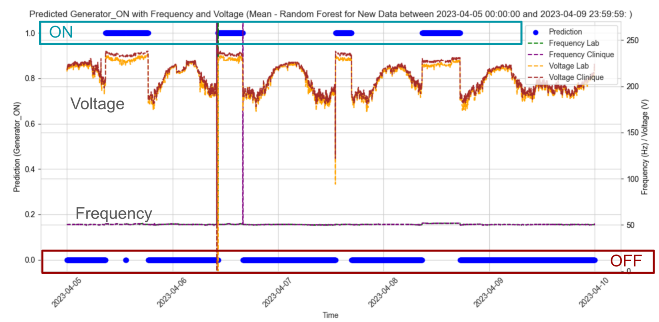


Figure 5. Voltage, frequency, and predicted generator usage at CH Bethesda Ndosho between 5-9 April 2023

We then validated the predicted generator power output from the ML model against actual fuel usage, as summarized in Table 3. Over the three-month period, we observed an average fuel-to-power conversion efficiency of approximately 23.4%. Month-to-month variations in efficiency may be attributed to factors such as fluctuating electrical loads (with lower efficiency under partial loading conditions), differences in generator power ratings, and frequent restarts, all of which can reduce overall performance.

Table 2. Generator time of utilization validation result

Time Period	Technician-logged (minutes)	Predicted (minutes)	Model overestimate (%)
January	3,159	2,968	-6.05%
March	5,158	5,010	-2.87%
April	12,030	11,874	-1.30%
Total	**20,347**	**19,852**	**-2.43%**

Table 3. Power usage validation result and generator fuel-to-power efficiency

Time Period	Logbook - fuel (kg)	Logbook - fuel (kWh)	Predicted (kWh)	Efficiency (%)
January	210	2,507.4	570.33	22.7
March	378	4,513.32	911.07	20.2
April	882	10,531.08	2,631.83	25
Total	**1,470**	**17551.8**	**4,113.23**	**23.4**

3) Applying Detection Models for additional facilities

We evaluated the reliability of our model in three additional facilities: HGR Walikale, HGR Kitatumba, and HGR Katwa (Table 3). These sites were selected because they provided a combination of reliable sensor data and operational diversity, offering an opportunity to test the model's performance across varying facility conditions. For each location, we trained the model using the same feature set—Voltage and Frequency, Mean Threshold, Time Series, and additional HMM features—and tested it to the site it was trained on.

The results demonstrate strong model performance across all sites (Table 3). For CH Bethesda Ndosho, the RF model achieved an accuracy of 98.7%, precision of 99.87%, and recall of 93.14%. At the Walikale facility, all models performed well with high accuracy and precision. The Kitatumba facility yielded near-perfect results, indicating excellent detection capability. In contrast, results for Katwa show slightly lower performance compared to other sites, which could be attributed to site-specific factors, such as signal noise, inconsistent generator use patterns, or fewer training samples.

Table 3. Assessment of feature group 4 to three other facilities

Facility	Model	Accuracy (%)	Precision (%)	Recall (%)
Walikale	Logistic Regression	99.29	99.86	93.12
	Random Forest	99.9	100	99.02
	XGBoost	99.91	100	99.14
Kitatumba	Logistic Regression	99.89	99.67	99.90
	Random Forest	100	100	100
	XGBoost	100	100	100
Katwa	Logistic Regression	96.97	87.69	83.66
	Random Forest	97.57	88.54	88.95
	XGBoost	96.98	84.02	88.95

4) Cross-validation within different facilities

Generator Operating Time Detection Error Analysis

The leave-one-out cross validation showed diversity in performance across different held out sites with Ndosho showing the best performance with a low detection error of 2.5% when evaluated using the XGBoost + feature group 4 model configuration. Additionally, across the other feature groups and for both the XGBoost and RF, the detection error for Ndosho was consistently below 25% as shown in Figure 6.

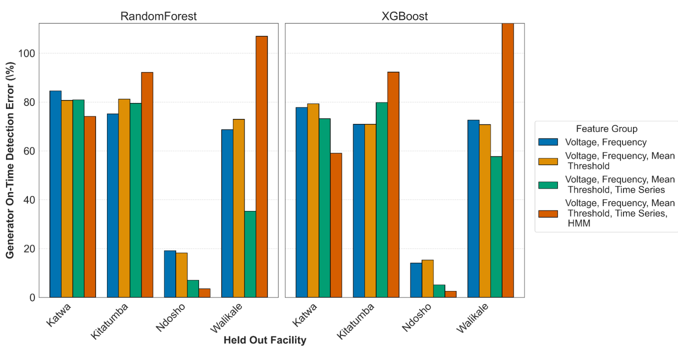


Figure 6. LOSO-CV: Absolute generator operating time detection error by held-out facility and feature group, shown for the RF and XGBoost models. Each bar represents the percentage error of the model in estimating generator operating time at a given facility using that particular feature group.

This performance is comparable to the site specific models. This suggests that while information from the Ndosho data may not generalize well to other sites, other facilities' data can effectively predict Ndosho generator behavior. Another plausible reason is the considerably larger size of the dataset from the Ndosho site in comparison to other sites. Notably, Kitatumba and Katwa showed consistently high detection errors above 75% across most feature combinations, indicating these sites may have unique operational characteristics that are difficult to predict using cross-site training data.

Site-Specific Operational Patterns

Figure 7 shows the significant F1 Score performance variations across held-out sites, suggesting distinct operational signatures at each facility. Ndosho's strong performance as a target site (low F1 when used for training others, high F1 when others train on it) indicates relatively predictable operational patterns that can be learned from diverse training data. In contrast, Kitatumba's poor performance in both directions suggests highly site-specific operational characteristics that neither generalize to other sites nor benefit significantly from cross-site knowledge transfer. The differential impact of time series features and HMM features across sites further supports this hypothesis, with temporal and latent patterns appearing more transferable for some facilities (Ndosho and Katwa) while potentially introducing noise for others (Kitatumba and Walikale).

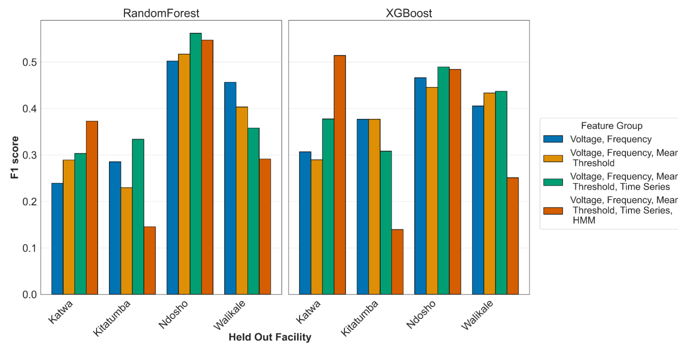


Figure 7. LOSO-CV: F1 score performance by facility and feature group across selected models. Each subplot corresponds to a different feature group, while bars within each subplot represent the F1 scores achieved on different held-out facilities by each model. This visualization illustrates the impact of feature selection and cross-site generalization on classification performance, with the F1 score summarizing the balance between precision and recall.

Feature Set Impact on Generalization

The baseline feature group comprising voltage and frequency only demonstrates relatively stable but modest performance across sites, particularly under the RF classifier, with F1 scores ranging from approximately 0.19 to 0.49 and correspondingly high detection error rates. The inclusion of mean threshold-based features generally enhances model performance across both RF and XGBoost, suggesting that aggregate statistics introduce informative patterns that are moderately transferable across operational contexts.

However, the addition of time series-derived features yields heterogeneous outcomes. While performance improves in certain cases, at Ndosho and Kitatumba, performance degrades substantially at others, such as Walikale. In some instances (Katwa and Ndosho under XGBoost), HMM features offer marginal improvements in both F1 score and generator detection accuracy. However, at Walikale and Kitatumba, HMM features consistently degrade performance across both classifiers, in some cases by more than 40% in F1 score or leading to an increase of over 30 percentage points in detection error. This suggests that temporal features and HMM-derived latent representations capture site-specific temporal dependencies that do not generalize well across heterogeneous facilities.

Feature Importance for Generalized Models

Considering the feature importance of the LOSO-CV analysis of XGBoost and RF in Figure 8 (a) and (b) respectively, aids in understanding the generalizability of the RF and XGboost models across sites. The feature importance of the models show clear interpretability and the trustworthiness of the model in performing for unseen sites. Understanding the way the models learn can aid in reducing noise, simplify the model, speed up training and lastly prevent overfitting. Find below a diagram that shows the most important features for the generalized XGBoost and RF across all facilities. Figure 8 shows that both models used are highly dependent on the data collected from sensors such as the voltage and frequency.

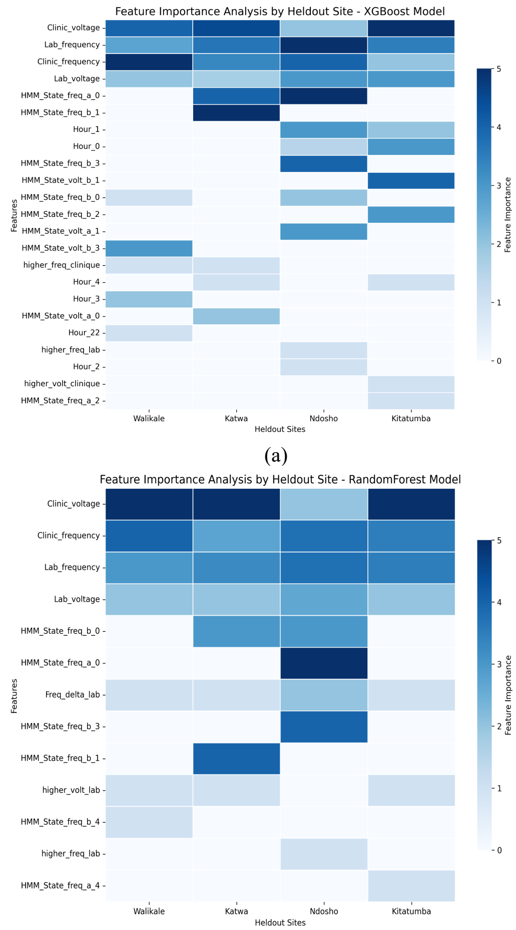


Figure 8. Feature importance by site for (a) XGBoost and (b) RF model

Implications for Facility-Specific vs Generalizable Models

The contrast between site-specific and LOSO-CV performance fundamentally presents a step towards deployment strategy considerations. The 4-10x performance degradation observed at some sites when using cross-site training suggests that there is a trade-off between model generalizability and performance. That is, site-specific models present far more precise generator operating time estimation (< 5% error). All sites demonstrate the ability to remain within an error bound of 0.5-21% error, but this requires dedicated models and training for each facility. In scenarios where site-specific models are impractical, Ndosho and Walikale emerge as potentially viable targets for cross-site models, with LOSO-CV errors achievable in the 5-15% range using XGBoost and RF respectively with appropriate feature selection.

Our LOSO-CV analysis informs our evaluation of the cross-site model transferability; however, the significant generalization penalties observed (up to 10x performance degradation) indicates avenues for future work. Overall, these findings bring to light the value of efficient feature engineering in cross-site generalization. We leave further exploration of generalizability across sites to future work, which could take advantage of expanded multi-site datasets, automatic feature extraction techniques beyond our handcrafted features, meta-learning and more sophisticated model architectures to bridge the current out-of-distribution and performance gaps.

B. CH Bethesda Ndosho 2023 power consumption

The average hourly load of the total facility was typically between 5-10 kWp during evening hours, but during the day consumption rises to 15-25 kWp, resulting in consumption of 200-300 kWh per day (Figure 9). We estimate that the facility used 22,295 kWh of energy from generators in 2023.

As such, we estimate that in 2023 generator use emitted 15.769 tonnes of CO_2, 326.247 kg of NOx, and 74.765 kg of CO, 109.972 kg of SOx, and 9.515 kg of $PM_{2.5}$ (a total of 113.066 tCO_2e). These emissions may pose substantial health risks, particularly in healthcare settings where sensitive populations—including pregnant women, infants, and immunocompromised patients—may be exposed.

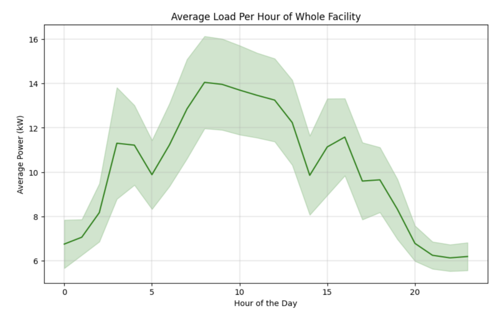


Figure 9. Hourly power consumption in CH Bethesda Ndosho. Data is averaged over approximately 6 months.

Diesel generators in sub-Saharan Africa contribute 15% of regional NO_x emissions and generate $PM_{2.5}$ volumes equivalent to 35% of vehicle emissions, releasing over 100 megatons of CO_2 annually [17]. Unlike dispersed vehicle traffic, hospital generators operate near inpatient wards, maternity units, and NICUs, creating concentrated exposure zones. Diesel-related $PM_{2.5}$ triggers acute asthma exacerbations and increases emergency visits among children and elderly patients [18]. A 10 $\mu g/m^3$ increase in $PM_{2.5}$ correlates with 9% higher infant mortality in sub-Saharan Africa, with 22% of infant deaths attributed to this exposure [19]. Prenatal exposure elevates risks of low birth weight and preterm birth at thresholds between 31–38 $\mu g/m^3$ [20], [21]. Despite these documented health impacts, less than 5% of global climate adaptation finance between 2009–2019 supported health systems [6], highlighting a critical funding gap that health-linked carbon credits could help address.

C. Potential climate financing for CH Bethesda Ndosho

Generator use at CH Bethesda Ndosho produced 113.066 tCO_2e in emissions. At a price of $8/$tCO_2e$—reflecting a conservative, low-end estimate typical of voluntary carbon market credits in LMICs [28]—these credits could generate \$904.53. In contrast, using \$70/tCO_2e, which is in line with recent EU Emissions Trading System (EU ETS) carbon prices for compliance markets [28], the credits would yield \$7,914.62. In 2023, the CH Bethesda Ndosho hospital used diesel generators to produce 22,295 kWh of energy at an average cost of \$0.086/kWh, totaling approximately \$1,917. A carbon credit project could offset these emissions at a cost of \$16.96/tCO_2e. We note that the low cost of electricity does not capture the true cost of using the generator, which would need to include maintenance and depreciation, increasing the per credit cost to offset generator usage.

Avoidance credits could be verified using our digital monitoring system, which distinguishes generator usage from solar or grid-based power signatures via voltage and frequency analysis. These offsets may be especially valuable in healthcare settings due to their co-benefits—improving air quality and reducing the burden of disease—alongside emissions reduction [7], [8], [9].

Projects that reduce emissions and deliver measurable health improvements are now commanding premium valuations in the voluntary carbon market. For instance, Gold Standard–certified credits—those verifying both climate and sustainable development impacts—can fetch prices 6–29% higher than conventional credits. A meta-analysis of recent transactions found that projects offering strong co-benefits averaged 30 % higher prices [10], [27].

Recent market data reinforce this trend. In the first half of 2025, more than 95 million carbon credits were retired—a record pace—with industrial and commercial project categories growing by 140% year-over-year [26]. Co-benefit-certified energy access projects, especially those integrating health outcomes, are trading at \$50–60/tCO_2e—well above the typical \$4–30/tCO_2e range [26], [27]. Based on our analysis, the generator displacement model aligns closely with these categories.

One of the key challenges for single-site health facilities entering carbon markets is the high cost of participation. However, we find that dMRV systems—like our IoT sensor and ML-based generator detection platform—offer a scalable, low-cost solution. Similar approaches have been used effectively in clean cookstove and mini-grid projects registered under Verra and Gold Standard [27]. Notably, fewer than 20% of current VCM projects are rated “A” or above in quality, indicating strong demand for reliable, well-monitored systems [26].

Beyond the voluntary carbon market, blended finance mechanisms that link climate and health outcomes are gaining momentum. The WHO–UNDP Climate and Health Co-Investment Facility, for instance, aims to mobilize \$120 million to support climate-resilient health systems [23]. Increasingly, results-based finance (RBF) models are being tied to measurable outcomes such as emissions reductions or DALYs averted.

In practical terms, electrification through solar-plus-battery systems could be deployed by hospital networks such as Médecins Sans Frontières (MSF), the International Committee of the Red Cross (ICRC), or the Baptist Community in Central Africa (CBCA). These institutions could measure and report reductions in CO_2, NO_x, and $PM_{2.5}$ while also documenting improved clinical outcomes, enabling the issuance of dual-benefit carbon credits. These are increasingly attractive to buyers and may offer a recurring revenue stream to support energy infrastructure or expand services.

Our ML model, trained on high-resolution voltage and frequency data, can detect generator usage with over 99% accuracy, 99% precision, and 96% recall—addressing ongoing concerns about transparency and verification in carbon accounting [11], [13]. While dMRV systems are already in use for cookstoves and electrification projects [4], [6], there is currently no standardized methodology for diesel displacement in health settings. Our work provides a foundational model that could support the development of such protocols.

Digital monitoring systems have advanced substantially in recent years. Through real-time IoT sensors and cloud-integrated platforms, we now have the ability to conduct high-frequency longitudinal monitoring of hospital energy use—enabling more rigorous impact measurement and faster intervention. These systems could and should be developed to also trigger alerts or corrective actions when anomalies are detected, enhancing operational continuity [11].

The main limitation of our study is the relatively small sample size, as data were collected from only four health facilities; substantially more data would be needed to enhance model robustness and generalizability. Expanding training datasets across more LMIC facilities will be crucial to improve the model's performance across diverse settings. There is also a need to standardize emissions modeling and integrate health outcome metrics—such as \$/DALY avoided—to support valuation frameworks. Industry bodies like Verra, Gold Standard, and Prospect should prioritize the development of diesel-specific methodologies for the health sector [5], [6].

To advance this work, we recommend that policymakers and implementers focus on two near-term priorities:

- Integrate high-accuracy ML models into dMRV systems to enable transparent, low-cost emissions tracking for climate finance [12];
- Elevate the health benefits of clean energy in financing mechanisms, particularly those targeting hospitals and other high-risk environments [2], [6].

IV. Conclusions

This study demonstrates that a data-driven approach combining voltage and frequency monitoring with ML models can reliably detect generator usage at health facilities in the DRC, achieving over 99% accuracy, 99% precision, and 96% recall. Based on our model, we calculated the total power consumption utilizing 3 months of validation data, and showing a generator efficiency of 23.4%. For our main facility CH Bethesda Ndosho, our 3 monthly validation data shows a maximum time underestimate of -6.05% in a month and average of -2.43%.

Our cross-validation results show that CH Bethesda Ndosho achieved the lowest detection error (<5%), while sites like Kitatumba showed high errors (>75%), highlighting the need for site-specific models. This underscores a trade-off between generalizability and accuracy across diverse facility operations. Future research could explore more sophisticated models, such as meta-learning or automated feature extraction, to improve cross-site performance.

Generator usage at CH Bethesda Ndosho was estimated to produce 113 tCO_2e emissions in 2023, translating into potential voluntary carbon market revenues ranging from $900 to $7,900 depending on credit prices. We estimated pollutants associated with generator usage at 326 kg of NO_x, 9.5 kg of $PM_{2.5}$, 110 kg of SO_x, and 75 kg of CO—potentially posing serious health risks in hospital settings where vulnerable groups like infants, pregnant women, and immunocompromised patients face heightened exposure. As a result, carbon credits with verified health co-benefits could command premium prices, offering hospitals a dual incentive: reducing fuel costs while unlocking climate finance to support clean energy transitions.

Scaling this methodology to additional facilities would help confirm the robustness of the detection model, though site-specific calibration may be needed due to varying operational patterns. By opening up new finance streams for cleaner energy transitions, this digital monitoring and verification approach offers a promising pathway for improving health outcomes and reducing climate impacts in low-resource settings.

Acknowledgments

We are grateful for the financial support of the CITRIS and the Banatao Institute, the Link Foundation, the Lao and Karsten Foundations, and the National Science Foundation under Grant No. DGE- 2125913. We gratefully acknowledge feedback from George Kibala Bauer and Zach White (GSMA) and the collaboration that made this work possible from Sejal Jinturkar, Charlie Huang, Bastian Herzog, Peter Flo Grinde-Hollevik, and Aashika Nair.

References

[1] M. Moner-Girona, G. Kakoulaki, G. Falchetta, D. J. Weiss, and N. Taylor, "Achieving universal electrification of rural healthcare facilities in sub-Saharan Africa with decentralized renewable energy technologies," *Joule*, vol. 5, no. 10, pp. 2687–2714, Oct. 2021, doi: 10.1016/j.joule.2021.09.010.

[2] S. Chawla et al., "Electricity and generator availability in LMIC hospitals: improving access to safe surgery," *J. Surg. Res.*, vol. 223, pp. 136–141, Mar. 2018, doi: 10.1016/j.jss.2017.10.016.

[3] D. Farquharson, P. Jaramillo, and C. Samaras, "Sustainability implications of electricity outages in sub-Saharan Africa," *Nature Sustainability*, vol. 1, no. 10, pp. 589–597, Oct. 2018, doi: 10.1038/s41893-018-0151-8.

[4] J. T. Lee and D. S. Callaway, "The cost of reliability in decentralized solar power systems in sub-Saharan Africa," *Nature Energy*, vol. 3, no. 11, pp. 960–968, Oct. 2018, doi: 10.1038/s41560-018-0240-y.

[5] S. Fisher et al., "Air pollution and development in Africa: impacts on health, the economy, and human capital," *Lancet Planetary Health*, vol. 5, no. 10, pp. e681–e688, Oct. 2021, doi: 10.1016/S2542-5196(21)00201-1.

[6] T. Alcayna, D. O'Donnell, and S. Chandaria, "How much bilateral and multilateral climate adaptation finance is targeting the health sector? A scoping review of official development assistance data between 2009–2019," *PLOS Global Public Health*, vol. 3, no. 6, p. e0001493, Jun. 2023, doi: 10.1371/journal.pgph.0001493.

[7] C. Blaufelder, C. Levy, P. Mannion, and D. Pinner, *A blueprint for scaling voluntary carbon markets to meet the climate challenge*, McKinsey and Company, 2021.

[8] A. Gill-Wiehl, D. M. Kammen, and B. K. Haya, "Pervasive over-crediting from cookstove offset methodologies," *Nature Sustainability*, vol. 7, no. 2, pp. 191–202, Jan. 2024, doi: 10.1038/s41893-023-01259-6.

[9] B. S. Probst et al., "Systematic assessment of the achieved emission reductions of carbon crediting projects," *Nature Communications*, vol. 15, no. 1, p. 9562, Nov. 2024, doi: 10.1038/s41467-024-53645-z.

[10] World Bank, *Digital Monitoring, Reporting, and Verification Systems and Their Application in Future Carbon Markets*, Washington, DC: World Bank, 2022.

[11] Z. Wang et al., "NOx emission prediction using a lightweight convolutional neural network for cleaner production in a down-fired boiler," *Journal of Cleaner Production*, vol. 389, p. 136060, Feb. 2023, doi: 10.1016/j.jclepro.2023.136060.

[12] S. Miles et al., "Internet of Things could shape healthcare facility electrification: Evidence from the Democratic Republic of the Congo," unpublished.

[13] M. Hanif, F. Sami, M. Hyder, and M. I. Iqbal, "Hidden Markov Model for Time Series Prediction," *Journal of Asian Scientific Research*, vol. 7, no. 5, pp. 196–205, 2017.

[14] B. Esmael, H. Abdulsalam, and H. El-Bakry, "Improving Time Series Classification Using Hidden Markov Models," ResearchGate, 2012.

[15] U.S. Environmental Protection Agency, "AP-42, Fifth Edition, Volume I, Chapter 3.4: Large Stationary Diesel and All Stationary Dual-Fuel Engines," Research Triangle Park, NC: U.S. EPA, 1995.

[16] T.F. Stocker et al., Eds., *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK: Cambridge University Press, 2013. [Online]. Available: https://www.ipcc.ch/report/ar5/wg1/

[17] International Finance Corporation, *The Dirty Footprint of the Broken Grid: The Impacts of Fossil Fuel Backup Generators in Developing Countries*, 2021.

[18] M. Park, M.-P. Kwan, Y. Choi, and S.-S. Hwang, "Short-term exposure to $PM_{2.5}$ and emergency hospital visits for asthma: A time-series study," *Environmental Health Perspectives*, 2020.

[19] S. Heft-Neal, J. Burney, E. Bendavid, and M. Burke, "Robust relationship between air quality and infant mortality in Africa," *Nature*, vol. 559, pp. 254–258, Jul. 2018.

[20] X. Zhu, Z. Wang, X. Zhang, and X. Xu, "Prenatal exposure to $PM_{2.5}$ and low birth weight: A global meta-analysis," *Toxics*, vol. 12, no. 1, Jan. 2024, doi: 10.3390/toxics12010013.

[21] N. Parasin, P. Chuenchom, and T. Noree, "Association between $PM_{2.5}$ exposure and preterm birth in Northern Thailand: A large cohort study," *Environmental Research*, vol. 226, p. 115572, 2023, doi: 10.1016/j.envres.2023.115572.

[22] E. Narassimhan, K. S. Gallagher, S. Koester, and J. R. Alejo, "Carbon pricing in practice: A review of existing emissions trading systems," *Climate Policy*, vol. 18, no. 8, pp. 967–991, 2018, doi: 10.1080/14693062.2018.1467827.

[23] World Health Organization and United Nations Development Program, *Climate and Health Co-Investment Facility: Driving financing to support low-carbon, resilient health systems*, WHO, 2023.

[24] Gold Standard Foundation, "Why co-benefits matter: Announcing Gold Standard for the Global Goals," Gold Standard, Jul. 2017.

[25] S. Piris-Cabezas, M. Lubell, S. H. Schneider, and M. Grasso, "Co-benefits of carbon markets: Pricing health and sustainability," *Nature Sustainability*, vol. 6, pp. 393–400, Apr. 2023, doi: 10.1038/s41893-023-01156-y.

[26] Minnesota Pollution Control Agency, "Diesel generator wet stacking," 2022.

[27] Sylvera, *Voluntary Carbon Market Data Q2–Q3 2025*, London: Sylvera Ltd., 2025.

[28] World Bank. 2023. State and Trends of Carbon Pricing 2023. © World Bank. http://hdl.handle.net/10986/39796 License: CC BY 3.0 IGO.