Bryan J. Stringham
Department of Mechanical Engineering,
Brigham Young University,
Provo, UT 84602
e-mail: bryan.stringham@byu.edu

Daniel O. Smith
Department of Mechanical Engineering,
Brigham Young University,
Provo, UT 84602
e-mail: dannyosmith@byu.net

Christopher A. Mattson
Professor
Department of Mechanical Engineering,
Brigham Young University,
Provo, UT 84602
e-mail: mattson@byu.edu

Eric C. Dahlin
Associate Professor
Department of Sociology,
Brigham Young University,
Provo, UT 84602
e-mail: eric.dahlin@byu.edu

Combining Direct and Indirect User Data for Calculating Social Impact Indicators of Products in Developing Countries

Evaluating the social impacts of engineered products is critical to ensuring that products are having their intended positive impacts and learning how to improve product designs for a more positive social impact. Quantitative evaluation of product social impacts is made possible through the use of social impact indicators, which combine the user data in a meaningful way to give insight into the current social condition of an individual or population. Most existing methods for collecting these user data for social impact indicators require direct human interaction with users of a product (e.g., interviews, surveys, and observational studies). These interactions produce high-fidelity data that help indicate the product impact but only at a single snapshot in time and are typically infrequently collected due to the large human resources and cost associated with obtaining them. In this article, a framework is proposed that outlines how low-fidelity data often obtainable using remote sensors, satellites, or digital technology can be collected and correlated with high-fidelity, infrequently collected data to enable continuous, remote monitoring of engineered products via the user data. These user data are critical to determining current social impact indicators that can be used in a posteriori social impact evaluation. We illustrate an application of this framework by demonstrating how it can be used to collect data for calculating several social impact indicators related to water hand pumps in Uganda. Key to this example is the use of a deep learning model to correlate user type (man, woman, or child statured) with the raw hand pump data obtained via an integrated motion unit sensor for 1200 hand pump users. [DOI: 10.1115/1.4047433]

Keywords: design for humans, machine learning, product development, sustainable design

1 Introduction

Approximately 4 billion people live on less than $8 per day and comprise what is often referred to as the *base of the pyramid* (BOP) [1]. Designing and producing products that meet the wants and needs of these income-poor individuals have the potential to improve their quality of life while also representing a significant market opportunity for companies of all sizes [2]. However, many who undertake such pursuits often fail to design products that have the intended impact on individuals in the developing world that comprise the BOP [3].

Evaluating a product's social impact is critical to ensuring that mechanical design results in positive impacts and avoids unintended negative impacts on people [2,4]. The social impact of a product refers to the effects that a product has on a person's daily quality of life [5]. Many different approaches are used for modeling the social impact of products or programs including a logic model [6,7], theory of change model [8], product impact metric [9], product social impact model [10], or social sustainability indicator model [11].

To at least some extent, each social impact model requires combining user data into social impact indicators, which are "what is measured or predicted in each impact category to understand a product's social impact" [10]. The change in these social impact indicators over time as calculated from the user data can capture the social impact of a product or a program as defined by Stevenson et al. and illustrated in Fig. 1 [10]. Collecting the user data to calculate social impact indicators is fundamental to quantifying the social impact of a product resulting from the mechanical design.

1.1 Direct Data.

Several different types of and methods for collecting the user data exist. The user data collected through direct interaction with or observation of users, termed *direct data* in the context of this article, is rich in information and has historically been the primary type of data that most organizations focus on obtaining when trying to understand social impacts [12]. Common methods used to collect the direct data include surveys, focus groups, interviews, observational studies, and ethnography studies [13,14]. The challenge with obtaining the direct data is that it can only realistically be collected once or at low-frequency intervals due to the high cost and extensive human facilitation required to collect such data [15,16]. This is often the case for those designing for global development who are frequently geographically removed from their customers [17]. Even short-term data collection efforts may require travel and cost thousands of dollars. Therefore, each snapshot of direct data has high value but also a large cost per data point for only a single point in time.

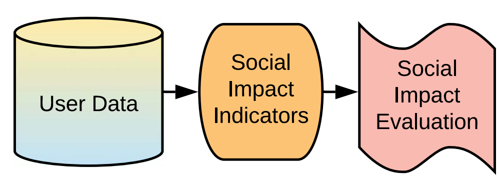

Fig. 1 Relationship among user data, social impact indicators, and social impact evaluation

Contributed by the Design Theory and Methodology Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received December 20, 2019; final manuscript received May 21, 2020; published online August 4, 2020. Assoc. Editor: Jitesh H. Panchal.

 Copyright © 2020 by ASME

1.2 Indirect Data. User data collected without direct interaction with or observation of users, termed *indirect data* in the context of this article, can also be obtained to determine the social impact indicators of a product. Advances in technology have enabled the collection of digital data that can be used to enhance and assist the use of direct data to determine the effects of products on individuals. Common methods for collecting these indirect data include sensors, social media platforms, satellites, and other Internet and digital technologies. The ability to rapidly collect and transmit vast amounts of these data has led to it also being referred to as *big data* [18]. The advantages of using indirect data are that it can be collected remotely, continuously, and often with a lower cost per data point than manual collection by individuals. For a large quantity of continuously collected data, it is generally less expensive to let a sensor collect data than to collect it manually. The challenge of using indirect data is that it can have little value without some way to interpret its meaning.

1.3 Proposed Framework. In this article, we propose a framework that demonstrates how direct and indirect data can be used together to enable the remote, continuous, and real-time collection of user data for calculating product-related social impact indicators through an approach that can be less expensive than manual data collection. These social impact indicators may then be used within one of the social impact models mentioned previously to identify the social impact of a product, or they may be used alone to simply understand the current social condition of an individual or population. The framework takes advantage of both types of user data, which complement each other well as presented in Table 1.

While the current typical approach to calculating social impact indicators is to manually collect direct data as shown in Fig. 2(*a*), the proposed approach shown in Fig. 2(*b*) is inspired by machine learning and includes special considerations for remotely and more easily collecting user data fundamental to calculating product-related social impact indicators. In machine learning, data is used to train a model that can then be used to make predictions about future data. Similarly, this framework shows how indirect data can be used to train a deep learning model, which can predict information-rich direct data for determining the social impact indicators related to products. The trained deep learning model can then be used to continuously predict direct data using indirect data without the need for frequent human facilitation. This approach is especially useful for products in the developing world where the costs and difficulty of simply collecting direct data can be prohibitive.

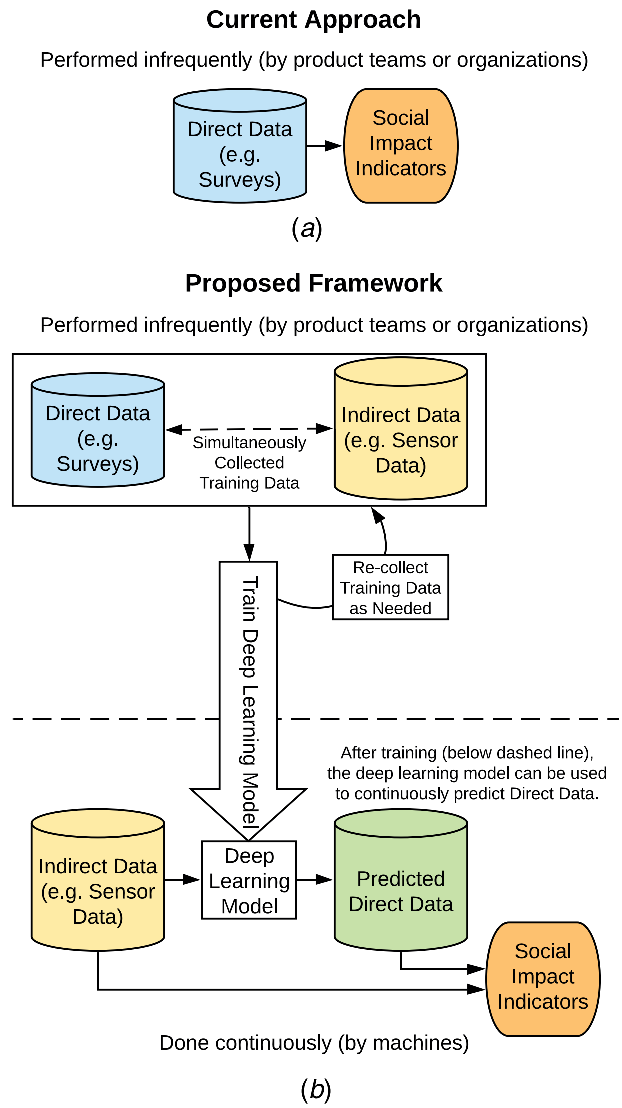

Fig. 2 (*a*) Current approach for calculating social impact indicators and (*b*) the proposed framework for calculating social impact indicators that uses deep learning and continuously collected indirect data

1.4 Deep Learning. We specifically advise the use of deep learning, a subset of machine learning, to provide the critical correlation between direct and indirect data because of the ability of deep learning to approximate complex functions or relationships more accurately than other machine learning approaches, especially when more than 10,000 data points are used in the model [19]. Deep learning is promising here due to the success it has had in many other applications. Existing successful applications of deep learning include both supervised and unsupervised data and vary widely from classification to generative tasks including image-based cancer detection [20], human and computer conversation [21,22], and image caption generation [23]. In another example, Bosco et al. enhanced geographically sparse survey data (direct data) with geospatial satellite data (indirect data) to predict physical growth stunting. They found that geospatial satellite data successfully corresponded to an impressive 60% of the variance in their model predicting the physical growth stunting of females across Nigeria [24]. Additional examples of effectively correlating direct and indirect data in other areas include improving efficiency of restaurant health code inspections in Boston using Yelp reviews [25],

Table 1 Data types for evaluating social impact of engineered products

	Amount of human effort to obtain each data point	Cost per data point	Independent or inherent value	Collection method or representation period	Example data sources
Direct data	Higher	Higher	Higher	Single snapshot in time	Surveys, focus groups, interviews, observational studies, ethnography studies
Indirect data	Lower	Lower	Lower	Continuous (in practice)	Online and social media activity data, remote sensor data, satellite data, digital purchase transaction data

tracking human activity using accelerometers in smart watches [26], creating individual health trends from monitoring human waste in toilets [27], aiding refugee settlement mapping with satellite imaging [28], tying farm production records to satellite imagery to help farmers conduct crop forecasting [29], and predicting regions of poverty using satellite imagery and census data [30].

Following the description and the process for using the proposed framework, we demonstrate the application of the framework in an example. In this example, a convolutional neural network deep learning algorithm is used to correlate direct and indirect user data that can be used to calculate several social impact indicators of using water hand pumps in Uganda (see Sec. 3).

2 Proposed Framework: Effective Collection of User Data for Social Impact Indicators

The purpose of the proposed framework is to describe how social impact indicators can be calculated from continuously and inexpensively collected user data. These social impact indicators may then provide the basis from which social impacts of products are evaluated. With social impact indicators calculated and monitored over time, designers may then use these data in the design process in at least three ways: (1) redesigning of existing products, (2) designing novel accessories for existing products that adapt or improve the functionality of a product relative to the product's original functionality, and (3) designing entirely new products, each based on insights provided by the data. For example, if location-specific social impact indicators for a water hand pump indicated that a significant portion of water pumping is completed by adolescents, these data would help designers possibly (1) redesign the pump to be more ergonomically efficient for adolescents to use, (2) design a handle accessory adapter that adds an additional pump handle at a more comfortable position for adolescents, or (3) design a completely new pump head and housing or design a new solar-powered pump that would run automatically so that the adolescents could spend time in school who previously spent in pumping.

Figure 2(*b*) illustrates each step of the *proposed framework*, which is shown in two parts–above and below the dashed line. Above the dashed line, a simultaneously but infrequently collected sample of direct and indirect data is used to train a deep learning model. Below the dashed line, the trained deep learning model predicts direct data given continuously collected indirect data. The predicted direct data and indirect data in some instances are then used to calculate social impact indicators to continuously monitor the effects of a product's use on the daily lives of individuals. The process for using the framework is presented in Secs. 2.1–2.6.

2.1 Step 1: Identify Use Context, Relevant Social Impact Categories, and Social Impact Modeling Approach for the Product. First, identify the context of the product's use and potential social impact categories relevant to the product. Ulrich and Eppinger provide some guidance for and suggest that those managing the development of a product must identify the *use environment*, or context, of the product [31].

As the context of a product's use is considered, potential impacts of the product will also become more apparent. Eleven possible social impact categories to consider, as outlined by Rainock et al., include health and safety, education, paid work, conflict and crime, family, gender, human rights, stratification, population change, social networks and communication, and cultural identity and heritage [32].

A review of industry practices shows that product designers and engineers of technology companies focus primarily on *health and safety* and generally do not consider other social impact categories nearly as much [33]; yet, it is important to consider the other ten categories because the likelihood of impact in more than one category can be high [34]. Less-apparent impact categories can be identified by utilizing the joint probability of a product having impact on more than one category, as presented by Ottosson et al. [34].

In addition, it is useful in the early stages of this process to identify the approach to social impact modeling that will be used to evaluate the social impacts of the product. Selecting the social impact modeling approach helps provide the structure for which indicators and hence user data will be needed to evaluate the social impact. A thorough consideration of which approach is best for a given application is deferred to the respective authors of the different approaches [6–11]. However, for product-related social impact modeling, we recommend the use of Stevenson et al.'s product social impact model approach due to its product focus [10].

2.2 Step 2: Identify Social Impact Indicators, Direct Data, and Indirect Data to Be Collected. The goal of this step is to identify the data to be gathered that will inform social impact assessment. As shown in Fig. 3, a correlation will be made between the indirect and direct data that will be used to calculate social impact indicators and inform social impact.

The activities of choosing social impact indicators, the direct data required to calculate those indicators, and indirect data sources required to predict the direct data are generally done in parallel based largely on the constraints of what information is feasible to collect.

Identify the social impact indicators that inform social impact evaluation. Stevenson et al. [10] describe a process for identifying appropriate social impact indicators. Their work focuses on calculating indicators from existing databases, such as the World Bank, while this work focuses on calculating pertinent indicators from user data collected from individuals relevant to the product. We choose to do this because (i) existing databases do not always have all of the important data relevant to an engineering design project, (ii) existing databases are often aggregated at a national level, which causes data about communities, families, or individuals within a country to be inaccessible to engineering teams, and (iii) existing databases typically include data collected once or only a few times a year. When selecting social impact indicators, at least one indicator is needed for each social impact category of interest, although as many indicators as are useful may be used.

Identify direct data that are used to calculate social impact indicators. The direct data are identified such that the change in the direct data *indicates* the effects of a product's use on an individual. The direct data are used to calculate the social impact indicators in one or more of the selected social impact categories. The choice of which direct data to collect is guided by the social impact category chosen and by the ability to correlate the direct data with the indirect data. Possible sources of the direct data include surveys, focus groups, interviews, observational studies, and ethnographic studies.

While the identification and collection of the direct data are sufficient to calculate the social impact indicators, the framework presented in this article proposes the collection of the *indirect data* that can be correlated with and used to predict the direct data. These predicted direct data can then be used to continuously, remotely, and inexpensively calculate social impact indicators.

Identify sources of indirect data related to direct data. Useful indirect data have inherent relationships to each of the direct data of interest such that the direct data can be predicted given the

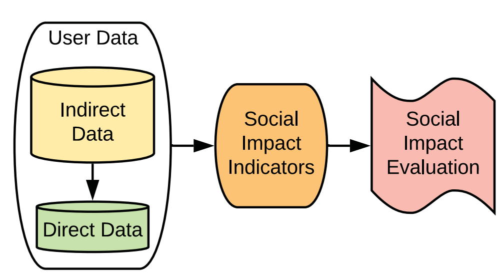

Fig. 3 Relationship among indirect data, direct data, social impact indicators, and social impact evaluation

indirect data. Some potential sources of the indirect data are sensor data (motion, vibration, temperature, flow, pressure, strain, among others), online and social media activity data (product reviews, Likes, Tweets and Retweets, Comments, etc.), satellite data (imaging, thermal imaging, geographic, precipitation, etc.), and transaction data (online and offline purchase data).

The relationship between the direct and indirect data will be established with a correlation model as discussed in Step 4. One potential issue to be aware of when selecting direct and indirect data sources is that the original correlation may be so specific to the initial data that they do not generalize well with the future data [35]. Therefore, when considering possible sources of the indirect data for predicting the direct data, it is important to consider the following:

(1) Is there reasonable evidence of a relationship between the direct and the indirect data?
(2) Does the relationship between the indirect and the direct data remain relatively constant with time?
(3) Is it possible to perform periodic resampling of the direct data to maintain an accurate correlation?
(4) Can the direct data be accurately predicted using only the indirect data?

If the answers to these questions are "yes," the source of the indirect data is likely to be well suited for predicting the direct data through a correlation model. If the answer is "no" to any of these questions, the relationship between the selected direct and indirect data may require more frequent validation, or different indirect data may be a better choice.

To illustrate this step, an example is provided. SweetSense Incorporated has deployed a cellular enabled data collection system for latrines that reports (a) each use of the latrine, (b) the approximate fill level of the waste receptacle, and (c) the GPS location of the latrine [36]. If the sensor was installed at a school, the recorded uses of the latrine could act as the indirect data and could be correlated with the number of students attending the school (direct data) as determined through survey or observation. The correlation model could then be used to predict number of students attending the school given uses of the latrine. If the predicted number of students attending the school were compared to the census of school-age children, the percentage of school-age children attending school could be calculated as a social impact indicator. These data could be collected remotely and continuously before and after a project or product introduction to assess the educational social impact.

2.3 Step 3: Collect Data and Construct Correlation Model.

Step 3 is composed of substeps 3a and 3b described in Secs. 2.3.1 and 2.3.2, respectively, because of the interdependent relationship between data collection and correlation model creation.

2.3.1 Step 3a: Collect Direct and Indirect Training Data.

After identifying the data needed to calculate the social impact indicators of interest, simultaneously collect an initial set of direct and indirect data (referred to as training data) that will be used to construct the correlation model for predicting the direct data given the indirect data.

One potential issue when collecting the data is class imbalance: when one or more of the classes being classified has a dramatically lower number of samples than the other class or classes. Class imbalance is important to be aware of during the data collection step because it can potentially lower the classification accuracy of the lower frequency class, which is especially important when it is critical to classify each class with similar accuracy. To avoid class imbalance, an approximately equal number of samples from each class should be collected when possible. If the sample class cannot be selected when collecting data, class imbalance can be addressed during the creation of the correlation model.

Also, in collecting the training data, it is important to consider the long-term indirect data collection process. It is beneficial to create a data collection process that will not only work for collecting training data but also enable the continuous, remote, and inexpensive collection of future indirect data, which will be used to predict the direct data in real time. Important considerations for long-term data collection are discussed in Sec. 2.5.

2.3.2 Step 3b: Construct Correlation Model Between Direct and Indirect Data.

Select and construct a correlation model that effectively captures the relationship between the direct and the indirect data. The intent of the correlation model is to predict the direct data given future indirect data. As a note, the correlation model is different than the social impact model mentioned previously. The relationship between the direct and indirect data is often nonlinear; thus, a machine learning tool, such as deep learning, is a promising candidate to model the relationship and recognize complex relationships between the direct and indirect data of interest [37]. However, any type of correlation model could be considered here. Some potential modeling approaches include logistic regression, support vector machines, regression trees, and supervised deep learning classification models (such as a convolutional neural network). It is also beneficial to consider multiple different correlation modeling approaches for comparison, exploring the fastest to implement options first to determine whether a simple logistic regression model, for example, will produce the needed performance. The number of models created and compared will likely be constrained by the minimum accuracy required for the correlation to be useful and the time and resources available to create such models.

As this work focuses on the use of deep learning to produce correlation models, an outline of the recommended process for creating a deep learning model is provided here. This process describes what can be done to achieve the theoretical maximum performance of a given correlation model and can be stopped as soon as desired model performance is reached with subsequent steps then being unnecessary. This process applies to most deep learning models generally and includes specific details for applying deep learning models to developing world applications. Every issue or detail of creating a desirable deep learning model is not addressed, but the process is described in further detail by Goodfellow et al. [38] and other researchers as specified in the relevant steps.

If completing the following steps do not result in the needed model performance, there are two possible reasons. The first possible cause is that there is a fundamental disconnect in the relationship between the direct and indirect data. One solution for this is to collect an additional source of the indirect data that could help explain the direct data of interest and repeat the entire process while including the new data source. The second possible cause is that the learning algorithm being used is not well suited to approximate the relationship of interest. This may be solved by using another existing algorithm or developing a new deep learning algorithm.

After acceptable accuracy has been achieved, the constructed model can then be used to predict the direct data given the future indirect data.

The recommended process for creating the model is as follows:

(1) *Identify appropriate error metrics and goal performance values for those error metrics.* Classification accuracy is the most common performance metric although precision, recall, F-score, or others may be more useful for a given application [38,39].
(2) *Create baseline model and end-to-end data processing method including initial performance values.* Hyperparameter tuning can be used to obtain the best initial model possible [38].
(3) *Consider the effects of class imbalance on model performance* as indicated by a lower classification accuracy for the class(es) for which there are a lower number of data samples. A confusion matrix of the predicted classifications can be used to identify the accuracy of each class prediction. Class imbalance may need to be dealt with when there is an

underrepresented class that critically needs to be classified accurately (as in the case of cancer cell detection in medical imaging). However, class imbalance is not always a problem if there is a class that does not occur as frequently as others and that class does not need to be classified with greater accuracy than is currently being attained. It may be best to accept a lower classification accuracy of the lower frequency class to preserve the overall accuracy because techniques for handling class imbalance can improve the accuracy of the lower frequency class at the expense of the overall accuracy. When it is determined that class imbalance is adversely affecting the model's performance, techniques for resolving class imbalance include oversampling [39], undersampling [39], SMOTE [39], focal loss [40], and ClusBUS [41].

(4) *Check quality of data.* If the data are noisy, corrupt, or has other avoidable issues based on the domain, *better* data should be collected instead of just collecting *more* data. Often it is tempting to use a new algorithm or change modeling approach if performance does not meet requirements. However, it is typically more important to collect better data than to try to improve the deep learning algorithm [38].

(5) *Determine whether to collect more data.* This can be done by plotting the test error versus log scale training set size for models created from all data and smaller subsets of data within the existing dataset. Because there is commonly a trend between test error and training set size, the trend can indicate if a larger training set would be beneficial. Collecting more data will be beneficial if the test error is trending downward with the increasing training set size [42].

During the process of creating the models from the reduced datasets, model layer sizes and hyperparameters will need to be adjusted to avoid overfitting and maintain an accurate representation of the true model performance.

In most deep learning modeling processes, it is often more beneficial and thus recommended to collect more data rather than trying to use a different learning algorithm once it is determined that more data will reduce test error for the current algorithm. However, due to the typically high cost of collecting additional training data in developing world situations, it may be more time effective to explore additional learning algorithms.

(6) *If collecting more data is projected to reduce the test error, simultaneously collect the additional data and check/refine model performance until required performance or theoretical max accuracy is achieved.* Collecting a small amount of additional samples is not likely to improve the model, so it is recommended to double the number of collected samples between experiments. The number of parameters in the model and the hyperparameters will need to be refined when creating new models after adding more data. Including additional data to reduce test error will become unproductive as it asymptotes.

2.4 Step 4: Calculate Social Impact Indicators From Training Data. With the initial set of direct data collected, the initial social impact indicators can be calculated. Calculating indicator values from the initial data will help ensure that all data necessary for calculating indicators has been collected and that there are no issues with the social impact indicator equations before moving into long-term data collection efforts.

2.5 Step 5: Continuously Collect Indirect Data to Predict Direct Data and Calculate Social Impact Indicators. The benefit of using this framework is the increased frequency of direct data predictions that offer near real-time insight into the social impact and enable quantification of a product's usage. The predicted direct data can act as a surrogate for direct data during the time between collecting direct data samples.

When preparing for continuous remote monitoring, consider the following:

- Collection and storage method for indirect data
- Power source and recharging of data collection system (if applicable)
- Data pipeline configuration including telecommunication technology
- Frequency of indirect data processing appropriate to the application
- User rights and data privacy

For example, if the indirect data source is a sensor, the indirect data could be transmitted remotely using the cellular network and stored using web servers. Then, the raw sensor data would be configured for use with the correlation model. Finally, the model would predict the direct data as the indirect data are fed into the correlation model at a frequency appropriate for the application.

Utilize the constructed correlation model to predict the direct data given the continuously collected indirect data. The indirect and thereby predicted direct data should be collected for a sufficient amount of time to be representative of the impacts to be identified. The predicted direct data can then be used to nearly continuously calculate social impact indicators and fed into the social impact model chosen in Step 1 to identify the social impacts of the product as shown in Fig. 3.

2.6 Step 6: Maintain Accuracy of Direct and Indirect Data Correlation. It is possible that the relationship between the direct and indirect data will change over time as human behavior and other social, political, environmental, and economic factors change. Therefore, it is important to periodically and as often as needed to collect samples of simultaneously collected direct and indirect data. The new batch of data can then be used to update the correlation model and maintain or improve the accuracy of the correlation between the direct and the indirect data. Data collection partnerships with individuals or organizations that live near the population of interest can further reduce the cost of maintaining the accuracy of the correlation.

2.7 Framework Summary. The proposed framework is an approach to enhancing information-rich, manually collected direct data with the predicted direct data for calculating social impact indicators. The predicted direct data are made possible through a correlation model that is trained using simultaneously collected direct and indirect data. The framework's effectiveness is realized in situations where manually collecting direct data is especially costly, yet where indirect data can be collected remotely and continuously, such as in developing countries. This framework provides a relatively inexpensive approach to remotely gather continuous, detailed data regarding the effects of a product's use on the lives of individuals in situations located remotely from the researcher. The following section illustrates one detailed example of how this framework was employed.

3 Example: Water Hand Pumps in Uganda

The proposed framework is used to demonstrate how direct and indirect data, correlated via a deep learning model, can be collected to continuously calculate social impact indicators relevant to water hand pump users. This example focuses on the application of only steps 1–4 due to the long-term maintenance focus of steps 5 and 6.

3.1 Step 1: Use Context, Relevant Social Impact Categories, and Social Impact Modeling Approach for Water Hand Pump. Regarding the context of the product in this example, harmful safety and gender-related impacts regarding water hand pump use can be more effectively addressed with a more complete picture of those who are using water pumps and when the pumps are being accessed [43].

The task of collecting clean water external from the home is a task that affects men, women, and children differently. In developing countries, it is reported that as many as 80% of women and girls are primarily responsible for drawing and transporting water for household consumption [44–46]. However, accurately measuring the water collection burden placed on women and children as it changes over time and over a widespread area is difficult and impractical.

Water pump usage practices have important gender implications by limiting female involvement in other activities or threatening the physical well-being of women and girls. When women and girls are tasked with obtaining water and other household chores, their involvement in other activities such as generating income, spending time with community or family members, childcare, leisure, or schooling can suffer. With respect to schooling, domestic responsibilities are more likely to prevent girls than boys from being on time or attending school altogether. One study in Morocco showed that projects designed to reduce girls' responsibility for collecting water increased girls' school attendance 21% over a 4-year period [47]. Another study by Assaad et al. examines the relationship between work (including paid work and domestic chores like collecting water) and schooling for a sample of 2,442 girls in Egypt ages 6 through 14 years. Assaad et al. found that an increased probability of working is negatively associated with attending school [48].

The implications for everyday female experiences related to retrieving water extend beyond school-related activities. Other consequences of fetching water, especially over long distances, may include physical strain, threats to physical safety, and animal attacks [49]. Pommells et al. discuss the threats to women's well-being associated with getting water in communities in East African countries. One informant who describes the risks of sexual assault states, "It is a good time to take advantage of women who are going to water sources to carry water home, especially peak hours, early in the morning…and late in the evening…on their way, these guys are waiting for them, and since it is generally accepted practice of the community, they will be raped" [50]. Because water collection may constitute such a large portion of the day, gendered water retrieval practices leaves women and girls more vulnerable to the threat of sexual violence. By gaining a more complete picture of those who are using water pumps and when the pumps are being accessed, communities can address these issues and begin to implement new practices that will help to ameliorate harmful and gendered outcomes.

Based on this context, the five relevant social impact categories from Rainock et al. [32] related to water hand pump usage are gender, health and safety, conflict and crime, education, and paid work.

We chose to follow the product social impact modeling approach as presented by Stevenson et al. [10]. In short, the product social impact model consists of identifying social impact indicators for each relevant category and using the direct data to calculate those social impact indicators. The example presented here includes the selection of social impact indicators and initial calculations, but does not extend through time.

3.2 Step 2: Identify Social Impact Indicators, Direct Data, and Indirect Data. The indicators chosen for the Gender social impact category were *total number of hours and fraction of daily pump usage* by each of the user groups (men, women, and children statured). These two indicators were selected because they reflect how the water collection burden varies by gender and age groups.

The indicator chosen for the Health and Safety social impact category was *average individual fraction of daily energy intake (from food consumption) expended* by users of each group. This was chosen as it indicates the potential health and nutritional challenge of each user group caused by lost calories due to pumping water.

The indicators chosen for the Conflict and Crime social impact category were total and fraction of *hours spent pumping water* by each of the user groups during the 2 h before sunrise and 2 h after sunset. These were chosen because the 2 h before sunrise and 2 h after sunset are the pump usage times during which individuals are more susceptible to the physical attack.

The indicator chosen for the Education social impact category was the *fraction of total pump usage by child-statured individuals during school hours* (8 a.m.–5 p.m.). This was chosen as it indicates the fraction of time during the day that children were using the pump instead of being in school.

Finally, the indicator chosen for the Paid Work social impact category was the *potential wages lost* for men and women due to using the pump as this indicates the financial implications that using the pump instead of working had on men versus women.

3.2.1 Social Impact Indicators. The social impact indicators identified for each of the relevant categories identified are shown in Eqs. (1)–(10). For the Gender social impact category, the imbalance in water collection roles across genders was considered.

For the Gender social impact category:

$$T_{U,i} = \sum_{j=1}^{N_i} t_{U,i_j} \tag{1}$$

where $T_{U,i}$ is the total pump usage time for the ith group, $i = 1$, 2, and 3 for men, women, and children, respectively, t_{U,i_j} is the pump usage time for the jth user of the ith group, and N_i is the number of individuals in the ith group.

$$F_{T_{U,i}} = \frac{T_{U,i}}{\sum_{i=1}^{3} T_{U,i}} \tag{2}$$

where $F_{T_{U,i}}$ is the fraction of time of total pump usage by the ith group and $T_{u,i}$ is the total pump usage time for the ith group.

For the Health and Safety social impact category:

$$F_{E_{U,i}} = \frac{\sum_{j=1}^{N_i} e_{U,i_j}}{N_i \cdot e_{S,i}} \tag{3}$$

where $F_{E_{U,i}}$ is the average individual fraction of energy expended by a user of the ith group, e_{U,i_j} is the energy expended by the jth user of the ith group while pumping, N_i is the number of users of the ith group, and $e_{S,i}$ is the average individual energy obtained through sustenance of the ith group.

$$e_{U,i_j} = V \cdot \rho \cdot g \cdot \frac{d}{2} \cdot \frac{\sum_{k=1}^{N_{P_i}} A_{j,k}}{A_{\max}} \tag{4}$$

where N_{p_i} is the number of pump strokes of the jth user in the ith group, $A_{j,k}$ is the pump amplitude of the kth pump stroke of the jth user, and $A_{\max}$ is the maximum possible pump stroke amplitude for the water hand pump being used.

For the Conflict and Crime social impact category:

$$T_{U_H,i} = \sum_{j=1}^{N_i} t_{U_H,i_j} \tag{5}$$

where $T_{U_H,i}$ is the total pump usage time spent by individuals of the ith group during the high risk hours 2 h after sunset and 2 h before sunrise, t_{U_H,i_j} is the pump usage time during high risk hours for the jth user of the ith group, and N_i is the number of individuals in the ith group that used the pump during high risk hours.

$$F_{T_{U_H,i}} = \frac{T_{U_H,i}}{T_H} \tag{6}$$

where $F_{T_{U_H,i}}$ is the fraction of total pump usage time during high risk hours by the ith group and T_H is the total time during the high risk hours.

The indicator values for groups 2 and 3 (women and children) are particularly important for this category since women and children are at a higher risk of being victims of crime or attack.

For the Education social impact category:

$$T_{U_D,3} = \sum_{j=1}^{N_3} t_{U_D,3_j} \tag{7}$$

where $T_{U_D,3}$ is the total pump usage time spent by individuals in group 3 (children) during the typical education period as specified by the Ugandan government of 8 a.m. to 5 p.m. on weekdays, $t_{U_D,3_j}$ is the pump usage time during the typical education period for the jth user of group 3 (children), and N_3 is the number of users in group 3 (children) that used the pump during the typical education period [51].

$$F_{T_{U_D,3}} = \frac{T_{U_D,3}}{T_D} \tag{8}$$

where $F_{T_{U_D,3}}$ is the fraction of total pump usage time during which the pump was being used by group 3 users (children) during typical education period and T_D is the total time during the typical education period.

For the Paid Work social impact category:

$$W_{U,i} = T_{U,i} \cdot R_i \tag{9}$$

where $W_{U,i}$ is the potential wages lost due to time of pump usage of the ith group and R_i is the average rate of pay for the ith group, which is the pay rate that the user could be earning if not pumping water [52].

$$F_{W_{U,i}} = \frac{W_{U,i}}{\sum_{i=1}^{3} W_{U,i}} \tag{10}$$

where $F_{W_{U,i}}$ is the fraction of potential wages lost by the ith group due to pump usage.

3.2.2 Direct Data. The primary source of the direct data required to calculate the social indicators identified is given by the usage of the water hand pump by user type (man, woman, or child) throughout the day as it indicates their relative water collecting responsibilities. User types were classified as man, woman, or child statured by an observational researcher familiar with dress and cultural norms that indicated the user type. The *stature* distinction was used due to the occasional difficulty of distinguishing between small-statured adults and/or large-statured children. The genders of children were not distinguished.

3.2.3 Indirect Data. The indirect data used here to predict the direct data were data from an inertial measurement unit that measured pump handle angle over time (see Fig. 4). At the outset of this study, it was observed that the speed of pumping and the magnitude of pump strokes differed between individuals, and we theorized it was distinct among men, women, and children. Therefore, this pair of direct and indirect data was chosen for this experiment.

3.3 Step 3a: Collect Direct and Indirect Training Data.

Observation of pump users at a distance of 10–15 m from the pump was our approach to direct data collection to avoid biasing the normal use of the pump by speaking to pump users. Interacting with pump users and asking them their age and gender may have resulted in a slightly more accurate classification but would have been obtrusive and could have affected normal pump usage.

The indirect data were obtained using an Arduino microcontroller setup equipped with a Bosch BNO055 9-axis (accelerometer, gyro, magnetometer) absolute orientation sensor that was mounted on the pump handle. This measured and stored the handle angle on an onboard microSD card at a frequency of 12–25 Hz. The variable collection frequency was due to an unexpected reduction in data writing speed that occurred as file size increased. To account for variable collection frequency, a second channel of time between data points (in milliseconds) was used along with the handle angle as an input channel to the deep learning model. Figure 5 shows the sensor mounted on the pump handle, and Fig. 4 shows a more detailed photo of the hardware. Figure 6 shows a representative 5-s segment of handle angle versus time data. To identify when one user would stop and the next begin, a remote control sensor connected to the handle sensor via Bluetooth allowed the data collector to increment the user number.

These data were collected from 1181 users including 115,000 pump strokes and 2.67 million data points obtained over a total of 4 full days at four pump sites in two different cities located at different ends of the country in Uganda. Table 2 presents a summary of the demographics of the data collected.

Because the users of the pump could not be chosen, class imbalance could not be treated at this stage of the process.

3.4 Step 3b: Construct Correlation Model Between Direct and Indirect Data.

The initial desired classification accuracy of the model chosen for this application by the practitioners is 75%. This was chosen based on the initial accuracy of deep learning

Fig. 4 Microcontroller setup composed of a custom PCB and sensors that tracked motion of the handle while researchers incremented users with a remote control

Fig. 5 Pump site in northern Uganda with the sensor mounted to the pump handle

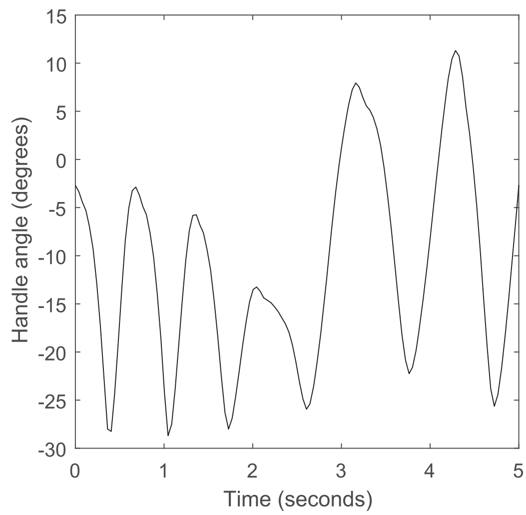

Fig. 6 Representative 5-s chunk of angle versus time data collected using the sensor system

Table 2 Demographics for data collected for this example

	Men	Women	Children	Total
Number of users	98	321	762	1181
Number of hours using pump	7.4	19.4	20.0	46.8

classification of other similar applications [53,54] The performance measure of accuracy was chosen over precision, recall, and F-score because we are primarily interested in overall classification accuracy without any particular interest in improving the classification accuracy of the least represented class, which was men in this example.

Three potential correlation models were considered to provide the correlation between the direct data of user type and the indirect data of pump handle angle versus time: logistic regression, convolutional neural network, and recurrent neural network. The convolutional neural network is given primary focus as it is the most accurate and useful model.

3.4.1 Data Preparation. Inherent in the training of a supervised classification or correlation model is the often manual classification of the dataset to generate the training data. To classify each user type, we worked with local assistants to establish visual indicators of gender based primarily on culturally indicative clothing. Each segment of sensor data was paired with its user type by observing pump users and classifying each user as man, woman, or child statured. To reduce classification bias between pump sites, the same researcher classified the data from all locations and pumping periods.

As customary in model creation, the data were prepared prior to use in the models. Approximately 5% of the original users whose data were severely corrupted by sensor drift were removed. Dormant time during which the pump handle was not being moved was also removed from the dataset.

3.4.2 Logistic Regression. A multinomial logistic regression model was created using JMP 14 statistical software because it provided a simple proof of concept and was straightforward to implement [55]. This model used the stroke level inputs of pump stroke period and amplitude to predict user type. A 70/30 train/predict split was used for this model and resulted in a classification accuracy of 51.5%. While the classification accuracy obtained by this logistic regression model is much better than random (33%), it is nevertheless significantly lower than that resulting from the convolutional neural network as shown next and justifies the use of deep learning in creating the correlation. This lower accuracy indicates that the information contained when using isolated strokes as used by the logistic regression model to predict user type is not as indicative of user type as when analyzing the raw data of multiple strokes together as done by the deep learning model. Therefore, the need for deep learning in this case is validated.

3.4.3 Convolutional Neural Network. The ability of deep learning models to identify the characteristic pumping features of the different user types more effectively than the traditional statistical approach of logistic regression made it well suited for use as the correlation model in this application. The specific deep learning model used for this application is a convolutional neural network with a one or more 1D convolutional layers followed by a final fully connected layer. This was selected due to its ability to extract data features more effectively than a basic fully connected deep neural network. One-dimensional convolutions were used to extract features from the 1D time series data similar to how 2D convolutions extract features from an image.

Two data channels of handle angle and sampling period were used as the inputs for the convolutional neural network. The data were normalized and adjusted to have a mean of zero before being fed into the deep learning model.

To use a convolutional neural network, a fixed size array input was required. This was achieved by selecting sequential and random *chunks* of a fixed number of sequential data points from the entire segment of a user's data. At the sampling frequencies stated, each chunk of data points represents approximately 4–8 s of pumping motion. In practice, this means that a man who pumped for 100 s was effectively represented in the data set by 12–25 unique chunks.

The PYTORCH open-source deep learning platform was used in conjunction with Colab, Google's cloud–based Jupyter notebook environment to provide the computational ability to create the model [56,57]. A randomized 70/30 train/test split at the user level was used to designate which user data were used to train versus validate the model. The loss function used was *cross entropy loss* as shown in Eq. (11) because, in practice, it typically provides more accurate results in deep learning classification models than other loss functions. The *Adam* optimizer algorithm with β values of 0.9 and 0.99 for β_1 and β_2, respectively, was used for all models due to its historically high performance on convolutional neural networks [58]. All models were run for 30 epochs, which was sufficient to allow the model sufficient time to learn.

$$loss = -\frac{1}{N}\left(\sum_{i=1}^{N} \log(\hat{y}_i)\right) \tag{11}$$

where N is the total number of chunks and $\hat{y}_i$ is the resulting softmax probability of the true class for each chunk [59].

Preliminary hyperparameter exploration was used to identify acceptable and reasonable baseline model hyperparameters. The hyperparameters that were considered include number, size, and shape of convolutional and fully connected layers, dropout (y/n; dropout percentage), chunk size, batch size, learning rate, and activation function. Systematic hyperparameter exploration was then completed to identify the values of these hyperparameters that resulted in the highest model accuracy. This process first involved holding other baseline hyperparameters constant while adding convolutional layers (stride = 1) with double the output channels as the previous layer until overfitting occurred. This was followed by the addition of regularization efforts in the form of adding dropout layers in various places of the model. Several different approaches of adding dropout were considered including having one dropout layer ($p = 0.25$) after the first convolutional layer, dropout layers ($p = 0.15$) after every convolutional layer, and dropout layers ($p = 0.25$) after every convolutional layer. Figure 7 shows the resultant average of the fourfold cross-validation classification accuracy of

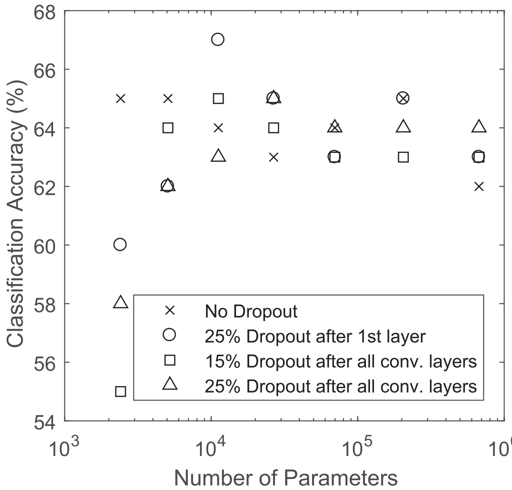

Fig. 7 Classification accuracy versus number of model parameters for models with the increasing number of convolutional layers and dropout

the models. Based on this process, the models with approximately 10,000 parameters provided the highest accuracy without significant overfitting.

The effect of other hyperparameters including chunk size (50, 100, 150, and 200), batch size (25, 40, and 55), learning rate (0.01, 0.001, and 0.0001), and activation function (ReLU, SELU, and LeakyReLU (negative slope = 0.01)) were also considered for the baseline 10,000 parameter model using threefold cross-validation to find the average accuracy. The results of this analysis indicated that the only changes that improved the accuracy were increasing the chunk size from 100 to 150 (67% accurate) and reducing the batch size from 40 to 25 (66% accurate).

The most accurate and best generalizing model identified as a result of this process discerned among men-, women-, and children-statured users with 67% test accuracy (where 33% is random) and between men and women/children statured combined with 84% test accuracy (where 50% is random). The model used three convolutional layers and one fully connected layer with the ReLU activation function and dropout ($p = 0.15$) after every convolutional layer with a chunk size of 150. This resulted in a model with 16,043 parameters. The topology of the network used is as follows:

- Input array: 100 × 2 (100 samples × 2 channels for current angle and time between sampling)
- 1D convolutional layer (input channels: 2, output channels: 8, kernel size: 3, stride: 1, padding: 0)
- ReLU
- Dropout (percent removed: 15%)
- 1D convolutional layer (input channels: 8, output channels: 16, kernel size: 3, stride: 1, padding: 0)
- ReLU
- Dropout (percent removed: 15%)
- 1D convolutional layer (input channels: 16, output channels: 32, kernel size: 3, stride: 1, padding: 1)
- ReLU
- Dropout (percent removed: 15%)
- Fully connected layer (input size: 4672, output size: 3 or 2 (for fully or reduced class model cases, respectively)

Due to the class imbalance of having a larger number of women and children for which data were collected, the classification accuracy of the model was 50%, 76%, and 65% for men, women, and children, respectively (34% and 92% for the men and combined women–children, respectively, for the combined model). Since women and children typically use the pump more than men, no efforts are necessary to implement class imbalance techniques to improve the classification accuracy of men at the expense of overall accuracy.

While 67% accuracy is a decent starting classification accuracy, this application would benefit by an improved classification model. This warrants completing the process to determine whether the model could be improved through further data collection. The first step to this process is examining the current data to determine whether it is noisy or corrupt. On inspection of the data, it was clear that approximately 5% of the data were unusable due to major drift; thus, it was removed. The remainder of the data were usable but also suffered with noise artifacts caused by sensor drift that seem likely to introduce error and reduce model accuracy. Therefore, the most beneficial next step to improve the model is to recollect the data using an improved sensor system void of drift artifacts. We plan to use a different sensing mechanism to treat the drift problem for future data collection.

3.4.4 Recurrent Neural Network. Another potential way to improve model accuracy is through the use of a recurrent neural network to provide the classification ability due to the ability of recurrent neural networks to capture long-term behavior compared

Table 3 Calculated current social impact indicators for two locations in Uganda

Category	Description	Indicator	Jinja average	Gulu average
Gender	Time men pumped (hours/day)	$T_{U,1}$	2.7	1.0
Gender	Time women pumped (hours/day)	$T_{U,2}$	2.6	7.1
Gender	Time children pumped (hours/day)	$T_{U,3}$	7.4	2.6
Gender	Fraction of pump usage time (men (%))	$F_{T_{U,1}}$	21	9
Gender	Fraction of pump usage time (women (%))	$F_{T_{U,2}}$	20	66
Gender	Fraction of pump usage time (children (%))	$F_{T_{U,3}}$	59	24
Education	Time children pumped during school hours (hours/day)	$T_{U_D,3}$	4.6	1.9
Education	Fraction of pump usage time during school hours (children (%))	$F_{T_{U_D,3}}$	55	23
Conflict and crime	Time men pumped during high risk hours (hours/day)	$T_{U_H,1}$	0.64	0.12
Conflict and crime	Time women pumped during high risk hours (hours/day)	$T_{U_H,2}$	0.62	0.87
Conflict and crime	Time children pumped during high risk hours (hours/day)	$T_{U_H,3}$	1.8	0.32
Health and Safety	Fraction of individual daily energy expended pumping (men (%))	$F_{E_{U,1}}$	0.84	0.43
Health and safety	Fraction of individual daily energy expended pumping (women (%))	$F_{E_{U,2}}$	0.62	0.80
Health and safety	Fraction of individual daily energy expended pumping (children (%))	$F_{E_{U,3}}$	0.17	0.46
Paid work	Potential wages lost due to time pumping (men (USD/day))	$W_{U,1}$	0.78	0.29
Paid work	Potential wages lost due to time pumping (women (USD/day))	$W_{U,2}$	0.26	0.71
Paid work	Potential wages lost due to time pumping (children (USD/day))	$W_{U,3}$	N/A	N/A
Paid work	Fraction of potential wages lost (men (%))	$F_{W_{U,1}}$	75	29
Paid work	Fraction of potential wages lost (women (%))	$F_{W_{U,2}}$	25	71
Paid work	Fraction of potential wages lost (children (%))	$F_{W_{U,3}}$	N/A	N/A

to a purely convolutional neural network approach [38]. However, the use of a recurrent neural network in this case would be more difficult to deploy than the convolutional neural network because, unlike the convolutional neural network, a recurrent neural network would require prior identification of when one user stops and another begins. Therefore, the development of a recurrent neural network model for this application is the subject of future work.

3.5 Step 4: Calculate Social Impact Indicators From Training Data.

The calculations of social impact indicators were completed using the direct data that were collected during the 2 days at pump sites in Jinja and 2 days at pump sites in Gulu. Table 3 presents the calculated social impact indicators by location as averages over the 2 days of data collected at each location. The following approaches were used in calculating the indicators:

Health and safety: The maximum possible pump stroke amplitude ($A_{\max}$) on the pumps observed was approximately 52 deg. The pump parameters (V and d) were specific to the India Mark II hand pump and found in Ref. [60]. Values for average individual energy obtained through sustenance for each group ($e_{S,i}$) were approximated using values from the Dietary Reference Intakes [61].

Conflict and crime. Our researchers were collecting data approximately 30 min before sunrise and 1 h after sunset, but the user type was not recorded during the dark hours, so the fraction of users of each type was assumed to be the same for light and dark hours of the day.

Paid work. The average hourly wage (R_i) was approximated for men using the “compensation of employees” in Uganda published by the World Bank [52]. The “compensation of employees” was divided by the number of working days (6 days per week $\times$ 52 weeks = 312 days) and working hours per day (8 h per day). According to this approximation, Ugandan men make \$0.29 USD per hour. According to the study by Campos et. al., Ugandan women salaries are approximately one-third of a typical Ugandan man’s salary [62]; thus, the women’s hourly wage is estimated to be \$0.10 USD per hour.

Education. The typical education period as specified by the Ugandan government is 8 a.m. to 5 p.m. on weekdays [51].

3.6 Steps 5 and 6: Continuously Collect Indirect Data to Predict Direct Data and Calculate Social Impact Indicators/Maintain Accuracy of Direct and Indirect Data Correlation.

These steps were not completed for this experiment due to the long-term maintenance focus of these steps. Long-term data collection will involve the deployment of a sensor that will transmit data via cellular networks. The device will be powered using an energy harvester that converts mechanical power during pumping into electrical energy. Additional training sets will be gathered during field studies or by employing local individuals.

3.7 Discussion of Example.

The example included provides a number of insights regarding the social impacts of water hand pumps in Uganda as well as the application of the framework proposed in this article.

Observing local Ugandans using the water hand pumps for 4 days showed some patterns. First, the water collection burden indeed falls primarily on the shoulders of women and children at least at these pump locations in Uganda. As presented in Table 3, men represented approximately 21% and 9% of pump users in Jinja and Gulu, respectively, where the women and children represented the remaining 79% and 91%, respectively. Men operated the pumps for a much lower percentage of time but were frequently present while women or children operated the pump.

Also, the Conflict and crime rows of Table 3 show that women and children continue to shoulder the water collection burden, even during higher risk hours, thus placing them in potential dangers discussed in Sec. 3.1.

Although women and children primarily carry the water collection burden in these areas of Uganda, the potential wages lost due to time pumping (see Paid Work rows of Table 3) is not as straightforward. Because men are on average paid three times more than women in Uganda, women have to put in more time pumping water to match men’s wages lost due to pumping. After accounting for this difference in hourly wage and the total pumping time by men and women, the potential wages lost by men and women varied greatly by location. In Jinja, men lost three times the potential wages as women, and in Gulu, women lost 2.4 times the potential wages as men.

Regarding energy lost by individuals in the different user groups, results varied by location. In Jinja, men typically expended the greatest proportion of their individual daily energy pumping (0.84%), whereas in Gulu, women typically expended the greatest proportion of individual daily energy (0.80%) as shown in the Health and safety rows of Table 3.

Regarding the Education social impact indicators, the fraction of pump users that were children during normal school hours was 55% in Jinja and 23% in Gulu. If this information was collected continuously, patterns could be observed over time that may suggest causes for children missing school. Perhaps more children miss school during harvest time, or more children help fetch water during breaks from school. The remote and continuous collection of these data could offer valuable insights.

Finally, although not captured quantitatively, the pump was a place of community. Groups of women or children would converse or play near the pump as they waited their turn to retrieve water. In terms of the social impact categories [32], the pump was a place of social network and communication for these local Ugandans. Compared to gathering water at the closest natural water source, individuals may choose to travel farther to fetch clean water at the community pump, thus connecting them to neighbors with whom they would not otherwise have had frequent contact.

3.8 Future Improvements.

Regarding the application of the proposed framework, this example showed that a correlation model can be created that predicts the direct data given the indirect data. The indirect data can be remotely collected and used to predict the direct data continuously. To improve the predictive power of the correlation model, our next steps would be to:

- Recollect the original data to eliminate the noise and drift challenges experienced with the original dataset
- Collect additional data (if needed)
- Reconstruct the deep learning model using a recurrent neural network learning algorithm (if needed)
- Collect indirect data from an additional source to supplement or replace inertial measurement unit data (if needed)

Individually, each of these steps have potential to improve the accuracy of the correlation model, so it may be only necessary to perform one, such as recollecting the inertial measurement unit data and attenuating noise. The order in which the steps are performed is driven by the resources necessary and available to perform each step.

4 Concluding Remarks

It is beneficial to evaluate a product’s social impact, or how a product affects an individual’s daily quality of life, to ensure the effects of mechanical design are positive and to identify design changes that could be made to new or existing products. Understanding a product’s social impact is made possible through calculating social impact indicators from the collected user data.

The framework presented herein provides a method for predicting the direct data–used to calculate social impact indicators–given the related indirect data. This framework employs correlation models, particularly deep learning models, to correlate information-rich but expensive and infrequently collected user data with raw, inexpensive, and continuously collected user data. The intent of this approach is that the correlation model predicts the direct data given indirect data, thus increasing the quantity and the frequency of the direct data for social impact indicator calculations.

An example is provided that illustrates how this framework can be used in a developing world setting to gather user data related to the social impact indicators of water hand pumps on individuals in Uganda. The example shows how data from an inertial measurement unit connected to the handle of the pump is used to train a deep learning model to predict if the user of the pump is man, woman, or child statured. This predicted direct data can then be used to calculate social impact indicators relevant to gender, conflict and crime, health and safety, paid work, and education social impact categories. The data collected for this example can be found online.[1]

A vital part of deciding how a design should be changed to improve its social impact is to first identify the product's current social impact, which is the focus of this framework. While it can be difficult to determine whether the calculated social impacts of a product are due to its design and not other factors, social impact indicators can nevertheless help designers identify beneficial design changes to the product. Furthermore, the potential for design changes are not limited to that product for which social impact indicator data are collected. For example, in the water pump application discussed in this article, the data collected indicate that much of the time spent using the pump was done by children. This points to at least two possible design changes: (1) a pump handle or handle adapter to make pumping easier for children (related to the product for which the data were measured) and (2) a better method for children to transport the heavy water containers (related to a different product). Other possible design changes indicated by the data collected include redesigning the pump for better ergonomics for pumping by women and designing a self-protection device for individuals who use the pump during high risk hours.

While the application of the proposed framework is only demonstrated for one application in this paper, it could nevertheless be applied to a wide range of situations to assist in the more effective collection of data for calculating the social impact indicators of products. Some potential developing world applications of this framework include the following:

- Measuring social impact indicators related to high-efficiency bee hives for honey farmers in Kenya. In this situation, raw audio data from the hive could be correlated with honey production as reported by farmers in surveys to continuously predict and track honey output and subsequent revenue generated as well as to test various design changes that could improve hive efficiency.
- Measuring social impact indicators related to road quality in India. In this case, GPS and accelerometer data could be correlated with the reported number of vehicle repairs reported by taxi drivers through interviews to identify the economic impact of poor road quality on taxi drivers as well as to assess vehicle suspension design changes that could make them more reliable.
- Measuring social impact indicators related to an electric cassava peeler on rural farmers in Brazil. Accelerometer and power usage data from the peeler could be paired with number of pounds of cassava peeled and subsequent revenue to determine the effectiveness of the peeler compared to manual peeling and to evaluate the effectiveness of various peeler designs.

Another objective of the framework presented in this article is to provide the basis from which additional social impact research questions may be answered including (1) when the validity of the link between the direct and indirect social impact data expires, (2) how frequently the direct data need to be collected to ensure its representation by the indirect data is valid, and (3) which social impacts can be effectively modeled using this approach. These and other follow-on research questions will provide the basis for the future critical work in the area and will benefit from the application of this framework to provide experimental validation.

One challenge of using the framework is that the estimate as to whether the indirect data source(s) will be an accurate predictor of the direct data cannot be determined until after Step 4 of the framework. This challenge should be considered because Steps 1–4 are costly. In addition, the cost of using the framework goes up as the number of indirect data sources increase.

Importantly, we are not suggesting that the indirect data and predicted direct data should completely replace the periodic collection of the direct data. There are invaluable insights gained from the data collected through direct interaction with or observation of users. The predicted direct data are intended to act as supplements to the direct data that are collected at appropriate intervals.

[1]gdi.byu.edu

Acknowledgment

The authors would like to recognize the National Science Foundation for providing the Graduate Research Fellowship Grant 1247046 and Grants CMMI–1632740 and CMMI–1761505 that funded this research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author on reasonable request. The data and information that support the findings of this article are freely available at: gdi.byu.edu.[1] The authors attest that all data for this study are included in the article.

Nomenclature

d = maximum distance of one pump stroke
g = acceleration due to gravity
R = average rate of pay
V = maximum volume of water in a single pump stroke
$e_{[\,]}$ = energy gained or lost during activity [ ] by an individual user
$t_{[\,]}$ = time spent performing activity [ ] by an individual user
$A_{k,l}$ = pump stroke amplitude for individual k and stroke l
$F_{[\,]}$ = fraction of [ ]
$N_{[\,]}$ = number of [ ], used in the upper limit of summations
$T_{[\,]}$ = total time spent performing activity [ ] for the specified user group
$W_{[\,]}$ = potential wages lost while doing activity [ ]
$[\,]_D$ = subscript for education hours (from 8 a.m. to 5 p.m. on weekdays)
$[\,]_H$ = subscript for high risk pumping hours (2 h after sunset and 2 h before sunrise)
$[\,]_P$ = subscript for pump strokes
$[\,]_S$ = subscript for sustenance
$[\,]_U$ = subscript for pump usage activity
i, j, k = counter indices
ρ = density of water

References

[1] United Nations Development Programme Growing Inclusive Markets Initiative, 2008, *Creating Value for All: Strategies for Doing Business With the Poor*, United Nations Development Programme, New York.
[2] Mattson, C. A., and Winter, A. G., 2016, "Why the Developing World Needs Mechanical Design," ASME J. Mech. Des., **138**(7), p. 070301.

[3] Wood, A. E., and Mattson, C. A., 2016, "Design for the Developing World: Common Pitfalls and How to Avoid Them," ASME J. Mech. Des., **138**(3), p. 031101.
[4] George, C., and Shams, A., 2007, "The Challenge of Including Customer Satisfaction Into the Assessment Criteria of Overseas Service-Learning Projects," Int. J. Serv. Learn. Eng., **2**(2), pp. 64–75.
[5] Burdge, R. J., 2004, *A Community Guide to Social Impact Assessment*, Vol. 2, Social Ecology Press.
[6] United Way of America, 1996, Measuring Program Outcomes: A Practical Approach, Alexandria, VA.
[7] Kellogg, W. K., 2006, *WK Kellogg Foundation Logic Model Development Guide*, WK Kellogg Foundation.
[8] Clark, H., and Anderson, A. A., 2004, "Theories of Change and Logic Models: Telling Them Apart," American Evaluation Association Conference, Atlanta, GA, July 2004.
[9] Stevenson, P. D., Mattson, C. A., Bryden, K. M., and MacCarty, N. A., 2018, "Toward a Universal Social Impact Metric for Engineered Products That Alleviate Poverty," ASME J. Mech. Des., **140**(4), p. 041404.
[10] Stevenson, P. D., Mattson, C. A., and Dahlin, E. C., 2020, "A Method for Creating Product Social Impact Models of Engineered Products," ASME J. Mech. Des., **142**(4), p. 041101.
[11] Hutchins, M. J., Gierke, J. S., and Sutherland, J. W., 2009, "Decision Making for Social Sustainability: A Life-Cycle Assessment Approach," International Symposium on Technology and Society, Proceedings, Tempe, AZ, May 2009, IEEE, pp. 1–5.
[12] Fontes, S. J., 2016, *Handbook for Product Social Impact Assessment*, PRe Sustainability, The Netherlands.
[13] Wood, A. E., and Mattson, C. A., 2019, "Quantifying the Effects of Various Factors on the Utility of Design Ethnography in the Developing World," Res. Eng. Des., **30**(3), pp. 317–338.
[14] He, L., Wang, M., Chen, W., and Conzelmann, G., 2014, "Incorporating Social Impact on New Product Adoption in Choice Modeling: A Case Study in Green Vehicles," Transp. Res. Part D: Transp. Environ., **32**(1), pp. 421–434.
[15] World Bank, 2019, "Methodologies—World Bank Data Help Desk." https://datahelpdesk.worldbank.org/knowledgebase/articles/906531-methodologies
[16] Chhipi-Shrestha, G. K., Hewage, K., and Sadiq, R., 2015, "'Socializing' Sustainability: A Critical Review on Current Development Status of Social Life Cycle Impact Assessment Method," Clean Technol. Environ. Policy, **17**(3), pp. 579–596.
[17] Donaldson, K., 2009, "The Future of Design for Development: Three Questions," Inf. Technol. Int. Dev., **5**(4), pp. 97–100.
[18] De Mauro, A., Greco, M., and Grimaldi, M., 2015, "What Is Big Data? A Consensual Definition and a Review of Key Research Topics," AIP Conference Proceedings, Madrid, Spain, Sept. 2014, Vol. 1644, American Institute of Physics Publishing, pp. 97–104.
[19] Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012, "ImageNet Classification With Deep Convolutional Neural Networks," *Neural Information Processing Systems*, Vol. 25, Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., eds., Curran Associates Inc, pp. 1097–1105.
[20] Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., and Sun, Q., 2018, "Deep Learning for Image-Based Cancer Detection and Diagnosis—A Survey," Pattern Recogn., **83**(1), pp. 134–149.
[21] Levianthan, Y., and Matias, Y., 2018, "Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone," Google AI blog, vol. 8.
[22] Yan, R., 2018, "Chitty-Chitty-Chat Bot: Deep Learning for Conversational AI," IJCAI International Joint Conference on Artificial Intelligence, Stockholm, Sweden, July, pp. 5520–5526.
[23] Hossain, Z., Sohel, F., Shiratuddin, M. F., and Laga, H., 2018, "A Comprehensive Survey of Deep Learning for Image Captioning," Comput. Res. Repository, **51**(6).
[24] Bosco, C., Alegana, V., Bird, T., Pezzulo, C., Hornby, G., Sorichetta, A., Steele, J., Ruktanonchai, C., Ruktanonchai, N., Wetter, E., Bengtsson, L., Tatem, A. J., Di Clemente, R., Luengo-Oroz, M., González, M. C., Nielsen, R., Baar, T., and Vacarelu, F., 2017, "Big Data and the Well-Being of Women and Girls Applications on the Social Scientific Frontier," Data2x, **1**(1). Washington, DC.
[25] DrivenData, 2016, "Using Yelp Reviews to Flag Restaurant Health Risks." https://www.drivendata.co/case-studies/using-yelp-reviews-to-flag-restaurant-health-risks/
[26] Ignatov, A., 2018, "Real-Time Human Activity Recognition From Accelerometer Data Using Convolutional Neural Networks," Appl. Soft Comput., **62**(1), pp. 915–922.
[27] Medic.Life, 2019, "MEDIC LAV," https://medic.life/technology/
[28] Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., and Luengo-Oroz, M., 2018, "Humanitarian Applications of Machine Learning With Remote-Sensing Data: Review and Case Study in Refugee Settlement Mapping," Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci., **376**(2128), p. 20170363.
[29] Satellite Imaging Corporation, 2017, "Agriculture Mapping," https://www.satimagingcorp.com/applications/natural-resources/agriculture/
[30] Pandey, S. M., Agarwal, T., and Krishnan, N. C., 2018, "Multi-Task Deep Learning for Predicting Poverty From Satellite Images," The 30th AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18), New Orleans, LA, Feb., IAAI, pp. 7793–7798.
[31] Ulrich, K. T., and Eppinger, S. D., 2008, "Identifying Customer Needs," *Product Design and Development*, Chap. 4, McGraw-Hill, New York, p. 54.
[32] Rainock, M., Everett, D., Pack, A., Dahlin, E. C., and Mattson, C. A., 2018, "The Social Impacts of Products: A Review," Impact Assess. Project Appraisal, **36**(3), pp. 230–241.
[33] Pack, A. T., Phipps, E. R., Mattson, C. A., and Dahlin, E. C., 2018, "Social Impact in Product Design: An Exploration of Current Industry Practices," Volume 2A: 44th Design Automation Conference, Quebec City, Canada, Aug., ASME, p. V02AT03A049.
[34] Ottosson, H. J., Mattson, C. A., and Dahlin, E. C., 2020, "Analysis of Perceived Social Impacts of Existing Products Designed for the Developing World, With Implications for New Product Development," ASME J. Mech. Des., **142**(5), p. 051101.
[35] Silva, S., and Gonçalves, I.and Sara, S., 2013, "Balancing Learning and Overfitting in Genetic Programming With Interleaved Sampling of Training Data," *European Conference on Genetic Programming*, pp. 73–84.
[36] Delea, M. G., Nagel, C. L., Thomas, E. A., Halder, A. K., Amin, N., Shoab, A. K., Freeman, M. C., Unicomb, L., and Clasen, T. F., 2017, "Comparison of Respondent-Reported and Sensor-Recorded Latrine Utilization Measures in Rural Bangladesh: A Cross-Sectional Study," Trans. R. Soc. Tropical Med. Hygiene, **111**(7), pp. 308–315.
[37] Guerra, L., McGarry, L. M., Robles, V., Bielza, C., Larrañaga, P., and Yuste, R., 2011, "Comparison Between Supervised and Unsupervised Classifications of Neuronal Cell Types: A Case Study," Dev. Neurobiol., **71**(1), pp. 71–82.
[38] Goodfellow, I., Bengio, Y., and Courville, A., 2016, *Deep Learning*, MIT Press. http://www.deeplearningbook.org.
[39] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., 2002, "SMOTE: Synthetic Minority Over-Sampling Technique," J. Artif. Intell. Res., **16**(1), pp. 321–357.
[40] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P., 2017, "Focal Loss for Dense Object Detection," Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, July 2016, pp. 2980–2988.
[41] Das, B., Krishnan, N. C., and Cook, D. J., 2013, "Handling Class Overlap and Imbalance to Detect Prompt Situations in Smart Homes," 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, Dec., IEEE, pp. 266–273.
[42] Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., and Zhou, Y., 2017, "Deep Learning Scaling is Predictable, Empirically," arXiv preprint arXiv:1712.00409, pp. 1–19.
[43] Jiménez, A., and Pérez-Foguet, A., 2010, "Challenges for Water Governance in Rural Water Supply: Lessons Learned From Tanzania," Int. J. Water Resour. Dev., **26**(2), pp. 235–248.
[44] Makoni, F. S., Manase, G., and Ndamba, J., 2004, "Patterns of Domestic Water Use in Rural Areas of Zimbabwe, Gender Roles and Realities," Phys. Chem. Earth, Parts A/B/C, **29**(15–18), pp. 1291–1294.
[45] Thompson, J. A., Folifac, F., and Gaskin, S. J., 2011, "Fetching Water in the Unholy Hours of the Night: The Impacts of a Water Crisis on Girls' Sexual Health in Semi-Urban Cameroon," Girlhood Stud., **4**(2), pp. 111–129.
[46] World Health Organization, 2017, "Progress on Drinking Water, Sanitation and Hygiene," World Health Organization, **1**(1). Geneva, Switzerland.
[47] World Bank, 2003, "Morocco—Rural Water Supply and Sanitation Project," World Bank, **1**(1). Washington, DC.
[48] Assaad, R., Levison, D., and Zibani, N., 2010, "The Effect of Domestic Work on Girls' Schooling: Evidence From Egypt," Feminist Econ., **16**(1), pp. 79–128.
[49] Sorenson, S. B., Morssink, C., and Campos, P. A., 2011, "Safe Access to Safe Water in Low Income Countries: Water Fetching in Current Times," Soc. Sci. Med., **72**(9), pp. 1522–1526.
[50] Pommells, M., Schuster-Wallace, C., Watt, S., and Mulawa, Z., 2018, "Gender Violence as a Water, Sanitation, and Hygiene Risk: Uncovering Violence Against Women and Girls as It Pertains to Poor WaSH Access," Violence Against Women, **24**(15), pp. 1851–1862.
[51] Kakooza, A., 2018, "Schools' and Other Institutions' Calendar—2019," Uganda Ministry of Education & Sports, **1**(1).
[52] The World Bank, 2017, "Compensation of Employees (Current LCU)—Uganda—Data," https://data.worldbank.org/indicator/GC.XPN.COMP.CN?locations=UG
[53] Meeker, M., 2018, "Internet Trends 2018," 2018 Code Conference, Rancho Palos Verdes, CA, May 2018, p. 25.
[54] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L., 2015, "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis., **115**(3), pp. 211–252.
[55] JMP, Version 13, SAS Institute Inc., Cary, NC, 1989-2019.
[56] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., 2017, "Automatic Differentiation in Pytorch," https://openreview.net/pdf?id=BJJsrmfCZ
[57] Google, 2018, "Colaboratory: Frequently Asked Questions," https://research.google.com/colaboratory/faq.html, Accessed June 21, 2018.
[58] Kingma, D. P., and Ba, J., 2014, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980.
[59] Zhang, Z., and Sabuncu, M. R., 2018, "Generalized Cross Entropy Loss for Training Deep Neural Networks With Noisy Labels," Neural Information Processing Systems, Montreal, Canada, pp. 8778–8788.
[60] Rural Water Supply Network, 2019, "Implementation – Handpump Technology, https://www.rural-water-supply.net/en/implementation/public-domain-handpumps/india-mark-ii.
[61] Institute of Medicine of the National Academies, 2002, "Panel on Macronutrients, Panel on the Definition of Dietary Fiber," National Academies of Sciences, **1**(1). Washington, DC.
[62] Campos, F., Goldstein, M., Mcgorman, L., Munoz Boudet, A. M., and Pimhidzai, O., 2015, "Breaking the Metal Ceiling Female Entrepreneurs Who Succeed in Male-Dominated Sectors," World Bank, https://www.econstor.eu/bitstream/10419/190011/1/wp2017-166.pdf