



OPEN ACCESS

EDITED BY

Joseph Okeibunor,
World Health Organization - Regional Office for
Africa, Republic of Congo

REVIEWED BY

Mohsin Sidat,
Eduardo Mondlane University, Mozambique
Delanyo Dovlo,
Ghana College of Physicians and Surgeons,
Ghana

*CORRESPONDENCE

Alma Fredriksson
fredriksson.alma@gmail.com

SPECIALTY SECTION

This article was submitted to Health
Technology Implementation, a section of the
journal Frontiers in Digital Health

RECEIVED 14 January 2022

ACCEPTED 25 July 2022

PUBLISHED 17 August 2022

CITATION

Fredriksson A, Fulcher IR, Russell AL, Li T,
Tsai Y-T, Seif SS, Mpembeni RN and
Hedt-Gauthier B (2022) Machine learning for
maternal health: Predicting delivery location in
a community health worker program in
Zanzibar.
Front. Digit. Health 4:855236.
doi: 10.3389/fdgth.2022.855236

COPYRIGHT

© 2022 Fredriksson, Fulcher, Russell, Li, Tsai,
Seif, Mpembeni and Hedt-Gauthier. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar

Alma Fredriksson^{1*}, Isabel R. Fulcher^{2,3}, Allyson L. Russell⁴,
Tracey Li⁴, Yi-Ting Tsai¹, Samira S. Seif⁴, Rose N. Mpembeni⁵
and Bethany Hedt-Gauthier^{1,2}

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States,

²Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United

States, ³Harvard Data Science Initiative, Cambridge, MA, United States, ⁴D-tree International, Dar es

Salaam, Tanzania, ⁵Department of Epidemiology and Biostatistics, Muhimbili University of Health and

Allied Sciences, Dar es Salaam, Tanzania

Background: Maternal and neonatal health outcomes in low- and middle-income countries (LMICs) have improved over the last two decades. However, many pregnant women still deliver at home, which increases the health risks for both the mother and the child. Community health worker programs have been broadly employed in LMICs to connect women to antenatal care and delivery locations. More recently, employment of digital tools in maternal health programs have resulted in better care delivery and served as a routine mode of data collection. Despite the availability of rich, patient-level data within these digital tools, there has been limited utilization of this type of data to inform program delivery in LMICs.

Methods: We use program data from 38,787 women enrolled in Safer Deliveries, a community health worker program in Zanzibar, to build a generalizable prediction model that accurately predicts whether a newly enrolled pregnant woman will deliver in a health facility. We use information collected during the enrollment visit, including demographic data, health characteristics and current pregnancy information. We apply four machine learning methods: logistic regression, LASSO regularized logistic regression, random forest and an artificial neural network; and three sampling techniques to address the imbalanced data: undersampling of facility deliveries, oversampling of home deliveries and addition of synthetic home deliveries using SMOTE.

Results: Our models correctly predicted the delivery location for 68%–77% of the women in the test set, with slightly higher accuracy when predicting facility delivery versus home delivery. A random forest model with a balanced training set created using undersampling of existing facility deliveries accurately identified 74.4% of women delivering at home.

Conclusions: This model can provide a “real-time” prediction of the delivery location for new maternal health program enrollees and may enable early provision of extra support for individuals at risk of not delivering in a health facility, which has potential to improve health outcomes for both mothers and their newborns. The framework presented here is applicable in other contexts and the selection of input features can easily be adapted to match data availability and other outcomes, both within and beyond maternal health.

KEYWORDS

maternal health, machine learning, digital health, global health, facility delivery, community health worker intervention, artificial intelligence

Introduction

Maternal and neonatal health outcomes in low- and middle-income countries (LMICs) have improved over the last two decades, in part due to increasing access to facilities for delivery and the quality of care therein (1–3). Facility-based delivery facilitates timely access to emergency obstetric care, and higher rates of facility delivery in sub-Saharan Africa (SSA) are significantly associated with lower maternal mortality (4–6). Between 2010 and 2018, it is estimated that nearly 67% of women in SSA delivered in health facilities (7); however, many pregnant women still deliver at home in LMICs (8), increasing the risk of poor outcomes for both the mother and the child should a complication occur during childbirth. Strategies to identify and target interventions for these women could help close the remaining gap.

Previous research exploring the likelihood of facility-based delivery in SSA has found that attendance of at least one antenatal care visit, proximity to a health facility, difficulty delivering in the past, lower parity, higher maternal education and partner involvement reduces the likelihood of home delivery (9–14). Further studies suggest that local tradition and beliefs, cost of delivery, household income, perceived quality of care and fear of discrimination are also important factors influencing the choice of delivery location (15–20). This literature provides crucial insights into strong predictors of health facility delivery but does not maximize prediction accuracy for the purposes of real-time programmatic use.

The uptick of digital tools in maternal health programs, such as mobile devices for community health workers (CHWs) (21–26), presents a unique opportunity for real-time implementation of predictive models (23). Digital tools capture data of individuals enrolled in these programs that can be used to build prediction models (24); specifically, for maternal health, machine learning techniques can capture complex, non-linear relationships between demographic or clinical characteristics of the woman and birth outcomes (27–29). These machine learning models could then be applied back into the program to guide the interventions in a more targeted, efficient manner. To date, few published papers have used machine learning techniques to predict maternal or child outcomes in SSA (26, 30–37), only one of which focused on facility-based delivery (31). Of note, this latter example from Ethiopia used national survey data and not data captured on individuals enrolled in a maternal health program.

To address these gaps, we applied several machine learning techniques to predict health facility delivery among pregnant women enrolled in the maternal health program, *Uzazi Salama*, in Zanzibar, Tanzania (38). We used existing program data, captured at the time of enrollment by CHWs equipped with a smartphone application, to build these models. In this paper, we first provide an overview of the *Uzazi Salama* program and data collection followed by rationale for choice of variables and machine learning models. We conclude with a discussion on the benefits of integrating prediction models within *Uzazi Salama* or similar maternal health programs with the goal of improving rates of facility delivery.

Materials and methods

Program overview

The *Uzazi Salama* (“Safer Deliveries”) program operated between 2016 and 2019 in Unguja and Pemba, the two main islands of Zanzibar. In January 2020, the program was adopted by the Government of Zanzibar as the national community health program, expanded to provide additional postpartum and child health services, and was renamed *Jamii ni Afya* (“Communities are Health”). The goal of the *Uzazi Salama* program was to increase the number of women delivering in a health facility by providing birth planning, counseling, and care referrals via in-person home visits with a community health volunteer. Henceforth, we will refer to community health volunteers as CHWs to be consistent with the literature. CHWs provided three scheduled health promotion home visits prior to delivery. CHWs used a program-specific cell phone application to guide the encounter, which also served as a real-time data collection tool. Further details regarding the *Uzazi Salama* program are available in [Supplementary Material](#).

Study population

We included women enrolled in *Uzazi Salama* between 1 January 2017–30 June 2019 with a recorded delivery in this analysis. Women had to deliver at a health facility or at home to be included. Women who delivered on the way to the facility and women with no birth location documented were excluded. A total of 38,787 women met the inclusion criteria and were considered in the analysis.

Key variables

The models presented here depend on individual- and community-level variables. All analyses are based *strictly* on information collected by CHWs prior to or during the enrollment visit. The considered variables are as follows: Woman's demographic characteristics included district of residence, age, and the following six socioeconomic indicators: whether all children are currently living, electricity in home, drinking water source, roof material, floor material and education level. CHWs collected these six socioeconomic indicators beginning in March 2017 during in-person home visits occurring one week postpartum. Although this information was collected later, it is assumed that the answers apply to the enrollment visit and in the new national program this data is being collected at the first visit. Self-reported clinical characteristics included details on (a) comorbidities (HIV status, cardiac disease, diabetes, high blood pressure, sickle cell anemia) and (b) past pregnancies as relevant (parity category (0, 1–2, 3–4, 5–7, 8+), previous delivery location, eclampsia, perineal tear, placenta previa, prolonged labor, retained placenta, vacuum, cesarean section, spontaneous abortions or stillbirths).

Current pregnancy characteristics included whether it was a twin pregnancy, whether the baby was larger than average (macrosomia), and the amount of money the family had currently saved for the delivery. It also included the total number of antenatal care (ANC) visits prior to the enrollment visit. If the first ANC visit took place prior to enrollment, information on the occurrence of convulsions, abdominal pain, bleeding, headache or difficulty breathing was recorded, along with whether the mother was charged a fee and whether the first ANC visit was a “negative experience”, defined as having both a privacy interruption and a wait time of more than 3 h. We also considered if the woman was referred back to a health facility by the CHW at the first visit and, if so, the reason for referral. Program-associated characteristics, also collected at the enrollment visit, included partner permission to deliver in a health facility (which required the partner to be present at one of the CHW visits, else not received), estimated gestational age at enrollment, type of recommended delivery facility, taxi price to delivery facility as a distance proxy, whether the woman had her reproductive and child health card available during the visit. Name of the recommended delivery facility was included as a proxy for perception of quality and to account for regional differences in the patient population.

Finally, community-level characteristics were created for inclusion in the model. For each woman, the following variables were calculated based on program-level information up to the time of the enrollment visit: cumulative level of experience of the woman's CHW measured in 6-month increments, rate of facility delivery in

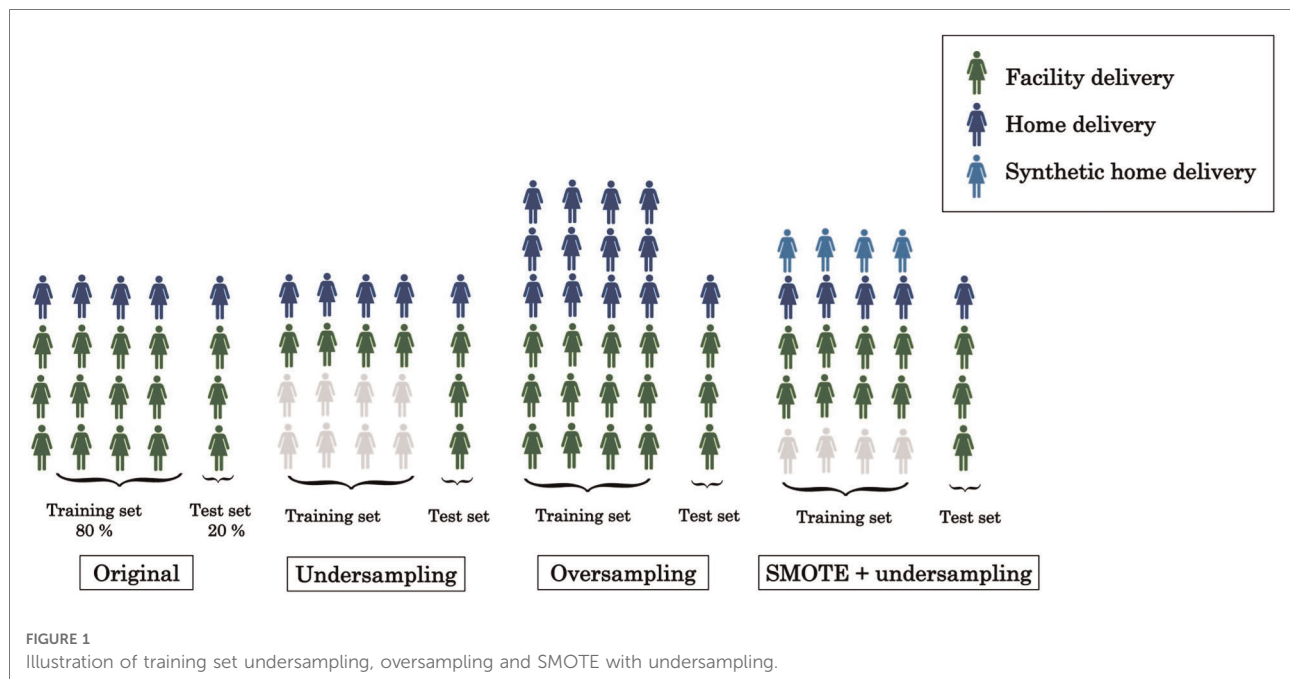
mother's shehia (local area), and rate of facility delivery among women with the same CHW. These were derived from the original data such that the rates were only based on women who delivered *prior* to the enrollment date for the woman of interest, including those enrolled in the program prior to 1 January 2017. Women living in the same shehia sometimes had the same CHW, resulting in some correlation between these variables.

Data preprocessing

Some women were missing information on savings (10%) and the six socioeconomic indicators (33%) because these variables were added to the smartphone app after the start of the program and because information was collected at a specific postpartum home visit, which some women may have missed. Data availability was high (>97%) across all other variables. Missing values in nominal variables were imputed with the most frequent category. *K*-Nearest Neighbor (KNN) imputation (39), which uses observations from *K* “similar” women in the dataset to impute a plausible value for the woman with a missing variable, was used to impute all other missing values. *K* = 1 was used for ordinal variables and *K* = 5 for all other variables.

We created two datasets: 80% of the women were randomly selected, stratified by home or facility delivery, for the *training set* to be used for model development and the remaining 20% were used as a *test set* after completion of model development. In both the training and test sets, 77% of women had delivered at a health facility, creating an outcome imbalance which could cause some prediction models to prioritize predicting all or most observations as facility deliveries. Three different techniques were employed to address this imbalance: undersampling, oversampling and a synthetic minority oversampling technique (SMOTE) (40). First, we undersampled from the facility deliveries in the training set by randomly selecting a subset of the same size as the number of home deliveries, as illustrated in **Figure 1**. This approach does not make full use of the information available but avoids the issue of poorer classification of home deliveries directly due to the imbalance of delivery facility types.

Second, we used oversampling from the minority class (home deliveries), whereby some observations occur multiple times in the training set. This increases the risk of overfitting but makes full use of the information in the sample. Finally, we also approached the imbalance issue using SMOTE, which creates synthetic records that are “similar” to actual records from women who delivered at home (see Section 4.2 of (40) for further technical details). To avoid relying too heavily on synthetic records, without compromising on the outcome balance, we simultaneously undersampled from the facility deliveries.



Statistical modeling

Prior to model development, we summarize the characteristics of women captured in the training dataset and use the chi-square statistic to identify the ten factors most associated with facility delivery. Continuous features (age, gestational age, taxi price to delivery facility, delivery savings, payment fee at first ANC visit, rate of facility delivery in shehia, rate of facility delivery among women with same CHW) were discretized for this calculation, but were left continuous for the prediction modeling.

We then used four different models, comprising two parametric and two non-parametric approaches, with varying levels of regularization, given that prior knowledge about the exact relationship between the outcome and the predictors was very limited. As a baseline model, we used multivariate logistic regression, assuming a linear relationship between the covariates and the log-odds of facility delivery. We also considered a logistic model using LASSO regularization (41), a means of feature selection to shrink the coefficients of “less predictive” covariates towards 0.

In addition, two non-parametric models were used: a random forest, which calculates an average prediction across a large set of decision trees based on bootstrapped samples and random subsets of features, and an artificial neural network, which consists of layers of neurons that each look specifically at certain features of data. The tree depth for the random forest was selected using cross validation to reduce the risk of overfitting to the training set and making insufficient use of the information available in the data.

While the logistic and the regularized logistic models are easily interpretable and the latter provides insight into variable importance (41), the random forest model is more flexible and may better capture complex non-linear relationships between the variables and the outcome (29) while offering a measure of relative variable importance (42). Deep learning models typically offer additional flexibility, but require larger sample sizes, are more computationally expensive, and offer limited interpretability on variable importance (43).

Model evaluation

We used multiple metrics to assess how well the test set predictions matched the true delivery locations. The baseline metric used was overall test set accuracy, the percent of women in the test set that were correctly classified. Given the skewed test set, with only ~23% being home deliveries, the individual performance on the minority and majority class was also measured using true positive rate (sensitivity), which measures the proportion of home deliveries that are classified correctly and true negative rate (specificity), which measures the proportion of facility deliveries that are classified correctly. Finally, the area under the receiver operating characteristic curve (AUC) is a measure that jointly accounts for true and false positives, thus offering an appropriate method for comparison among models with similar performance across the other metrics. We computed the chi-square statistic to estimate which predictors best discriminated between facility

and home deliveries. All analysis was conducted using R version 3.6.1 and Python packages scikit-learn 0.23.1 and tensorflow 2.4.1. The code used for model training and testing is available in the GitHub repository [alfredriksson/facility-delivery](https://github.com/alfredriksson/facility-delivery).

Results

Among the 31,435 women in the training set, 76.6% delivered at a health facility and 23.4% delivered at home after taking part in Uzazi Salama. A summary of the demographic, programmatic and self-reported clinical characteristics for the full dataset is presented in **Tables 1, 2**. We find that the 10 predictors that *individually* best discriminate between home and facility deliveries are the following, in order of magnitude of association:

- (1) Previous delivery location

TABLE 1 Demographic characteristics (n = 38,787).

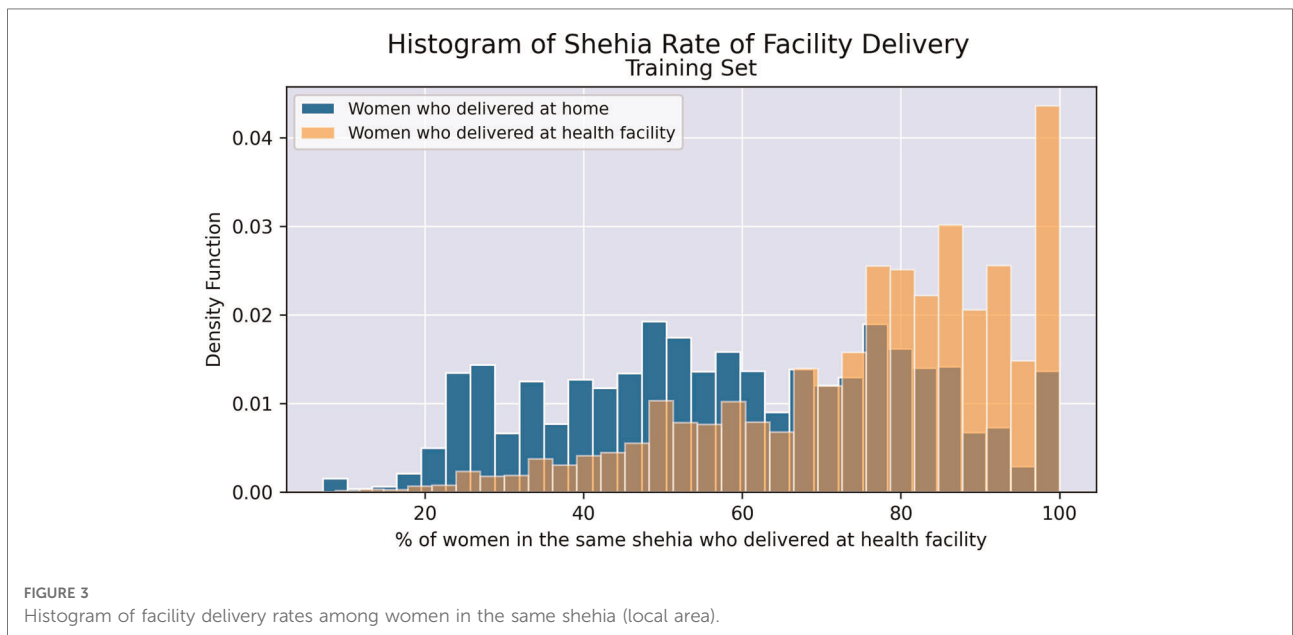
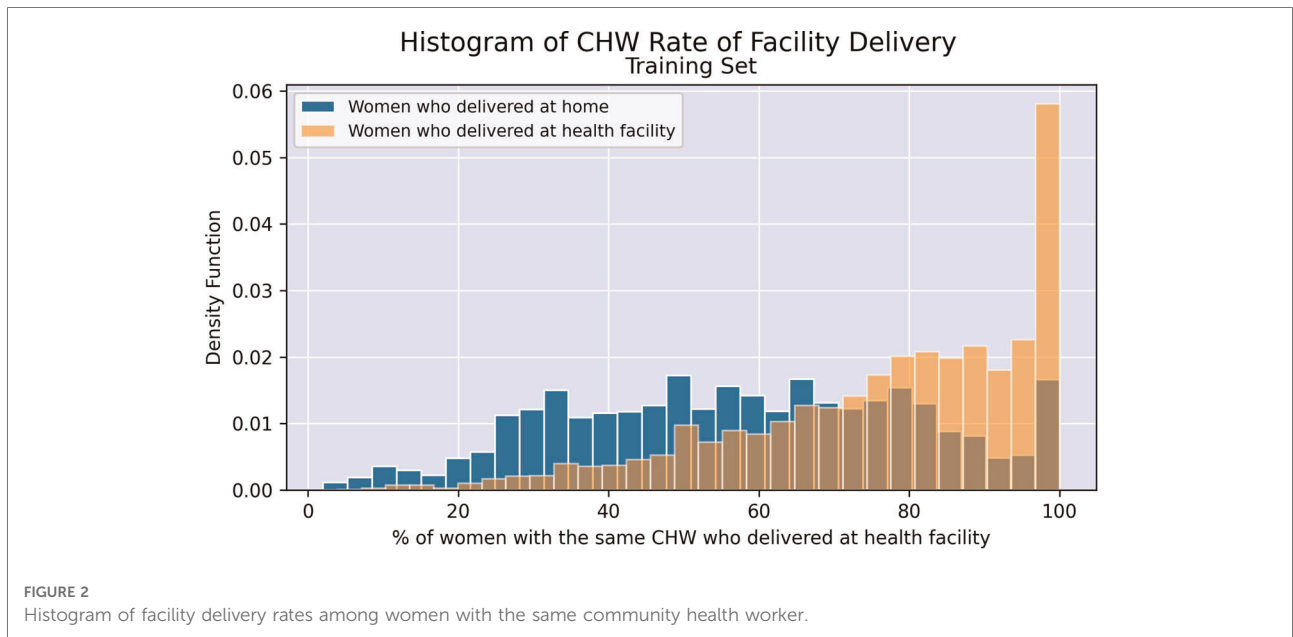
Variable	N (%)
Overall	38787 (100%)
Delivery location	
Health facility	29679 (76.5%)
Home	9108 (23.5%)
District	
Kaskazini A	5799 (15.0%)
Kaskazini B	2831 (7.3%)
Kati	3529 (9.1%)
Magharibi	4853 (12.5%)
Kusini	1861 (4.8%)
Mkoani	6135 (15.8%)
Wete	3917 (10.1%)
Micheweni	4471 (11.5%)
Chake Chake	5391 (13.9%)
Age of mother	
10–20	5730 (14.8%)
21–30	22406 (57.9%)
31–40	9710 (25.1%)
41+	868 (2.2%)
Parity	
0	8916 (23.0%)
1	6648 (17.2%)
2–4	14980 (38.7%)
5+	8168 (21.1%)
Previous delivery location	
At home/in community	7946 (20.5%)
On the way to health facility	338 (0.9%)
Health facility	21514 (55.6%)
No previous delivery	8916 (23.0%)

- (2) Rate of facility delivery among women with the same CHW
- (3) Rate of facility delivery in shehia
- (4) Name of recommended delivery facility
- (5) Residential district
- (6) Parity
- (7) Education level
- (8) Floor material
- (9) Access to electricity in home
- (10) Payment fee charged at first ANC visit

Figures 2, 3 illustrate the difference in distribution between women who delivered in a facility and women who delivered at home for key continuous predictors above. Further details regarding test statistics for all top variables and figures visualizing the difference in socioeconomic distribution between women who delivered at a health facility and women who delivered at home are available in **Supplementary Material**.

TABLE 2 Programmatic and self-reported clinical characteristics (N = 38,787).

Variable	N (%)
Overall	38787 (100%)
Gestational age at enrollment (weeks)	
0–10	1599 (4.1%)
11–20	15231 (39.3%)
21–30	18612 (48.0%)
31–40	3331 (8.6%)
Partner permission to enroll	
Yes	16881 (43.5%)
No	21906 (56.5%)
Taxi price to delivery facility (TZS)	
0–10,000	14101 (36.6%)
10,001–20,000	13573 (35.3%)
20,001–30,000	7280 (18.9%)
30,000+	3544 (9.2%)
CHW experience level (months)	
0–6	8567 (22.3%)
7–12	9176 (23.9%)
13–18	8779 (22.8%)
19–24	7394 (19.2%)
25+	4522 (11.8%)
HIV status	
Positive	443 (1.2%)
Negative	37367 (96.3%)
Unknown	977 (2.5%)
Cardiac disease	
Yes	52 (0.1%)
No	38662 (99.9%)
High blood pressure	
Yes	132 (0.3%)
No	38582 (99.7%)



Undersampling

After performing undersampling, the training set consisted of 14,572 women; 7,286 women from each outcome class. The logistic regression model, its regularized counterpart and the random forest model achieve an overall classification accuracy of 72%–74% on the test set (see **Table 3**). While these three models have a slight difference in accuracy across the two outcome classes, the neural network shows a major performance gap, where facility deliveries are more

accurately predicted than home deliveries. The random forest model shows the most promising performance on the home deliveries with correct classification in more than 74% of the cases. The associated variable importance measure, reflecting how often a variable is used in the condition that splits a tree branch into two, suggested that the rate of facility delivery among women in the same shehia or with the same CHW are key predictors. As previously noted, there is some correlation between these variables as women in the same shehia may have the same CHW.

TABLE 3 Model performance on test set by training set type.

Training set	Classifier	True positive rate ^a	True negative rate ^b	Overall accuracy	AUC
Undersampled $N = 14,572$	Logistic	71.7%	74.0%	73.5%	0.801
	Regularized Logistic	71.0%	74.5%	73.7%	0.799
	Random Forest	74.4%	71.8%	72.4%	0.800
	Neural Network	56.1%	80.3%	74.6%	0.744
Oversampled $N = 47,486$	Logistic	71.1%	74.5%	73.7%	0.802
	Regularized Logistic	71.0%	74.6%	73.8%	0.802
	Random Forest	71.1%	73.8%	73.1%	0.799
	Neural Network	75.1%	65.9%	68.1%	0.772
SMOTE with minor undersampling $N = 30,000$	Logistic	71.8%	74.0%	73.5%	0.801
	Regularized Logistic	71.1%	74.5%	73.7%	0.802
	Random Forest	68.6%	76.5%	74.6%	0.798
	Neural Network	58.9%	80.8%	75.7%	0.778

^aProportion of home deliveries that are classified correctly.

^bProportion of facility deliveries that are classified.

Oversampling

After performing oversampling, the training set consisted of 47,486 women; 23,743 women who had delivered at a facility, 7,368 unique observations from women who delivered at home, and 16,735 observations from women who delivered at home that were sampled from the 7,368 unique observations. The neural network shows a promising true positive rate and correctly identifies more than 75% of home deliveries (see [Table 3](#)). The logistic and regularized logistic models both achieve an overall accuracy of close to 74%. The similar performance across the two logistic models suggests that little or no feature selection occurred.

SMOTE with minor undersampling

After performing SMOTE with minor undersampling, the training set consisted of 30,000 women; 15,000 women who had delivered at a facility, 7,368 unique observations from women who delivered at home, and 7,632 synthetic observations similar to women who delivered at home. The unregularized logistic and regularized logistic regression models trained on synthetic records performed well with true positive rates just below 72% (see [Table 3](#)), but generally predicted facility deliveries more accurately than home deliveries.

Further details about the false positive rate and false negative rate of each model are available in [Supplementary Material](#).

Discussion

Our paper is one of few using machine learning approaches to predict facility delivery in SSA. Across varying modeling and balancing techniques, we found the accuracy to be 68%–77%,

with slightly higher accuracy when predicting facility delivery versus home delivery. One other paper has previously shown the utility of these types of models and successfully predicted skilled birth attendance in Ethiopia using retrospectively collected survey data (31). The model we present operates on program data collected at enrollment, including community-level factors, which enables prospective implementation and integration into existing maternal health programs. It could also be used in real-time to assess the risk of home delivery and provide tailored intervention. We found education, age and parity to be important, which corroborates findings from other papers (9–14). However, we also found community-level variables to be highly predictive, highlighting their importance for inclusion in the prediction models.

For all but two combinations of modeling and balancing techniques, the performance was higher when predicting facility delivery, which is consistent with the fact that all the models based on oversampling and synthetic records have less real information about home deliveries available regardless of the balanced training set. Importantly, the classification used in this paper is optimized for overall prediction accuracy, but in practice, the classification thresholds could be altered manually to increase or reduce the sensitivity and specificity of the algorithm according to maternal health program priorities and the cost of specific interventions used. Across the undersampled and the oversampled training set, the random forest model and the neural network respectively obtain the highest true positive rate or, equivalently, shows the strongest performance on home deliveries (74.4% and 75.1% respectively). The former achieves this without compromising noticeably on the performance for facility deliveries. For the partly synthetic training set, the logistic regression model achieves the highest classification accuracy on the home deliveries at 71.8%. Among these three candidate models, the undersampled random forest is the only one to show strong performance on the home deliveries, which we are particularly interested in identifying, without noticeably

sacrificing performance on the facility deliveries. This strong performance of the non-parametric random forest model relative to the parametric methods suggests that there may be non-linear relationships present that the traditional parametric models may not fully capture.

While the training data is specifically obtained from *Uzazi Salama*, the framework is applicable in other contexts and the selection of input features can easily be adapted to match data availability and other outcomes, both within and beyond maternal health. The insights obtained here will be leveraged to predict delivery location for pregnant women who are enrolled in the *Jamii ni Afya* program. We hypothesize that the prediction accuracy will be improved as socioeconomic data will be collected for all individuals. Extensions of this work include updating predictions after additional data is collected at future CHW home visits, which would allow for multi-stage program tailoring.

There are several limitations to our findings, including missing data, which is a common issue in these contexts. In the *Uzazi Salama* data, we had missingness in several variables, primarily among the socioeconomic indicators. We used the KNN algorithm to impute values, given that it generally performs well compared to similar techniques and the risk of model misspecification is low, as it is fully non-parametric (40). However, applying it to a dataset with a vast number of predictors increases the risk of relying on features with limited relevance to the outcome, which can add randomness to the distance measure used to select neighbors and subsequently reduce the quality of the imputation (44). We further recognize that although several socioeconomic indicators were included in the analysis, there may be nuances of each woman's socioeconomic status that have not been fully captured.

Similar to most maternal health programs, *Uzazi Salama* also experienced loss to follow-up. Approximately 14% of enrollees did not have a delivery outcome recorded and were excluded from the analysis, as we did not feel comfortable imputing these. We hypothesize that women who are lost to follow-up have either traveled to another location, such as a family home, to deliver and CHWs were not able to reach them or they are hesitant to tell their CHW that they delivered at home. However, as this study does not claim causality, the main concern of this potential data bias is primarily reduced prediction accuracy rather than confounding.

Conclusion

We compared four machine learning models and three sampling techniques to display the utility of using routinely collected maternal health program data to predict delivery location. Our models correctly captured the delivery location for 68%–77% of the women in our test set, with slightly higher accuracy when predicting facility delivery versus home delivery. We demonstrate that prediction models can effectively be used

to obtain a “real-time” prediction of the delivery location for new maternal health program enrollees. Embedded in a cell phone application, this model can be leveraged to immediately alert the CHW to individuals at particularly high risk of delivering at home, allowing for provision of extra support or further visits, which has potential to improve health outcomes for both mothers and children. While our training data originates from the *Uzazi Salama*, the framework is applicable in other contexts and the selection of input features can easily be adapted to match data availability and other outcomes, both within and beyond maternal health.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The dataset presented in this article is not readily available as it contains personal health data and is subject to data privacy regulation. Requests to access these datasets should be directed to Tracey Li, tli@d-tree.org.

Ethics statement

The studies involving human participants were reviewed and approved by the Harvard Human Research Protection Program on May 24, 2017 (IRB16–1024). Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

AF conducted data analysis and led manuscript creation. IF, AR, BHG were responsible for conceptualization and methodology. IF, BHG supervised the project. TL, YT, RM, SS helped with interpretation of data and results. All authors contributed to the article and approved the submitted version.

Funding

This study did not receive any funding from an external source. The first and second authors were respectively supported by the Rose Traveling Fellowship Program in Chronic Disease Epidemiology and Biostatistics and the Harvard Data Science Initiative Postdoctoral Fellows Program. The data was collected as a byproduct of a global health program made possible through the generous support of the Saving Lives at Birth partners: the United States Agency for International Development (USAID), the Government of Norway, the Bill & Melinda Gates Foundation, Grand

Challenges Canada, and the UK Government. It was prepared by the authors and does not necessarily reflect the views of the Saving Lives at Birth partners.

Acknowledgments

We would like to thank the Zanzibar Ministry of Health for their collaboration in the *Uzazi Salama* and *Jamii ni Afya* programs.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Thomas TN, Gausman J, Lattof SR, Wegner MN, Kearns AD, Langer A. Improved maternal health since the ICPD: 20 years of progress. *Contraception*. (2014) 90(6):S32–8. doi: 10.1016/j.contraception.2014.06.026
2. Kalipeni E, Iwelunmor J, Grigsby-Toussaint D. Maternal and child health in Africa for sustainable development goals beyond 2015. *Glob Public Health*. (2017) 12(6):643–7. doi: 10.1080/17441692.2017.1304622
3. UNICEF. Trends in Maternal Mortality: 2000 to 2017 (2019). Available at: <https://data.unicef.org/resources/trends-maternal-mortality-2000-2017/> (Accessed November 8, 2021).
4. Moyer CA, Dako-Gyeke P, Adanu RM. Facility-based delivery and maternal and early neonatal mortality in sub-saharan Africa: a regional review of the literature. *Afr J Reprod Health*. (2013) 17(3):30–43.
5. Campbell OM, Graham WJ, Lancet Maternal Survival Series steering group. Strategies for reducing maternal mortality: getting on with what works. *Lancet*. (2006) 368(9543):1284–99. doi: 10.1016/S0140-6736(06)69381-1
6. Chinkhumba J, De Allegri M, Muula AS, et al. Maternal and perinatal mortality by place of delivery in sub-saharan Africa: a meta-analysis of population-based cohort studies. *BMC Public Health*. (2014) 14. doi: 10.1186/1471-2458-14-1014. [Epub ahead of print]
7. Adde KS, Dickson KS, Amu H. Prevalence and determinants of the place of delivery among reproductive age women in sub-saharan Africa. *PLoS One*. (2020) 15(12):e0244875. doi: 10.1371/journal.pone.0244875
8. Doctor HV, Nkhana-Salimu S, Abdulsalam-Anibilowo M. Health facility delivery in sub-saharan Africa: successes, challenges, and implications for the 2030 development agenda. *BMC Public Health*. (2018) 18(1):765. doi: 10.1186/s12889-018-5695-z
9. Rockers PC, Wilson ML, Mbaruku G, Kruk ME. Source of antenatal care influences facility delivery in rural Tanzania: a population-based study. *Matern Child Health J*. (2009) 13(6):879–85. doi: 10.1007/s10995-008-0412-7
10. Anyait A, Mukanga D, Oundo GB, Nuwaha F. Predictors for health facility delivery in busia district of Uganda: a cross sectional study. *BMC Pregnancy Childbirth*. (2012) 12(1):1–9. doi: 10.1186/1471-2393-12-132
11. Mochache V, Lakhani A, El-Busaidy H, Temmerman M, Gichangi P. Correlates of facility-based delivery among women of reproductive age from the digo community residing in kwale, Kenya. *BMC Res Notes*. (2018) 11(1):715. doi: 10.1186/s13104-018-3818-3
12. Boah M, Mahama AB, Ayamga EA. They receive antenatal care in health facilities, yet do not deliver there: predictors of health facility delivery by women in rural Ghana. *BMC Pregnancy Childbirth*. (2018) 18(1):125. doi: 10.1186/s12884-018-1749-6
13. Moyer CA, Mustafa A. Drivers and deterrents of facility delivery in sub-saharan Africa: a systematic review. *Reprod Health*. (2013) 10:40. doi: 10.1186/1742-4755-10-40

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdgth.2022.855236/full#supplementary-material>.

14. Moshi F, Nyamhanga T. Understanding the preference for homebirth; an exploration of key barriers to facility delivery in rural Tanzania. *Reprod Health*. (2017) 14(1):132. doi: 10.1186/s12978-017-0397-z
15. Huda TM, Chowdhury M, El Arifeen S, Dibley MJ. Individual and community level factors associated with health facility delivery: a cross sectional multilevel analysis in Bangladesh. *PLoS One*. (2019) 14(2):e0211113. doi: 10.1371/journal.pone.0211113
16. Cofie LE, Barrington C, Singh K, Sodzi-Tettey S, Ennett S, Maman S. Structural and functional network characteristics and facility delivery among women in rural Ghana. *BMC Pregnancy Childbirth*. (2017) 17(1):425. doi: 10.1186/s12884-017-1611-2
17. Gitonga E, Muiruri F. Determinants of health facility delivery among women in tharaka nithi county, Kenya. *Pan Afr Med J*. (2016) 25(Suppl 2):9.
18. Kohi TW, Mselle LT, Dol J, Aston ML. When, where and who? Accessing health facility delivery care from the perspective of women and men in Tanzania: a qualitative study. *BMC Health Serv Res*. (2018) 18. doi: 10.1186/s12913-018-3357-6. [Epub ahead of print]
19. Bohren MA, Hunter EC, Munthe-Kaas HM, Souza JP, Vogel JP, Gülmezoglu AM. Facilitators and barriers to facility-based delivery in low- and middle-income countries: a qualitative evidence synthesis. *Reprod Health*. (2014) 11(1):71. doi: 10.1186/1742-4755-11-71
20. Kruk ME, Kujawski S, Mbaruku G, Ramsey K, Moyo W, Freedman LP. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. *Health Policy Plan*. (2018) 33(1):e26–33. doi: 10.1093/heapol/czu079
21. Haddad SM, Souza RT, Cecatti JG. Mobile technology in health (mHealth) and antenatal care-searching for apps and available solutions: a systematic review. *Int J Med Inform*. (2019) 127:1–8. doi: 10.1016/j.ijmedinf.2019.04.008
22. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. *BMC Health Serv Res*. (2017) 17(1):704. doi: 10.1186/s12913-017-2664-7
23. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. *Matern Child Health J*. (2012) 16(5):1092–101. doi: 10.1007/s10995-011-0836-3
24. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of Mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review. *Health Promot Pract*. (2019) 20(6):805–17. doi: 10.1177/1524839919855391
25. Hategeka C, Ruton H, Law MR. Effect of a community health worker mHealth monitoring system on uptake of maternal and newborn health services in Rwanda. *Glob Health Res Policy*. (2019) 4:8. doi: 10.1186/s41256-019-0098-y.
26. Agarwal S, Glenton C, Tamrat T, Henschke N, Maayan N, Fønhus MS, et al. Decision-support tools via mobile devices to improve quality of care in primary

healthcare settings. *Cochrane Database Syst Rev.* (2021) 7(7):CD012944. doi: 10.1002/14651858.CD012944

27. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* (2018) 319(13):1317–8. doi: 10.1001/jama.2017.18391
28. Espinosa C, Becker M, Marić I, Wong RJ, Shaw GM, Gaudilliere B, et al. Data-Driven modeling of pregnancy-related complications. *Trends Mol Med.* (2021) 27:762–76. doi: 10.1016/j.molmed.2021.01.007
29. Auret L, Aldrich C. Interpretation of nonlinear relationships between process variables by use of random forests. *Miner Eng.* (2012) 35:27–42. doi: 10.1016/j.mineng.2012.05.008
30. Mboya IB, Mahande MJ, Mohammed M, Obure J, Mwambi HG. Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania. *BMJ Open.* (2020) 10(10):e040132. doi: 10.1136/bmjopen-2020-040132
31. Tesfaye B, Atique S, Azim T, Kebede MM. Predicting skilled delivery service use in Ethiopia: dual application of logistic regression and machine learning algorithms. *BMC Med Inform Decis Mak.* (2019) 19(1):209. doi: 10.1186/s12911-019-0942-5
32. Kwizera A, Kissoon N, Musa N, Urayenzeza O, Mujiyarugamba P, Patterson AJ, et al. “Sepsis in resource-limited nations” task force of the surviving sepsis campaign. A machine learning-based triage tool for children with acute infection in a low resource setting. *Pediatr Crit Care Med.* (2019) 20(12):e524–30. doi: 10.1097/PCC.0000000000002121
33. Rittenhouse KJ, Vwalika B, Keil A, Winston J, Stoner M, Price JT, et al. Improving preterm newborn identification in low-resource settings with machine learning. *PLoS One.* (2019) 14(2):e0198919. doi: 10.1371/journal.pone.0198919
34. Tuti T, Agweyu A, Mwaniki P, Peek N, English M, Clinical Information Network Author Group. An exploration of mortality risk factors in non-severe pneumonia in children using clinical data from Kenya. *BMC Med.* (2017) 15(1):201. doi: 10.1186/s12916-017-0963-9
35. Sazawal S, Ryckman KK, Das S, Khanam R, Nisar I, Jasper E, et al. Machine learning guided postnatal gestational age assessment using new-born screening metabolomic data in south Asia and sub-saharan Africa. *BMC Pregnancy Childbirth.* (2021) 21(1):609. doi: 10.1186/s12884-021-04067-y
36. Ogallo W, Speakman S, Akinwande V, Varshney KR, Walcott-Bryant A, Wayua C, et al. Identifying factors associated with neonatal mortality in sub-saharan Africa using machine learning. *AMIA Annu Symp Proc.* (2021) 2020:963–72.
37. Shukla VV, Eggleston B, Ambalavanan N, McClure EM, Mwenechanya M, Chomba E, et al. Predictive modeling for perinatal mortality in resource-limited settings. *JAMA Netw Open.* (2020) 3(11):e2026750. doi: 10.1001/jamanetworkopen.2020.26750
38. Fulcher IR, Nelson AR, Tibaijuka JI, Seif SS, Lilienfeld S, Abdalla OA, et al. Improving health facility delivery rates in zanzibar, Tanzania through a large-scale digital community health volunteer programme: a process evaluation. *Health Policy Plan.* (2020) 35(10):1–11. doi: 10.1093/heapol/czaa068
39. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat.* (1992) 46(3):175–85. doi: 10.1080/00031305.1992.10475879
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
41. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B Stat Methodol.* (1996) 58(1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
42. Breiman L. Random forests. *Mach Learn.* (2001) 45(1):5–32. doi: 10.1023/A:1010933404324
43. Paliwal M, Kumar UA. Neural networks and statistical techniques: a review of applications. *Expert Syst Appl.* (2009) 36(1):2–17. doi: 10.1016/j.eswa.2007.10.005
44. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak.* (2016) 16(3):197–208. doi: 10.1186/s12911-016-0318-z