BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Causal machine learning models for predicting low birth weight in midwife-led continuity care intervention in North Shoa Zone, Ethiopia

Wudneh Ketema Moges[1,2,3*], Awoke Seyoum Tegegne[1], Aweke A. Mitku[1,6], Esubalew Tesfahun[4] and Solomon Hailemeskel[5]

Abstract

Background Low birth weight (LBW) is a critical global health issue that affects infants disproportionately, particularly in developing countries. This study adopted causal machine learning (CML) algorithms for predicting LBW in newborns, drawing from midwife-led continuity care (MLCC).

Methods A quasi-experimental study was carried out in the North Shoa Zone of Ethiopia from August 2019 to September 2020. A total of 1166 women were allocated into two groups. The first group, the MLCC group, received all their antenatal, labor, birth, and immediate post-natal care from a single midwife. The second group received care from various staff members at different times throughout their pregnancy and childbirth. In this study, CML was implemented to predict LBW. Data preprocessing, including data cleaning, was conducted. CML was then employed to identify the most suitable classifier for predicting LBW. Gradient boosting algorithms were used to estimate the causal effect of MLCC on LBW. Moreover, meta-learner algorithms were utilized to estimate the individual treatment effect (ITE), the average treatment effect (ATE), and performance. Moreover, meta-learner algorithms were utilized to estimate the individual treatment effect (ITE), the average treatment effect (ATE), and performance.

Results The study results revealed that Causal K-Nearest Neighbors (CKNN) was the most effective classifier based on accuracy and estimated LBW using a 94.52% accuracy, 90.25% precision, 92.57% recall, and an F1 score of 88.2%. Meconium aspiration, perinatal mortality, pregnancy-induced hypertension, vacuum babies in need of resuscitation, and previous surgeries on their reproductive organs were identified as the top five features affecting LBW. The estimated impact of MLCC versus other professional groups on LBW was analyzed using gradient boosting algorithms and was found to be 0.237. The estimated ATE for the S-learner was 0.284, which is lower than the true ATE of 0.216. Additionally, the estimated ITE for both the T-learner and X-learner was less than -0.5, indicating that mothers would not choose to participate in the MLCC program.

*Correspondence:
Wudneh Ketema Moges
wudnehketema@gmail.com
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions Based on these findings, the CKNN classifier demonstrated a higher accuracy and effectiveness. The S-learner and R-learner models, utilizing the XGBoost Regressor and BaseSRegressor, provided accurate estimations of ITE for assessing the impact of the MLCC program. Promoting the MLCC program could help stabilize LBW outcomes.

Keywords Low Birth Weight, ATE, Machine learning, Causal machine learning, Meta–learner

Background

The World Health Organization (WHO) defines low birth weight (LBW) as a birth weight below 2500 g. This condition presents a significant health challenge in low- and middle-income countries (LMICs), as it is associated with increased rates of infant mortality and various long-term health complications [1]. LBW ranks as the second most significant contributor to global perinatal mortality, trailing only behind premature birth. LBW may result from either preterm delivery, which occurs before 37 weeks of gestation, or intrauterine growth restriction (IUGR). However, this is more commonly seen in LMICs, where it accounts for approximately 60% of LBW cases. In these areas, IUGR affects about 11% of all births [2].

The worldwide prevalence of LBW is a considerable public health issue, with estimates indicating that between 15% and 20% of all births globally result in LBW, which translates to more than 20 million infants each year [3, 4]. Notably, more than 95% of these LBW babies are born in LMICs, highlighting the substantial impact of this condition in resource-limited settings [5–8].

The prevalence of LBW exhibits considerable variation across different regions. Asia reports the highest prevalence at 18.3%, approximately three times greater than Europe's rate of 6.4%, which is the lowest in the world. Within Asia, there is a significant in LBW rates, ranging from 5.9% in Eastern Asia to 27% in South-central Asia, underscoring the substantial disparities within the continent [5, 9].

A newborn's birth weight significantly influences its development, healthy growth, and survival [10]. LBW neonates have a 20 times higher risk of dying at birth compared to newborns with a standard birth weight of 2,500 to 4,000 g [11]. Brain development retardation, poor language development, and intellectual disabilities are more common in LBW newborns [12]. LBW babies often need extra medical attention in environments where their future health outcomes remain a persistent concern and source of uncertainty [13]. In developing countries like Ethiopia, LBW poses a severe threat to public health since it causes the infant's growth to be stunted, which raises morbidity and death [14].

Therefore, the WHO recommends using midwife-led continuity of care (MLCC) models, where a known midwife supports a woman throughout pregnancy, birth, and the postnatal period. This recommendation applies to settings with midwife education programs that train midwives to practice in an MLCC setting [15, 16]. MLCC ensures that a single midwife provides continuous care throughout pregnancy, birth, and postnatal, offering individualized education and counseling. The midwife also identifies and refers women needing specialist care to appropriate professionals [17]. MLCC is particularly relevant in Ethiopia and other LMICs due to its significant impact on maternal and neonatal health outcomes. Studies have shown that MLCC improves maternal and neonatal outcomes where a known midwife or a team of midwives provides continuous care throughout the antenatal, intrapartum, and postnatal periods [18, 19]. This ongoing support alleviates stress and encourages timely interventions that are crucial for preventing LBW. Implementing MLCC in Ethiopia can significantly improve neonatal outcomes, making it a vital strategy in LMICs [20, 21]. Current newborn and maternal healthcare practices frequently fall short of the personalized and continuous care offered by MLCC. This gap can lead to inconsistent treatment and poorer outcomes, such as higher rates of LBW. Midwives used a standard tool to record the baseline characteristics through interviews and antenatal cards. Blinded data collectors gathered post-birth outcomes, while intervention and continuity data were obtained from medical records and early postnatal interviews with eight midwives and four supervisors trained for three days. The MLCC was compared with other professional groups without random assignment in a quasi-experimental study design. This design evaluates the effect of the intervention by observing the differences between groups.

Causal inference is essential in healthcare because it allows us to understand the true impact of interventions rather than just identifying correlations. For instance, when comparing the effects of the MLCC (maternal and child health care) program to other professional groups on LBW outcomes, it's crucial to determine whether the MLCC program directly improves LBW rates [22, 23].

Machine learning (ML), which was once viewed as a lesser player in the field of statistics, has gained significant recognition in recent years. In the realm of digital medicine, ML models have made remarkable progress, largely driven by technological advancements and improved data collection methods. As these parallel developments continue, the potential for ML to revolutionize healthcare remains on the rise [24]. One reason for its underdog status is that biostatistics and other sciences focus not only on predictions but also on establishing causal inferences. Thus, ML algorithms solve a fundamentally different problem [25, 26].

Standard ML excels at finding patterns and making predictions based on data. However, this does not necessarily explain why those patterns exist or what would happen if we changed one of the variables. This is where Causal Machine Learning (CML) occurs. CML focuses on understanding cause-and-effect relationships, which is vital for making informed decisions and effective interventions in healthcare [24, 27, 28].

Integrating CML to assess and predict the impact of an MLCC program can significantly enhance healthcare strategies. This approach aims to reduce LBW rates by providing consistent, personalized care throughout pregnancy and delivery, thereby improving health outcomes for both mothers and newborns.

By applying CML, we can directly measure the impact of participation in the MLCC program on LBW outcomes compared to other professional groups. This insight enables healthcare providers and policymakers to make informed decisions regarding which programs to implement and how to allocate resources effectively. For instance, if CML demonstrates that the MLCC program significantly reduces LBW rates, it provides strong evidence to support the expansion of this program [27, 29, 30].

Furthermore, considering the consistently elevated rates of LBW globally, particularly in resource-constrained countries, there is an urgent need for innovative solutions. Implementing advanced CML in these environments is essential, as it facilitates more precise, data-driven healthcare practices that can greatly decrease LBW rates and enhance the health outcomes for both mothers and their newborns [27, 31].

Therefore, CML combines all these concepts by adjusting ML techniques to provide clearly defined causal questions using relevant data [32, 33]. CML provides diverse methods for causal inference and uplift modeling using ML algorithms. These methods are grounded in current studies and allow users to estimate the Individual Treatment Effect (ITE). Specifically, CML facilitates uplift optimization by analyzing quasi-experimental data [34]. Therefore, this study aimed to adopt CML techniques to predict LBW in newborns, drawing from MLCC treatment.

Methods

Data

From August 2019 to September 2020, a quasi-experimental study was conducted in the North Shoa Zone of the Amhara Regional State, Ethiopia. The zone has a population of over two million people, with approximately 2,393,877 living in this zone, of which 1,207,839 are males and 1,186,038 are females. The health infrastructure consists of nine hospitals, with one serving as a referral center for comprehensive emergency obstetric care. Additionally, there were 95 health centers and 389 health posts. Primary hospitals have 10–15 midwives each and at least one integrated emergency surgical officer. Obstetricians and gynecologists are available only at the referral hospital.

Consequently, we included a total of 1,166 mothers who visited the prenatal and antenatal care clinics during the study's data collection phase. Four primary hospitals in the study area were randomly selected: Shoa Robit, Ataye, Mehal Meda, and Alem Ketema Enat Hospital. These hospitals provide delivery services to visitors from both urban and rural areas.

As mentioned above, we used a two-stage stratified cluster sampling technique to select four primary hospitals. The samples were equally distributed, and systematic random sampling with an interval of two was used to select participants. Shoa Robit and Ataye hospitals were intervention sites with MLCC, while Mehal Meda and Alemketema Enat served as control sites. Eligible pregnant women were approached and selected until the required sample size was achieved.

Data collection

The baseline characteristics of the study participants, including socio-demographics and obstetric, gynecologic, medical, and surgical histories, were recorded by midwives using a standard tool. Data were collected through face-to-face interviews using maternal antenatal cards. An independent, blinded data collector obtained post-birth, maternal, and neonatal outcomes from the birth registry. Intervention exposure and continuity of care data were gathered from medical records and interviews conducted during the early postnatal period. To prevent the Hawthorne effect, the healthcare providers were blinded to the outcome data. Eight midwife data collectors and four supervisors were trained in data collection and extraction over three days.

Eligibility criteria

The studies included pregnant women who were less than 24 weeks gestational age at their first antenatal care booking, had a singleton pregnancy, and were considered to have a low obstetric risk. Women with multiple pregnancies, those intending to register with a different care provider, or those with a history of medical or obstetric complications were excluded from the study.

Quasi-experimental setup

Treatment Group (MLCC) The WHO refers to this intervention as an MLCC. MLCC describes a model in which women are followed through the continuum of pregnancy and contact a single midwife, the primary midwife responsible for the entire process. This continued contact through the intrapartum and postnatal period facilitated the relationship between women and known midwives. The two primary forms of MLCC are the caseload and team midwifery models. In the caseload model, one midwife cared for up to 45 women and facilitated relational continuity.

In addition, an MLCC is described as the care provided to women throughout pregnancy, birth, and early parenting by a known midwife or a small team of midwives. Midwives operating within midwifery continuity of care models are known as caseload midwives, as they serve as the primary professionals for a specific group of women from their initial antenatal booking through the postnatal period. Care is typically delivered by a small group of midwives, known as team midwifery. Within this model, midwives work in partnership with women and lead in the planning, organization, and delivery of care, including referrals to other professionals as appropriate.

Control (other professional) Depending on the stage of pregnancy for the women in the control group, the responsibility for organizing and delivering care from the time of initial booking through the postnatal period is shared among several healthcare professionals.

Variables in the study The outcome variable in this research was LBW which has been defined as the classification of LBW, based on the weight of an infant at birth, in this study, the target variable was categorized into two groups based on birth weight. LBW for infants with a birth weight of 2499 g or less, and not LBW for infants with a birth weight of 2500 g or more [35], which was defined as with an allocated group of mothers (MLCC and other professional groups). The analytical framework for LBW highlights MLCC and other professional groups, explained through various demographic, obstetric, medical, and neonatal outcome variables (Fig. 1).

Data analysis

The research employed CML techniques to identify patterns and predict LBW, allowing for comparison with the outcomes derived from CML methods. We computed the chi-square test analysis to evaluate important factors in LBW. CML employs model selection techniques and algorithms to find patterns in data. A version of Python software 3.10 worked for all statistical analyses. The maternal, and newborn data were initially split into training, and test data sets, with 20% for validation and 80% for training, respectively. To run the RFregressor module, Python requirements for Scipy, Pandas, Sklearn, Matplotlib, Seaborn, and NumPy were required. Training a CML allows it to learn from historical data, adjust internal parameters, and generalize patterns. Evaluation using test data validates its real-world effectiveness. In short, training ensures learning, and testing validates prediction quality.

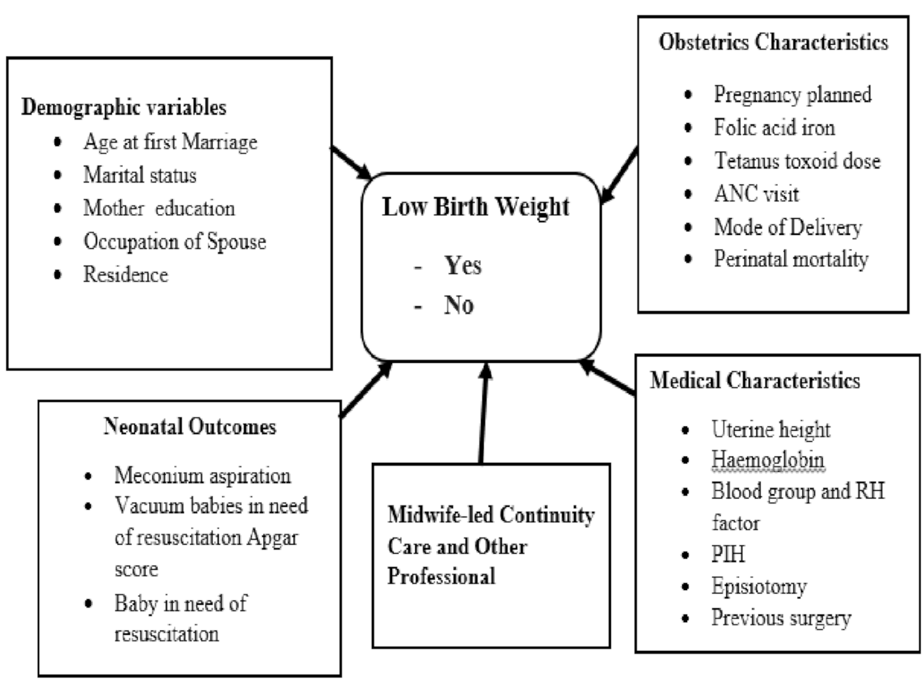

Fig. 1 Analytical framework

Causal machine learning models to estimate the causal effect of low birth weight

Causal analysis and ML originally developed as separate disciplines; however, they have recently begun to intersect, resulting of a beneficial exchange of ideas and heightened interest in both areas. The integration of causal analysis can improve ML by elucidating the causal relationships and effects of MLCC on LBW.

We trained and tested six CML models to predict LBW and compared their performance. The models included Causal Logistic Regression, Causal Random Forest (CRF), Causal Decision Tree, Causal Naïve Bayes, Causal K-Nearest Neighbors (CKNN), and Causal Support Vector Machine (CSVM) classifiers.

In causal inference techniques, CML algorithms such as CKNN and CRF can enhance the understanding of causal relationships within data. Specifically, CKNN leverages the KNN algorithm within a causal inference framework to estimate causal effects by comparing outcomes of similar individuals who participated in the MLCC program or were treated by other professionals [36].

In the context of MLCC, CKNN can be used to analyze the impact of different care models on LBW outcomes. By considering care practice features and potential causal factors, CKNN can help identify effective strategies for improving LBW and inform evidence-based decision-making for maternal and newborn care [37]. In addition, CRF is an extension of the traditional random forest algorithm, specifically designed for causal inference. It adapts the random forest algorithm to estimate causal effects by using decision tree outcomes to predict the potential impact of variables.

Gradient boosting algorithms (GBM)

Gradient boosting algorithms are powerful ensemble learning techniques that iteratively combine weak learners to create a robust prediction model. By focusing on the errors of previous iterations, the predictions were sequentially improved, making them highly effective for regression and classification tasks. Boosting combines several base estimators' predictions to enhance the single estimator's robustness. We used gradient-boosting algorithms, specifically XGBoost, LightGBM, and CatBoost, to estimate the causal effect of LBW for an allocated group of mothers [38, 39].

To emphasize using CML methods designed to handle confounding and selection bias and provide interpretability for reliable causal inference [40, 41]. XGBoost was employed within ensemble methods, such as T-learner and S-learner, to estimate the causal effect of MLCC interventions compared to other professional groups. It is crucial to predict counterfactual outcomes in which individuals receive MLCC treatment to enhance causal understanding [42].

Recursive elimination method

Recursive Feature Elimination (RFE) is an effective feature selection algorithm in ML that enhances the model performance by removing irrelevant features and focusing on key predictors. In this study, with numerous features involved, RFE simplifies the model and improves accuracy by reducing noise. In addition, Wrapper methods in causal inference involve evaluating subsets of features using the model itself to identify the most relevant data for accurate predictions. These methods iteratively add or remove features based on their impact on the model's performance, ensuring that only the most significant variables are included [43].

However, to elucidate the causes of LBW, it is essential to integrate causal inference methods with feature-importance analysis is essential. A quasi-experimental design helps identify causal effects and control confounders, ensuring accurate interpretation by distinguishing predictive importance from causality. Therefore, RFE enhances model performance in CML by systematically removing irrelevant or less important features from the dataset. By focusing on features that have a significant impact on LBW, REF helps create more interpretable models that capture essential causal relationships by focusing on features that significantly impact LBW. This process not only improves prediction accuracy but also aids in understanding which variables genuinely influence LBW, thereby supporting better decision-making [44, 45].

Software implementation

There was different uplift modeling algorithms in the state-of-the-art for the current version of CausalML implements. Due to the length and structure of the manuscript, we present only a selection of the algorithms in Fig. 2.

CML was used as a tool for causal inference ease of use, and to estimate the average treatment effect (ATE), the underlying method is expanded upon by the meta-learner algorithms. CML supports four main algorithms: S-learner, T-learner, X-learner, and R-learner. Every meta-algorithm uses a distinct method to estimate the ATE and ITE.

Causal machine learning primarily backs uplift-based causal inference tree-based and methods for meta-learning. The most important algorithms were the causalML-inference tree, causalML-inference, meta-packages, and causalML-feature selection, and their familiar package was causalML-inference in Python statistical language [46].

CausalML is a software library designed for causal inference tasks using flow charts to identify causal relationships between variables. CML algorithms, a broad category for data predictions, include tree-based

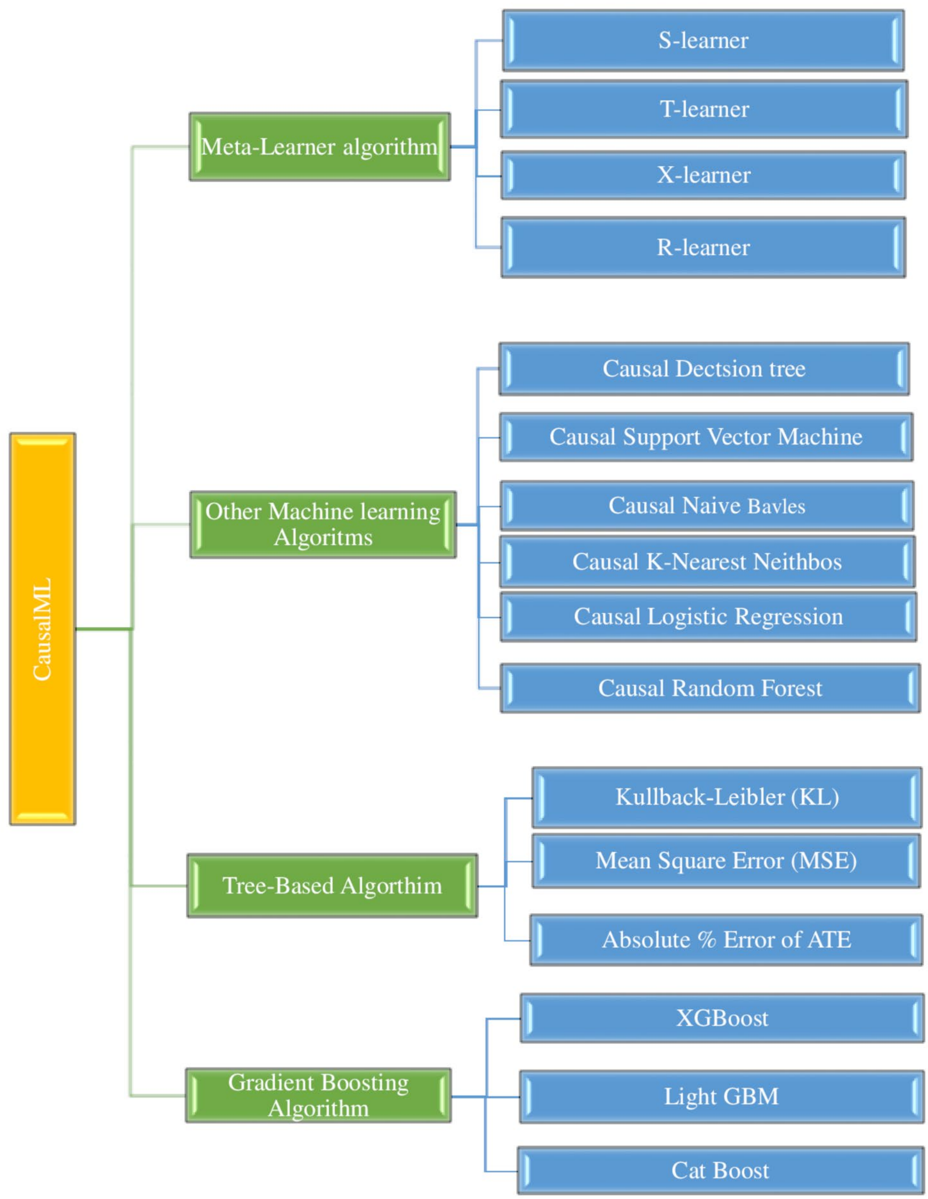

Fig. 2 Causal machine learning algorithm diagram

algorithms, gradient-boosting algorithms, and meta-learners. Tree-based algorithms use decision trees as their foundation, while gradient boosting algorithms are a specific type of tree-based ensemble method [46].

Furthermore, the flowchart visually organizes the algorithms based on their utility in causal analysis, emphasizing meta-learners and gradient-boosting methods. Meta-learner algorithms are central to learning causal relationships between variables. The specific algorithm used is given in Fig. 2.

Individual treatment effect This is the different outcome when a mother is exposed to MLCC versus a mother exposed to other professionals.

$$ITE = P(LBW|MLCC) - P(LBW|\,other\ professional\,)$$

$$ITE = P(y_i = 1|\boldsymbol{X}_i, w_i = 1) - P(y_i = 1, \boldsymbol{X}_i, w_i = 0)$$

$$ITE = P(y_i = 1,\ w_i = 1|\boldsymbol{X}_i) = P(y_i = 1,\ w_i = 0\,|\boldsymbol{X}_i)$$

Where: $P(w_i = 1|\boldsymbol{X_i})$ is the probability that a mother $\boldsymbol{X}_i$ is assigned to MLCC.

$P\,(\,w_i = 0|\boldsymbol{X_i})$ is the probability that a mother $\boldsymbol{X}_i$ is assigned to another professional group.

$$ITE = \frac{P(y_i = 1, w_i = 1|\boldsymbol{X}_i)}{0.5} - \frac{P(y_i = 1, w_i = 0|\boldsymbol{X}_i)}{0.5}$$

$$ITE = 2[P(y_i = 1,\ w_i = 1|\boldsymbol{X}_i) - P(y_i = 1,\ w_i = 0|\boldsymbol{X}_i)]$$

$$ITE = 2\left[P(Z_i = 1|\boldsymbol{X}_i) + (\ P(y_i = 1,\ w_i = 1|\boldsymbol{X}_i)) + P(y_i = 0, w_i = 1|\boldsymbol{X}_i)) - 1\right]$$

$$ITE = 2\left[P(Z_i = 1|\boldsymbol{X}_i) + P(w_i = 1|\boldsymbol{X}_i) - 1\right]$$

$$= 2[P(z_i = 1|\boldsymbol{X}_i) + 0.5 - 1]$$

$$= 2p(z_i = 1|\boldsymbol{X}_i) - 1$$

Where represents the outcome variable of interest. In this context, it refers to a binary outcome related to birth weight (LBW or not).

$\boldsymbol{X_i}$ Represents the set of features or covariates

$\boldsymbol{w_i}$ Represents the treatment assignment variable, it takes binary values , $\boldsymbol{w_i} = 1$, which indicates that the mother is assigned to MLCC, $\boldsymbol{w_i} = 0$, this indicates that the mother is assigned to care-managed by other professionals.

Meta-algorithms for learning

Meta-learner is a model-agnostic algorithm that provides a recipe for estimating conditional average treatment (CATE) using any ML method (called base learners) [47]. We applied a meta-learning algorithm to estimate the causal effect. This meta-algorithm can either use a single base learner with the treatment indicator as a feature (S-learner) or multiple base learners, addressing a specific aspect of MLCC and other professional groups (T-learner, X-learner, and R-learner). To estimate the ATE using the XGBTRegressor, we fit the XGBTRegressor model. Then, we use the estimate ATE method to obtain the ATE along with its upper and lower bounds.

T-Learner In the initial steps, we estimate the control response using data from other professional groups. We use a base learner, which can be any supervised learning or regression estimator, to deepen our understanding of causal relationships. Next, we estimate the response function for MLCC care.
To estimate the ATE, we explore various scenarios and carefully analyze the data where MLCC is not applied and applied $(Y = 0) : \mu_0(\boldsymbol{X}) = E[Y(0)|\ \boldsymbol{X} = \boldsymbol{x}]$, $(Y = 1) : \mu_1(\boldsymbol{X}) = E[Y(1)|\ \boldsymbol{X} = \boldsymbol{x}]$ respectively, so **X** represents the features, and Y represents LBW.

Step 1
Estimate the average outcome $\mu_0(\boldsymbol{x}) = E[Y(0)|\ \boldsymbol{X} = \boldsymbol{x}]$
$\mu_1(\boldsymbol{x}) = E[Y(1)|\ \boldsymbol{X} = \boldsymbol{x}]$ Using ML.

Step 2
Define the CATE estimates as:

$$\hat{\tau}(\boldsymbol{x}) = \hat{\mu}_1(\boldsymbol{x}) - \hat{\mu}_0(\boldsymbol{x})$$

S-learner estimates the MLCC as follows, using a single ML model.

Step 1
Estimates the average outcome $\mu(\boldsymbol{x})$ using an indicator variable, and covariate **X** covariates for MLCC, Z: $\mu(\boldsymbol{x}, z) = E[Y|\ \boldsymbol{X} = \boldsymbol{x},\ Z = z]$ Using an ML model.

Step 2
The CATE estimate is defined as:

$$\hat{\tau}(\boldsymbol{x}) = \mu(\boldsymbol{x}, Z = 1) - \hat{\mu}(\boldsymbol{x}, Z = 0)$$

Including the propensity score in the model can reduce bias from the regularization-induced confounding [48].

To estimate the ATE, we explore various scenarios and carefully analyze the data using supervised learning or regression algorithms, we can denote the estimated functions as follows. To estimate the CATE, we can use a weighted average derived from the response functions produced by supervised learning or regression algorithms. This approach allows us to capture the nuanced impact of treatments across different scenarios. X-learner [49], is a T-learner extension that involves the following three steps:

Step 1
Estimate the average outcomes $\mu_0(\boldsymbol{x}) = E[Y(0)|\ \boldsymbol{X} = \boldsymbol{x}]$
$\mu_1(\boldsymbol{x}) = E[Y(1)|\ \boldsymbol{X} = \boldsymbol{x}]$ Using ML models;

Step 2
Impute the user-level treatment effects, $D_i{}^1$ and $D_j{}^0$ for use in the MLCC group based on
$\mu_0(\boldsymbol{x})$, and

$$D_i{}^1 = Y_i^1 - \hat{\mu}_0\left(\boldsymbol{X}_i{}^1\right)$$

$$D_i{}^0 = \hat{\mu}_1\left(\boldsymbol{X}_i{}^1\right) - Y_i^0$$

Then estimate $\tau_1(\boldsymbol{x}) = E\left[D^1|\boldsymbol{X} = \boldsymbol{x}\right]$, and $\tau_0(\boldsymbol{x}) = E\left[D^0|\boldsymbol{X} = \boldsymbol{x}\right]$ using ML models

Step 3
Define the CATE estimates by a weighted average of; $\tau_1(\boldsymbol{x})$ and $\tau_0(\boldsymbol{x})$

$$\tau\ (\boldsymbol{x}) = g\ (\boldsymbol{x})\ \tau\ _0\ (\boldsymbol{x}) + (1 - g\ (\boldsymbol{x}))\tau\ _1\ (\boldsymbol{x})$$

Where $g \in [0,1]$.

The X-learner is a meta–learner that is an extension of the T-learner.

The R-learner leverages out-of-fold estimations of outcomes and propensity scores to enhance its predictions. The R-learner leverages out-of-fold estimates of outcomes to enhance its predictions $\widehat{m}^{(-1)}\ (\boldsymbol{x}_i)$ and propensity scores $\widehat{e}^{(-i)}\ (\boldsymbol{x}_i)$. There are two steps in this process:

Step 1
Fit $\widehat{m}\ (\boldsymbol{x})$ and $\widehat{e}\ (\boldsymbol{x})$ use ML cross-validation models;

Step 2
Calculate the effects of the treatment by minimizing the R-loss, $\widehat{L}_n\ (\tau\ (x))$

$$\hat{L}_n\ (\tau\ (x)) = \frac{1}{n}\sum_{i=1}^{n}(\left(Y_i - \hat{m}^{(-i)}\ (X_i)\right) - \left(Z_i - \hat{e}^{(-i)}\ (X_i)\right)\tau\ (X_i))^2$$

Where: $\widehat{e}^{(-i)}\ (\boldsymbol{x_i})$, the R-learner leverages out-of-fold held-out predictions (made without using the ith training sample) to enhance its predictions.

Tree-based algorithms
Uplift tree modeling techniques assess the incremental impact of an action or treatment, combining causal inference and ML. This approach uses tree-based algorithms, splitting based on differences in uplift, and introduces three methods to quantify divergence gain from splitting [50].

$$D_{gain} = D_{after,\ split}\left(P^T, P^C\right) - D_{before,\ split}\left(P^T, P^C\right)$$

Where: **D** measures the divergence , $\boldsymbol{P^T}$, and $\boldsymbol{P^C}$ refers to the outcome of interest's probability distribution in MLCC and other professionals, respectively. The package implements three methods for calculating divergence: Kullback-Leibler (KL), mean square error (MSE), and absolute error of ATE.

The divergence of Kullback-Leibler (KL) is provided by:

$$KL\,(P:Q) = \sum_{k=left,\ right} pk_{log}\frac{p_k}{q_k}$$

Where: **p** is the sample mean in the MLCC, **q** is the sample mean in the other professionals, and the leaf in which **p** and **q** are calculated is indicated by **k**.

The fitness of the tree-based algorithm was assessed using the MSE and the mean absolute error of ATE. The MSE of a predicted model relates to the estimated variables and is defined as follows [51].

$$MSE\ = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,j} - X_{model,i}\right)^2}{n}}$$

Where: $\boldsymbol{X_{obs}}$ is observed values, $\boldsymbol{X_{model}}$ is modeled values at the time $\boldsymbol{i}$.

The MSE and statistical approaches such as CML comparison, feature importance, and cross-validation were used to evaluate the results and select the best model for predicting LBW.

The mean absolute error has been calculated as:

$$Abs\ \%\ error\ of\ ATE\ = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n}$$

Where: $\boldsymbol{x_i}$ is the prediction and y_i is the true value.

Data preprocessing
Data preprocessing was a crucial step in ML that involved transforming raw data into a more suitable format for building and training models.

In this study, we utilized 10 continuous variables that were scaled. We employed grid search to systematically assess a predefined set of hyperparameter values, selecting the combination that delivers the best performance on the validation set.

To enable quicker model convergence, numerical features were standardized using the Scikit-learn Standard Scaler Python package. Subsequently, the dataset was split into two groups, with 20% for validation and 80% for training, to assess the CML models.

Despite this, all models with the same ratio of test-to-train splits had nearly the same performance evaluation measures. The CML model was fitted using a training dataset, and the test dataset was used to assess how well the CML model fits the data. The performance of the predictive models was evaluated and compared using various metrics. Training and testing metrics, such as accuracy, precision, recall, and F1 score, are essential for learning from data and assessing real-world performance. A smaller gap between training and testing accuracy indicates better generalization, as illustrated in Fig. 3.

Results
We computed the chi-square statistics to estimate which covariates best discriminated between LBW. The prevalence of LBW was the lowest among mothers aged less than 20 years (16.11%) for mothers 20–29 years(66.41%), and 16.33% for mothers aged greater than or equal to 30 years old ($p < 0.05$). The prevalence of LBW was the lowest among mothers who had not been given folic acid/iron (4.9%) compared with mothers who had been given folic acid/Iron (95.08%), ($X^2 = 3.265, p_{value} = 0.0071$), and others such as

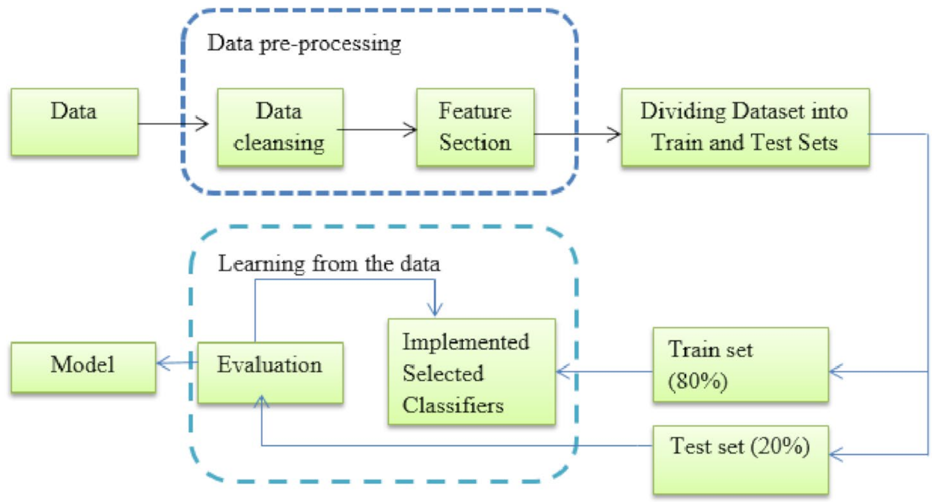

Fig. 3 Implementation of the predictive models

tetanus toxoid dose $(X^2 = 0.773, p_{value} = 0.002)$, ANC visit $(X^2 = 4.16, p_{value} = 0.044)$, counseling on nutrition $(X^2 = 1.192,\ p_{value} = 0.0275)$, blood pressure measurement taken$(X^2 = 109.72,\ p_{value} = 0.000)$, uterine height$(X^2 = 5.448,\ p_{value} = 0.043)$, venereal disease research laboratory (syphilis test) $(X^2 = 9.711,\ p_{value} = 0.002)$, blood group and Rh factor $(X^2 = 36.639,\ p_{value} = 0.000)$,pregnancy-induced hypertension(PIH)$(X^2 = 3.62,\ p_{value} = 0.036)$, meconium aspiration$(X^2 = 9.496,\ p_{value} = 0.002)$, prevention of mother-to-child transmission of HIV$(X^2 = 42.62,\ p_{value} = 0.000)$, fetal heart beat $(X^2 = 0.036,\ p_{value} = 0.018)$, vacuum$(X^2 = 0.576, p_{value} = 0.045)$, babies ≤ 7 Apgar score with 5 min $(X^2 = 1.778,\ p_{value} = 0.0282)$, and post-natal care $(X^2 = 1.02, p_{value} = 0.03)$ were significantly associated with LBW of newborn babies. Therefore, the significant covariates identified through chi-square statistics were employed to build ML algorithms on the training dataset.

Recursive elimination method

Figure 4 presents the important feature selection algorithms for 87 features. Based on those features, higher scores indicate higher importance features. Figure 4 shows that REF is a feature-selection technique that iteratively removes less important features from a dataset. In CML, RFE helps identify relevant features by considering their impact on causal relationships. By gradually eliminating features, RFE enhances model interpretability and avoids spurious correlations, making it valuable for selecting features that directly influence the LBW. From the 87 sets of features, a subset of the top 20 useful features was shortlisted for importance, and there was only a very gradual drop in the number of cases misclassified. As a result, 15 features were selected to balance computational efficiency and classification accuracy effectively. Factors such as meconium aspiration, perinatal mortality, PIH, vacuum baby in need of resuscitation, and previous surgeries on their reproductive organs were the top five important predictors of LBW; however, the normal last menstrual period, preterm labor, and early postnatal care score were the lowest predictive variables in our model.

Perinatal mortality, occurring from the 22nd week of gestation up to seven days post-birth was a crucial predictor in the context of LBW. Infants face a higher risk of perinatal death due to health complications. Additionally, the causes of perinatal mortality may impact LBW rates, such as induction of labor, preterm labor, neonatal infections, maternal diseases, PIH, and birth complications are closely associated with perinatal mortality.

The EFE method was implemented in Scikit-learn using Python software. This method randomly shuffles each feature and computes the change in the model's performance. The features that influence performance most are considered the most critical. Their impact was pivotal in determining the model's effectiveness.

In the model, the score of a variable reflects its relative importance compared to others. Therefore, if a categorical variable indicates a high score, the variable as a whole plays a significant role in the model performance. For meconium aspiration, the reference group had no respiratory problems; for perinatal mortality, the reference group had preterm death; for PIH, the reference group had gestational hypertension during pregnancy; and for maternal age, the reference group was less than 20 years. For folic acid, the reference group consisted of individuals who had not received folic acid supplementation. Nutrition counseling had not been provided. The

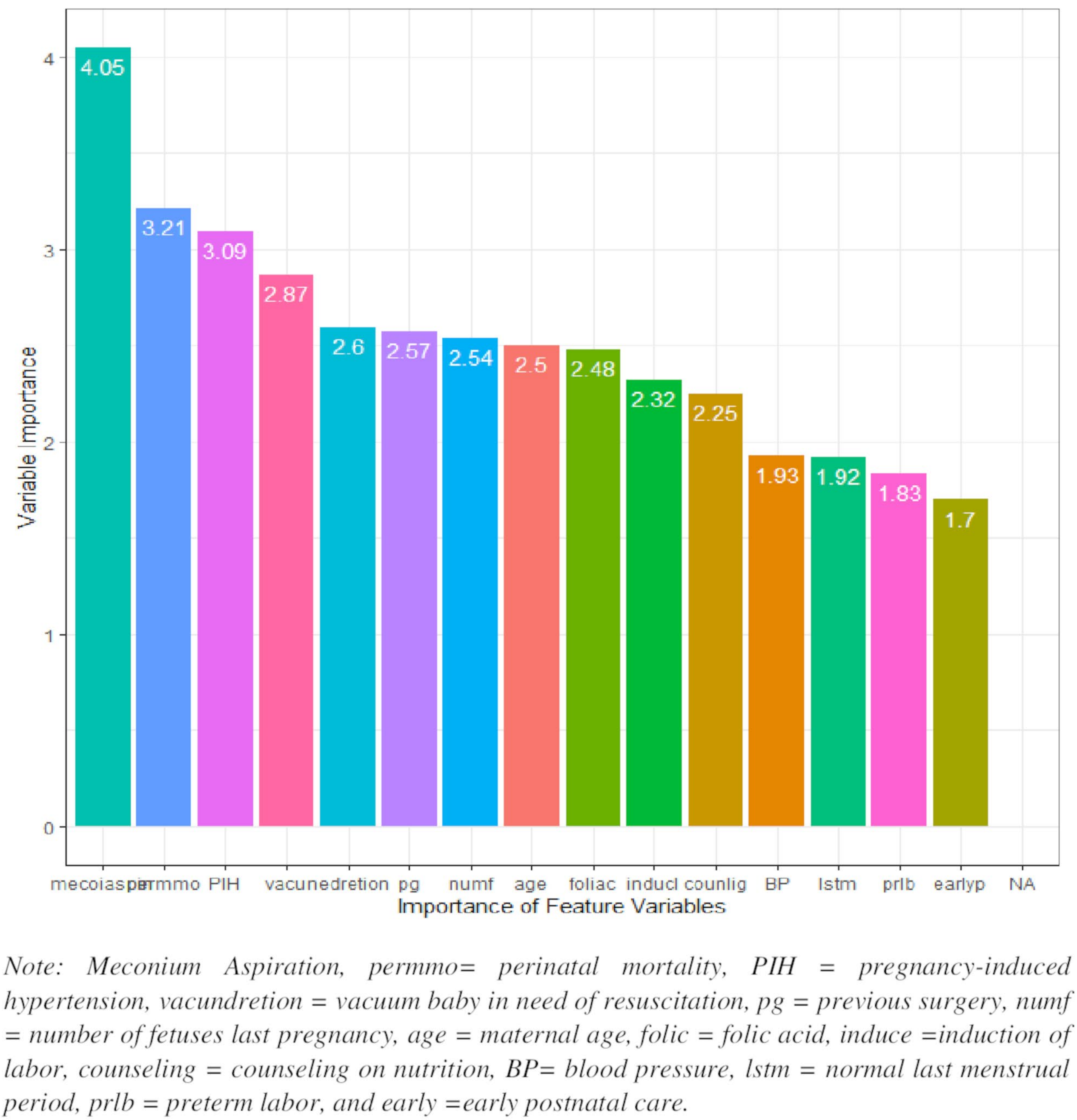

Fig. 4 Top 20 mean variable importance from the recursive elimination method

reference group for previous surgeries consisted of mothers who have had surgeries on their reproductive organs.

Predictive causal machine learning models to estimate LBW

As shown in Table 1, the CKNN model emerged as the most accurate predictor of LBW. As a method for causal inference, CKNN aims to estimate the causal effect of MLCC interventions on LBW. By leveraging the k-nearest neighbor principle, CKNN effectively addresses the issues of confounding and selection bias, providing robust estimates of causal effects. This makes CKNN a powerful tool for understanding and improving maternal and neonatal health outcomes [52].

The CKNN and random forest models performed well in training and testing, with CKNN displaying a minimal accuracy difference of 1.4% points (training: 96.0%, testing: 94.6%). In contrast, the decision tree accurately

Table 1 Predictive models of performance of low Birth Weight Training and Testing Data

Predictive Models	Model Performance							
	Training data				Test data			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Causal Logistics Regression	92.45	91.5	95.0	89.5	89.6	90.6	92.5	87.0
Causal Support Vector Machine	91.0	85.1	89.5	81.8	88.5	82.6	87.6	78.3
Causal Decision Tree	91.0	92.5	93.6	87.6	87.5	88.4	92.6	78.6
Causal Random Forest	92.8	94.9	95.0	93.2	90.3	92.2	94.3	89.2
Causal K-Nearest Neighbors	96.0	92.5	94.6	92.6	94.5	90.2	92.5	88.2
Causal Naïve Bayes	91.0	89.8	91.6	88.5	89.3	87.5	87.3	82.3

identified most actual LBW cases but had lower precision. The causal random forest method performed well, achieving a training accuracy of 92.8% in the training data. However, the accuracy of unseen testing data decreased slightly to 90.3%, indicating potential overfitting. This suggests that while CKNN and causal random forest both generalize effectively, there is room for improvement in handling new instances. By addressing overfitting and analyzing data distribution, model reliability in practical scenarios can be enhanced, ensuring better generalization to new data (Table 1).

The causal random forest demonstrated robust predictive capabilities, correctly predicting 525 LBW babies and 615 normal birth weight babies, with an accuracy of 90.3%, precision of 92.2%, recall of 94.3%, and an F1 score of 89.2%. This model outperformed the CML based on these metrics. Meanwhile, the CKNN model correctly predicted 820 LBW babies and 271 normal birth weight babies, achieving an accuracy of 94.5%, precision of 90.2%, recall of 92.5%, and an F1 score of 88.2%. Despite the CKNN's higher accuracy, the causal random forest showed better precision and recall, making it a strong contender for predicting LBW, as given in Table 1.

Among the models in this study, the CKNN model demonstrated the highest predictive accuracy for LBW (Table 2). The causal naïve Bayes model correctly identified 87.3% of them, which is the lowest performance among the models in this study. The sensitivity increased by 0.0698 when using the causal random forest compared to the causal naïve Bayes. This indicates that, based on sensitivity, the causal random forest was a better choice for investigating LBW in our dataset.

Ultimately, the accuracy, precision, and recall exceeded 85%, reflecting a 15% error rate. There are two types of error rates: Type I and Type II. In our scenario, the Type I error rate (specificity) was 16.4%, and the Type II error rate (sensitivity) was 10.57%. This means that Type I errors occurred in the MLCC prediction of LBW, whereas Type II errors occurred in predictions by other professionals. Because Type II errors are more costly, we should aim to reduce them, even if it means increasing Type I errors, as shown in Table 2.

Gradient boosting algorithms ATE of LBW

As shown in Table 3, the LBW rate for the allocated group of mothers was 0.2347, indicating a significant variation between MLCC and other professional groups. XGBoost can be used within ensemble methods to estimate the causal effect of MLCC versus the control group on LBW. The average predicted counterfactual estimate for mothers who received MLCC is 0.21606.

Meta-learner algorithms to estimate ATE and ITE of LBW

We can observe that the treatment has a positive impact on some meta-algorithms and a negative impact on other meta-algorithms of other individuals. The ATE was around 0.5. Most individuals in the dataset showed positive treatment effects. The ATE for the population was the average of the ITE. Therefore, the ATE values for the S-learner, T-learner, X-learner, and R-learner were 0.284, 0.3191, 0.2256, and 0.51, respectively.

The estimated ATE for the S-learner was 0.284, which is lower than the true ATE of 0.216. The 95% confidence interval for the S-learner's ATE was between 0.274 and 0.293. In contrast, the R-learner's estimated ATE was 0.51, slightly exceeding the true ATE by 0.01. The confidence interval for the R-learner's ATE spanned from 0.04293 to 0.7223.

The estimated ITE for T and X-learners was less than – 0.5, indicating that mothers without MLCC (maternal lifestyle change counseling) units experienced a negative impact on the birth weight of their newborns. In contrast, the estimated ITE for S and R-learners was greater than – 0.5, suggesting that mothers with MLCC units saw stabilization in their newborn birth weight. Therefore, the S and R-learners model using XGBRegressor with BaseSRegressor provides a more accurate estimation of ITE, as shown in Table 4.

We assessed the ATE estimation using various meta-learner algorithms in the Meta algorithm. To assess variability, we calculated the absolute error of the ATE, along with the MSE and KL divergence. As shown in Table 5, during the validation test of ATE for the training data, the X-learner (XGB) had the lowest absolute error estimate, whereas the S-learner (LG) had the highest. The

Table 2 Predictive models with confusion matrix for low Birth Weight as evaluated on Test Data

Evaluation Matrix			Predictive Models											
			CLR		CSVM		CDT		CNB		CRF		CKNN	
			Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Confusion matrix	Observed	Yes	100	20	90	30	90	30	80	125	525	50	820	60
		No	10	1036	20	1026	20	1026	25	936	40	615	15	271

Table 3 Causal effect of LBW using gradient boosting algorithm

LGBM Regressor	XGB Regressor	Cat boost Regressor	Mean
0.18045	0.21606	0.10243	0.2347

Table 4 Meta Learner Algorithm to Estimate ATE and ITE

Meta Learner Algorithm	Average Treatment Effect(ATE)		Individual Treatment Effect(ITE)	
	Estimate	95% CI	Estimate	95% CI
S-learner	0.284	(0.274,0.293)	0.7347	(0.13188,1.3822)
T-learner	0.3191	(0.1925,0.7336)	-0.55042	(-0.2565,-1.1169)
X-learner	0.2256	(0.1383,0.6485)	-0.68032	(-0.0362,-1.54997)
R-learner	0.50272	(0.04293,0.7223)	0.6423	(0.3345,1.12087

Table 5 Validating Meta-Learner Accuracy for Training Data

Meta-Learner Algorithm	Abs % Error of ATE	MSE	KL Divergence
S-learner (LG)	0.37381	0.076792	3.7732272
S-learner (XGB)	0.022471	0.027847	0.048112
T-learner (LR)	0.354732	0.03398	0.300478
T-learner (LR)	0.018514	0.131973	0.272192
X-learner (LR)	0.354732	0.033398	0.300478
X-learner (XGB)	0.014297	0.064749	0.097801
R-learner (LR)	0.293485	0.032854	0.285591
R-learner (XGB)	0.139719	0.152496	1.255079

Table 6 Validating Meta-Learner Accuracy for Validation Data

Meta-Learner Algorithm	Abs % Error of ATE	MSE	KL Divergence
S-learner (LG)	0.370911	0.076236	3.745188
S-learner (XGB)	0.020455	0.029570	0.060167
T-learner (LR)	0.353449	0.033313	0.299904
T-learner (LR)	0.015387	0.131514	0.285431
X-learner (LR)	0.353449	0.033313	0.299904
X-learner (XGB)	0.013689	0.064864	0.104117
R-learner (LR)	0.293852	0.033015	0.286671
R-learner (XGB)	0.136497	0.151742	0.259589

R-learners (LR) recorded the lowest MSE, whereas the R-learners (XGB) had the highest MSE.

For the validation data, the ATE on the testing dataset showed that the lowest absolute error estimates were found for the T-learner (XGB), X-learner (XGB), and S-learner (XGB). Conversely, the highest absolute error estimates were observed in the S-learner (LR), X-learner (LR), and T-learner (LR). The lowest MSE was recorded by the S-learners (XGB), T-learners (LR), and X-learners (LR), while the highest MSE was noted for the R-learners (XGB) and T-learners (XGB), as shown in Table 6.

The testing validation dataset demonstrated the best performance when comparing validation accuracy between training and testing with the meta-learning algorithm. Finally, when comparing the MSE of the ATE

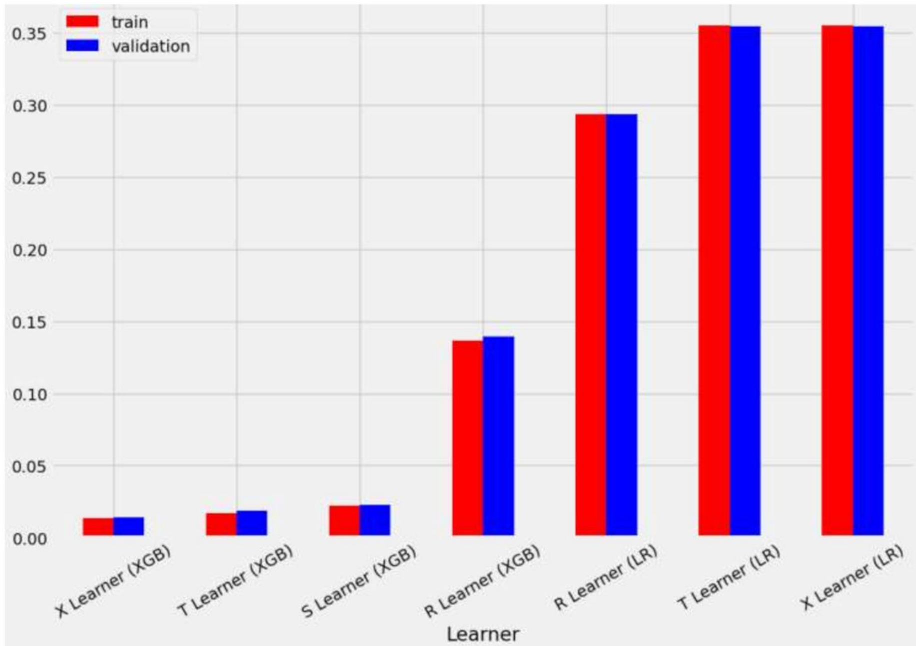

Fig. 5 Learner performance of absolute % error for ATE

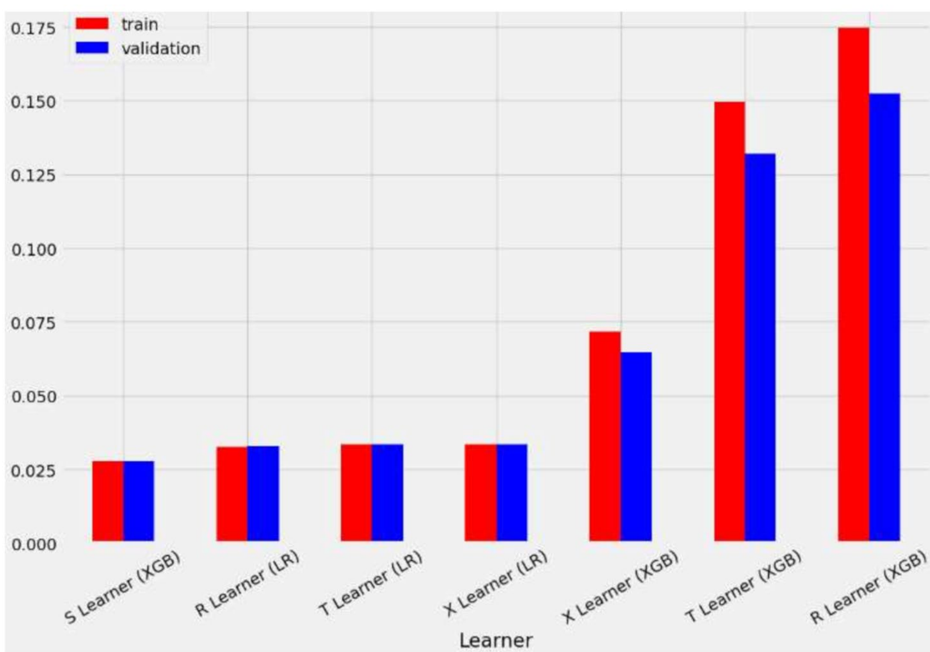

Fig. 6 Learner performance of mean squared error

estimates, we observed a trade-off between variance and bias. Models with low bias often exhibited higher variance, resulting in a slightly increased MSE in the testing dataset, as shown in Table 6.

Figure 5 shows the absolute percentage error performance measure for ATE on the training and validation datasets. XGBoost-based learners (S-learner, T-learner, and X-learner) achieved lower absolute percentage errors for ATE on both datasets. In contrast, learners using logistic regression (S-learner, T-learner, and X-learner) exhibited the highest absolute percentage errors for ATE, indicating lower performance than XGBoost.

Figure 6 illustrates the MSE for ATE across both training and validation datasets. The lowest MSE for ATE was achieved by the S-learner (XGB), R-learner (LR), and T-learner (LR) in both datasets. Conversely, the highest

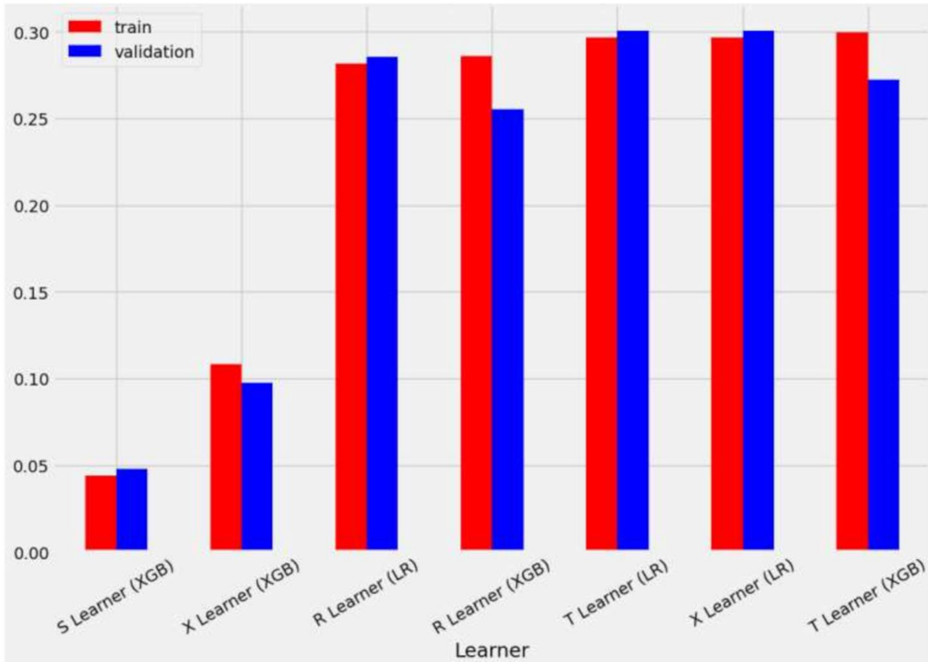

Fig. 7 Learner performance of KL divergence

MSE for ATE was observed in the R-learner (XGB) and T-learner (XGB). Notably, the smallest MSE in the validation dataset was recorded by the S-learner (XGB). Therefore, the S-learner (XGB) demonstrated the best performance in terms of MSE for ATE in the meta-learning algorithm for LBW compared to other performance measures.

Figure 7 illustrates ATE's KL divergence across training and validation datasets. XGBoost-based learners (S-learner, X-learner) achieved the lowest KL divergence for ATE, indicating a smaller distribution difference between the true and estimated ATE. The S-learner (XGB) recorded the smallest KL divergence in the validation set. In contrast, learners using logistic regression (X-learner, T-learner) and the T-learner (XGBoost) showed higher KL divergence, suggesting a larger difference between the true and estimated ATE compared to XGBoost-based learners.

Discussion

This study effectively demonstrated the practical application of CML methods in predicting LBW outcomes under MLCC intervention in the North Shoa Zone, Ethiopia. By employing CausalML, this study analyzed the effects of MLCC on LBW using quasi-experimental data. Using CML algorithms, the impact of MLCC on LBW outcomes is estimated. This approach has great potential for improving neonatal healthcare and enhancing maternal care practices. These findings align with broader efforts to advance maternal and child health [53].

Maternal age was considered one of the major factors in the healthy outcome of pregnancy. This study found a statistical association between maternal age and LBW, which contradicts findings from a study conducted in Nepal [54].

Maternal age is a critical factor influencing LBW. CML models adaptable to specific clinical contexts are valuable for healthcare providers. By leveraging these models, we can enhance neonatal healthcare and reduce mortality risks in targeted settings. This study's significance lies in its pioneering application of CML algorithms to predict LBW outcomes, aligning with previous research [55].

The divergence between this study and previous research on LBW risk highlights the need for further investigation using CML. Unlike traditional statistical models, CML explicitly considers causal relationships, allowing us to discern cause-and-effect associations and estimate counterfactual outcomes. This approach offers a fresh understanding of LBW causality [7, 14, 54, 56, 57]. PIH emerged as a top critical predictor of LBW in the North Shoa Zone, Ethiopia using mean rank-based identification for feature selection. This finding aligns with prior studies conducted in Ethiopia, Kenya, and Afghanistan [57–59].

This article investigates the impact of training and testing ratios on the performance of six well-known CML models, to predict LBW [60, 61], and we experimented with different training and validation data split ratios.

In our study, the causal random forest demonstrated commendable diagnostic performance with an accuracy of 90.3% and a sensitivity of 94.3%, aligning with previous research findings [62, 63]. Similarly, the causal naïve Bayes model showed favorable performance, achieving an accuracy of 89.3% and a sensitivity of 87.3%, consistent with prior investigations [63, 64]. However, both findings were limited by a small sample size (fewer than 1000 mothers), similar to our study.

The causal random forest model, with 90.3% accuracy, 92.2% precision, 94.3% recall, and an 89.2% F1 score, proved highly effective in predicting LBW in the North Shoa Zone, Ethiopia. Its suitability, balancing key metrics, aligns with similar findings in Afghanistan, showcasing the model's robustness in maternal and neonatal data analysis [59].

The CKNN model demonstrated the highest predictive accuracy for LBW, correctly predicting 820 LBW cases and 271 normal birth weight cases. It achieved 94.5% accuracy, 90.2% precision, 92.5% recall, and an 88.2% F1 score, outperforming the other CML models. These results align with previous research findings [65]. Our study, using gradient boosting algorithms, found significant variation in the causal effect of MLCC on LBW compared to other professional groups, aligning with previous research findings [66].

The meta-learning algorithm estimated the ATE for LBW using an S-learner, T-learner, X-learner, and R-learner, resulting in ATEs of 0.284, 0.3191, 0.2256, and 0.51, respectively. These findings contradicted studies conducted in France and the Netherlands [67, 68].

Our study suggests that the S-learner might miss some true treatment impacts. At the same time, the R-learner's estimated ATE could be higher than the actual ATE, indicating a potentially higher estimation of the magnitude of the treatment effect. This finding aligns with studies conducted in the USA [69, 70]. Our analysis revealed that the MSE for estimating the ATE occurred with S-learners (XGB), T-learners (LR), and X-learners (LR). These findings are consistent with previous studies [71, 72].

Contemporary CML models, such as XGBoost, significantly enhance the prediction of LBW by incorporating treatment effects for ATE and ITE. The evolution of CML has advanced causal inference in ML, integrating mature methods, tree-based approaches, and meta-learners. The selection of the most appropriate algorithm depends on the specific problem context, data characteristics, and research objectives, positioning CML as superior to the existing methods for predicting LBW [46].

This study demonstrates that CML, models such as XGBoost, offer several advantages over traditional methods. Firstly, they provide superior predictive accuracy, as evidenced by lower absolute percentage errors, MSE, and KL divergence, leading to more reliable and precise predictions. Secondly, CML models excel at capturing complex, non-linear relationships between variables that traditional methods might miss, offering a deeper understanding of the factors influencing LBW. Additionally, CML's ability to estimate ITE allows for personalized insights into how different interventions might impact individual outcomes, a level of detail typically not provided by traditional methods. Furthermore, the CML models effectively balance the trade-off between bias and variance, ensuring better generalization to new data. These contributions enhance predictive performance and advance the literature by showcasing the practical utility of advanced ML techniques in healthcare, particularly in predicting and managing LBW [35, 73].

Strengths, limitations, and future work

The strengths of this study using CML to examine MLCC include explicitly considering causal relationships, which allows us to understand the true impact of MLCC care on LBW. Additionally, it helps identify critical factors influencing maternal and neonatal health and ensures robust feature selection.

In our study on predicting LBW using CML models within MLCC interventions, we faced a significant limitation due to the absence of critical variables like birth order and body mass index (BMI) in many birth records. Despite having a substantial dataset with extensive maternal and neonatal information, the lack of these key variables hindered our ability to incorporate them into the feature selection process. Birth order and BMI are known to be influential factors in determining LBW, and their absence could affect our predictive models' accuracy and robustness. This limitation underscores the challenges of working with incomplete data in real-world settings, particularly in resource-limited environments where data collection may be inconsistent.

Nevertheless, our study still provides valuable insights by leveraging the available data and applying advanced CML techniques to identify other significant predictors of LBW.

Therefore, we employ deep learning techniques, such as causal convolutional neural networks, which could further improve the accuracy of LBW predictions by capturing more complex patterns and relationships within the data. Future work could also investigate the impact of additional maternal and neonatal health factors, such as maternal nutrition status or fetal heart rate, on LBW prediction. Ultimately, these efforts may result in the creation of more advanced and accurate models, greatly improving the quality of care for pregnant women and their infants.

Conclusions

The CKNN model achieved the highest predictive accuracy for LBW among CML algorithms, with 94.52% accuracy, 90.25% precision, 92.57% recall, and an F1 score of

88.2%. The top five features influencing LBW were meconium aspiration, perinatal mortality, PIH, vacuum babies' need for resuscitation, and previous surgeries on their reproductive organs. In contrast, the least predictive features included a normal last menstrual period, preterm labor, and early postnatal factors. The impact of MLCC and other professional groups on the causal relationship with LBW exhibited significant variation. S-learner and R-learner models with XGBoostRegressor and BaseSRegressor provide accurate ITE estimations for assessing the impact of MLCC on LBW.

The MLCC positively impacts stabilizing LBW, as shown by S-Learner and R-Learner estimates. The S-learner and R-learner models using XGBRegressor and BaseSRegressor provide better ITE estimations. In the validation dataset, the lowest absolute error estimates for ATE were achieved by the X-learner (XGB), T-learner (XGB), and S-learner (XGB).

Therefore, CML models significantly enhance LBW prediction by accurately estimating the impact of MLCC interventions. These models offer personalized insights and improve predictive accuracy, making them valuable tools for healthcare decision-making.

Recommendations

Given the finding that MLCC positively contributes to stabilizing LBW, we recommend prioritizing the implementation and promotion of MLCC practices. Ensuring access to skilled midwives, monitoring high-risk factors like meconium aspiration and perinatal mortality, and emphasizing early intervention can further enhance LBW outcomes. Additionally, ongoing evaluation of meta-learner algorithms, such as S-learner and R-learner, will help refine care strategies and improve maternal and newborn health.

Abbreviations

LBW	Low Birth Weight
CML	Causal Machine Learning
ML	Machine Learning
ATE	Average Treatment Effect
ITE	Individual Treatment Effect
MSE	Mea Square Error
XGB	Extreme Gradient Boosting
SVD	Spontaneous Vaginal Delivery
PIH	Pregnancy-induced hypertension

Acknowledgements

The authors would like to recognize the early contributions of expectant mothers to the study design and the efforts of the MLCC and other professional groups. The authors also wish to express their profound gratitude to the pregnant women who contributed their time and changes to this work, which any funding agencies did not support. The researchers also thank the health service personnel who correctly collected the data. Finally, I'd like to thank Debre Berhan University and Bahir Dar University for supporting my monthly wage and entrance, respectively. Many thanks to Dr. Solomon and Dr. Esubalew for providing/sharing the Data source.

Author contributions

W.K. led the study, data management, and data exploration. All authors conceived and designed the study. W.K. conducted the statistical analysis and drafted the article.A.S., and A.A. took responsibility for the integrity of the data and the accuracy of the data analysis. S.H. and E.S. provided the data. All authors made critical revisions to the manuscript for important intellectual content and gave final approval of the manuscript.

Funding

This project was unfunded.

Data availability

Upon a reasonable request, the corresponding author will release the datasets used in the current work to the public.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Review Board of Bahir Dar University, Ethiopia, Ethical Review Board referencing number RCS/1412/2022. All experiments adhered to relevant guidelines and regulations, including the Declaration of Helsinki. In the context of the current investigation, informed consent has been waived due to the use of secondary data. Informed consent was waived by the Board of Bahir Dar University, Institutional Review Board (IRB). Researchers were provided with the patient's ID number and other essential factors related to the experiment.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Statistics, College of Science, Bahir Dar University, P.O.Box 79, Bahir Dar, Ethiopia
[2]Department of Statistics, College of Science, Debre Berhan University, P.O.Box 445, Debre Berhan, Ethiopia
[3]Department of Data Science, College of Computing, Debre Berhan University, P.O.Box 445, Debre Berhan, Ethiopia
[4]Department of Public Health, College of Health Science, Debre Berhan University, P.O.Box 445, Debre Berhan, Ethiopia
[5]Department of Midwifery, College of Health Science, Debre Berhan University, P.O.Box 445, Debre Berhan, Ethiopia
[6]Global Change Institute (GCI), Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Received: 1 April 2024 / Accepted: 3 February 2025
Published online: 07 February 2025

References

1. Blencowe H, et al. National, regional, and worldwide estimates of low birth weight in 2015, with trends from 2000: a systematic analysis. Lancet Global Health. 2019;7(7):e849–60.
2. Valderrama CE, et al. Estimating birth weight from observed postnatal weights in a Guatemalan highland community. Physiol Meas. 2020;41(2):025008.
3. Weise A. Global Nutrition Targets 2025: Low Birth Weight Policy Brief. World Health Organization: Geneva, Switzerland, 2012: pp. 1–7.
4. Barreto CTG, et al. Low birth weight, prematurity, and intrauterine growth restriction: results from the baseline data of the first indigenous birth cohort in Brazil (Guarani Birth Cohort). BMC Pregnancy Childbirth. 2020;20:1–19.
5. Laopaiboon M, et al. An outcome-based definition of low birth weight for births in low and middle-income countries: a secondary analysis of the WHO global survey on maternal and perinatal health. BMC Pediatr. 2019;19:1–9.
6. Patterson JK, et al. Building a predictive model of low birth weight in low and middle-income countries: a prospective cohort study. BMC Pregnancy Childbirth. 2023;23(1):600.

7. Thapa P, et al. Prevalence of low birth weight and its associated factors: hospital-based cross-sectional study in Nepal. PLOS Global Public Health. 2022;2(11):e0001220.
8. Valderrama CE, et al. A proxy for detecting IUGR based on gestational age estimation in a Guatemalan rural population. Front Artif Intell. 2020;3:56.
9. Cutland CL, et al. Low birth weight: case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine. 2017;35(48Part A):6492.
10. Bendhari ML, Haralkar SJ. Study of maternal risk factors for low birth weight neonates: a case-control study. Int J Med Sci Public Health. 2015;4(7):987–90.
11. Linnér A. Immediate skin-to-skin contact for very preterm and low birth weight infants: from newborn physiology to mortality reduction. Inst för kvinnors och barns hälsa/Dept of Women's and Children's Health; 2022.
12. Saeidi R, Rahmani S, Saeidi M. Developmental outcomes of premature and low Birth Weight infants. Iran J Neonatology, 2016. 15(1).
13. Barros FC, et al. Global report on preterm birth and stillbirth (3 of 7): evidence for the effectiveness of interventions. BMC Pregnancy Childbirth. 2010;10:1–36.
14. Toru T, Anmut W. Assessment of low birth weight and associated factors among neonates in Butajira General Hospital, South Ethiopia, cross-sectional study, 2019. International Journal of Pediatrics, 2020. 2020.
15. Mose A, et al. Pregnant women's perception of midwifery-led continuity care model in Ethiopia: a qualitative study. BMC Womens Health. 2023;23(1):304.
16. WHO, L. WHO recommendations on antenatal care for a positive pregnancy experience. WHO Recomm. Antenatal care Posit. Pregnancy Exp; 2016.
17. Cbe JS. The contribution of continuity of midwifery care to high-quality maternity care. The Royal College of Midwives; 2017.
18. Fikre R, et al. Effectiveness of midwifery-led care on pregnancy outcomes in low-and middle-income countries: a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2023;23(1):386.
19. Butler MM, et al. Evaluating midwife-led antenatal care: choice, experience, effectiveness, and preparation for pregnancy. Midwifery. 2015;31(4):418–25.
20. Hailemeskel S, et al. Midwife-led continuity of care improved maternal and neonatal health outcomes in north Shoa Zone, Amhara regional state, Ethiopia: a quasi-experimental study. Women Birth. 2022;35(4):340–8.
21. Bradford BF, et al. Midwifery continuity of care: a scoping review of where, how, by whom and for whom? PLOS Global Public Health. 2022;2(10):e0000935.
22. Gutierrez P, Gérardy J-Y. Causal inference and uplift modeling: A review of the literature. In International conference on predictive applications and APIs. 2017. PMLR.
23. Gérardy jean-yves J-Y. Causal Inference and Uplift modeling a review of the literature.
24. Muse VP, et al. Seasonally adjusted laboratory reference intervals to improve the performance of machine learning models for classification of cardiovascular diseases. BMC Med Inf Decis Mak. 2024;24(1):62.
25. Masís S. Interpretable machine learning with Python: build explainable, fair, and robust high-performance models with hands-on, real-world examples. Packt Publishing Ltd.; 2023.
26. Agarwal N, Das S. Interpretable machine learning tools: A survey. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI). 2020. IEEE.
27. Leist AK, et al. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. Sci Adv. 2022;8(42):peabk1942.
28. Mosqueira-Rey E, et al. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev. 2023;56(4):3005–54.
29. Feuerriegel S, et al. Causal machine learning for predicting treatment outcomes. Nat Med. 2024;30(4):958–68.
30. Cui P et al. Causal inference meets machine learning. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020.
31. Prosperi M, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell. 2020;2(7):369–75.
32. Sanchez P, et al. Causal machine learning for healthcare and precision medicine. Royal Soc Open Sci. 2022;9(8):220638.
33. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. Nat Commun. 2020;11(1):3923.
34. Künzel SR, et al. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci. 2019;116(10):4156–65.
35. Ranjbar A, et al. Machine learning-based approach for predicting low birth weight. BMC Pregnancy Childbirth. 2023;23(1):803.
36. Kricke M, Peschenz T. Applied Predictive Analytics Seminar-Causal KNN. Google Scholar; 2019.
37. Zhou X, Kosorok MR. Causal nearest neighbor rules for optimal treatment regimes. arXiv Preprint arXiv:1711.08451, 2017.
38. Nhu V-H, et al. Shallow landslide susceptibility mapping: a comparison between logistic model tree, logistic regression, naïve Bayes tree, artificial neural network, and support vector machine algorithms. Int J Environ Res Public Health. 2020;17(8):2749.
39. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis Analytics J. 2022;3:100071.
40. Fan J, et al. Light gradient boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agric Water Manage. 2019;225:105758.
41. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):1–45.
42. Mooney SJ, Keil AP, Westreich DJ. Thirteen questions about using machine learning in causal research (you won't believe the answer to number 10!). Am J Epidemiol. 2021;190(8):1476–82.
43. Hoffman SR, et al. A step-by-step guide to causal study design using real-world data. Health Services and Outcomes Research Methodology; 2024. pp. 1–15.
44. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet. 2018;19:1–6.
45. Guyon I, Aliferis C. Causal feature selection, in Computational methods of feature selection. 2007, Chapman and Hall/CRC. pp. 79–102.
46. Zhao Y, Liu Q, Causal ML. Python package for causal inference machine learning. SoftwareX. 2023;21:101294.
47. Salditt M, Eckes T, Nestler S. A Tutorial introduction to Heterogeneous Treatment Effect Estimation with Meta-learners. Adm Policy Mental Health Mental Health Serv Res, 2023: pp. 1–24.
48. Learner S. Doctors, and mental health. BMJ, 2011. 342.
49. He Y et al. X-learner: Learning cross sources and tasks for universal visual representation. In European Conference on Computer Vision. 2022. Springer.
50. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014;23(13):3133–57.
51. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci Model Dev. 2014;7(3):1247–50.
52. Lecca P. Machine learning for causal inference in biological networks: perspectives of this challenge. Front Bioinf. 2021;1:746712.
53. Chernozhukov V et al. Applied causal inference powered by ML and AI. arXiv preprint arXiv:2403.02467, 2024.
54. KC A, Basel PL, Singh S. Low birth weight and its associated risk factors: Health facility-based case-control study. PLoS ONE. 2020;15(6):e0234907.
55. Margret IN et al. Machine learning-based Box models for pregnancy care and maternal mortality reduction: a Literature Survey. IEEE Access, 2024.
56. Lingani M, et al. Low birth weight and its associated risk factors in a rural health district of Burkina Faso: a cross-sectional study. BMC Pregnancy Childbirth. 2022;22(1):1–8.
57. Bekele WT. Machine learning algorithms for predicting low birth weight in Ethiopia. BMC Med Inf Decis Mak. 2022;22(1):232.
58. Muchemi OM, Echoka E, Makokha A. Factors associated with low birth weight among neonates born at Olkalou District Hospital, Central Region, Kenya. Pan Afr Med J, 2015. 20(1).
59. Zahirzada A, Lavangnananda K. Implementing predictive model for low birth weight in Afghanistan. In 2021 13th International Conference on Knowledge and Smart Technology (KST). 2021. IEEE.
60. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805–14.
61. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. Emerg Artif Intell Appl Comput Eng. 2007;160(1):3–24.
62. Ahmadi P, et al. Prediction of low birth weight using Random Forest: a comparison with logistic regression. Archives Adv Biosci. 2017;8(3):36–43.
63. Cho H, et al. Machine learning-based risk factor analysis of adverse birth outcomes in very low birth weight infants. Sci Rep. 2022;12(1):12119.
64. Desiani A et al. Naive Bayes classifier for infant weight prediction of hypertension mother. in Journal of Physics: Conference Series. 2019. IOP Publishing.

65. Rácz A, Bajusz D, Héberger K. Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. Molecules. 2021;26(4):1111.
66. Fikre R, et al. Effectiveness of midwifery-led care on pregnancy outcomes in low-and middle-income countries: a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2023;23(1):1–10.
67. Acharki N et al. Comparison of meta-learners for estimating multi-valued treatment heterogeneous effects. In International Conference on Machine Learning. 2023. PMLR.
68. Jacob D. CATE meets ML–The Conditional Average Treatment Effect and Machine Learning, (2021).
69. Chernozhukov V, et al. Double/debiased/neyman machine learning of treatment effects. Am Econ Rev. 2017;107(5):261–5.
70. Chernozhukov V et al. Double/debiased machine learning for treatment and causal parameters. 2017.
71. Datta S et al. A comprehensive review of the application of machine learning in fabrication and implementation of photovoltaic systems. IEEE Access, 2023.
72. Iban MC, Sekertekin A. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: a case study of Adana and Mersin provinces. Turk Ecol Inf. 2022;69:101647.
73. Reza TB, Salma N. Prediction and feature selection of low birth weight using machine learning algorithms. J Health Popul Nutr. 2024;43(1):157.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mr. Wudneh Ketema is a biostatistician and researcher on different public health problems using appropriate statistical methods. He also has expertise in statistical software and has published research articles in indexed journals. He is now a PhD candidate at Bahir Dar University, Ethiopia.Prof. Awoke Seyoum is a Professor of Biostatistics at Bahir Dar University, Ethiopia. He has published several articles regarding public health problems using appropriate statistical methods in highly cited journals. Currently, he is the director of research at Bahir Dar University.

Aweke A. Mitku is a Biostatistics Assistant Professor at Bahir Dar University, Ethiopia. He is an expert in statistical methods relating to public health issues and has published several articles in highly indexed journals. In addition to statistical software, he is currently employed as a postdoctoral fellow at the University of the Witwatersrand in South Africa.

Esubalew Tesfahun is an associate professor of public health at Debre Berhan University. He has also published some articles related to public health problems in well-indexed journals. He is an expert in determining the determinants of diseases and public health problems, and he currently works at the Debre publication and documentation directorate.

Solomon Hailemeskel is an Assistant professor of midwifery at Debre Berhan University at Ethiopia's College of Health Science. He is an expert in maternal and neonatal scenarios and other public health issues and has published several studies in highly indexed journals. He is currently a scientific director at Debre Berhan University in Ethiopia's College of Health Sciences.