Predicting Perinatal Mortality Based on Maternal Health Status and Health Insurance Service using Homogeneous Ensemble Machine Learning Methods

Dawit S Bogale(PhD Student) dawit.shibabaw@uog.edu.et
Tesfamariam M Abuhay tesfamariam.mabuhay@uog.edu.et
Belayneh E Dejene (PhD Student) belayneh.endalamaw@uog.edu.et

University of Gondar, Gondar, Ethiopia

Abstract

Background: Perinatal mortality in Ethiopia is the highest in Africa, with 68 per 1000 pregnancies intrapartum deaths (death during the delivery). It is mainly associated with home delivery, which contributes to more than 75% of perinatal deaths. Financial constraints have a significant impact on timely access to maternal health (MH) care. Financial incentives, such as health insurance, may address the demand- and supply-side factors. This study, hence, aims to predict perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods

Methods: The data was collected from the Ethiopian demographic health survey from 2011 to 2019 G.C. The data were pre-processed to get quality data that are suitable for the homogenous ensemble machine-learning algorithms to develop a model that predicts perinatal mortality. We have applied filter (chi-square and mutual information) feature selection and wrapper (sequential forward and sequential backward) feature selection methods. After selecting all the relevant features to develop a predictive model, we developed a predictive model using cat boost, random forest, and gradient boosting algorithms. After developing the predictive model, we evaluate the model using both subjective (domain expert) evaluation and objective (accuracy, precision, recall, F1_score, ROC) based evaluation techniques.

Results: we have develop the proposed model, by conducting three experiments using random forest, gradient boosting, and cat boost algorithms. The overall accuracy of random forest, gradient boosting, and cat boost with 17 features (which is selected by sequential backward feature selections and domain experts recommendation) is 89.95%, 90.24%, and 82%, respectively.

Conclusions: We decided to use gradient boosting algorithms for further use in the development of artifacts, model deployment, risk factor analysis, and generating relevant rules because it has registered better performance with 90.24% accuracy. The most determinant risk factors of perinatal mortality were identified using feature importance. Some of them are community-based health insurance, mother's educational level, residence, mother's age, wealth status, distance to the health facility, preterm, smoking cigarette, anemia level, hemoglobin level, and marital status.

Keywords: - homogenous ensembles, machine learning, perinatal mortality, maternal health, health insurance.

1. BACKGROUND

Perinatal mortality refers to a fatal death at or after 28 weeks of pregnancy (stillbirth) and includes death within 7 days of life after birth [1][2]. According to the World Health Organization (WHO) 2019 report, there were 2.6 million newborn infants globally, but more than 8200 died within a day [3]. Among the 133 million newborn infants alive each year, 2.8 million died in the first week of life after birth/at birth, and the majority occurred in low-income level countries [3]. Given the fast approaching deadlines for reaching the Millennium Development Goals, the international community supports low- and middle-level income countries to renew their commitment towards reducing maternal and infant mortality rates by improving access to maternal, neonatal, and perinatal health services [4].

Perinatal mortality in Ethiopia is the highest in Africa, with 68 per 1000 pregnancies and Intrapartum deaths (death during the delivery) [5]. Ethiopia shares and values the Sustainable Development Goals and has been trying to achieve the target of reducing neonatal mortality to below 12 per 1000 live births, by 2030 [6]. However, the reduction of neonatal, infant and under-five mortalities cannot be realized without a substantial reduction of perinatal mortality [7]. It is mostly associated with home deliveries, which contributed to more than 75% of all perinatal deaths due to the lack of awareness about health insurance services during birth, and it continued to be an essential part of the third sustainable development goal which aims to end preventable children's deaths by 2030 [6].

Financial constraints have a significant impact on timely access to maternal health care, such as Antenatal Care (ANC), skilled care at delivery, access to facility-based deliveries, postnatal care, and perinatal [8]. Over 100 million individuals pay out-of-pocket payments to get health treatments that have proven difficult to obtain for millions of poor people, resulting in increased morbidity and mortality [9]. WHO recommends community-based health insurance as one of the approaches for reducing pay out-of-pocket expenditures for registered families which, in turn, reduces morbidity and mortality [10]. The association of community-based health insurance with reduced maternal and infant mortality was apparent but it is impossible to reduce the infant mortality rate, without reducing the perinatal mortality [8]. Financial incentives, such as health insurance, can address the demand- and supply factors that may impact maternal, neonatal, and perinatal health results [11]. To this end, the Ethiopian Ministry of Health has been working for years to make health services accessible for women through community and facility-based interventions to increase the survival of newborns and children [6]. Despite these interventions, perinatal death remains an issue in Ethiopia, in particular; home delivery remains the challenge to reduce perinatal mortality [11]. Still, 74% of women give birth outside health institutions without skilled care attendants in Ethiopia [5][12][13]. Developing perinatal mortality prediction model using machine learning algorithms facilitates preventive actions. Machine learning, which aims to teach computers to perform what naturally involves humans learning from experience [5], includes supervised learning, unsupervised learning, and reinforcement learning [5]. Machine learning allows processing of large healthcare datasets and extracting clinical insights that support clinicians in planning and giving care, resulting in better outcomes, lower healthcare costs, and increased patient satisfaction [14].

This study, hence, aims at predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods by investigating the following research questions (1) what is the underline structure and evolution of perinatal mortality in Ethiopia over time? (2) Which homogeneous ensemble machine learning method is suitable to predict perinatal mortality in Ethiopia effectively? (3) What are the determinant factors of perinatal

mortality in Ethiopia? (4) What are the important rules that may shape strategies, policies, and interventions toward preventing and/or reducing perinatal mortality in Ethiopia?
The rest of this paper is organized as follows: Section 2 presents related works, Section 3 discusses materials and methods used, Section 4 mentions experimental setup, result, and discussion, and Section 5 presents conclusion.

2. RELATED WORK

Several studies investigated perinatal mortality in Ethiopia using different methods. Getachew et al. [14] investigated perinatal mortality and associated risk factors using a case-control study between 2008 and 2010 using a total of 1356 newborns' data (452 cases and 904 controls). Subgroup binary logistic regression analyses was done to identify associated risk factors for perinatal mortality, stillbirths, and early neonatal deaths. The study reported that the perinatal mortality rate was 85/1000, and after or at 28 weeks of birth death accounts for 87% [14]. Adjusted odds ratios revealed that obstructed labor, malpresentation, preterm birth, death during the delivery hemorrhage, and hypertensive disorders of pregnancy was independent predictor for high perinatal mortality.
Emmanuel et al [15]. Conducted research on Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset. Researcher used cross-sectional study design was conducted using the 2014–15 Rwanda Demographic and Health Survey. Researcher also used the total data set 30058 for their research and employed the machine learning algorithms such as Random forest, decision tree, support vector machine and logistic regression scored the accuracy of 84.3%, 83%, 68.5% and 61.8% respectively. Finally the researcher that applying the machine learning extract some hidden patterns from data but researchers give the direction for future work ensemble machine learning which might be increase the accuracy of algorithms.
Bitew et al [16]. Conducted on Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey. Researcher's also used the total dataset of 624 and applied machine learning techniques such as random forests, logistic regression, and K-nearest neighbor's and get accuracy, of 67.2%, 59.9% and 46.3%, respectively. Researchers gave the direction to work on algorithm optimization to reinforce the need to focus on the most important predicted factors including breastfeeding, birth interval control.
Another study was conducted by Yemisrach et al. [17] on factors associated with perinatal mortality among public health deliveries in Addis Ababa, Ethiopia using an unmatched case-control study and secondary data that was collected between 1st January to 30th February 2015. In this study, a total of 1113 (376 cases and 737 controls) maternal charts were reviewed and the mean age of the mothers for cases and controls were 26.47 ± 4.87 and 26.95 ± 4.68, respectively. Five hundred ninety-seven (53.6%) mothers delivered for the first time and factors that are significantly associated with increased risk of perinatal mortality were birth interval less than 2 years, preterm delivery, anemia, congenital anomaly, previous history of early neonatal death, and low birth weight. This study also reported that the use of a partograph was also associated with decreased risk of perinatal mortality. Bekele et al. [18] studied the effect of community-based health insurance on the utilization of outpatient health care services in Yirgalem town, Southern Ethiopia. This study used both quantitative and qualitative (mixed) research approaches using a comparative cross-sectional study design. A randomly selected sample of 405 (135 members and 270 non-members) household heads was used for quantitative analysis. Multivariate logistic regression was employed to identify the effect of community-based health insurance on healthcare utilization. This study revealed that members of households with community-based health insurance were about three times more likely to utilize outpatient care than their non-member counterparts [AOR: 2931; 95% CI (1.039, 7.929); p-value=0.042]. Finally, the researchers conclude that community-based health insurance is an effective tool to increase the utilization of healthcare services and provide the scheme to member households.
However, the aforementioned studies focused on identifying determinant risk factors only with a small dataset that cover limited geographical areas. Besides, these studies did not develop a predictive model, did not design an artifact that can be used by potential users, and did not generate rules that allow the development of evidence-based preventive strategies, policies, and interventions. On the other hand, machine learning algorithms have proven to be effective and efficient in predicting child mortality in African countries such as South Africa [19] and Uganda [20]. This study, hence, is motivated to fill these gaps by constructing a predictive model, identifying risk factors, designing artifacts, and generating relevant rules that help to develop evidence-based strategies, policies and interventions towards preventing, controlling and/or ending perinatal mortality in Ethiopia.

3. MATERIALS AND METHODS

Figure 1 depicts methodological flow chart that was implemented in this study to construct a perinatal mortality prediction model, identify risk factors, extract relevant rules, and design artifacts.

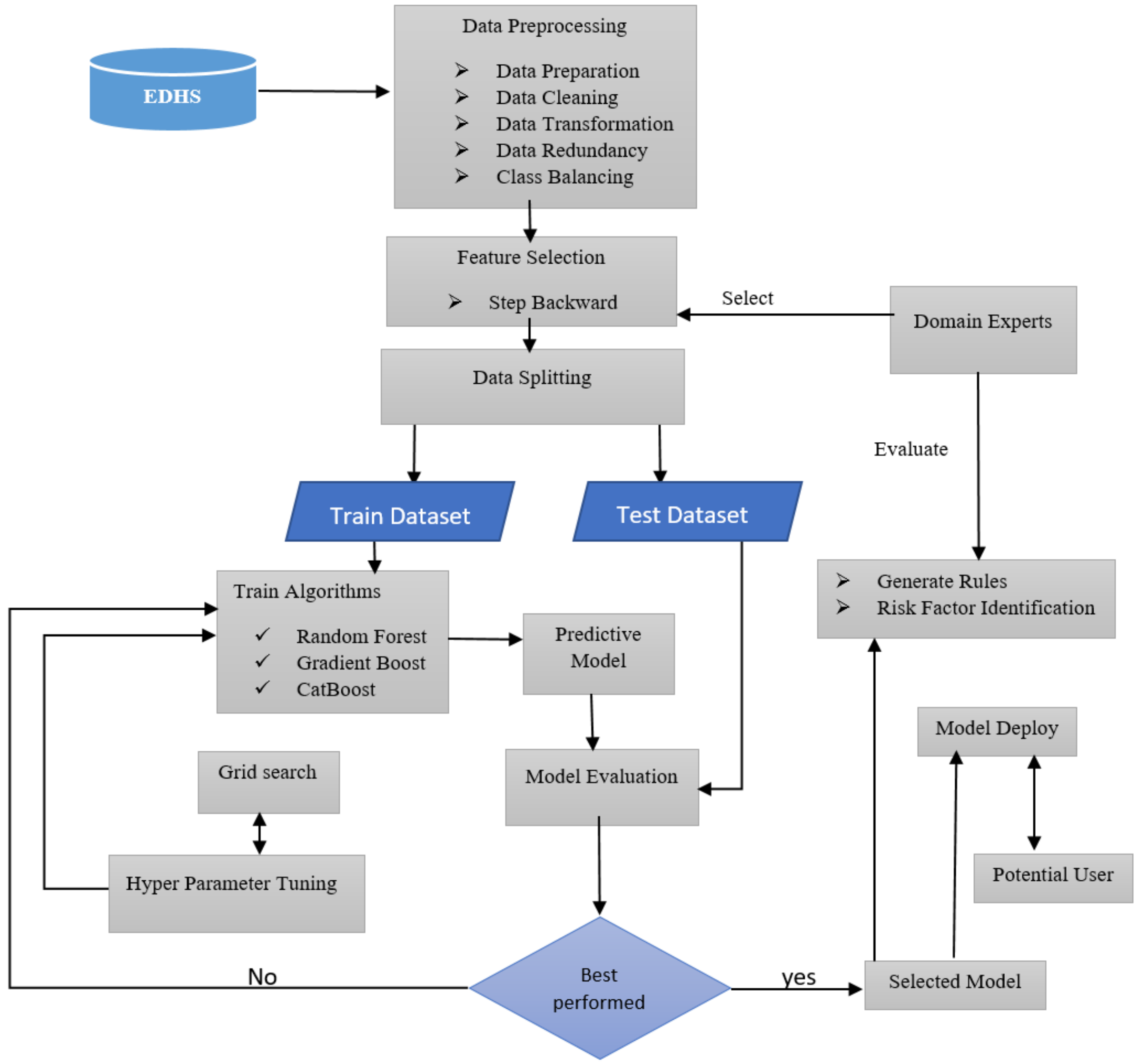

Figure 1:- The proposed model architecture

3.1. DATA COLLECTION and PREPROCESSING

In this study, a secondary data from the Ethiopia Demographic and Health Surveys (EDHS) which was collected by the Ethiopian Central Statistical Agency in 2011, 2016, and 2019 G.C, in five years intervals were used. The EDHS is a nationally representative household survey that collect data about demography, health, and nutrition for the purpose of impact monitoring and evaluation. To conduct this study, we have used the women file data set only on the EDHS as a predictor variables which focuses on the maternal determinants and the perinatal status as an outcome variable.

The raw data contains a total of 45 columns and 109531 instances before applying any feature selection techniques. From these instances and features 44 features were predictor variables and the last feature is the outcome variables, which is the status of the perinatal. The data imputation method (mode for categorical data and mean for continuous data) was employed to substitute the missing values. Outliers were identified using a boxplot and replaced using the Interquartile Range (IQR) scores. Data discretization was applied to transform some of the features. For example, the feature 'education level of mothers (v106)' has 8 different values which were transformed into five different values (illiterate (1), grade 1-8 (elementary), grade9-12 (secondary), grade 12+ (tertiary), and higher education (university and college)). The synthetic minority over-sampling technique (SMOTE) was implemented to handle the class imbalance in the training dataset. The main reason that we used SMOTE is it avoids loss of valuable information [21][22]. Before applying SMOTE, the data were 109531 instances, while after applying SMOTE, the data becomes 148659 instance, and we used the balanced dataset to develop the final predictive model. We have conducted four feature selection experiments to select the most relevant predictor variables or features using filter (mutual information and chi-square) and wrapper (sequential forward and sequential backward) methods from the total of 44 features. As a result, the sequential backward feature selection method has registered the highest performance with 90.5% of accuracy and produced 13 important features, see Table 1. The baseline fitness improvement (BFI) is used to limit the amount/number of features to be select using different feature selection methods[23]. So we are decided to take 13 feature based on the baseline fitness improvement suggest, and the accuracy the feature selection techniques scored by taking different number of feature. Besides, the domain experts who work at the University of Gondar comprehensive specialized hospital recommended 4 additional features namely preterm, the highest educational level, visited health facility in last 12 weeks, and birth interval. Finally, the total features selected by both sequential backward and domain experts that were used for further analysis are 17, see Table 1.

No	Chi_best	Mi_best	SFFS	SBFS
1.	Maternal age	Maternal age	ever-married sample	Community health in surance
2.	Literacy	Preterm	currently pregnant	smokes cigarettes
3.	frequency of reading newspapers or magazine	highest educational level	current contraceptive method	Region
4.	Family size	Religion	current use by metho d type	type of cooking fuel
5.	Wealth index	Ethnicity	heard family planning on radio last few mon ths	Wealth index
6.	Preterm	Family size	heard family plannin g on TV last few mon ths	current contraceptive method
7.	current contraceptive method	frequency of listening to the radio	heard family plannin g in newspaper/maga zine last few months	Hemoglobin level
8.	current use by method type	type of cooking fuel	visited health facility last 12 months	Occupation
9.	visited health facility last 12 months	wanted last children	Smokes pipe full of t obacco (women)	Maternal age
10.	currently breastfeeding	currently breastfeeding	Chews tobacco	Marital status
11.	when children put to the breast	Hemoglobin level	smokes other	anemia level
12.	Chews tobacco	anemia level	Community health in surance	Chews tobacco
13.	smokes other nicotine	Marital status	Health insurance type : provided by the emp loyer (women)	Place of residence
Accuracy	84.43	85.32	85.5	90.5

Table 1:- Features selected by sequential -backward feature selection

1	V228	Preterm
2	V106	Highest educational level
3	V394	visited health facility last 12 weeks
4	Bord	Birth interval

Table 2:-Remaining four feature were recommended by domain experts

3.2. MODEL DEVELOPMENT METHODS

To conduct this study, a design science research approach was implemented. Secondary data that was extracted from the Ethiopian Health Demographic Survey that was gathered in 2011, 2016, and 2019 was employed. We have conducted all the experiments using Python which a general purpose programming language. Different data pre-processing techniques such as handling missing values for categorical data using the mode imputation method, handling data transformation using binning, discretization, and normalization according to the data type, and handling class imbalance using SMOTE were employed. Several experiments were conducted using different feature selection methods such as filter (Mutual information and Chi-square), wrapper (step forward and step backward) and domain experts. And features that were selected by step backward feature selection methods and features that were recommended by domain experts were used to develop the classification model. To develop the predictive model, we have implemented homogeneous ensemble machine learning algorithms such as Random Forest, CatBoost, and gradient Boost, and the relevant parameters of each algorithm were tuned using grid search. We have selected those three homogeneous ensemble machine learning algorithms due to their strengths of preventing bias and over fittings and the nature of the data. Finally, the performance of the predictive models was evaluated using objective (accuracy, precision, recall, f1_score, ROC, and K-fold cross-validation (k=10)) and subjective (domain experts' opinion) methods. The best performing model was used to identify risk factors using feature importance, generate relevant rules for decision making, and develop an artifact that can be used by potential users such as health care professionals and parents.

4. RESULTS and DISCUSSION

In this section, the research questions that were raised at the introduction section of this study are discussed sequentially.

4.1. *What is the underline structure and evolution of perinatal mortality in Ethiopia over time?*

Perinatal mortality has been reducing over time in Ethiopia. This is because of the increase in the number of hospitals, especially in rural areas, and the introduction of community-based health insurance which encourages pregnant women to visit the hospital to give birth. But, due to the COVID-19 pandemic, the data collected in 2019 were twice as smaller as in previous years, as shown in Figure 2.

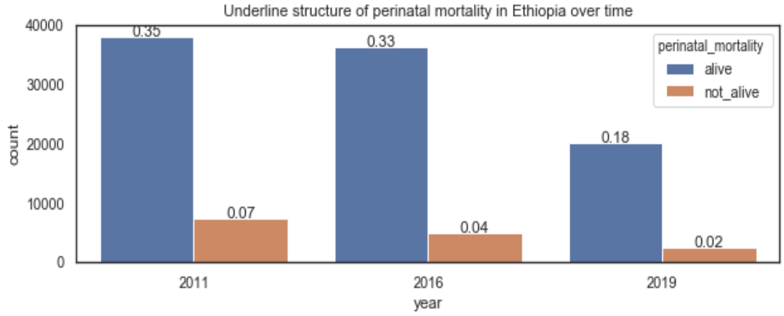

Figure 2:- Perinatal mortality over time in Ethiopia

The perinatal mortality across the regions of Ethiopia was also investigated. Among nine regions and two city administrations of Ethiopia, Amhara and Oromia regions registered higher perinatal mortality compared to other regions, as shown in Figure 3.

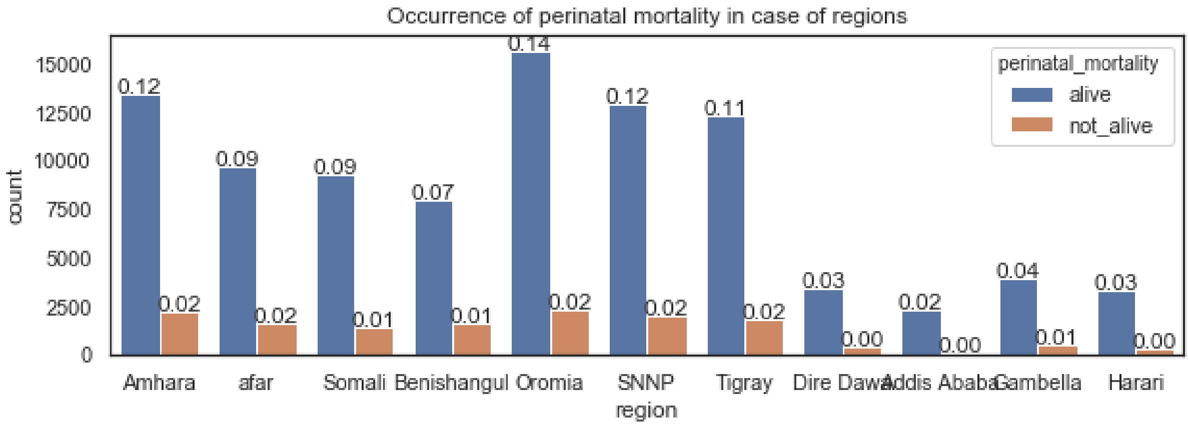

Figure 3: The Perinatal Mortality across the Regions in Ethiopia

Having health insurance service leads to almost zero perinatal mortality because it helps mothers to get affordable access to a health facility in time, as shown in Figure 4.

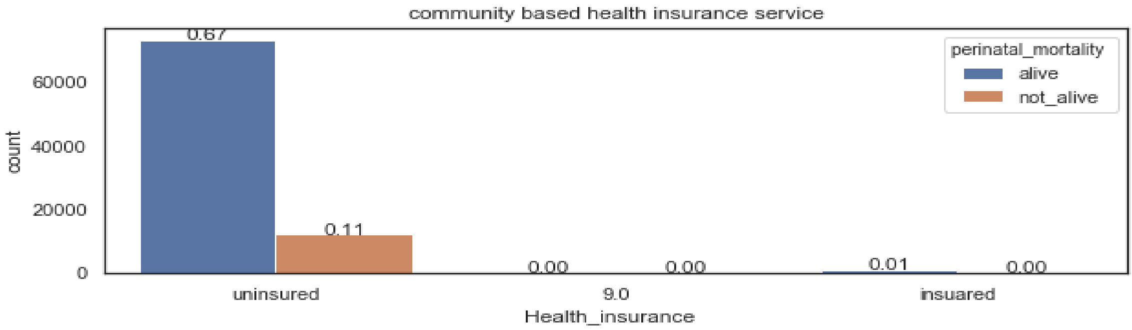

Figure 4 Perinatal Mortality Compared to Health Insurance

4.2. *Which homogeneous ensemble machine learning algorithm predicts perinatal mortality in Ethiopia effectively?*

Three experiments were conducted to build a perinatal mortality prediction model using classification algorithms namely; Gradient Boost, CatBoost, and random forest classifiers. A tenfold cross-validation method was implemented during the training. To select an optimal model with better performance, grid search was used to tune the best parameters of each classification algorithm. The values for parameters of each homogeneous ensemble machine learning algorithm are identified by Grid search for all experiments.

Table 3: Tuned parameters using grid-search

Algorithm	Parameter	values
Gradient Boost	Criterion	friedman_mse
	max_depth	15
	n_estimators	100

	Random_state	42
Random forest	Criterion	entropy
	max_features	sqrt
	min_samples_split	3
	n_estimators	200
	random_state	0
	max_depth	20
	max_leaf_nodes	400
	n_jobs	-1
Cat boost	Verbose	False
	eval_metric	AUC
	Iterations	500
	thread_count	None
	random_state	1.0

As a result, gradient boosting with the aforementioned, see Table 2, parameters performed better with 99.72% recall, 90.24% accuracy, 92.80% f1-score, 86.96% ROC and 87.24% precision. The recall indicates that there is a maximized true positive rate and a minimized false-negative rate meaning; there is a minimum false-negative rate. The confusion matrix of Gradient Boosting, Cat Boost and Random Forest algorithms is presented in Table 3.

Gradient boosting		*Predicted Class*	
		Died	***Alive***
Actual class	***Died***	*8125*	*2736*
	Alive	*167*	*18704*
Cat Boost		Predicted Class	
		Died	**Alive**
Actual class	**Died**	7056	3764
	Alive	1750	17162
Random Forest		*Predicted Class*	
		Died	***Alive***
Actual class	***Died***	*8033*	*2902*
	Alive	*36*	*18711*

Table 4: Confusion matrix

By using all the tuned parameters (see table 2), we have developed a predictive model using gradient boosting, cat boost, and random forest algorithms. Based on the tuned parameters, we have also evaluated the developed predictive model using different evaluation metrics, and the gradient boost algorithm is selected as the best homogenous ensemble machine learning algorithm for predicting perinatal mortality based on maternal health status and health insurance service in Ethiopia. The overall results of each experiment are summarized in Table 4.

Evaluation	**Algorithms**		
	Gradient Boost (%)	Cat Boost (%)	Random forest (%)
Accuracy	**90.24**	81.45	89.95
Precision	87.24	82.01	86.42
Recall	99.72	90.75	99.54
ROC	96.96	87.98	96.50
F1_Score	92.80	86.16	92.72
K-fold cross-validation for k = 10	94.72	88.26	90.78
Specificity	74.81%	82.01%	86.57%

Table 5: Overall performance of models

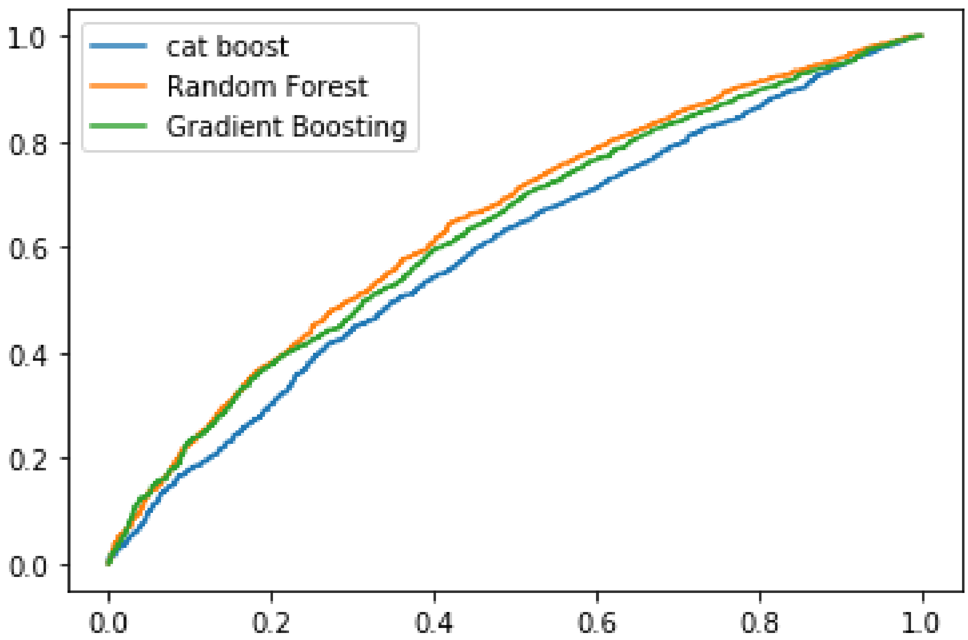

Figure 5: ROC AUC curve

4.3. What are the determinant factors of perinatal mortality in Ethiopia?

Feature importance analysis was conducted to identify risk factors of perinatal mortality in Ethiopia using the best-performing model which was developed using gradient boosting. As a result, factors that are significantly associated with perinatal mortality in Ethiopia Birth interval, Region, Maternal age, Wealth index, Maternal occupation, Anemia level, Visited health facility last 12 week, Marital status, Current contraceptive, Types of cooking fuel, Educational level, Preterm, Hemoglobin level, Place of residence, Community/mutual health insurance, Smoke cigarettes and Chews tobacco, see Table 5. Besides, community/mutual health insurance is one of the risk factors that are associated with perinatal mortality in Ethiopia. We have also conducted the feature importance's analysis for using the models that were developed using Cat Boost and Random Forest algorithms and the results are presented in appendix II.

No	Feature code	Feature description	Feature importance value
1.	Bord	Birth interval	0.291119
2.	V024	Region	0.122834
3.	V013	Maternal age	0.077887
4.	V190	Wealth index	0.072292
5.	V717	Maternal occupation	0.055206
6.	V457	Anemia level	0.054080
7.	V394	Visited health facility last 12 week	0.038728
8.	V501	Marital status	0.030207
9.	V312	Current contraceptive	0.026625
10.	V161	Types of cooking fuel	0.026059
11.	V106	Educational level	0.021210
12.	V228	Preterm	0.019435
13.	V455	Hemoglobin level	0.016383
14.	V025	Place of residence	0.009877
15.	V481a	Community/mutual health insurance	0.007053
16.	V463a	Smoke cigarettes	0.006037
17.	V463c	Chews tobacco	0.002598

Table 6:- Risk factors that were identified using feature importance analysis

4.4. What are the important rules that may shape strategies, policies, and interventions toward reducing and/or preventing perinatal mortality in Ethiopia?

The most relevant rules were generated from the best-performed algorithm (gradient boost) model, and the rules were validated by the domain experts who works at the University of Gondar comprehensive specialized hospital. Sample rules are presented in Appendix I and Figure 5 presents a decision tree of relevant rules that were generated by the best-performing algorithm:

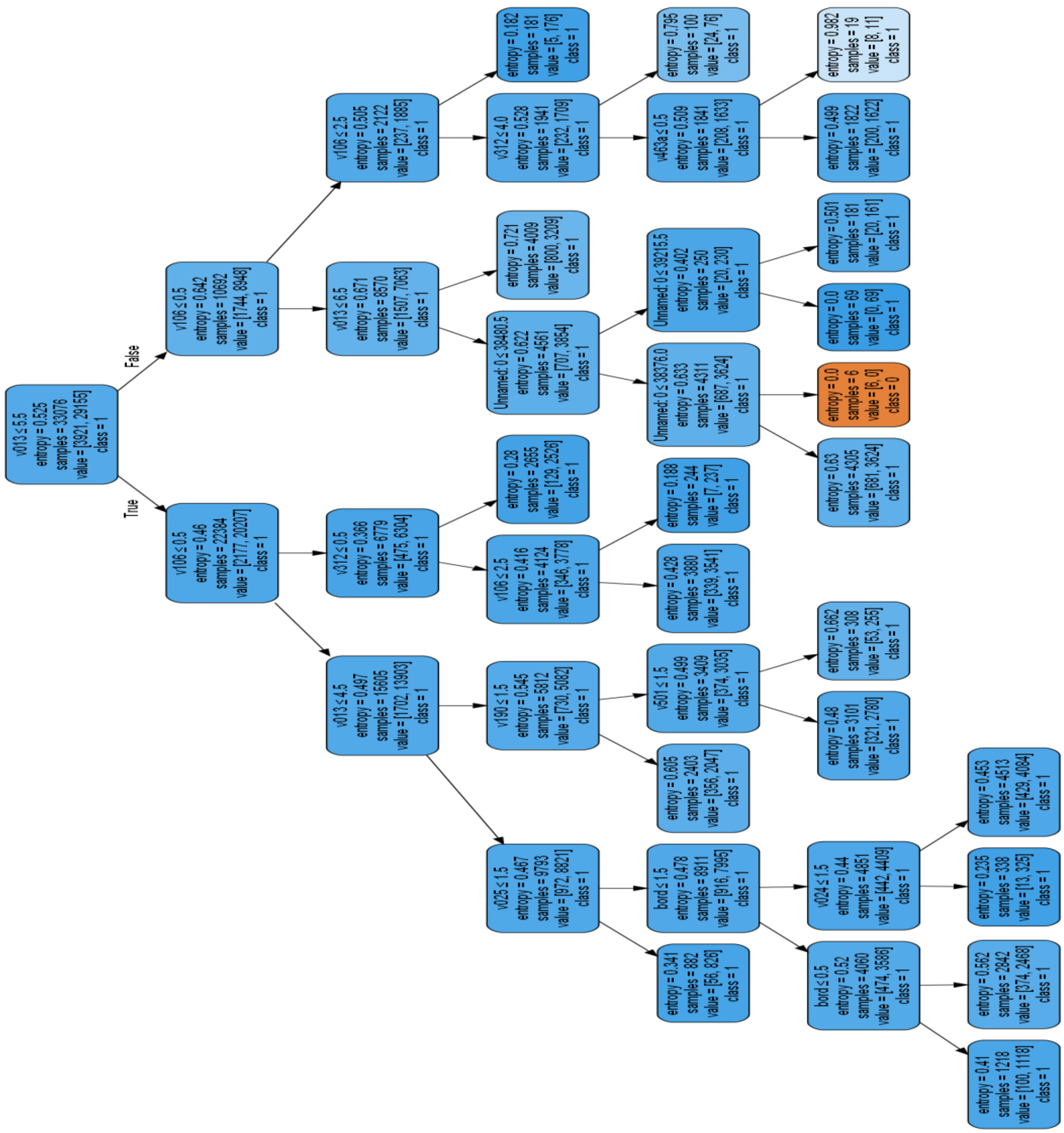

Figure 6:- generated rule by decision tree

5. DISCUSSION

Predicting perinatal mortality plays an important role to prevent and/or control child mortality in Ethiopia and maintain the health of future generations. This study, hence, targets to investigate the underlying structure of perinatal mortality, develop a machine learning model that predicts perinatal mortality, identify risk factors of perinatal mortality in Ethiopia and recommend relevant rules that help different stakeholders to make an evidence-based decision towards combating perinatal mortality in Ethiopia. To this end, data were collected from the Ethiopian demographic health survey and pre-processed using different methods. The explanatory analysis shows that perinatal mortality has been reducing over time in Ethiopia due to an increase in the number of hospitals in rural areas and the introduction of community-based health insurance which encourages pregnant women to visit the hospital to give birth. The proposed model, which was developed using a gradient boosting machine learning algorithm, achieved an overall performance of 90.24%, which is a better result compared to a result achieved by previous studies [15][16][24] which achieved 83% overall performance using gradient boosting machine learning algorithm. Based on the feature importance analysis, perinatal mortality in Ethiopia is significantly associated with health insurance and maternal health status such as birth interval of less than 2 years, preterm delivery, anemia, congenital anomaly, educational status, family size, occupation, marital status, perceived quality of care, the first choice of place for treatment during illness, previous history of perinatal death, not receiving tetanus toxoid immunization, and lack of iron supplementation. Besides, relevant rules, that facilitate evidence-based decision, were generated using the best-performing algorithm. The proposed model was deployed on the cloud using Heroku and Flask framework and can be freely accessed by potential users via this link: http://perinatal-mortality.herokuapp.com/

6. CONCLUSION

Perinatal mortality refers to fatal death at or after 28 weeks of pregnancy (stillbirth) and includes death within 7 days of life after birth. Among the 7.7 million deaths in 2018 attributed to children aged below five years, 3.1 million were in their perinatal period within 28 weeks and the first week of life. This study, thus, aimed to develop a machine learning model that predicts perinatal mortality in Ethiopia using homogeneous ensemble machine learning methods. To this end, design science research approach was employed and the proposed model was constructed using homogeneous ensemble machine learning algorithms namely gradient boost, random forest, and cat boost. The gradient boost algorithm has registered the highest performance with 99.72% recall, 90.24% accuracy, 92.80% f1-score, 86.96% ROC and 87.24% precision. We have also identified the determinant risk factors by conducting a feature importance analysis on the best-performed algorithms and some of the most determinant risk factors were maternal residence, level of education, birth interval, and community-based health insurance. The most relevant rules, that help to formulate evidence-based strategies and policies towards preventing, controlling and/or ending perinatal mortality in Ethiopia, were generated from the best-performing model, and the rules were validated by the domain experts. In the future, additional and recent data will be collected and heterogeneous ensemble machine learning algorithms will be considered.

LIST OF ABBREVIATIONS

ANC	Antenatal Care	PNC	Postnatal Care
CBHI	Community-Based Health Insurance	ROC	Receiver Operating Characteristics
EDHS	Ethiopia Demographic Health Survey	SDGs	Sustainable Development Goals
MH	Maternal Health	SMOTE	Synthetic Minority Over-Sampling Technique
OOP	Out-Of-Pocket	WHO	World Health Organization

Acknowledgment

We would like to acknowledge the Ethiopia central statistical agency for providing us with the data.

Funding

The research was supported by the University of Gondar research and community service vice president's office.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the 'perinatal_dataset-' repository on GitHub: https://github.com/dawitemu1/perinatal_dataset-

Declarations

Ethics approval and consent to participate

All methods used in this study followed guidelines and regulations that were approved by the institutional review board of the University of Gondar. Members of the board are Professor Feleke Moges, Mr. Niguse Yigzaw, Mr. Abiyot Endale, Dr. Misaye Mulate, Dr. Alemayehu Tekelu, and Dr. Bimerew Admasu.

Authors' Contribution

Bogale D.S and Abuhay T.M conceived and designed the study, participated in data analysis, wrote the report, finished the model refinements, carried out deep analysis of the experiment results, drafted and revised the manuscript; Abuhay T.M also managed the quality and progress of the whole study, and revised the manuscript; Dejene B.E revised the manuscript. All authors read and approved the final manuscript.

Consent for publication

Not applicable.

Competing interests

The authors report that they have no conflicts.

Author Details

All authors of this study are affiliated with the University of Gondar, College of Informatics, Gondar, Ethiopia.

Reference

[1] A. B. Comfort, L. A. Peterson, and L. E. Hatt, "Effect of health insurance on the use and provision of maternal health services and maternal and neonatal health outcomes: A systematic review," J. Heal. Popul. Nutr., vol. 31, no. 4 SUPPL.2, 2013, doi: 10.3329/jhpn.v31i4.2361.

[2] V. Jain and J. M. Chatterjee, "Machine Learning with Health Care Perspective: Machine Learning and Healthcare," no. March, 2020, doi: 10.1007/978-3-030-40850-3.

[3] R. Rasaily et al., "Effect of home-based newborn care on neonatal and infant mortality: A cluster randomised trial in India," BMJ Glob. Heal., vol. 5, no. 9, pp. 1–11, 2020, doi: 10.1136/bmjgh-2017-000680.

[4] J. R. Daw, T. N. A. Winkelman, V. K. Dalton, K. B. Kozhimannil, and L. K. Admon, “Medicaid expansion improved perinatal insurance continuity for low-income women,” Health Aff., vol. 39, no. 9, pp. 1531–1539, 2020, doi: 10.1377/hlthaff.2019.01835.

[5] B. J. Akombi and A. M. Renzaho, “Perinatal mortality in sub-saharan africa: A meta-analysis of demographic and health surveys,” Ann. Glob. Heal., vol. 85, no. 1, pp. 1–8, 2019, doi: 10.5334/aogh.2348.

[6] P. R. Ghimire, K. E. Agho, A. M. N. Renzaho, M. K. Nisha, M. Dibley, and C. Raynes-Greenow, “Factors associated with perinatal mortality in Nepal: Evidence from Nepal demographic and health survey 2001-2016,” BMC Pregnancy Childbirth, vol. 19, no. 1, pp. 1–12, 2019, doi: 10.1186/s12884-019-2234-6.

[7] Z. A. Hassan and M. J. Ahmed, “Factors associated with immunisation coverage of children aged 12- 24 months in Erbil / Iraq 2017-2018,” vol. 24, no. 08, pp. 12222–12235, 2020, doi: 10.37200/IJPR/V24I8/PR281205.

[8] N. Haven et al., “Community-based health insurance increased health care utilization and reduced mortality in children under-5, around Bwindi Community Hospital, Uganda between 2015 and 2017,” Front. Public Heal., vol. 6, no. OCT, 2018, doi: 10.3389/fpubh.2018.00281.

[9] E. Jukes, Encyclopedia of Machine Learning and Data Mining (2nd edition), vol. 32, no. 7/8. 2018. doi: 10.1108/rr-05-2018-0084.

[10] W. Soors, N. Devadasan, V. Durairaj, and B. Criel, “Community Health Insurance and Universal Coverage: Multiple paths, many rivers to cross,” 48, pp. 1–122, 2010.

[11] R. A. Knuppel and J. H. Shepherd, “Perinatal mortality rates,” Br. Med. J., vol. 280, no. 6228, p. 1376, 1980, doi: 10.1136/bmj.280.6228.1376.

[12] B. H. Jena, G. A. Biks, K. A. Gelaye, and Y. K. Gete, “Magnitude and trend of perinatal mortality and its relationship with inter-pregnancy interval in Ethiopia: A systematic review and meta-analysis,” BMC Pregnancy Childbirth, vol. 20, no. 1, pp. 1–13, 2020, doi: 10.1186/s12884-020-03089-2.

[13] G. T. Debelew, “Magnitude and Determinants of Perinatal Mortality in Southwest Ethiopia,” J. Pregnancy, vol. 2020, 2020, doi: 10.1155/2020/6859157.

[14] G. Bayou and Y. Berhan, “Perinatal mortality and associated risk factors: a case control study.,” Ethiop. J. Health Sci., vol. 22, no. 3, pp. 153–62, 2012, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23209349%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3511893

[15] E. Mfateneza, P. C. Rutayisire, E. Biracyaza, S. Musafiri, and W. G. Mpabuka, “Application of machine learning methods for predicting infant mortality in Rwanda : analysis of Rwanda demographic health survey 2014 – 15 dataset,” BMC Pregnancy Childbirth, pp. 1–13, 2022, doi: 10.1186/s12884-022-04699-8.

[16] F. H. Bitew, S. H. Nyarko, L. Potter, and C. S. Sparks, “Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey,” Genus, vol. 76, no. 1, 2020, doi: 10.1186/s41118-020-00106-2.

[17] Y. Getiye and M. Fantahun, “Factors associated with perinatal mortality among public health deliveries in Addis Ababa, Ethiopia, an unmatched case control study,” BMC Pregnancy Childbirth, vol. 17, no. 1, pp. 1–7, 2017, doi: 10.1186/s12884-017-1420-7.

[18] B. Demissie and K. G. Negeri, “Effect of community-based health insurance on utilization of outpatient health care services in southern ethiopia: A comparative cross-sectional study,” Risk Manag. Healthc. Policy, vol. 13, pp. 141–153, 2020, doi: 10.2147/RMHP.S215836.

[19] J. Chodziwadziwa Kabudula, Reesha Kara, Henry Wandera, Fidelia A. A. Dake and C. M. Dansou, Dereje Danbe Debeko, “Evaluation of machine learning methods for predicting the risk of child mortality in South Africa,” vol. 2019, no. 2019, pp. 1–5, 2019, [Online]. Available: https://www.uaps2019.popconf.org

[20] G. Nguyen, “Evaluating statistical and machine learning methods to predict risk of in-hospital child mortality in Uganda,” 2016, [Online]. Available: https://digital.lib.washington.edu/researchworks/handle/1773/36994?show=full

[21] K. H. Haseeb Ali, Mohd Najib Mohd Salleh, Rohmat Saedudin and M. F. M. Faheem, “Imbalance class problems in data mining: A review,” Indones. J. Electr. Eng. Comput. Sci., vol. 14, no. 3, pp. 1552–1563, 2019, doi: 10.11591/ijeecs.v14.i3.pp1552-1563.

[22] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, “SMOTE for regression,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8154 LNAI, pp. 378–389, 2013, doi: 10.1007/978-3-642-40669-0_33.

[23] A. D. Keromytis, “Comparative Analysis,” SpringerBriefs Comput. Sci., vol. 1, pp. 57–60, 2011, doi: 10.1007/978-1-4419-9866-8_5.

[24] D. P. Gnawali et al., “The effect of community-based health insurance on the utilization of modern health care services: Evidence from Burkina Faso,” Health Policy (New. York)., vol. 90, no. 2–3, pp. 214–222, 2009, doi: 10.1016/j.healthpol.2008.09.015.