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

Dawit S Bogale
University of Gondar

Tesfamariam M Abuhay (tesfamariam.mabuhay@uog.edu.et)
University of Gondar

Belayneh E Dejene
University of Gondar

Research Article

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

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1445740/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on December 28th, 2022. See the published version at https://doi.org/10.1186/s12911-022-02084-1.

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 for 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, can 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 Ethiopian demographic health survey from 2011 to 2019 G.C. The data were pre-processed to get quality data that are suitable for a homogenous ensemble machine-learning algorithm to develop a model that predicts perinatal mortality.

Results

For constructing the proposed model, three experiments were conducted using random forest, gradient boosting, and cat boost algorithms. The overall accuracy of random forest, gradient boosting, and cat boost with 17 features is 89.95%, 90.24%, and 82%, respectively.

Conclusions

We finally concluded that perinatal mortality over time in Ethiopia is decreasing. We found out that perinatal mortality in Ethiopia is associated with risk factors such as community-based health insurance, mother's educational level, residence, mother age, wealth status, distance to the health facility, preterm, smoke cigarette, anemia level, haemoglobin level, and marital status.

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 reaching deadlines for reaching the Millennium Development Goals, the international community supports low- and middle-level income countries to renew their commitment to reducing maternal and infants 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 Intrapartum deaths (death during the delivery) [8]. Ethiopia shared and valued the Sustainable Development Goals (SDGs) and has been trying to achieve the target of reducing neonatal mortality to below 12 per 1000 live births, by 2030 [9]. However, reduction of neonatal, infant and under-five mortalities cannot be realized without substantial reduction of perinatal mortality [10]. It is mostly associated with home deliveries, which contributed for 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 [9].

Financial constraints have a significant impact on timely access to maternal health (MH) care, such as Antenatal Care (ANC), skilled care at delivery, access to facility-based deliveries, postnatal care (PNC), and perinatal [7]. Over 100 million individuals pay out-of-pocket (OOP) payments to get health treatments that have proven difficult to obtain for millions of poor people, resulting in increased morbidity and mortality [5]. WHO recommends community-based health insurance (CBHI) as one of the approaches for reducing OOP expenditures for registered families which, in turn, reduce morbidity and mortality [6]. The association of CBHI with reduced maternal and infant mortality was apparent but it is impossible to reduce the infant mortality rate, without reducing the perinatal mortality [7]. Financial incentives, such as health insurance, can address the demand- and supply factors that may possibly impacting 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 survival of newborn and children [9]. 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 [8][12][13].

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 methods 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 towards 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 and result 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 were 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 haemorrhage, and hypertensive disorders of pregnancy was an independent predictor for high perinatal mortality.

Another study was conducted by Yemisrach et al. [15] 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 up 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. Use of partograph was also associated with decreased risk of perinatal mortality. Bekele et al. [16] studied the effect of community-based health insurance on utilization of outpatient health care services in Yirgalem town, Southern Ethiopia. This study used both quantitative and qualitative (mixed) approaches using a comparative cross-sectional study design. Randomly selected sample of 405 (135 members and 270 non-members) household heads were used for quantitative analysis. Multivariate logistic regression was employed to identify the effect of community-based health insurance on healthcare utilization. This study reveals 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. Besides, these studies did not develop a predictive model, did not design an artefact 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 their effectiveness and efficiency in predicting child mortality in African countries such as South Africa [17] and Uganda [18]. This study, hence, motivated to fill these gaps by identifying risk factors, constructing a predictive model, design artifact, and generate rules that help to develop evidence-based policies and interventions towards perinatal mortality in Ethiopia.

3. Materials And Methods

Figure 1 depicts the proposed model architecture that was implemented in this study to construct a predictive model, identify risk factors, extract relevant rules, and design artifacts.

3.1. DATA COLLECTION

In this study we used secondary data, 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. The EDHSs are nationally representative household surveys that collect data for a variety of demographic, health, and nutrition monitoring and impact evaluation purposes.

3.2. DATA PREPROCESSING

The raw data contains 45 columns and 109531 instances. Data imputation (mode for categorical data and mean for continuous data) method was employed to substitute the missing values. Outliers were identified using a boxplot and replaced using the Interquartile Range (IQR) scores. Binning 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 use SMOTE is it avoids loss of valuable information [22][23]. Then, four experiments were conducted using filter and wrapper methods to select the relevant features for developing a perinatal mortality prediction model. As a result, the sequential backward feature selection method has registered the highest performance with 90.5% of accuracy and produced 13 important features. Besides, the domain expert’s recommended additional 4 features and the total features selected for further analysis are 17, see Table 1.

Table 1
Features selected by sequential forward feature selection

No	Feature code	Feature description
1.	Bord	Birth interval
2.	V024	Region
3.	V013	Maternal age
4.	V190	Wealth index
5.	V717	Maternal occupation
6.	V457	Anemia level
7.	V394	Visited health facility last 12 week
8.	V501	Marital status
9.	V312	Current contraceptive
10.	V161	Types of cooking fuel
11.	V106	Educational level
12.	V228	Preterm
13.	V455	Hemoglobin level
14.	V025	Place of residence
15.	V481a	Community/mutual health insurance
16.	V463a	Smoke cigarettes
17.	V463c	Chews tobacco

4. Experimental Setup And Results In Discussion

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

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

The perinatal mortality across the regions (Tigray, afar, Amhara, Oromia, Somali, Benishangul, SNNP, Gambella, Harari, Addis Ababa, and Dire Dawa) of Ethiopia was also investigated. Among nine regions

and two administrative cities of Ethiopia, Amhara and Oromia regions registered higher perinatal mortality compared to other regions, as shown in Fig. 3.

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

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

Three experiments were conducted to build a perinatal mortality predictive model using classification algorithms namely: Gradient Boost, CatBoost, and random forest classifiers. Grid search was applied to tune the hyperparameters of these algorithms. As a result, gradient boosting (with parameters: criterion='entropy', max_depth = 15, max_Depth = max_Depth, bootstrap = True, N_estimators = warn, N_jobs = none, random state = 42) 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 algorithms is presented in Table 2.

Table 2
Confusion matrix of gradient boosting

		Predicted Class	
		Died	**Alive**
Actual class	**Died**	8125	2736
	Alive	167	18704

Therefore, 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 the study area. The overall results of each experiment are summarized in Table 3.

Table 3
Overall performance of models

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	86.96	77.98	86.50
F1_Score	92.80	86.16	92.72

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

Feature importance analysis was conducted to identify determinant 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 increased risk of perinatal mortality are birth interval less than 2 years, preterm delivery, anemia, congenital anomaly, educational status, family size, occupation, marital status, traveling time to the nearest health institution, perceived quality of care, the first choice of place for treatment during illness and expected healthcare cost of recent treatment, prematurity, low birth weight, previous history of perinatal death, not receiving tetanus toxoid immunization, and lack of iron supplementation, see Table 4.

Table 4
Risk factors with feature importance

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

4.4 What are the important rules that may shape strategies, policies and interventions towards 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. Sample rules are presented here below and Fig. 5 presents decision tree of relevant rule that were generated by the best performing algorithm:

Rule1:- if currently breast feeding and preterm == 'no' AND maternal education== 'no education' AND wanted least children == 'wanted then' AND smoke ciggrate == 'no' AND health insurance provide by employer == 'no' AND smoke Tabaco == 'never in union' AND types of cooking fuel == 'wood' AND occupation == 'not working AND wealth index== 'poorest' AND maternal age == '35–39' AND place of residence == 'rural' AND Community based health insurance == 'no' AND Then child=='Alive'

Rule 2:- if currently breast feeding and preterm == 'no' AND maternal education== 'no education' AND wanted least children == 'wanted then' AND smoke ciggrate == 'no' AND health insurance provide by employer == 'no' AND smoke Tabaco == 'never in union' AND types of cooking fuel == 'wood' AND occupation == 'not working AND wealth index== 'poorest' AND maternal age == '40–44' AND place of residence == 'urban' AND Community based health insurance == 'no' AND Then child=='Died'

Rule3:- if currently breast feeding and preterm == 'no' AND maternal education== 'no education' AND wanted least children == 'wanted then' AND smoke ciggrate == 'no' AND health insurance provide by employer == 'no' AND smoke Tabaco == 'never in union' AND types of cooking fuel == 'wood' AND occupation == 'not working AND wealth index== 'poorest' AND maternal age == '45–49' AND place of residence == 'rural' AND Community based health insurance == 'no' AND Then children=='Alive'

Rule4:- if currently breast feeding and preterm == 'no' AND maternal education== 'no education' AND wanted least children == 'wanted then' AND smoke ciggrate == 'no' AND health insurance provide by employer == 'no' AND smoke Tabaco == 'never in union' AND types of cooking fuel == 'wood' AND occupation == 'not working AND wealth index== 'poorest' AND maternal age == '45–49' AND place of residence == 'rural' AND Community based health insurance == 'no' AND Then children=='Alive'

Rule 5:- if currently breast feeding and preterm == 'no' AND maternal education== 'no education' AND wanted least children == 'wanted then' AND smoke ciggrate == 'no' AND health insurance provide by employer == 'no' AND smoke Tabaco == 'never in union' AND types of cooking fuel == 'wood' AND occupation == 'not working AND wealth index== 'poorest' AND maternal age == '15–19' AND place of residence == 'rural' AND Community based health insurance == 'no' AND Then children== 'Died'

5. Discussion

As we discussed in risk factors identification section, the risk factors were identified using feature importance techniques and rules were generated using best performing algorithm which is gradient

boosting as show in the experimental section. As we have discussed in the experimental result section, the proposed system achieved an overall performance of 90.24%, which is better a result compared to a result achieved by previous studies using gradient boosting machine learning algorithm which was 83% overall performance. We have deployed the model on cloud using Heroku and Flask framework and can be freely accessed via this link: http://perinatal-mortality.herokuapp.com/

6. Conclusion

This study aims at developing a model that predicts perinatal mortality in the case of Ethiopia by using homogeneous ensemble machine learning methods. 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 identified the determinant risk factors of perinatal mortality with feature importance techniques such as maternal residence, level of education, birth interval, and community-based health insurance. The most relevant rules, that helps to formulate evidence-based strategies and policies towards maintaining perinatal mortality, were generated from the best performing model, and the rules were validated by the domain experts.

Abbreviations

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

Declarations

Acknowledgements

We would like to acknowledge the Ethiopia central statistical agency for providing us 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 analysed during the current study are available in the ‘perinatal_dataset-’ repository, https://github.com/dawitemu1/perinatal_dataset-.

Ethics approval and consent to participate

All methods used in this study followed guidelines and regulationsthatwere approved by the institutional reviewboardof the University of Gondar. Members of the board are Professor FelekeMoges, Mr. NiguseYigzaw, Mr. AbiyotEndale, Dr. MisayeMulate, Dr. AlemayehuTekelu and Dr. BimerewAdmasu.

Consent for publication

Not applicable.

Competing interests

The authors report that they have no conflicts.

References

[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] E. Jukes, Encyclopedia of Machine Learning and Data Mining (2nd edition), vol. 32, no. 7/8. 2018.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[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.

[15] 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.

[16] 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.

[17] C. Kabudula et al., “Evaluation of machine learning methods for predicting the risk of child mortality in South Africa,” 2019.

[18] G. Nguyen, “Evaluating statistical and machine learning methods to predict risk of in-hospital child mortality in Uganda,” 2016.

[19] D. D. Atnafu, H. Tilahun, and Y. M. Alemu, “Community-based health insurance and healthcare service utilisation, North-West, Ethiopia: A comparative, cross-sectional study,” BMJ Open, vol. 8, no. 8, pp. 1–6,

2018, doi: 10.1136/bmjopen-2017-019613.

[20] K. Shiferaw, B. Mengiste, T. Gobena, and M. Dheresa, “The effect of antenatal care on perinatal outcomes in Ethiopia: A systematic review and meta-analysis,” PLoS One, vol. 16, no. 1 January, pp. 1–19, 2021, doi: 10.1371/journal.pone.0245003.

[21] D. Prasad et al., “The effect of community-based health insurance on the utilization of modern health care services : Evidence from Burkina Faso,” vol. 90, pp. 214–222, 2009, doi: 10.1016/j.healthpol.2008.09.015.

[22] I. Journal and C. Science, “Class Imbalance Problem in Data Mining : Review,” vol. 2, no. 1, 2013.

[23] R. P. Ribeiro, “SMOTE for Regression,” no. October 2015, 2013, doi: 10.1007/978-3-642-40669-0.

Figures

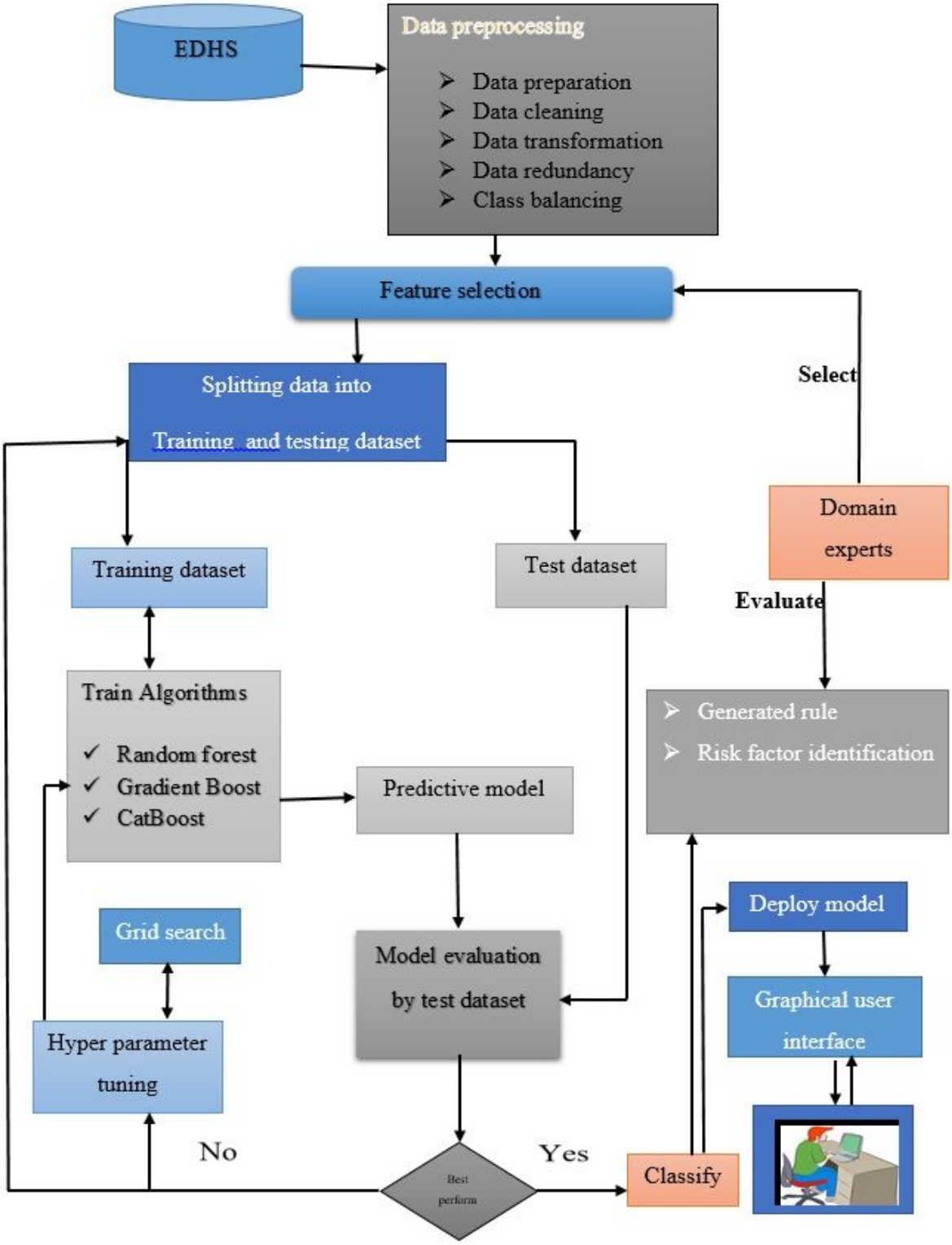

Figure 1

The proposed model architecture

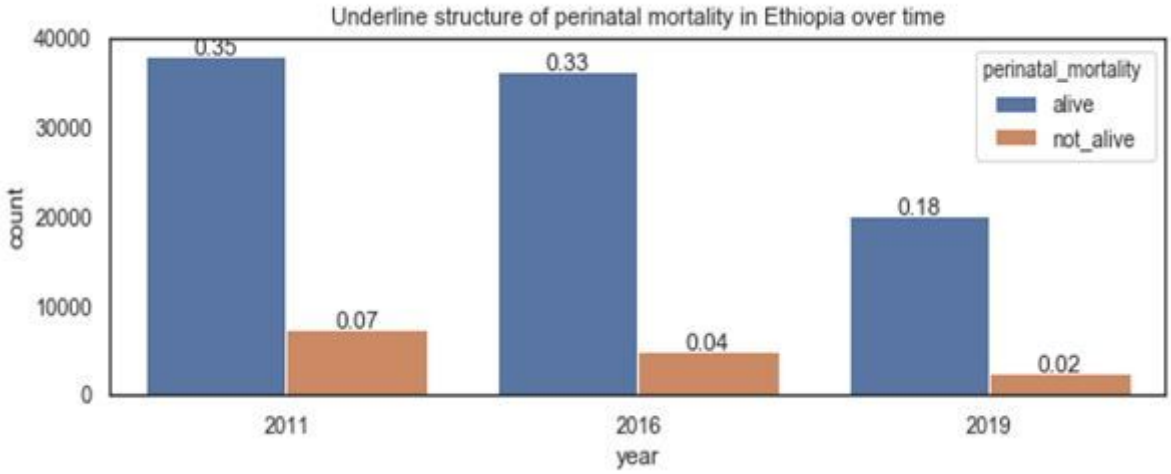

Figure 2

Perinatal mortality over time in Ethiopia

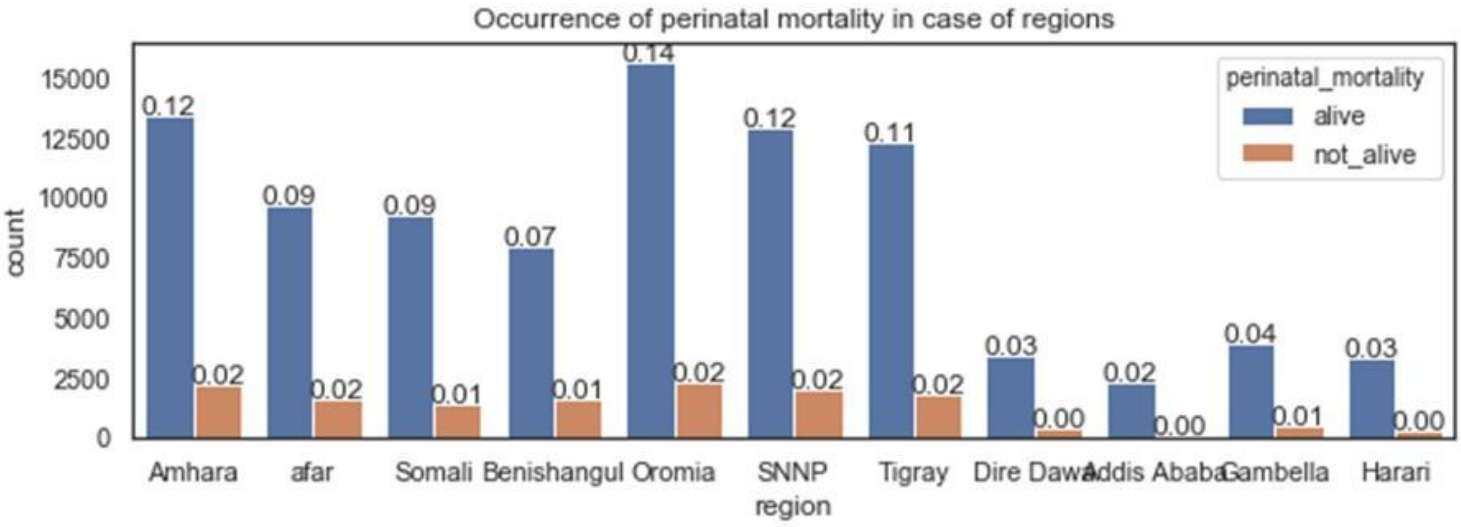

Figure 3

The Perinatal Mortality across the Regions in Ethiopia

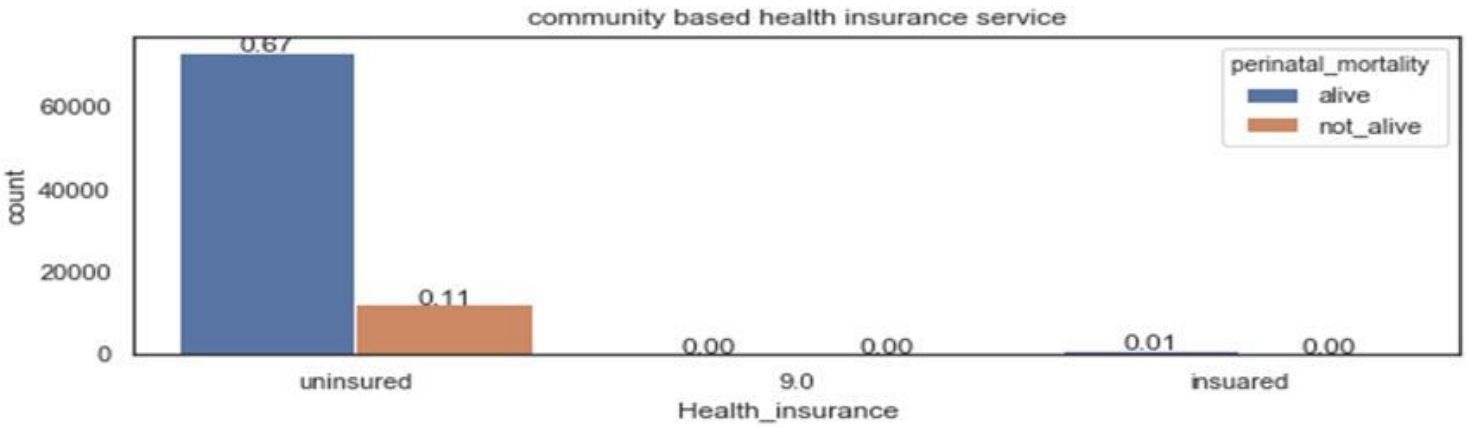

Figure 4

Perinatal Mortality Compared to Health Insurance

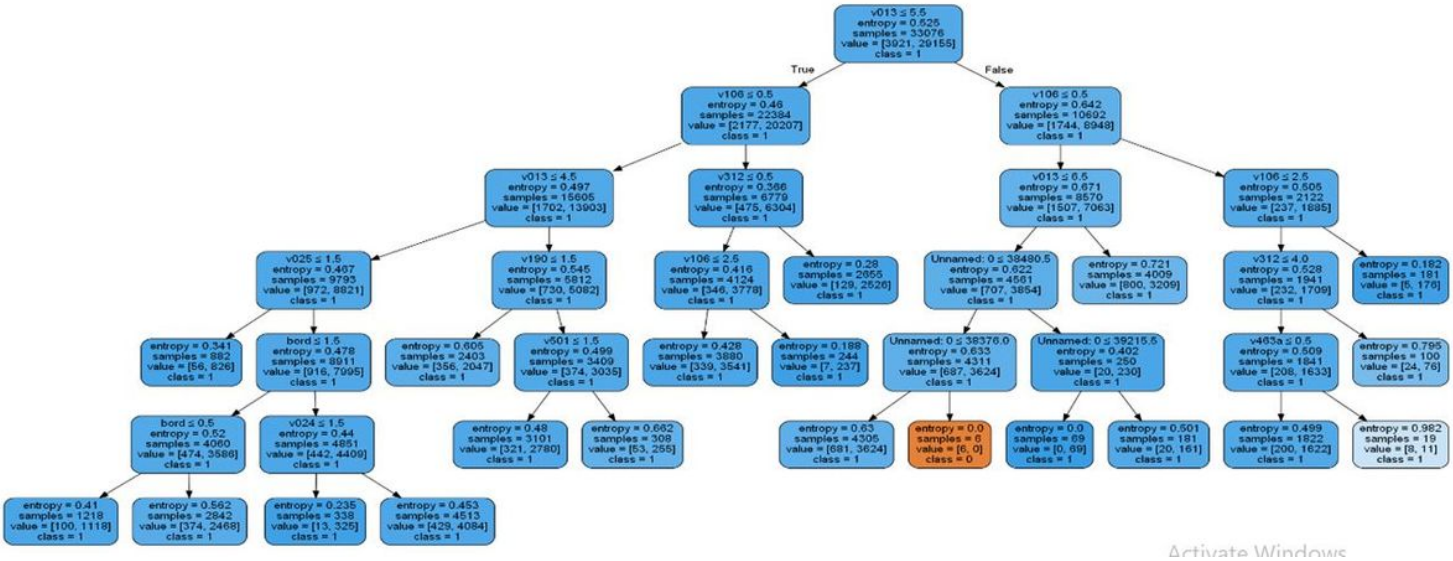

Figure 5

generated rule by decision tree