Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

# Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia

Herdiantri Sufriyana, MD[a,b], Yu-Wei Wu, PhD[b,a,c], Emily Chia-Yu Su, PhD[a,c,d,*]

[a] Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan
[b] Department of Medical Physiology, College of Medicine, University of Nahdlatul Ulama Surabaya, Surabaya 60237, Indonesia
[c] Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei 11031, Taiwan
[d] Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei 11031, Taiwan

ARTICLE INFO

*Article History:*
Received 29 November 2019
Revised 24 February 2020
Accepted 24 February 2020
Available online xxx

*Keywords:*
Artificial intelligence
Preeclampsia
Clinical prediction rule
Machine learning
Natural language processing
Health insurance dataset

ABSTRACT

*Background:* We developed and validated an artificial intelligence (AI)-assisted prediction of preeclampsia applied to a nationwide health insurance dataset in Indonesia.
*Methods:* The BPJS Kesehatan dataset have been preprocessed using a nested case-control design into preeclampsia/eclampsia ($n$ = 3318) and normotensive pregnant women ($n$ = 19,883) from all women with one pregnancy. The dataset provided 95 features consisting of demographic variables and medical histories started from 24 months to event and ended by delivery as the event. Six algorithms were compared by area under the receiver operating characteristics curve (AUROC) with a subgroup analysis by time to the event. We compared our model to similar prediction models from systematically reviewed studies. In addition, we conducted a text mining analysis based on natural language processing techniques to interpret our modeling results.
*Findings:* The best model consisted of 17 predictors extracted by a random forest algorithm. Nine~12 months to the event was the period that had the best AUROC in external validation by either geographical (0.88, 95% confidence interval (CI) 0.88–0.89) or temporal split (0.86, 95% CI 0.85–0.86). We compared this model to prediction models in seven studies from 869 records in PUBMED, EMBASE, and SCOPUS. This model outperformed the previous models in terms of the precision, sensitivity, and specificity in all validation sets.
*Interpretation:* Our low-cost model improved preliminary prediction to decide pregnant women that will be predicted by the models with high specificity and advanced predictors.
*Funding:* This work was supported by grant no. MOST108-2221-E-038-018 from the Ministry of Science and Technology of Taiwan.

## 1. Introduction

Predicting preeclampsia may prevent neonatal morbidity because this disorder can lead to neonatal prematurity [1,2]. Admission to neonatal intensive care units (ICUs) was not reduced (odds ratio [OR] 0.93, 95% confidence interval (CI) 0.55–1.59), although preterm/early preeclampsia was prevented by aspirin administration at 11–13 weeks' gestation [3]. This is because term/late preeclamptic women are more common than preterm/early ones and only 85% of those were detected with a false positive rate of 10% at 35–37 weeks in high-resource settings [4]. A nationwide health insurance dataset of the BPJS Kesehatan in Indonesia can provide big data to develop artificial intelligence (AI)-assisted predictions that may reduce false positives. However, the predictive performance of predicting preeclampsia developed based on this health insurance dataset is still unclear.

Preeclampsia is one of the pregnancy-induced hypertension (PIH) and placenta dysfunction-related disorders [5,6]. Preeclampsia affects 4.6% (95% uncertainty range 2.7–8.2) of pregnant women [7], and may also impair fetal growth, which leads to low birth weights as a predisposing factor to neonatal deaths [8,9]. Although many conditions in pregnant women contribute to premature and low-birth-weight infants [10], preeclampsia is the major contributor, because early delivery is the only cure for this disease. Yet, the decision to deliver early may be based on a false positive that leads to inefficient

[*] Corresponding author at: Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan
*E-mail addresses:* herdiantrisufriyana@unusa.ac.id (H. Sufriyana), yuwei.wu@tmu.edu.tw (Y.-W. Wu), emilysu@tmu.edu.tw (E.C.-Y. Su).

https://doi.org/10.1016/j.ebiom.2020.102710

## Research in context

*Evidence before this study*

Preeclampsia is a subtype of pregnancy-induced hypertension (PIH) that is a common cause of maternal mortality. The etiology and pathogenesis are not well understood, but it was evidenced that the only cure is delivery; thus, false positives of preeclampsia predictions might lead to unnecessary early deliveries. This contributes to premature and low-birth-weight babies, which in turn, increases inefficient utilization of neonatal intensive care units (ICUs). There are four problems with preeclampsia prediction models from previous studies: (1) no robust prediction of all subtypes of preeclampsia; (2) biased predictive performances; (3) low precision or positive predictive values; and (4) a need for a high-resource setting to apply the prediction model. Most models achieved greater than 90% sensitivity and specificity only for early-onset but not all subtypes of preeclampsia. Meanwhile, late-onset preeclampsia is more common than early-onset cases; thus, admission to neonatal ICUs was not significantly reduced. A previous study showed promising predictive performance of a prediction model for late-onset preeclampsia. Yet, this did not apply recent standards that have been developed for predictive modeling, which were designed to avoid risks of bias. Ultimately, low precision of prediction models is common in preeclampsia predictions because of class imbalances in which preeclampsia outcomes were very low compared to normotensive controls (mostly 1:9). In low-precision prediction models, a predicted preeclampsia case is likely to be a false positive, and this leads to unnecessary early deliveries. Although there are highly precise prediction models limited to early-onset preeclampsia, these require expensive, inaccessible biophysical and biochemical markers such as the pulsatility index of the uterine artery by ultrasound measurement, soluble fms-like tyrosine kinase-1 (sFlt-1), and/or placental growth factor (PlGF). We need a prediction model with low false positive rate and low-cost predictors with high sensitivity at the same specificity compared to the others with low-cost predictors. This model will be a preliminary model to decide utility of prediction models with advanced predictors. Therefore, we can reduce both maternal mortality and neonatal morbidity as well as the prediction cost of preeclampsia at community level.

*Added value of this study*

The prediction model proposed in this study was robust for preeclampsia in both internal and external validation sets. This model was developed based on the Prediction Model Risk of Bias Assessment Tool (PROBAST) which contains recent guidelines for prediction model development to avoid risks of bias. We compared the precision, including sensitivity and specificity, to similar prediction models from previous studies. These were systematically reviewed among 879 records from PUBMED, EMBASE, and SCOPUS within the last 5 years (since 2015). Our model applied a machine learning algorithm that uses demographic variables and diagnoses on previous visits which are conceivably applicable in low-resource settings. To develop and validate this model, we utilized big data from a nationwide health insurance dataset in Indonesia ($n$ = 2,641,096) with preeclampsia/eclampsia ($n$ = 3318) vs. non-PIH nested control ($n$ = 19,883) outcomes. Our model outperformed those from systematically reviewed studies in terms of both internal and external validation sets. For external validation, our precision levels were 0.59 (95% confidence interval (CI) 0.58−0.60; by geographical splitting) and 0.72 (95% CI 0.72−0.72; by temporal splitting) compared to the best previous model 0.17 (95% CI 0.17−0.17) at sensitivity ~95%. Meanwhile, the specificities were 0.47 (95% CI 0.45−0.49; by geographical splitting) and 0.44 (95% CI 0.43−0.45; by temporal splitting) compared to the best previous model of 0.47 (95% CI 0.40−0.55). The areas under the receiver operating characteristics curve of our model were 0.88 (95% CI 0.88−0.89) and 0.86 (95% CI 0.85−0.86) for the geographical and temporal splits, respectively. Subjects predicted as preeclampsia/eclampsia was ~80% in both external validation sets, which imply potential reduction of ~20% cost for prediction by highly-specific models with advanced predictors. We also applied natural language processing techniques to assist interpretation of our model, which is considered one of the most important artificial intelligence applications.

*Implications of all the available evidence*

Since our model showed an acceptable predictive performance using information from a health insurance dataset that came from multiple healthcare facilities, we encourage health insurance companies to facilitate this model deployment in order to be used by inter-healthcare facilities in privacy-aware information systems. We expect this model to have an impact on improving efficient neonatal ICU utilization and in turn reduce expenses of insurance companies. Our prediction model also supported several recent findings on preeclampsia pathogenesis. The best predictive performance of our model used predictors during 9−<12 months to the event. This supports recent findings from bioinformatics studies which revealed that preeclampsia pathogenesis possibly starts before pregnancy rather than during the first trimester. Approximately one-third of text profiling results from diagnoses on previous visits were bacterial infection-related conditions, as inferred by natural language processing techniques. This also corresponds to recent findings from microbiology and microbiome studies that provided evidence of the role of bacterial infections or specific microbial communities in several organs of women with preeclampsia.

utilization of neonatal ICUs for preventable premature babies. Although neonatal ICU admission was not reduced by babies from preeclamptic women given aspirin at 11−13 weeks' gestation [3], the length of stay in neonatal ICU was reduced by 20.3 days (95% CI 7.0−38.6) [11]. However, this was because of decreased birth rates at <32 weeks' gestation (OR 0.42; 95% CI, 0.19−0.93), or prevention of early preeclampsia. Meanwhile, the number of babies that were admitted to neonatal ICU was larger from term/late ($n$ = 14/102, 13.72%) preeclamptic women compared to those from preterm/late ones ($n$ = 7/102, 6.86%). By reducing length of stay without reducing neonatal ICU admission, the cost reduced mostly at individual but not at community level. Therefore, reducing false positives from preeclampsia predictions may improve the efficiency of utilization of these scarce facilities.

Predicting preeclampsia is important because the effective prevention is only applied for preterm preeclampsia (risk ratio [RR] 0.92, 95% CI 0.45−0.87) in which aspirin is given at ≤16 weeks' gestation [12]. Ninety predictors and 52 prediction models were compared by 126 systematic reviews, and 63.49% of them included advanced biomarkers, genomics, and/or ultrasound measures [13]. Nevertheless, few of those tests had both sensitivity and specificity above 90% in the external validation. Although there was an externally validated prediction model with a sensitivity of 93% (95% CI 76%−99%; at a specificity of 90%), this was true only for early but not all preeclampsia (sensitivity 49%, 95% CI 43%−56%) [14]. This model also used advanced biomarkers, while another model only achieved a sensitivity of 47.6% (95% CI 44.0–51.1%; at a specificity of 89.4%) using maternal characteristics and medical histories [15]. A prediction model with high precision and sensitivity but low-cost predictors is needed for preeclampsia. The model should have high sensitivity at the same or better specificity compared to the others with low-cost predictors. The model is intended to decide which patients will be predicted by other highly-specific models with advanced predictors. The preliminary prediction model will improve efficiency of neonatal ICU utilization and reduce the prediction cost at community level without sacrificing either maternal or neonatal patient safety.

The poor performances of preeclampsia prediction may be caused by the complexity of this disease at the transcriptomic level [16]. Machine learning can potentially deal with this problem [17]; however, it needs big data to achieve good predictive performances. A recent machine learning prediction study demonstrated a promising predictive performance in internal validation by a stochastic gradient-boosting algorithm for late preeclampsia (*c*-statistics 0.924; with a sensitivity of 0.60 and a specificity of 0.99) [18]. That study utilized electronic medical records consisting of 24 clinical and biochemical predictors, but there were only 474 events of preeclampsia. It lacked events per variable (EPVs) which may cause overfitting to several machine learning algorithms [19]. However, there are no previous studies that developed and externally validated prediction models for preeclampsia that utilized big data with sufficient EPVs for machine learning algorithms.

The Nationwide Health Insurance Dataset of BPJS Kesehatan (NHID-BPJSKes) in Indonesia can provide big data for developing machine learning prediction models. Health insurance datasets have been utilized for association studies involving PIH in Taiwan [20–23] and a predictive study for postpartum women in the UK [24]. Although only demographic data and diagnoses are provided by the NHID-BPJSKes, machine learning prediction models can be developed using this dataset since it provides sufficient EPVs. This is because a systematic review showed that Indonesia as one of the countries with high incidences of preeclampsia based on two studies (8.6%; *n* = 43,464) [7]. Although there is no effective prevention for late preeclampsia, machine learning trained on big data may provide a predictive model with better precision. By reducing false positives of pregnancy termination because of preeclampsia, it may eventually improve utilization of ICUs. In addition, this predictive model can be used to efficiently construct prospective cohorts of preeclampsia for further development of machine learning predictions, especially in poor-resource settings. This study attempted to develop and validate an AI-assisted prediction of preeclampsia by machine learning applied to the NHID-BPJSKes in Indonesia.

## 2. Materials and methods

This study developed a prognostic prediction model utilizing publicly-accessed dataset; thus, our study was non-interventional and non-observational. We applied guidelines extended from Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD), which is widely accepted for diagnostic/prognostic studies. The guidelines were extended by several TRIPOD authors specifically for multivariable prediction models instead of a single predictor to minimize risks of bias and optimism for prediction model development [25]. The extended guidelines were called as Prediction Model Risk of Bias Assessment Tool (PROBAST). We applied the PROBAST in conjunction with the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [26].

### 2.1. Data source

This study utilized NHID-BPJSKes which is a cross-sectional dataset representing real-world data of insurance-based healthcare in Indonesia on 2015 and 2016, but, we preprocessed this dataset in order to build a tidy dataset for a nested case-control design. Until 2018, the health insurance covered 200,259,147 (75.8%) individuals in Indonesia [27]. This reflected coverage of this insurance on the pregnant women in this study.

We utilized the initial version of the original dataset that had no accession number but published with the dataset code book for each version [27]. This book described details on cross-sectional sampling procedures of the original dataset. Briefly, all individuals covered by this insurance were sampled randomly stratified by 66,072 combinations of primary care (22,024 facilities) and family category (3 classes). The family category consisted of family of which members: (1) never using the health insurance; (2) using it in primary care only, and (3) using it in both primary care and hospital. The dataset only included the strata combination that consisted at least one family. It also included maximum 10 families; thus, if the combination consisted >10 families, then these were undersampled into 10 families. In the end, the sampling procedures resulted 586,969 families and 1,697,452 individuals.

Before we reconstructed the NHID-BPJSKes dataset for this study, it had been sampled from overall data in the insurance database by the owner which was the social security administrator for health or *badan penyelenggara jaminan sosial (BPJS) kesehatan* in Indonesia. None of the authors were parts of the BPJS Kesehatan and the sample dataset has been also deidentified before it was made publicly accessed by request; thus, there was neither issue of patient privacy or need of informed consent for this study. Permission to the dataset for this study has been granted by the owner (dataset request approval no.: 12047/I.2/0919). The datasets for model development and validation in this study are available for public access by request to the corresponding author and by approval of the BPJS kesehatan in Indonesia.

There were approximately 2.6 million instances from 34 provinces of Indonesia. These consisted of four tables of claims data, which describe membership, primary care visits by capitation, primary care visits by case-based group (CBG) payments, and hospital visits by CBG payments. Diagnoses in this dataset were coded according to the International Classification of Diseases 10th Revision (ICD10).

### 2.2. Study design

We applied a nested case-control design for this study. Inclusion criteria were pregnant women with and those without preeclampsia/eclampsia. All women with pregnancy records were included without considering whether the subjects were first-time mothers or not. Exclusion criteria were all subtypes of PIH except preeclampsia/eclampsia. Cases were defined as women with preeclampsia/eclampsia and one pregnancy but without other PIH diagnoses, while controls were defined as women with one pregnancy but without PIH diagnoses including preeclampsia/eclampsia.

To approach these definitions, we applied data preprocessing by ICD10 codes. Case groups consisted of all visits from subjects that had codes of both O14–15 and Z33–37, while controls consisted of those from subjects having codes of Z33–37 only (Table 1). Neither cases

**Table 1**
Diagnosis codes for nested case-control sampling.

| ICD10 codes | Description |
|---|---|
| O | Pregnancy, childbirth, and puerperium |
| O10–16 | Oedema, proteinuria, and hypertensive disorders in pregnancy, childbirth, and puerperium |
| O14–15 | Preeclampsia and eclampsia |
| O80–82 | Encounter for delivery |
| Z33–37 | Pregnant state, encounter for supervision of normal pregnancy, encounter for antenatal screening of mother, and outcome of delivery |

nor controls included visits from subjects that had other codes within O10−16. The pregnancy period was defined between the earliest and latest dates of visits coded by Z33−37 or O. This period was applied for feature extraction. We also removed all records that possibly had more than one episode of pregnancy within a 2-year period in the dataset. This was achieved by removing subjects with differences of known earliest and latest delivery codes (O80−82) that were greater than zero. In the end, the age range of control group was matched with the case group (12−55 years old).

### 2.3. Feature extraction, representation, and selection

Features consisted of demographic variables and diagnoses on previous visits. We conducted a time-to-event analysis to extract diagnosis predictors. The event was delivery of which the time was considered as comparable time of outcome between case and control. Demographic variables were age (years), marital status (married/single/divorced or widowed/undefined), family role (wife/child/primary member/additional member), membership strata (first/second/third), and membership type (government-paid labor/company-paid labor/self-paid labor/non-labor). Diagnoses were derived from encounters coded by A to N (Table S1 in Supplementary materials). To capture specific codes to several causes of disease and organ-related diseases, 15 features were also added. All of the diagnostic features were accounted for in each period of either 1 year before the event or during gestation. In addition, the time to the event (months) and distinct diagnoses compared to all available three-digit codes were included. We also included diagnoses within 2 years before the event along with an additional feature of censored time to the event (months) for those with event times in 2015. All continuous variables were normalized.

In total, there were 95 candidate features included in the predictive model. To avoid irrelevant and redundant features, we applied several feature selection techniques in a multivariate logistic regression model (MvLRM). These could be forward, backward, or stepwise feature selection. We applied 0.05 as the significance level for retaining the candidate feature in the model. Each of these methods might or might not be preceded by feature representation as either polynomial terms or principal components. Forward, backward, and stepwise selection were not used at the same time. We compared the MvLRM performances using any of these combinations of feature selection and/or representation with those with neither feature selection or representation. Instead of using one feature selection technique, comparison of the multiple combinations allowed us to have larger search space for model optimization. However, we limited this search space within feasible size by applying only 2-degree polynomial terms. The goal for the feature selection was to reach the predefined number of candidate features that fit a sufficient EPV for machine learning development [19,28]. We would define this number after finishing the subject selection (see Section 2.2). We applied the number as starting number of features included in backward selection that would stop at two remaining features. Contrarily, we started the forward and stepwise selection from two features and stopped at the predefined number. We chose at least two features because we applied conditional MvLRM to force time to the event being retained in the feature selection. This feature should be in the model anyway because we would conduct subgroup analysis using this feature in the best model (see Section 2.7). In addition, we also limit maximum principal components as much as the predefined number of candidate features. By this predefined number, we estimated the sample size having sufficient power for developing a predictive model, including the machine learning ones.

### 2.4. Model development

We compared six state-of-the-art machine learning algorithms using SAS Enterprise Miner 14.3 (SAS Institute, Cary, NC, US) to develop prognostic prediction model. These included the machine learning-optimized logistic regression (LR), decision tree (DT), artificial neural network (ANN), random forest (RF), support vector machine (SVM), and the ensemble (Ens.) algorithm that combined other algorithms. We conducted parameter tuning of the algorithms by comparing 726 configurations (Table S2 in Supplementary materials). Each algorithm included the best feature set from a previous selection. The best parameter tuning of each of the models was used for the final comparison. We also applied a critical appraisal to the best model based on the domain knowledge.

### 2.5. Model evaluation

We evaluated all models using both calibration and discrimination tests. Calibration was assessed by a linear regression of the predicted and true probabilities, while discrimination was assessed by the area under curve of the receiver operating characteristics curve (AUROC). In the end, we also compared the positive predictive value or information retrieval precision (Prec.) for the false positive rate (FPR or 1-specificity) of 10%.

### 2.6. Model validation

We split up the tidy dataset into training and test sets for internal and external validation, respectively. Data partitioning for external validation was further split geographically and temporally. In the geographical split for external validation (GEV), one city in each province was randomly sampled. The city list was used to filter the dataset into the test set, while the rest was split for the training and another validation set. Geographical randomization from each province of Indonesia is important to avoid racial/ethnic disparities associated between preeclampsia and its risk factors [29]. In the temporal split for external validation (TEV), 25% of the days in each month were randomly selected. All visits from subjects with delivery time on those selected days were split for the validation set; thus, the subjects were completely external to those in the training set and internal validation set. Temporal randomization was intended to avoid a seasonality effect on preeclampsia [30]. Women who delivered during winter (non-tropical regions) or rainy season (tropical regions) have higher prevalence of preeclampsia/eclampsia. Both geographical and temporal randomization are different to simple randomization which may leave subjects in the training set, that lived in the same cities or were delivered at the same time periods with those in the test sets. Using geographical and temporal randomization, the test sets would have subjects with unobserved features in the training set, that were related to the city and time period. Meanwhile, since preeclampsia was associated with racial/ethnicity and seasonality [29,30], no randomization for city and time period selection might cause biases in the predictive performance. We also conducted 10-time bootstrapping to iterate the external validation. Therefore, we could estimate the predictive performance of our model for future dataset.

Before feature selection, a tidy dataset with balanced cases and controls was constructed for internal and external validations and analyzed for missing data. We conducted oversampling of cases by stratified random sampling. All candidate features and outcomes were used as stratification variables. Chi-square and *t*-tests were conducted on the dataset before and after oversampling to ensure that there was no significant effect of oversampling of the case group. Statistical tests were also conducted on the dataset before and after removing missing data.

We conducted 10-fold cross-validation when developing the models. The training set and internal validation (IV) set were randomly assigned into 10 groups by stratified random sampling. The stratification variable for this randomization was the target outcome. The models were trained/fitted by aggregating nine groups and were internally validated by the remaining group each time. This process was repeated 10 times until all groups were used as the validation set.

We applied 10-fold cross-validation starting from feature selection. However, to efficiently search for the best parameter tuning for each algorithm, we applied test split validation with a 9:1 ratio for both the training and validation sets. In the final comparison among all algorithms, we also applied 10-fold cross validations. External validations were applied to feature selection, parameter tuning, and final comparison, but these datasets had no role in parameter updating in each model.

### 2.7. Subgroup and text mining analyses

A subgroup analysis was conducted using the time-to-event. This determines the period before the event that has the highest discrimination ability according to the AUROC. To do so, we split all external validation sets after prediction. Datasets were split by the time-to-event into four groups which were 2 days to <6 months, 6∼<9 months, 9∼<12 months, and 12−24 months. The groupings were intended to imply known prediction periods and pathogenesis paradigms from previous studies within the range of available data. These were second- or third-trimester predictions [4], first-trimester prediction/pathogenesis [14], near-pregnancy pathogenesis (related to endometrial maturation) [31], and genetic paradigms of pathogenesis (related to vascular susceptibility) [32].

In addition, we also conducted a text mining analysis using SAS Enterprise Miner 14.3 (SAS Institute). It was based on natural language processing techniques on the internal validation set to interpret the best model. We extracted all ICD10 codes in any visits that had non-zero values in each diagnosis predictor and that had codes classified to the predictor. The visits were limited to those that had true predictions by the best model. Text profiling of ICD10 codes was conducted for each diagnosis predictor in the case group.

### 2.8. Comparison to previous studies

We also compared the best period for prediction of our model with those from previous studies. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for the comparison purpose was applied. We searched both research and review articles (systematic review/meta-analysis) in PUBMED, EMBASE, and SCOPUS within the last 5 years (since 2015) that developed and/or validated clinical prediction models. The models had to match our eligibility criteria as defined by PICOTS: [28] (1) population: women or pregnant women without specializing the population; (2) index: multivariable, prognostic clinical prediction model using demographic and/or clinical predictors in a poor-resource setting; (3) comparator: the best model in this study; (4) outcome: preeclampsia without differentiating early- or late-onset and with or without fetal growth restriction; (5) timing: before or during pregnancy until 2 days before onset or delivery; and (6) setting: survey, primary care, or hospital. The studies had to report the point and interval estimates of predictive performance, sample sizes in either case or control, and model validation methods. These were parts of quality assessment we followed from Prediction Model Risk of Bias Assessment Tool (PROBAST) [25], and the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [26]. All authors independently assessed the criteria in order as described. If there was a non-matching criterion, then we did not assess the next criterion. If there was a disagreement among authors, this was resolved through discussion. The data of eligible articles were extracted by HS, and the extracted data were reviewed by YWW and ECYS.

### 2.9. Decision curve analysis

To determine cut off values, we applied decision curve analysis that showed FPR, negative predictive value (NPV), positive predictive value (PPV), proportion of predicted positives, and true positive rate (TPR) or sensitivity. We identified cut off value with either sensitivity ∼0.95 and specificity ∼0.90 to compare the predictive performances with those of prediction models from previous studies. The cut off values were chosen based on the internal validation set to apply on the other sets. The final model used only cut off value at sensitivity ∼0.95 to achieve a sufficient preliminary prediction model that will be combined with other highly-specific models. Inverse of the proportion of predicted positives might imply potential reduction of further prediction by the models with advanced predictors.

### 2.10. Statistical analysis

We used SAS Enterprise Guide 7.1 (SAS Institute) to conduct all statistical analyses. Evaluation metrics were expressed as point and interval estimates with the 95% confidence interval (CI). The results from 10-fold cross-validation and 10-time bootstrapped external validation were used to calculate the interval estimate. We used the interval estimates to compare evaluation metrics of the models. The best model was determined by the AUROC and PPV or IR's precision from both external validations. In addition, significance of the selected candidate features was expressed as adjusted *p*-value. To describe the continuous features, we applied mean and standard deviation as the center and dispersion metrics, respectively, while frequency and the proportion was applied for the categorical features.

## 3. Results

### 3.1. Characteristics of the dataset

Datasets were constructed for internal and external validations (*n* = 23,201; Fig. 1). From these datasets, the proportion of visits by women who would be delivered in primary care and hospital (ratio of IV, GEV, and TEV) were 43.82% (9035:717:414) and 56.17% (11,940:605:490), respectively. Missing data were minor in both cases (0.33%, *n* = 11/3329) and controls (1.44%, *n* = 290/20,173). The numbers of missing data were the number of either cases or controls that had missing value on any of the features only. The outcome, either case or control, was complete in this study. Differences in predictor and outcome candidates were described before and after removing missing data or balancing data (Tables S3 and S4 in Supplementary materials).

Censored diagnoses on previous visits in the training set were also minor in both cases (*n* = 878, 28.75%) or and controls (*n* = 5856, 32.67%). This is important because censored diagnoses can be viewed as negatives, while these may actually be positives but not recorded due to data availability. In real world data, this situation can occur with a new member to the health insurance program. We added the censored time-to-event to tell the algorithms how many months a subject had censored diagnoses on previous visits. Nonetheless, this candidate feature was not chosen in the selection process.

### 3.2. Selected feature candidates

To achieve sufficient EPV for model development in all machine learning algorithms, we selected up to 17 features as the predefined number of candidates. Several feature candidates were selected from the MvLRM using forward selection of both original features and principal components (Table 2). We also forced the principal component analysis to obtain 17 components. In addition to the original set of 95 feature candidates, the principal components made a feature

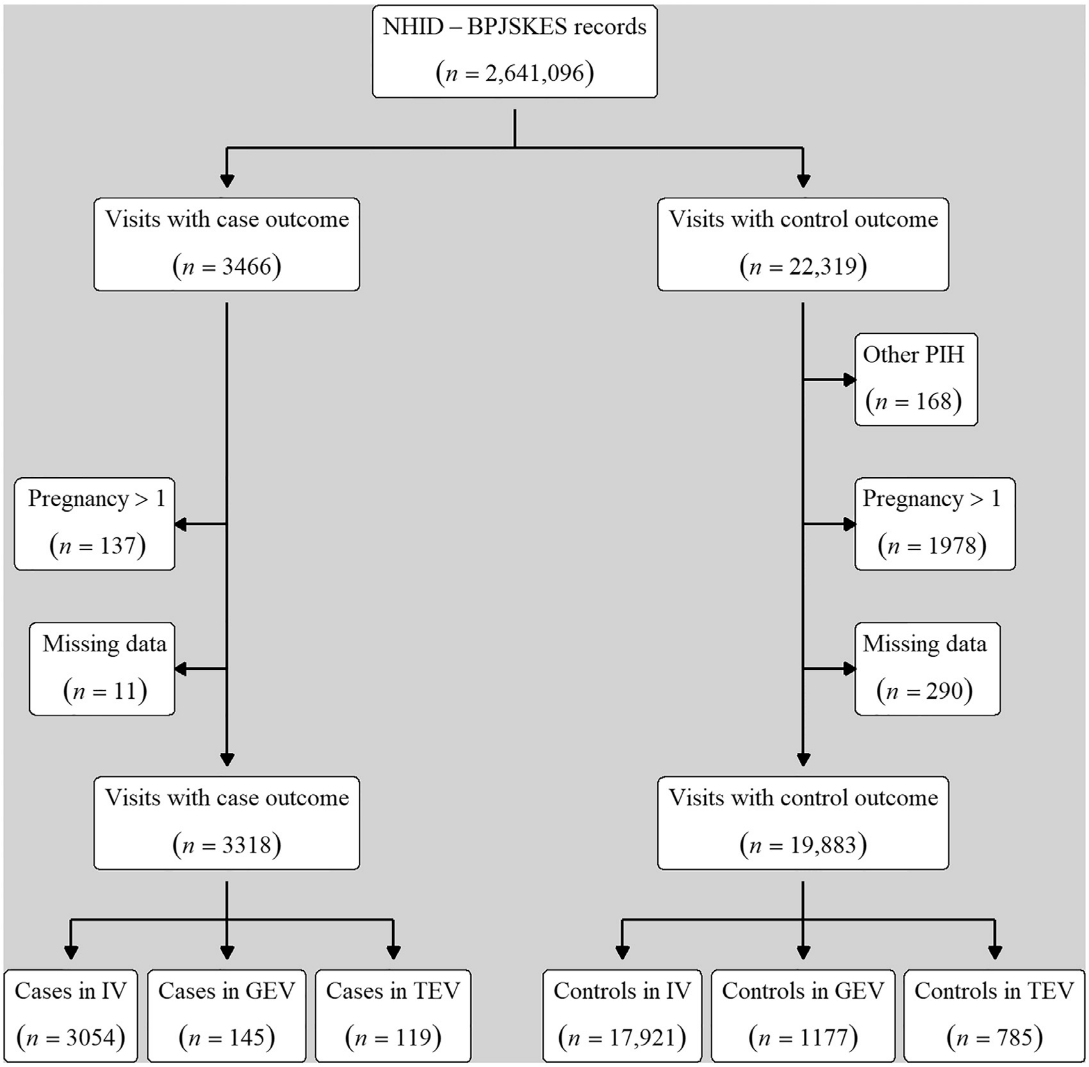

**Fig. 1.** Dataset constructed for model development. The original dataset was constructed with a nested case-control design. Controls were sampled within the same age range of case groups (12–55 years old). NHID-BPJSKes, nationwide health insurance dataset of BPJS Kesehatan; PIH, pregnancy-induced hypertension; IV, internal validation; GEV, geographical split for external validation; TEV, temporal split for external validation.

set with 112 feature candidates. With this number of candidates, the EPV for the training set was 27.3. This number was sufficient for preliminary feature selection by the MvLRM in order to avoid optimism according to the standard, which was 20~50 EPV for logistic regressions.

### 3.3. Model comparison

Six machine learning models were compared (Table 3). The best model used the random forest algorithm consisting of 500 trees. This model was consistently superior in terms of both internal and external validations by geographical and temporal splits, compared to the other machine learning models, including the ensemble model, which did not outperform the random forest model. Three algorithms were combined in the ensemble model: the logistic regression, decision tree, and artificial neural network. This configuration had the best predictive performance among other configurations of ensemble models.

Calibration slope of the model with random forest algorithm was significantly different to 1 as demonstrated by the estimates (Table 3). This was also shown by the models with other algorithms. The receiver operating characteristics (ROC) curves and the area under curve (AUROCs) are also shown for the random forest model (Fig. 2). The ROC curve and AUROC for the training set were similar to internal validation set, but not external validation sets. For a specificity of ~90%, the detection rates were 0.58 (95% CI 0.57–0.59), 0.44 (95% CI 0.43–0.46), and 0.37 (95% CI 0.36–0.38) for IV, GEV, and TEV, respectively. For the same specificity, the precisions were 0.86 (95% CI 0.85–0.86), 0.82 (95% CI 0.81–0.83), and 0.78 (95% CI 0.78–0.79) for IV, GEV, and TEV, respectively.

### 3.4. Subgroup analysis by the time-to-event of the best model

Instances in each external validation set were subgrouped by period of the time-to-event. The AUROCs were re-computed in each subgroup (Fig. 3). The period of 9–<12 months to the event was the period which showed the highest AUROC both for GEV (0.89, 95% CI 0.88–0.89) and TEV (0.86, 95% CI 0.85–0.86). The wide discrepancy of AUROCs between internal and external validation were contributed by instances subgrouped by the period of 12–24 months, 6–<9 months, and 2 days–<6 months, but not 9–<12 months. These indicated that the features from this period might be more important than those from other periods.

### 3.5. Text mining analysis of the best model

A text mining analysis was conducted for all instances that were true-predicted by the random forest model in the internal validation

**Table 2**
Feature candidates selected by the multivariate logistic regression model with forward selection from original candidates and principal components.

| # | Feature | Cases (*n* = 3054) | Controls (*n* = 17,921) | *p* value |
|---|---|---|---|---|
| 1 | Time-to-event (months) ± SD [*] | 4.56 ± 5.19 | 4.16 ± 4.43 | 0.08 |
| | **Demographic variables** | | | |
| 2 | Age (years) ± SD | 32 ± 12 | 30 ± 12 | <0.0001 |
| 3 | Family role, *n* (%) | | | |
| | . Wife[†] | 1895 (62.05) | 10,953 (61.12) | – |
| | . Primary member | 849 (27.80) | 4381 (24.45) | 0.06 |
| | . Child | 214 (7.01) | 2161 (12.05) | 0.01 |
| | . Additional member | 96 (3.14) | 426 (2.38) | <0.0001 |
| 4 | Member stratum, *n* (%) | | | |
| | . First | 459 (15.03) | 2494 (13.92) | <0.0001 |
| | . Second | 1306 (42.76) | 8114 (45.28) | 0.45 |
| | . Third [†] | 1289 (42.21) | 7313 (40.80) | – |
| 5 | Member type, *n* (%) | | | |
| | . Company-paid labor | 1517 (49.67) | 8720 (48.66) | <0.0001 |
| | . Government-paid labor | 769 (25.18) | 4997 (27.88) | <0.0001 |
| | . Self-paid labor [†] | 747 (24.46) | 4173 (23.29) | – |
| | . Non-labor | 21 (0.69) | 31 (0.17) | <0.0001 |
| | **Diagnoses within the last 2 years to the event (partially censored)** | | | |
| 6 | A codes - Certain infectious and parasitic diseases, visits ± SD; *n* (%) [‡] | 2.72 ± 1.79; 248 (8.12) | 1.54 ± 1.06; 1412 (7.88) | 0.02 |
| 7 | E codes - Endocrine, nutritional and metabolic diseases, visits ± SD; *n* (%) [‡] | 5.00 ± 5.41; 187 (6.12) | 2.65 ± 2.33; 310 (1.73) | <0.0001 |
| 8 | I codes - Diseases of the circulatory system, visits ± SD; *n* (%) [‡] | 4.05 ± 3.75; 570 (18.66) | 2.63 ± 2.33; 609 (3.40) | <0.0001 |
| 9 | Immune-related codes, visits ± SD; *n* (%) [‡] | 2.97 ± 2.15; 308 (10.09) | 1.77 ± 1.39; 1142 (6.37) | <0.0001 |
| 10 | Eye-related codes, visits ± SD; *n* (%) [‡] | 2.81 ± 1.62; 57 (1.87) | 1.78 ± 1.10; 444 (2.48) | <0.0001 |
| | **Diagnoses within the last year to the event** | | | |
| 11 | N codes - Diseases of the genitourinary system, visits ± SD; *n* (%) [‡] | 3.94 ± 3.37; 172 (5.63) | 1.95 ± 1.99; 856 (4.78) | <0.0001 |
| 12 | Eye-related codes, visits ± SD; *n* (%) [‡] | 2.28 ± 1.37; 248 (8.12) | 1.71 ± 0.99; 1412 (7.88) | <0.0001 |
| | **Diagnoses within the pregnancy period to the event** | | | |
| 13 | Breast-related codes, visits ± SD; *n* (%) [‡] | 5.85 ± 2.58; 13 (0.43) | 1.00 ± 0.00; 6 (0.03) | <0.0001 |
| 14 | Digestive system-related codes, visits ± SD; *n* (%) [‡] | 2.48 ± 2.35; 186 (6.09) | 1.85 ± 1.60; 768 (4.29) | <0.0001 |
| 15 | Skin and subcutaneous-related codes, visits ± SD; *n* (%) [‡] | 1.81 ± 0.71; 36 (1.18) | 1.52 ± 1.14; 287 (1.60) | <0.0001 |
| | **Principal components** | | | |
| 16 | Principal components 8 (see Table 4 for the profile) | 2.72 ± 1.79; 248 (8.12) | 1.54 ± 1.06; 1412 (7.88) | <0.0001 |
| 17 | Principal components 10 (see Table 4 for the profile) | −0.09 ± 0.03 | 0.09 ± 0.01 | <0.0001 |

[*] Forced into the multivariate logistic regression model.
[†] Comparator.
[‡] Non-zero visits.

set. Text profiles are shown for all diagnosis predictors (Table 4). Several codes in the text profiles were classified to one or more diagnosis predictors in the case group. Therefore, we could identify specific diagnoses in the true-predicted instances to interpret the best model in this study.

### 3.6. *Comparison to previous studies*

We found 879 records from PUBMED, EMBASE, and SCOPUS for 'preeclampsia prediction model' within the last 5 years, and seven studies were eligible for comparison to our random forest model in

**Table 3**
Calibration and discrimination tests of six machine learning models by both internal and external validations.

| Validation | Algorithm | Calibration | | Discrimination tests | |
|---|---|---|---|---|---|
| | | Slope (95% CI) | Intercept (95% CI) | AUROC (95% CI) | Prec. (95% CI) [*] |
| Internal | LR | 1.08 (1.08, 1.09) | −0.04 (−0.04, −0.03) | 0.70 (0.69, 0.70) | 0.78 (0.78, 0.78) |
| | DT | 0.99 (0.99, 1.00) | 0.01 (0.01, 0.01) | 0.66 (0.66, 0.67) | 0.73 (0.72, 0.74) |
| | ANN | 0.64 (0.63, 0.64) | 0.14 (0.14, 0.15) | 0.65 (0.64, 0.67) | 0.74 (0.73, 0.75) |
| | RF | 1.54 (1.54, 1.54) | −0.27 (−0.27, −0.26) | 0.86 (0.85, 0.86) | 0.86 (0.85, 0.86) |
| | SVM | 2.68 (2.66, 2.70) | −0.89 (−0.90, −0.88) | 0.68 (0.67, 0.68) | 0.78 (0.76, 0.79) |
| | Ens. | 1.21 (1.21, 1.22) | −0.13 (−0.13, −0.12) | 0.70 (0.70, 0.71) | 0.78 (0.77, 0.78) |
| External, geographical split | LR | 1.80 (1.76, 1.83) | −0.34 (−0.35, −0.32) | 0.74 (0.73, 0.76) | 0.68 (0.67, 0.70) |
| | DT | 0.69 (0.67, 0.71) | 0.15 (0.14, 0.16) | 0.60 (0.59, 0.61) | 0.80 (0.79, 0.81) |
| | ANN | 0.75 (0.73, 0.77) | 0.08 (0.07, 0.09) | 0.67 (0.64, 0.70) | 0.55 (0.52, 0.58) |
| | RF | 1.47 (1.45, 1.50) | −0.19 (−0.21, −0.18) | 0.76 (0.76, 0.77) | 0.82 (0.81, 0.83) |
| | SVM | 3.12 (3.02, 3.21) | −1.07 (−1.12, −1.02) | 0.62 (0.61, 0.62) | 0.54 (0.52, 0.57) |
| | Ens. | 1.52 (1.49, 1.55) | −0.28 (−0.30, −0.26) | 0.72 (0.71, 0.73) | 0.70 (0.68, 0.72) |
| External, temporal split | LR | 0.74 (0.72, 0.76) | 0.16 (0.15, 0.17) | 0.62 (0.62, 0.63) | 0.77 (0.76, 0.77) |
| | DT | 0.92 (0.90, 0.93) | 0.08 (0.08, 0.09) | 0.63 (0.62, 0.63) | 0.69 (0.68, 0.70) |
| | ANN | 0.30 (0.29, 0.31) | 0.34 (0.33, 0.35) | 0.58 (0.58, 0.59) | 0.71 (0.70, 0.72) |
| | RF | 1.09 (1.08, 1.11) | 0.02 (0.02, 0.03) | 0.70 (0.70, 0.70) | 0.78 (0.78, 0.79) |
| | SVM | 2.25 (2.20, 2.30) | −0.65 (−0.67, −0.62) | 0.63 (0.63, 0.63) | 0.72 (0.71, 0.73) |
| | Ens. | 0.74 (0.72, 0.76) | 0.15 (0.14, 0.16) | 0.61 (0.61, 0.62) | 0.74 (0.73, 0.74) |

AUROC, area under the receiver operating characteristic curve; LR, machine learning-optimized logistic regression; DT, decision tree; ANN, artificial neural network; RF, random forest; SVM, support vector machine; Ens., ensemble algorithm.
[*] For a specificity of ~90%.

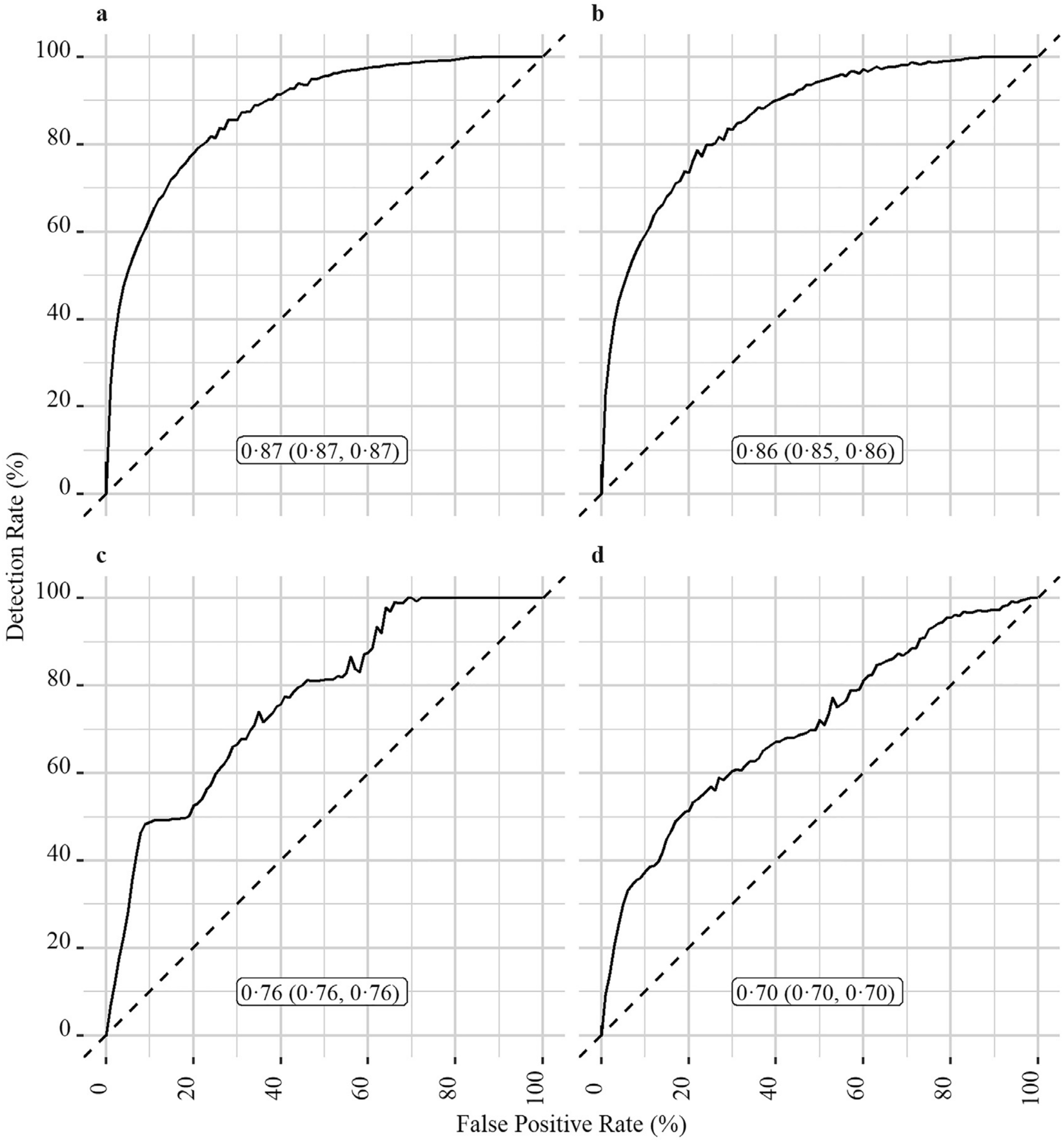

**Fig. 2.** Receiver operating characteristics (ROC) curves for the random forest model. Four panels show the ROC curves with AUROCs and 95% CIs using these datasets: (a) training set; (b) internal validation set; (c) external validation set by geographical split; and (d) external validation set by temporal split. The dashed line is a reference line. AUROC, area under the receiver operating characteristics curve.

the subgroup of 9–12 months to the event (Supplementary materials). Compared to most previous models, our model in this subgroup had the best predictive performances in the AUROC competing with those from MacDonald-Wallis (2015), including the predictive performance using GEV and TEV (Table 5) [33]. The precision and sensitivity of our model were also the highest ones among those with a specificity of ~90% [15,34,35]. For a sensitivity of ~95%, our model had higher precision and competing specificity compared to that of MacDonald-Wallis (2015) [33]. For comparison purpose, we applied 0.34 and 0.54 as cut off values for the model at sensitivity (~0.95) and specificity (~0.90), respectively. The cut off values were determined based on internal validation (Fig. S2–S7 in Supplementary materials). However, we recommend cut off value of 0.34 to get highly-sensitive performance using our model as the preliminary prediction model to decide which patient will be predicted by other models with high specificity. We also recommend to apply prediction model from MacDonald-Wallis (2015) [33], which had NPV 1.00 (95% CI 0.99–1.00), to confirm predicted negatives by our model. Using cut off value of 0.34, the proportions of predicted positives were 77% (95% CI 75–78%) in GEV and 78% (95% CI 77–78%) in TEV. This imply potential reduction of ~20% cost needed for prediction models with advanced predictors.

## 4. Discussion

Our model included predictors mostly from the medical history. It was the most frequent predictor used in preeclampsia prediction models [38]. Some of our predictors were also used in clinical prediction models from previous studies. Age, chronic hypertension (I10, I159), and diabetes mellitus (E118, E119) were used in the NICE and ACOG guidelines with or without modification [37]. In addition, a previous meta-analysis also showed associations between some of these predictors and preeclampsia [39]. These were a maternal age of >40 years (OR 1.50, 95% CI 1.20–2.00; $I^2$=95%; *df*=15) and chronic hypertension (ICD10 I10, I159; OR 5.10, 95% CI 4.00–6.50; $I^2$=98%; *df*=20). The random forest and text profiling algorithms included these diseases that were available in our training set.

Nevertheless, there was no systemic lupus erythematosus (SLE), antiphospholipid syndrome or other thrombophilia, or chronic kidney disease in our training set, while previous models included those

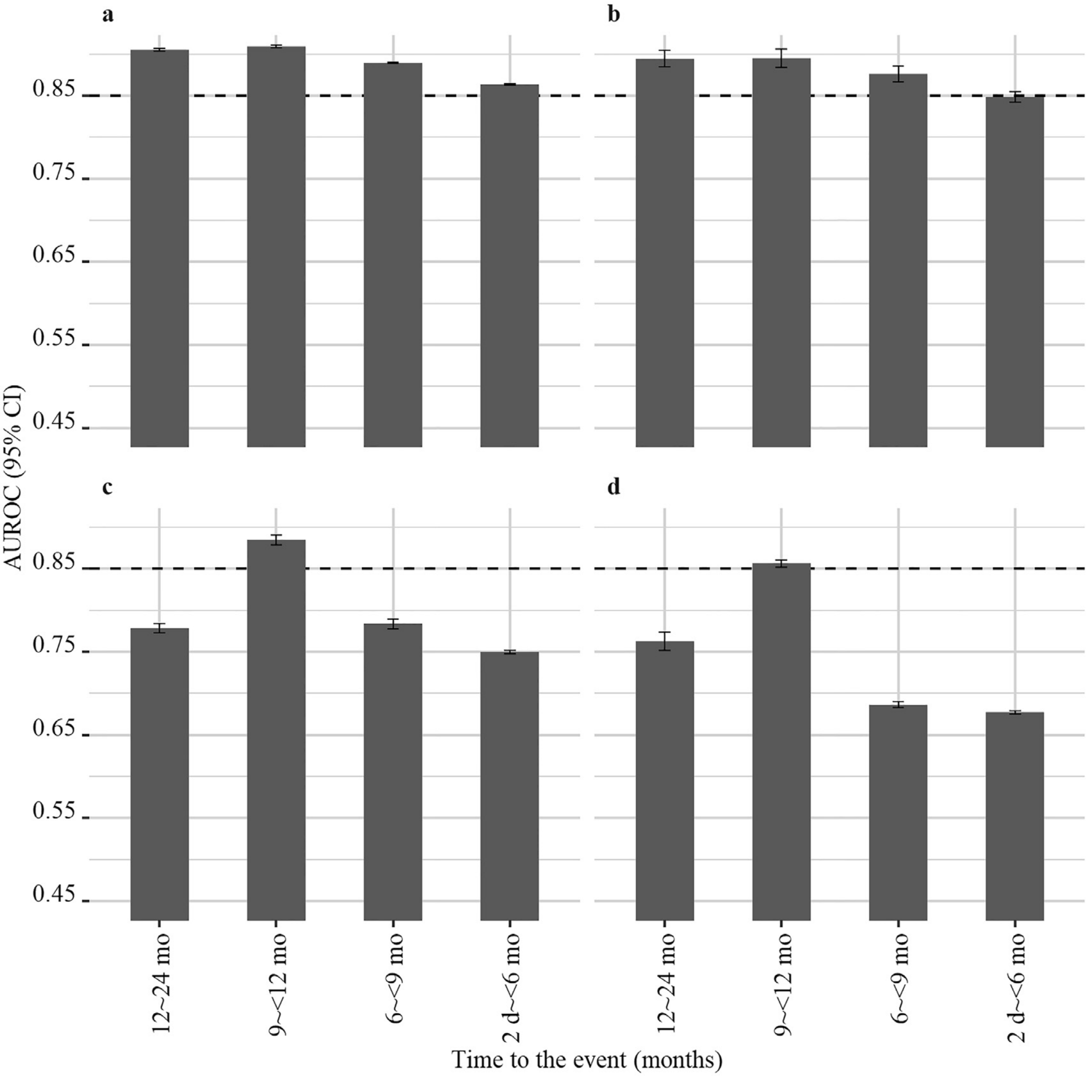

**Fig. 3.** Area under receiver operating characteristics curve (AUROC) of subgroups by the time-to-event from the random forest model. Four panels show the AUROCs using these datasets: (a) training set; (b) internal validation set; (c) external validation set by geographical split; and (d) external validation set by temporal split. The error bar and 95% confidence interval are shown. To improve readability, the y-axis scale was begun from 0•45; all of the data are completely shown. The dashed line shows the minimum AUROC among those using training and IV sets. AUROC, area under the receiver operating characteristics curve.

diseases along with age, chronic hypertension, and diabetes mellitus [15,33–36]. Although there was no SLE, erosive arthritis was available in our training set and was found in the text profiling results for immune-related codes. Erosive arthritis can be determined by the anti-citrullinated peptide and anti-carbamylated protein antibodies that were predictive among four other predictors in a predictive model for SLE (AUROC 0.81, 95% CI 0.80–0.81).

In the same diagnosis predictor including chronic hypertension, congestive heart failure (I500) was also shown in the text profile of I codes (diseases of the circulatory system). This disease shared the same prognostic factor with preeclampsia, which is serum homocysteine. This factor was predictive for early-onset preeclampsia (AUROC 0.87; OR 1.54, 95% CI 1.30–1.84) [40]. Meanwhile, the serum homocysteine level was elevated in patients with congestive heart failure compared to controls ($p<0.01$) [41]. Preeclampsia was also associated with future congestive heart failure (RR 3.62, 95% CI 2.25–5.85; $I^2$=83%; *df*=6) [42]; however, this may be related to common genetic backgrounds between the diseases [32].

Beyond diabetes mellitus, other endocrine, nutritional, and metabolic diseases were also included as diagnosis predictors in our model. Thyrotoxicosis or hyperthyroidism (E059) was associated with preeclampsia in a nationwide population-based study (OR 1.21, 95% CI 1.14–1.29) [43]. A meta-analysis also showed significant mean differences of total cholesterol (ICD10 E780; 20.20 mg/dL, 95% CI 8.70–31.70; $I^2$=99%; *df*=45) and triglycerides (ICD10 E785; 80.29 mg/dL, 95% CI 51.45–109.13; $I^2$=99%; *df*=43) in women with preeclampsia compared to those of controls, especially in the third trimester.

Interestingly, there was an atopic pattern in the text profile of immune-related and skin-related codes, which consisted of allergic rhinitis (J304), asthma (J459), and atopic dermatitis (L208, L209). Women with atopic dermatitis had higher risks of severe preeclampsia (OR 1.27, 95% CI 1.01–1.58), and eclampsia (OR 2.08, 95% CI 1.09–3.98) [44]. Women with preeclampsia had higher incidences of having a child with allergic rhinitis (incidence rate ratio [IRR] 1.29, 95% CI 1.11–1.50), asthma (IRR 1.17, 95% CI 1.11–1.25), and atopic dermatitis (IRR 1.15, 95% CI 1.01–1.32) at ≥14 days old. Women with asthma had a higher risk of having a child with asthma if either the mother developed preeclampsia (adjusted hazard ratio [aHR] 4.73, 95% CI 2.20–10.70) or not (aHR 2.18, 95% CI 1.46–3.26) compared to neither asthma nor preeclampsia [45]. This genetic tendency could appear in both mother and offspring, so it is not easy to identify the cause or effect between preeclampsia and maternal atopy. In addition, asthma might be correlated with *Brucella abortus* infection

**Table 4**
Text profile for ICD10 codes of diagnosis predictors in the true-predicted case group by the random forest model.

| Time-to-event | Diagnosis predictor | ICD10 codes and description |
|---|---|---|
| Diagnoses within the last 2 years to the event (partially censored) | A codes - Certain infectious and parasitic diseases | A010 (Typhoid fever)<br>A09 (Infectious gastroenteritis and colitis, unspecified)<br>A182 (Tuberculous peripheral lymphadenopathy)<br>A231 (Brucellosis due to *Brucella abortus*)<br>A78 (Q fever)<br>A91 (Dengue haemorrhagic fever) |
| | E codes - Endocrine, nutritional, and metabolic diseases | E059 (Thyrotoxicosis, unspecified)<br>E118 (Type 2 diabetes mellitus with unspecified complications)<br>E119 (Type 2 diabetes mellitus without complications)<br>E780 (Pure hypercholesterolemia)<br>E785 (Hyperlipidaemia, unspecified)<br>E86 (Volume depletion) |
| | I codes - Diseases of the circulatory system | I10 (Essential [primary] hypertension)<br>I159 (Secondary hypertension, unspecified)<br>I500 (Congestive heart failure) |
| | Immune-related codes | J304 (Allergic rhinitis, unspecified)<br>J329 (Chronic sinusitis, unspecified)<br>J459 (Asthma, unspecified)<br>L208 (Other atopic dermatitis)<br>L209 (Atopic dermatitis, unspecified)<br>M154 (Erosive [osteo]arthrosis) |
| | Eye-related codes | H000 (Hordeolum and other deep inflammation of eyelid)<br>H055 (Retained [old] foreign body following penetrating wound of orbit)<br>H109 (Conjunctivitis, unspecified)<br>H521 (Myopia)<br>H527 (Disorder of refraction, unspecified) |
| Diagnoses within the last year to the event | Diseases of the genitourinary system | N300 (Acute cystitis)<br>N309 (Cystitis, unspecified)<br>N601 (Diffuse cystic mastopathy)<br>N608 (Other benign mammary dysplasias)<br>N609 (Benign mammary dysplasia, unspecified)<br>N61 (Inflammatory disorders of breast) |
| | Eye-related codes | H000 (Hordeolum and other deep inflammation of eyelid)<br>H055 (Retained [old] foreign body following penetrating wound of orbit)<br>H109 (Conjunctivitis, unspecified)<br>H521 (Myopia)<br>H527 (Disorder of refraction, unspecified) |
| Diagnoses within the pregnancy period to the event | Breast-related codes | N61 (Inflammatory disorders of breast) |
| | Digestive system-related codes | A09 (Infectious gastroenteritis and colitis, unspecified)<br>K029 (Dental caries, unspecified)<br>K040 (Pulpitis)<br>K045 (Chronic apical periodontitis)<br>K047 (Periapical abscess without sinus)<br>K053 (Chronic periodontitis)<br>K30 (Excessive attrition of teeth) |
| | Skin and subcutaneous-related codes | L209 (Atopic dermatitis, unspecified) |
| Principal components | Principal components 8 | H000 (Hordeolum and other deep inflammation of eyelid)<br>H109 (Conjunctivitis, unspecified)<br>H521 (Myopia)<br>H527 (Disorder of refraction, unspecified)<br>H608 (Other otitis externa)<br>H609 (Otitis externa, unspecified)<br>H811 (Benign paroxysmal vertigo)<br>H814 (Vertigo of central origin) |
| | Principal components 10 | D509 (Iron deficiency anemia, unspecified)<br>D648 (Other specified anaemias)<br>D649 (Anemia, unspecified) |

(A231), since the numbers of *B. abortus* in the lungs were higher in asthma-induced murine models compared to the controls ($p<0.001$) [46].

There were other specific infections included in the A codes (certain infectious and parasitic diseases). The immune response specifically mediated by *Salmonella typhi* (A010) harboured more-diverse microbial communities in the gut of individuals with a multiphasic response compared to those with a late response [47]. This may be related to preeclampsia because there was also a significant shift in the gut microbial communities in women with this disease [48]. Tuberculous peripheral lymphadenopathy (A182) symptoms were frequent in human immunodeficiency virus (HIV) infection that shared dysregulation of the complement system with preeclampsia [49,50]. Infections by *Coxiella burnetii* or Q fever (A78) were associated with adverse maternal outcomes related to preeclampsia, such as intrauterine growth retardation and preterm delivery [51]. Dengue

**Table 5**
Predictive performances of the random forest model in the subgroup of 9−12 months to the event with cut off value at either similar sensitivity or specificity based on internal validation compared to those from previous studies.

| Algorithm | Validation | AUROC (95% CI) | Prec. (95% CI) | Sens. (95% CI) | Spec. (95% CI) |
|---|---|---|---|---|---|
| **Interval validation** | | | | | |
| *At sensitivity ~0.95* | | | | | |
| RF 9−<12 mo.; cut off value of 0.34 | 10-fold CV | 0.90 (0.88, 0.91) | 0.71 (0.68, 0.73) | 0.98 (0.97, 0.99) | 0.52 (0.49, 0.55) |
| MacDonald-Wallis et al. (2015)[33] | Bootstrapping | 0.88 (0.86, 0.90) | 0.04 (0.03, 0.04) | 0.95 [*] | 0.37 (0.31, 0.42) |
| *At specificity ~0.90* | | | | | |
| RF 9−<12 mo.; cut off value of 0.54 | 10-fold CV | 0.90 (0.88, 0.91) | 0.88 (0.87, 0.90) | 0.70 (0.67, 0.73) | 0.89 (0.87, 0.91) |
| Guy et al. (2017)[34] | No IV | 0.80 (0.75, 0.85) | 0.09 (0.07, 0.12) | 0.41 (0.29, 0.54) | 0.90 [*] |
| Viguiliouk et al. (2017)[36] | No IV | 0.76 (0.72, 0.81) | NA | NA | NA |
| Wright et al. (2015)[15] | 5-fold CV | 0.76 [†] | 0.08 [†] | 0.40 (0.39, 0.42) | 0.89 [†] |
| Rocha et al. (2017)[35] | No IV | 0.75 (0.72, 0.79) | 0.18 [†] | 0.44 [†] | 0.90 [*,†] |
| **External validation** | | | | | |
| *At sensitivity ~0.95* | | | | | |
| RF 9−<12 mo.; cut off value of 0.34 | Bootstrapped GEV | 0.88 (0.88, 0.89) | 0.59 (0.58, 0.60) | 1.00 (1.00, 1.00) | 0.47 (0.45, 0.49) |
| MacDonald-Wallis et al. (2015)[33] | Bootstrapping | 0.88 (0.84, 0.93) | 0.05 (0.04, 0.06) | 0.95 [*] | 0.47 (0.40, 0.55) |
| RF 9−<12 mo.; cut off value of 0.34 | Bootstrapped TEV | 0.86 (0.85, 0.86) | 0.72 (0.72, 0.72) | 0.90 (0.90, 0.90) | 0.44 (0.43, 0.45) |
| ACOG (2017)[37] | Bootstrapping | 0.57 (0.54, 0.61) | 0.17 [†] | 0.87 [†] | 0.27 [†] |
| *At specificity ~0.90* | | | | | |
| RF 9−<12 mo.; cut off value of 0.54 | Bootstrapped GEV | 0.88 (0.88, 0.89) | 0.82 (0.80, 0.85) | 0.52 (0.52, 0.52) | 0.91 (0.90, 0.93) |
| RF 9−<12 mo.; cut off value of 0.54 | Bootstrapped TEV | 0.86 (0.85, 0.86) | 0.89 (0.89, 0.89) | 0.70 (0.70, 0.70) | 0.86 (0.86, 0.86) |
| NICE (2015)[15] | Bootstrapping | 0.76 [†] | 0.07 [†] | 0.39 (0.33, 0.37) | 0.89 [†] |
| NICE (2017)[37] | Bootstrapping | 0.61 (0.58, 0.65) | 0.09 [†] | 0.38 [†] | 0.85 [†] |

AUROC, area under the receiver operating characteristic curve; Prec., precision; Sens., sensitivity; Spec., specificity; RF, random forest; NA, not available; NICE, National Institute for Health and Care Excellence; ACOG, American College of Obstetrics and Gynaecology.

[*] Fixed specificity.

[†] Interval estimate was not reported.

haemorrhagic fever (A91) may make pregnant women more susceptible to endothelial dysfunction and volume depletion (E86) in preeclampsia [52,53]. In addition, both intranasal bacteria and dysbiosis of microbiomes play putative roles in chronic sinusitis (J329) in immune-related codes [54]. Meanwhile, the means of hematogenous spread, including from the respiratory tract, were shown to be involved in great obstetrical syndromes like preeclampsia [55]. However, associations of these diseases with preeclampsia are still poorly understood.

Conversely, preeclampsia and infectious diseases of the urinary tract as well as periodontal diseases are well studied. Maternal infections associated with preeclampsia were included as a diagnosis predictor by the N codes (diseases of the genitourinary system within the last year to the event). Urinary tract infections, including cystitis (N300, N309), were associated with preeclampsia (OR 1.57, 95% CI 1.45−1.70; $I^2$=79%; *df*=16) [56]. Preeclampsia was also associated with periodontal diseases (ICD10 K040, K045, K047, K053; OR 1.76, 95% CI 1.43−2.18; $I^2$=80%; *df*=5) [56]. Several microbes were identified on placental tissue samples from women with preeclampsia (*n* = 7; who underwent an elective caesarian delivery) by a polymerase chain reaction (PCR), 16S ribosomal (r)RNA gene, and next-generation sequencing, while all samples from the controls were negative (*n* = 48; *p* = 0.006). Generic levels of microbiomes were associated with periodontal disease, including *Variovorax, Prevotella, Porphyromonas*, and *Dialister* [57].

In a previous study, *Bacillus cereus* was also found in >90% of microbial communities from all women with late-onset preeclampsia (*n* = 4), but not in those from all women with early-onset preeclampsia (*n* = 3). 16S rRNA genes were negative in all venous blood, urine, and amniotic fluid samples, except in one woman whose amniotic fluid had *B. cereus*. This bacterium is an opportunistic pathogen among gastrointestinal infections (A09), and is widely recognized as a challenging problem in the food industry [58]. Outbreaks of *B. cereus* gastroenteritis were reported [59−61]. Interestingly, the numbers of patients with *B. cereus* bloodstream infections were higher in summer, and the source was urinary catheters [62]. Meanwhile, women with a month of conception during the summer (OR 1.22, 95% credible interval [CrI] 0.89−1.65) also had higher incidences of preeclampsia, but those who delivered during winter (OR 3.33, 95% CrI 0.31−35.48) had higher incidences of eclampsia in both the northern and southern hemispheres [63].

Inflammatory disorders of the breast (N61) was found as text profiling result of either N codes (diseases of the genitourinary system within 1 last year to the event) or breast related codes (within the pregnancy period to the event). However, most studies investigated breast neoplasms (N601, N608, and N609) as effects of preeclampsia instead of focusing on the inflammation alone. The incidences were lower in women with preeclampsia compared to non-preeclampsia if these were adjusted by the sex of the fetus (RR 0.85, 95% CI 0.77−0.95; $I^2$=49%; *df*=5) [64]. Inflammatory disorders of the breast might be related to neoplasms in that study, since there are strong linkages between toll-like receptor (TLR)-mediated regulation of inflammation during breast cancer [65]. In context of this study, the inflammatory disorders of the breast (N61) before pregnancy might be more related to preeclampsia compared to those during pregnancy. This was implied by the finding that the period of 9~<12 months to the event was the period which showed the best predictive performance.

Preeclampsia is also associated with other digestive system-related codes, such as dental caries (K029) and excessive attrition of the teeth (K30). Pregnant women with dental caries had a higher prevalence of preeclampsia compared to normotensive controls (adjusted odds ratio [aOR] 1.76, 95% CI 1.43−2.18) [66]. However, excessive attrition of the teeth might not be directly associated with preeclampsia, but with age, because attrition of the teeth was greater in the age group of 51−60 years compared to those in either the age group of 20−30 years or other younger age groups (*p*<0.003) [67].

Surprisingly, eye-related codes were included in several diagnosis predictors in our model. Codes included those for disorders of refraction, especially myopia. These disorders are likely associated with age. A meta-regression of age and the year of birth with the prevalence of myopia demonstrated a U-shaped relationship (*p*<0.05) with an increasing prevalence from the age of 30 years [68]. The prevalence of hypermetropia increased from the age group of 41~50 (aOR 2.7, 95% CI 1.3−5.7; *p* = 0.007) to 61−70 years (aOR 5.8, 95% CI 2.7−12.7; *p*<0.001) [69]. Age was also correlated with astigmatisms of ≥1.00 D, in terms of both corneal (OR 1.007, 95% CI 1.001−1.013; *p* = 0.02) and refractive astigmatism (OR 1.043, 95% CI 1.036−1.051; *p*<0.0001) [70]. This diagnosis predictor was a part of the principal components that also included age, as shown by text profiling results. In addition, other eye-related codes had unclear associations with preeclampsia. But, these diseases are common in clinical practice and might involve bacterial or viral infections, such as hordeolum and other deep inflammation of the eyelid (H000) [71], a retained foreign body following a penetrating wound of the eye orbit (H055) [72], conjunctivitis (H109) [73], otitis externa (H608 and H609) [74], and vertigo (H811 and H814) [75,76].

Another principal component included D codes (neoplasms or diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism). The text profiling results showed codes for anemia (D509, D648, and D649). Severe anemia was associated with preeclampsia/eclampsia in both nulliparous (aOR 3.74, 95% CI 2.90−4.81) and multiparous (aOR 3.45, 95% CI 2.79−4.25) women [77]. The effect of anemia in our model may have been adjusted by other conditions, since it was also a part of the principal components. A qualitative assessment from a systematic review described increases in anemia and eclampsia during periods of greater rainfall, which included studies suggesting that the seasonality of those diseases was associated with malaria [78].

The random forest outperformed other machine learning algorithms in this study. This algorithm was also the best model with superior predictive performance in several studies that developed clinical predictive models for such conditions as end-stage renal disease [79], incident delirium [80], H3K27M mutations in brainstem gliomas [81], prostate cancer [82], in-hospital mortality [83], chemoradiotherapy outcomes [84] and acute kidney injuries [85]. Features included demographic characteristics, comorbidities, medical histories, clinical predictors, laboratory findings, medical imaging, treatments, and biomarkers. Two studies also utilized routine registry data that were preprocessed by a nested case-control design [80,83].

Although machine learning algorithms did not show higher AUROC compared to logistic regression for clinical prediction models, particularly those with low risk of bias [86], the previous systematic review was limited. It did not compare the algorithms using the same datasets. Modern machine learning algorithms, including random forest, are data hungry; thus, the algorithms need a dataset with higher EPV than those for logistic regression [19]. Problem of low EPV causes overfitting in turn causing the predictive performance far poorer using external validation set compared to those using either training or internal validation set [28]. Meanwhile, the systematic review demonstrated that the most common cause of high risk of bias is the external validation method. The comparison of predictive performance was confounded by factors other than the model algorithm, such as sample size and number of predictors. These factors were our reasons to utilize a dataset with larger EPV and to apply external validation method rigorously based on the PROBAST guidelines.

Calibration slopes of all models in this study were significantly different to 1 based on the 95% confidence interval although the AUROCs of models with several algorithms were considerably moderate to high, including the random forest algorithm. These may happen in a prediction model using nonlinear machine learning algorithm. [87]. We can expect a model with both well-calibrated and high AUROC if the model uses a linear function. Logistic regression predicts the outcome probability as a function where predictors come into the model linearly [88]. Depend on sample size, a prediction model that assumes linearity between predictors and outcome may have a poor predictive performance [86]. If a training set lacks of number of event outcome for one predictor adjusted by the others, linearity will be unlikely found between the predictor and the outcome although the association may be linear in larger sample size. Meanwhile, nonlinearity was found in many associations between several predictors and preeclampsia outcome [89−93]. Preeclampsia prediction using a nonlinear machine learning algorithm may outperform that using the linear algorithm depend on other factors. These made random forest outperformed the other algorithms. The factors were the dominance of predictors derived from those with binomial probability (the medical histories) [94], and high-dimensional training set with large sample size [95].

All of the selected candidate features for the random forest model were continuous variables, except three features which were categorical in origin from the NHID-BPJSKES. Continuous variables were the proportion of days with a visit to total days since recorded in the database. Instances with missing values of features were simply removed because of their minority. There was no significant difference in any selected candidate features before and after removing missing data. Categorization of continuous variables can lead to optimism as it uses cut offs based on the same dataset, but handling missing data by exclusion is still acceptable if it does not significantly change the distributions of predictors and outcomes [28].

The random forest algorithm in the best model did not apply built-in feature selection for the 17 features. These were taken from preliminary filtering of 95 candidates and 17 principal components by the MvLRM with forward selection. Principal components 8 and 10 were selected over 1 and 2. This was because those were not only compared to the other principal components but also the original candidates; thus, selection would not follow the way the principal components were ranked and selected. Using this workflow, we achieved ≥20 EPVs for the MvLRM and 180 EPVs for the random forest. These algorithms need 20−50 and 50−200 EPVs, respectively, utilizing three datasets [19].

In our dataset, candidate features like blood pressure and proteinuria were not available. However, unlike diagnostic prediction model, the use of candidate features that are parts of outcome definition should be avoided, i.e. blood pressure and proteinuria that are parts of preeclampsia definition. This situation is called as outcome leakage in either prognostic prediction modeling or machine learning prediction [26,28]. Nonetheless, we validate our best model by comparing the predictive performance in external validation with those of the traditional clinical scoring models for prognostic prediction of pregnancy outcome resulting preeclampsia. The models were those from NICE and ACOG in external validation of previous studies (Table 5) [15,37].

We applied cross-validation and geographical/temporal splitting for internal and external validation, respectively. Cross-validation was applied from feature selection to model selection with parameter updating. This technique was recommended by PROBAST guidelines as an unbiased method rather than using a non-repeated random splitting [28]. Our external validation sets included >100 instances in the case group, as recommended by the same guidelines. We applied geographical and temporal randomization instead of simple randomization to split the dataset for external validation. Simple randomization was not recommended because the training and external validation sets that only differed by chance would probably have similar predictive performances [96]. By geographical and temporal splitting, our external validations were similar to those from independent validation studies because the variance of features in our training set, that were related to the city and time period, were unobserved in our external validation sets. Therefore, this study applied standards for feature extraction, feature selection, model validation, and others that were designed to avoid bias and overfitting in development of either prognostic factors or the prediction model [25,28,96,97].

Our random forest model had the best predictive performance for the period of 9−<12 months to the event compared to 2 days−<6 months, 6−<9 months, and 12−24 months to the event. The first trimester was the most frequent period in which most studies developed a preeclampsia prediction model ($n$ = 42/70, 61.43%) [38]. This is approximately equivalent to <9 months to the event. Only one of the studies developed the model before conception, and two studies used only non-time-varying maternal characteristics (i.e., ethnicity or social class). This was probably due to a common belief about preeclampsia that the pathogenesis begins from 11 to 13 weeks' gestation [98]. In the two-stage model of preeclampsia pathogenesis, this disease is initiated by placental dysfunction, followed by endothelial dysfunction; yet, various theories have attempted to explain the cause of placental dysfunction. However, pregnant women with preeclampsia and endothelial dysfunction have been reported without placental disease [99]. Impaired endometrial maturations before and during early pregnancy was also demonstrated in preeclamptic women [31,100,101]. Only the late secretory phase of the menstruation cycle was impaired, and this phase is the only one enabling a successful pregnancy. This evidence suggests that an event may impair the endometrium in the last menstruation period before pregnancy.

Our best model outperformed prognostic prediction models that only used demographic and/or clinical predictors from previous studies. The models included preeclampsia risk scoring from the NICE and ACOG. In particular, the precision or positive predictive value of our model was distinguished compared to those of previous models. This is because many preeclampsia prediction models were developed with an imbalanced dataset in which preeclampsia group was minor compared to the control group. Imbalanced outcomes impair the precision of prediction models, which can be handled by oversampling [102]. The most widely used oversampling method, which is called the synthetic minority over-sampling technique (SMOTE), can also improve the sensitivity in the minor-positives training set, although slightly reducing the accuracy and specificity [103]. However, SMOTE may cause problems in the distribution of the dataset [104]. We applied naïve random oversampling that randomly sampled the minority outcome with replacement. A machine learning predictive model using a dataset with this oversampling technique had a fairly better improvement in the AUROCs of many machine learning algorithms compared to those with an imbalanced dataset [105]. We also provide evidence that this technique did not affect the distribution of predictors and outcome in this study.

However, several limitations of our prediction model should be considered. The training set in this study consisted of patients with medical histories that might be recorded by several healthcare facilities. The predictive performance might be poor in a healthcare facility at a certain visit if the patient had medical histories that were recorded mostly in databases of other healthcare facilities. The model deployment will need an information system that can be used by inter-healthcare facilities. Transforming this model into a risk calculator in clinical practice need further validation. Since we have no explicit information in our dataset to identify the healthcare facilities where the visits took place, we could not construct another external validation set that approached prediction with medical histories from a patient retrieved from single database of a healthcare facility. Nearly one-third of instances in our training set were also back-censored; thus, medical histories were not observed by our model, particularly during 12−24 months to the event.

Other predictors, such as body mass index and gestational age at diagnosis or at delivery, were not available in the original dataset we utilized for model development in this study. This model also could not differentiate early- vs. late-onset and preterm vs. term preeclampsia with or without intrauterine growth restrictions. However, the development of prediction model in this study was not intended to predict either subtypes of preeclampsia or the adverse events. The model was intended to be preliminary prediction model to determine pregnant women that will be predicted by the other models with high specificity and advanced predictors. Our model had a high sensitivity and low-cost predictors. The model also had a better precision compared to the other model with low-cost predictor. This model may reduce false positives of preeclampsia. At same specificity, the sensitivity of our model was also higher than those from previous models with low-cost predictors. Our preliminary prediction model may reduce ~20% of the cost at community level to use the highly-specific prediction model with advanced predictors. This could be achieved if the advanced prediction is only applied to the predicted positives by our model. We also recommend to confirm the predicted negatives by our model using the low-cost prediction model from MacDonald-Wallis (2015) [33]. Conceivably, only by combined prediction, the low-cost prediction will improve neonatal morbidity and ICU utilization without sacrificing the mother safety.

Indeed, Indonesian people had rich human genetic diversity [106,107]. The genetic variation may affect characteristics of many diseases; yet, the geographic variation may be different from the variants affecting the diseases [108]. Nevertheless, the Indonesian genomic variation only covered those of Asian and Austronesian [107,109]. The generalizability of this model may be limited to Indonesia or other regions that share similar race/ethnicity and climate conditions. More external validations are still needed to consider this model being valid in other populations. This was not feasible for this study because we did not have access to health insurance dataset in other countries, that have similar data structure and the same classification system of diseases.

In addition, because gestational period information was not available, we could not determine whether 9~<12 months to the event was equivalent to the near or far periods before pregnancy. Selected features may be bystanders rather than causal or risk factors for preeclampsia; thus, this model should be interpreted carefully. Further investigation at the molecular level should be conducted to confirm this association.

In conclusion, the best model in this study had robust performances on all validation sets, including external validations. This may describe the generalizability of a prediction model to unobserved samples. Our model also outperformed previous models, especially in precision, which used maternal characteristics and medical histories without biophysical or biochemical markers. This may reduce false positives for the decision for an early delivery, especially in poor-resource settings. But, a future study is needed to investigate the impact of our prediction model for reducing the false positives. In addition, it applied the random forest algorithm on features that were best to predict preeclampsia/eclampsia within 9−<12 months to the event and corresponded to findings of previous studies. This may give more insights into the preeclampsia pathogenesis; however, future investigations are needed to confirm these insights. Because medical histories used by our model were recorded from multiple healthcare facilities, we also recommend health insurance companies, particularly in Indonesia, facilitate this model deployment in privacy-aware information systems used by inter-healthcare facilities. Using our prediction model that showed acceptable performances, expense efficiency of insurance management may be improved in addition to preventing inefficient use of neonatal ICUs as expected.

## Declaration of Competing Interest

Dr. Sufriyana, Dr. Wu, and Dr. Su have nothing to disclose.

## Acknowledgments

The social security administrator for health or *badan penyelenggara jaminan sosial (BPJS) kesehatan* in Indonesia gave permission to access the sample dataset in this study (dataset request approval no.: 12047/I.2/0919).

## Funding sources

This study was funded by the Ministry of Science and Technology (MOST) in Taiwan (grant number MOST107-2221-E-038-016 and MOST108-2221-E-038-018) and the Higher Education Sprout Project by Ministry of Education (MOE) in Taiwan (grant number DP2-107-21121-01-A-01 and DP2-108-21121-01-A-01-04) to Emily Chia-Yu Su. The sponsor had no role in the research design or contents of the manuscript for publication. Emily Chia-Yu Su had full access to all the data

in the study and had final responsibility for the decision to submit for publication.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2020.102710.

## References

[1] Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for preeclampsia: us preventive services task force recommendation statement. JAMA 2017;317:1661–7. doi: 10.1001/jama.2017.3439.

[2] Henderson JT, Thompson JH, Burda BU, Cantor A. Preeclampsia screening: evidence report and systematic review for the us preventive services task force. JAMA 2017;317:1668–83. doi: 10.1001/jama.2016.18315.

[3] Rolnik DL, Wright D, Poon LC, et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. N Engl J Med 2017;377:613–22. doi: 10.1056/NEJMoa1704559.

[4] Huluta I, Panaitescu AM. Prediction of preeclampsia developing at term. Ginekol Pol 2018;89:217–20. doi: 10.5603/GP.a2018.0037.

[5] Nissaisorakarn P, Sharif S, Jim B. Hypertension in pregnancy: defining blood pressure goals and the value of biomarkers for preeclampsia. Curr Cardiol Rep 2016;18:131. doi: 10.1007/s11886-016-0782-1.

[6] Schneider H. Placental dysfunction as a key element in the pathogenesis of preeclampsia. Dev Period Med 2017;21:309–16. PMID: 29291358.

[7] Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur J Obstet Gynecol Reprod Biol 2013;170:1–7. doi: 10.1016/j.ejogrb.2013.05.005.

[8] Souza MA, de Lourdes Brizot M, Biancolin SE, et al. Placental weight and birth weight to placental weight ratio in monochorionic and dichorionic growth-restricted and non-growth-restricted twins. Clinics 2017;72:265–71. doi: 10.6061/clinics/2017(05)02.

[9] Yu J, Flatley C, Greer RM, Kumar S. Birth-weight centiles and the risk of serious adverse neonatal outcomes at term. J Perinat Med 2018;46:1048–56. doi: 10.1515/jpm-2017-0176.

[10] Nijkamp JW, Sebire NJ, Bouman K, Korteweg FJ, Erwich J, Gordijn SJ. Perinatal death investigations: what is current practice. Semin Fetal Neonatal Med 2017;22:167–75. doi: 10.1016/j.siny.2017.02.005.

[11] Wright D, Rolnik DL, Syngelaki A, et al. Aspirin for evidence-based preeclampsia prevention trial: effect of aspirin on length of stay in the neonatal intensive care unit. Am J Obstet Gynecol 2018;218:612.e1–6. doi: 10.1016/j.ajog.2018.02.014.

[12] Roberge S, Bujold E, Nicolaides KH. Aspirin for the prevention of preterm and term preeclampsia: systematic review and metaanalysis. Am J Obstet Gynecol 2018;218:287–93 e1. doi: 10.1016/j.ajog.2017.11.561.

[13] Townsend R, Khalil A, Premakumar Y, et al. Prediction of pre-eclampsia: review of reviews. Ultrasound Obstet Gynecol 2019;54:16–27. doi: 10.1002/uog.20117.

[14] Wright D, Tan MY, O'Gorman N, et al. Predictive performance of the competing risk model in screening for preeclampsia. Am J Obstet Gynecol 2019;220 199.e1−99.e13. doi: 10.1016/j.ajog.2018.11.1087.

[15] Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol 2015;213:62.e1–62.e10. doi: 10.1016/j.ajog.2015.02.018.

[16] Nair TM. Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia. Comput Biol Chem 2018;75:222–30. doi: 10.1016/j.compbiolchem.2018.05.011.

[17] Agarwal V, Podchiyska T, Banda JM, et al. Learning statistical models of phenotypes using noisy labeled training data. J Am Med Inform Assoc 2016;23:1166–73. doi: 10.1093/jamia/ocw028.

[18] Jhee JH, Lee S, Park Y, et al. Prediction model development of late-onset preeclampsia using machine learning-based methods. PLoS ONE 2019;14: e0221202. doi: 10.1371/journal.pone.0221202.

[19] van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14:137. doi: 10.1186/1471-2288-14-137.

[20] Lin LT, Wang PH, Tsui KH, et al. Increased risk of systemic lupus erythematosus in pregnancy-induced hypertension: a nationwide population-based retrospective cohort study. Medicine 2016;95:e4407. doi: 10.1097/md.0000000000004407.

[21] Lin LT, Hu LY, Tang PL, et al. Do racial differences exist in the association between pregnancy-induced hypertension and breast cancer risk. Hypertens Pregnancy 2017;36:138–44. doi: 10.1080/10641955.2016.1258411.

[22] Li JY, Wang PH, Vitale SG, et al. Pregnancy-induced hypertension is an independent risk factor for meconium aspiration syndrome: a retrospective population based cohort study. Taiwan J Obstet Gynecol 2019;58:396–400. doi: 10.1016/j.tjog.2018.11.034.

[23] Chen SN, Wang PH, Hsieh MF, Tsai HW, Lin LT, Tsui KH. Maternal pregnancy-induced hypertension increases the subsequent risk of neonatal candidiasis: a nationwide population-based cohort study. Taiwan J Obstet Gynecol 2019;58:261–5. doi: 10.1016/j.tjog.2019.01.017.

[24] Sultan AA, West J, Grainge MJ, et al. Development and validation of risk prediction model for venous thromboembolism in postpartum women: multinational cohort study. BMJ 2016;355:i6253. doi: 10.1136/bmj.i6253.

[25] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8. doi: 10.7326/m18-1376.

[26] Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18:e323. doi: 10.2196/jmir.5870.

[27] Ariawan I, Sartono B, Jaya C, et al. Sample dataset of the BPJS kesehatan 2015-2016. Jakarta: BPJS Kesehatan; 2019.

[28] Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1–w33. doi: 10.7326/m18-1377.

[29] Nakagawa K, Lim E, Harvey S, Miyamura J, Juarez DT. Racial/ethnic disparities in the association between preeclampsia risk factors and preeclampsia among women residing in Hawaii. Matern Child Health J 2016;20:1814–24. doi: 10.1007/s10995-016-1984-2.

[30] TePoel MR, Saftlas AF, Wallis AB. Association of seasonality with hypertension in pregnancy: a systematic review. J Reprod Immunol 2011;89:140–52. doi: 10.1016/j.jri.2011.01.020.

[31] Rabaglino MB, Post Uiterweer ED, Jeyabalan A, Hogge WA, Conrad KP. Bioinformatics approach reveals evidence for impaired endometrial maturation before and during early pregnancy in women who developed preeclampsia. Hypertension 2015;65:421–9. doi: 10.1161/hypertensionaha.114.04481.

[32] Lisowska M, Pietrucha T, Sakowicz A. Preeclampsia and related cardiovascular risk: common genetic background. Curr Hypertens Rep 2018;20:71. doi: 10.1007/s11906-018-0869-8.

[33] Macdonald-Wallis C, Silverwood RJ, De Stavola BL, et al. Antenatal blood pressure for prediction of pre-eclampsia, preterm birth, and small for gestational age babies: development and validation in two general population cohorts. BMJ 2015;351. doi: 10.1136/bmj.h5948.

[34] Guy GP, Ling HZ, Garcia P, Poon LC, Nicolaides KH. Maternal cardiac function at 35-37 weeks' gestation: prediction of pre-eclampsia and gestational hypertension. Ultrasound Obstet Gynecol 2017;49:61–6. doi: 10.1002/uog.17300.

[35] Rocha RS, Alves JAG, Maia EHMSB, et al. Simple approach based on maternal characteristics and mean arterial pressure for the prediction of preeclampsia in the first trimester of pregnancy. J Perinat Med 2017;45:843–9. doi: 10.1515/jpm-2016-0418.

[36] Viguiliouk E, Park AL, Berger H, Geary MP, Ray JG. A simple clinical method to identify women at higher risk of preeclampsia. Pregnancy Hypertens 2017;10:10–3. doi: 10.1016/j.preghy.2017.07.145.

[37] Rocha RS, Gurgel Alves JA, Bezerra Maia EHMS, et al. Comparison of three algorithms for prediction preeclampsia in the first trimester of pregnancy. Pregnancy Hypertens 2017;10:113–7. doi: 10.1016/j.preghy.2017.07.146.

[38] De Kat AC, Hirst J, Woodward M, Kennedy S, Peters SA. Prediction models for preeclampsia: a systematic review. Pregnancy Hypertens 2019;16:48–66. doi: 10.1016/j.preghy.2019.03.005.

[39] Bartsch E, Medcalf KE, Park AL, et al. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. BMJ 2016;353. doi: 10.1136/bmj.i1753.

[40] Cheng PJ, Huang SY, Su SY, Hsiao CH, Peng HH, Duan T. Prognostic value of cardiovascular disease risk factors measured in the first-trimester on the severity of preeclampsia. Medicine (Baltimore) 2016;95:e2653. doi: 10.1097/md.0000000000002653.

[41] Fournier P, Fourcade J, Roncalli J, Salvayre R, Galinier M, Causse E. Homocysteine in chronic heart failure. Clin Lab 2015;61:1137–45 pmid: 26554232.

[42] Wu P, Haththotuwa R, Kwok CS, et al. Preeclampsia and future cardiovascular health: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2017;10. doi: 10.1161/circoutcomes.116.003497.

[43] You SH, Cheng PJ, Chung TT, Kuo CF, Wu HM, Chu PH. Population-based trends and risk factors of early- and late-onset preeclampsia in Taiwan 2001−2014. BMC Pregnancy Childbirth 2018;18:199. doi: 10.1186/s12884-018-1845-7.

[44] Hamann CR, Egeberg A, Wollenberg A, Gislason G, Skov L, Thyssen JP. Pregnancy complications, treatment characteristics and birth outcomes in women with atopic dermatitis in denmark. J Eur Acad Dermatol Venereol 2019;33:577–87. doi: 10.1111/jdv.15256.

[45] Mirzakhani H, Carey VJ, McElrath TF, et al. Impact of preeclampsia on the relationship between maternal asthma and offspring asthma. an observation from the vdaart clinical trial. Am J Respir Crit Care Med 2019;199:32–42. doi: 10.1164/rccm.201804-0770OC.

[46] Machelart A, Potemberg G, Van Maele L, et al. Allergic asthma favors brucella growth in the lungs of infected mice. Front Immunol 2018;9:1856. doi: 10.3389/fimmu.2018.01856.

[47] Eloe-Fadrosh EA, McArthur MA, Seekatz AM, et al. Impact of oral typhoid vaccination on the human gut microbiota and correlations with s. Typhi-specific immunological responses. PLoS ONE 2013;8:e62026. doi: 10.1371/journal.pone.0062026.

[48] Liu J, Yang H, Yin Z, et al. Remodeling of the gut microbiota and structural shifts in preeclampsia patients in south china. Eur J Clin Microbiol Infect Dis 2017;36:713–9. doi: 10.1007/s10096-016-2853-z.

[49] Lieberman TD, Wilson D, Misra R, et al. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated mycobacterium tuberculosis. Nat Med 2016;22:1470–4. doi: 10.1038/nm.4205.

[50] Pillay Y, Moodley J, Naicker T. The role of the complement system in HIV infection and preeclampsia. Inflamm Res 2019;68:459–69. doi: 10.1007/s00011-019-01240-0.

[51] Eldin C, Melenotte C, Mediannikov O, et al. From q fever to *Coxiella burnetii* infection: a paradigm change. Clin Microbiol Rev 2017;30:115–90. doi: 10.1128/cmr.00045-16.

[52] Malavige GN, Ogg GS. Pathogenesis of vascular leak in dengue virus infection. Immunology 2017;151:261–9. doi: 10.1111/imm.12748.

[53] Boeldt DS, Bird IM. Vascular adaptation in pregnancy and endothelial dysfunction in preeclampsia. J Endocrinol 2017;232:R27–44. doi: 10.1530/joe-16-0340.

[54] Sivasubramaniam R, Douglas R. The microbiome and chronic rhinosinusitis. World J Otorhinolaryngol Head Neck Surg 2018;4:216–21. doi: 10.1016/j.wjorl.2018.08.004.

[55] Solt I. The human microbiome and the great obstetrical syndromes: a new frontier in maternal-fetal medicine. Best Pract Res Clin Obstet Gynaecol 2015;29:165–75. doi: 10.1016/j.bpobgyn.2014.04.024.

[56] Conde-Agudelo A, Villar J, Lindheimer M. Maternal infection and risk of preeclampsia: systematic review and metaanalysis. Am J Obstet Gynecol 2008;198:7–22. doi: 10.1016/j.ajog.2007.07.040.

[57] Amarasekara R, Jayasekara RW, Senanayake H, Dissanayake VH. Microbiome of the placenta in pre-eclampsia supports the role of bacteria in the multifactorial cause of pre-eclampsia. J Obstet Gynaecol Res 2015;41:662–9. doi: 10.1111/jog.12619.

[58] Ehling-Schulz M, Frenzel E, Gohar M. Food-bacteria interplay: pathometabolism of emetic *Bacillus cereus*. Front Microbiol 2015;6:704. doi: 10.3389/fmicb.2015.00704.

[59] Al-Abri SS, Al-Jardani AK, Al-Hosni MS, Kurup PJ, Al-Busaidi S, Beeching NJ. A hospital acquired outbreak of *Bacillus cereus* gastroenteritis, oman. J Infect Public Health 2011;4:180–6. doi: 10.1016/j.jiph.2011.05.003.

[60] Bennett SD, Walsh KA, Gould LH. Foodborne disease outbreaks caused by *Bacillus cereus*, *Clostridium perfringens*, and *Staphylococcus aureus*—United States, 1998—2008. Clin Infect Dis 2013;57:425–33. doi: 10.1093/cid/cit244.

[61] Thirkell CE, Sloan-Gardner TS, Kacmarek MC, Polkinghorne B. An outbreak of *Bacillus cereus* toxin-mediated emetic and diarrhoeal syndromes at a restaurant in Canberra, Australia 2018. Commun Dis Intell (2018) 2019;43. doi: 10.33321/cdi.2019.43.40.

[62] Kato K, Matsumura Y, Yamamoto M, et al. Erratum to: seasonal trend and clinical presentation of *Bacillus cereus* bloodstream infection: association with summer and indwelling catheter. Eur J Clin Microbiol Infect Dis 2016;35:875–83. doi: 10.1007/s10096-016-2618-8.

[63] Beltran AJ, Wu J, Laurent O. Associations of meteorology with adverse pregnancy outcomes: a systematic review of preeclampsia, preterm birth and birth weight. Int J Environ Res Public Health 2013;11:91–172. doi: 10.3390/ijerph110100091.

[64] Sun M, Fan Y, Hou Y, Fan Y. Preeclampsia and maternal risk of breast cancer: a meta-analysis of cohort studies. J Matern Fetal Neonatal Med 2018;31:2484–91. doi: 10.1080/14767058.2017.1342806.

[65] Bhatelia K, Singh K, Singh R. Tlrs: linking inflammation and breast cancer. Cell Signal 2014;26:2350–7. doi: 10.1016/j.cellsig.2014.07.035.

[66] Khader Y, Jibreal M, Burgan S, Amarin Z. Risk indicators of pre-eclampsia in north Jordan: is dental caries involved? Gynecol Obstet Invest 2007;63:181–7. doi: 10.1159/000097633.

[67] Sarig R, Hershkovitz I, Shpack N, May H, Vardimon AD. Rate and pattern of interproximal dental attrition. Eur J Oral Sci 2015;123:276–81. doi: 10.1111/eos.12198.

[68] Pan CW, Dirani M, Cheng CY, Wong TY, Saw SM. The age-specific prevalence of myopia in Asia: a meta-analysis. Optom Vis Sci 2015;92:258–66. doi: 10.1097/opx.0000000000000516.

[69] Hashemi H, Nabovati P, Yekta A, Shokrollahzadeh F, Khabazkhoob M. The prevalence of refractive errors among adult rural populations in Iran. Clin Exp Optom 2018;101:84–9. doi: 10.1111/cxo.12565.

[70] Sanfilippo PG, Yazar S, Kearns L, Sherwin JC, Hewitt AW, Mackey DA. Distribution of astigmatism as a function of age in an Australian population. Acta Ophthalmol 2015;93:e377–85. doi: 10.1111/aos.12644.

[71] Bragg KJ, Le JK. Hordeolum. [Accessed 10/16/2019]. StatPearls [Internet]. Treasure IslandFL: StatPearls Publishing; 2019. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441985/.

[72] Turliuc DM, Costan VV, Cucu AI, Costea CF. Intraorbital foreign body. Rev Med Chir Soc Med Nat Iasi 2015;119:179–84 pmid: 25970964.

[73] Alfonso SA, Fawley JD, Alexa Lu X. Conjunctivitis. Prim Care 2015;42:325–45. doi: 10.1016/j.pop.2015.05.001.

[74] Wipperman J. Otitis externa. Prim Care 2014;41:1–9. doi: 10.1016/j.pop.2013.10.001.

[75] Nuti D, Masini M, Mandala M. Benign paroxysmal positional vertigo and its variants. Handb Clin Neurol 2016;137:241–56. doi: 10.1016/b978-0-444-63437-5.00018-2.

[76] Choi JY, Lee SH, Kim JS. Central vertigo. Curr Opin Neurol 2018;31:81–9. doi: 10.1097/wco.0000000000000511.

[77] Chen C, Grewal J, Betran AP, Vogel JP, Souza JP, Zhang J. Severe anemia, sickle cell disease, and thalassemia as risk factors for hypertensive disorders in pregnancy in developing countries. Pregnancy Hypertens 2018;13:141–7. doi: 10.1016/j.preghy.2018.06.001.

[78] Hlimi T. Association of anemia, pre-eclampsia and eclampsia with seasonality: a realist systematic review. Health Place 2015;31:180–92. doi: 10.1016/j.healthplace.2014.12.003.

[79] Liu Y, Zhang Y, Liu D, et al. Prediction of esrd in iga nephropathy patients from an Asian cohort: a random forest model. Kidney Blood Press Res 2018;43:1852–64. doi: 10.1159/000495818.

[80] Corradi JP, Thompson S, Mather JF, Waszynski CM, Dicks RS. Prediction of incident delirium using a random forest classifier. J Med Syst 2018;42:261. doi: 10.1007/s10916-018-1109-0.

[81] Pan CC, Liu J, Tang J, et al. A machine learning-based prediction model of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features. Radiother Oncol 2019;130:172–9. doi: 10.1016/j.radonc.2018.07.011.

[82] Xiao LH, Chen PR, Gou ZP, et al. Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. Asian J Androl 2017;19:586–90. doi: 10.4103/1008-682x.186884.

[83] Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 2016;23:269–78. doi: 10.1111/acem.12876.

[84] Deist TM, Dankers F, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. Med Phys 2018;45:3449–59. doi: 10.1002/mp.12967.

[85] Chiofolo C, Chbat N, Ghosh E, Eshelman L, Kashani K. Automated continuous acute kidney injury prediction and surveillance: a random forest model. Mayo Clin Proc 2019;94:783–92. doi: 10.1016/j.mayocp.2019.02.009.

[86] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22. doi: 10.1016/j.jclinepi.2019.02.004.

[87] Niculescu-Mizil A., Caruana R.Predicting good probabilities with supervised learning. Proceedings of the twenty-second international conference on Machine learning. 2005:625—32. doi: 10.1145/1102351.1102430.

[88] Deo RC. Machine learning in medicine. Circulation 2015;132:1920–30. doi: 10.1161/circulationaha.115.001593.

[89] Metsala J, Stach-Lempinen B, Gissler M, Eriksson JG, Koivusalo S. Risk of pregnancy complications in relation to maternal prepregnancy body mass index: population-based study from Finland 2006-10. Paediatr Perinat Epidemiol 2016;30:28–37. doi: 10.1111/ppe.12248.

[90] Kosinska-Kaczynska K, Wielgos M. Do normal-weight women pregnant with twins are at the lowest risk of developing preeclampsia? J Matern Fetal Neonatal Med 2017;30:191–3. doi: 10.3109/14767058.2016.1166358.

[91] Jung J, Rahman MM, Rahman MS, et al. Effects of hemoglobin levels during pregnancy on adverse maternal and infant outcomes: a systematic review and meta-analysis. Ann N Y Acad Sci 2019;1450:69–82. doi: 10.1111/nyas.14112.

[92] Baca KM, Simhan HN, Platt RW, Bodnar LM. Low maternal 25-hydroxyvitamin d concentration increases the risk of severe and mild preeclampsia. Ann Epidemiol 2016;26:853–7 e1. doi: 10.1016/j.annepidem.2016.09.015.

[93] Aune D, Saugstad OD, Henriksen T, Tonstad S. Physical activity and the risk of preeclampsia: a systematic review and meta-analysis. Epidemiology 2014;25:331–43. doi: 10.1097/ede.0000000000000036.

[94] Au TC. Random forests, decision trees, and categorical predictors: the "absent levels" problem. J Mach Learn Res 2017;19 URL: http://www.jmlr.org/papers/volume19/16-474/16-474.pdf.

[95] Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. Int J Comp Sci Issues 2012;9. http://ijcsi.org/articles/Random-forests-and-decision-trees.php.

[96] Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73. doi: 10.7326/m14-0698.

[97] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the tripod statement. Ann Intern Med 2015;162:55–63. doi: 10.7326/m14-0697.

[98] Jim B, Karumanchi SA. Preeclampsia: pathogenesis, prevention, and long-term complications. Semin Nephrol 2017;37:386–97. doi: 10.1016/j.semnephrol.2017.05.011.

[99] Redman C. Pre-eclampsia: a complex and variable disease. Pregnancy Hypertens 2014;4:241–2. doi: 10.1016/j.preghy.2014.04.009.

[100] Conrad KP, Rabaglino MB, Post Uiterweer ED. Emerging role for dysregulated decidualization in the genesis of preeclampsia. Placenta 2017;60:119–29. doi: 10.1016/j.placenta.2017.06.005.

[101] Rabaglino MB, Conrad KP. Evidence for shared molecular pathways of dysregulated decidualization in preeclampsia and endometrial disorders revealed by microarray data integration. FASEB J 2019:fj201900662R. doi: 10.1096/fj.201900662R.

[102] Chicco D. Ten quick tips for machine learning in computational biology. BioData Min 2017;10:35. doi: 10.1186/s13040-017-0155-3.

[103] Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F, Khalili D. The impact of oversampling with smote on the performance of 3 classifiers in prediction of type 2 diabetes. Med Decis Making 2016;36:137–44. doi: 10.1177/0272989x14560647.

[104] Zheng Z. Oversampling method for imbalanced classification. Comput Inform 2015;34:1017–37. http://www.cai.sk/ojs/index.php/cai/article/view/1277/724.

[105] Suh Y, Yu J, Mo J, Song L, Kim C. A comparison of oversampling methods on imbalanced topic classification of Korean news articles. J Cogn Sci 2017;18:391–437. doi: 10.17791/jcs.2017.18.4.391.

[106] Lipson M, Cheronet O, Mallick S, et al. Ancient genomes document multiple waves of migration in southeast Asian prehistory. Science 2018;361:92–5. doi: 10.1126/science.aat3188.

[107] Hudjashov G, Karafet TM, Lawson DJ, et al. Complex patterns of admixture across the Indonesian archipelago. Mol Biol Evol 2017;34:2439–52. doi: 10.1093/molbev/msx196.

[108] Lu YF, Goldstein DB, Angrist M, Cavalleri G. Personalized medicine and human genetic diversity. Cold Spring Harb Perspect Med 2014;4:a008581. doi: 10.1101/cshperspect.a008581.

[109] McColl H, Racimo F, Vinner L, et al. The prehistoric peopling of southeast Asia. Science 2018;361:88–92. doi: 10.1126/science.aat3628.