J Korean Med Sci. 2024 Oct 14;39(39):e271
https://doi.org/10.3346/jkms.2024.39.e271
eISSN 1598-6357·pISSN 1011-8934

JKMS

Original Article
Obstetrics & Gynecology

Identification of Preeclamptic Placenta in Whole Slide Images Using Artificial Intelligence Placenta Analysis

Young Mi Jung [1*] **Seyeon Park** [2,3*] **Youngbin Ahn** [3,4] **Haeryoung Kim** [5] **Eun Na Kim** [5] **Hye Eun Park** [6] **Sun Min Kim** [1,7] **Byoung Jae Kim** [7] **Jeesun Lee** [1] **Chan-Wook Park** [1] **Joong Shin Park** [1] **Jong Kwan Jun** [1] **Young-Gon Kim** [3,8,9] **and Seung Mi Lee** [1,9,10,11]

[1]Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul, Korea
[2]Department of Medical Device Development, Seoul National University College of Medicine, Seoul, Korea
[3]Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul, Korea
[4]Interdisciplinary Program of Medical Informatics, Seoul National University College of Medicine, Seoul, Korea
[5]Department of Pathology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea
[6]Department of Pathology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Korea
[7]Department of Obstetrics and Gynecology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Korea
[8]Department of Medicine, Seoul National University College of Medicine, Seoul, Korea
[9]Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Korea
[10]Department of Obstetrics and Gynecology, Seoul National University Hospital, Seoul, Korea
[11]Medical Big Data Research Center & Institute of Reproductive Medicine and Population, Medical Research Center, Seoul National University, Seoul, Korea

Received: Mar 15, 2024
Accepted: Jul 22, 2024
Published online: Aug 20, 2024

Address for Correspondence:
Young-Gon Kim, PhD
Department of Transdisciplinary Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
Email: younggon2.kim@gmail.com

Seung Mi Lee, MD, PhD
Department of Obstetrics and Gynecology, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
Email: smleemd@hanmail.net

*Young Mi Jung and Seyeon Park contributed equally as the first authors of this study.

ORCID iDs
Young Mi Jung
https://orcid.org/0000-0001-7877-4450
Seyeon Park
https://orcid.org/0009-0000-6267-958X
Youngbin Ahn
https://orcid.org/0009-0000-3588-4537
Haeryoung Kim
https://orcid.org/0000-0002-4205-9081

ABSTRACT

Background: Preeclampsia (PE) is a hypertensive pregnancy disorder linked to placental dysfunction, often involving pathological lesions like acute atherosis, decidual vasculopathy, accelerated villous maturation, and fibrinoid deposition. However, there is no gold standard for the pathological diagnosis of PE and this limits the ability of clinicians to distinguish between PE and non-PE pregnancies. Recent advances in computational pathology have provided the opportunity to automate pathological analysis for diagnosis, classification, prediction, and prediction of disease progression. In this study, we assessed whether computational pathology could be used to identify PE placentas.
Methods: A total of 168 placental whole-slide images (WSIs) of patients from Seoul National University Hospital (comprising 84 PE cases and 84 normal controls) were used for model development and internal validation. For external validation of the model, 76 placental slides (including 38 PE cases and 38 normal controls) were obtained from the Boramae Medical Center (BMC). To establish standard criteria for diagnosing PE and distinguishing it from controls using placental WSIs, patch characteristics and quantification of terminal and intermediate villi were employed. In unsupervised learning, *K*-means clustering was conducted as a feature obtained through an Auto Encoder to extract the ratio of each cluster for each WSI. For supervised learning, quantitative assessments of the villi were obtained using a U-Net-based segmentation algorithm. The prediction model was developed using

Eun Na Kim
https://orcid.org/0000-0003-2992-7881
Hye Eun Park
https://orcid.org/0000-0003-1048-2827
Sun Min Kim
https://orcid.org/0000-0002-3369-3937
Byoung Jae Kim
https://orcid.org/0000-0002-2734-997X
Jeesun Lee
https://orcid.org/0000-0001-8567-2364
Chan-Wook Park
https://orcid.org/0000-0002-1290-6438
Joong Shin Park
https://orcid.org/0000-0002-5246-0477
Jong Kwan Jun
https://orcid.org/0000-0002-0242-1736
Young-Gon Kim
https://orcid.org/0000-0003-2148-1299
Seung Mi Lee
https://orcid.org/0000-0002-2543-0408

Funding

This research was supported by a grant from the Seoul National University Hospital (SNUH) research fund (Grant Number: 0320230180) and an Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (Grant Number: 23ZR1100: A Study of Hyper-Connected Thinking Internet Technology by Autonomous Connection, Control, and Evolutionconnecting, Controlling and Evolving Ways). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1F1A1046707).

Disclosure

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Jung YM, Lee SM. Data curation: Jung YM, Park S, Ahn Y. Formal analysis: Jung YM, Park S, Ahn Y, Kim YG. Funding acquisition: Kim YG, Lee SM. Investigation: Jung YM, Park S, Kim H, Ahn Y, Kim EN, Kim YG, Lee SM. Methodology: Jung YM, Kim H, Lee SM. Project administration: Lee SM. Resources: Park CW, Park JS, Jun JK. Software: Park S, Ahn Y. Supervision: Kim H, Kim EN, Kim SM, Kim BJ, Kim YG, Lee SM. Validation: Park HE, Kim SM, Kim BJ. Visualization: Park S, Kim YG. Writing - original draft: Jung YM, Park S. Writing - review & editing: Kim EN, Park HE, Lee J, Park JS, Kim YG, Lee SM.

an ensemble method and was compared with a clinical feature model developed by using placental size features.

Results: Using ensemble modeling, we developed a model to identify PE placentas. The model showed good performance (area under the precision-recall curve [AUPRC], 0.771; 95% confidence interval [CI], 0.752–0.790), with 77.3% of sensitivity and 71.1% of specificity, whereas the clinical feature model showed an AUPRC 0.713 (95% CI, 0.694–0.732) with 55.6% sensitivity and 86.8% specificity. External validation of the predictive model employing the BMC-derived set of placental slides also showed good discrimination (AUPRC, 0.725; 95% CI, 0.720–0.730).

Conclusion: The proposed computational pathology model demonstrated a strong ability to identify preeclamptic placentas. Computational pathology has the potential to improve the identification of PE placentas.

Keywords: Placenta; Preeclampsia; Artificial Intelligence; Unsupervised Learning

INTRODUCTION

Preeclampsia (PE) is a hypertensive disorder specific to pregnancy and occurs in 3–5% of pregnancies globally.[1-3] It leads to varying placental malperfusion levels and the release of substances into the maternal circulation.[4] A set of distinctive pathological lesions have been identified in preeclamptic placentas, including acute atherosis, decidual vasculopathy, accelerated villous maturation, and fibrinoid deposition, collectively offering insights into the intricate interplay between placental abnormalities and PE development.[5-7]

Despite these critical pathological insights, it is currently difficult to distinguish PE and non-PE cases because of the absence of a universally accepted gold standard for the pathological diagnosis of PE. This lack of consensus regarding a definitive diagnostic pathologic criteria hinders the precise delineation of PE and non-PE cases.

Recent advances in artificial intelligence (AI), including whole-slide image (WSI) and AI solutions, have brought about significant changes in the field of pathology.[8-12] These innovations allow us to delve into information beyond what human vision can perceive. This usage can be broadly categorized into two types of research: supervised and unsupervised learning. The primary distinction between these approaches lies in the presence of annotations for regions of interest. Supervised learning is effective when there are clear differentiators between regions with and without lesions, as observed in cases such as lung cancer segmentation through the study of various cancerous and normal patches.[8] Additionally, the development of an end-to-end deep learning framework for WSI segmentation and subsequent analysis of various histopathological tasks has been explored.[9] On the other hand, unsupervised learning offers the ability to discover hidden patterns and structures in unlabeled data, enabling insights that may go unnoticed in supervised approaches, while also having the advantages of being scalable and effective for anomaly detection and data preprocessing tasks.

Similar to these approaches, we used unsupervised learning to cluster patches based on the most compact features, while incorporating both supervised and unsupervised learning techniques to extract meaningful features from WSIs. In this study, we aimed to investigate the potential of harnessing these emerging computational pathology techniques to effectively differentiate preeclamptic and normal placentas.

METHODS

Study population

In this study, we enrolled women who underwent singleton deliveries at Seoul National University Hospital between 2004 and 2019 and underwent placental pathological examinations. Among them, we selected patients for whose placental pathology slides were clearly identifiable for the presence or absence of PE by clinical diagnosis. For model development and internal validation, we matched 84 patients with PE to 84 controls based on age, underlying conditions, and gestational age at delivery. For external validation, we used placental slides from a geotemporal cohort of women who delivered their babies at the Boramae Medical Center (BMC).

Outcomes

PE was defined as the occurrence of hypertension, proteinuria, and multiorgan damage (thrombocytopenia, renal insufficiency, liver involvement, cerebral symptoms, and pulmonary edema). Participants without hypertensive disease during pregnancy (gestational hypertension, PE, eclampsia, chronic hypertension, and superimposed PE) were defined as controls.

Image acquisition

The techniques we employed are described below.

Unsupervised learning

In the unsupervised learning phase, 84 cases were employed, evenly divided into 42 cases of PE and 42 normal control cases. This stage involved the utilization of an Auto Encoder (AE)[13] and *K*-means clustering.[14] Half of the 84 cases were designated for AE training, and the remainder for *K*-means clustering. The AE model was trained on a dataset that was divided into 90% for training and 10% for testing. Within the training set, 10% was allocated for validation.

Supervised learning

Of the 84 WSIs utilized in the unsupervised learning phase, 18 WSIs that had been annotated with villi information were selected. Specifically, 90% of the dataset was allocated for training, and the remaining 10% was evenly divided: half for validation and the other half for testing.

For machine learning (ML) purposes, another set of 84 WSIs (comprising 42 cases of PE and 42 normal controls) were employed to develop and assess a classification model. This model aimed to predict PE using the features extracted in the previous two steps. Due to the limited size of the dataset, the training and test sets were created across 100 iterations. Of the 84 WSIs, 80% were allocated for training, and the remaining 20% were reserved for testing. External validation was performed using a separate set of 76 WSIs (including 38 PE cases and 38 normal controls).

Data pre-processing

The original image size of WSIs (approximately 20 billion pixels; 100 K × 100 K) had put limits on computing memory that hindered its use in model learning. We therefore divided each WSI into lower-sized patch units (pWSI). For unsupervised learning, patches were extracted at a 2.5× magnification to a size of 128 × 128 pixels to identify pathological patterns, observe cellular-level details, while considering memory and computational costs. Intermediate and terminal villi annotations were performed for supervised learning. To minimize labeling time, two specific 2,048 × 2,048 size boxes were selected for each pWSI, with both villi within these boxes labeled. A patch size of 256 × 256 at 10× magnification was employed to capture the villi structures.

https://doi.org/10.3346/jkms.2024.39.e271

JKMS

Patch clustering (unsupervised learning)

To extract the most essential information from each pWSI, we employed the AE algorithm to compress the key details within each patch. Clustering was performed using the *K*-means clustering method. For the AE hyperparameters, we utilized the Adam optimizer with a learning rate of 0.001 and employed the mean squared error as the loss function. The batch size was set to 64, and the model was trained for 1,000 epochs, with early stopping if there was no improvement in 10 consecutive epochs. The latent vector dimension was 256 that was reduced to 50 feature dimensions using principal component analysis. We determined the optimal number of clusters, denoted as *K*, using the elbow method to identify the point at which the inertia experiences a significant change. Consequently, we categorized the pWSI into seven clusters using *K*-means clustering (**Fig. 1**). We assessed cluster distribution in WSIs and statistically compared normal and PE groups using the Mann-Whitney *U* test (**Fig. 2**).

Villi quantification (supervised learning)

We aimed to quantify terminal and intermediate villi using pWSI through supervised learning. Segmentation models for terminal and intermediate villi classes were developed using the U-Net algorithm[15] with a ResNet34[16] as the backbone that was pre-trained with ImageNet.[17] For the hyperparameters of both models, we selected Adam as the optimizer, set the learning rate to 1e-5, and employed dice loss as the loss function. The batch size was configured as 4, and the models were trained for 1,000 epochs, with early stopping in case there was no improvement for 10 consecutive epochs.

ML

In the final step, features extracted from pWSI, including patch clustering and villi quantification results, were utilized as inputs for ML. We evaluated the performance of these features using ML methods to assess their significance in predicting PE. In parallel, we compared the performance of models derived from clinical features, which incorporated placental width, length, and height.

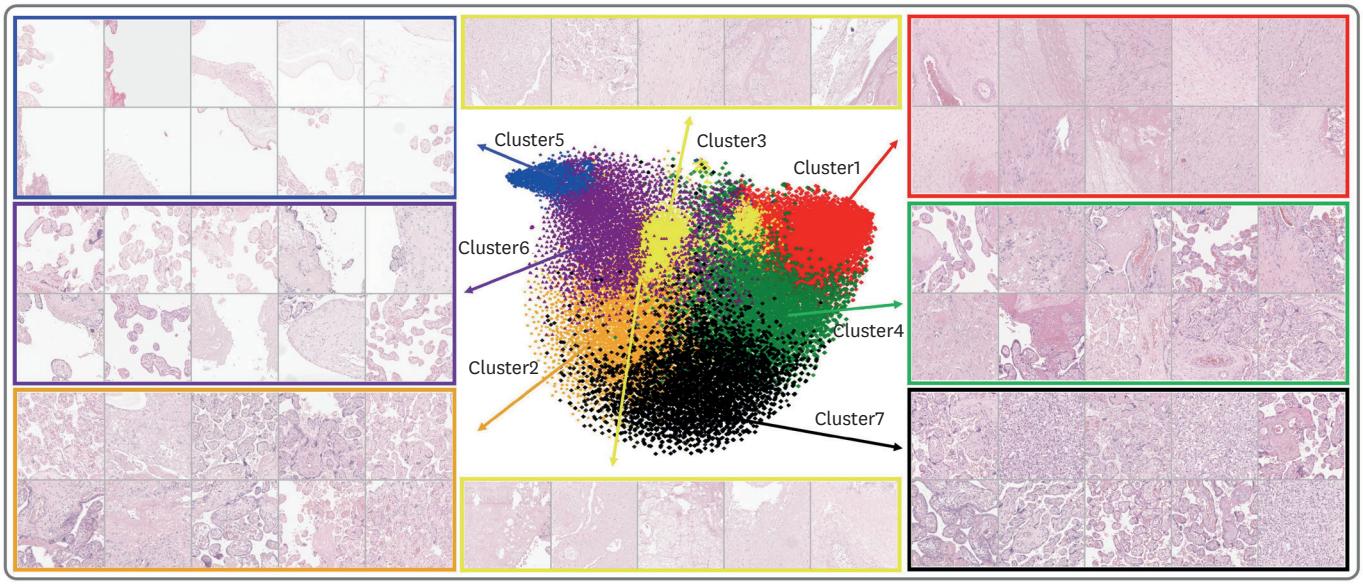

Fig. 1. An example of unsupervised learning-based patches by *K*-means clustering (*K* = 7).

https://doi.org/10.3346/jkms.2024.39.e271

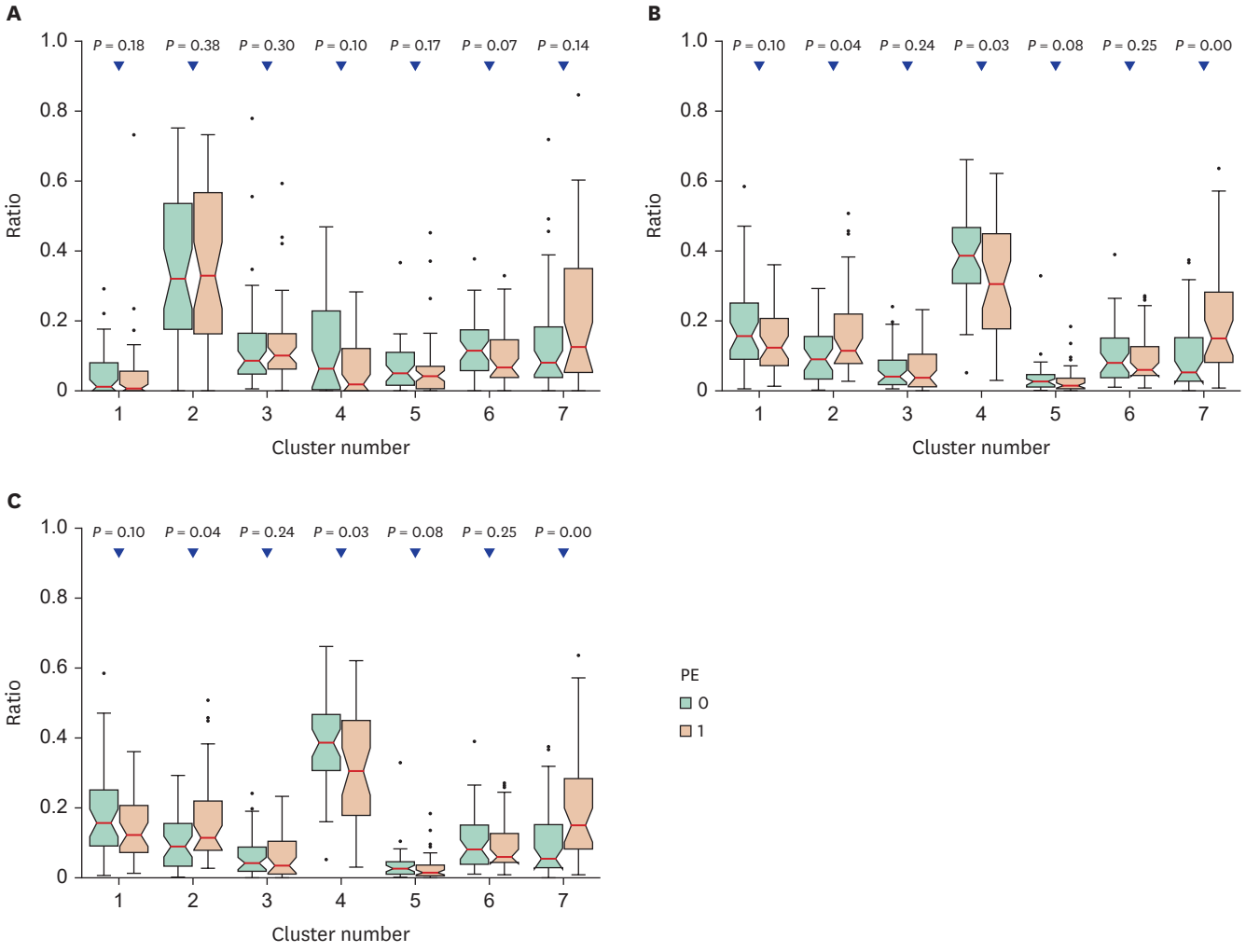

Fig. 2. Cluster distribution of the internal and external dataset. (**A**) Internal validation set. (**B**) External validation set. (**C**) Whole validation set (internal validation set + external validation set).
PE = preeclampsia (1: preeclampsia, 0: normal).

In this study, a total of five ML methods were employed. Random forest (RF)[18] combines multiple decision trees to mitigate overfitting and deliver robust performance in complex datasets. Support vector machine[18] effectively classifies data in high-dimensional spaces by maximizing margins to enhance classification accuracy. Gradient boosting machine[18] progressively corrects errors and learns intricate data patterns, ensuring high predictive accuracy. Logistic regression[18] models linear decision boundaries, making it useful for binary classification with interpretable results and rapid learning. Lastly, CatBoost[19] automates handling of categorical variables, streamlining preprocessing while maintaining excellent predictive performance.

A final ensemble of the three highest efficiency models was developed by comparing the area under the precision-recall curve (AUPRC).[20] Instead of using all extracted features, we applied feature selection, employing an embedded method that assesses feature importance through a ML model. Their importance was determined using the feature importance attributes of the RF model.

https://doi.org/10.3346/jkms.2024.39.e271

JKMS

Statistical analysis

As evaluation metrics,[21] area under the receiver operating characteristic (AUROC), AUPRC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used. External validation was conducted to evaluate the generalization performance of the model, compare the performance using seven metrics as in the internal test set, and obtain the final model and performance by assembling the three models with the highest average AUPRC. The Delong test calculated *P* values to assess the statistical significance in AUPRCs and AUROCs between the deep learning feature model and the clinical feature model.

Experiment environment

All processing in this study was carried out in a shared computing facility consisting of Intel Core i7-10700 CPUs and NVIDIA RTX 2080Ti GPUs.

Ethics statement

This study was approved by the Institutional Review Boards (IRBs) of Seoul National University Hospital (IRB Number: 2311-039-1482). Informed consent was waived because of the retrospective nature of the study.

RESULTS

Table 1 shows the baseline characteristics of the study population. Women with PE were more likely to be nulliparous, have higher body mass index scores, more likely to deliver a small for gestational age baby and more likely to deliver a lower 5-minute Apgar score baby. **Supplementary Table 1** presents the characteristics of the external validation cohort based on the presence or absence of preeclampsia. The baseline characteristics of the model development, internal validation, and external validation cohorts are presented in **Supplementary Table 2**.

The model development process involved three stages: 1) unsupervised learning (patch clustering), 2) supervised learning (villi quantification), and 3) ML (**Fig. 3**). In the first step,

Table 1. Characteristics and pregnancy outcomes of the study population (Seoul National University Hospital)

Characteristics	PE (−) (n = 84)	PE (+) (n = 84)	*P* value
Age, yr	32.2 ± 6.6	34.5 ± 4.1	0.404
Nulliparity	46 (54.8)	64 (76.2)	0.003
BMI (before delivery)	25.0 ± 4.6	27.0 ± 4.0	< 0.001
Pregnancy outcomes			
Gestational diabetes	3 (3.6)	7 (8.3)	0.192
Cesarean section	78 (92.9)	75 (89.3)	0.417
Gestational age at delivery	33.9 ± 2.9	33.9 ± 3.1	0.422
Neonatal outcomes			
Neonatal sex, male	39 (46.4)	42 (50.0)	0.643
Birth weight	2,198 ± 744	1,650 ± 656	< 0.001
SGA	24 (28.6)	58 (69.0)	< 0.001
1-min AS < 7	42 (50.0)	38 (45.2)	0.537
5-min AS < 7	22 (2.62)	9 (10.7)	0.010
NICU admission	42 (50.0)	37 (44.0)	0.440
Cord pH	7.275 ± 0.133	7.240 ± 0.096	0.038

Data are presented as number (%) or mean ± standard deviation.
PE = preeclampsia, BMI = body mass index, SGA = small for gestational age, AS = Apgar score, NICU = neonatal intensive care unit.

https://doi.org/10.3346/jkms.2024.39.e271

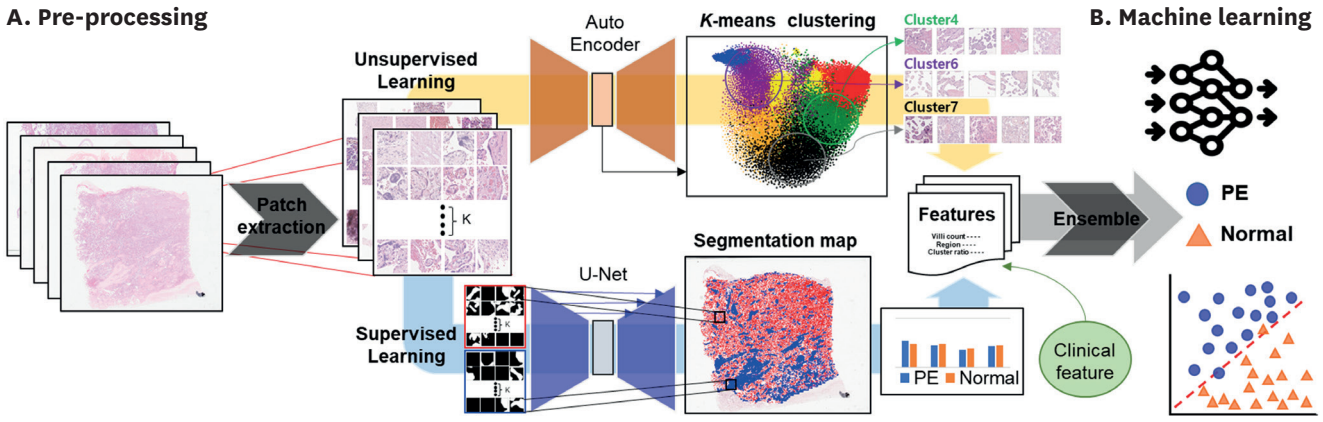

Fig. 3. Overview of the proposed method.
PE = preeclampsia.

the patches were classified into seven clusters. In **Fig. 1** only eight patches within each cluster were randomly selected and visualized. **Fig. 2** shows the cluster distributions of both the internal and external datasets. In this step, cluster distribution was used to obtain seven cluster ratios ('cluster1,' 'cluster2,' 'cluster3,' 'cluster4,' 'cluster5,' 'cluster6,' 'cluster7') to be used as features for each pWSI.

In the second step, villi were quantified to compare differences in the ratio of terminal villi to intermediate villi observed in PE and normal pWSI. The mean dice scores for each segmentation model were 0.802 (95% confidence interval [CI], 0.745–0.859) and 0.570 (95% CI, 0.270–0.684), respectively, for the test set. During this step, six features were extracted, including the total tissue region ('foreground region'), the region and number of terminal villi ('terminal villi count') and the region of intermediate villi ('terminal villi region,' 'intermediate villi region'), as well as the ratio of the corresponding villus area to the total pWSI ('terminal villi region ratio,' 'intermediate villi region ratio').

In the third step, the model was trained and validated by adding clinical features to the nine features selected from the features derived from the previous two steps (deep learning features). Within the model ensemble, the performance of the model combining deep learning features with clinical features exhibited a higher AUPRC value than the model that relied solely on clinical features (**Tables 2** and **3**).

Table 2 shows that in the internal validation set, the model ensemble combining deep learning features with clinical features exhibited the highest AUPRC value of 0.771 (95% CI, 0.752–0.790) and AUROC of 0.744 (95% CI, 0.724–0.764). In contrast, the model ensemble consisting solely of clinical features displayed a relatively lower AUPRC performance of 0.713 (95% CI, 0.694–0.732) and AUROC performance of 0.647 (95% CI, 0.625–0.669).

The external validation set exhibited a similar pattern to the internal validation set results as displayed in **Table 3**. The AUPRC value for the model ensemble combining deep learning features with clinical features was 0.725 (95% CI, 0.720–0.730), while the AUPRC value for the model relying solely on clinical features was 0.668 (95% CI, 0.665–0.671). The difference in AUROC values between the two models was statistically significant.

https://doi.org/10.3346/jkms.2024.39.e271

Table 2. Performance metrics for deep learning features combined with clinical features and clinical features alone using the internal dataset

Machine learning	AUPRC (95% CI)	AUROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Deep learning feature + clinical feature							
Ensemble	0.771**** (0.752–0.790)	0.744**** (0.724–0.764)	0.761 (0.746–0.776)	0.773 (0.739–0.807)	0.748 (0.711–0.785)	0.653 (0.601–0.705)	0.801 (0.777–0.825)
RF	0.750 (0.731–0.770)	0.738 (0.718–0.757)	0.752 (0.739–0.765)	0.743 (0.704–0.783)	0.761 (0.728–0.794)	0.649 (0.596–0.703)	0.785 (0.759–0.811)
SVM	0.750 (0.730–0.769)	0.720 (0.698–0.741)	0.748 (0.732–0.763)	0.734 (0.701–0.768)	0.761 (0.726–0.797)	0.635 (0.580–0.690)	0.767 (0.744–0.790)
GBM	0.733 (0.714–0.753)	0.721 (0.702–0.740)	0.747 (0.733–0.761)	0.762 (0.731–0.794)	0.732 (0.701–0.764)	0.663 (0.617–0.709)	0.782 (0.758–0.805)
LR	0.751 (0.733–0.769)	0.690 (0.670–0.711)	0.730 (0.716–0.744)	0.640 (0.606–0.674)	0.820 (0.787–0.853)	0.579 (0.513–0.644)	0.713 (0.695–0.732)
CatBoost	0.743 (0.722–0.763)	0.731 (0.711–0.750)	0.756 (0.741–0.770)	0.760 (0.725–0.787)	0.751 (0.716–0.787)	0.661 (0.611–0.712)	0.792 (0.768–0.817)
Clinical feature							
Ensemble	0.713 (0.694–0.732)	0.647 (0.625–0.669)	0.712 (0.697–0.727)	0.556 (0.523–0.589)	0.868 (0.845–0.891)	0.528 (0.457–0.599)	0.672 (0.656–0.688)
RF	0.609 (0.590–0.628)	0.528 (0.504–0.552)	0.643 (0.629–0.657)	0.483 (0.437–0.529)	0.802 (0.763–0.841)	0.523 (0.460–0.586)	0.633 (0.613–0.653)
SVM	0.719 (0.698–0.739)	0.658 (0.637–0.679)	0.719 (0.703–0.735)	0.611 (0.574–0.648)	0.828 (0.791–0.864)	0.509 (0.439–0.580)	0.702 (0.681–0.723)
GBM	0.624 (0.605–0.642)	0.536 (0.513–0.559)	0.639 (0.625–0.653)	0.500 (0.447–0.553)	0.778 (0.726–0.829)	0.445 (0.379–0.510)	0.654 (0.628–0.679)
LR	0.721 (0.703–0.740)	0.661 (0.642–0.680)	0.713 (0.699–0.727)	0.604 (0.568–0.641)	0.822 (0.787–0.857)	0.537 (0.470–0.604)	0.698 (0.678–0.719)
CatBoost	0.663 (0.643–0.683)	0.576 (0.553–0.599)	0.674 (0.660–0.688)	0.492 (0.453–0.531)	0.856 (0.824–0.887)	0.442 (0.372–0.512)	0.643 (0.626–0.659)

AUPRC = area under the precision-recall curve, AUROC = area under the receiver operating characteristic, PPV = positive predictive value, NPV = negative predictive value, CI = confidence interval, RF = random forest, SVM = support vector machine, GBM = gradient boost machine, LR = logistic regression.
****$P < 0.0001$ (P value with ensemble model AUPRC and AUROC of clinical features).

Table 3. Performance metrics for deep learning features combined with clinical features and clinical features alone using the external dataset

Machine learning	AUPRC (95% CI)	AUROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Deep learning feature + clinical feature							
Ensemble	0.725**** (0.720–0.730)	0.717**** (0.711–0.723)	0.692 (0.688–0.696)	0.616 (0.596–0.636)	0.767 (0.748–0.786)	0.737 (0.726–0.748)	0.672 (0.664–0.680)
RF	0.667 (0.659–0.674)	0.664 (0.655–0.672)	0.647 (0.641–0.653)	0.639 (0.605–0.673)	0.655 (0.622–0.688)	0.668 (0.655–0.680)	0.670 (0.652–0.687)
SVM	0.735 (0.731–0.739)	0.722 (0.719–0.725)	0.704 (0.701–0.708)	0.592 (0.575–0.609)	0.817 (0.802–0.831)	0.774 (0.763–0.785)	0.670 (0.664–0.677)
GBM	0.649 (0.641–0.658)	0.631 (0.623–0.640)	0.636 (0.629–0.643)	0.622 (0.581–0.662)	0.650 (0.609–0.690)	0.673 (0.655–0.690)	0.659 (0.643–0.674)
LR	0.703 (0.700–0.706)	0.715 (0.712–0.718)	0.694 (0.691–0.698)	0.619 (0.604–0.634)	0.770 (0.753–0.787)	0.737 (0.727–0.747)	0.672 (0.666–0.678)
CatBoost	0.678 (0.670–0.685)	0.671 (0.663–0.679)	0.660 (0.654–0.666)	0.619 (0.588–0.650)	0.701 (0.671–0.732)	0.695 (0.681–0.710)	0.663 (0.651–0.676)
Clinical feature							
Ensemble	0.668 (0.665–0.671)	0.695 (0.693–0.697)	0.690 (0.687–0.693)	0.700 (0.683–0.717)	0.680 (0.664–0.696)	0.691 (0.685–0.697)	0.701 (0.693–0.709)
RF	0.618 (0.612–0.624)	0.667 (0.661–0.672)	0.671 (0.668–0.675)	0.728 (0.706–0.751)	0.614 (0.592–0.637)	0.661 (0.653–0.669)	0.708 (0.696–0.720)
SVM	0.655 (0.653–0.658)	0.703 (0.701–0.705)	0.693 (0.691–0.696)	0.725 (0.708–0.741)	0.661 (0.646–0.677)	0.685 (0.680–0.691)	0.713 (0.705–0.721)
GBM	0.623 (0.616–0.631)	0.633 (0.625–0.640)	0.642 (0.636–0.648)	0.648 (0.619–0.678)	0.635 (0.607–0.663)	0.652 (0.641–0.662)	0.659 (0.647–0.671)
LR	0.644 (0.643–0.646)	0.666 (0.664–0.668)	0.675 (0.670–0.679)	0.624 (0.608–0.641)	0.725 (0.708–0.742)	0.699 (0.693–0.706)	0.662 (0.656–0.668)
CatBoost	0.644 (0.660–0.667)	0.687 (0.664–0.690)	0.687 (0.683–0.692)	0.684 (0.667–0.702)	0.691 (0.675–0.706)	0.693 (0.686–0.699)	0.693 (0.684–0.701)

AUPRC = area under the precision-recall curve, AUROC = area under the receiver operating characteristic, PPV = positive predictive value, NPV = negative predictive value, CI = confidence interval, RF = random forest, SVM = support vector machine, GBM = gradient boost machine, LR = logistic regression.
****$P < 0.0001$ (P value with ensemble model AUPRC and AUROC of clinical features).

The results of comparing pregnancy outcomes based on the presence or absence of PE, as confirmed by the predictive model, are presented in **Supplementary Table 3**. According to the results, the group predicted to diagnosed as PE delivered at an earlier gestational age and gave birth to infants with lower birth weights.

We conducted an experiment to compare AUPRC and AUROC by the number of features in the ensemble model to identify the appropriate number of features through feature importance (**Supplementary Fig. 1**). As a result, the most robust AUPRC and AUROC were obtained when using nine features (**Supplementary Fig. 2**). **Supplementary Table 3** presents the clinical outcomes between the groups diagnosed with PE and those not diagnosed with PE using the prediction model. It was observed that the group predicted to have PE delivered at an earlier gestational age and had lower birth weights of neonataes.

https://doi.org/10.3346/jkms.2024.39.e271

DISCUSSION

Utilizing an ensemble modeling approach, we successfully developed a model to identify PE placentas. This model demonstrated AUPRC value 0.771 (95% CI, 0.752–0.790), with a sensitivity of 77.3% and a specificity of 74.8%. In contrast, the clinical features model exhibited a lower AUPRC value of 0.713 (95% CI, 0.694–0.732), with a sensitivity of 55.6% and specificity of 86.8%. Furthermore, external validation using a BMC-derived patient dataset confirmed the robust discrimination ability of the model, with an AUPRC value of 0.725 (95% CI, 0.720–0.730).

PE is characterized by distinctive placental findings, including vascular lesions such as atherosis and infarction and non-vascular lesions such as hyperplasia. Additionally, inflammatory lesions such as umbilical vasculitis, chorionic plate vasculitis, acute chorioamnionitis, and chronic villitis are commonly observed in preeclamptic placentas.[7,22] These pathological changes in the placenta provide valuable insights into the underlying mechanisms of this pregnancy complication.[6,23] For these reasons, many obstetricians, gynecologists, and pathologists have conducted extensive research on the placenta, aiming to unravel the pathophysiology of PE. However, even when certain distinctive placental findings (such as inflammatory and vascular lesions) accompany PE, confirming the presence of PE based solely on placental pathology using conventional methods has proven to be a challenging task.

The **Tables 2** and **3** reveal that the AUPRC of the ensemble model, trained with clinical features, is 71.3% in the internal validation set and 66.8% in the external validation set. These results suggest the relevance of the actual size of the placenta in differentiating between PE and normal groups. As depicted in **Table 2** for the internal validation set, the ensemble model exhibited improvements of 5.8% in AUPRC, 9.7% in AUROC, 4.9% in accuracy, 21.7% in sensitivity, 12.5% in PPV, and 12.9% in NPV. Similarly, in **Table 3** for the external validation set, the ensemble model displayed improvements in AUPRC by 5.7%, AUROC by 2.2%, accuracy by 0.2%, specificity by 8.7%, and PPV by 4.6%. Hence, it can be inferred that the extracted deep learning features contribute significantly to the prediction of PE.

We selected a model based on a high AUPRC for the ensemble of models. Various metrics are commonly employed to assess model performance in binary classification, metrics other than AUPRC and AUROC heavily depend on the chosen thresholds for dichotomizing prediction probabilities.[24,25] In contrast, AUPRC and AUROC operate independently of specific thresholds, offering a comprehensive reflection of overall model performance, thereby elevating their significance as performance indicators. Sensitivity (recall) holds paramount importance in medical diagnostics, particularly in the early detection of diseases. AUPRC places a greater emphasis on recall, rendering it a more suitable assessment metric in the medical domain.

AI-based placental analyses have several clinical implications. An AI-Clinical Decision Support Systems (CDSS) improves assessment and management, enhancing convenience and sensitivity/specificity. Digital pathology, ideal for AI-CDSS, enhances clinical assessment. ML-derived features in WSIs surpass standard pathology methods. Their implications on neonatal/pregnancy outcomes require further study. PE links to long-term maternal cardiovascular outcomes. Placental implantation's role in PE pathology suggests AI-based analysis could aid outcome prediction.

Recent technological advancements have led to the development of techniques that utilize histopathological image analysis and ML for data analysis. A key feature contributing to the remarkable success of deep learning algorithms in the field of digital pathology is their end-to-end learning approach. Unlike previous methods that relied on incorporating pre-existing knowledge into the algorithm design, these adaptive algorithms have the ability to learn directly from datasets and predict labels from pixel values.

While there is an international consensus established based on the Amsterdam criteria,[26] it is important to acknowledge that the diagnosis still presents many challenges. In contrast to the typical cancer diagnoses, placental pathology presents a unique aspect where the integration of clinical manifestations in obstetrics is of utmost importance. Insufficient understanding of obstetric clinical situation often leads to challenges in making accurate placental diagnoses.[27]

In addition, especially in the pathological diagnosis of preeclamptic placentas, quantitative parameters often assume a critical role, yet they are beset by indistinct and ill-defined cutoffs. For example, there exists no established normative threshold for the lumen size of spiral arteries at different gestational stages, rendering it challenging to define, and diagnose the point of failure of the spiral artery transformation. Additionally, the assessment of complete muscle loss in spiral arteries[28] relying solely on hematoxylin and eosin staining is sometimes ambiguous if alpha smooth muscle actin immunohistochemical staining cannot not be performed. Moreover, in the diagnosis of hypoplastic villi, there remains a conspicuous absence of precise quantitative criteria determining the success of spiral artery transformation based on the quantification of spiral artery lumen size poses difficulties. This underscores the complexity of manual, human-driven reading and interpretation in placental pathology.

However, as vascular pathologies such as acute atherosis, observed in preeclamptic placenta, continue to be crucial indicators for predicting long-term maternal outcomes,[29] the precise diagnosis of preeclamptic placenta remains critical. In this regard, AI-driven digital pathology computational analysis is poised to play a significant role.

To our knowledge, the current study is the first to identify the characteristics of pre-eclamptic placentas using AI-based placental analysis. The results of this study demonstrate the potential to diagnose PE using a hybrid-based approach combining unsupervised and supervised learning rather than relying solely on traditional histopathological interpretation. This approach effectively harnesses the strengths of unsupervised and supervised learning to partition histopathological images into clusters, thereby unlocking new possibilities for analyzing such images across diverse scenarios. It also offers a novel perspective for tackling topics necessitating intricate pathological image analysis.

Our study had certain limitations since we were unable to compare the performance of our automated diagnostic system with that of actual pathologists when examining the placental pathology of PE. In future work, we aim to analyze and compare the diagnostic performance and correlation between pathologists and AI using only WSI. Additionally, we plan to conduct further studies to validate the practicality and accuracy of the AI model, not only on the dataset already collected but also on prospectively collected real clinical case. Secondly, the dice score for intermediate villi segmentation was 57%, indicative of a performance level that did not reach a notably high standard. Due to the considerable time and financial resources required for meticulous annotation, obtaining a sufficient number of labeled datasets for villi

segmentation proved challenging. Consequently, we posit that the constrained performance can be attributed to the limitations imposed by learning with a restricted labeled dataset. We anticipate that with an increased number of labels in future iterations, a higher dice score can be achieved. Thirdly, we did not consider quantitative measures to evaluate clustering quality. Relying solely on qualitative assessments based on expert opinions, we aim to introduce quantitative measures such as silhouette score and Davies-Bouldin index[30] in future studies for a more objective approach.[31]

The computational pathology model we developed in this study demonstrated a strong ability to predict PE.[32] Computational pathology has the potential to improve the identification of PE placentas.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Characteristics and pregnancy outcomes of the study population (Boramae Metropolitan Hospital)

Supplementary Table 2
Characteristics and pregnancy outcomes of the study population

Supplementary Table 3
Pregnancy outcomes based on the results of the predictive model

Supplementary Fig. 1
Feature importance of random forest using training set.

Supplementary Fig. 2
Comparison of AUPRC and AUROC for ensemble model based on the number of features.

REFERENCES

1. 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):1-7. PUBMED | CROSSREF
2. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller AB, Daniels J, et al. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Health* 2014;2(6):e323-33. PUBMED | CROSSREF
3. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. *BMJ* 2013;347:f6564. PUBMED | CROSSREF
4. Chappell LC, Cluver CA, Kingdom J, Tong S. Pre-eclampsia. *Lancet* 2021;398(10297):341-54. PUBMED | CROSSREF
5. Kulkarni VG, Sunilkumar KB, Nagaraj TS, Uddin Z, Ahmed I, Hwang K, et al. Maternal and fetal vascular lesions of malperfusion in the placentas associated with fetal and neonatal death: results of a prospective observational study. *Am J Obstet Gynecol* 2021;225(6):660.e1-12. PUBMED | CROSSREF
6. Roland CS, Hu J, Ren CE, Chen H, Li J, Varvoutis MS, et al. Morphological changes of placental syncytium and their implications for the pathogenesis of preeclampsia. *Cell Mol Life Sci* 2016;73(2):365-76. PUBMED | CROSSREF
7. Assibey-Mensah V, Parks WT, Gernand AD, Catov JM. Race and risk of maternal vascular malperfusion lesions in the placenta. *Placenta* 2018;69:102-8. PUBMED | CROSSREF
8. Li Z, Zhang J, Tan T, Teng X, Sun X, Zhao H, et al. Deep learning methods for lung cancer segmentation in whole-slide histopathology images-the ACDC@LungHP challenge 2019. *IEEE J Biomed Health Inform* 2021;25(2):429-40. PUBMED | CROSSREF

9. Khened M, Kori A, Rajkumar H, Krishnamurthi G, Srinivasan B. A generalized deep learning framework for whole-slide image segmentation and analysis. *Sci Rep* 2021;11(1):11579. PUBMED | CROSSREF
10. Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology image analysis using segmentation deep learning algorithms. *Am J Pathol* 2019;189(9):1686-98. PUBMED | CROSSREF
11. Kanavati F, Toyokawa G, Momosaki S, Rambeau M, Kozuma Y, Shoji F, et al. Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci Rep* 2020;10(1):9297. PUBMED | CROSSREF
12. Muhammad H, Sigel CS, Campanella G, Boerner T, Pak LM, Buttner S, et al. Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder. Proceedings of Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference; 2019 October 13–17; Shenzhen, China. Berlin, Germany: Springer International Publishing; 2019.
13. Bank D, Koenigstein N, Giryes R. Autoencoders. In: Rokach L, Maimon O, Shmueli E, editors. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*. Berlin, Germany: Springer Nature; 2023, 353-74.
14. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans Pattern Anal Mach Intell* 2002;24(7):881-92. CROSSREF
15. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Proceedings of Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference; 2015 October 5–9; Munich, Germany. Berlin, Germany: Springer; 2015.
16. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016 June 27–30; Las Vegas, NV, USA. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2016.
17. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. Proceedings of 2009 IEEE Computer Vision and Pattern Recognition; 2009 June 20–25; Miami, FL, USA. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE); 2009.
18. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2021;2(3):160. PUBMED | CROSSREF
19. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Proceedings of 32nd International Conference on Neural Information Processing Systems; 2018 December 3–8; Montreal, Canada. Red Hook, NY, USA: Curran Associates Inc.; 2018.
20. Boyd K, Eng KH, Page CD. Area under the precision-recall curve: point estimates and confidence intervals. Proceedings of Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013; Prague, Czech Republic; 2013 September 23–27. Berlin, Germany: Springer; 2013.
21. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowl Manag Process* 2015;5(2):1-11. CROSSREF
22. Nelson DB, Ziadie MS, McIntire DD, Rogers BB, Leveno KJ. Placental pathology suggesting that preeclampsia is more than one disease. *Am J Obstet Gynecol* 2014;210(1):66.e1-7. PUBMED | CROSSREF
23. Hauspurg A, Redman EK, Assibey-Mensah V, Tony Parks W, Jeyabalan A, Roberts JM, et al. Placental findings in non-hypertensive term pregnancies and association with future adverse pregnancy outcomes: a cohort study. *Placenta* 2018;74:14-9. PUBMED | CROSSREF
24. Wang Q, Guo A. An efficient variance estimator of AUC and its applications to binary classification. *Stat Med* 2020;39(28):4281-300. PUBMED | CROSSREF
25. Tharwat A. Classification assessment methods. *Appl Comput Inform* 2021;17(1):168-92. CROSSREF
26. Khong TY, Mooney EE, Ariel I, Balmus NC, Boyd TK, Brundler MA, et al. Sampling and definitions of placental lesions: Amsterdam Placental Workshop Group consensus statement. *Arch Pathol Lab Med* 2016;140(7):698-713. PUBMED | CROSSREF
27. Redline RW. Placental pathology: a systematic approach with clinical correlations. *Placenta* 2008;29 Suppl A:S86-91. PUBMED | CROSSREF
28. Burton GJ, Woods AW, Jauniaux E, Kingdom JC. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* 2009;30(6):473-82. PUBMED | CROSSREF
29. Pitz Jacobsen D, Fjeldstad HE, Johnsen GM, Fosheim IK, Moe K, Alnæs-Katjavivi P, et al. Acute atherosis lesions at the fetal-maternal border: current knowledge and implications for maternal cardiovascular health. *Front Immunol* 2021;12:791606. PUBMED | CROSSREF
30. Ashari IF, Dwi Nugroho E, Baraku R, Novri Yanda I, Liwardana R. Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta. *J Appl Inform Comput* 2023;7(1):95-103. CROSSREF

31. Lee Y, Kim SY. Potential applications of ChatGPT in obstetrics and gynecology in Korea: a review article. *Obstet Gynecol Sci* 2024;67(2):153-9. PUBMED | CROSSREF
32. Ahn TG, Hwang JY. Preeclampsia and aspirin. *Obstet Gynecol Sci* 2023;66(3):120-32. PUBMED | CROSSREF