

Interpretable learning predicts miscarriage using pre-pregnancy serum metabolites

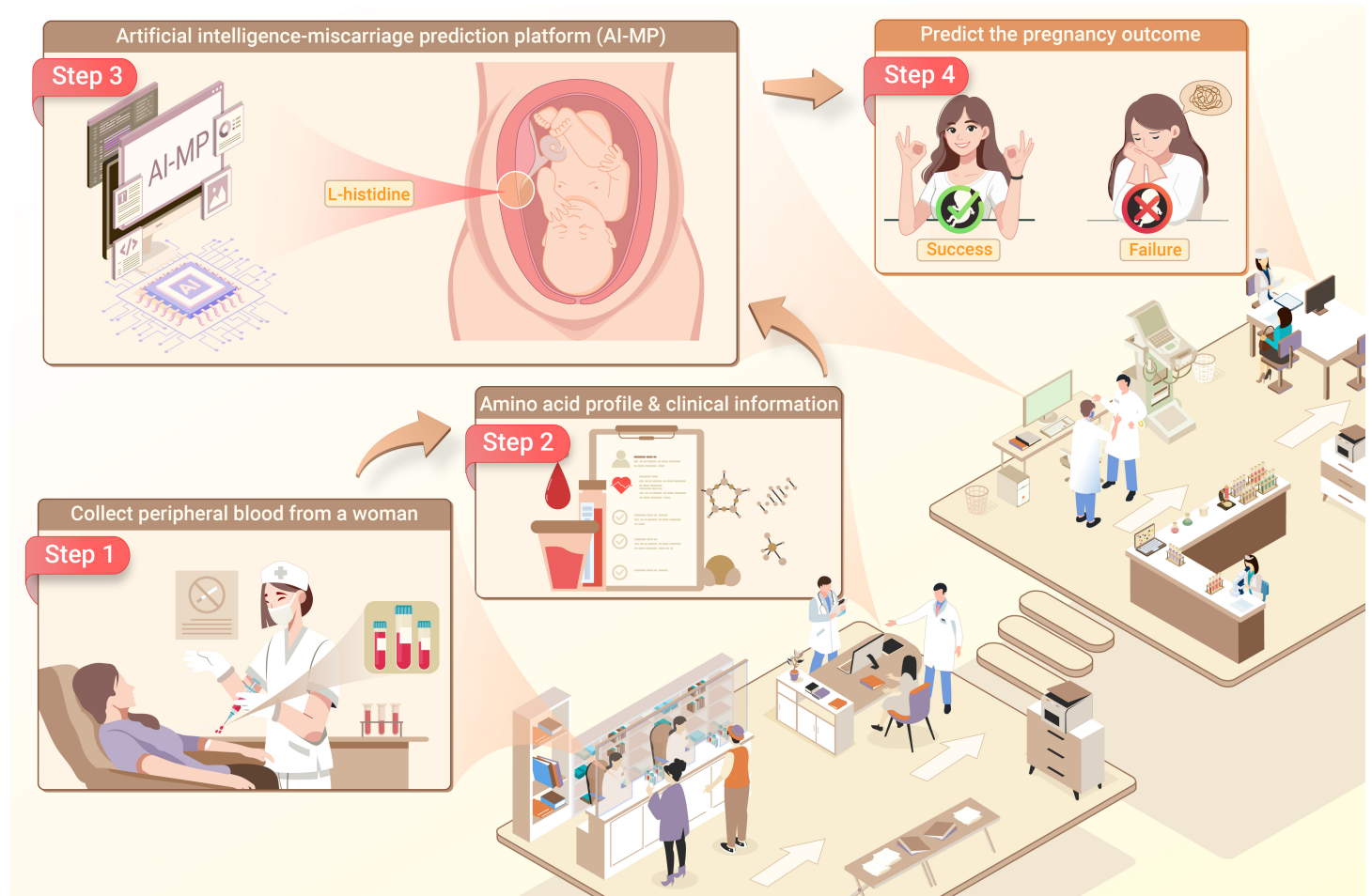
Linchen Tang,^{1,2,6} Zhilong Wang,^{3,4,6} Xianghong Xu,^{2,6} Yawei Yuan,^{5,6} Fan Hao,² An Chen,^{3,4} Shengnan Wu,² Wenxuan Li,² Yanqiang Han,^{3,4} Jiaxue Sun,² Kehao Tao,^{3,4} Zeying Liu,² Zhiyun Wei,^{2,*} Jinjin Li,^{3,4,*} and Liping Jin^{1,2,*}

*Correspondence: jinlp01@163.com (L.J.); lijijin@sztu.edu.cn (J.L.); zhiyun_wei@163.com (Z.W.)

Received: March 31, 2024; Accepted: August 28, 2024; Published Online: September 5, 2024; <https://doi.org/10.59717/j.xinn-med.2024.100093>

© 2024 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Amino acids, especially the L-histidine, play a pivotal role in pregnancy.
- Pre-pregnancy serum amino acid profiles can predict miscarriage.
- An interactive platform is established to predict miscarriage prior to conception.

Interpretable learning predicts miscarriage using pre-pregnancy serum metabolites

Linchen Tang,^{1,2,6} Zhilong Wang,^{3,4,6} Xianghong Xu,^{2,6} Yawei Yuan,^{5,6} Fan Hao,² An Chen,^{3,4} Shengnan Wu,² Wenxuan Li,² Yanqiang Han,^{3,4} Jiaxue Sun,² Kehao Tao,^{3,4} Zeying Liu,² Zhiyun Wei,^{2,*} Jinjin Li,^{3,4,*} and Liping Jin^{1,2,*}

¹Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Hospital of Obstetrics and Gynecology, Shanghai Medical School, Fudan University, Shanghai 200090, China

²Shanghai Key Laboratory of Maternal Fetal Medicine, Clinical and Translational Research Center, Department of Biobank, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai 201204, China

³National Key Laboratory of Advanced Micro and Nano Manufacture Technology, Shanghai Jiao Tong University, Shanghai 200240, China

⁴Department of Micro/Nano Electronics, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

⁵Department of Anesthesiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

⁶These authors contribute equally

*Correspondence: jinlp01@163.com (L.J.); lijijin@sytu.edu.cn (J.L.); zhiyun_wei@163.com (Z.W.)

Received: March 31, 2024; Accepted: August 28, 2024; Published Online: September 5, 2024; <https://doi.org/10.59717/j.inn-med.2024.100093>

© 2024 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Tang L., Wang Z., Xu X., et al., (2024). Interpretable learning predicts miscarriage using pre-pregnancy serum metabolites. *The Innovation Medicine* 2(4): 100093.

In preconception counseling, there has been a lack of quantitative approaches to predict the outcome of an upcoming pregnancy, which would greatly benefit women and society. By applying state-of-the-art artificial intelligence algorithms to clinical and metabolome data from 481 women, we have, for the first time, proposed a pre-pregnancy classifier that predicts miscarriage with a high precision rate of 87%. Our embedded feature engineering revealed the critical impact of serum histidine level, further supported by its elevation in recurrent spontaneous miscarriage (RSM). Mechanistically, elevated histidine level, combined with compromised diamine oxidase (DAO) expression, led to a fatal accumulation of histamine at the maternal-fetal interface (157 specimens from 113 donors). Additionally, a high-histidine diet induced significant embryo loss in mice without causing malabsorption of other amino acids. This pilot study shows promise in predicting pregnancy outcomes prior to conception, opening an important window for early warning that is particularly meaningful given the global decline in birth rates.

INTRODUCTION

Pregnancy establishes the starting of a life. Current clinical guidelines for pregnancy apply ultrasound and human chorionic gonadotrophin (hCG) test as routine approaches during gestation to provide information of fetal development.^{1,2} Such monitoring tests often reveal irreversible defects, and are too late to avoid damages to women. Developing early predictor or even pre-conception classifier for pregnancy outcome is therefore practically urgent, yet has been barely investigated or discussed thus far.

Miscarriage, one of the most devastating consequences of pregnancy, occurs with increasing frequency in fertile women in most countries, causing approximately 23 million pregnancy losses every year worldwide.³ Women who have experienced two or more consecutive spontaneous abortions in their first trimesters of pregnancy, diagnosed as recurrent spontaneous miscarriage (RSM), are facing 2.21 times more risks of miscarriage than others.⁴ Current pathophysiology can only explain less than half of RSM with multiple hypotheses combined, leaving most patients without any effective prevention.^{5,6} It has been suggested that early extra care could improve live birth rates for RSM.^{5,7-9} Therefore, it would be of great benefit to be able to identify miscarriage risk at the earliest possible time.¹⁰

Here, we established an artificial intelligence miscarriage prediction (AI-MP) platform to predict miscarriage via ensemble learning. The proposed study provides the first proof of principle for predictivity of pre-pregnancy serum metabolites, a step forward comparing to pioneering post-conception works, by advancing a precise prediction up to six months. The established AI model to predict miscarriage achieves an unprecedented and very important predictive target, which suggests a likely underlying mechanism of histidine on miscarriage, with supporting evidences from 481 blood samples, 157 specimens of 3 maternal-fetal tissues, and one mouse model. The proposed AI-MP could predict miscarriage outcome before pregnancy within milliseconds, with a predictive precision of 87%. The ensemble learning in AI-MP

revealed the critical impact of serum histidine level on miscarriage, which was further supported by its significant elevation in patients with RSM than normal subjects. Our study discovered a groundbreaking value of ignored human pre-pregnancy metabolome on early prediction, and represented a broadly applicable strategy for AI-auxiliary precision medicine.

MATERIALS AND METHODS

Human sample collection

Human pre-pregnancy peripheral blood was obtained from 240 women with clinical normal pregnancy history (the NOR cohort) and 241 women with recurrent miscarriage history for unexplained reasons (the RSM cohort), recruited from September, 2020 to March, 2021. Serum was separated by centrifugation at 1,000 rcf for 15 min at 4°C. Within the RSM cohort, full-term tracking was accomplished from 154 women, whose data were used for establishing AI-MP model. Demographics of donors were summarized in Table 1 and Table S1.

The third cohort included 100 pregnant women recruited from January, 2021 to July, 2021 (the pregnant cohort): 57 normal subjects who terminated their pregnancies for nonmedical reasons, and 43 RSM patients who experienced spontaneous miscarriages for unexplained reasons again. Their pregnancies were confirmed by ultrasound and blood HCG tests. In this cohort, three types of specimens were collected: peripheral blood (n=40 vs 20), decidual tissues (n=45 vs 41), and villus tissues (n=29 vs 29) during gestational weeks 8-12.

The fourth cohort included 13 normal subjects recruited in July, 2021, who underwent hysteroscopic surgery for benign reasons unrelated to endometrial dysfunction, such as mild intrauterine adhesion, endometrial polyps, and hysteromyoma (the endometrium cohort). Endometrium samples were collected during the proliferative phase in this cohort. Demographics of donors in the pregnant cohort and the endometrium cohort were summarized in Table 2.

All published potential clinical predictors were considered as modelling input, including maternal age, body mass index (BMI), previous number of miscarriages (PNM), genetic factors, immunological factors, thrombophilia, uterine factors, and endocrine factors. In details, genetic factors included maternal/paternal chromosome abnormality (Boolean) and fetal chromosome abnormality in her previous miscarriage(s) (Boolean). Immunological factors included anti-phospholipid syndrome (Boolean), autoimmune disorders (Boolean), and immune cell dysfunction (Boolean). Thrombophilia included factor V Leiden mutation (Boolean), protein S/C deficiency (Boolean), and MTHFR mutation (Boolean). Uterine factors included congenital uterine malformation (Boolean), and acquired uterine malformation (Boolean). Endocrine factors included insulin resistance (Boolean), polycystic ovary syndrome (Boolean), and hypothyroidism (Boolean). Other factors (Boolean) included folate deficiency (Boolean), smoking (Boolean), and alcohol consumption (Boolean). The diagnostic criteria of above features followed clinical guidelines (the 2022 ESHRE guidelines¹¹ and the 2012 ASRM guidelines about RSM¹²).

Table 1. Demographics of the NOR and RSM cohorts

Characteristics	NOR cohort ^a	RSM cohort ^b	P value ^c
Cohort size	240	241	
Age (years) (mean±s.e.m.)	30.14±0.24	30.84±0.20	0.07
BMI (mean±s.e.m.)	22.44±0.22	21.98±0.19	0.14
First trimester loss (mean±s.e.m.)	0.00±0.00	2.64±0.07	0.00
Prior live birth (mean±s.e.m.)	0.13±0.02	0.04±0.01	0.00

Notes: ^a The NOR cohort were nonpregnant women with clinical normal pregnancy history.

^b The RSM cohort were nonpregnant women with recurrent miscarriage history for unexplained reasons.

^c P-value was calculated by two-tailed unpaired student's t-test for normally distributed data (Age, BMI) or two-tailed Mann Whitney test for non-normally distributed data (First trimester loss, Prior live birth).

Table 2. Demographics of the pregnant and endometrium cohorts

Characteristics	NOR in prepregnancy ^a	NOR in pregnancy ^b	RSM in pregnancy ^c	P value ^{ab}	P value ^{bc}
Cohort size	13	57	43		
Age (years) (mean±s.e.m.)	32.38±0.87	30.51±0.57	32.23±0.88	0.15	0.20
First trimester loss (mean±s.e.m.)	0.00±0.00	0.00±0.00	3.00±0.21		0.000
Gestational age (day) (mean±s.e.m.)	0.00±0.00	49.95±0.85	50.49±0.69		0.40

Notes: ^a The NOR in prepregnancy were the endometrium cohort. All women underwent hysteroscopic surgery for benign reasons unrelated to endometrial dysfunction, such as mild intrauterine adhesion, endometrial polyps, and hysteromyoma.

^{b,c} The NOR and RSM in pregnancy were the pregnant cohort. The NOR subjects terminated their pregnancies for nonmedical reasons, and RSM subjects were abortion patients who had experienced more than two unexplained and consecutive spontaneous abortions. Their pregnancies were confirmed by ultrasound and blood HCG tests.

^{ab} P-value between NOR in prepregnancy and NOR in pregnancy was calculated by two-tailed unpaired student's t-test for normally distributed data (Age).

^{bc} P-value between NOR in pregnancy and RSM in pregnancy was calculated by two-tailed unpaired student's t-test for normally distributed data (Age) or two-tailed Mann Whitney test for non-normally distributed data (First trimester loss and Gestational age).

This study was approved by the Scientific and Ethical Committee of the Shanghai First Maternity and Infant Hospital affiliated with Tongji University. All study participants provided written informed consents.

Animal experiments

Barrier-raised SPF C57/BL6 mice were purchased from the Model Organism (Shanghai, China), and housed in the animal facility of Model Organism, Shanghai, China with a controlled environment (22–24 °C and 60–70% relative humidity) and a light/dark cycle (12 h light/12 h dark). All animal experiments were performed in accordance with National Institutes of Health published Guide for the Care and Use of Laboratory Animals, and were approved by the Institutional Animal Care and Use Committee of Shanghai Model Organisms. To verify the roles of histidine on pregnancy *in vivo*, a high histidine diet was applied for the natural pregnancy mouse model. Female mice of eight-weeks age (weight: 19–22 g; n = 24) were divided into two equal groups using a random number table by body weight, age and litter: standard diet (Ctrl) group (4.5 g L-histidine/kg diet) and high histidine diet (Treated) group (9 g L-histidine/kg diet).¹³ After two weeks of feeding with the corresponding diet, four mice from each group were sacrificed for collecting peripheral blood and uterus tissues for metabolites detection. Mild histidine supplementation in relatively short duration was adopted with the purpose of avoiding disturbance of other amino acids. The remaining female mice were mated with eight C57/BL6 male mice (ten weeks old) to establish pregnancy. In details, one male mouse was mated with two female mice (one from control group and the other from treated group) per cage. Copulation was confirmed by the observation of vaginal plugs on the following morning, which was designated as gestational day (GD) 0.5. Immediately, the mated female mice were transferred back to the cages with the corresponding diet, and housed until sacrificed on GD12.5. The average weight of fetus, litter size and absorption rate were recorded. Embryo absorption rate was calculated as follows: percentage fetal loss = $R / (R + V) \times 100$, where R represented the

number of hemorrhagic implantations (sites of fetal loss), and V represented the number of viable fetuses.

Quantification and statistical analysis

Data are presented as the mean ± s.d./s.e.m. indicated in each figure legend. Statistical differences were analyzed using GraphPad Prism version 8 (GraphPad software, USA). Two-sided Student's t-tests were performed for continuous data. If unpaired data did not have similar variances, Welch's t-test was performed. Differences were considered statistically significant if the P values were less than 0.05. Significance was indicated as follows: *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001. No significance was indicated as ns.

RESULTS

Establishment of AI model to predict upcoming miscarriage before conception

In this pilot study, we established two single-center prospective cohorts: the RSM cohort (n=241) with pre-pregnancy blood draws as well as up to 16 months follow-up, and the NOR cohort (n=240) including non-gestational blood samples from age-matched women without RSM history. In the RSM cohort, 87 women were excluded due to no conception within six months after blood sampling with (n=64) or without (n=15) persistent contraception, or failure of full-length follow-up (n=8). Among the remaining 154 donors, 41 achieved successful fertilization in 3.1 months (s.d. 0.21 months), extrapolated by their menstrual cycles, but had miscarriage within 7.34 weeks (s.d. 0.37 weeks) of gestational ages (Figure 1A). The other 113 RSM donors succeeded on delivery after taking similar time to fertilization (3.12±0.14 months, no significance; Figure 1A). Pre-pregnancy serum markers were explored in such cohort of high miscarriage risk.

A few leading studies, including ours, highlighted the metabolism dysregulation in miscarriage and RSM.^{14,15} We recently discovered rewiring, and

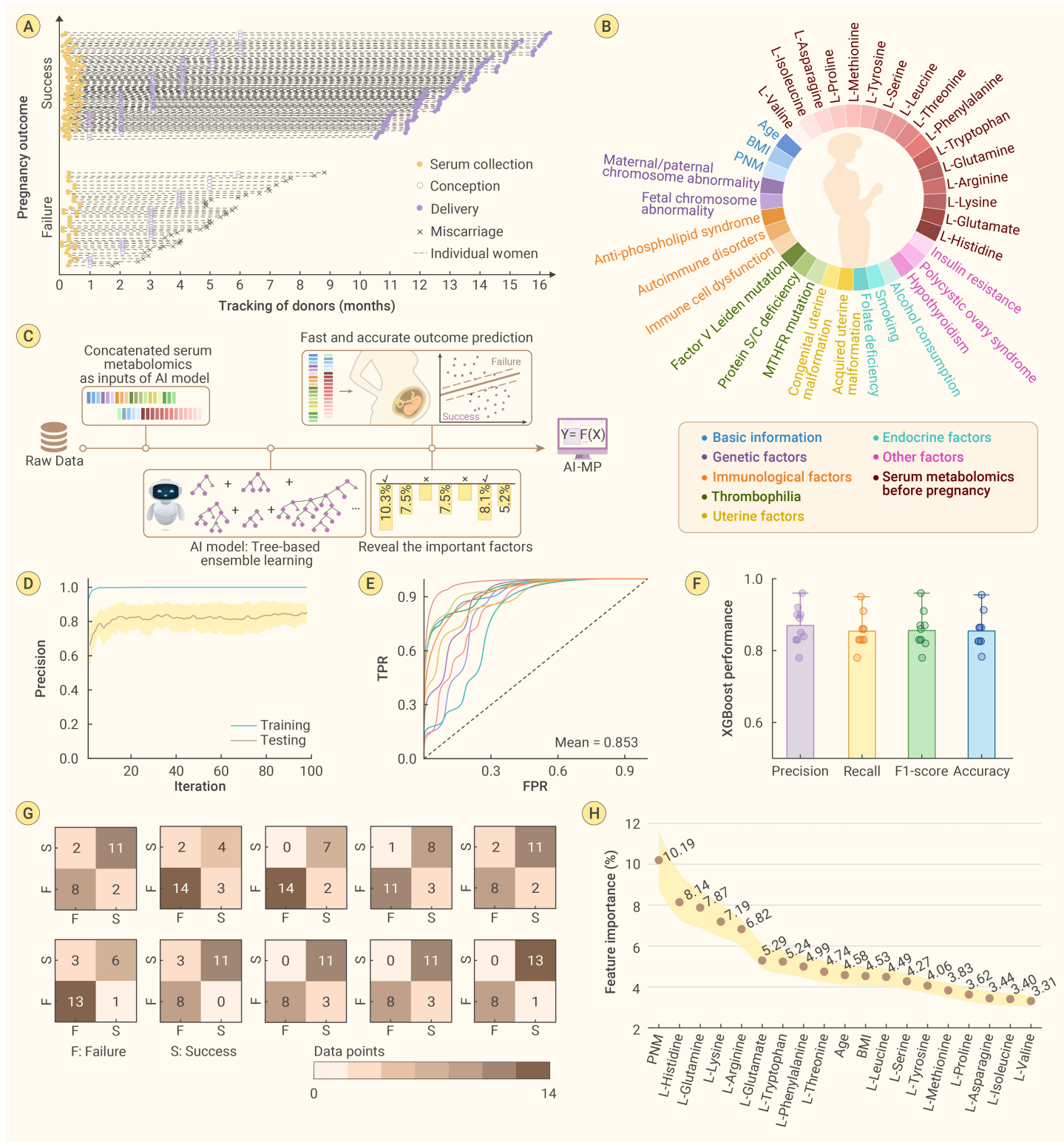


Figure 1. Establishment of AI model to predict miscarriage before pregnancy (A) Timeline of sample collection, conception (last menstrual period), and delivery or miscarriage. Tracking history of each donor was aligned based on her last menstrual period before the blood draw, denoted as zero on the x axis. (B) Scheme of AI input data. (C) Machine learning pipeline for miscarriage prediction. XGBoost was applied for training model, and the feature importance was obtained during the iterations. The AI-MP platform was then developed based on the well-trained model. (D) Precision on the training and testing data during the iterations, shown as means±s.d. (E) ROC curves on testing data of XGBoost model from 10-fold CVs, where the mean AUC was 0.853 (s.d. 0.061). (F) Four evaluation metrics of XGBoost model from 10-fold CVs, shown as means±s.d. (G) Confusion matrices on testing data from 10-fold CVs. (H) Importance of each feature to the prediction model, shown as means±s.d.

dysregulation in RSM, of glutamine metabolism during endometrial decidualization, implicating amino acids as important maternal influencers on pregnancy.¹⁶ Therefore, we investigated whether amino acid profile in pre-pregnancy serum could estimate the risk of miscarriage using targeted metabolomics, with specificity and accuracy guaranteed by in-parallel full-

coverage standards (Figure 1B). All published potential clinical predictors were also considered as modelling input, with the purpose of optimizing machine learning (ML) performance, including the basic information (maternal age, BMI, PNM), genetic factors (e.g., maternal/paternal chromosome abnormality), immunological factors (e.g., anti-phospholipid syndrome, APS),

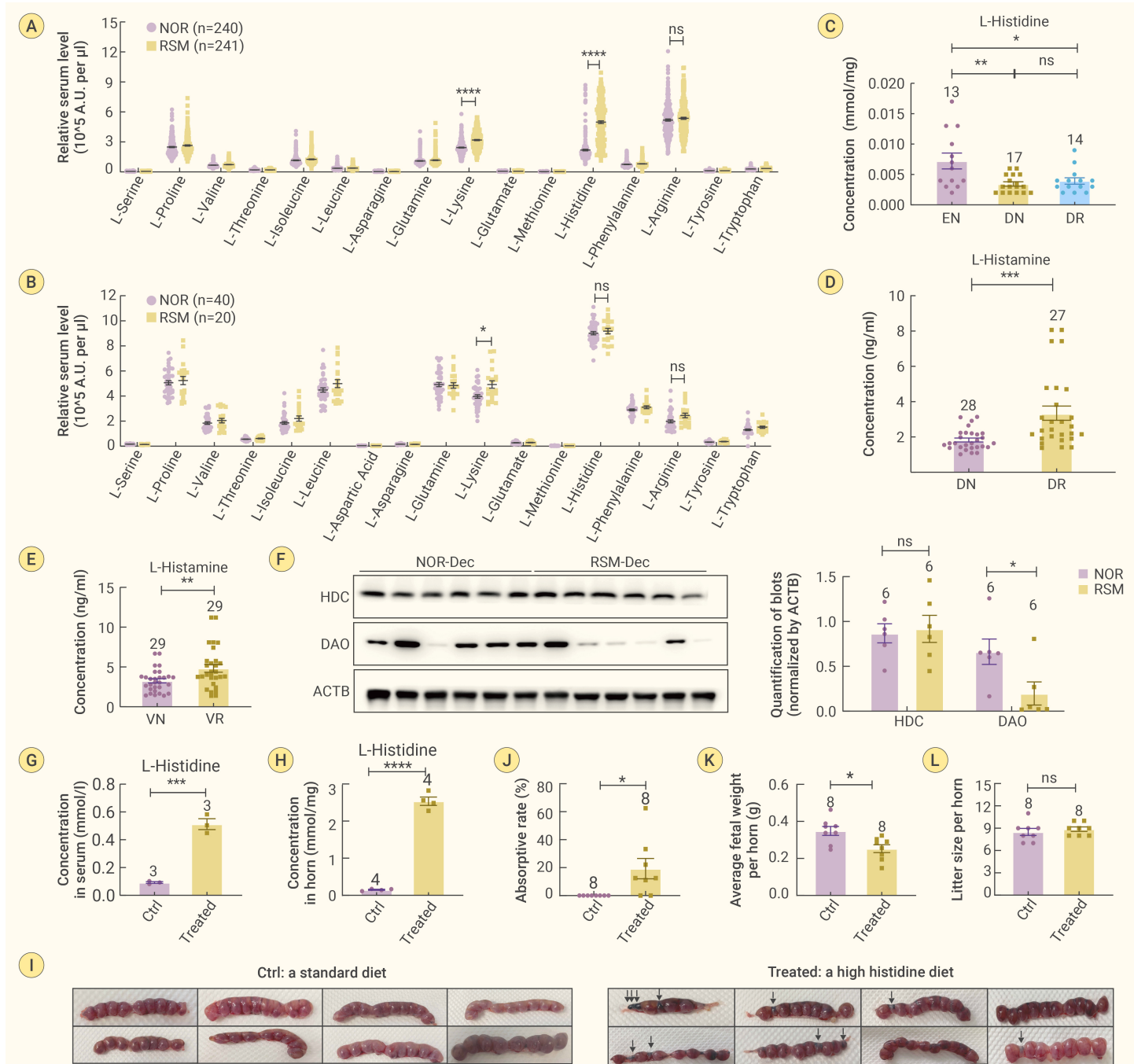


Figure 2. Hyperactive histidine-histamine flux contributed to fetal loss in both human and mice (A) Targeted metabolome comparison between RSM patients and normal subjects in pre-pregnancy serum. (B) Targeted metabolome comparison between RSM patients and normal subjects in in-pregnancy serum. (C) Comparison of tissue histidine concentration among endometrium from normal subjects (EN), decidua from normal subjects (DN), and decidua from RSM patients (DR). (D) Decidual histamine concentration was significantly higher in RSM patients. (E) Villous histamine concentration was significantly higher in RSM patients. (F) Expression of HDC and DAO enzymes in decidua from RSM patients or normal subjects. (G, H) Histidine concentration was significantly elevated in pre-pregnancy serum (G) and uterus (H) from mice with high histidine diet (Treated), comparing to the standard diet group (Ctrl). (I-L) Photos of uterine horns at GD12.5 from mice fed with two different diets, where absorption points were marked by black arrows (I). Embryo absorptive rate (J), average fetus weight per horn (K), and implantation sites per horn (L) were compared between diets. Number above each bar indicated sample size. Error bar represented mean $\pm$ s.e.m. EN, endometrium from normal subjects. DN, decidual tissues from NOR subjects. DR, decidual tissues from RSM patients. Dec, Decidua. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, non-significant; two-tailed t-test.

thrombophilia (e.g., factor V Leiden mutation), uterine factors (e.g., congenital uterine malformation), endocrine factors (e.g., insulin resistance) and other factors (e.g., smoking) (Figure 1B). To select the most predictive features, we attempted to construct the model using individual and a limited combination of features for comparative purposes. Different combinations of features with different numbers of features were chosen for establishing the predictive model, which was realized by the randomly sampling function in Python. The model was trained five times randomly for each case (from test 1 to 5). It is obvious that as the number of features increases, the model performance

(precision) generally improves. Finally, through iterations and optimizations, nineteen features were testified by ML to be the most predictive (Figure S1 and Table S2).

Given the instinctive imbalance of sample sizes in cohorts for nearly all clinical phenotypes, limiting performance of logistic prediction models with bias towards the larger group, we adopted SMOTE to realize the data augmentation and obtained a balanced data (named BalData) between successful and failed pregnancies.¹⁷⁻¹⁹ BalData maintained the data distribution after the SMOTE oversampling (Figure S2). ML was then applied to

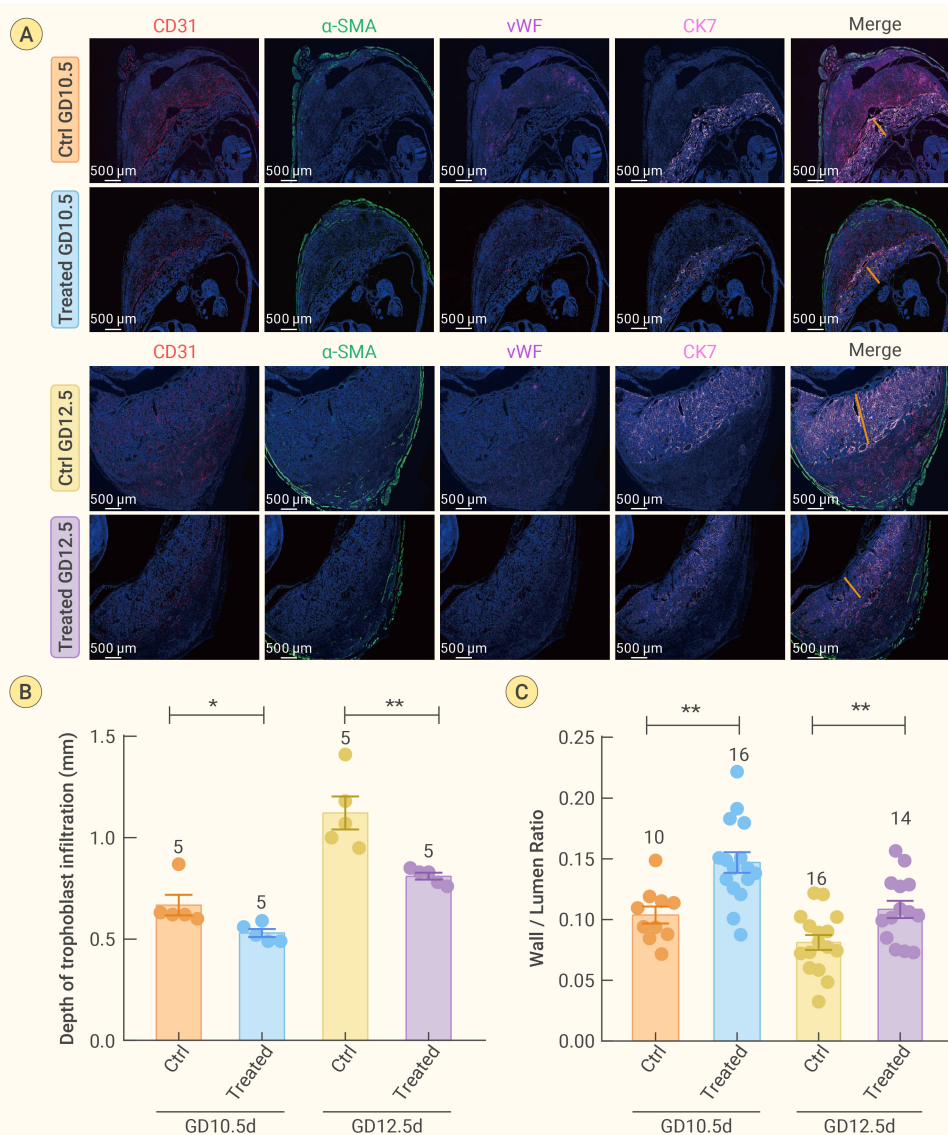


Figure 3. Uterine spiral artery remodeling and trophoblast invasion in mice under different diets (A) Representative photomicrographs depicting mIF labelings for angiogenesis markers CD31 (Red) and α-SMA (Green), endothelial marker vWF (Violet) and trophoblast infiltration marker CK7 (Pink) observed in Ctrl and Treated groups at GD10.5 (n=5) and GD12.5 (n=5). (B) The depth of CK7⁺ trophoblast infiltration into uterus was compared between diets at GD10.5 and GD12.5. (C) The assessment of spiral artery remodeling by the ratio of wall/lumen between groups at GD10.5 and GD12.5. α-SMA was strongly expressed in the spiral artery walls and used to measure artery wall thickness and lumen diameter. DAPI (Blue) was labeled to visualize nuclei. Scale bar = 500 μm for all panels. Orange bar represented the depth of CK7⁺ trophoblast infiltration into uterus. Error bar represented mean±s.e.m. *, p<0.05; **, p<0.01; two-tailed t-test.

success and failure of pregnancy. These high classification scores proved the reliability and robustness of our model, and also indicated that the tested pre-pregnancy markers had a strong predictive ability for pregnancy outcomes. Meanwhile, we tested and compared the performance of other three tree-based ML models: random forest (RF), gradient boosting decision tree (GBDT), and decision tree (DT), shown in Figure S5 and Table S3. Although RF, GBDT, and DT models were inferior to XGBoost on the testing performance (precision = 0.807, 0.859, 0.792, respectively), they all supported the predictive ability of pre-pregnancy serum metabolome on miscarriage risk. Although our AI-MP model would open new exciting windows for early prevention to effectively protect maternal and neonatal health, it was limited by the miscarriage-enriched cohort all from Chinese Han population. Further validation in a broader, unselected fertile women population of a larger sample size with diverse ethnicities would be essential

establish a high-precision classification model, with all features as input and miscarriage outcome as output (Figure 1C). Due to dimensional differences in amino acid levels and clinical characteristics, we considered XGBoost as the model for training this task,²⁰ where features were computed in the internal tree nodes (Figure 1C). During the training stage, we applied best-subset selection with 10-fold cross-validations (CVs, where the data set was divided into ten parts, alternately using nine as training data and one as testing data), and used grid search method to optimize the hyper-parameters (Figure S3). Moreover, the implicit feature selection was incorporated into XGBoost training (embedded feature engineering), which made the model interpretable and quantified contribution of each feature (feature importance). Consequently, based on the well-trained model, we developed an end-to-end AI-MP, which could be used for identifying miscarriage outcome in the pre-pregnancy stage within milliseconds (Figure 1C).

The XGBoost algorithm established a stable model within 100 steps of iterations, shown by precision of prediction on each iteration in training data set as well as in 10-fold CVs of testing set (Figure 1D). The average testing precision of 0.870 (s.d. 0.050) indicated the high reliability of our AI-MP model. The performance of AI-MP was further confirmed by the receiver operating characteristic (ROC) curve analysis with the mean area under the curve (AUC) as 0.853 (s.d. 0.061; Figure 1E), as well as by other three metrics: mean recall as 0.854 (s.d. 0.045), mean F1-score as 0.856 (s.d. 0.048), and mean accuracy as 0.855 (s.d. 0.047; Figure 1F). The confusion matrixes of testing set and total data set from 10-fold CVs (Figure 1G and Figure S4) further demonstrated the generalization ability of the model to successfully discriminate

to develop AI-MP to a clinical test.

Our pilot study revealed the ability of pre-pregnancy markers to precisely assess the miscarriage risk in the upcoming gestation, independent from any post-conception indicators, consolidating the strong influence of maternal factors on successful pregnancy. To explore the underlying biological mechanism of AI-MP, the embedded feature engineering was accomplished to decode contribution of each marker to the model. The leading contributor was previous number of miscarriages with a feature importance of 10.19% (Figure 1H), which agreed well with the previous clinical observation that miscarriage risk rose approximate 10% for each additional miscarriage, up to 42%, in women with more than two previous miscarriages.⁵ In accordance with our previous findings,²¹ glutamine was indeed a predictor for miscarriage risk with a feature importance of 7.87%. More intriguingly, pre-pregnancy serum level of histidine contributed the second largest (8.14%) to AI-MP, serving as the most important feature among all amino acids (Figure 1H).

Hyperactive histidine-histamine flux contributed to fetal loss in both human and mice

We next asked how histidine was involved in pregnancy and dysregulated in miscarriage. Given that pre-pregnancy histidine level was predictive to the following miscarriage outcome in RSM patients, we further compared this indicator between the RSM cohort, as a collectivity of substantial miscarriage risk, and the NOR cohort. A notable accumulation of histidine was indeed observed in serums from RSM patients before pregnancy (Figure 2A, Figure S6A).

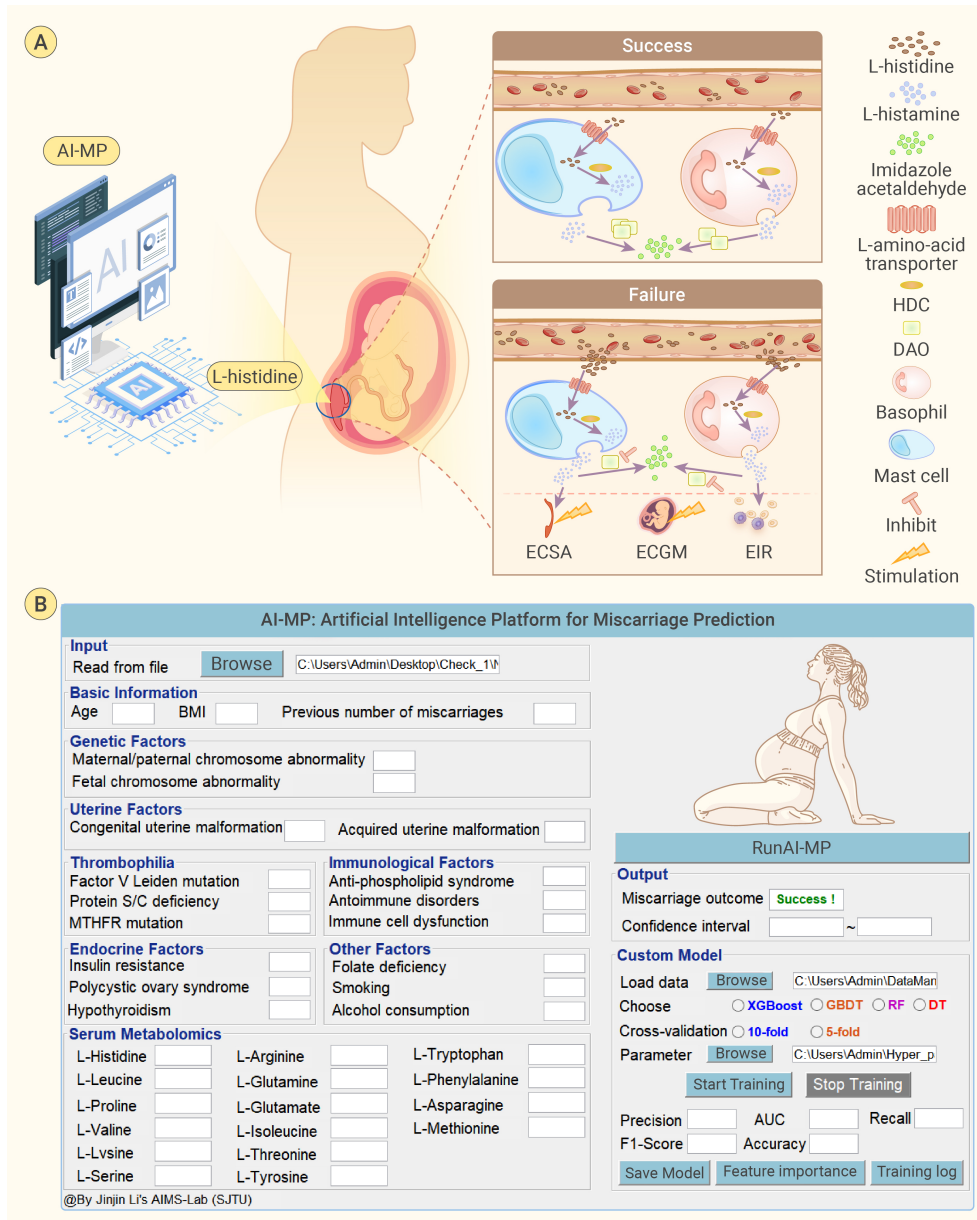


Figure 4. Revelation of AI-MP and its platform interface (A) Working model for the excessive histidine as the predictor of miscarriage. ECSA, excessive constriction on spiral artery; ECGM, excessive constriction on gestational myometrium; EIR, excessive inflammatory response. (B) Interface of AI-MP platform. The Software Copyright for AI-MP has been registered in China with No. 2022R11S 0268771.

after reduction of histamine degradation would contribute to its accumulation synergically (Figure 2F). These findings suggested that the clinically insightful histidine level in pre-pregnancy serum would cause miscarriage by hazardous histamine excess in maternal-fetal interface at early pregnancy.

To further validate the causality between histidine and miscarriage, beyond their strong correlation in human serum and maternal-fetal interface tissues, we examined fertility of the mice modelled with a high-histidine diet. Twenty-four female mice were randomized into cages with either Treated group or Ctrl group. Two weeks later, high-histidine diet led to a notable accumulation of histidine in both serum (Figure 2G) and uterus (Figure 2H), without malabsorption of other amino acids (Figure S6C and S6D). Female mice were then re-assigned to twelve new cages, each of which hosted one Treated female mouse, one Ctrl female mouse and one standard-diet male mouse for mating in a randomized way. After the confirmation of vaginal plugs, female mice were transferred back to the corresponding cages with differential diets, and their wombs were surgically examined on GD12.5. Embryo loss was prominently induced in female mice fed with the high-histidine diet (Figure 2I, 2J), along with inhibited average fetus weight (Figure 2K). Meanwhile, the implantation sites were similar between the two diets, indicating unaffected ovulation and fertilization (Figure 2L).

During pregnancy, extravillous trophoblast remodeling of the uterine spiral arteries is essential for promoting blood flow to the placenta and fetal development.²⁷ Inadequate expansive remodeling of the spiral arteries and abnormal trophoblast invasion may lead to early pregnancy loss.²⁸ To further analyze the possible mechanism of histidine intake on early pregnancy loss, we next performed multiplex immunofluorescence (mIF) staining to estimate uterine spiral artery remodeling and trophoblast invasion. We observed that the depth of CK7⁺ trophoblast infiltration into the uterus was decreased in the Treated group compared with the Ctrl group (Figure 3A, 3B). Using CD31 and α -SMA to label the wall and lumen of spiral artery, respectively, poor spiral artery remodeling can be represented by thick wall and narrow lumen. Our staining data indicated that the high histidine diet led to impaired spiral arteries remodeling (Figure 3A, 3C), suggesting that the high histidine diet might increase the risk of pregnancy loss through deficient extravillous trophoblast remodeling of the uterine spiral arteries. These observations deciphered the mechanism of histidine on miscarriage, behind its predictive contribution to AI-MP.

Development of visual tools for clinical diagnosis

Our pilot study unveiled the implication of hyperactive histidine-histamine metabolism flux in miscarriage using human maternal tissues and a mouse model (Figure 4A). More importantly, histidine serum level, along with other pre-pregnancy biomarkers, could precisely predict miscarriage risk of the

Surprisingly, such deviation became non-significant in serums collected during the first trimester (Figure 2B, Figure S6B), highlighting the unique, yet often ignored, predictive value of pre-pregnancy serum. In the decidua, histidine level was also trending to elevate in RSM subjects, without statistical significance though, possibly due to limited sample size (Figure 2C). Considering the decline of histidine level during normal endometrial decidualization (Figure 2C), we hypothesized that excessive circulating histidine might be converted in uterus to overstocked downstream metabolites. Physiologically, after transported into cells from peripheral blood, histidine is catabolized by HDC into histamine, which can be released into extracellular space by mast cells upon specific stimuli and degraded by DAO.²¹ Both these two enzymes are highly induced during the normal pregnancy.²²⁻²⁴ Inordinate maternal histamine levels and reduced plasma DAO activities have been reported to associate with pregnancy disorders, such as RSM, intrauterine growth retardation, and premature rupture of membranes,^{23,25,26} likely through excessive inflammatory response, and immoderate constriction on spiral artery or gestational myometrium.²³ Accordingly, a notable accumulation of histamine was observed in decidua of RSM patients (Figure 2D), as well as in its partner within the maternal-fetal interface, the villus tissues (Figure 2E). Conversion from histidine to histamine was probably dysregulated in RSM by chemical potential of raised histidine concentration, instead of the catalyzing rate, evidenced by comparable expression of HDC between RSM and normal subjects (Figure 2F). Additionally, compromised DAO expression and there-

upcoming pregnancy via our proposed AI-MP model. To facilitate the broad validation and thus translational application of our model, we developed an interactive platform for miscarriage prediction based on the well-trained AI-MP, of which the interface was shown in Figure 4B (see command-line interface in Figure S7, a detailed manual in Note S1, and codes public in GitHub). Clinicians could input data with ease and use our model to predict (click the "Run AI-MP" button) the miscarriage outcome in negligible time (5 ms). Moreover, the AI-MP platform would allow investigators to further optimize models with emerging data and customized parameters for better performance.

DISCUSSION

Histidine is an essential amino acid involved in various biological processes, including antioxidant properties, proton-buffering power, and metal ion chelation.^{21,29} It has been reported that dietary histidine supplementation can improve tissue repair, defend the skin from atopic dermatitis, and regulate neurological disorders, among other benefits.^{30,31} However, excess histidine supplementation may also induce side effects, such as hypercholesterolemia and liver enlargement.^{31,32} Increased ammonia level, resulting from excess histidine catabolism, may exert detrimental effects, especially in individuals with impaired liver function.³³ It is worth noting that RSM patients, particularly those with APS, commonly use anticoagulant drugs such as aspirin and low-molecular-weight heparin (LMWH).³⁴ Hepatotoxicity with LMWH is a relatively common adverse reaction in RSM patients.³⁵ Given that RSM patients have significantly higher serum histidine levels, it appears that they might be more susceptible to high histidine-induced miscarriage.

Liver is one of the first organs to form during embryonic development.³⁶ During early embryogenesis, the liver bud emerges from the foregut endoderm and begins to grow and differentiate.³⁷ In our mice experiments, we found that a high-histidine diet could lead to embryo loss. Studies have suggested that impaired liver development is associated with an increased risk of fetal loss or miscarriage.³⁸ These conditions often involve defects in bile duct formation, hepatic vasculature, or hepatobiliary development, which can compromise liver function and disrupt essential metabolic processes crucial for fetal growth and development. It could be hypothesized that a high-histidine diet might impair embryonic liver development, thus resulting in fetal loss. Histidine overload might impede early liver development and be harmful to embryonic health, even lead to embryo loss. Due to the limitation that we could not dissect embryonic liver tissues from absorbed embryos, earlier embryos need to be collected in future studies to further investigate the effects of histidine intake on embryonic liver development.

Excitingly, the establishment of the AI-MP model relies only on a noninvasive pre-pregnancy blood test and basic clinic information, making it cost-effective for healthcare systems with limited resources. Notably, the potential clinical utility of the AI-MP model is enormous due to its high accuracy, early detection prior to pregnancy establishment, and clinical benefits, such as reducing miscarriage incidence, improving pregnancy outcomes, and guiding personalized treatments. Nonetheless, the predictive performance of the AI-MP model in the general population remains unclear, as it was trained in the RSM cohort. Validation beyond RSM patients is highly valuable yet largely challenging due to the following reasons: (1) The miscarriage rate is very low among NOR women (approximately 3% in our hospital), thus requiring an enormous sample size to evaluate model performance in the NOR cohort. (2) Unlike RSM patients, who are more likely to actively prepare for pregnancy, participants of the NOR cohort are uncertain about their pregnancy plans. This causes huge difficulty to obtain serum samples within six months before conception. (3) After blood draw for various clinical reasons, only a small portion of the NOR cohort, if they conceive, choose the same hospital for prenatal check-ups, increasing the difficulty and error rate of outcome follow-up. Additionally, even if a large NOR cohort with enough tracked miscarriages is accomplished, conducting targeted metabolomics on such a large sample size poses a significant budgetary challenge.

It is worth noting that SMOTE is widely used to address class imbalance in machine learning by generating synthetic samples for the minority class. It helps models learn more effectively from limited data by providing a more balanced training set. However, SMOTE may introduce noise, especially if the minority class is highly diverse or the synthetic samples do not represent real-

world scenarios accurately. Over-sampling can also lead to overfitting, particularly when combined with high-dimensional data, which is an important consideration when handling clinical medical data.

Overall, the AI-MP model has the potential to streamline and improve healthcare practices by reducing costs and optimizing resource allocation. However, several ethical concerns must be addressed prior to the implementation of the AI-MP model as well as other predictive models.^{39,40} Informed consent is a critical ethical principle. Patients must be fully informed about the optional use of the predictive model in their care and which personal information and test results are involved. Individuals have an inherent right to privacy and confidentiality regarding their reproductive health information, necessitating robust data security to prevent breaches and unauthorized use. To be noted, the application of established models generally requires minimal computer performance and can be achieved within the intranet of a hospital. Patient voluntariness is essential, ensuring that they can choose whether or not to participate in using the predictive model without feeling coerced, thus respecting their autonomy. The psychological impact of predictive information on patients must be considered, as predictions of high miscarriage risk can cause significant distress. It is crucial to provide clear explanations of how the AI-MP model reaches its conclusions and how to interpret its prediction result. While the AI-MP model may reduce healthcare costs in the long term, evaluating its cost-effectiveness requires comprehensive input from clinical economics experts and policymakers. Regulatory frameworks must evolve to keep pace with technological advancements, ensuring that AI systems meet ethical standards and do not pose undue risks.^{41,42} By addressing these ethical considerations, the AI-MP model can be responsibly integrated into healthcare, enhancing patient outcomes and system efficiency while maintaining high ethical standards.

CONCLUSION

Currently, the continuing conviction that miscarriages are unavoidable causes a pervasive attitude of acceptance of miscarriage in the healthcare system and broader society.³ The pressure is mounting to change the status quo. Our study holds the promise of early prevention and/or intervention against miscarriage by providing a precise prediction up to six months in advance, compared to relying on biomarkers or clinical symptoms during pregnancy. A noninvasive pre-pregnancy blood test is likely to be an informative and reliable indicator for both maternal and neonatal health.

REFERENCES

- Kramer, M.S., McLean, F.H., Boyd, M.E., et al. (1988). The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. *JAMA* **260**: 3306–3308. DOI: 10.1001/jama.1988.03410220090034.
- Doubilet, P.M., Benson, C.B., Bourne, T., et al. (2013). Diagnostic criteria for nonviable pregnancy early in the first trimester. *N. Engl. J. Med.* **369**: 1443–1451. DOI: 10.1056/NEJMra1302417.
- Quenby, S., Gallos, I.D., Dhillon-Smith, R.K., et al. (2021). Miscarriage matters: The epidemiological, physical, psychological, and economic costs of early pregnancy loss. *Lancet* **397**: 1658–1667. DOI: 10.1016/s0140-6736(21)00682-6.
- Magnus, M.C., Wilcox, A.J., Morken, N.H., et al. (2019). Role of maternal age and pregnancy history in risk of miscarriage: Prospective register based study. *BMJ (Clinical research ed.)* **364**: l869. DOI: 10.1136/bmj.l869.
- Coomarasamy, A., Dhillon-Smith, R.K., Papadopolou, A., et al. (2021). Recurrent miscarriage: Evidence to accelerate action. *Lancet* **397**: 1675–1682. DOI: 10.1016/s0140-6736(21)00681-4.
- Najafi, K., Mehrooz, Z., Ardalan, F., et al. (2021). Identifying the causes of recurrent pregnancy loss in consanguineous couples using whole exome sequencing on the products of miscarriage with no chromosomal abnormalities. *Sci. Rep.* **11**: 6952. DOI: 10.1038/s41598-021-86309-9.
- Van Dijk, M.M., Kolte, A.M., Limpens, J., et al. (2020). Recurrent pregnancy loss: Diagnostic workup after two or three pregnancy losses? A systematic review of the literature and meta-analysis. *Hum. Reprod. Update* **26**: 356–367. DOI: 10.1093/humupd/dmz048.
- Kuehn, B.M. (2021). More comprehensive care for miscarriage needed worldwide. *JAMA* **325**: 2335. DOI: 10.1001/jama.2021.8549.
- Slomski, A. (2020). Mixed results for progesterone for miscarriage prevention. *JAMA* **324**: 1129. DOI: 10.1001/jama.2020.17525.
- Pearson, H. (2004). Predicting miscarriage. *Nature* DOI: 10.1038/news040105-11.
- Bender Atik, R., Christiansen, O.B., Elson, J., et al. (2023). Eshre guideline: Recurrent pregnancy loss: An update in 2022. *Hum. Reprod. Open* **2023**: hoad002. DOI: 10.1093/hropen/hoad002.

12. Practice Committee of the American Society for Reproductive Medicine. (2012). Evaluation and treatment of recurrent pregnancy loss: A committee opinion. *Fertil. Steril.* **98**: 1103–1111. DOI: 10.1016/j.fertnstert.2012.06.048.
13. De Fabo, E.C., Webber, L.J., Ulman, E.A., et al. (1997). Dietary L-histidine regulates murine skin levels of trans-urocanic acid, an immune-regulating photoreceptor, with an unanticipated modulation: Potential relevance to skin cancer. *J. Nutr.* **127**: 2158–2164. DOI: 10.1093/jn/127.11.2158.
14. Cuny, H., Rapadas, M., Gereis, J., et al. (2020). Nad deficiency due to environmental factors or gene-environment interactions causes congenital malformations and miscarriage in mice. *Proc. Natl. Acad. Sci. U.S.A.* **117**: 3738–3747. DOI: 10.1073/pnas.1916588117.
15. Wang, X.H., Xu, S., Zhou, X.Y., et al. (2021). Low chorionic villous succinate accumulation associates with recurrent spontaneous abortion risk. *Nat. Commun.* **12**: 3428. DOI: 10.1038/s41467-021-23827-0.
16. Tang, L., Xu, X.H., Xu, S., et al. (2023). Dysregulated gln-glu- α -ketoglutarate axis impairs maternal decidualization and increases the risk of recurrent spontaneous miscarriage. *Cell Rep. Med.* **4**: 101026. DOI: 10.1016/j.xcrm.2023.101026.
17. Kim, M., Chen, C., Wang, P., et al. (2022). Detection of ovarian cancer via the spectral fingerprinting of quantum-defect-modified carbon nanotubes in serum by machine learning. *Nat. Biomed.* **6**: 267–275. DOI: 10.1038/s41551-022-00860-y.
18. Megahed, F.M., Chen, Y.J., Megahed, A., et al. (2021). The class imbalance problem. *Nat. Methods* **18**: 1270–1272. DOI: 10.1038/s41592-021-01302-4.
19. Yang, H., Li, J., Lim, K.Z., et al. (2022). Automatic strain sensor design via active learning and data augmentation for soft machines. *Nat. Mach. Intell.* **4**: 84–94. DOI: 10.1038/s42256-021-00434-8.
20. Yue, Q., Yuan, A., Che, X., et al. (2016). Batch model for batched timestamps data analysis with application to the ssa disability program. *KDD* **2016**: 343–352. DOI: 10.1145/2939672.2939706.
21. Holeček, M. (2020). Histidine in health and disease: Metabolism, physiological importance, and use as a supplement. *Nutrients* **12**: 848. DOI: 10.3390/nu12030848.
22. Velicky, P., Windsperger, K., Petroczi, K., et al. (2018). Pregnancy-associated diamine oxidase originates from extravillous trophoblasts and is decreased in early-onset preeclampsia. *Sci. Rep.* **8**: 6342. DOI: 10.1038/s41598-018-24652-0.
23. Maintz, L., Schwarzer, V., Bieber, T., et al. (2008). Effects of histamine and diamine oxidase activities on pregnancy: A critical review. *Hum. Reprod. Update* **14**: 485–495. DOI: 10.1093/humupd/dmn014.
24. Weisburger, W.R., Mendelsohn, G., Eggleston, J.C., et al. (1978). Immunohistochemical localization of histaminase (diamine oxidase) in decidual cells of human placenta. *Lab. Invest.* **38**: 703–706. DOI.
25. Legge, M. and Duff, G.B. (1981). Plasma diamine oxidase levels in pregnancy complicated by threatened abortion. *J. Clin. Pathol.* **34**: 187–188. DOI: 10.1136/jcp.34.2.187.
26. Brew, O. and Sullivan, M.H. (2006). The links between maternal histamine levels and complications of human pregnancy. *J. Reprod. Immunol.* **72**: 94–107. DOI: 10.1016/j.jri.2006.04.002.
27. Fitzgerald, J.S., Poehlmann, T.G., Schleussner, E., et al. (2008). Trophoblast invasion: The role of intracellular cytokine signalling via signal transducer and activator of transcription 3 (stat3). *Hum. Reprod. Update* **14**: 335–344. DOI: 10.1093/humupd/dmn010.
28. Díaz-Hernández, I., Alecsandru, D., García-Velasco, J.A., et al. (2021). Uterine natural killer cells: From foe to friend in reproduction. *Hum. Reprod. Update* **27**: 720–746. DOI: 10.1093/humupd/dmaa062.
29. Vera-Aviles, M., Vantana, E., Kardinassari, E., et al. (2018). Protective role of histidine supplementation against oxidative stress damage in the management of anemia of chronic kidney disease. *Pharmaceuticals (Basel, Switzerland)* **11**: 111. DOI: 10.3390/ph11040111.
30. Gibbs, N.K. (2020). L-histidine supplementation in adults and young children with atopic dermatitis (eczema). *J. Nutr.* **150**: 2576s–2579s. DOI: 10.1093/jn/nxaa200.
31. Thalacker-Mercer, A.E. and Gheller, M.E. (2020). Benefits and adverse effects of histidine supplementation. *J. Nutr.* **150**: 2588s–2592s. DOI: 10.1093/jn/nxaa229.
32. Holeček, M. and Vodeníčarovová, M. (2020). Effects of histidine supplementation on amino acid metabolism in rats. *Physiol. Res.* **69**: 99–111. DOI: 10.33549/physiolres.934296.
33. Holeček, M. and Vodeníčarovová, M. (2019). Effects of histidine load on ammonia, amino acid, and adenine nucleotide concentrations in rats. *Amino Acids* **51**: 1667–1680. DOI: 10.1007/s00726-019-02803-5.
34. Deng, T., Liao, X., and Zhu, S. (2022). Recent advances in treatment of recurrent spontaneous abortion. *Obstet. Gynecol. Surv.* **77**: 355–366. DOI: 10.1097/ogx.0000000000001033.
35. Yang, X., Li, N., Guo, T., et al. (2020). Comparison of the effects of low-molecular-weight heparin and fondaparinux on liver function in patients with pulmonary embolism. *J. Clin. Pharmacol.* **60**: 1671–1678. DOI: 10.1002/jcph.1686.
36. Zaret, K.S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nat. Rev. Genet.* **3**: 499–512. DOI: 10.1038/nrg837.
37. Shin, D. and Monga, S.P. (2013). Cellular and molecular basis of liver development. *Compr. Physiol.* **3**: 799–815. DOI: 10.1002/cphy.c120022.
38. Tanaka, M., Fuentes, M.E., Yamaguchi, K., et al. (1999). Embryonic lethality, liver degeneration, and impaired nf-kappa b activation in ikk-beta-deficient mice. *Immunity* **10**: 421–429. DOI: 10.1016/s1074-7613(00)80042-4.
39. Liu, X., Zhang, S., Shao, L., et al. (2024). Improving prediction of treatment response and prognosis in colorectal cancer with ai-based medical image analysis. *The Innovation Medicine* **2**: 100069. DOI: 10.59717/j.xinn-med.2024.100069.
40. Xu, Y., Liu, X., Cao, X., et al. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation* **2**: 100179. DOI: 10.1016/j.xinn.2021.100179.
41. Hantel, A., Walsh, T.P., Marron, J.M., et al. (2024). Perspectives of oncologists on the ethical implications of using artificial intelligence for cancer care. *JAMA Netw. Open* **7**: e244077. DOI: 10.1001/jamanetworkopen.2024.4077.
42. Huang, T., Xu, H., Wang, H., et al. (2023). Artificial intelligence for medicine: Progress, challenges, and perspectives. *The Innovation Medicine* **1**: 100030. DOI: 10.59717/j.xinn-med.2023.100030.

FUNDING AND ACKNOWLEDGMENTS

The work is supported by the National Key R&D Program of China (Nos. 2022YFC2702204, 2021YFC2100100, 2022YFC2704702), the National Natural Science Foundation of China (Nos. 82271701, 81730039, 81901069, 21901157, 82071653, 82171657, 81971384), the Shanghai Science and Technology Project (Nos. 21JC1403400, 22ZR1480400, 23ZR1450400, 22140903400, 23JC1402300), the SJTU Global Strategic Partnership Fund (2020 SJTU-HUJI), the Shanghai Municipal Medical and Health Discipline Construction Projects (No. 2017ZZ02015), Shenkang Technology standardization management and promotion project (SHDC 22023204) and Science and Technology Development Fund institutions in Pudong New Area (PKJ2022-Y17). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Lincheng Tang, Zhilong Wang, Xianghong Xu, Yawei Yuan: Conceptualization, methodology, visualization, data curation, and writing-original draft. Fan Hao, An Chen, Shengnan Wu, Wenxuan Li, Yanqiang Han: Methodology, validation, and writing. Jiaxue Sun, Kehao Tao, Zeying Liu: Methodology and validation. Zhiyun Wei, Jinjin Li, Liping Jin: Conceptualization, methodology, supervision, resources, and writing-review and editing.

DECLARATION OF INTERESTS

We have registered a Software Copyright for AI-MP platform in China with No. 2022R11S0268771. The authors declare no other competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

The human subjects research was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Shanghai First Maternity and Infant Hospital (No. KS18133). All study participants provided written informed consents. All animal experiments were performed in accordance with National Institutes of Health published Guide for the Care and Use of Laboratory Animals. Approval was granted by the Institutional Animal Care and Use Committee of Shanghai Model Organisms (No. 2023-004).

DATA AND CODE AVAILABILITY

The well-trained XGBoost models and the codes of command-line interface are available at <https://github.com/CodingWZL/AI-MP/tree/CodingWZL-patch-1> and https://www.aimslab.cn/#/miscarriage_prediction.

SUPPLEMENTAL INFORMATION

It can be found online at <https://doi.org/10.59717/j.xinn-med.2024.100093>

LEAD CONTACT WEBSITE

Liping Jin: <https://med.tongji.edu.cn/info/1428/5079.htm>