

Clinical And Laboratory Determinants Of Preeclampsia Onset Using Ensemble Machine Learning Prediction And Comparison Among Stacking Ensemble, Neural Network With And Without Stacking Ensemble Approach

Sangita Bhasme^{1*}, Parminder Kaur², Dr Shrinivas N Gadappa³

^{1*}Department of Information Technology, Faculty of Engineering, MGM University, Maharashtra 431001, India

²Faculty of Engineering, MGM University, Maharashtra 431001, India

³Professor & Head, Department of Obstetrics & Gynaecology, Government Medical College, Chhatrapati Sambhajinagar, Maharashtra 431001, India

*Corresponding E-mail: sangitabhasme20@gmail.com

ABSTRACT

Preeclampsia is a pregnancy-related hypertensive illness that bears increased maternal and fetal morbidity and mortality. For early risk stratification of preeclampsia, the development of models that can forecast its onset is crucial. This work analyzes maternal clinical and laboratory features using bagging- and boosting-based ensemble models in comparison with a neural network stacking framework that incorporates these models. Experimental results show that Random Forest yields the highest performance of 81% for late-onset preeclampsia and maintains robust performance for early-onset preeclampsia. In contrast, the boosting models XGBoost and CatBoost yield competitive performance with an accuracy of 77% and 71%, respectively. However, their performance is still inferior to that from Random Forest. Hematological parameters, especially total leukocyte count and platelet count, are major predictors influencing model performance and misclassification patterns across all the ensemble models. On the other hand, neural network-based stacking ensemble models yielded more balanced and reliable predictions, with great value for clinical risk assessment. The developed models can enhance the early risk stratification of preeclampsia and support decisions for obstetric care.

Keywords: Machine Learning Ensemble Models; Neural Network-Based Stacking Ensemble; Neural Networks; Clinical and Laboratory Risk Factors and Onset Preeclampsia prediction.

1. INTRODUCTION

Preeclampsia is a multisystem disorder of pregnancy and the puerperium, characterized by hypertension more than or equal to 140/90 mmHg growing after 20 weeks of gestation in formerly normotensive women (Sibai, 1999). It affects approximately 2–10% of pregnancies worldwide and is up to seven times more prevalent in growing countries compared to developed nations (Kooffreh et al., 2010). Based on the onset of symptoms, preeclampsia is classified as Early-Onset Preeclampsia (EOPE), occurring before 34 weeks, and Late-Onset Preeclampsia (LOPE), occurring after 34 weeks (Hellaman et al., 2014). EOPE, the more severe variant accounting for 5–20% of cases, is associated with poor neonatal outcomes, whereas LOPE, representing 75–80% of cases, is linked to maternal comorbidities such as metabolic syndrome, obesity, and chronic hypertension, with generally normal placental volume and birth weight (Madazli et al., 2014). In a systematic review performed by the World Health Organization, hypertensive disorders including preeclampsia accounted for accounting for approximately 14% of maternal deaths between

2003 and 2009, with a sharp rise in severe obstetric morbidities among affected women (Shu-Han et al., 2001–2014). Variations in diagnostic the criteria between the American College of Obstetricians and Gynecologists and the International Society for the Study of Hypertension in Pregnancy, have contributed to variability in reported maternal and fetal outcomes (ACOG & TFHP, 2013; Tranquilli et al., 2014; Magee et al., 2014). Both Organizations have updated the definitions recently, yet some uncertainties still linger in regard to the consequences of EOPE and LOPE A.L. Tranquilli et al., 2013; Poon & Nicolaides, 2014; PCSMF & Sibai, 2011; Pettit et al., 2015). Generally, EOPE poses a higher risk to both mother and fetus, whereas LOPE tends to present with less severe clinical features (Paruk & Moodley, 2000; Aksornphusitaphong & Phupong, 2013; I. Stergiotou et al., 2013).

A cross-sectional study in the Department of Obstetrics and Gynaecology at Bangabandhu Sheikh Mujib Medical University and Dhaka Medical College Hospital, Dhaka, over a period of six months. The incidence of preeclampsia in this study was 3.4%. The most common maternal symptoms were pedal edema and abdominal pain 38%. Nearly 46% of pregnancies were terminated between 29 and 34 weeks of gestation. Major maternal complications included abruptio placenta 4%, postpartum eclampsia 2%, and disseminated intravascular coagulation 2%. Among the affected pregnancies, 80% resulted in live births, while stillbirths and intrauterine deaths each accounted for 6%, and abortions occurred in 8% of cases. There were many fetal complications, including intrauterine growth restriction 50% of in newborns and prematurity in 52%. The conclusion of this study was abruption placenta, eclampsia, and Disseminated Intravascular Coagulation. These were the most common maternal complications, whereas perinatal mortality and stillbirths emerged as major fetal issues (Khan et al., 2013). A retrospective, descriptive, cross-sectional study conducted at the Istanbul Medeniyet University Goztepe Teaching and Research Hospital, analyzing preeclampsia diagnoses from 2012 to 2014. Of the 116 participants, 43 (37%) had severe preeclampsia, and 73 (62%) had mild preeclampsia. EOPE was present in 29 cases, while 71% of the subjects in the other. Maternal adverse outcomes occurred in 18 cases, including six cases of placental abruption and two cases of eclampsia. The need for blood transfusion was the most frequent maternal complication, and one case of maternal mortality was reported due to Disseminated Intravascular Coagulation and Hemolysis, Elevated Liver enzymes, and Low Platelet count. The most common adverse fetal outcomes were Neonatal Intensive Care Unit admission, low Apgar scores, and neonatal death (Bozdogan et al., 2015). An investigated hypertensive disorders of pregnancy among women beyond 20 weeks of gestation admitted to the Kamineni Institute of Medical Sciences over one year. Hypertensive disorders were identified in 7.9% of total deliveries, comprising gestational hypertension 2.1%, preeclampsia 4.9%, and eclampsia 0.9%. Hemolysis, Elevated Liver enzymes, and Low Platelet count was the most common maternal complication 3.4%, though no maternal mortality was reported. Preterm delivery occurred in 3.47% of cases, perinatal mortality in 23%, and intrauterine growth restriction was the most frequent fetal complication, affecting 33.4% of neonates (Kolluru et al., 2016). A retrospective study at King George Hospital, Visakhapatnam, over a six-month period. Most patients 62.3% were aged 20–25 years, and 67.7% were primigravidae, with many referred from peripheral centers. Deranged liver function tests were observed in 12.1% and abnormal renal function in 6.5% of patients. Preterm delivery occurred in 30.96% of cases, 7.1% of neonates required Neonatal Intensive Care Unit admission, and there were 28 neonatal deaths (Madhuri & Varalakshmi, 2019). A cross-sectional study at the Lady

Reading Hospital, Peshawar, between March and September 2016, involving 254 patients with preeclampsia. Of these, 172 (67.7%) had LOPE and 82 (32.3%) had EOPE. The mean Body Mass Index for EOPE and LOPE groups was 25.25 ± 1.70 kg/m² and 25.97 ± 2.03 kg/m², respectively. Multigravidity was more common in both groups. Preterm birth was significantly higher among EOPE patients 59.75% compared to LOPE 34.88% ($p=0.0001$), while cesarean sections were more frequent in LOPE 36.02% than EOPE 19.53% ($p=0.008$). Neonatal Intensive Care Unit admissions were also higher in EOPE 60.92% than LOPE 41.82% ($p=0.004$). Neonatal deaths and small for gestational age rates did not differ significantly. The study concluded that EOPE was associated with higher rates of preterm birth and Neonatal Intensive Care Unit admission, whereas LOPE was more frequently managed via cesarean delivery (Gohar & Syed, 2021).

Although traditional clinical and laboratory assessments remain central to preeclampsia diagnosis, they often identify the disorder only after clinical manifestation, limiting opportunities for early intervention. Due to the complex and multifactorial as well as nonlinear interaction of various maternal demographic, clinical, and laboratory variables, using conventional statistical methods may not suffice for accurate early prediction, especially when differentiating between EOPE and LOPE. Recent breakthroughs in machine learning (ML), therefore, offer a chance to overcome these drawbacks in order to enable the analysis of high dimensional, heterogeneous datasets and identifying complex patterns which are difficult to identify using conventional approaches. Ensemble learning methods, bagging and boosting algorithms, and have shown to improve prediction and robustness in medical classification tasks. In addition, the stacking ensemble method with the artificial neural network model has the advantage for the integration of the strengths of various models, for improving generalizability and minimizing misclassification risk.

Based on this rationale, the primary hypothesis in this investigation is that ensemble machine learning models, especially neural network stacking models that combine bagging and boosting algorithms, which can enhance prediction and risk stratification of early and late onset preeclampsia relative to person-based models, as well as identifying clinical predictors of interest to the disease development.

The specific objectives of the present study are : To demonstrate the potential of machine learning based decision support systems for early risk stratification and proactive clinical management of preeclampsia in obstetric care by analyzing maternal clinical and laboratory characteristics associated with preeclampsia, with particular emphasis on distinguishing early-onset and late-onset disease. First to develop and evaluate bagging- and boosting-based ensemble machine learning models, for predicting EOPE and LOPE and identify key predictive features driving model performance and misclassification patterns, thereby enhancing interpretability and clinical relevance. Second to construct and assess a neural network-based stacking ensemble framework that integrates multiple ensemble models and compares its performance with individual classifiers.

2. MATERIALS AND METHODS

2.1 Description of Data

In this prospective observational study, 103 records of pregnant women from which 34 records of early onset preeclampsia and 69 records of late onset preeclampsia outcome of preeclampsia since 2020 to 2022 from tertiary care center, Chhatrapati Sambhajinagar in the state of Maharashtra, India. This study is being conducted on hospital records to know

the different risk factors, demographic, vital signs, lab results. Preprocessing steps used to handle missing values, categorical encoding. For pure boosting, XGBoost, LightGBM, CatBoost, AdaBoost, Gradient Boosting Model no scaling is needed.

2.2 Data collection

Data were collected on Demographic Features, these variables describe the basic maternal characteristics, Age (years), Gravida (number of pregnancies), Body Mass Index (BMI, kg/m²). Clinical Features, these parameters are obtained from physical examination and bedside monitoring including Signs and Symptoms e.g., headache, epigastric pain, visual disturbances, nausea, vomiting. Peripheral capillary oxygen saturation (SpO₂, %), Pulse Rate (PR, bpm), Systolic Blood Pressure (SBP, mmHg), Diastolic Blood Pressure (DBP, mmHg), Respiratory Rate (RR, breaths/min), Fetal Heart Sounds (FHS, bpm). Biochemical or Laboratory Features, these variables are derived from hematological, biochemical, and urine investigations. Hematological Parameters as Hemoglobin (g/dL), Platelet Count (cells/cumm). Liver Function Parameters includes Total Bilirubin (mg/dL), Serum Glutamic Oxaloacetic Transaminase (SGOT, U/L), Serum Glutamic Pyruvic Transaminase (SGPT, U/L). Renal Function Parameters are Urea (mg/dL), Creatinine (mg/dL). Coagulation and Urinary Parameters includes International Normalized Ratio (INR), Urinary Albumin (proteinuria) Preeclampsia, the diagnosis of preeclampsia will be made using the current American College of Obstetricians and Gynaecologists guidelines (ACOG, 2002). These recommendations outline preeclampsia as sustained pregnancy-triggered high blood pressure with proteinuria. Hypertension became described as sustained blood stress readings of greater than 140/90 mmHg. ACOG defines proteinuria as urine protein concentrations of greater than 30 mg/dL on or greater random specimens gathered greater than 4 hours apart. The onset of pre-eclampsia may be divided into early and overdue onset; early onset is gestational age much less than 34 weeks, and late onset is gestational age of 34 weeks or greater.

Table 1. Distribution of Maternal Characteristics According to Early and Late Onset of Preeclampsia

Maternal characteristics of 103 records	Categories/ Range of Maternal characteristics	Early Onset PE (34)	Late Onset PE (69)
Age in years	18 to 20	11 (32.35%)	14 (20.29%)
	21 to 25	13 (38.24%)	36 (52.17%)
	26 to 30	7 (20.59%)	15 (21.74%)
	31 to 35	1 (2.94%)	2 (2.90%)
	>35	2 (5.88%)	1 (1.45%)
Gravida in number	G1	14 (41.18%)	38 (55.07%)
	G2	15 (44.12%)	11 (15.94%)
	G3	3 (8.82%)	15 (21.74%)
	G4	3 (8.82%)	4 (5.80%)
BMI in kg/m ²	Normal (18.5 to 24.9)	29 (85.29%)	47 (68.12%)
	Pre-Obesity (25.0 to 29.9)	3 (8.82%)	13 (18.84%)

	Obesity (>30.0)	3 (8.82%)	8 (11.59%)
Signs and Symptoms	Yes	19 (55.88%)	33 (47.83%)
	No	15 (44.12%)	36 (52.17%)
SPO2 in %	Normal: 95–100%	34 (100.00%)	69 (100.00%)
	Mild hypoxemia: 90–94%	-	-
	Severe hypoxemia : <90%	-	-
PR in bpm	< 60	24 (70.59%)	58 (84.06%)
	>100	7 (20.59%)	11 (15.94%)
SBP in mmHg	Non- Sever >140 and <160	14 (41.18%)	39 (56.52%)
	Sever >=160	20 (58.82%)	30 (43.48%)
DBP in mmHg	Non-Sever >90 and <110	26 (76.47%)	65 (94.20%)
	Sever >=110	7 (20.59%)	5 (7.25%)
RR breaths/min	<12	-	-
	12–20	16 (47.06%)	38 (55.07%)
	>20	18 (52.94%)	31 (44.93%)
FHS in bpm	<110	-	-
	Normal: 110–160	34 (100.00%)	69 (100.00%)
	>160	-	-
Hemoglobin in g/dL	Very Severe (<4.0)	-	-
	Severe (4.0 to 6.9)	2 (5.88%)	5 (7.25%)
	Moderate (7.0 to 9.9)	7 (20.59%)	19 (27.54%)
	Mild (10.0 to 10.9)	3 (8.82%)	16 (23.19%)
	Normal (>11.0)	24 (70.59%)	27 (39.13%)
Platelet in cell/cumm	<1 lakh	1 (2.94%)	5 (7.25%)
	>1 lakh	33 (97.06%)	64 (92.75%)
Total Bilirubin in mg/dL	<5	34 (100.00%)	69 (100.00%)
	>5	-	-
SGOT in U/L	<45	33	27 (39.13%)
	>45	1 (2.94%)	5 (7.25%)
SGPT in U/L	<45	33	27 (39.13%)
	>45	1 (2.94%)	5 (7.25%)
Urea in mg/dL	<60	34 (100.00%)	69 (100.00%)
	>60	-	-
Creatinine in mg/dL	<2	34 (100.00%)	69 (100.00%)
	>2	-	-
INR	<2	34 (100.00%)	69 (100.00%)
	>2	-	-

Albumin in Urine	Absent	2 (5.88%)	12 (17.39%)
	Traces	4 (11.76%)	4 (5.80%)
	Plus	24 (70.59%)	36 (52.17%)
	Plus Plus	4 (11.76%)	17 24.64%)

2.3 Statistical analysis of data

Table 1, the study revealed that 71.7% of the participants were diagnosed with late-onset preeclampsia (≥ 34 weeks of gestation), while 28.3% presented with early-onset preeclampsia (< 34 weeks). Most mothers were aged 21–25 years, were primigravida, and had a normal BMI, showing no significant association between these demographic parameters and the onset of preeclampsia. Clinical symptoms were more prominent in early-onset cases, with epigastric pain and jaundice showing statistically significant associations. The mean SBP and DBP were significantly higher among early-onset cases (SBP: 160.53 ± 15.62 mmHg; DBP: 100.27 ± 7.91 mmHg) compared to late-onset cases (SBP: 153.95 ± 14.80 mmHg; DBP: 96.84 ± 6.98 mmHg), indicating a clear relationship between blood pressure severity and the timing of onset. Regarding hematological and biochemical parameters, lower hemoglobin and platelet counts were observed among mothers with late-onset preeclampsia, both showing statistically significant associations. SGOT levels were also significantly higher in late-onset cases, while SGPT abnormalities were more frequent in this group. In contrast, total bilirubin, blood urea, serum creatinine, and INR levels showed no significant differences between groups. Conversely, urine albumin levels were significantly higher among mothers with early-onset preeclampsia, highlighting its potential as an indicator of disease severity. Overall, the findings suggest that blood pressure, hemoglobin, platelet count, SGOT, and urine albumin serve as important differentiating markers between early- and late-onset preeclampsia, whereas most other biochemical parameters remain unaffected by onset variation.

3. RESULTS

3.1 The Correlation Among Maternal Characteristics

The correlation heatmap (see Fig. 1) provides an overview of the pairwise linear relationships among the maternal clinical, laboratory, and vital-sign variables included in the study. Overall, most correlations are weak to moderate, indicating that the dataset contains largely independent predictors, which is advantageous for machine learning model performance. Hemodynamic and Vital Parameters, SBP and DBP confirmed 51% the strongest positive relationship in the dataset, reflecting their anticipated physiological interdependence. RR showed weak-to-moderate associations with SBP 25% and DBP 14%, signifying mild co-variation with hemodynamic status. PR exhibited minimal correlations with other vital parameters, indicating its relative independence as a clinical indicator. Liver and Kidney Function Markers, A cluster of moderate-to-high correlations was observed among the biochemical parameters related to hepatic and renal function. SGOT and SGPT confirmed 60% as a strong positive correlation, which follows the usual patterns of hepatic stress or

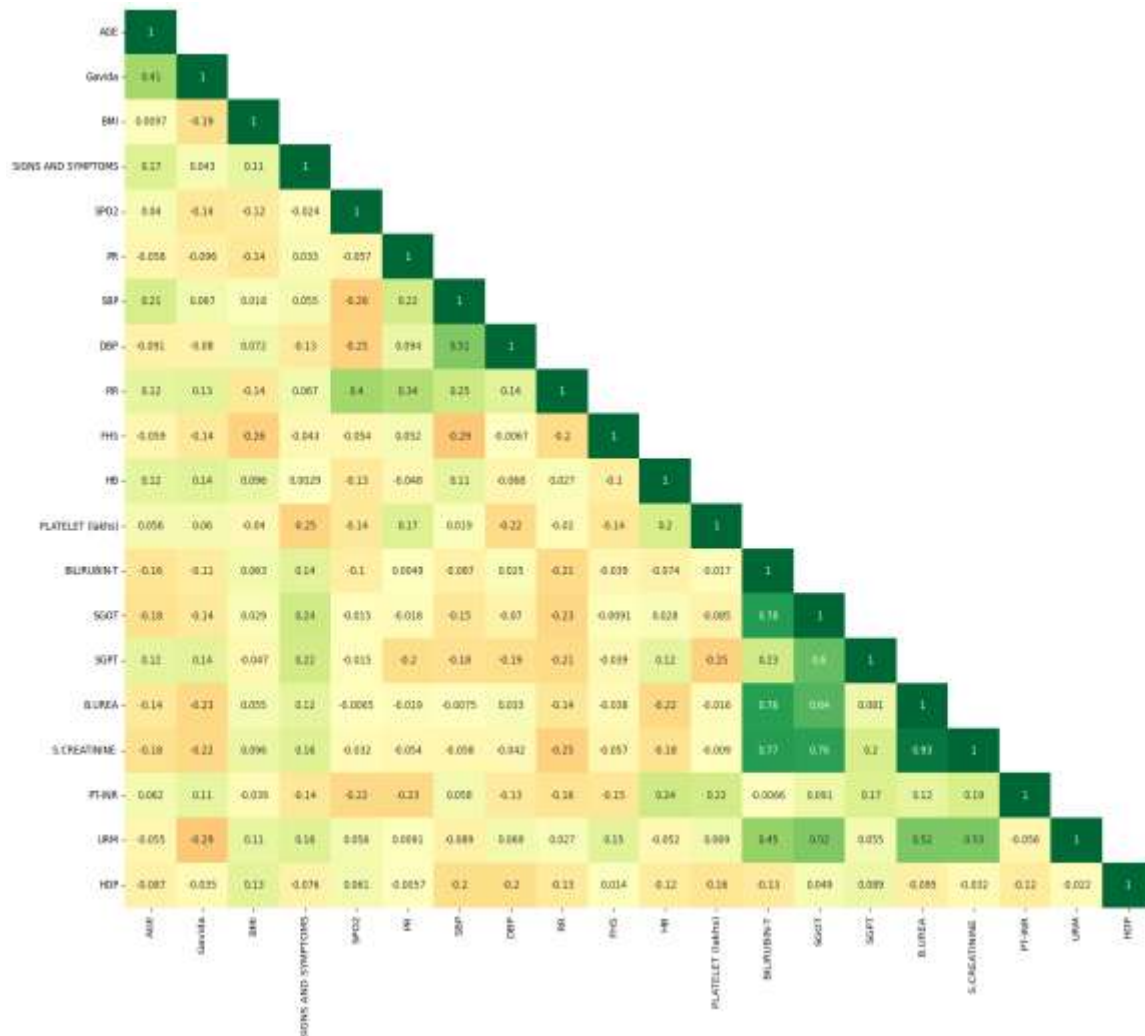


Fig. 1. The Heatmap of Maternal Characteristics

dysfunction. Similarly, serum creatinine was highly correlated with urea 93 %, SGOT 76%, and SGPT 64%, thus associating impaired renal function with increases in liver enzymes. Patterns typical of severe maternal complications, such as preeclampsia and HELLP Syndrome. Most of the hematological indicators such as HB and platelet count showed weak correlations with most variables. Platelet count showed a moderate poor correlation with both SGOT by -7% and SGPT by -12%, which follows expected clinical trends where thrombocytopenia may occur along with Liver enzymes are elevated in the hypertensive disorders of pregnancy.

Maternal History and Demographic-Mat Variables, Maternal age correlated poorly with all clinical variables, which indicates that age is not correlated with the physiological data in this database. Gravida showed limited linear relation between the heights and the resting relationships with BMI 19% and URM by negative 29%, suggesting a subtle variation in metabolic or renal markers across parity. Pregnancy Outcomes: Hypertensive Disorder of pregnancy (HDP) had low correlations with most predictors, indicating that the condition tends to be complex and nonlinear: by one single variable. It looks like a condition having

multiple factors and not linear, complex interactions. Thus, learning algorithms requiring the ability to find non-linear mappings in high-dimensional spaces are most suitable.

Pregnancy risk prediction is a process of evaluating a number of interconnected variables including maternal age, BMI, blood pressure, biochemical markers, ultrasound parameters, lifestyle habits, and genetics predispositions. In general, these variables interact in complex ways, often in nonlinear patterns, making difficult statistical models like logistic regression less effective in capturing their underlying relationships. In Reasons other machine learning algorithms such as Random Forests, Gradient Boosting Machines, and Neural Networks—are able to spot complex patterns and high dimensional feature interactions in the clinic data (Liu et al., 2022). Recent work has shown the ability of ML-based predictive models to demonstrate superiority to traditional risk-scoring systems in the assessment of pregnancy-related complications such as preeclampsia and gestational diabetes. These models, including XGBoost and Random Forest, have been able to achieve better performance in classification tasks, indicating AUC values from 0.85 to 0.90, outperforming traditional regression-based approaches (Ye et al., 2020).

Machine learning early risk detection of women at high risk of developing hypertensive disorders of pregnancy as early as the first trimester of pregnancy, allowing preventive interventions such as low-dose aspirin treatment, intensive surveillance, and referral to specialized care (Sovio et al., 2020). In contrast to fixed clinical scoring systems, ML models may be retrained and optimized for the local data sets, enabling flexibility for differences in population demographics, disease prevalence, and healthcare infrastructure (Jhee et al., 2019). In addition, the combination of the incorporation of data sources such as clinical, lab, and imaging data helps to improve the predictive accuracy and reliability of early risk detection models compared to single-modality approaches (Thangaratnam et al., 2011).

3.2 The Comparative Analysis of Ensemble Models

Since the dataset included 103 patient entries, several procedural steps needed to be carried out in relation to experiment. A cross validation approach was adopted, wherein both 70% training and 30% testing instances were adopted, because the number of classes in the data was imbalance, Class-weight corrections and imbalance handling were also used within each fold to ensure the reliability of the performance of the model. In the study, some of the learning techniques that were applied through the use of Ensemble learning techniques that included Random Forests (RF), GBM, and XGBoost have demonstrated robustness in predicting preeclampsia by incorporating multiple weak learners to increase accuracy, sensitivity, and robustness. Boosting algorithms are built trees sequentially, with each successive learner trying to correct the mistakes made by the previous one in their training trees. This is an iterative approach that often results in good predictive results but is prone to overfitting. It is commonly used in the presence of noisy or limited data sets. On the other hand, RF generates many trees independently using Bootstrap sampling and aggregate predictions of the bootstrap to decrease variance.

Table 2 shows the comparative results of various ensemble models, where RF is performed better than all other algorithms in the evaluation measures. For the LOPE scenario, RF consistently performed better than all other algorithms in achieving the highest accuracy of 0.81 with balanced precision, recall, and F1-score of 0.87, showing good predictive stability. The model was also able to perform well under the EOPE condition. However, boosting algorithms showed relatively higher

Table 2. Performance of Individual Ensemble -Based Models

ML Model	EOPE			LOPE			Accuracy
	precision	recall	f1-Score	precision	recall	f1-Score	
RF	0.62	0.62	0.62	0.87	0.87	0.87	0.81
GBM	0.33	0.25	0.29	0.76	0.83	0.79	0.68
XGBoost	0.56	0.62	0.59	0.86	0.83	0.84	0.77
LightBoost	0.33	0.38	0.35	0.77	0.74	0.76	0.65
CatBoost	0.43	0.38	0.40	0.79	0.83	0.81	0.71

variability GBM marked a remarkably low recall of 0.25 for EOPE, while LightGBM obtained the lowest accuracy of 0.65. Though XGBoost and CatBoost had competitive F1-scores of 0.84 and 0.81, it was still marginally inferior to RF with respect to predictive balance and robustness. Owing to its minimal difference in performance the need for hyperparameter tuning, RF stood out as the most successful and robust model in this sternal impotence clinical setting.

RF's high AUC value further validates its classification robustness, whereas lower AUCs for

LightGBM and XGBoost indicate reduced generalization capability (see Fig. 2).

Comparing the feature importances between ensemble models, it is observed that XGBoost gave feature importance much greater importance to some predictors, including Platelet Count, Bilirubin-D, and SGPT—emphasis on their significant discriminative power. By contrast, RF, GBM, LightGBM and CatBoost have shown more balanced distribution on feature contributions, which indicate equal importance assignments by the model variables. These results suggest that although XGBoost effectively learns nonlinear interactions between features, the superior overall generalization and reliability of RF confirm again its aptness for use in cases of preeclampsia risk prediction (see. Fig. 3).

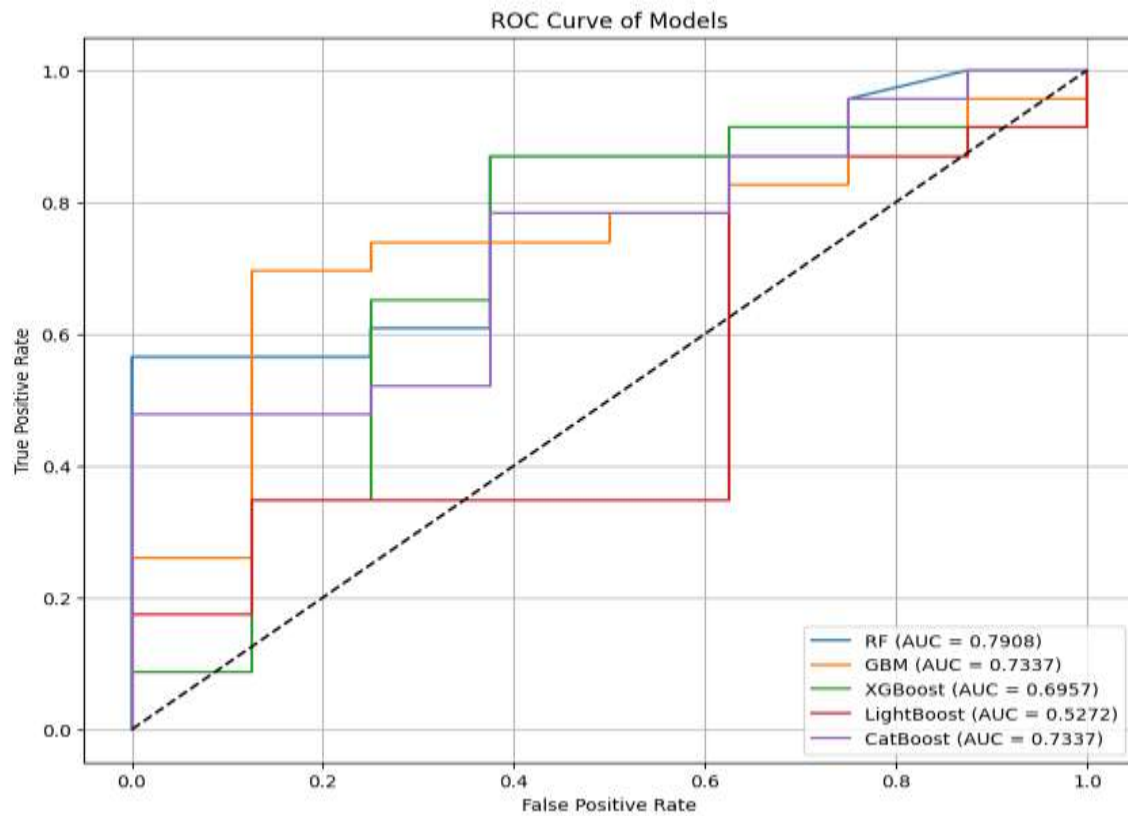


Fig. 2. ROC curve of Ensemble -Based Models.

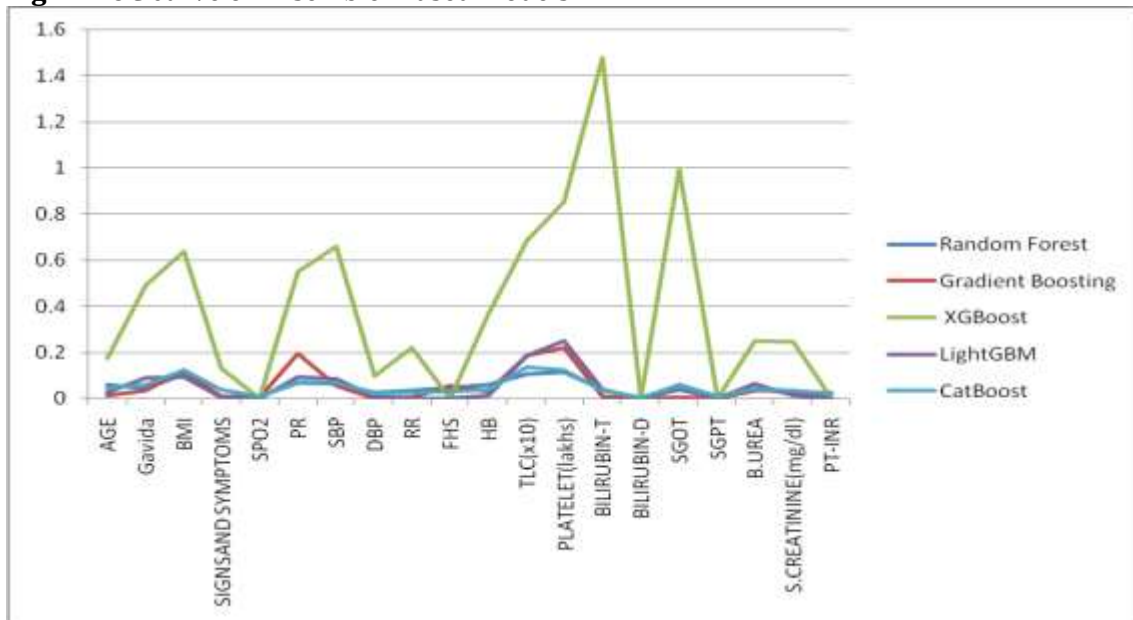


Fig. 3. Feature Importance of ML Models

Overall Trends: A large number of models (Random Forest, Gradient Boosting, XGBoost, CatBoost) give low to moderate weights to most variables, which implies that HDP risk is

the product of multifactorial causes, and there is no leading factor in these models. Models demonstrate consistent significance for blood pressure variables (SBP, DBP), platelet count, SGOT, SGPT, urea, and serum creatinine, in agreement with their relevance in hypertensive disorders. Model-Level Comparison: Random Forest and Gradient Boosting are rather smooth and comparable importance distributions, suggesting stable identification of many moderately influential predictors. XGBoost and CatBoost are slightly better at discriminating liver enzymes, such as SGOT, SGPT) and renal indicators (urea, serum creatinine). LightGBM gives top importance to a subset of Laboratory biomarkers, indicating greater sensitivity to medical variables related to the organs dysfunction. Clinical Implications: The overall pattern suggests that HDP is associated with a spectrum of Clinical parameters including blood pressure, renal function, liver enzymes, and hematological parameters. The distinct peaks in LightGBM thus give evidence of detecting subtle and high-impact risk contributors which may be less clear-cut with traditional tree-based boosting methods. This suggests that boosting algorithms, especially LightGBM or XGBoost, to achieve better results in the predictive performance about maternal risk modeling

SHAP force plots of four incorrectly classified cases, In all instances, the interplay between the risk-enhancing (red) and protective (blue). The features led to the ultimate result because misclassification was due to the dominant protective factors, however, factors outweighed clinically significant risk indicators. For the GBM ($f(x)=-5.72$), strong contributions from TLC: 7300, platelet count: 2.76 lakhs, BMI: 23.0, PT-INR: It reduced the effect of high systolic pressure (140mmHg) and hence led to underestimation of risk. The XGBoost model ($f(x) = -2.54$) showed positive effects of PR (94.0), age (21), and SBP (140) were offset by adverse factors of TLC, platelet count, SGOT (27.0), and BMI (23.0), which lowered sensitivity to cardiovascular and age-related risk. The LightGBM model ($f(x) = -2.34$) also showed comparable results, wherein the good aspects like PR and SBP were overpowered by the bad influences of TLC, platelet count, blood urea (45.0), BMI, and age. Finally, the CatBoost model ($f(x) = -1.26$) exhibited a problem of misclassification owing to the strong protective factors provided by TLC platelet count, PT-INR, blood urea, BMI, bilirubin-T (0.9), and SGOT Metabolic stability (see Fig. 4).

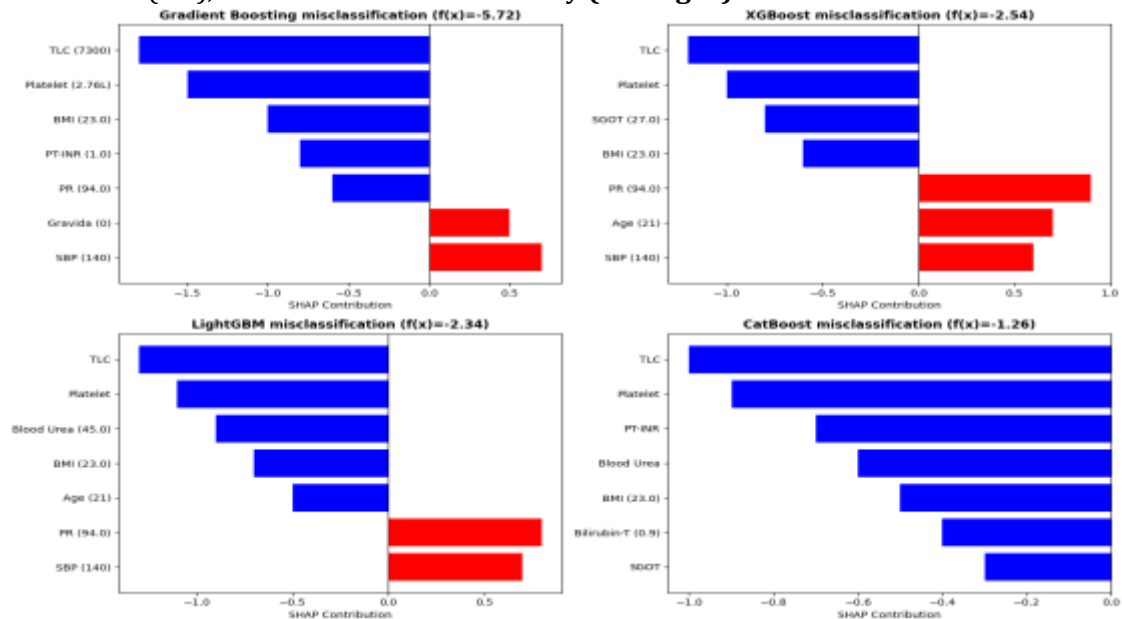


Fig. 4. SHAP force plots for four misclassified cases

Collectively, these results reveal a systematic bias toward protective hematological and biochemical markers at the expense of hemodynamic and clinical risk indicators such as high SBP, elevated pulse rate, and young maternal age. This finding highlights the need for feature rebalancing or clinically informed weighting strategies to improve interpretability and reduce misclassification in maternal risk prediction models.

3.3 Comparative Interpretation of SHAP Explanations for Misclassified Cases

In this study ensemble learning methods RF shows highest performances as it captures non-linear interactions extremely well and handles mixed clinical variables very well by constructing multiple trees independently using bootstrap sampling and aggregating their predictions, thereby reducing variance and improving stability. XGBoost, CatBoost, Gradient Boosting and LightBoost have shown strong performance in predicting preeclampsia by combining multiple weak learners to enhance accuracy, sensitivity, and robustness. Boosting algorithms build trees sequentially, with each new learner focusing on correcting the errors of the previous one. This approach often yields strong predictive performance but increases the risk of overfitting, particularly in the presence of noisy or limited datasets.

Further to want a more flexible non-linear model, hybrid models like Neural network (NN) with and without stacking ensemble based on performances of boosting algorithms which automatically learns complex, hidden relationships and also works well with the many features that contribute weakly.

3.4 Comparative Analysis of Hybrid Models

As shown in **Table 3**, both Stacking Ensemble and NN-based Stacking Ensemble models demonstrated comparable and stable performance across EOPE and LOPE settings. Under EOPE, both achieved moderate results with the value of precision 0.50, recall 0.38, F1-score 0.43, while under LOPE they performed strongly with the value of precision 0.80, recall 0.87, F1-score 0.83 and accuracy 0.74. These findings highlight the robustness and generalizability of stacking-based approaches, which effectively reduce variance and over fitting. In contrast, the standalone Neural Network model exhibited poor generalization under EOPE with all metrics values 0.00, likely due to over fitting, but performed competitively under LOPE with the highest value of recall 1.00 and F1-score 0.85. Overall, stacking ensembles provided more balanced and reliable performance across varying conditions.

Table 3. The performance of the three hybrid models

Hybrid Models	EOPE			LOPE			Accuracy
	precision	recall	f1-Score	precision	recall	f1-Score	
Stacking Ensemble	0.50	0.38	0.43	0.80	0.87	0.83	0.74
NN-Based Stacking Ensemble	0.50	0.38	0.43	0.80	0.87	0.83	0.74
NN without Stacking	0.00	0.00	0.00	0.74	1.00	0.85	0.74

4 DISCUSSION

The comparative analysis of the results among the different ensemble models showed that the RF algorithm consistently outperformed in predicting both early- and late-onset preeclampsia. For RF, the highest overall accuracy of 0.81 and relatively well-balanced performance with respect to precision, recall, and F1-score all being 0.87 under the LOPE setting, while still being robust with respect to generalization in the case of EOPE. This superior performance can be explained by RF's ensemble averaging and low sensitivity to overfitting, which enhance model stability even with moderately sized datasets. Boosting algorithms, in contrast, can provide higher variability across evaluation metrics. For example, GBM provides weak recall under EOPE 0.25, while LightGBM reported the lowest overall accuracy 0.65. XGBoost and CatBoost performing better, performed competitively, yielding F1 scores of 0.84 and 0.81, respectively. They required extensive hyperparameter tuning and were more likely to overfit, particularly with noisy or unbalanced data. The higher AUC values obtained for RF further established its better discriminative capability, while the lower AUCs for LightGBM and XGBoost indicated limited generalization on given dataset.

Feature importance analysis showed that whereas XGBoost selected a smaller number of strong features, particularly platelets, bilirubin D, and SGPT, reflecting its increased sensitivity to distinct biochemical indices. RF and other ensemble methods showed relatively equal importance values, reflecting their moderate attention to several clinical features. This difference indicates that while the power of identifying non-linear relations lies with the XGBoost classifier, the ease of interpretation lies with the RF classifier.

SHAP force plot analysis of misclassified cases demonstrated a systematic bias of all the models toward protective hematological and biochemical features such as platelet count, TLC, BMI, and INR, leading to underestimation of important clinical risk factors including increased systolic blood pressure, pulse rate, and lower maternal age. These reflect the need for feature rebalancing or domain-specific weighting strategies that ensure that clinically important variables are not masked by statistically dominant but less important factors. Analysis of misclassifications using the SHAP force plot showed that there was a consistent bias among the models toward the protective hematological and biochemical variables of platelet count, TLC, BMI, and INR, such that there was a likely underestimate of important clinical risk factors such as systolic blood pressure, pulse rate, and maternal youth. Just how important the best-performing factors can be across the dataset can be demonstrated by the relative balance between the two modes of pathophysiological perturbation.

The comparative research on hybrid models showed that both Stacking Ensemble and NN-based Stacking Ensemble had stable and balanced performances in both EOPE and LOPE settings with F1-score 0.43 and 0.83, respectively. The ability of Stacking Ensemble and NN-based Stacking Ensemble to include multiple base learners reduced variance and enhanced robustness, making them more suitable for clinical applications where reliability is paramount. Conversely, the standalone NN failed to generalize effectively under EOPE, likely due to overfitting on a limited dataset, though it achieved the highest recall 1.00 under LOPE. This contrast underlines the need for hybrid and ensemble learning approaches to achieve reliable clinical prediction outcomes.

5 CONCLUSION

This study has proved the effectiveness of the RF model over the best performing ensemble methods, specifically in the EOPE case, as it is able to achieve better results compared to the other ensemble/hybrid techniques, even though it requires simpler preprocessing. Another benefit of the RF model is its ability to handle the imbalance in the dataset, which is a common issue in real-world datasets requirements, and the increased AUC values confirm the clinical reliability. Even though the boosting algorithms such as XGBoost and CatBoost showed competence with regard to the performance, they were found to be sensitive to the imbalance and noise present in the data. Analysis from the interpretation of features further proved the clinically significant contribution of the major parameters such as the platelet count, bilirubin D value, SGPT, and blood pressure variables. The findings emphasize the value that should be placed on the integration of machine learning-based predictive tools in maternal healthcare in order to provide an opportunity to diagnose high-risk pregnancies early in relation to the case of early-onset preeclampsia. As a result of ensemble learning approaches in healthcare, better diagnosis can be achieved with interventions as well. Future studies should include a broader dataset in multiple populations to better generalize, combining multiple types of input, e.g., clinical, laboratory, and imaging variables, to improve prediction accuracy, and feature weighting/explainable AI techniques may be employed to mitigate biases in machine learning models. The application of the optimized Random Forest model for early detection and intervention of preeclampsia through a real-time decision support system is feasible.

ACKNOWLEDGMENT

We acknowledge with gratitude the support provided by the Department of Obstetrics & Gynaecology, Government Medical College, Chhatrapati Sambhajnagar, Maharashtra 431001, India to the study by giving access to the clinical data as well as cooperating with the study throughout the process. We also thank one of the authors who is affiliated with the department who assisted us with the data preprocessing and analysis process that contributed to the success of the work.

REFERENCES

1. American College of Obstetricians and Gynecologists. (2002). ACOG practice bulletin No. 33: Diagnosis and management of preeclampsia and eclampsia. *Obstetrics & Gynecology*, 99(1), 159–167.
2. American College of Obstetricians and Gynecologists, & Task Force on Hypertension in Pregnancy. (2013). Hypertension in pregnancy: Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstetrics & Gynecology*, 122(5), 1122–1131.
3. Aksornphusitaphong, A., & Phupong, V. (2013). Risk factors of early and late onset pre-eclampsia. *Journal of Obstetrics and Gynaecology Research*, 39(3), 627–631.
4. Bozdag, H., Ogutcuoglu, F. B., Guzin, K., Kilic, S. R., Duran, E. A., & Aydin, I. T. (2015). The frequency and fetomaternal outcomes of early- and late-onset preeclampsia: Experience of a single tertiary health center in Istanbul. *Medeniyet Medical Journal*, 30(4), 163–169.
5. F. Pettit, Mangos, G., Davis, G., Henry, A., & Brown, M. A. (2015). Pre-eclampsia causes adverse maternal outcomes across the gestational spectrum. *Pregnancy Hypertension*, 5(2), 198–204.

6. Gohar, S., & Syed, W. (2021). Comparison of adverse fetomaternal outcome in early and late onset pre-eclampsia. *Journal of Postgraduate Medical Institute*, 35(1), 15–18.
7. Hellman, L. M., & Pritchard, J. A. (2014). *Williams obstetrics* (14th ed.). World Health Organization.
8. Jhee, J. H., et al. (2019). Prediction of pregnancy outcomes using machine learning: A retrospective cohort study. *PLOS ONE*.
9. Khan, S., Aleem, N. T., Hasna, A. U., Husein, S., Khatun, K., & Ara, R. (2013). A clinical study on risk factors and fetomaternal outcome of pre-eclampsia in Bangladesh. *Journal of the Bangladesh Society of Anaesthesiologists*, 26(1), 39–44.
10. Kolluru, V., Harika, R. Y., & Kaul, R. (2016). Maternal and perinatal outcomes associated with pregnancy-induced hypertension. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, 5(10), 3367–3372.
11. Kooffreh, M. E., Ekott, M., & Ekpoudom, D. O. (2010). The prevalence of preeclampsia among pregnant women in the University of Calabar Teaching Hospital, Calabar. *Saudi Journal of Health Sciences*, 3(3), 133–136.
12. Liu, L., et al. (2022). Machine learning models for early prediction of preeclampsia: A systematic review. *Frontiers in Public Health*.
13. Madazli, R., Yuksel, M. A., Imamoglu, M., Tuten, A., Oncul, M., et al. (2014). Comparison of clinical and perinatal outcomes in early- and late-onset preeclampsia. *Archives of Gynecology and Obstetrics*, 290(1), 53–57.
14. Madhuri, C., & Varalakshmi, Y. (2019). Retrospective study on fetomaternal outcome in gestational hypertension, pre-eclampsia and eclampsia in a tertiary care centre. *Indian Journal of Basic and Applied Medical Research*, 8(4), 246–255.
15. Magee, L. A., Pels, A., Helewa, M., Rey, E., & von Dadelszen, P. (2014). Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy. *Pregnancy Hypertension*, 4(2), 105–145.
16. Paruk, F., & Moodley, J. (2000). Maternal and neonatal outcome in early- and late-onset pre-eclampsia. *Seminars in Neonatology*, 5(3), 197–207.
17. Poon, L. C., & Nicolaides, K. H. (2014). Early prediction of preeclampsia. *Obstetrics and Gynecology International*, Article ID 297397.
18. Publications Committee, Society for Maternal-Fetal Medicine, & Sibai, B. M. (2011). Evaluation and management of severe preeclampsia before 34 weeks' gestation. *American Journal of Obstetrics and Gynecology*, 205(3), 191–198.
19. Sibai, B. M. (1999). Hypertension in pregnancy. *Clinical Obstetrics and Gynecology*, 421–436.
20. Sovio, U., et al. (2020). Prediction of preeclampsia using first-trimester biomarkers and machine learning models. *Hypertension*.
21. Stergiotou, I., Crispi, F., Valenzuela-Alcaraz, B., Bijnsens, B., & Gratacos, E. (2013). Patterns of maternal vascular remodeling and responsiveness in early- versus late-onset preeclampsia. *American Journal of Obstetrics and Gynecology*, 209(6), 558.e1–558.e14.
22. Thangaratinam, S., et al. (2011). Prediction models for pre-eclampsia: A systematic review. *BMJ*.
23. Tranquilli, A. L., Brown, M. A., Zeeman, G. G., Dekker, G., & Sibai, B. M. (2013). The definition of severe and early-onset preeclampsia: Statements from the

- International Society for the Study of Hypertension in Pregnancy (ISSHP). Pregnancy Hypertension, 3(1), 44–47.
24. Tranquilli, A. L., Dekker, G., Magee, L., et al. (2014). The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. Pregnancy Hypertension, 4(2), 97–104.
 25. World Health Organization. (2018). Maternal mortality. Retrieved March 24, 2018, from <http://www.who.int/mediacentre/factsheets/fs348/en>
 26. Ye, C., et al. (2020). Prediction of pregnancy-related hypertensive disorders using machine learning. BMC Medical Informatics and Decision Making.
 27. You, S.-H., Cheng, P.-J., Chung, T.-T., Kuo, C.-F., Wu, H.-M., & Chu, P.-H. (2014). Population-based trends and risk factors of early- and late-onset preeclampsia in Taiwan.