DOI: 10.7860/JCDR/2026/82583.23356

Review Article

Obstetrics and Gynaecology Section

From Data to Diagnosis: The Role of Machine Learning in Preeclampsia Prediction

R SAROASH ZULFISHAAN[1], R ANUSHA[2], PREET AGARWAL[3], LEENA CHAND[4]

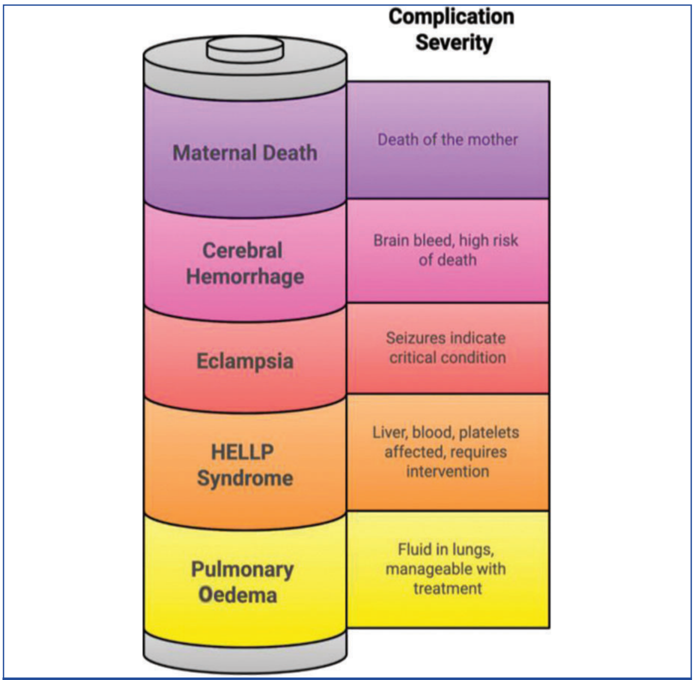
ABSTRACT

Preeclampsia (PE) is a complex, potentially life threatening hypertensive disorder of pregnancy, that affects approximately 2-5% of pregnancies worldwide. It is characterised by elevated Blood Pressure (BP), Proteinuria and multiorgan dysfunction, posing significant risks to both maternal and foetal health. Timely identification of high-risk pregnancies is crucial to prevent severe complications. However, conventional diagnostic approaches relying on clinical assessments and biomarkers analysis often fall short in early and accurate prediction. Recent advancements in Machine Learning (ML) have opened new avenues for improving early prediction and diagnosis of PE, offering the potential for more targeted and timely interventions. This review explores the integration of ML algorithms with clinical, laboratory and demographic data to develop robust predictive models for PE. Studies employing Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF) and Deep Neural Networks (DNN) has demonstrated the ability to uncover complex non linear patterns associated with PE risk. These models often surpass traditional diagnostic methods in accuracy and offer transformative potential in prenatal care. By incorporating ML based tools into routine obstetrics practice, it can help healthcare providers predict PE early and implement preventive strategies to reduce the risk of complications such as eclampsia and organ failure. This review offers a comprehensive, evidence-based review of ML models in PE prediction by integrating clinical, biochemical, imaging and genomic data. By bridging obstetrics and data sciences, it highlights model performance, personalisation, and clinical relevance, making it a timely and translational resource for improving maternal-foetal outcomes through AI driven healthcare.

Keywords: Early prediction, Predictive models, Reproductive health (SDG 3)

INTRODUCTION

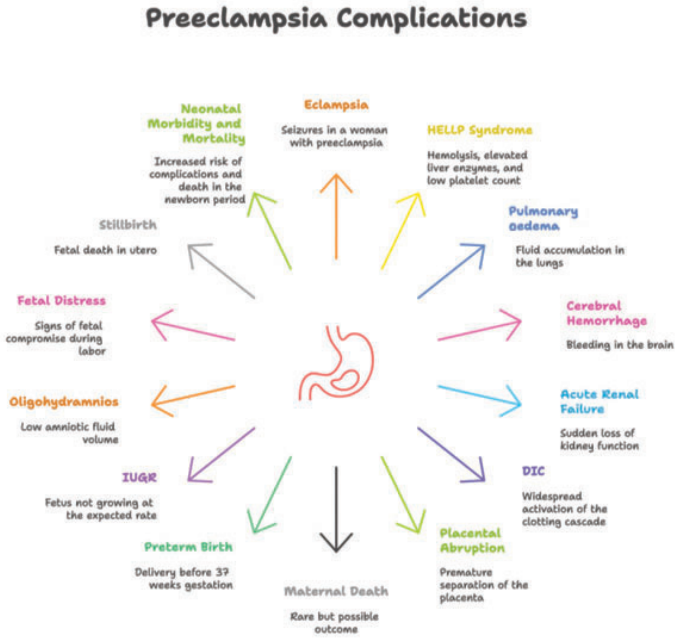
Understanding the underlying cause of the disease remains one of the central pursuits of modern medicine, PE, a multifaceted and elusive hypertension disorder of pregnancy, continues to challenge clinicians and researchers alike. Often referred to as the "Disease of Theories", PE exemplifies the complexity of pregnancy related conditions, falling under the broader category of 'Great Obstetrical Syndrome' [1]. These syndromes are characterised by the convergence of multiple potentially overlapping pathological processes that culminate in a common clinical presentation [2]. As illustrated in [Table/Fig-1], the extensive maternal and foetal morbidity are associated with PE complications such as HELLP Syndrome, renal failure, placental abruption, foetal distress, intrauterine growth restriction (IUGR), and stillbirth emphasise that the condition is not confined to BP alone but affects virtually every organ system. Similarly [Table/Fig-2] highlights the complications of PE span a wide spectrum - from manageable conditions such as pulmonary oedema to life threatening outcomes like eclampsia, cerebral haemorrhage and even maternal death. This increase of severity visually shows how a single disease can escalate into diverse multisystem complications. PE shares certain physiological patterns with other obstetrical conditions such as preterm labour particularly through the involvement of the final common parturition pathway [3]. This pathway is typified by increased uterine contractility, cervical remodelling and activation of the membrane and decidua. In case of PE, the pathophysiology primarily involves syncytiotrophoblast stress, systemic endothelial cell activation and a heightened intravascular inflammatory response [2,4]. However, emerging clinical guidelines, as proposed by leading professional organisations, have redefined the diagnostic criteria. It is now recognised that in the context of multisystem involvement, PE can be diagnosed even in the absence of the proteinuria, reflecting a broader understanding of its systemic nature and diverse manifestation [5].



[Table/Fig-1]: Complications of PE (Image created by the authors using Napkin AI based on literature derived concepts).

Importance of Early Prediction in Maternal and Foetal Health

Early prediction of PE is crucial for improving maternal and foetal outcomes, as the condition can quickly progress to life threatening



[Table/Fig-2]: Systemic impact of PE: PE affects multiple organ systems, causing maternal complications such as hypertension, kidney dysfunction, stroke and anaemia along with foetal risks including growth restriction and preterm birth (Image created by the authors using Napkin AI based on literature derived concepts).

stages if not identified and managed promptly. Timely prediction enables preventive interventions, reducing severe maternal complications and adverse neonatal outcomes. Accurate risk stratification in early pregnancy can enable vigilant monitoring and proactive care [6,7].

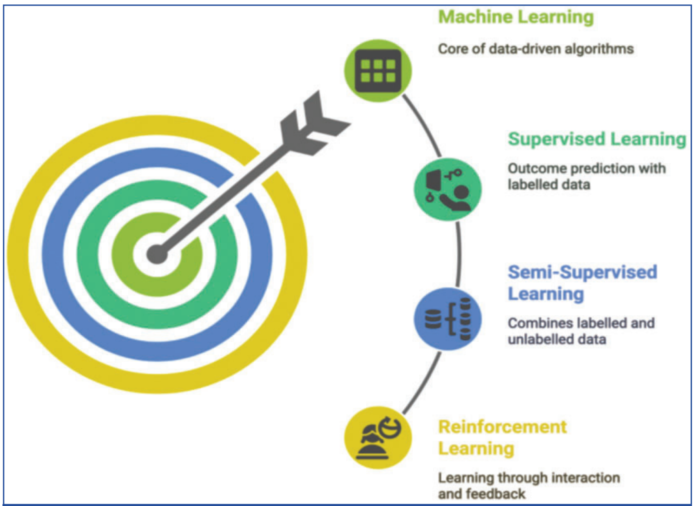
Uterine artery doppler ultrasound in the first trimester is a key non invasive tool that assesses uteroplacental blood flow and predicts PE risk [8]. Studies report that abnormal doppler findings correlate strongly with later PE development with high specificity and sensitivity, integrating maternal demographic factors with biochemical and biophysical markers further improve prediction models [7-9]. Research shows combining maternal history with presentation assessments enhances early identification of high-risk women, enabling personalised interventions [7,8].

Recently, ML has emerged as a promising approach in obstetrics, overcoming limitation of traditional statistic models, ML algorithm process, higher dimensional, heterogeneous data, including biochemical markers, imaging and clinical records, to uncover complex, non-linear patterns enabling more accurate risk stratification [8,10].

The ML, a branch of AI, develops algorithms that learn from data to identify patterns and make predictions without explicit programming [11]. It includes unsupervised, supervised, semi supervised and reinforcement learning, each suited to specific data types and predictive goals [Table/Fig-3]. In PE, prediction ML enables integration of diverse data sources, revealing intricate risk factor interactions and supporting individualised early interventions. This positions ML as a transformative tool in prenatal care and prevention strategies [8,10,11].

Deep Learning Architecture

Recent advancements in ML have given rise to more sophisticated and efficient computational methodologies, among which deep learning stands out as a transformative development. Deep Learning represents a specialised branch of ML that leverages multilayer artificial neural networks to model complex, high dimensional data [12]. It has also revolutionised various fields, particularly computer vision, and natural language processing by significantly enhancing the performance of tasks that require advanced pattern recognition. Key architectures within deep learning such as Convolutional Neural networks (CNNs) and Recurrent Neural Networks (RNNs) have enabled the effective processing of structured and unstructured data



[Table/Fig-3]: Overview of the three main types in Machine Learning paradigms: Supervised, unsupervised, and reinforcement learning (Image created by the authors using Napkin AI based on literature derived concepts).

alike [10,11]. CNNs are especially effective in image Recognition and classification due to their ability to spatial hierarchies in visual data [13]. Meanwhile, RNNs, with their feedback connections, are well suited for sequential data processing, making them ideal for applications such as time-series analysis, language modeling and speech recognition. In modern artificial intelligence deep learning serves as a foundational technology [12,13]. Its capability to extract meaningful representations from raw data has led to substantial improvements in accuracy across a range of domains, including medical diagnostics, autonomous systems and predictive analytics. These innovations have particular relevance for healthcare applications, where the interpretation of complex biomedical data is critical for early disease prediction and personalised treatment strategies [12-14].

Automated ML (AutoML)

Automated ML represents a major advancement in making ML more accessible and efficient. By automating key steps such as feature selection and hyper parameter tuning. AutoML allows even non experts to build effective predictive models [15]. This reduces development time, improves performance and lowers the barrier to adopting AI driven solutions across diverse industries. In healthcare, including PE prediction, AutoML enables rapid creation of robust models without deep expertise in data sciences, supporting clinicians and researchers in developing reliable, data driven tools for improved patient care [15-17].

Advantages of ML in PE Prediction: Prediction Accuracy

One of the most significant advantages of applying ML in PE prediction is its ability to improve predictive accuracy by analysing complex datasets. Unlike traditional statistical approaches, which rely on predefined assumptions and linear relationships ML algorithms can uncover subtle, non linear interactions among diverse risk factors [18]. This enables nuanced modelling of PE risk and supports earlier, more reliable identification of high-risk pregnancies. ML techniques are particularly well suited for integrating heterogeneous data, including maternal demographic characteristics, biophysical assessments (such as uterine artery Doppler measurements), and biochemical markers [19]. Recent studies have demonstrated that by combining these features in ML based models significantly enhance predictive performance, often outperforming conventional clinical models [20]. These algorithms can identify early warning signs and subtle indicators that may otherwise be overlooked by standard diagnostic protocols, thereby facilitating timely clinical interventions. Furthermore, the continuous evolution of ML tools and their incorporation into modern healthcare workflows contributes to more precise risk stratification and personalised monitoring

[21]. As the volume and variety of healthcare data increase ML it offers scalable and adaptive solutions that enhance diagnostic precision, optimise patient outcomes and support evidence-based decision making prenatal care. Early identification of at-risk individuals enables connections to intervene during the initial stages of pregnancy. Therefore improving both maternal and foetal outcomes, ML based models demonstrate a superior capacity to identify vulnerable patients [22]. By analysing large volumes of clinical demographic and biochemical data to prove collected during early gestation [22,23]. Unlike conventional risk assessment methods, which are limited to a narrow set of known risk factors such as high BP, proteinuria, previous history of PE, obesity, diabetes and maternal age, ML algorithms can detect complex non obvious patterns associated with disease progression. One key benefit of this early detection is the opportunity to initiate evidence-based preventive measures [24]. For instance, low dose aspirin before 16 weeks reduces PE risk in women identified through traditional screening, which relies on key maternal risk factors such as PE, chronic hypertension, diabetes, renal or autoimmune disease, multifoetal gestation, obesity, advanced age, and nulliparity and basic early pregnancy tools like maternal history assessment, mean arterial pressure and uterine artery doppler [24]. ML models that integrate maternal history, biophysical and biochemical indicators enable more precise risk stratification and improved identification of candidates for such interventions [23]. This targeted approach enhances this efficient clinical decision-making. Supporting the development of a personalised treatment plan, moreover, early interventions strategies such as Aspirin therapy not only improve maternal and neonatal outcomes also help to reduce the burden on theme healthcare systems, preventing complications like preterm birth, severe hypertension and organ dysfunction [25]. The continued advancement of ML techniques is reshaping obstetric care by improving early diagnosis, strengthening risk management, and supporting more proactive, personalised models of maternal health [24,25].

Personalised Risk Assessment using ML in Maternal Healthcare

The ML has revolutionised risk evaluation by shifting from standard models to individualised predictions. Unlike traditional methods, ML algorithms integrate genetic, clinical, and demographic data, to generate more precise risk assessment [26]. In maternal healthcare, ML models excel at identifying complications like PE by analysing complex risk factors, enabling targeted monitoring and interventions [27]. Personalised prenatal care including lifestyle changes and low dose aspirin can reduce complications severity and improve outcomes. ML driven assessments also enhance the healthcare efficiency by optimising resources and reducing unnecessary procedures. Future progress depends on collaboration between healthcare professionals and data scientists to refine models improve detection and support better pregnancy outcomes [26,27].

ML Models for PE Prediction

Recent ML advances show promise in early PE detection, improving maternal and neonatal outcomes; several studies report strong predictive performance. Here, recent studies are highlighted that use the ML models in PE prediction [Table/Fig-4] [28-41].

Zhang X et al., conducted a retrospective study at West China Second University Hospital, leveraging ML to identify early biomarkers of severe PE between 8th to 20th weeks of gestation [28]. Their dataset encompassed blood test results from 19,653 pregnant individuals collected between 2017 and 2019. From this cohort, 248 subjects were selected, 124 diagnosed with severe PE, and 124 matched controls based on demographic characteristics, the researchers applied a Light Gradient Boosting machine (GBM) and two additional ML models to analyse 43 blood based variables using 35% of the data set for internal validation, the light GBM model outperformed the others, achieving an area under the receiver operating characteristic curve, AUC of 89.74% with a sensitivity of 88.37% and specificity of 77.27%. Key predictive biomarkers identified, including direct bilirubin, aspartate

Author	Work done	ML Models used	Key data/Features	Highlights
Zhang X et al., [28]	Early detection of severe PE using blood indices from 19,653 patients	Light Gradient Boosting machine, DT, RF	Haematological markers	Achieved strong performance in identifying severe PE
Marić I et al., [29]	Screening tool using routine prenatal data (16,370 births)	Gradient boosting, elastic net	Clinical visit data	Efficient early risk stratification
Liu M et al., [30]	PE prediction from clinical/lab data (11,152 patients)	LR, DNN, SVM, RF, DT	18 clinical variables	RF achieved best results (AUC 0.86)
Li YX et al., [31]	PE risk model from 3,759 cases	RF, LR, SVM, Extreme Gradient Boosting (XGBoost)	Clinical parameters	XGBoost yielded AUC of 0.955
Marin I et al., [32]	Smart bracelet-based monitoring with mobile app	Viterbi, HMM	BP, age, weight	Accuracy: 80%, Sensitivity: 92.5%
Schmidt LJ et al., [33]	Biomarker and sonographic-based PE prediction (1,647 patients)	RF, Gradient Boosted Trees	PIGF, sFlt-1, ultrasound	Strong model interpretability using 10-fold CV
Liu L et al., [34]	Multicentre validation of wearables	Random Forest, LSTM	HR, BP, activity and SpO_2	AUC 0.89; 78% fewer visits; r=0.91 BP correlation
Sufriyana H et al., [35]	Prediction using Doppler and biomarker data	Regression-based Model	sFlt-1, Placental Growth Factor (PLGF), Doppler	AUC 0.97, Sensitivity 95%
Edvinsson C et al., [36]	Severe PE risk modeling (81 patients)	Logistic Regression (LR)	BMI, uric acid, ASAT	AUC 0.91, Accuracy 88%
Kovacheva VP et al. [37]	Genomic+EHR fusion for PE risk	LR, XGBoost	Polygenic Risk Scores (PRS)	AUC 0.91 (late pregnancy)
Wesson JL and Smith N [38]	South African population-based model	LR, Gradient Boosting	BP, BMI, diabetes, age	Effective in early prediction context
Serra B et al., [39]	First-trimester PE screening (6,893 pregnancies)	Gaussian model	PGF, MAP, PI, maternal data	AUC 0.96, false positive rate 5-10%
Melinte-Popescu AS et al., [40]	Prospective study of 233 women	DT, Naive Bayes, SVM, RF	Biomarkers, ultrasound, maternal risk	DT had 94.1% accuracy for early-onset PE
Jhee JH et al., [41]	Late-onset PE prediction (11,006 cases)	LR, DT, NB, SVM, RF, SGB	BP, BUN, creatinine, platelets	SGB model: 97.3% accuracy, FPR 0.9%

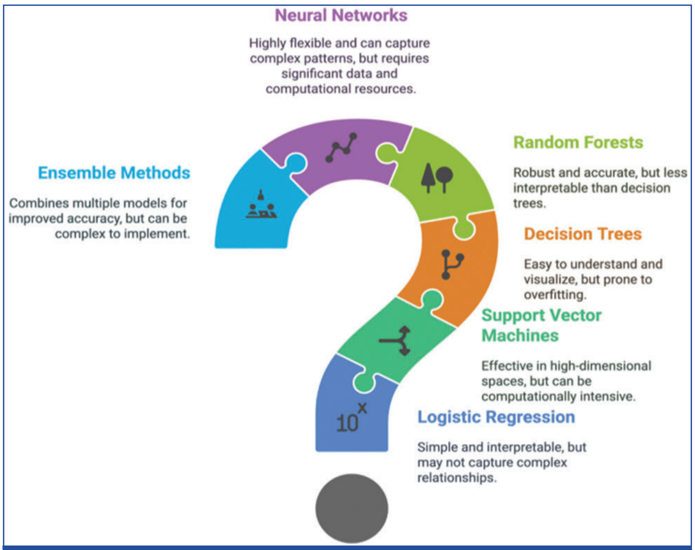
[Table/Fig-4]: A review of key studies 2019-2024.
DT: Decision tree; RF: Random forest; LR: Logistic regression; DNN: Deep neural network; SVM: Support vector machine; HMM: Hidden markov model; SGB: Stochastic gradient boosting

aminotransferase, and activated partial thromboplastin time ratio. These findings underscore the potential of ML to facilitate rapid and accurate prediction of severe PE to routine blood biomarkers supporting earlier clinical intervention. However, the authors noted the external validation using larger and more diverse population is essential to enhance the model's generalisability and clinical applicability [28]. Similarly, another pivotal study by researchers at Lucile packard children hospital explored ML driven approaches to identify predictors of early onset PE using Electronic Medical Records (EMRs) [29]. The study analysed 16,370 pregnancy cases between April 2014 and January 2018, integrating 67 features, including patient demographics, obstetrics history, clinical lab results and treatment records. The auto developed models using gradient boosting and elastic net regularisation techniques. With an AUC of 0.79, a false positive rate of 8.1%, and a sensitivity of 45.2%, the elastic net model performed the best overall. Interestingly, the early onset PE subgroup, which comprised 98 patients, about 1.9% of the entire cohort, produced an AUC of 0.89, or (95% CI: 0.84-0.95), a true positive rate of 72.3%, and an FPR of 8.8%. These results highlight the enhanced predictive potential of ML when applied to structured EMR data, especially in identifying high-risk individuals early in pregnancy. The study concluded that Integrating ML models into prenatal screening workflows could facilitate earlier diagnosis and allow for the implementation of personalised surveillance and preventive care strategies. Nevertheless, the author stresses the importance of validating their models across different geographic and demographic populations to ensure robustness and fairness in diverse clinical environments. Together these studies exemplify how ML methodologies can uncover complex associations among clinical variables that might be overlooked by conventional statistical tools by improving the accuracy of early PE prediction. ML not only empowers screening with timely insights, but also provides a shift towards personalised and proactive maternal healthcare [29].

A Laboratory Database Models

The ML models leveraging clinical and laboratory data have shown promising results in predicting PE, especially when applied to enlarge, real world patient data sets. These models are particularly valuable due to their ability to analyse routinely collected clinical variables and offer scalable diagnostic solutions. Liu M et al., developed a PE prediction framework using healthcare records from 11,152 pregnant individuals treated at Jinan university hospital between December 2015 to September 2019, their data set included 143 confirmed PE cases, 95 gestational hypertension cases and 10,914 on for normotensive pregnancies [30]. Models SVM, LR, DNN, DT, and RF. Among these, the RF model showed the best performance achieving an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.86 predictive performance with the accuracy of 74% recall of 42% precision of 82%, an area under the receiver operating characteristic curve, AUC of 0.86, and a brier score of 0.17, these results underscore the potential RF models in effectively utilising standard clinical data for PE risk stratification [30]. [Table/Fig-5] shows representative machine learning algorithms.

In another study, Li YX et al., applied a suite of ML models-RF, SVM, LR, and Extreme Gradient Boosting (XGBoost)- to predict PE using clinical data from 3,759 pregnant patients at Xinhua Hospital between July 2016 and December 2019 [31]. Among the tested models, XGBoost exhibited the highest predictive accuracy, by achieving an AUC of 0.955, an F1-score of 0.571, recall of 78.9%, precision of 44.7%, and an overall accuracy of 92%. The authors emphasised the significance of feature important analysis in identifying critical predictors of PE, which can enhance model interpretability and clinical utility [32]. These findings demonstrate how well ML models- particularly ensemble approaches like XGBoost and RF- predict PE based on clinical and laboratory data, allowing for the early detection of pregnancies at risk and facilitating their inclusion into standard care.



[Table/Fig-5]: Representative Machine Learning (ML) algorithms. ML encompasses supervised, unsupervised, and reinforcement learning approaches. These paradigms enable tasks such as prediction, pattern recognition, and decision-making, with broad applications in healthcare, including preeclampsia risk modelling. (Image created by the authors using Napkin AI, based on concepts derived from the literature).

Wearable Technology and Remote Monitoring

Integrating wearable devices with ML enables real time, remote monitoring of maternal health improving early PE detection and enhancing traditional settings Marin I et al., conducted a study utilising the viterbi ML algorithms, which is grounded in Hidden Markov Model (HMM) principles to evaluate the effectiveness of wearable technology in predicting PE [32]. The study involved 105 pregnant participants who wore an i-bracelet device designed to measure and transmit BP data via Bluetooth to a mobile application. This application processed, achieving an accuracy of 80%, a sensitivity of 92.5%, and a specificity of 72%. These results underscore the utility of combining monitoring of PE risk. Such system can offer proactive alerts and facilitate timely intervention, especially in resource-limited settings or for patients with restricted access to frequent in clinic visits [33]. Wearables with ML capabilities have the potential to revolutionise prenatal care by enabling remote, customised risk assessment; however, further data and broader validation are required to bolster these models.

Biomarkers and Sonographic Measurements based Models

Recent advances in PE prediction show that biomarkers and sonographic measurements improve the accuracy of ML models, enabling precise risk stratification by combining biochemical and imaging data.

Schmidt LJ et al., conducted a comprehensive study using 2472 clinical samples obtained from 1,647 pregnant women at the charite universitatsmedisin Obstetrics, department spanning July 2010 to March 2019 [33]. The data set included 114 clinical features notably, the biomarkers PIGF and soluble fms like tyrosine kinase 1 along with first and second trimester samples primarily between 11th to 22nd weeks of gestation. In conjunction with sonographic measurements to analyse, this high dimensional data set the researchers Employed RF classifiers and gradient boosted trees, performance evaluation was carried out through a robust methodology, 10 fold cross validation on a 10 × 10 data split. The study demonstrated that integrating biomarkers and ultrasound data significantly enhanced the ability to predict high-risk pregnancies, requiring clinical interventions underscoring the diagnostic utility of these features [33].

Doppler Measurements and Clinical Risk Factors

Doppler ultrasound assessment, when combined with maternal clinical risk factors and biochemical markers, significantly enhances the predictive performance of ML models in detecting

PE. These models offer a non invasive, data rich approach to improving the identification of at risk pregnancies. Sufriyana H et al., developed an advanced ML prediction model, incorporating maternal characteristics uterine doppler measurements and angiogenic biomarkers such as soluble fms-like tyrosine kinase I and Placental Growth Factor (PLGF) measured between 24 and 37 weeks of gestation the model developed using the weka software [35]. Suite underwent 10 fold cross validation and demonstrated strong predictive capabilities with an Area Under the Curve (AUC) of 0.97. It achieved 100% specificity and 95% sensitivity with a mature correlation coefficient of 0.93, these findings underscore the substantial value of Doppler metrics and biomarker integration in identifying pregnancies at high risk for PE [35].

In a related study, Edvinsson C et al., constructed a ML model that integrated worth standard clinical indicators and routine biomarkers collected during mid-pregnancy between 20th to 28th weeks of gestation among a cohort of 81, pregnant patients, including 41 cases of severe PE [36]. Key predictors, identified in this model were Body Mass Index (BMI), serum uric acid and aspartate aminotransferase, the cross validated model achieved an AUC of 0.91 and an overall accuracy of 88%. This study confirmed that even commonly measured clinical parameters can serve as powerful predictors when leveraged through ML algorithms [36].

Genomic Data and Electronic Health Records Integration

Integrating genomic data with electronic health records offers a promising approach to enhancing personalised risk stratification of PE. Kovacheva VP et al., conducted a study involving 1,125 pregnant women who delivered at Mass General Brigham Hospital between May 2015 and May 2022 [37]. Their research utilised both LR and eXtreme Gradient boosting (XGBoost) algorithms to predict PE risk by analysing Polygenic Risk Scores (PRS) derived from systolic BP related genetic variants. XGBoost models applied to first trimester laboratory data (between 11th to 13+6 weeks of gestation) achieved an AUC of 0.74, While models applied to later pregnancy stages (≥20 weeks) reached an AUC of 0.91. These results demonstrated that integrating genetic testing with longitudinal clinical data from EHRs can significantly improve PE risk prediction accuracy. The study highlights the feasibility of developing individualised PE screening tools that combine genetic predisposition, with dynamic clinical variables [37].

Demographic and Maternal Health Characteristics

Demographic data and maternal health history remain fundamental to PE risk assessments, especially in regions where access to advanced biomarker and genomic testing maybe limited. Wesson JL and Smith N investigated the prediction of PE within the South African obstetrics population, the study identified eight key predictors, systolic and diastolic pressure, maternal age, BMI, diabetes history of hypertension, nulliparity, and pre-existing maternal disease [38]. ML models trained on this dataset showed strong predictive performance, particularly when used during early pregnancy (before 20th weeks of gestation), the finding underscored the importance of region specific model development and the inclusion of a localised epidemiological data in pe risk prediction strategies [38].

In another large scale study, Serra B et al., developed a multivariate Gaussian distribution model integrating maternal risk factors, mean arterial pressure, uterine artery pulsatility index, and PlGF to predict early onset PE [39]. The analysis was conducted on data from 6893 Singleton pregnancy collected during first trimester screening (11th to 13th weeks of gestation), the model achieved an AUC of 0.96 and detected 59% of early onset PE cases at a 5% false positive rate and 94% at a 10% false positive rate. These results provide compelling evidence that routine PE screening protocols should incorporate placental biomarkers alongside standard material assessment to enhance early detection and intervention planning [39].

Subtype Specific Prediction Models and Clinical Applications

A prospective study by Melinte-Popescu AS et al., examined 233 Romanian pregnant women to forecast PE ultrasound assessment and clinical subtypes using ML algorithms, the researchers integrated data from system maternal risk profiles, ultrasound assessments and serum biomarker collected during the second trimester (between 20th to 28th weeks of gestation) analysis to train multiple predictive models among the evaluated methods, the DT algorithm achieved a notable 94.1% accuracy in predicting early onset, PE in additional the RF and naive bayes classifiers demonstrated high overall performance across all PE subtypes [40]. This study underscores the clinical value of ML based screening frameworks that incorporate paraclinical elements to enhance predictive capabilities for PE and its variants [40].

Jhee JH et al., extended this approach by focusing on the prediction of late onset PE using data from 11,006 pregnant women treated at Yongsan university hospital. This was a retrospective study in which clinical and laboratory data from routine antenatal visits were analysed. The cohort included women who later developed late onset PE as a case group (between 20th to 28th weeks) and normotensive pregnancies as a comparative control group, allowing the performance, 97.3% accuracy and maintained a low false positive rate of 0.9%. Key predictive variables included systolic BP, blood urea, nitrogen reaction, platelet count and urinary protein levels factors commonly available in standard clinical practice. These findings support the feasibility of deploying robust ML Models in real world clinical settings to identify patients at risk for late-onset PE enabling earlier and more targeted interventions [41].

This review highlights the growing potential of ML in predicting PE, demonstrating its ability to integrate diverse data sources, clinical biochemical and genomic imaging to build robust predictive models, ensemble algorithms such as RF, gradient boosting, and XG boost have consistently shown strong predictive accuracy underscoring the capacity to handle complex high dimensional healthcare data. However, the generalisability of models across different populations is still limited, highlighting the need for large-scale multicentric validation studies.

CONCLUSION(S)

The ML models show strong promise in early PE prediction before 20th weeks of gestation using clinical, laboratory, biomarker, genetic and demographic data. Algorithms like RF, XGBoost, LightGBM consistently deliver high accuracy. Innovation such as wearable devices and genomic integration expand opportunities for personalised, real-time monitoring. Preprocessing strategies, including normalisation and outlier detection improve the model reliability. Despite these advances, challenges remain in generalisability and clinical adoption. Future work should emphasise external validation, interpretability and seamless integration into healthcare workflows with continued progress. ML based tools can significantly improve maternal and foetal outcomes through timely, accurate and personalised PE risk assessment.

Acknowledgement

The author(s) would like to thank all the colleagues and reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

REFERENCES

[1] Rakityanskaya EA, Panova IA, Malyshkina AI, Fetisova IN, Fetisov NS, Kharlamova NV, et al. Technologies for prediction of PE. Sovren Technol Med. 2020;12(5):78.
[2] Jung E, Romero R, Yeo L, Gomez-Lopez N, Chaemsaithong P, Jaovisidha A, et al. The etiology of preeclampsia. Am J Obstet Gynecol. 2022;226(2 Suppl):S844-S66.
[3] Bisson C, Dautel S, Patel E, Suresh S, Dauer P, Rana S. Preeclampsia pathophysiology and adverse outcomes during pregnancy and postpartum. Front Med (Lausanne). 2023;10:1144170.

[4] Aljameel SS, Alzahrani M, Al Musharraf R, Altukhov M, Alshaia S, Sahloul H, et al. Prediction of PE using machine learning and deep learning models: A review. Big Data Cogn Comput. 2023;7(1):32.
[5] Ranjbar A, Montazeri F, Ghamsari SR, Mehrnoush V, Roozbeh N, Darsareh F. Machine learning models for predicting preeclampsia: A systematic review. BMC Pregnancy Childbirth. 2024;24(1):6.
[6] Cnossen JS, Vollebregt KC, De Vrieze N, Ter Riet G, Mol BW, Franx A, et al. Accuracy of mean arterial pressure and blood pressure measurements in predicting pre-eclampsia: Systematic review and meta-analysis. BMJ. 2008;336(7653):1117-20.
[7] Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: A systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth. 2022;22(1):348.
[8] Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, Hosseinzadeh M, et al. Machine learning (ML) in medicine: Review, applications, and challenges. Mathematics. 2021;9(22):2970.
[9] Leslie K, Thilaganathan B, Papageorghiou A. Early prediction and prevention of pre-eclampsia. Best Pract Res Clin Obstet Gynaecol. 2011;25(3):343-54.
[10] Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutas V, Atiah F, et al. A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl Based Syst. 2020;194:105596.
[11] Tufail S, Riggs H, Tariq M, Sarwat AI. Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms. Electronics. 2023;12(8):1789.
[12] Li T, Xu M, Wang Y, Wang Y, Tang H, Duan H, et al. Prediction model of preeclampsia using machine learning based methods: A population based cohort study in China. Front Endocrinol (Lausanne). 2024;15:1345573.
[13] Layton AT. Artificial intelligence and machine learning in preeclampsia. Arterioscler Thromb Vasc Biol. 2025;45(2):165-71.
[14] Malik V, Agrawal N, Prasad S, Talwar S, Khatuja R, Jain S, et al. Prediction of preeclampsia using machine learning: A systematic review. Cureus. 2024;16(12).
[15] Torres-Torres J, Villafan-Bernal JR, Martinez-Portilla RJ, Hidalgo-Carrera JA, Estrada-Gutierrez G, Adalid-Martinez-Cisneros R, et al. Performance of machine-learning approach for prediction of pre-eclampsia in a middle-income country. Ultrasound Obstet Gynecol. 2024;63(3):350-57.
[16] Parrales-Bravo F, Caicedo-Quiroz R, Rodríguez-Larraburu E, Barzola-Monteses J. ACME: A classification model for explaining the risk of preeclampsia based on Bayesian network classifiers and a non-redundant feature selection approach. Informatics. 2024;11(2):31.
[17] Tiruneh SA, Rolnik DL, Teede H, Enticott J. Temporal validation of machine learning models for pre-eclampsia prediction using routinely collected maternal characteristics: A validation study. Comput Biol Med. 2025;191:110183.
[18] Munyao MM, Maina EM, Mambo SM, Wanyoro A. Real-time pre-eclampsia prediction model based on IoT and machine learning. Discov Internet Things 2024;4(1):10.
[19] Hennessy A, Tran TH, Sasikumar SN, Al-Falahi Z. Machine learning, advanced data analysis, and a role in pregnancy care? How can we help improve preeclampsia outcomes? Pregnancy Hypertens. 2024;37:101137.
[20] Kravtsova OM, Kuznetsov PA, Dobrokhotova YE, Raba DP. Prediction of preeclampsia: A retrospective study. Gynecol. 2025;27(1):12-15.
[21] Ansbacher-Feldman Z, Syngelaki A, Meiri H, Cirkin R, Nicolaides KH, Louzoun Y. Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers. Ultrasound in Obstetrics & Gynecology. 2022;60(6):739-45.
[22] Wei X, Li Q, Wu F, Qin C, Dong J, Chen C, et al. Development and validation of preeclampsia predictive models using key genes from bioinformatics and machine learning approaches. Front Immunol. 2024;15:1416297.
[23] Wang H, Zhang Z, Li H, Li J, Li H, Liu M, et al. A cost-effective machine learning-based method for preeclampsia risk assessment and driver genes discovery. Cell Biosci. 2023;13(1):41.
[24] Swathikrishna MR, Sriram S, Subha B. Preeclampsia risk prediction using machine learning algorithms. In: Joshi, A., Mahmud, M., Ragel, R.G., Karthik, S. (eds) ICT: Innovation and Computing. ICTCS 2023. Lecture Notes in Networks and Systems, vol 879. Springer, Singapore. 2023; 71-80. https://doi.org/10.1007/978-981-99-9486-1_5
[25] Gomathi R, Menaka K. A support vector machine classifier approach for predicting preeclampsia and gestational hypertension. In: Vimal, V., Perikos, I., Mukherjee, A., Piuri, V. (eds) Multi-Strategy Learning Environment. ICMSLE 2024. Algorithms for Intelligent Systems. Springer, Singapore. 2024;99-112. https://doi.org/10.1007/978-981-97-1488-9_8
[26] Eberhard BW, Cohen RY, Wheeler N, Kleinlein R, Rigoni J, Bates DW, et al. Development and validation of an interpretable longitudinal preeclampsia risk prediction using machine learning. PLoS One. 2025;20(6):e0323873.
[27] Wright D, Syngelaki A, Akolekar R, Poon LC, Nicolaides KH. Competing risks model in screening for preeclampsia by maternal characteristics and medical history. Am J Obstet Gynecol. 2015;213(1):62.e1-10.
[28] Zhang X, Chen Y, Salerno S, Li Y, Zhou L, Zeng X, et al. Prediction of severe preeclampsia in machine learning. Med Novel Technol Devices. 2022;15:100158.
[29] Marić I, Tsur A, Aghaeepour N, Montanari A, Stevenson DK, Shaw GM, et al. Early prediction of preeclampsia via machine learning. Am J Obstet Gynecol MFM. 2020;2(2):100100.
[30] Liu M, Yang X, Chen G, Ding Y, Shi M, Sun L, et al. Development of a prediction model on preeclampsia using machine learning-based method: A retrospective cohort study in China. Front Physiol. 2022;13:896969.
[31] Li YX, Shen XP, Yang C, Cao ZZ, Du R, Wang JP, et al. Novel electronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms. Pregnancy Hypertens. 2021;26:102-09.
[32] Marin I, Pavaloiu BI, Marian CV, Racovita V, Goga N. Early detection of preeclampsia based on a machine learning approach. In: 2019 E-Health Bioeng Conf (EHB). 2019;1-4.
[33] Schmidt LJ, Rieger O, Neznansky M, Hackelöer M, Dröge LA, Henrich W, et al. A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes. Am J Obstet Gynecol. 2022;227(1):77.e1.
[34] Liu L, Pu Y, Fan J, Yan Y, Liu W, Luo K, et al. Wearable sensors, data processing, and artificial intelligence in pregnancy monitoring: A review. Sensors (Basel). 2024;24(19):6426. Doi: 10.3390/s24196426.
[35] Sufiyana H, Wu YW, Su EC. Prediction of preeclampsia and intrauterine growth restriction: Development of machine learning models on a prospective cohort. JMIR Med Inform. 2020;8(5):e15411.
[36] Edvinsson C, Björnsson O, Erlandsson L, Hansson SR. Predicting intensive care needs in women with preeclampsia using machine learning–A pilot study. Hypertens Pregnancy. 2024;43(1):2312165.
[37] Kovacheva VP, Eberhard BW, Cohen RY, Maher M, Saxena R, Gray KJ. Preeclampsia prediction using machine learning and polygenic risk scores from clinical and genetic risk factors in early and late pregnancies. Hypertension. 2024;81(2):264-72.
[38] Wesson JL, Smith N. A machine learning model to predict preeclampsia in pregnant women. Procedia Comput Sci. 2024;239:1645-52.
[39] Serra B, Mendoza M, Scazzocchio E, Meler E, Nolla M, Sabrià E, et al. A new model for screening for early-onset preeclampsia. Am J Obstet Gynecol. 2020;222(6):608.e1.
[40] Melinte-Popescu AS, Vasilache IA, Socolov D, Melinte-Popescu M. Predictive performance of machine learning-based methods for the prediction of preeclampsia—A prospective study. J Clin Med. 2023;12(2):418.
[41] Jhee JH, Lee S, Park Y, Lee SE, Kim YA, Kang SW, et al. Prediction model development of late-onset preeclampsia using machine learning-based methods. PLoS One. 2019;14(8):e0221202.

PARTICULARS OF CONTRIBUTORS:
1. PhD Scholar, Department of Biochemistry, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
2. Professor and Head, Department of Biochemistry, Panimalar Medical College and Research Hospital, Chennai, Tamil Nadu, India.
3. Professor, Department of Obstetrics and Gynaecology, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
4. Associate Professor, Department of Biochemistry, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Dr. Leena Chand,
Associate Professor, Department of Biochemistry, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
E-mail: leena@sriramachandra.edu.in

PLAGIARISM CHECKING METHODS: [Jain H et al.]
- Plagiarism X-checker: Sep 01, 2025
- Manual Googling: Dec 19, 2025
- iThenticate Software: Dec 22, 2025 (1%)

ETYMOLOGY: Author Origin

EMENDATIONS: 5

AUTHOR DECLARATION:
- Financial or Other Competing Interests: None
- Was informed consent obtained from the subjects involved in the study? NA
- For any images presented appropriate consent has been obtained from the subjects. NA

Date of Submission: **Aug 06, 2025**
Date of Peer Review: **Nov 18, 2025**
Date of Acceptance: **Dec 26, 2025**
Date of Publishing: **May 01, 2026**