

# Pre-eclampsia Indicators as Predictors: A Pathway to Early Postpartum Depression Detection

S-22 : Charan B K, Darshan Srivatsa A, Estarth R Gupta, Hemant Chandrashekhar Wade  
Project Guide: Dr.Shruthi M L J (ECE)

## Abstract

Using **Pre-eclampsia (PE)** as a crucial factor, this work develops a **two-step method** for early **Postpartum depression (PPD)** prediction using the **PRAMS dataset (2012–2020)**. The following machine learning models were used: **SVM, Random Forest (RF), KNN, and Logistic Regression (LR)**. **MICE** and **SMOTE** were used in data preprocessing to handle **NULL values** and **class imbalance**, respectively. Two folds were used to evaluate the models, and RF consistently performed exceptionally well in terms of **precision, recall, F1-score, accuracy, and AUC**. These results **demonstrate the potential value of RF in clinical settings for precise and early PPD prediction**.

## Methodology

### Fold I: Predicting Pre-eclampsia

- **Data:** PRAMS dataset (353,827 participants, 613 features).
- **Features:** Selected **53** clinically significant features for PE.
- **Preprocessing:** Removed rows with **>20% null values**, applied **MICE** for imputation, and addressed imbalance with **SMOTE**.
- **Labelling & Splitting:** PE-positive samples identified using a weighted scoring system; data split **60:20:20** for training, validation, and testing.

### Fold II: Predicting PPD Using PE

- **Dataset:** True PE-positive cases from Fold I, with **39 shared features for PE and PPD**.
- **Preprocessing:** **MICE** and **SMOTE** applied, maintaining the same data split, highlighting the **PE-PPD connection**.

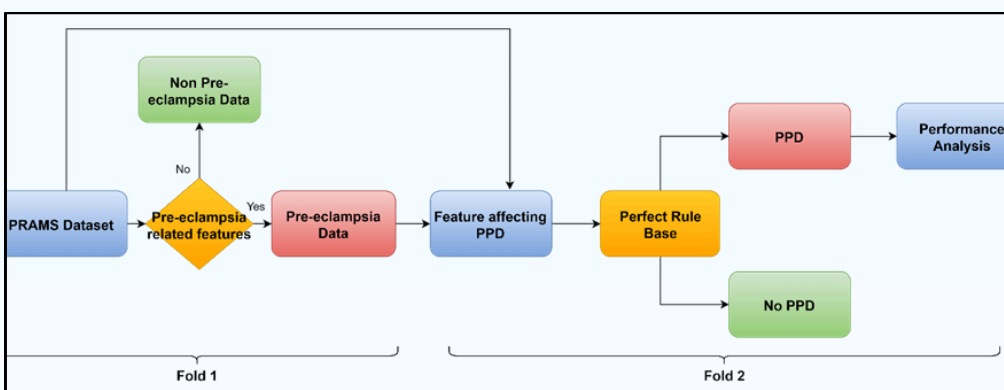


Figure 1: Block Diagram of the two-fold process

## Background

- Existing research on pre-eclampsia (PE) and postpartum depression (PPD) is often limited to **questionnaire-based studies or data from specific hospitals**, making findings hard to generalize.
- **No prior studies have explored the predictive relationship between PE and PPD**, despite its potential for early intervention and improved maternal health outcomes.

## Results & Analysis

- **SVM:** With PE-PPD% decreasing slightly from **18.86% to 17.2%** after addressing class imbalance.
- **Random Forest: Best performer**, increasing true positives and raising PE-PPD% from **17.48% to 19.26%**.
- **Logistic Regression:** Significant improvement post-imbalance handling, with PE-PPD% rising from **9.45% to 15.66%**.
- **KNN:** Considerable improvement, with PE-PPD% increasing from **11.80% to 16.33%**.

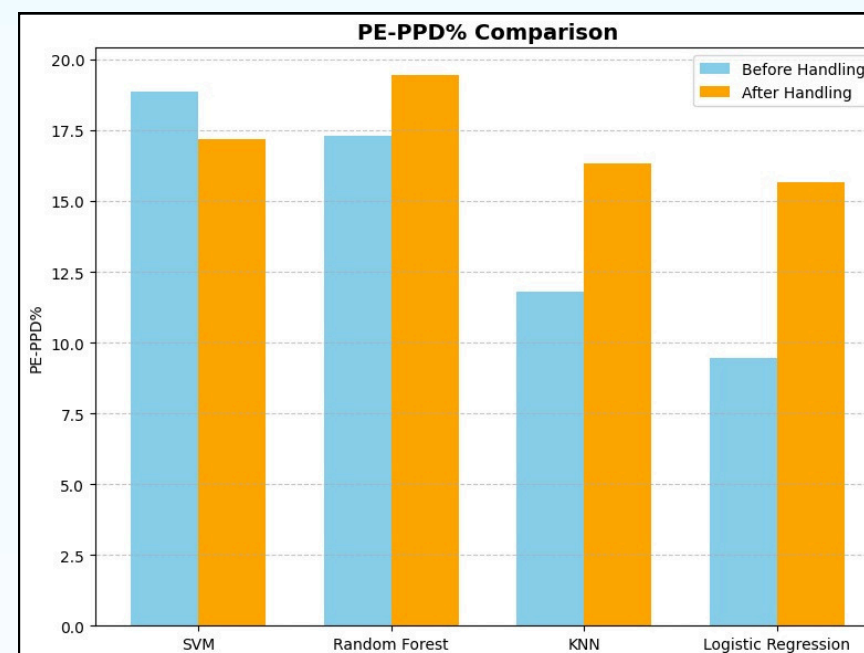


Figure 2: PE-PPD % comparison

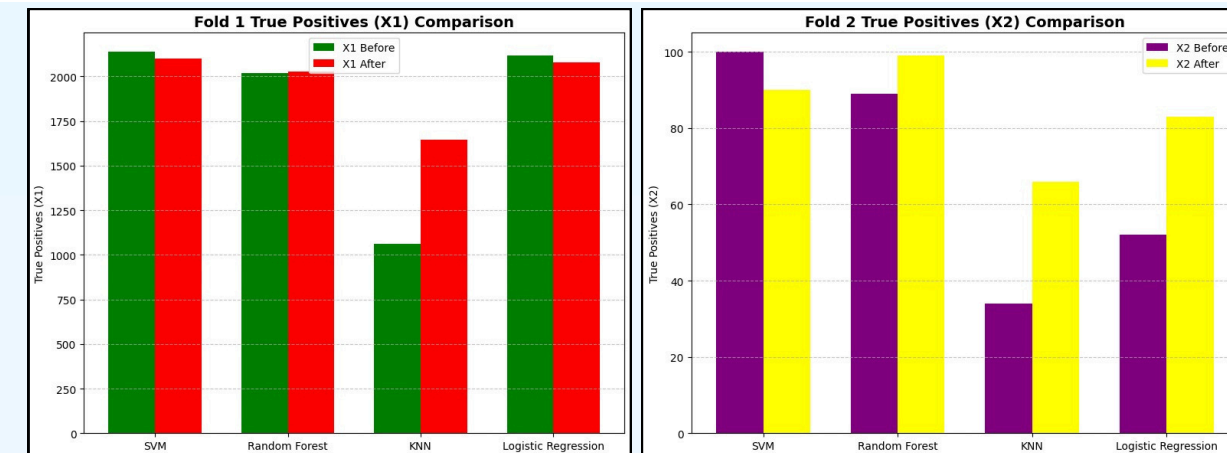


Figure 3: True positives of Fold1 and Fold2

Model	Class Imbalance	Fold	Precision	Recall	F1-Score	AUC	Accuracy
SVM	Before Handling	1	0.77	0.99	0.87	0.97	0.97
		2	0.73	0.95	0.83	0.91	0.90
	After Handling	1	0.63	0.97	0.76	0.95	0.94
		2	0.51	0.93	0.66	0.83	0.78
Random Forest	Before Handling	1	0.47	0.93	0.63	0.97	0.90
		2	0.59	0.85	0.70	0.85	0.81
	After Handling	1	0.50	0.93	0.65	0.97	0.91
		2	0.77	0.97	0.86	0.98	0.92
KNN	Before Handling	1	0.89	0.49	0.63	0.92	0.95
		2	0.97	0.62	0.76	0.94	0.89
	After Handling	1	0.66	0.76	0.71	0.92	0.94
		2	0.65	0.82	0.73	0.90	0.85
Logistic Regression	Before Handling	1	0.78	0.99	0.87	1.00	0.97
		2	0.68	0.50	0.57	0.87	0.82
	After Handling	1	0.68	0.97	0.80	0.99	0.96
		2	0.84	0.81	0.82	0.97	0.92

Table 1: Performance metrics of the ML models

## Conclusion

PPD, a global health issue, is linked to PE as a key risk factor. Using the PRAMS dataset, this study applied SMOTE and MICE to refine data, with models like Random Forest and SVM identifying PE cases to predict PPD. **Random Forest excelled post-imbalance handling, detecting 19.26% PPD subjects using PE as an indicator**, while SVM remained consistent. **Future ML advancements could enhance PPD prediction by integrating broader risk factors like genetics and lifestyle.**