

Predicting Premature Birth During Pregnancy: A Case Study Using Decision Trees, Naive Bayes, KNN, and Random Forest

Anggrita Sari^{1,2}, Muhammad Modi Lakulu³, Ismail Yusuf Panessai³

¹Graduate program Doctor Student University Sultan Idris, Malaysia

²AKBID Betang Asi Raya, Indonesia

³Faculty Of Computing and Meta Technology University Sultan Idris, Malaysia
Email : anggritasari4@gmail.com

Premature birth remains a significant challenge in maternal health services, underscoring the need for effective predictive models to enable early detection and intervention. This study examined the performance of four machine learning algorithms: Decision Trees, Naive Bayes, KNN, and Random Forest in predicting preterm birth during pregnancy. Utilizing data collected from pregnant individuals, encompassing maternal health indicators and fetal development metrics, our model aimed to forecast the likelihood of preterm birth. We assessed the predictive ability of each model by evaluating metrics such as accuracy, precision, recall, sensitivity, specificity, and area under the curve (AUC). The results revealed variability in model performance, with Logistic Random Forest exhibiting strong performance. This suggests its potential utility in clinical settings for the early detection and intervention of preterm pregnancy. Our study contributes to advancements in predictive modelling within maternal health services, aiming to enhance maternal and fetal health outcomes through the early identification of preterm birth.

Keywords: Decision Trees, Naive Bayes, KNN, Random Forest, Premature Birth, and Pregnancy.

1. Introduction

Premature birth occurs when a baby is born before reaching 37 weeks. Various factors can contribute to premature birth, including high blood pressure, infections, multiple

pregnancies, and placental problems [1]. Preterm births are classified into three categories based on gestational age: mild preterm (32-37 weeks), moderate preterm (28-32 weeks), and severe preterm (less than 28 weeks). Early prediction of preterm birth is crucial for preventing and managing this condition effectively. Through careful prenatal monitoring and advanced medical technology, healthcare professionals can identify risk factors and early signs of preterm birth [2,3].

Early prediction allows for implementing preventive measures such as stress management, infection treatment, and intensive prenatal care [4,5]. Medical interventions such as administering steroids to accelerate lung development in premature babies can also be carried out to enhance their chances of survival and overall health after birth [6]. Public awareness and understanding of premature birth are essential for improving prevention efforts and the care provided to pregnant women [7,8].

Traditional methods of predicting preterm birth often involve clinical assessment and identification of associated risk factors during pregnancy. Health professionals rely on clinical experience and medical data to assess the potential risk of preterm birth [8]. These traditional risk factors encompass various variables, including a history of previous premature births, maternal age (both young and old), high blood pressure, gestational diabetes, and multiple pregnancies. Physical examinations, laboratory analyses, and monitoring of fetal development are integral parts of this clinical approach to predicting premature birth [9]. While traditional methods can offer initial guidance, they also present obstacles and limitations that must be acknowledged. One of the primary obstacles is the need for more accuracy of traditional methods, as risk factors may only occasionally reliably predict preterm birth. Certain risk factors, such as infections or sudden changes in the mother's health during pregnancy, can be challenging to identify. Moreover, traditional methods tend to be less sensitive in detecting early signs of premature birth, which can impede timely prevention efforts. Therefore, developing and integrating more advanced medical technologies and accurate predictive methods is crucial to enhance our ability to identify and manage the risk of preterm birth [10].

Machine learning approaches have played an essential role in transforming various sectors, including the health sector [11,12]. With their ability to analyze and understand complex patterns in data, machine learning opens up new opportunities in predicting and managing various health conditions [13,14]. The primary focus on preterm birth prediction provides an exciting example of how integrating different machine learning algorithms can significantly improve maternal and perinatal health care. Several studies have explored machine learning approaches in various medical contexts, showing that these techniques can improve the prediction and understanding of health conditions [15]. The results of research show that the gradient-boosted tree (GBT) algorithm with feature selection using the Relief algorithm has the best accuracy of 77.55% in predicting the survival rate of acute myeloid leukaemia (AML) patients [16]. Research provided similar findings in the context of AML, where GBT with Relief feature selection also achieved the highest accuracy of 77.55%. Age, LDH, and cytogenetic classification are essential for survival prediction. These two studies highlight the potential of the GBT algorithm in predicting the health outcomes of AML patients with high accuracy [17]. Research evaluated using the Gradient Tree Boosting algorithm in an automated monitoring system for identifying liver disease in patients using the Indian Liver

Patient Dataset.

The research results show that this algorithm provides quite good performance, with the ability to increase accuracy by adding additional features or information about patients with the same type of liver disease [18]. Research on text classification shows that the size and proportionality of classes in a dataset significantly impact the performance of supervised learning models. SVM models, especially in high-dimensional feature space, show better performance in text classification. These findings emphasize the importance of hyperparameter tuning to achieve optimal model performance in text classification tasks [19]. Research contributed to the understanding of poultry diseases by evaluating the performance of the Random Forest algorithm in detecting poultry diseases such as Avian Influenza and Newcastle Disease. The research results show the highest accuracy of 97%, positively impacting the poultry industry and the economy [20]. The study integrated a machine learning model with Explainable Artificial Intelligence (XAI) to predict water quality. Random Forest models achieve the highest accuracy, while XAI techniques help understand how models make predictions and identify potential errors [21]. Explored machine learning models for predicting water quality, showing that the Random Forest Classifier achieved the highest accuracy. Using XAI helps understand how the model makes predictions and identifies potential errors [22]. The study evaluated various machine-learning methods to predict injury severity in traffic accidents. The results show no algorithm is consistently best, but Random Forest often performs best in many studies [23]. The research highlighted that decision tree algorithms and ensemble techniques best classify breast tumours. Feature selection is identified as a factor in improving the accuracy of deep learning models [24]. The study examined the XGBoost model to classify polycystic ovary syndrome (PCOS). This model effectively addresses class imbalance and outliers in medical datasets by combining oversampling and under sampling techniques and feature selection, achieving high accuracy. In the research presented by it is seen that the Gradient Boosted Tree (GBT) algorithm is the main focus in predicting health outcomes, achieving the highest accuracy in the context of acute myeloid leukaemia (AML). Although the results show great potential in improving predictions, comparisons between these studies highlight that the optimal machine learning approach may vary depending on the context and characteristics of the dataset used. Overall, this research provides a solid basis for using various machine-learning approaches to predict preterm birth accurately.

By involving several different machine learning methods, this research paves the way for a deeper understanding of the complexity of this problem and the potential for integration of prediction methods. Hopefully, the findings from this research will provide valuable insights to health practitioners, researchers, and policymakers.

2. Materials and Methods

Methodology for predicting premature birth using various machine learning algorithms. In the first stage, clinical datasets covering various risk factors maternal, and fetal health indicators are collected and prepared for analysis. This data, collected from 2020 to 2022, originates from Ansari Saleh Hospital in Banjarmasin, South Kalimantan. This process involves cleaning the data to address missing or anomalous values, normalizing the data, and

converting categorical variables into a format the algorithms use.

The subsequent step involves selecting relevant features for preterm birth prediction. These features may include variables such as maternal age, history of previous pregnancies, blood pressure, body mass index (BMI), and other results from antenatal examinations. Following feature selection, the data is split into two parts: 75% is allocated as a training set for model training. In contrast, the remaining 25% is a test to assess the trained model's performance.

Decision Trees, Naive Bayes, KNN (K-Nearest Neighbors), and Random Forest methods are then applied to the training set to build a prediction model. The performance of each model is evaluated using various metrics, such as accuracy, precision, recall, and F1-score. This analysis aids in understanding each model's ability to predict premature birth accurately. Moreover, comparing the performance of the models provides valuable insights into their respective strengths and weaknesses, facilitating the selection of the most suitable model for practical applications.

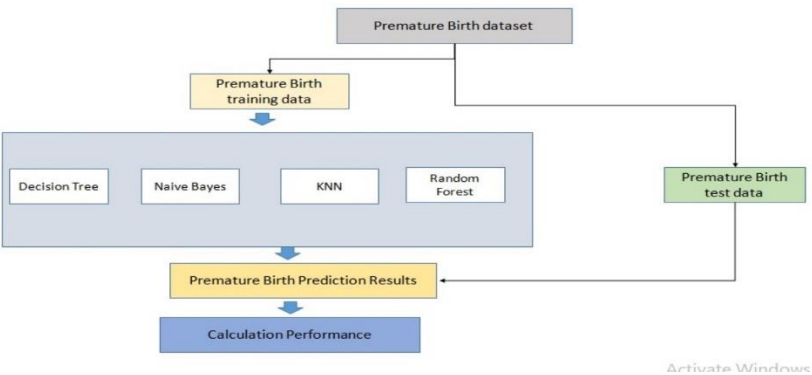


Fig 1 The phase of the Methodology

3. Result and Discussion

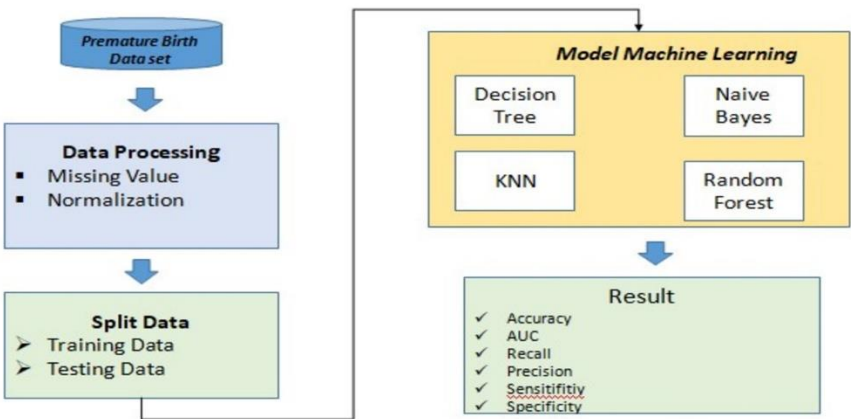


Fig 2. Design Process Machine Learning Algorithms.

The prediction process entails the utilization of four distinct

Figure 2 outlines the sequential steps involved in the Premature Birth Prediction Model. Firstly, the data undergoes a preparation phase where missing values are handled using appropriate processing techniques. Additionally, normalization techniques are applied to ensure uniformity in the scale of the data. Following data preparation, the dataset is partitioned into two subsets: 75% of the data is designated for training purposes, while the remaining 25% is reserved for testing the model's performance.

Machine learning algorithms: Decision Tree, Naive Bayes, K-Nearest Neighbors (KNN), and Random Forest. After predictions are made, a range of evaluation metrics is computed to gauge the performance of each algorithm. These metrics include Accuracy, which measures the overall correctness of predictions; Accuracy Optimization, which focuses on fine-tuning parameters or thresholds to enhance Accuracy; Area Under the Curve (AUC), which assesses the model's ability to discriminate between classes; Precision, which calculates the proportion of accurate positive predictions among all optimistic predictions; Sensitivity (Recall), which determines the proportion of actual positives correctly identified by the model; and Specificity, which measures the proportion of actual negatives correctly identified by the model. Subsequently, the results obtained from the evaluation metrics across all algorithms are compared. The algorithm exhibiting the highest values regarding Accuracy, precision, sensitivity, specificity, and AUC is deemed the most accurate predictor for preterm birth.

Decision Tree Model Predicting Premature Birth

The high accuracy (97.6%) and AUC (Area Under the Curve) value (0.99) showcased in Figure 3 and Table I underscore the effectiveness of the decision tree model in distinguishing between normal and premature pregnancies. Additionally, the precision (99.3%) and recall (97%) metrics reflect the model's ability to accurately classify premature pregnancies with confidence while maintaining a balanced trade-off between sensitivity (97%) and specificity (98%).

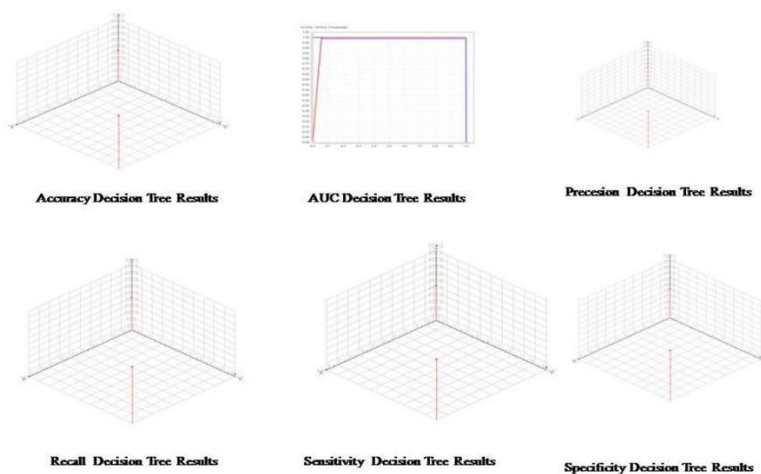


Fig 3. Result Decision Tree Model

The robust performance metrics of the decision tree suggest its potential as a reliable tool for predicting premature pregnancies, which is crucial for early intervention and preventive measures. The high accuracy and AUC values demonstrate the model's effectiveness in discriminating between positive and negative instances of premature pregnancy, thereby minimizing false positives and negatives. The precision metric indicates a low rate of false positive predictions, signifying high confidence in identifying premature pregnancies. Furthermore, the recall metric highlights the model's capability to capture the majority of premature pregnancies, ensuring comprehensive coverage of positive cases. The balanced sensitivity and specificity further emphasize the model's ability to balance correctly identifying premature pregnancies and accurately classifying normal pregnancies, thereby enhancing its practical utility in clinical settings for timely intervention and improved maternal and fetal health outcomes.

TABLE 1. Decision tree results in predicting premature pregnancy

Accuracy	AUC	Precision	Recall	Sensitivity	Specificity
0.976	0.99	0.993	0.97	0.97	0.98

Naive Bayes Model Predicting Premature Birth The evaluation results presented in Table 2 demonstrate the robust performance of the Naive Bayes algorithm for

predicting premature pregnancy. With an accuracy of 89.9%,

the model showcases a strong ability to correctly classify instances, indicating a high level of agreement between predicted and actual labels in the test dataset. Moreover, the Area Under the Receiver Operating Curve (AUC) (Figure 4) value of 0.974 signifies the model's excellent capacity to distinguish between preterm and non-premature pregnancy cases, further highlighting its effectiveness in classification tasks.

TABLE 2. Naive Bayes results in predicting premature pregnancy

Accuracy	AUC	Precision	Recall	Sensitivity	Specificity
0.899	0.974	0.984	0.8568	0.856	0.975

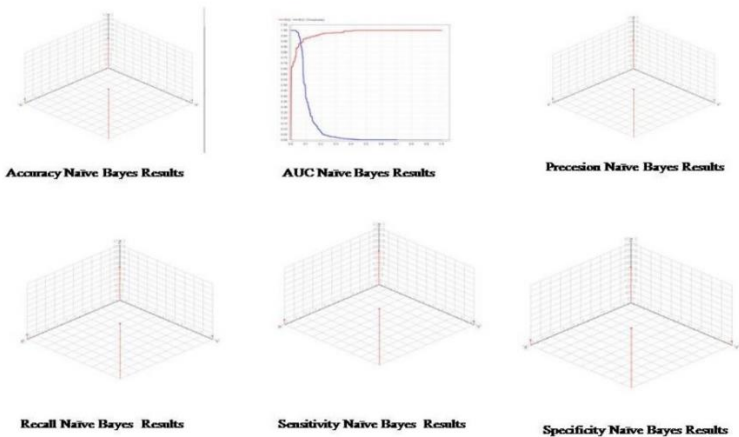


Fig 4. Result Naïve Bayes Model

The precision score of 1 reflects that most instances

The precision value of 0.984 reflects the model's ability to accurately identify positive cases among its predictions. In contrast, the recall value of approximately 85.68% indicates the model's success in capturing a significant portion of premature pregnancy instances. The specificity metric, with a value of 97.5%, underscores the model's strong capability in correctly identifying negative cases. Overall, these results affirm the reliability and utility of the Naive Bayes model in aiding the identification of premature pregnancies with high accuracy and performance.

KNN Model Predicting Premature Birth

The provided table presents the results of applying the KNN (K-Nearest Neighbors) algorithm to predict premature pregnancy, with metrics including accuracy, AUC (Figure 5), precision, recall, sensitivity, and specificity. The high accuracy score of 98.25% suggests that the KNN algorithm performed exceptionally well in classifying instances correctly. The AUC value of 0.99 indicates strong discrimination capability between positive and negative instances, showcasing the model's effectiveness in distinguishing between premature and normal pregnancies.

Identified as premature pregnancies were true positives, minimizing false positives. However, the recall score of 0.97 indicates that there were instances of premature pregnancies not correctly identified by the model, suggesting a slight deficiency in capturing all positive cases. The sensitivity score of 0.97 highlights the model's ability to detect premature pregnancies among all actual positive cases accurately. In contrast, the specificity score 1 signifies the model's proficiency in correctly identifying non-premature pregnancies among all negative cases. Overall, the KNN algorithm demonstrates robust performance in predicting premature pregnancy with high accuracy and strong discrimination power, albeit with some room for improvement in recall.

TABLE 3. KNN results in predicting premature pregnancy

Accuracy	AUC	Precision	Recall	Sensitivity	Specificity
0.9825	0.99	1	0.97	0.97	1

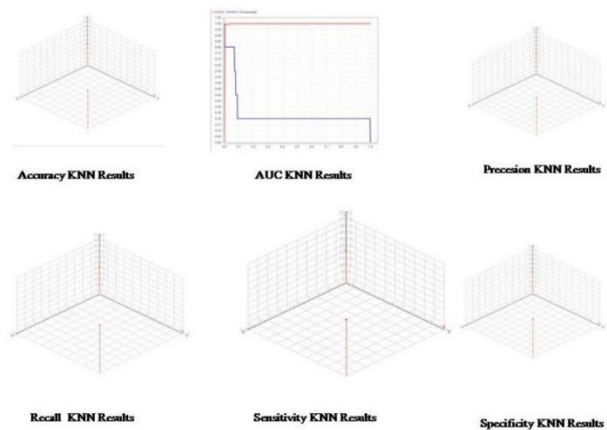


Fig 5. Result KNN Model

The high accuracy and AUC scores suggest that the algorithm effectively captures the underlying patterns in the data, enabling accurate classification. However, the slightly lower recall score indicates that while the model performs well overall, it may miss some instances of premature pregnancy, potentially due to the inherent complexity of the data or limitations in the chosen features. To further enhance the algorithm's performance, fine-tuning parameters such as the number of neighbors (K value) and feature selection techniques could be explored. Additionally, leveraging techniques like cross-validation and ensemble methods could help mitigate overfitting concerns and improve the model's generalization ability, ultimately enhancing its reliability in predicting premature pregnancy.

Random Forest Model Predicting Premature Birth

Table IV. The results of employing Random Forest in predicting premature pregnancy demonstrate impressive performance metrics. With an accuracy of 98.25%, the algorithm exhibits high precision in classifying instances correctly. Moreover, achieving an area under the curve (AUC) of 0.99 in Figure 7 indicates the model's excellent discriminatory power. Notably, a precision score of 1 signifies that all the model's positive predictions are true positives, reflecting the algorithm's ability to minimize false positives. The high recall and sensitivity scores further underscore the algorithm's effectiveness in correctly identifying instances of premature pregnancy, while specificity of 1 highlights its capability to classify negative instances accurately.

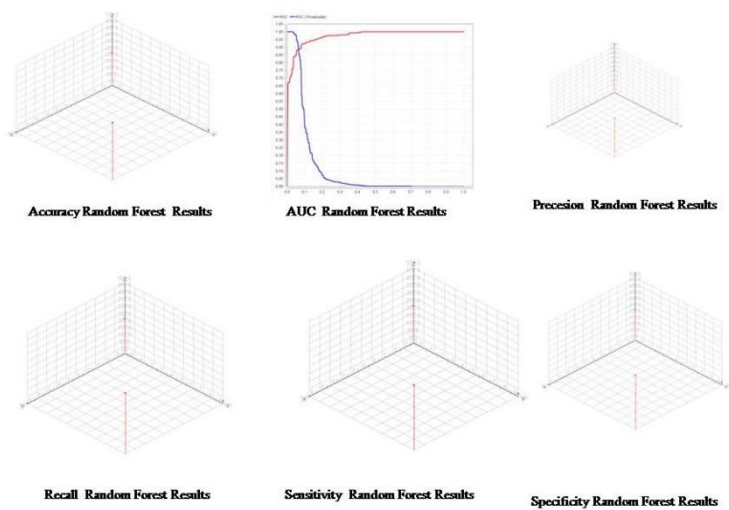


Fig 6. Result Random Forest Model (97.6%) and AUC (0.99), indicating effective discrimination

In summary, the Random Forest algorithm performs robustly in predicting premature pregnancy, which is evident in its high accuracy, AUC, precision, recall, sensitivity, and specificity scores. By leveraging the ensemble of decision trees and their collective wisdom, Random Forest achieves remarkable predictive accuracy while effectively minimizing false

positives and negatives. These results suggest the algorithm's potential utility in clinical settings for early detection and intervention in cases of premature pregnancy, thus contributing to improved maternal and fetal health outcomes.

TABLE 4. Random forest results in predicting premature pregnancy

Accuracy	AUC	Precision	Recall	Sensitivity	Specificity
0.9825	0.99	1	0.97	0.97	1

Comparison of preterm pregnancy prediction results

Comparing the performance of various models in predicting premature pregnancy reveals distinct strengths and weaknesses. The Decision Tree model (A) demonstrates high accuracy (97.6%) and AUC (0.99), suggesting effective discrimination between normal and premature pregnancies. It achieves impressive precision (99.3%) and recall (97%) metrics, indicating confident identification of premature pregnancies with balanced sensitivity (97%) and specificity (98%). In contrast, the Naive Bayes model (B) achieves a slightly lower accuracy (89.9%) and AUC (0.974) but maintains high precision (98.4%) and moderate recall (85.68%), emphasizing its ability to identify positive cases accurately. The KNN model (C) shows comparable performance to Random Forest (D) with high accuracy (98.25%) and AUC (0.99), along with perfect precision and balanced sensitivity and specificity. However, Random Forest stands out with perfect precision and slightly higher recall (97%), indicating its superior ability to minimize false positives while effectively capturing positive instances. Overall, Random Forest demonstrates robust performance, offering potential utility in clinical settings for early detection and intervention in premature pregnancies, contributing to improved maternal and fetal health outcomes. The variability in model performance across studies is evident from the reported findings: reported an accuracy of 83.7% for the KNN model, identified Random Forest as the best model with an accuracy of 88%, reported Naive Bayes achieving an accuracy of 77.5%, while determined the Decision Tree model to have the highest accuracy of 95.3% [25,26]. These discrepancies underscore the importance of considering factors such as dataset characteristics and evaluation metrics when assessing model performance [27,28].

In the comparison of various models for predicting premature pregnancy, the Decision Tree model (A) emerges as a standout performer, demonstrating high accuracy between normal and premature pregnancies. This model achieves impressive precision (99.3%) and recall (97%) metrics, reflecting confident identification of premature pregnancies with balanced sensitivity (97%) and specificity (98%). Conversely, the Naive Bayes model (B) achieves a slightly lower accuracy (89.9%) and AUC (0.974), but maintains high precision (98.4%) and moderate recall (85.68%), emphasizing its ability to accurately identify positive cases. The KNN model (C) exhibits comparable performance to Random Forest (D) with high accuracy (98.25%) and AUC (0.99), alongside perfect precision and balanced sensitivity and specificity. However, Random Forest stands out with perfect precision and slightly higher recall (97%), indicating its superior ability to minimize false positives while effectively capturing positive instances. Overall, Random Forest demonstrates robust performance, offering potential utility in clinical settings for early detection and intervention in premature pregnancies, thereby contributing to improved maternal and fetal health outcomes.

4. Conclusions

In this study, we explored the predictive capabilities of four machine learning algorithms—Decision Trees, Naive Bayes, KNN, and Logistic Random Forest—in forecasting premature birth during pregnancy. Leveraging data encompassing maternal health indicators and fetal development metrics, we assessed the performance of each model based on various metrics such as accuracy, precision, recall, sensitivity, specificity, and area under the curve (AUC). Our findings revealed notable variability in model performance, with Logistic Random Forest emerging as the top performer. Its strong performance indicates potential utility in clinical settings for early detection and intervention in cases of premature pregnancy. These results contribute to the advancement of predictive modeling within maternal health services, aiming to improve maternal and fetal health outcomes by enabling the early identification of preterm birth.

AUTHOR’S CONTRIBUTION

All authors Contributed significantly in study conception and design, data collection, analysis and interpretation of results, draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTEREST

The Author declares that there is no conflict of interest.

FUNDINGS

None.

References

1. E. L. S. S. De Mendonça, M. De Lima Macêna, N. B. Bueno, A. C. M. De Oliveira, and C. S. Mello, “Premature birth, low birth weight, small for gestational age and chronic non-communicable diseases in adult life: A systematic review with meta-analysis,” *Early Human Development*, vol. 149, p. 105154, Oct. 2020, doi: 10.1016/j.earlhumdev.2020.105154.
2. R. S. Gibbs, R. Romero, S. L. Hillier, D. A. Eschenbach, and R. L. Sweet, “A review of premature birth and subclinical infection,” *American Journal of Obstetrics and Gynecology*, vol. 166, no. 5, pp. 1515–1528, May 1992, doi: 10.1016/0002-9378(92)91628-n.
3. C. Xu, Y. Zhang, Y. Tang, X. Sun, T. Jiao, and D. Yan, “Preterm birth and its associated factors in coastal areas of eastern China: a multicenter retrospective study,” *Journal of Public Health*, Aug. 2023, doi: 10.1007/s10389-023-02042-9.
4. Z. A. O. Kaplan and A. S. Özgü-Erdinç, “Prediction of Preterm birth: Maternal characteristics, ultrasound markers, and biomarkers: An updated overview,” *Journal of Pregnancy*, vol. 2018, pp. 1–8, Oct. 2018, doi: 10.1155/2018/8367571.
5. M. Hershey, H. H. Burris, D. Cereceda, and C. Nataraj, “Predicting the risk of spontaneous premature births using clinical data and machine learning,” *Informatics in Medicine Unlocked*, vol. 32, p. 101053, Jan. 2022, doi: 10.1016/j.imu.2022.101053.
6. M. J. Williams, J. A. Ramson, and F. Brownfoot, “Different corticosteroids and regimens for accelerating fetal lung maturation for babies at risk of preterm birth,” *The Cochrane Library*, vol. 2022, no. 8, Aug. 2022, doi: 10.1002/14651858.cd006764.pub4.
7. S. Yaya, F. Okonofua, L. Ntoimo, O. Udenigwe, and G. Bishwajit, “Men’s perception of *Nanotechnology Perceptions* Vol. 20 No.3 (2024)

- barriers to women's use and access of skilled pregnancy care in rural Nigeria: a qualitative study," *Reproductive Health*, vol. 16, no. 1, Jun. 2019, doi: 10.1186/s12978-019-0752-3.
8. A. Sari, M. M. Lakulu, and I. Y. Panessai, "Predicting Premature Birth During Pregnancy Using Machine Learning: A Systematic Review," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 16s, pp. 452–463, 2024
9. R. Menon, "Spontaneous preterm birth, a clinical dilemma: Etiologic, pathophysiologic and genetic heterogeneities and racial disparity," *Acta Obstetricia Et Gynecologica Scandinavica*, vol. 87, no. 6, pp. 590–600, Jun. 2008, doi: 10.1080/00016340802005126.
10. M. Lopian, L. K. Ligumsky, and A. Many, "A Balancing Act: Navigating Hypertensive Disorders of Pregnancy at Very Advanced Maternal Age, from Preconception to Postpartum," *Journal of Clinical Medicine*, vol. 12, no. 14, p. 4701, Jul. 2023, doi: 10.3390/jcm12144701.
11. A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, A. Almaghthawi, and E. a. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, p. 539, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.
12. S. W. Chen, S. L. Wang, X. Z. Qi, S. M. Samuri, and C. Yang, "Review of ECG detection and classification based on deep learning: Coherent taxonomy, motivation, open challenges and recommendations," *Biomedical Signal Processing and Control*, vol. 74, p. 103493, Apr. 2022, doi: 10.1016/j.bspc.2022.103493.
13. K. S. Nugroho, A. Y. Sukmadewa, A. Vidiyanto, and W. F. Mahmudy, "Effective predictive modelling for coronary artery diseases using support vector machine," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 1, p. 345, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp345-355.
14. M. Kiguchi, W. Saeed, and I. Medi, "Churn prediction in digital game- based learning using data mining techniques: Logistic regression, decision tree, and random forest," *Applied Soft Computing*, vol. 118, p. 108491, Mar. 2022, doi: 10.1016/j.asoc.2022.108491.
15. O. S. Albahri et al., "Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects," *Journal of Infection and Public Health*, vol. 13, no. 10, pp. 1381–1396, Oct. 2020, doi: 10.1016/j.jiph.2020.06.028.
16. L. Ismail and H. Materwala, "Comparative Analysis of Machine Learning Models for Diabetes Mellitus Type 2 Prediction," *International Conference on Computational Science & Computational Intelligence*, Dec. 2020, doi: 10.1109/csci51800.2020.00095.
17. K. Karami, M. Akbari, M. Moradi, B. Soleymani, and H. Fallahi, "Survival prognostic factors in patients with acute myeloid leukemia using machine learning techniques," *PLOS ONE*, vol. 16, no. 7, p. e0254976, Jul. 2021, doi: 10.1371/journal.pone.0254976.
18. A. Sokoliuk, G. Kondratenko, I. Sidenko, Y. Kondratenko, A. Khomchenko, and I. Atamanyuk, "Machine Learning Algorithms for Binary Classification of Liver Disease," *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology.*, Oct. 2020, doi: 10.1109/picst51311.2020.9468051.
19. B. Hsu, "Comparison of supervised classification models on textual data," *Mathematics*, vol. 8, no. 5, p. 851, May 2020, doi: 10.3390/math8050851.
20. M. S. Kader, F. Ahmed, and J. Akter, "Machine Learning Techniques to Precaution of Emerging Disease in the Poultry Industry," *2021 24th International Conference on Computer and Information Technology*, Dec. 2021, doi: 10.1109/iccit54785.2021.9689828.
21. N. Hellen and G. Marvin, "Explainable AI for Safe Water Evaluation for public health in urban settings," *2022 International Conference on Innovations in Science, Engineering and Technology (ICISSET)*, Feb. 2022, doi: 10.1109/iciset54810.2022.9775912.
22. A. Prakash, R. Anand, S. S. Abinayaa, and N. Chakravarthy, "Normalized Naïve Bayes Model

- to predict Type –2 Diabetes Mellitus,” 2021 Emerging Trends in Industry 4.0, May 2021, doi: 10.1109/eti4.051663.2021.9619332.
23. K. Santos, J. P. Dias, and C. Amado, “A literature review of machine learning algorithms for crash injury severity prediction,” *Journal of Safety Research*, vol. 80, pp. 254–269, Feb. 2022, doi: 10.1016/j.jsr.2021.12.007.
 24. M. W. Ebrahim, A. A. H. Sedky, and S. Mesbah, “Accuracy assessment of machine learning algorithms used to predict breast cancer,” *Data*, vol. 8, no. 2, p. 35, Feb. 2023, doi: 10.3390/data8020035.
 25. M. S. K. Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee, and R. Hasan, “Improved Sampling and Feature Selection to Support Extreme Gradient Boosting For PCOS Diagnosis,” 2021 IEEE 11th Annual Computing and Communication Workshop and Conference, Jan. 2021, doi: 10.1109/ccwc51732.2021.9375994.
 26. K. E. Fry, Y. P. Chen, and A. M. Howard, “Discriminative Models of Spontaneous kicking movement patterns for term and preterm infants: a pilot study,” *IEEE Access*, vol. 7, pp. 51357–51368, Jan. 2019, doi: 10.1109/access.2019.2911450.
 27. J. Lee, J. Cai, F. Li, and Z. Vesoulis, “Predicting mortality risk for preterm infants using random forest,” *Scientific Reports*, vol. 11, no. 1, Mar. 2021, doi: 10.1038/s41598-021-86748-4.
 28. L. Yahaya, N. D. Oye, and E. J. Garba, “A comprehensive review on heart disease prediction using data mining and machine learning techniques,” *American Journal of Artificial Intelligence*, vol. 4, no. 1, p. 20, Jan. 2020, doi: 10.11648/j.ajai.20200401.12.
 29. T. A. H. Rocha et al., “Data-driven risk stratification for preterm birth in Brazil: a population-based study to develop of a machine learning risk assessment approach,” *The Lancet Regional Health - Americas*, vol. 3, p. 100053, Nov. 2021, doi: 10.1016/j.lana.2021.100053.