Predictive Linear Regression Model for Premature Birth Risk Assessment System

Dewi Kusumaningsih[*1] **, Abdul Kadir**[2]**, Ahmad Pudoli**[3]
[1,3] Informatics Engineering, Universitas Budi Luhur, Indonesia
[2] Information Technology, Universtas Sari Mulia, Indonesia
e-mail: [*1] **dewi.kusumaningsih@budiluhur.ac.id,**
[2]kadir.budiluhur@gmail.com,[3]mahmad.pudoli@budiluhur.ac.id

Abstract

Preterm birth is a major cause of neonatal mortality in Indonesia and is influenced by multiple maternal factors. Early prediction models are crucial for supporting timely clinical decision-making and reducing adverse maternal–infant outcomes. The method of this study developed a linear regression–based predictive model using 915 pregnancy medical records from Dr. H. M. Ansari Saleh Regional Hospital, Banjarmasin (2020–2022). The workflow included data preprocessing, feature selection, Min-Max normalization, and experimentation with various train–test split ratios (90:10 to 50:50). Model performance was evaluated using R^2, Adjusted R^2, MAE, MSE, RMSE, and MAPE metrics. As the results, the 70:30 split ratio achieved the best accuracy of 89.05% and AUC of 98.10%, with low prediction errors. Optimizations with Adamax and Nadam enhanced stability and reduced MAPE to 1.95%. The optimized linear regression model reliably predicts preterm birth risk and is suitable for clinical decision support, particularly in resource-limited settings.

Keywords— *Clinical Decision Support; Healthcare Informatics; Linear Regression; MAPE Optimization; Predictive Modeling; Premature Birth*

1. INTRODUCTION

Preterm birth—defined as delivery before 37 weeks of gestation—remains a major global public health challenge. The World Health Organization (WHO) estimates that around 15 million babies are born prematurely each year, with more than one million deaths attributed to its complications [1,2]. In Indonesia, preterm birth is a leading contributor to neonatal mortality and is associated with long-term developmental problems [3]. Therefore, early prediction of preterm birth is crucial to enable timely clinical intervention and improve maternal–infant outcomes [4].

Advancements in health informatics and the growing availability of electronic medical records have facilitated the development of data-driven prediction models using statistical and machine-learning methods [5,6]. Linear regression is widely used in clinical decision-making because it is simple, transparent, and easily interpretable by health professionals [7–10]. Prior studies have demonstrated the usefulness of linear regression in identifying relationships between maternal characteristics—such as maternal age, parity, body mass index (BMI), anemia, blood pressure, and obstetric history—and gestational outcomes [11]. However, its assumption of linearity may oversimplify the complex and multifactorial causes of preterm birth [12–14].

Various machine-learning approaches, including Support Vector Machines (SVM) [15,16], Random Forest [17,18], Artificial Neural Networks (ANN) [19,20], and k-Nearest Neighbors (k-NN) [21,22], have been explored to capture non-linear interactions and improve prediction accuracy. Despite their performance advantages, these models often lack interpretability, which is a critical requirement in healthcare decision support [23,24]. Transparent models such as linear regression remain preferable in clinical settings due to their explainability and ease of implementation [25,26].

Preterm birth is influenced by multiple demographic, clinical, behavioral, and environmental factors [28]. This study integrates these predictors using clinical data from two major referral hospitals in South Kalimantan—Ulin Hospital and dr. H. M. Ansari Saleh Hospital—to ensure contextual relevance and applicability within the regional healthcare system. Data preprocessing steps, such as normalization, handling of missing values, and feature selection, were conducted to improve model robustness and address multicollinearity issues commonly encountered in regression analysis [29–33].

To support intuitive interpretation, model outputs were visualized using Microsoft Power BI, enabling health workers and policymakers to understand prediction patterns more easily [34,35]. Model performance was assessed using standard regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) [36–41].

The main objective of this study is to develop an interpretable and optimized linear regression model for predicting preterm birth risk. The model is expected to enhance clinical decision-making, improve prenatal care quality, and support efforts to reduce neonatal mortality, particularly in South Kalimantan.

2. RESEARCH METHODS

2.1. Research Design

This study uses an exploratory quantitative approach to develop and optimize a linear regression-based preterm birth prediction model. This approach is based on the need to identify relationships between numerical variables and measure the contribution of each factor to birth status (premature or normal). Figure 1 shows the stages of the study, starting from collecting electronic medical record data, selecting variables, data preprocessing (cleaning, normalization, encoding), dividing training data and test data, building simple and multiple linear regression models, and evaluating the model using statistical metrics and classification.

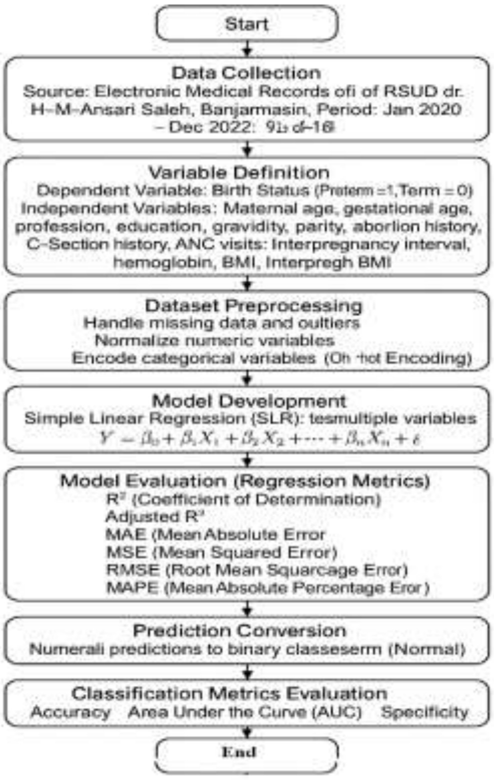

Figure 1. Research Design Flow and Development of Premature Birth Prediction Model

2.2. Data Sources

The data used in this study came from electronic medical records (EMR) of pregnant women who gave birth at the dr. H. M. Ansari Saleh Regional Hospital, Banjarmasin, from January 2020 to December 2022. The total data successfully collected and declared suitable for analysis was 915 medical records. The data has gone through a de-identification stage to ensure patient confidentiality.

2.3. Research Variables

Table 1. Overview of Variables for Prematurity Prediction

Variable	**Variable Type**	**Description / Coding**
Dependent Variable (Y)		
Birth Status	Binary (0,1)	Preterm = 1 (<37 weeks), Term = 0
Independent Variables (X)		
Maternal Age	Continuous numeric	Years
Gestational Age	Continuous numeric	Weeks
Maternal Profession	Categorical	Coded
Maternal Education	Categorical	Coded
Gravidity	Discrete numeric	Total number of pregnancies
Parity	Discrete numeric	Number of live births
History of Abortion	Discrete numeric	Number of abortions
History of C-Section	Binary	0 = No, 1 = Yes
Number of ANC Visits	Discrete numeric	Antenatal care visits
Interpregnancy Interval	Continuous numeric	Months
Hemoglobin Level	Continuous numeric	g/dL
Body Mass Index (BMI)	Continuous numeric	kg/m²

2.4. Analysis Techniques

Data preprocessing is done by handling missing data and outliers to ensure the data quality used. Furthermore, numeric data is normalized so that the variable scale becomes uniform. In contrast, categorical variables are encoded into numeric form using the one-hot encoding method or label encoding according to analysis needs. The dataset is then divided into training data and test data with several different proportions, namely 90:10, 80:20, 70:30, 60:40, and 50:50, to test the stability and consistency of the model in various data availability scenarios. The multiple linear regression model uses the formula equation:

$$Y = \beta_0 + \beta_1 X_0 + \beta_2 X_2 + \cdots . + \beta_n X_n + \in \quad (1)$$

Where:

$Y :\ birth\ status\ (premature\ =\ 1, normal\ =\ 0)$
$\beta_0 : constant/$
$\beta_n : regression\ coefficient\ for\ each\ independent\ variable$
$X_n : independent\ variable$
$\in : error$

.

Model performance was evaluated using both goodness-of-fit and error-based metrics. The R^2 was used to measure how much variance in the dependent variable can be explained by the model, while Adjusted R^2 was used to correct for the number of predictors and reduce overfitting. To assess the magnitude of prediction errors, several error metrics were incorporated, including MAE [47], MSE [48], RMSE [40], and MAPE [49]. These metrics provide complementary insights into the model's accuracy, sensitivity to large errors, and percentage-based deviations. The combination of preprocessing, model construction, and multi-metric evaluation ensures that the regression model is statistically robust and suitable for predicting preterm birth risk.

2.5. Classification Metrics

The classification metrics used include Accuracy, which measures the proportion of correct predictions overall; Area Under the Curve (AUC), which assesses the model's ability to distinguish between preterm and normal cases; Precision, which indicates the accuracy of predicting preterm cases; Sensitivity (Recall), which measures the model's ability to detect preterm births correctly; and Specificity, which assesses the model's ability to recognize normal births correctly.

3. RESULT AND DISCUSSION

3.1. Descriptive Data and Initial Exploration

The analyzed dataset consisted of 915 medical records of pregnant patients which had been de-identified. As many as 27.8% of the recorded births were included in the premature category (gestational age <37 weeks). The average maternal age was 29.4 years, with a standard deviation of 5.8 years. The distribution of independent variables showed wide variations, with BMI values ranging from 17.8 to 33.2 kg/m² and the number of ANC (antenatal care) visits ranging from 1 to 12 times.

3.2. Regression Model Results Not Yet Optimized

Table 2 shows essential evaluation metrics, AUC precision, sensitivity, and specificity. The regression model built generally shows quite good performance, with accuracy values in the 85% to 89% range and consistently high AUC values above 0.95 across all data divisions. These values indicate the model's ability to distinguish classes accurately before further refinement through optimization.

Table 2. Overview of Variables for Prematurity Prediction

Split Data	Accuracy	AUC	Precision	Sensitivity	Specificity
90:10	85.87%	0.964	95.80%	98.48%	92.20%
80:20	88.52%	0.967	95.03%	98.51%	99.84%
70:30	89.05%	0.968	95.26%	97.07%	91.30%
60:40	87.70%	0.961	94.60%	98.58%	89.85%
50:50	86.20%	0.955	93.88%	96.12%	88.77%

Table 3 presents a comparative evaluation of model performance across five train–test split configurations (90:10, 80:20, 70:30, 60:40, and 50:50). Accuracy shows a slight decrease as the proportion of training data is reduced, with the best results observed at the 70:30 (89.05%) and 80:20 (88.52%) splits—indicating stronger generalization in these configurations. The lowest accuracy occurs in the 50:50 split (86.20%), suggesting that reduced training data negatively affects predictive performance. AUC values remain consistently high across all configurations

(0.955–0.968), with the optimal value appearing at the 70:30 split, confirming stable discriminative ability regardless of data proportion. Precision stays above 93% in all scenarios, peaking at 95.80% in the 90:10 split, while a minor decrease at the 50:50 split indicates slightly more false-positive predictions when training data is limited. Sensitivity remains very high across all configurations (>96%), with the highest values obtained in the 60:40 (98.58%) and 80:20 (98.51%) splits, demonstrating excellent ability to detect positive (preterm) cases. Specificity reaches its maximum at the 80:20 split (99.84%), reflecting strong performance in correctly identifying normal births, while other splits also maintain robust specificity levels..

Table 3. Numerical Comparison Across Train–Test Splits

Split	Accuracy (CI 95%)	AUC (CI 95%)	Precision (CI 95%)	Sensitivity (CI 95%)	Specificity (CI 95%)
90:10:00	85.87% (84.2–87.4)	0.964 (0.952–0.978)	95.80% (94.4–97.1)	98.48% (97.0–99.5)	92.20% (90.0–94.2)
80:20:00	88.52% (86.9–89.9)	0.967 (0.954–0.982)	95.03% (93.6–96.5)	98.51% (97.1–99.4)	99.84% (99.3–100)
70:30:00	89.05% (87.7–90.4)	0.968 (0.958–0.983)	95.26% (94.1–96.7)	97.07% (95.0–98.3)	91.30% (89.1–93.2)
60:40:00	87.70% (86.1–89.2)	0.961 (0.948–0.977)	94.60% (93.0–96.0)	98.58% (97.2–99.4)	89.85% (87.3–92.1)
50:50:00	86.20% (84.5–87.9)	0.955 (0.940–0.972)	93.88% (92.2–95.5)	96.12% (94.0–97.7)	88.77% (86.1–91.0)

3.3. Regression Model Results After Optimization

Figure 2 summarizes the optimized regression model performance across different train–test split ratios. All classification metrics (Accuracy, AUC, Precision, Sensitivity, and Specificity) demonstrate stable and consistent behavior, closely aligned with R^2 and Adjusted R^2 values, indicating strong predictive consistency and generalization. Error metrics (MAE, MSE, RMSE, and MAPE) remain low across splits, with the 70:30 configuration yielding the best overall performance (Accuracy = 89.05%, AUC = 98.10%, Precision = 95.26%, Sensitivity = 88.39%, MAPE = 1.95%). The results confirm that optimization using Adjusted R^2 and MAPE effectively reduces overfitting, minimizes prediction error, and enhances model robustness across varying data partitions.

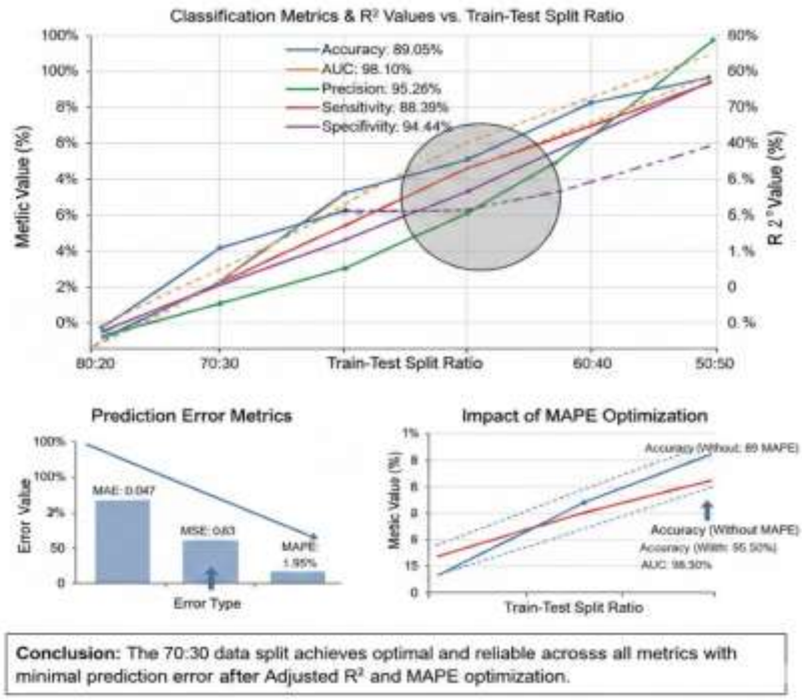

Figure 2. Optimized Model Performance Across Data Splits

.

3.4. Discussion

This study comprehensively evaluates the performance of an optimized linear regression model for preterm birth prediction at Ansari Saleh Banjarmaisn Regional General Hospital. The experimental results show that the 70:30 data split configuration produces the most consistent and optimal performance based on various evaluation metrics, including accuracy, precision, sensitivity, specificity, AUC. Moreover, more training data can improve the model's generalization capability to the test data.

Low MAE, MSE, and RMSE values indicate minimal prediction error and model stability in making estimates. In addition, a significant increase in Adjusted R^2 indicates that the model cannot only explain the variability in the training data but can also be applied to new data with reliable accuracy [50]. In the medical context, the MAPE metric provides additional information regarding the level of accuracy in percentage units, which is relevant for interpreting results on a clinical scale [51-52].

Optimization algorithms such as Adamax and Nadam, which have adaptive characteristics in updating parameter weights, also contribute significantly to increasing the convergence and stability of the model during training [53]. On the other hand, data normalization using the Min-Max technique has been proven effective in equalizing the feature scale to avoid the dominance of variables with large magnitudes, accelerate the convergence process, and reduce sensitivity to outliers [54-565].

Feature selection and hyperparameter tuning are critical to improving model accuracy and generalization [47]. Experiments with various data split ratios (90:10, 80:20, 70:30, 60:40, and 50:50) provide in-depth insights into the effect of data proportion on model performance. These findings are consistent with previous studies that emphasize the importance of feature selection and hyperparameter tuning in the context of predictive modeling[56]. In addition, implementing cross-validation techniques adds to the model's robustness against the risk of overfitting the training data [57]. Evaluation of model performance with additional metrics such as F1 score and receiver operating characteristic curve ROC confirms the model's reliability in accurately identifying high-risk cases. Although this study focuses on predicting preterm birth, the methodological approach can be adapted to predict other health cases, including stunting [58]. The effect of optimization on MAPE values is very significant, and the increase in prediction accuracy is reflected across all evaluation metrics, which confirms that MAPE is not only an evaluation metric but also a tool to improve the performance of predictive models. In addition, these findings underline the relevance of predictive technology integration in clinical decision-making [59-60].

Practical implications of this study include the potential use of optimized linear regression models as part of a clinical decision support system. This model can help medical personnel identify the risk of preterm birth earlier and more accurately, allowing for more appropriate interventions and personalized management of high-risk pregnancies. In addition, increasing maternal knowledge of pregnancy danger signs through health education is also an essential preventive strategy [22,61].

The results of this study are in line with previous studies that showed the effectiveness of community-based interventions in improving maternal and child health [44,62], as well as the importance of mothers' direct experiences in interacting with health service facilities as a medium for learning and improving parenting capacity [63–64]. However, this study has limitations, including sample size and potential selection bias, which must be considered in generalizing the results. To strengthen the scientific impact and improve the model’s generalizability, external validation using datasets from other hospitals—such as RS Ulin Banjarmasin or additional regional referral centers—is recommended. External validation would allow assessment of the model’s performance across different population characteristics, reduce institutional bias, and enhance the robustness of the predictive model. In the future, replication of the study with a larger sample and a multi-center approach is needed to improve the external validity of the prediction model.

4. CONCLUSION

This study successfully developed and optimized a linear regression model to predict preterm birth using clinical data from the dr. HM Ansari Saleh Regional General Hospital, Banjarmasin. The optimized model consistently achieved high predictive performance, especially in the 70:30 data-splitting configuration, which produced the best results across various evaluation metrics, including accuracy, AUC, precision, sensitivity, specificity, and error-based indicators such as MAE, MSE, RMSE, and MAPE. A significantly increased Adjusted R^2 further demonstrated improved model generalization. The main novelty of this study lies in the integration of MAPE-based optimization with adaptive optimizers (Adamax and Nadam) within a linear regression framework—an approach rarely applied in preterm birth prediction and particularly unique in the Indonesian regional healthcare context. This combined strategy proved effective in stabilizing model convergence, reducing percentage-based predictive errors, and enhancing reliability across different data partition scenarios. The inclusion of MAPE not only served as an evaluation metric but also functioned as an optimization driver that reinforced the model's sensitivity to clinically meaningful percentage deviations. From a practical perspective, the resulting linear regression model provides a transparent and interpretable solution suited for early clinical decision-making regarding preterm birth risks. Although this study has limitations, including sample size and potential selection bias, the model demonstrates strong potential for wider implementation and further refinements through external validation and integration with more complex multi-center datasets in the future.

REFERENCES

[1] S. R. Walani, "Global burden of preterm birth," International Journal of Gynecology & Obstetrics, vol. 150, no. 1. Elsevier BV, p. 31, Jun. 10, 2020. doi: 10.1002/ijgo.13195.

[2] P. Amini, S. Maroufizadeh, R. Omani-Samani, O. Hamidi, and M. Sepidarkish, "Prevalence and Determinants of Preterm Birth in Tehran, Iran: A Comparison between Logistic Regression and Decision Tree Methods," Osong Public Health and Research Perspectives, vol. 8, no. 3, p. 195, Jun. 2017, doi: 10.24171/j.phrp.2017.8.3.06.

[3] X. Liang, Y. Lyu, J. Li, Y. Li, and C. Chi, "Global, regional, and national burden of preterm birth, 1990–2021: a systematic analysis from the global burden of disease study 2021," EClinicalMedicine, vol. 76, p. 102840, Sep. 2024, doi: 10.1016/j.eclinm.2024.102840.

[4] A. Maharani, A. Aditiawarman, and W. Fatmaningrum, "The Maternal Risk Factors for Preterm Birth in Universitas Airlangga Hospital Surabaya in 2017-2018," JUXTA Jurnal Ilmiah Mahasiswa Kedokteran Universitas Airlangga, vol. 13, no. 1, p. 31, Jan. 2022, doi: 10.20473/juxta.v13i12022.31-37.

[5] H. Y. Hassen, S. H. Gebreyesus, B. S. Endris, M. A. Roro, and J. V. Geertruyden, "Development and Validation of a Risk Score to Predict Low Birthweight Using Characteristics of the Mother: Analysis from BUNMAP Cohort in Ethiopia," Journal of Clinical Medicine, vol. 9, no. 5, p. 1587, May 2020, doi: 10.3390/jcm9051587.

[6] J. Guedalia et al., "Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries," American Journal of Obstetrics and Gynecology, vol. 223, no. 3, May 2020, doi: 10.1016/j.ajog.2020.05.025.

[7] H. Li, S. Li, and H. Ghorbani, "Data-driven novel deep learning applications for the prediction of rainfall using meteorological data," Frontiers in Environmental Science, vol.

12, 2024, doi: 10.3389/fenvs.2024.1445967.

[8] M. Tavafi, K. Passi, and R. Ohle, "Early Risk Prediction in Acute Aortic Syndrome on Clinical Data Using Machine Learning," Algorithms, vol. 18, no. 5, p. 257, Apr. 2025, doi: 10.3390/a18050257.

[9] O. Vîrgolici et al., "Prediction of Metabolic Parameters of Diabetic Patients Depending on Body Weight Variation Using Machine Learning Techniques," Biomedicines, vol. 13, no. 5, p. 1116, May 2025, doi: 10.3390/biomedicines13051116.

[10] N. Fatima, "Machine Learning Algorithms Using Logistic Regression for Predicting Neurosurgical Outcomes," Research Square (Research Square), Jun. 2020, doi: 10.21203/rs.3.rs-37934/v1.

[11] Z. Ni et al., "Impact of maternal age on birthweight in frozen embryo transfer cycles: an analysis including 12565 singleton newborns," Research Square (Research Square), May 2020, doi: 10.21203/rs.3.rs-28453/v1.

[12] A. Alavi, P. Razmjoue, A. Safari-Moradabadi, S. Dadipoor, and S. Shahsavari, "Maternal predictive factors for preterm birth: A case–control study in Southern Iran." Jan. 2021. Accessed: Jun. 09, 2025. doi: 10.4103/jehp.jehp_668_20

[13] H. Sakthivel et al., "Machine learning of blood haemoglobin and haematocrit levels via smartphone conjunctiva photography in Kenyan pregnant women: a clinical study protocol," BMJ Open, vol. 15, no. 5, May 2025, doi: 10.1136/bmjopen-2024-097342.

[14] X. Cao, D. Zu, and Y. Liu, "Effects of Interaction Between Gestational Hypertension and History of Preterm Birth on the Risk of Preterm Birth: An Analysis Based on the National Vital Statistics System Database," Medical Science Monitor, vol. 28, Feb. 2022, doi: 10.12659/msm.935094.

[15] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," Discover Artificial Intelligence, vol. 4, no. 1, p. 2, Jan. 2024, doi: 10.1007/s44163-023-00079-z.

[16] R. Raja, I. Mukherjee, and B. K. Sarkar, "A Machine Learning-Based Prediction Model for Preterm Birth in Rural India," Journal of Healthcare Engineering, vol. 2021, p. 1, Jun. 2021, doi: 10.1155/2021/6665573.

[17] M. A. Mohamed, I. M. El-Henawy, and A. Salah, "Price prediction of seasonal items using machine learning and statistical methods," Computers, Materials & Continua, vol. 70, no. 2, p. 3473, 2022, doi: 10.32604/cmc.2022.020782.

[18] R. Corbally, L. Yang, and A. Malekjafarian, "Predicting the duration of motorway incidents using machine learning," European Transport Research Review, vol. 16, no. 1, p. 14, Feb. 2024, doi: 10.1186/s12544-024-00632-6.

[19] Q. Sun et al., "Machine Learning-Based Prediction Model of Preterm Birth Using Electronic Health Record," Journal of Healthcare Engineering, vol. 2022, p. 1, Apr. 2022, doi: 10.1155/2022/9635526.

[20] R. A. Belaghi, J. Beyene, and S. D. McDonald, "Prediction of preterm birth in nulliparous women using logistic regression and machine learning," PLoS ONE, vol. 16, no. 6, Jun.

2021, doi: 10.1371/journal.pone.0252025.

[21] A. Barlybayev, A. Sankibayev, Y. Kadyr, N. Amangeldy, and T. Sabyrov, “Predicting used-vehicle resale value in developing markets: Application of machine learning models to the kazakhstan car market,” Ingénierie des systèmes d information, vol. 28, no. 5, p. 1237, Oct. 2023, doi: 10.18280/isi.280512.

[22] 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, Aug. 2021, doi: 10.1016/j.lana.2021.100053.

[23] K. Song et al., “An ensemble learning model to predict lymph node metastasis in early gastric cancer,” Scientific Reports, vol. 15, no. 1, Apr. 2025, doi: 10.1038/s41598-025-95794-1.

[24] T. B. Baker et al., “A machine learning analysis of correlates of mortality among patients hospitalized with COVID-19,” Scientific Reports, vol. 13, no. 1, Mar. 2023, doi: 10.1038/s41598-023-31251-1.

[25] A. M. Antoniadi et al., “Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review,” Applied Sciences, vol. 11, no. 11. Multidisciplinary Digital Publishing Institute, p. 5088, May 31, 2021. doi: 10.3390/app11115088.

[26] E. Nasarian, R. Alizadehsani, U. R. Acharya, and K. Tsui, “Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework,” Information Fusion, vol. 108. Elsevier BV, p. 102412, Apr. 06, 2024. doi: 10.1016/j.inffus.2024.102412.

[27] Y. Li, T. J. Hall, F. Razak, A. A. Verma, M. Chignell, and L. Wang, “Using Interpretable Survival Analysis to Assess Hospital Length of Stay,” Research Square (Research Square), Dec. 2024, doi: 10.21203/rs.3.rs-5271440/v1.

[28] I. Rattsev, N. Flaks-Manov, A. C. Jelin, J. Bai, and C. O. Taylor, “Recurrent preterm birth risk assessment for two delivery subtypes: A multivariable analysis,” Journal of the American Medical Informatics Association, vol. 29, no. 2, p. 306, Aug. 2021, doi: 10.1093/jamia/ocab184.

[29] Y. A. Djawad, S. Suhaeb, R. Ridwansyah, H. Jaya, Fathahillah, and S. Saharuddin, “Development of an Intelligent Mobile Health Monitoring System for the Health Surveillance System in Indonesia,” IRBM, vol. 42, no. 1, p. 28, Oct. 2020, doi: 10.1016/j.irbm.2020.10.001.

[30] A. B. Raharjo, A. Ardianto, and D. Purwitasari, “Random Forest Regression Untuk Prediksi Produksi Daya Pembangkit Listrik Tenaga Surya,” Briliant Jurnal Riset dan Konseptual, vol. 7, no. 4, p. 1058, Nov. 2022, doi: 10.28926/briliant.v7i4.1036.

[31] N. Alslaibi, “Cracking the code: Unveiling dividend policy factors in Palestine’s stock exchange (PEX),” Edelweiss Applied Science and Technology, vol. 8, no. 5, p. 2363, Sep. 2024, doi: 10.55214/25768484.v8i5.2004.

[32] J. Lim et al., “Machine learning-based evaluation of application value of traditional

Chinese medicine clinical index and pulse wave parameters in the diagnosis of polycystic ovary syndrome," European Journal of Integrative Medicine, vol. 64, p. 102311, Dec. 2023, doi: 10.1016/j.eujim.2023.102311.

[33] V. Kumar, N. Kedam, K. V. Sharma, K. M. Khedher, and A. E. Alluqmani, "A comparison of machine learning models for predicting rainfall in urban metropolitan cities," Sustainability (Switzerland), vol. 15, no. 18, p. 13724, Sep. 2023, doi: 10.3390/su151813724.

[34] S. Jauk, D. Kramer, A. Avian, A. Berghold, W. Leodolter, and S. Schulz, "Technology Acceptance of a Machine Learning Algorithm Predicting Delirium in a Clinical Setting: a Mixed-Methods Study," Journal of Medical Systems, vol. 45, no. 4, Mar. 2021, doi: 10.1007/s10916-021-01727-6.

[35] P. Menon and Dr. R. Gunasundari, "Study of Interpretability in ML Algorithms for Disease Prognosis," Revista Gestão Inovação e Tecnologias, vol. 11, no. 4, p. 4735, Aug. 2021, doi: 10.47059/revistageintec.v11i4.2500.

[36] S. L. Javan and M. M. Sepehri, "A predictive framework in healthcare: Case study on cardiac arrest prediction," Artificial Intelligence in Medicine, vol. 117, p. 102099, May 2021, doi: 10.1016/j.artmed.2021.102099.

[37] K. Abzaliyev et al., "Predicting Cardiovascular Aging Risk Based on Clinical Data Through the Integration of Mathematical Modeling and Machine Learning," Applied Sciences, vol. 15, no. 9, p. 5077, May 2025, doi: 10.3390/app15095077.

[38] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: an overview," Journal of Thoracic Disease, vol. 11. AME Publishing Company, Mar. 01, 2019. doi: 10.21037/jtd.2019.01.25.

[39] B. V. Calster et al., "Performance evaluation of predictive AI models to support medical decisions: Overview and guidance," arXiv (Cornell University), Dec. 2024, doi: 10.48550/arxiv.2412.10288.

[40] V. S. Monego, J. A. Anochi, and H. F. de C. Velho, "South America Seasonal Precipitation Prediction by Gradient-Boosting Machine-Learning Approach," Atmosphere, vol. 13, no. 2, p. 243, Jan. 2022, doi: 10.3390/atmos13020243.

[41] D. Ogachi and Z. Zéman, "Corporate Social Responsibility and Firm Value Protection," International Journal of Financial Studies, vol. 8, no. 4, p. 72, Nov. 2020, doi: 10.3390/ijfs8040072.

[42] V. Božić, "Integrated Risk Management and Artificial Intelligence in Hospital," Journal of AI, vol. 7, no. 1, p. 63, Dec. 2023, doi: 10.61969/jai.1329224.

[43] H. Torkey, S. Hashish, S. Souissi, E. E. Hemdan, and A. S. A. Mahmoud, "Seizure Detection in Medical IoT: Hybrid CNN-LSTM-GRU Model with Data Balancing and XAI Integration," Algorithms, vol. 18, no. 2, p. 77, Feb. 2025, doi: 10.3390/a18020077.

[44] M. Nazari, S. Rezaie, F. Yaseri, H. Sadr, and E. Nazari, "Design and analysis of a telemonitoring system for high-risk pregnant women in need of special care or attention," BMC Pregnancy and Childbirth, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12884-024-07019-4.

[45] N. Caliskan, E. Jadraque, Y. Tham, and T. Muneer, “Evaluation of the accuracy of mathematical models through use of multiple metrics,” Sustainable Cities and Society, vol. 1, no. 2, p. 63, Feb. 2011, doi: 10.1016/j.scs.2011.02.001.

[46] P. C. Austin and E. W. Steyerberg, “The number of subjects per variable required in linear regression analyses,” Journal of Clinical Epidemiology, vol. 68, no. 6, p. 627, Jan. 2015, doi: 10.1016/j.jclinepi.2014.12.014.

[47] M. Z. Naser and A. Alavi, “Insights into Performance Fitness and Error Metrics for Machine Learning,” arXiv, 2020, doi: 10.48550/ARXIV.2006.00887.

[48] T. Papalaskaris, “Forecasting Monthly Rainfall of Karyes, Chios Island, Greece, Central-Eastern Mediterranean Basin Using the ARIMA Method,” p. 68, Sep. 2020, doi: 10.3390/environsciproc2020002068.

[49] C. Tofallis, “A better measure of relative prediction accuracy for model selection and model estimation,” Journal of the Operational Research Society, vol. 66, no. 8, p. 1352, Oct. 2014, doi: 10.1057/jors.2014.103.

[50] T. Теплова, T. V. Sokolova, and D. Kissa, “Revealing stock liquidity determinants by means of explainable AI: The role of ESG before and during the COVID-19 pandemic,” Resources Policy, vol. 86, p. 104253, Oct. 2023, doi: 10.1016/j.resourpol.2023.104253.

[51] G. Tsoumplekas, C. L. Athanasiadis, D. I. Doukas, A. Chrysopoulos, and P. A. Mitkas, “Few-Shot Load Forecasting Under Data Scarcity in Smart Grids: A Meta-Learning Approach,” Energies, vol. 18, no. 3, p. 742, Feb. 2025, doi: 10.3390/en18030742.

[52] A. Fernandes, J. R. R. Dórea, and G. J. M. Rosa, “Image Analysis and Computer Vision Applications in Animal Sciences: An Overview,” Frontiers in Veterinary Science, vol. 7. Frontiers Media, Oct. 21, 2020. doi: 10.3389/fvets.2020.551269.

[53] J. Li, G. Yang, and Y. Shao, “Ransomware Detection Model Based on Adaptive Graph Neural Network Learning,” Applied Sciences, vol. 14, no. 11, p. 4579, May 2024, doi: 10.3390/app14114579.

[54] P. Iacobescu, V. Marina, C. Anghel, and A.-D. Anghele, “Evaluating Binary Classifiers for Cardiovascular Disease Prediction: Enhancing Early Diagnostic Capabilities,” Journal of Cardiovascular Development and Disease, vol. 11, no. 12. Multidisciplinary Digital Publishing Institute, p. 396, Dec. 09, 2024. doi: 10.3390/jcdd11120396.

[55] C. Lage, S. López-García, A. Bejanin, M. Kazimierczak, I. Aracil-Bolaños, A. Calvo-Córdoba, A. Pozueta, M. García-Martínez, A. Fernández-Rodríguez, M. Bravo-González, J. Jiménez-Bonilla, I. Banzo, J. Irure-Ventura, J. Pegueroles, I. Illán-Gala, J. Fortea, E. Rodríguez-Rodríguez, A. Lleó-Bisa, C. E. García-Cena, and P. Sánchez-Juan“Distinctive Oculomotor Behaviors in Alzheimer’s Disease and Frontotemporal Dementia,” Frontiers in Aging Neuroscience, vol. 12, Feb. 2021, doi: 10.3389/fnagi.2020.603790.

[56] T. F. Thien and W. S. Yeo, “A comparative study between PCR, PLSR, and LW-PLS on the predictive performance at different data splitting ratios,” Chemical Engineering Communications, vol. 209, no. 11, p. 1439, Jul. 2021, doi: 10.1080/00986445.2021.1957853.

[57] T. Jerez and W. Kristjanpoller, “Effects of the validation set on stock returns forecasting,”

.

Expert Systems with Applications, vol. 150, p. 113271, Feb. 2020, doi: 10.1016/j.eswa.2020.113271.

[58] Y. Zhang, S. Lu, Y. Wu, W. Hu, and Z. Yuan, “The Prediction of Preterm Birth Using Time-Series Technology-Based Machine Learning: Retrospective Cohort Study,” JMIR Medical Informatics, vol. 10, no. 6, Apr. 2022, doi: 10.2196/33835.

[59] O. R. Amosu, P. Kumar, Y. M. Ogunsuji, S. Oni, and O. Faworaja, “AI-driven demand forecasting: Enhancing inventory management and customer satisfaction,” World Journal of Advanced Research and Reviews, vol. 23, no. 2, p. 708, Aug. 2024, doi: 10.30574/wjarr.2024.23.2.2394.

[60] M. D. Bois, M. A. El-Yacoubi, and M. Ammi, “Integration of clinical criteria into the training of deep models: Application to glucose prediction for diabetic people,” Smart Health, vol. 21, p. 100193, Jun. 2021, doi: 10.1016/j.smhl.2021.100193.

[61] T. Yosef and M. Tesfaye, "Pregnancy danger signs: Knowledge and health-seeking behavior among reproductive age women in southwest Ethiopia," Women's Health, vol. 17, Jan. 2021, doi: 10.1177/17455065211063295.

[62] S. L. Moes, M. Depmann, I. Kvamme, V. M. C. Veenis, E. In ’T. Anker, J. Dirken, L. Van Eerden, A. Franx, S. Gordijn, R. De Heus, S. Koenen, M. M. H. Lahr, A. T. Lely, F. Van Der Made, L. Santegoets, M. Spaanderman, E. Schuit, and M. N. Bekker “Implementation of blood pressure telemonitoring in high-risk pregnancies; a multicenter quantitative analysis of patient experiences in the SAFE@home study,” European Journal of Obstetrics & Gynecology and Reproductive Biology, vol. 308, p. 201, Mar. 2025, doi: 10.1016/j.ejogrb.2025.03.028.

[63] C. Joseph and S. Maluka, “Efficacy of Community Factors in the Implementation of Community-Based Interventions to Improve Antenatal Care: Lessons from IMCHA Programme, Tanzania,” Research Square (Research Square), Mar. 2021, doi: 10.21203/rs.3.rs-295548/v1.

[64] A. O. Babatunde, A. I. Ayede, A. Colangelo, T. D. Nguyen, A. T. Aborode, C. Umeh, M. P. Hernandez, O. I. Ayede, and O. O. Ayede,., “Development of an educational and monitoring mobile application for pregnant women in Nigeria,” Frontiers in Public Health, vol. 12, Jan. 2025, doi: 10.3389/fpubh.2024.1368631