Pakistan Journal of Statistics and Operation Research

Predictive Accuracy of Logistic Regression and Support Vector Machine for Short Interpregnancy Interval

Asif Hanif[1*], Tahira Ashraf[2], Nyi Nyi Naing[3], Nadiah Wan-Arfah[4], Mirza Rizwan Sajid[5]

* Corresponding author

1. Department of Biostatistics, Faculty of medicine, Sakarya University, Sakarya-Türkiye; asifhanif@sakarya.edu.tr
2. Medical Research and Statistical Consultancy Training Centre, Pakistan; tahira.ashraf.biostats@gmail.com
3. Faculty of Medicine, Universiti Sultan Zainal Abidin, Malaysia; syedhatim@unisza.edu.my
4. Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Malaysia; wanwaj@unisza.edu.my
5. Department of Statistics, University of Gujrat mirza.rizwan@uog.edu.pk

Abstract

Support vector machine (SVM) is considered a robust machine learning (ML) algorithm. In contrast, Logistic regression (LR) is the most preferred statistical model especially in healthcare and medical field due to its interpretability and mathematical foundations. Considering the competitive characteristics of these models, the predictive and discriminative strength of these models have been tested in this study. Short interpregnancy interval (SIPI) is a global public health issue and is associated with several feto-maternal complications. This study aims to identify the risk factors of SIPI and compare the predictive accuracy of LR vs SVM. Further, feature importance of both models will also be computed and compared. This study was conducted on 528 Pakistani pregnant females and their status of SIPI was predicted through number of risk factors. Various evaluation matrices have been computed to assess the superiority of model. Results have shown that the overall accuracy for LR was 83.14, while Sensitivity, Specificity, PPV, and NPV were 81.6%, 85.23%, 84.58% and 81.82%, respectively. The discriminating strength of this model is 92.1% and examined through receiver operating characteristic (ROC) curve. SVM yielded 94.70% accuracy, with Sensitivity, Specificity, PPV, and NPV as 95.08%, 94.32%, 94.36% and 95.04%, respectively. Further, ROC value was 98.83%. These findings suggests that SVM is better algorithm in predicting SIPI. All measures of predictive analysis as well as model fit indices were better in SVM. Hence, SVM is a comprehensive, interactive, flexible and accurate ML tool that can be used for better predictions of risk factors of SIPI compared to LR. Further, this ML algorithm is free from certain statistical assumptions like linearity of logits, model specification and weak multicollinearity as required in LR models.

Key Words: Feature Importance, Short Interpregnancy Interval, Predictive Accuracy, Logistic Regression, Support Vector Machine.

Introduction

Machine learning (ML) models have significantly revolutionized the various domains of knowledge especially Statistics (Maryam, Anggoro, Tika, & Kusumawati, 2022). In statistical modeling, there are various associated assumption which are hard to fulfill in real world scenarios. In classification problems, conventional logistic regression (LR) is one of the most preferred model to predict the categorical outcome variable. However, this model assumes linear relationships between continuous predictors and the logits $(\frac{p}{1-p})$, which may not always be the case in complex, real-world data (Hazra & Gogtay, 2017; Sajid et al., 2021). This model is also influenced by multicollinearity issues (Algamal & Hammood, 2022) and need correct model specification for unbiased parameters estimation (Fernandes, Figueiredo Filho, Rocha, & Nascimento, 2021). In contrast, ML models are flexible in such assumptions and provide the best alternative with more predictive accuracies. Support vector machine (SVM) is one of the robust ML model. The nature of the identified relationships is non-linear, and currently, SVM has many advantages over traditional regression models, in particular in dealing with high-dimensional data (Huang et al., 2018; Van Messem, 2020). SVM finds the optimal boundary between different classes by constructing hyperplanes in higher-dimensional

spaces (Xu, Zomer, & Brereton, 2006). It has shown encouraging results in different classification tasks, including medical predictions (Cervantes, Garcia-Lamont, Rodríguez-Mazahua, & Lopez, 2020; Huang et al., 2018).

There is vast literature in recent past in which superiority of ML models can be witnessed. Especially, public health experts, practitioners are now preferring these models over statistical models due to their better performance. Researchers have used ML models in early detection of cardiovascular diseases (Ogunpola, Saeed, Basurra, Albarrak, & Qasem, 2024; Sajid et al., 2021), multiple disease prediction (Arumugam et al., 2023), air quality and smart ehealth system (Rahi, Sood, Bajaj, & Kumar, 2021), lung cancer classification (Gomiasti, Warto, Kartikadarma, Gondohanindijo, & Setiadi, 2024) and many more domains of healthcare. Considering these applications of SVM and other ML models, this study will provide a comparison between two models; LR and SVM by applying them on a very serious public health issue.

The interval between pregnancies referred to as interpregnancy interval (IPI), plays an important role in maternal and neonatal health outcomes. A short interpregnancy interval (SIPI), defined as an interval of less than 24 months by WHO, has been associated with various fetal, neonatal, and maternal adverse outcomes including gestational diabetes, gestational hypertension, preterm labor, premature birth, low birth weight, neurodevelopment disorders, prolonged hospitalizations, and perinatal, neonatal, and infant deaths (Mamo, Dagnaw, Sharew, Brhane, & Kotiso, 2021; A. Weiss, Sela, Rotem, Grisaru-Granovsky, & Rottenstreich, 2021). Despite ongoing awareness campaigns, the prevalence rate of SIPI remains high, especially in developing countries like Pakistan. Approximately 25% of all births worldwide occur at a short interval (Byamukama et al., 2022). In 2022, the reported prevalence rate of SIPI was 22.9% in Pakistan (Nausheen et al., 2021). Therefore, there is need to early detection of this condition with more accuracy.

Various risk factors of SIPI have been reported in the previous literature and identifying these is important for the development of effective interventions for maternal and child well-being (Sakai-Bizmark et al., 2024; Sujan, Class, Rickert, Van Hulle, & D'Onofrio, 2021). Accurate prediction models can improve outcomes by identifying high risk populations. LR is the most commonly used statistical method in medical sciences to predict various health outcomes (Hanley, Hutcheon, Kinniburgh, & Lee, 2017; Ari Weiss, Sela, Grisaru-Granovsky, & Rottenstreich, 2023). It is a powerful tool for analyzing binary outcomes, where the dependent variable is categorical (Harris, 2021), such as whether the IPI is short or not. However, findings of this conventional model will be compared with SVM based model.

Therefore, this study was designed to explore risk factors for SIPI in the local population. In addition, this study will serve as a cornerstone for comparing the predictive accuracy and model fit indices of LR and SVM in the Pakistani population. The findings of this research will inform future methodological approaches to predictive modeling in maternal health and contribute to the broader understanding of reproductive health risk factors. This study also seeks to highlight the potential of ML methods in improving health outcomes and advancing the precision of public health interventions aimed at reducing the incidence of short interpregnancy intervals.

Materials and Methods

This study used a retrospective cohort study design and two groups were made i.e. Short interpregnancy interval (SIP) and normal interpregnancy interval (NIPI). These two groups acted as binary dependent variable which will be predicted through conventional LR and SVM model. The study was completed in 2 years after approval of the proposal (20th Feb 2020 to 20th Feb 2022) and conducted at Lady Aitchison Hospital Lahore. Data was collected through simple random sampling. This study data was derived from a project the sample size was n = 528 (264 for NIPI and 264 for SIPI). In medical studies, inclusion and exclusion criteria is very important for generalization of study findings.

a. Inclusion criteria
For Short IPI: 18-45 years old females having interpregnancy interval less than 24 months.
For normal IPI: 18 - 45 years old females with ≥ 24 months interpregnancy interval.

b. Exclusion criteria
Women who had a history of sub-fertility before their index pregnancy had an unknown last menstrual period, or forgot their last delivery month and year.

The study included a total of 528 pregnant females. They were randomly divided into two groups, i.e. SIPI and NIPI. IPI was calculated as the difference between the delivery date of the previous pregnancy and the conception date of the new pregnancy, as examined by ultrasound (Khan & Khanam, 2023). In the current study, information was noted on a questionnaire developed with the help of a literature review and validated questionnaires to gather sociodemographic and clinical details and evaluate knowledge about family planning and birth spacing importance (A Hanif, Naing, Nadiah, & Ashraf, 2022; A Hanif, Naing, Wan-Arfah, & Ashraf, 2023). Throughout the procedure,

it was ensured that the questionnaire and its administration adhered to ethical guidelines, including obtaining informed consent from participants and maintaining confidentiality. Approval was taken for data collection from the Human Research Ethics Committee [UniSZA.C/2/UHREC/628-2 Jld 2(4)] and the Medical Superintendent office of Lady Aitchison Hospital, Lahore [Letter # 28/744 LAH].

All data was entered and analyzed using DTREG – Predictive modeling program version 10.10.0. In this commercial predictive modeling software, "logit" link function was used with regularization options. Further, AIC and BIC approach were used to remove irrelevant or insignificant predictors. On the other side, Linear and Radial Basis Kernel were used with varying cost functions (0.1 to 100). In linear kernel decision function is represented with following equation

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i < x_i, x > + b$$

Where

α_i are the support vector coefficients,

y_i are the actual class labels (+1 or -1),

x_i are the support vectors, and $< x_i, x >$ is the dot product between training vector and new input x.

b is the bias term.

On the other side, RBF kernel function use the given equation

$$K(x, x_i) = \exp\left(-\gamma ||x - x_i||^2\right)$$

Where:

x = input vector (new observation),

xi = support vector from training data,

$||x - x_i||^2$ = squared Euclidean distance between x and x_i,

γ = kernel parameter that controls the width of the RBF.

Further, Gamma function was also tuned with varied values of 0.01 to 10.0. Grid search strategy was used for parameter optimization for SVM. The MLR and SVM models were all performed and the area under the ROC curve was calculated. P-value ≤ 0.05 was considered significant where hypothesis testing was applicable. Lastly, SVM model was also validated through 10 fold cross validation. After developing these models, feature importance of both models are also computed and compared.

Results

This study has a main focus on the prediction of SIPI which has two possibilities 0 and 1. Therefore, as per prediction model the focus is on the prediction of SIPI =1. According to the predictive accuracy of conventional MLR, the overall accuracy was 83.1%, while Sensitivity, Specificity, PPV, and NPV were 81.6%, 85.2%, 84.6%, and 81.8%, respectively, with an area under the ROC curve as 92.1%. The Geometric mean of sensitivity and specificity was 83.1% and for PPV and NPV, it was 83.2%. Precision measures the proportion of positive class predictions that are correctly classified as belonging to the positive class. Recall is a metric that measures the proportion of positive class predictions correctly identified among all positive examples present in the dataset. Hence, the Precision and recall for LR were 84.6% and 81.1%, respectively (Table-1).

For SVM, the overall accuracy was 94.7%, with Sensitivity, Specificity, PPV, and NPV as 95.1%, 94.3%, 94.4%, and 95.0%, respectively, with the area under the ROC curve as 98.8%. The Geometric mean of sensitivity and specificity was 94.7% and for PPV and NPV, it was 94.8%. The Precision and recall for MLR were 94.4% and 95.1%, respectively. Comparatively, SVM gave better predictive accuracy (in terms of Overall accuracy, Sensitivity, Specificity, PPV, NPV and Area under ROC curve) compared to MLR. According to model fit indices, for LR the values of C-statistics, F-statistics, and correctly classified percentages were 92.0%, 82.8%, and 83.1%, while these model fit indices for SVM were 98.8% (C-statistics), 94.7% (F statistics) and 0.9 (correctly classified percentage). Comparatively, SVM gave better model fit indices (in terms of C-statistics, F-statistics, and correctly classified) as compared to MLR. (Table-1)

LR identified 9 predictors according to their relative importance for SIPI. These included Knowledge of SIPI (100%), Knowledge of family planning (90.7%), desire of more male babies (51.9%), breastfeeding of less than 1 year (18.5%), family doesn't like SIPI (14.8%), Own desire of more babies (13.0%), Parity < 3 (13.0%), Desire of equal male and female babies (5.6%) and History of miscarriage (3.7%). Similar to MLR, SVM also identified 11 factors compared to 9 factors identified by LR that showed the relative importance of predictors associated with SIPI. These included Desire of more male babies (100%), Own desire of more babies (93.8%), Breastfeeding of less than 1 year (61.7%), Desire of equal male and female babies (60.5%), Knowledge of SIPI (45.7%), Family doesn't like SIPI (37.0%),

Knowledge of family planning (30.9%), Parity<3 (21.0%), rural area of residence (19.8%), current marital age (6.2%) and History of miscarriage (3.7%). (Table- 2)

ROC analysis is a graphical method used to evaluate the effectiveness of a classifier in terms of discriminating power. This particular statistic is developed by using sensitivity and specificity of model. For medical practitioners a classifier's performance is characterized by a pair of statistic: the true positive rate and the false positive rate. The probability threshold chart visually represents the probability threshold report, which informs how outcome category assignments are influenced by various probability thresholds. Ideally this curve should be close to top left corner of the graph or numerically close to 1. LR based model reported 0.92 which is significantly less than 0.98 for SVM based model. Further, the Probability Calibration Chart displays the distribution of predicted probabilities for a specific target category and allows for the assessment of the accuracy of these predictions in various deciles or group of dataset. Gain and lift charts are graphical tools used to assess the effectiveness of classification models. The Gain and Lift chart assesses the performance of a model within a specific subset of the population, in contrast to the confusion matrix which evaluates the overall population.

Table-1: Comparison of predictive accuracy and model fit indices of MLR and SVM

| Parameters	MLR	SVM
Accuracy	83.1%	94.7%
Sensitivity	81.6%	95.1%
Specificity	85.2%	94.3%
Geometric mean of sensitivity and specificity	83.1%	94.7%
PPV	84.6%	94.4%
NPV	81.8%	95.0%
Geometric mean of PPV and NPV	83.2%	94.7%
F statistics	82.8%	94.7%
Area under ROC curve (C-statistics)	92.0%	98.8%
C-statistics	92.0%	98.8%
Precision	84.6%	94.4%
Recall	81.1%	95.1%
Correctly classified %	83.1%	
Hosmer – Lemeshow (p-value)	7.31, (0.503)	
[a] Likelihood ratio +	5.51	16.48
[a] Likelihood ratio -	0.2159	0.051
Log likelihood of model	-187.4365	
Deviance (-2 * Log likelihood)	374.873036	
AIC	398.873036	
BIC	450.102191	

a| these were calculated using formulas LR+ = sensitivity/(1 − specificity), LR− = (1 − sensitivity)/specificity

Table -2: Comparison of Variable Importance between MLR and SVM

MLR		SVM	
Variables	Importance	**Variables**	Importance
Knowledge SIPI	100.0%	**Desire of more male**	100.0%
Knowledge FP	90.7%	**Own desire more kids**	93.8%
Desire of more male	51.8%	**Breastfeeding < 1 year**	61.8 %
Breast feeding	18.5%	**Desire of equal male female**	60.5%
Family doesn't like	14.8%	**Knowledge SIPI**	45.7%
Own desire more kids	13.0%	**Family doesn't like**	37.0%
Parity < 3	13.0 %	**Knowledge FP**	30.9%
Desire of equal male female	5.6 %	**Parity <3**	21.0%
History of Miscarriage	3.7 %	**Rural residence**	19.8%
		Current age	6.2%
		History of Miscarriage	3.7%

MRL

SVM
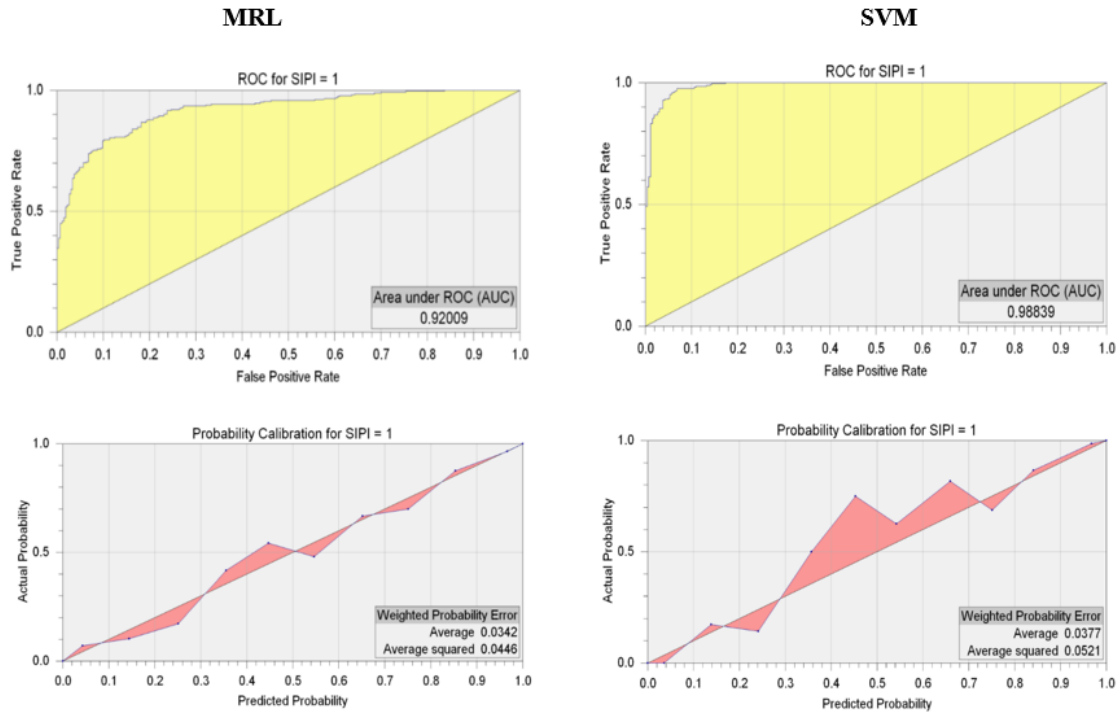
Figure 1: Comparison of ROC and Probability calibrations in MLR and SVM

MRL

SVM
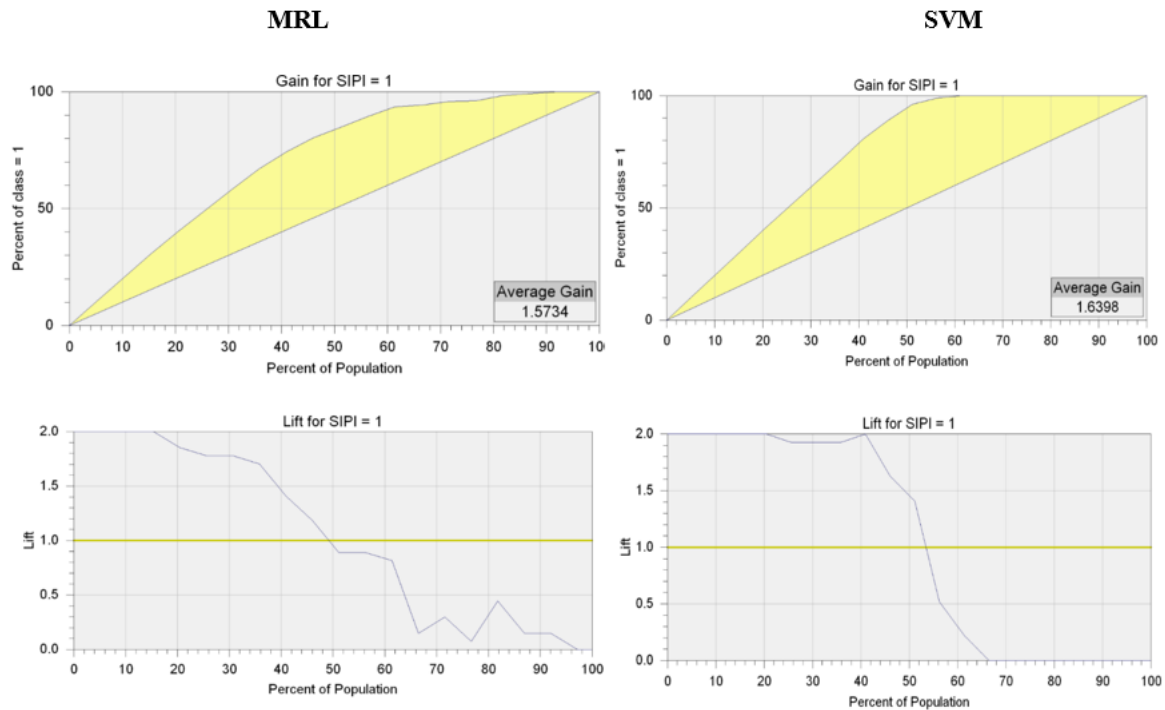
Figure 2: Comparison of Gain and Lift in MLR and SVM

Discussion

Several supervised machine learning models are available and being used in data sciences today which are Decision Trees, Artificial Neural Networks, Support Vector Machine, and Random Forest (Pranckevičius & Marcinkevičius, 2017). Of these, the SVM algorithm is considered significant in solving health challenges because of the high accuracy of computations and its mathematical robustness (Qayyum, Qadir, Bilal, & Al-Fuqaha, 2020). This study findings confirms the superiority of SVM over conventional MLR model. All evaluation matrices especially accuracy and ROC values of SVM are significantly higher than MLR. These findings suggests that SVM can be used as an alternative approach if researchers needs more predictive strength and data possibly have non-linear patterns. A recent literature also augmented these findings indirectly. SVM was utilized for the prediction of feto-maternal conditions, including the prediction of low birth weight, pre-term birth, and intrauterine growth restriction (Zhang et al., 2024), while the current study is the first that attempts to predict the risk factors associated with SIPI. According to the authors' best knowledge, no study has used and compared the modeling effectiveness of SVM to the conventional approach of logistic regression.

SVMs are employed to carry out regression, classification and identification of outliers in data. Some scholars have also indicated that SVM can assist in identifying health conditions, for instance, cancer, blood pressure, and diabetes through medical big data analysis (Guido, Ferrisi, Lofaro, & Conforti, 2024; Huang et al., 2018). The increased SVM use in controlling and responding to global health concerns is anticipated to cause a massive revolution (Razzaghi, Roderick, Safro, & Marko, 2016). MLR and SVM are two different techniques of binary classification. The first one relies on a model and then fits the data to a logistic curve; while SVM, is a relatively complex approach, which converts the input data into a high-dimensional space and forms a hyperplane that best separates the classes (Pranckevičius & Marcinkevičius, 2017). The main advantage of the presented SVM approach is that it is powerful for cases where the size of the training data is small, while the data is high-dimensional, and there is no need for a specific model. This enables it to have strong discriminant power. This technique has been lately used for automated disease classification as well as improving disease detection in healthcare settings (Guido et al., 2024; Naraei, Abhari, & Sadeghian, 2016).

This study has also addressed a potential challenge associated with ML models; black box nature of models. For this purpose, feature importance was also computed for SVM model. In the current study, eleven significant risk factors were identified after applying MLR. These included desire of a male baby, low knowledge of SIPI, low knowledge of family planning, maternal age, own desire to have more kids, rural residence, parity >3, miscarriage, not breastfeeding up to 1 year, and family doesn't like birth spacing. Whereas, in SVM, significant predictors were the desire for more male babies, own desire to have more kids, breastfeeding < 1 year, desire of equal male and female babies, family doesn't like SIPI, low knowledge of SIPI, low knowledge of family planning, parity <3, rural residence, current maternal age, and history of miscarriage. Both models have reported almost similar risk factors. However, the predictive accuracy of SVM was statistically better than the indices of MLR.

Although, no study so far has compared the predictive ability of SVM with LR for determinants of SIPI, however, literature published on other diseases has proved the better accuracy of SVM compared to LR (Safitri, Chamidah, Saifudin, Firmansyah, & Alpandi, 2024; Xiong et al., 2022). A study conducted by Xiong et al., to compare LR and SVM for predicting the severity of COVID-19, revealed that SVM had better predictive accuracy compared to LR (88.5% vs. 85.2%). The sensitivity and specificity were 93.9% and 79%, respectively, for SVM, and 92.3% and 72.3%, respectively, for LR (Xiong et al., 2022). Similarly, in another study to predict the risk of stroke, it was reported that SVM performed better in predicting the overall stroke risk compared to LR (84% vs. 81%) (Safitri et al., 2024). However, another study to predict preterm birth documented superior predictability of LR (AUC = 77.96%) compared to SVM (AUC = 71.70%) (Zhang et al., 2024).

This study has potential limitations as data was collected from single centered and only married females were consider due to the nature of objective of study.

Conclusion

This study concluded that the predictive performance of SVM was statistically better than LR. This superiority is in almost all evaluation matrices. Most importantly, very high C-statistic value of SVM model reflected its far better discriminating strength than MLR. Further, SVM identified different pattern of feature importance in the study which is probably due to non-linearity and some hidden patterns of data which have not been gauged through MLR model. These findings specifically important for practitioners and encourage them to use assumption-free ML models to reveal un-orthodox patterns of data that can lead to more accurate decision making especially in medical domain. Overall, SVM has proven its accuracy and flexibility that can be used for better predictions of risk factors of SIPI compared to conventional MLR.

Reference

1. Algamal, Z. Y., & Hammood, N. (2022). A new Jackknifing ridge estimator for logistic regression model. *Pakistan Journal of Statistics and Operation Research*, 955-961. https://doi.org/10.18187/pjsor.v18i4.3748

2. Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. *Materials today: proceedings, 80*, 3682-3685. https://doi.org/10.1016/j.matpr.2021.07.361

3. Byamukama, O., Migisha, R., Kalyebara, P. K., Tibaijuka, L., Lugobe, H. M., Ngonzi, J., . . . Kanyesigye, H. (2022). Short interbirth interval and associated factors among women with antecedent cesarean deliveries at a tertiary hospital, Southwestern Uganda. *BMC Pregnancy and Childbirth, 22*(1), 268. https://doi:10.1186/s12884-022-04611-4

4. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing, 408*, 189-215. doi:https://doi.org/10.1016/j.neucom.2019.10.118

5. Fernandes, A. A. T., Figueiredo Filho, D. B., Rocha, E. C. d., & Nascimento, W. d. S. (2021). Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política, 28*, 006.

6. Gomiasti, F. S., Warto, W., Kartikadarma, E., Gondohanindijo, J., & Setiadi, D. R. I. M. (2024). Enhancing lung cancer classification effectiveness through hyperparameter-tuned support vector machine. *Journal of Computing Theories and Applications, 1*(4), 396-406. http://dx.doi.org/10.62411/jcta.10106

7. Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information, 15*(4), 235. doi:https://doi.org/10.3390/info15040235

8. Hanif, A., Naing, N., Nadiah, W. A., & Ashraf, T. (2022). Development and validation of a questionnaire about knowledge of birth spacing among females. *Science International Lahore, 34*(6), 589-594.

9. Hanif, A., Naing, N. N., Wan-Arfah, N., & Ashraf, T. (2023). Development and Validation of Knowledge about Family Planning Questionnaire among Pakistani Females. *Journal of Xi'an Shiyou University, Natural Science Edition, 19*(2), 923-935.

10. Hanley, G., Hutcheon, J., Kinniburgh, B., & Lee, L. (2017). Interpregnancy Interval and Adverse Pregnancy Outcomes: An Analysis of Successive Pregnancies. *Obstetrics and gynecology, 129*(3), 1891. doi:10.1097/AOG.0000000000001891

11. Harris, J. K. (2021). Primer on binary logistic regression. *Family Medicine and Community Health, 9*(Suppl 1), e001290. doi:10.1136/fmch-2021-001290

12. Hazra, A., & Gogtay, N. (2017). Biostatistics Series Module 10: Brief Overview of Multivariate Methods. *Indian Journal of Dermatology, 62*(4), 358-366. doi:10.4103/ijd.IJD_296_17

13. Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics, 15*(1), 41-51. doi:10.21873/cgp.20063

14. Khan, M. N., & Khanam, S. J. (2023). The effectiveness of WHO's interpregnancy interval advice. *The Lancet Global Health, 11*(10), e1476-e1477. doi:10.1016/S2214-109X(23)00402-3

15. Mamo, H., Dagnaw, A., Sharew, N. T., Brhane, K., & Kotiso, K. S. (2021). Prevalence of short interpregnancy interval and its associated factors among pregnant women in Debre Berhan town, Ethiopia. *Plos one, 16*(8), e0255613. doi:https://doi.org/10.1371/journal.pone.0255613

16. Maryam, M., Anggoro, D. A., Tika, M. F., & Kusumawati, F. C. (2022). An intelligent hybrid model using artificial neural networks and particle swarm optimization technique for financial crisis prediction. *Pakistan Journal of Statistics and Operation Research*, 1015-1025. https://doi.org/10.18187/pjsor.v18i4.3927

17. Naraei, P., Abhari, A., & Sadeghian, A. (2016). *Application of multilayer perceptron neural networks and support vector machines in classification of healthcare data.* Paper presented at the 2016 Future Technologies Conference (FTC). https://doi.org/10.1109/FTC.2016.7821702

18. Nausheen, S., Bhura, M., Hackett, K., Hussain, I., Shaikh, Z., Rizvi, A., . . . Soofi, S. (2021). Determinants of short birth intervals among married women: a cross-sectional study in Karachi, Pakistan. *BMJ Open, 11*(4), e043786. doi:10.1136/bmjopen-2020-043786

19. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics, 14*(2), 144. https://doi.org/10.3390/diagnostics14020144

20. Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing, 5*(2), 221. http://dx.doi.org/10.22364/bjmc.2017.5.2.04

21. Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2020). Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering, 14*, 156-180. https://doi.org/10.1109/rbme.2020.3013489

22. Rahi, P., Sood, S. P., Bajaj, R., & Kumar, Y. (2021). Air quality monitoring for Smart eHealth system using firefly optimization and support vector machine. *International Journal of Information Technology, 13*(5), 1847-1859. https://doi.org/10.1007/s41870-021-00778-9

23. Razzaghi, T., Roderick, O., Safro, I., & Marko, N. (2016). Multilevel weighted support vector machine for classification on healthcare data with missing values. *Plos one, 11*(5), e0155119. https://doi.org/10.1371/journal.pone.0155119

24. Safitri, L. R., Chamidah, N., Saifudin, T., Firmansyah, M., & Alpandi, G. T. (2024). Comparison of Logistic Regression and Support Vector Machine in Predicting Stroke Risk. *Inferensi, 7*(2), 123-128. doi:10.12962/j27213862.v7i2.20420

25. Sajid, M. R., Muhammad, N., Zakaria, R., Shahbaz, A., Bukhari, S. A. C., Kadry, S., & Suresh, A. (2021). Nonclinical features in predictive modeling of cardiovascular diseases: a machine learning approach. *Interdisciplinary Sciences: Computational Life Sciences, 13*, 201-211. https://doi.org/10.1007/s12539-021-00423-w

26. Sakai-Bizmark, R., Jackson, N. J., Wu, F., Marr, E. H., Kumamaru, H., Estevez, D., . . . Henwood, B. F. (2024). Short Interpregnancy Intervals Among Women Experiencing Homelessness in Colorado. *JAMA Netw Open, 7*(1), e2350242. doi:10.1001/jamanetworkopen.2023.50242

27. Sujan, A. C., Class, Q. A., Rickert, M. E., Van Hulle, C., & D'Onofrio, B. M. (2021). Risk factors and child outcomes associated with short and long interpregnancy intervals. *Early Child Dev Care, 191*(14), 2281-2292. doi:10.1080/03004430.2019.1703111

28. Van Messem, A. (2020). Chapter 10 - Support vector machines: A robust prediction method with applications in bioinformatics. In A. S. R. Srinivasa Rao & C. R. Rao (Eds.), *Handbook of Statistics* (Vol. 43, pp. 391-466): Elsevier.

29. Weiss, A., Sela, H. Y., Grisaru-Granovsky, S., & Rottenstreich, M. (2023). Short Interpregnancy Interval Following a Multifetal Pregnancy: Maternal and Neonatal Outcomes. *Journal of Clinical Medicine, 12*(7), 2576. doi:https://doi.org/10.3390/jcm12072576

30. Weiss, A., Sela, H. Y., Rotem, R., Grisaru-Granovsky, S., & Rottenstreich, M. (2021). Recurrent short interpregnancy interval: Maternal and neonatal outcomes. *Eur J Obstet Gynecol Reprod Biol, 264*, 299-305. doi:10.1016/j.ejogrb.2021.07.040

31. Xiong, Y., Ma, Y., Ruan, L., Li, D., Lu, C., Huang, L., & the National Traditional Chinese Medicine Medical, T. (2022). Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious Diseases of Poverty, 11*(1), 19. doi:10.1186/s40249-022-00946-4

32. Xu, Y., Zomer, S., & Brereton, R. (2006). Support Vector Machines: A Recent Method for Classification in Chemometrics. *Critical Reviews in Analytical Chemistry, 36*, 177-188. doi:10.1080/10408340600969486

33. Zhang, F., Tong, L., Shi, C., Zuo, R., Wang, L., & Wang, Y. (2024). Deep Learning in Predicting Preterm Birth: A Comparative Study of Machine Learning Algorithms. *Maternal-Fetal Medicine, 6*(3), 141-146. doi:10.1097/fm9.0000000000000236