**Regularized logistic regression models for breast cancer**

**Abstract**

Breast cancer classification is a critical facet of modern oncology, employing advanced statistical learning algorithms to distinguish between benign and malignant cases. Accurate classification is imperative for early detection, treatment planning, and patient outcomes. The importance of effective breast cancer classification lies in its potential to enhance diagnostic precision, guide treatment strategies, and ultimately improve survival rates and quality of life for affected individuals. In this paper we implemented three sophisticated models, such as Lasso, Ridge, and Elastic Net logistic regression which play a pivotal role in achieving high accuracy, precision, and recall. To ensure robust and reliable results, a comprehensive step of preprocessing techniques was implemented, encompassing data cleaning to address null values and duplicate records, data scaling for feature normalization, random over-sampling to tackle class imbalance, and an 80:20 data splitting ratio for training and testing. Additionally, cross-validation was employed to assess model generalization and robustness. The paramount importance of accurately diagnosing cancer types lies in its potential to significantly impact patient outcomes and guide treatment strategies. Lasso logistic regression emerges as the top performer, achieving 97.2% accuracy, balanced precision and recall at 97%, and an F1-score of 97%. Ridge logistic regression follows closely with 95.8% accuracy and balanced precision, recall, and F1-score at 96%. Elastic Net logistic regression, positioned between Lasso and Ridge, attains 96.5% accuracy, with precision and recall both at 96.5%, and an F1-score of 96.2%. These findings underscore the trade-offs and strengths of each algorithm, offering guidance to practitioners for optimal model selection in breast cancer classification. In conclusion, this research provides a comprehensive analysis of statistical learning algorithms for breast cancer classification, offering insights into their performance metrics, trade-offs, and discriminative abilities. These findings empower practitioners to make informed decisions based on specific objectives and requirements in the context of breast cancer diagnosis.