

INTRODUCTION

Machine learning has transformed healthcare predictive analytics, enabling improved disease diagnosis and risk assessment. However, challenges like class imbalance, interpretability, and generalizability remain. Traditional methods often struggle with complex data, while deep learning demands high computational resources. This study focuses on ensemble learning to enhance model accuracy and robustness in healthcare predictions. To address these challenges, we provide a suite of tools for efficient model optimization and selection.

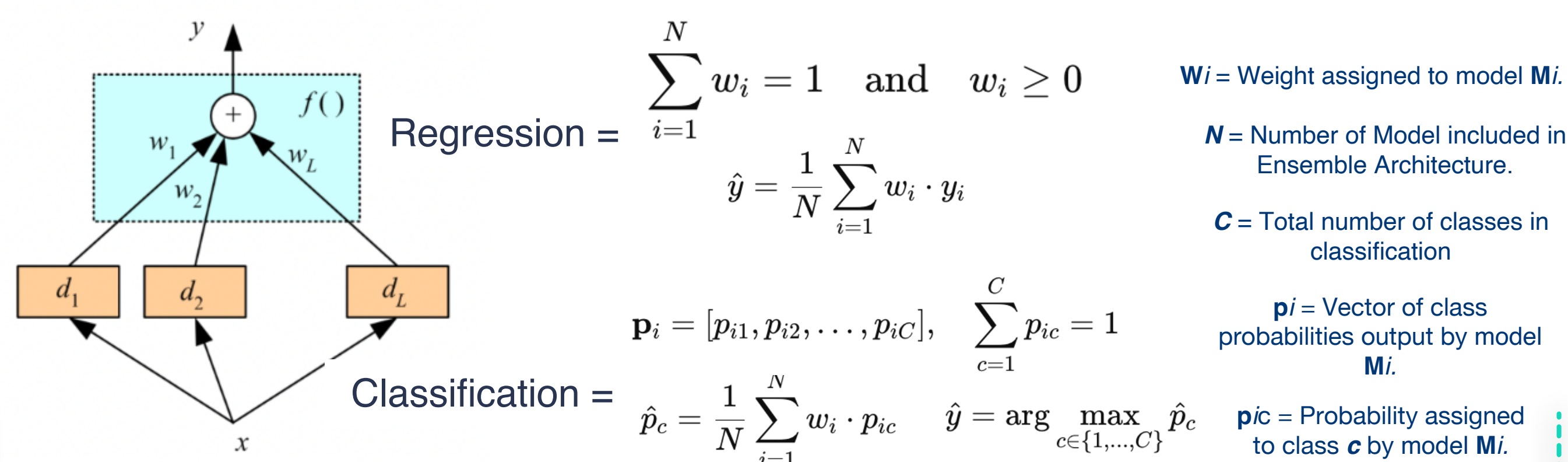
OBJECTIVES

TweenMe Single fine-tunes hyperparameters for individual models, enhancing accuracy in classification and regression. **TweenMe Best** evaluates and ranks multiple algorithms to identify the top-performing model. **TweenMe Ensemble** combines outputs from several models through a stacked ensemble, improving overall prediction accuracy and robustness. Together, these tools ensure reliable, well-tuned models that effectively handle data quality issues in healthcare analytics.

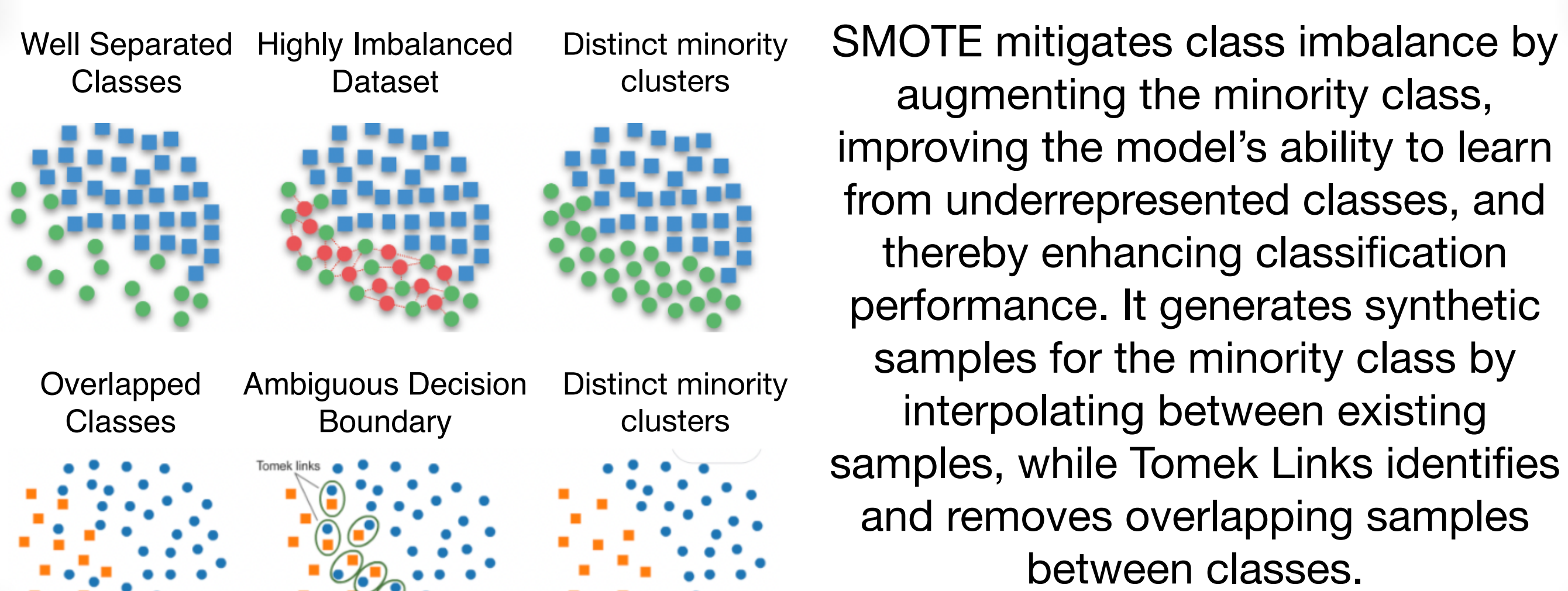
METHODS

Ensemble Architecture

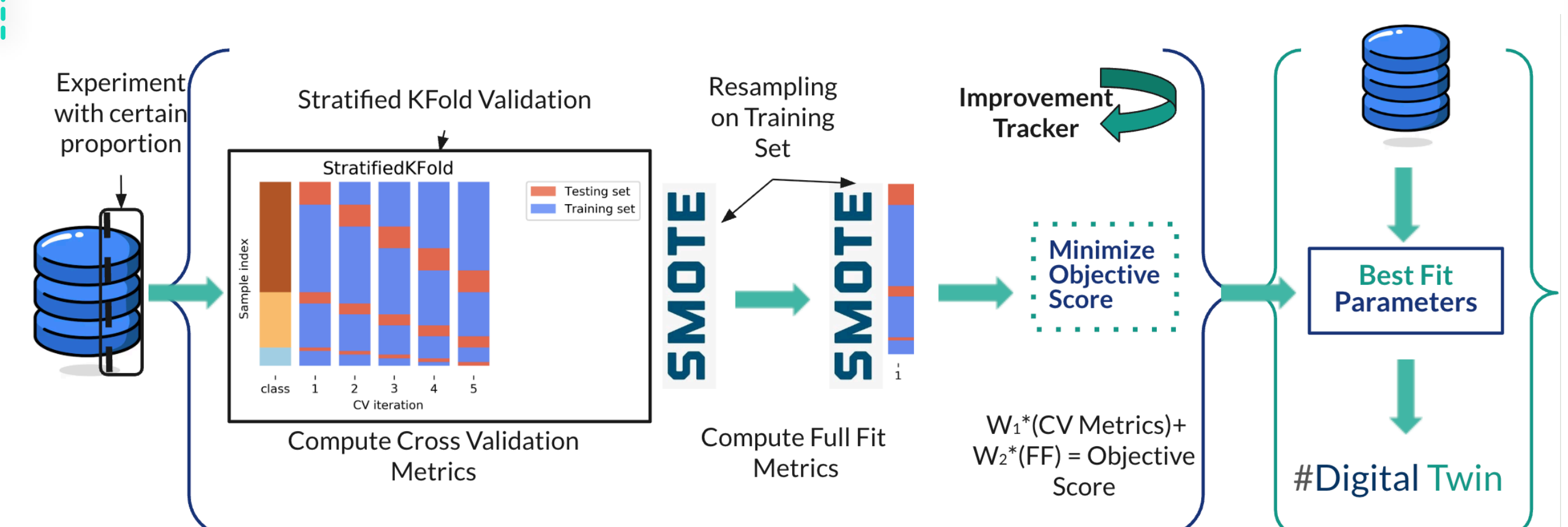
Ensemble learning architectures provide a substantial advantage in disease diagnosis by integrating multiple predictive models to reduce bias and variance. This hybrid approach enhances early and accurate diagnosis, enabling early interventions to mitigate health risks. In this method, models are assigned weights based on their performance, and the final prediction is determined by aggregating the predictions of individual models, influenced by their respective assigned weights. This approach ensures that models with higher accuracy have a greater impact on the outcome, enhancing overall performance.



Resampling Method



Optimization Flow



OptiTween with OPTUNA implements Bayesian optimization to efficiently navigate the hyperparameter space. This method balances exploration (sampling diverse hyperparameter configurations) and exploitation (focusing on promising regions) to minimize a composite objective function—a weighted loss aggregating multiple evaluation metrics.

A core component, the ImprovementTracker, monitors trial performance in real time, enabling adaptive guidance of the optimization process and ensuring convergence toward optimal hyperparameter settings.

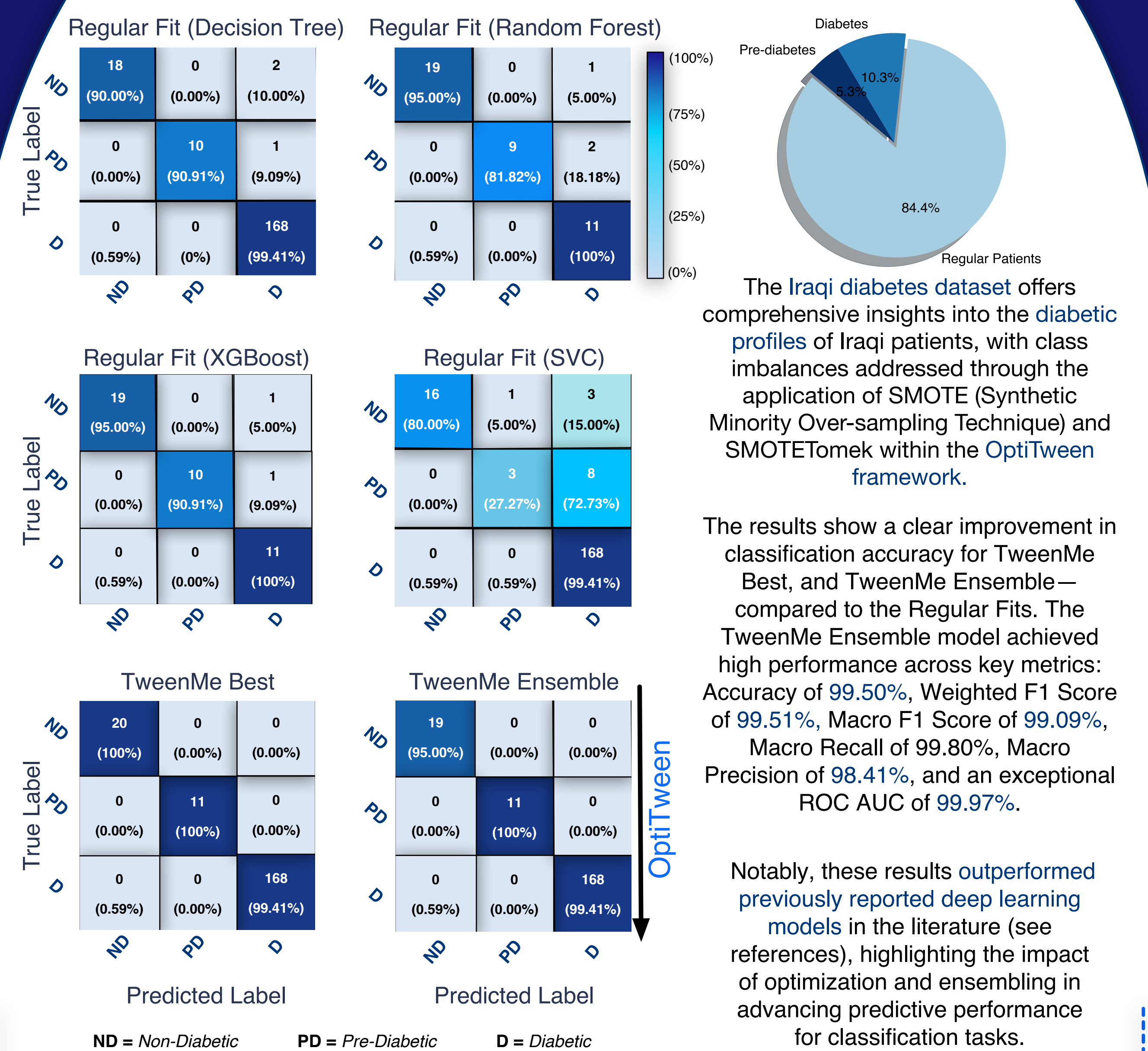
CONCLUSION

Model Performance and Ensemble Advantage: TweenMe Ensemble consistently deliver the most robust performance across tasks, outperforming traditional models through precise optimization and strategic ensembling. While RF and XGBoost are powerful alone, the ensemble approach further reduces overfitting and improves generalization, which is crucial for high-dimensional and/or imbalanced datasets like Parkinson's and Iraqi Diabetes. SVR, RR, and SVC struggle with wide residuals or non-separable classes, further confirming the superiority of optimized, ensemble-based strategies.

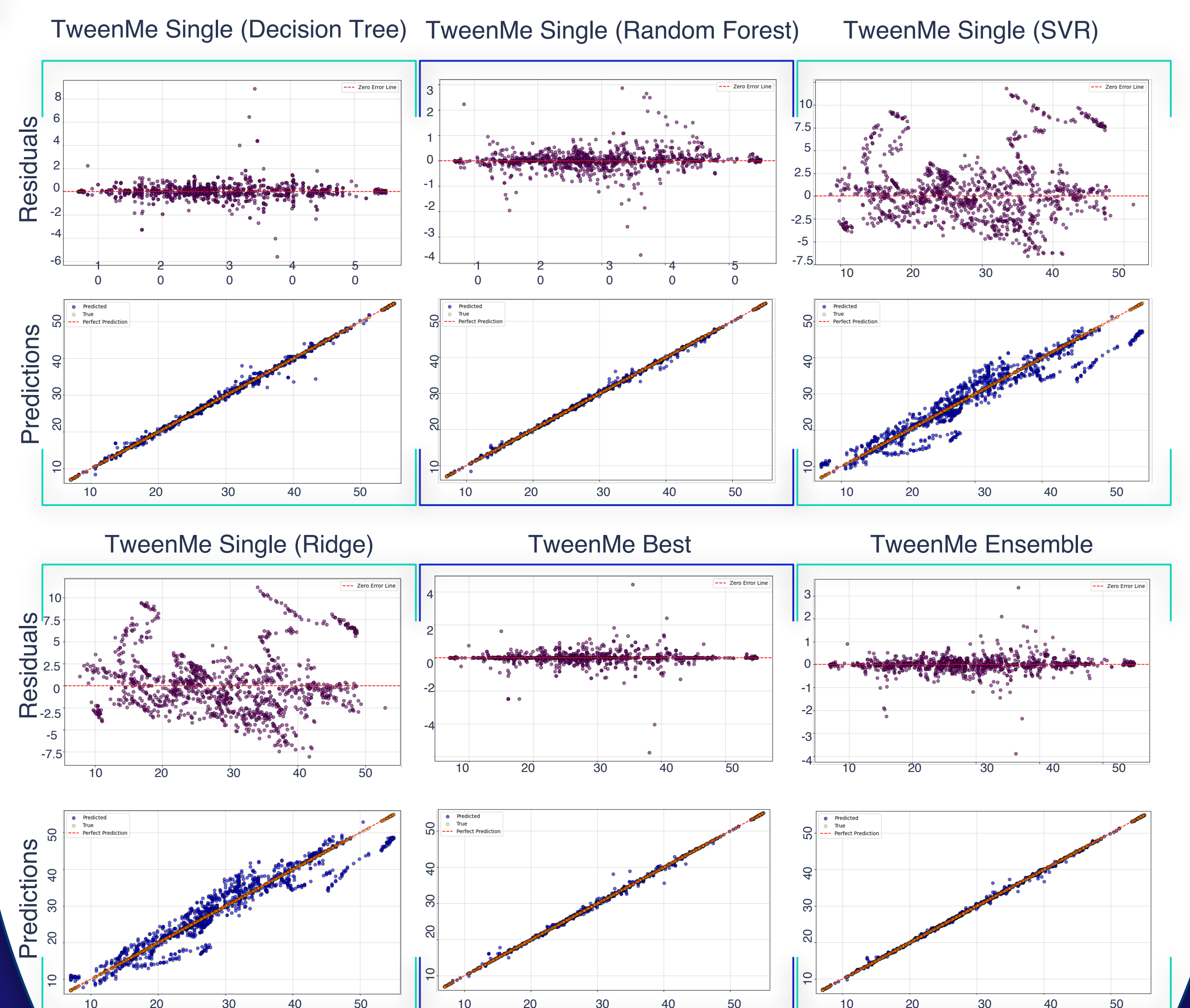
Value of Optimization Framework: These findings highlights the critical role of model selection and fine-tuning in medical machine learning. The TweenMe Digital Twin Bakery provides an integrated solution by systematically refining model performance, thereby improving prediction accuracy across complex clinical applications.

RESULTS

TweenMe ML Classifiers



TweenMe ML Regressors



REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining.

Al-Jumaily, D., Hussain, A., & Khalaf, O. I. (2020). Iraqi Diabetes Dataset [Data set]. PLOS ONE. <https://doi.org/10.1371/journal.pone.0300785>

Fahmy, N. L., Syafudin, M., Alfian, G., Fhee, J. (2019). Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension. IEEE Access, 7, 144777-144789. Article 8854986. <https://doi.org/10.1109/ACCESS.2019.2945129>

García-Ordás, M. T., Benavides, C., Benítez-Andrades, J. A., Albiz-Moreno, H., García-Rodríguez, I. (2021). Diabetes detection using deep learning techniques with oversampling and feature augmentation. Computer methods and programs in biomedicine, 202, 105968. <https://doi.org/10.1016/j.cmpb.2021.105968>

Karami, H., Aharthi, R., Anzi, A.A., Alhebshi, R.M., Eshmaei, A.A., Alsbai, S., Umer, M. (2023). Improving Prediction of Cervical Cancer Using KNN Imputed SMOTE Features and Multi-Model Ensemble Learning Approach. Cancers, 15.

Kavitha, M., Ganeswar, G., Dinesh, R., Rohith Sai, Y. and Sai Suraj, R. (2021) Heart Disease Prediction Using Hybrid Machine Learning Model. 2021 IEEE 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 20-22 January 2021, 1329-1333. <https://doi.org/10.1109/ICICT50816.2021.9358597>

Ijaz, M.F., Alique, M., Son, Y. (2020). Data-Driven Cervical Cancer Prediction Model with Outlier Detection and Over-Sampling Methods. Sensors (Basel, Switzerland), 20.

Watanabe, S. (2023). Tree-structured Parzen estimator: Understanding its algorithm components and their roles for better empirical performance. ArXiv:2304.11127.