ASEC 2025 Conference

The 6th International Electronic Conference on Applied Sciences

09–11 December 2025 | Online

TransQSAR-pf: A Bio-Informed QSAR Framework Using *Plasmodium falciparum* Stress Signatures for Enhanced Antiplasmodial Activity Prediction

Favour O. Igwezeke[1]*, Charles O. Nnadi[2]

[1] Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka 410001, Enugu, Nigeria

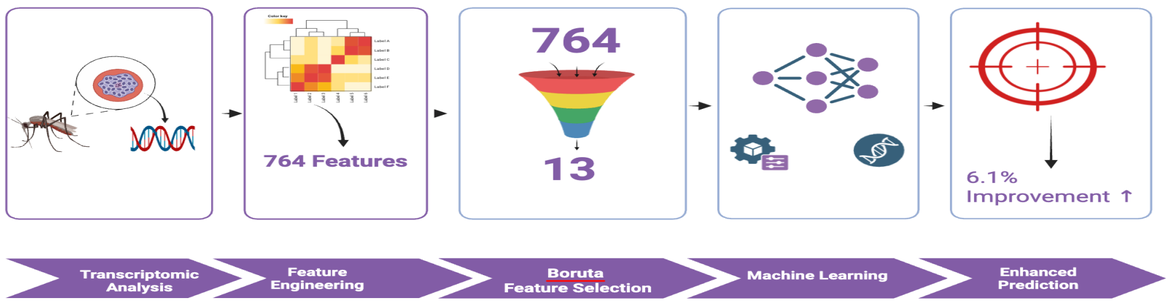
[2] Department of Pharmaceutical and Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka 410001, Enugu, Nigeria

INTRODUCTION & AIM

The Challenge: Traditional QSAR models predict drug activity from molecular structure alone, completely ignoring the biological state of the target organism. This limitation becomes critical for antimalarial drug discovery, where *Plasmodium falciparum*'s stress response to drug treatment plays a pivotal role in determining compound efficacy.

Our Innovation: We developed **TransQSAR-pf**, a framework that integrates parasite transcriptomic stress signatures with classical QSAR descriptors. By studying both the "key" (drug molecule) and the "lock's internal state" (parasite biology), we achieve superior predictive accuracy and uncover novel drug targets hidden in unexplored biology.

METHOD



Data Sources & Integration

Data Sources

- **Transcriptomics:** GSE10022 (24,563 probes, 18 samples, 3 genotypes under chloroquine pressure)
- **Compounds:** 125 triazolopyrimidine derivatives with IC_{50} values + 15 QSAR descriptors

Our 4-Step Integrated Workflow

Step 1: Transcriptomic Analysis

- Differential expression via limma with FDR correction
- GSEA via fgsea (PlasmoDB, 18,000 annotations)
- **Key pathways:** Conserved *Plasmodium* proteins (p=0.005), RNA-binding proteins (p=0.020), PfEMP1 (p=0.036)

Step 2: Feature Engineering

Created 764 transcriptomic features capturing: 600 differential expression signatures, 3 pathway enrichment scores, 100 expression variability metrics, 61 functional group profiles

Step 3: Intelligent Feature Selection

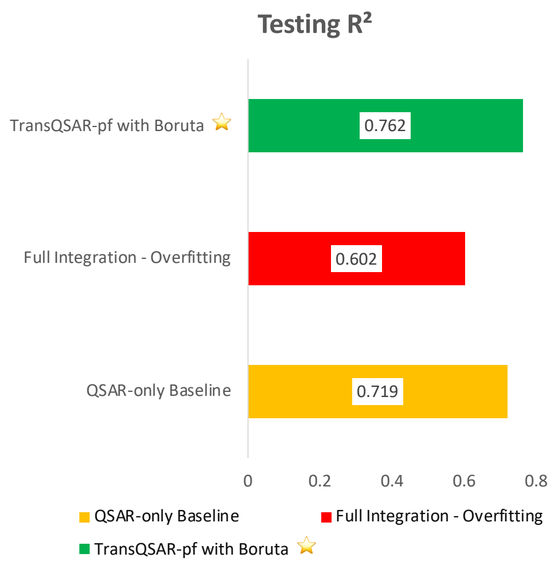
- Applied **Boruta algorithm** (200 iterations, 5-fold CV)
- Reduced 764 → **13 critical biological predictors**
- **98.3% feature reduction** while preserving predictive power

Step 4: Machine Learning

- Trained Random Forest, SVM (RBF kernel), and Elastic Net models
- Hyperparameter optimization via grid search
- Rigorous 5-fold cross-validation to prevent overfitting

RESULTS & DISCUSSION

Finding 1: Strategic Feature Selection Drives Success

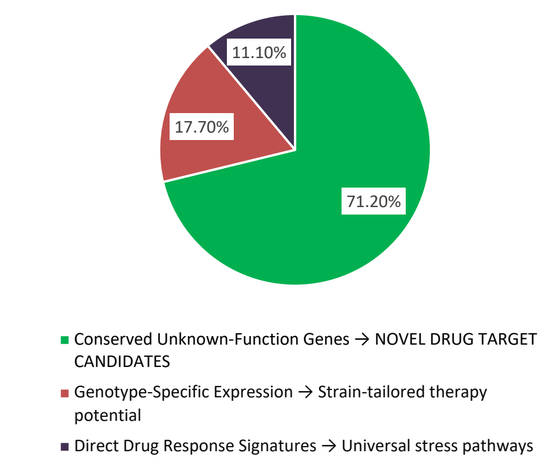


TransQSAR-pf achieved R^2=0.762 (RMSE=0.470)—a 6.1% improvement over QSAR-only baseline (R^2=0.719).

Adding all 764 features without selection caused overfitting (R^2=0.602), proving **biology-guided selection, not data dumping, is essential**.

Model robustness: Training R^2=0.899 with minimal train-test gap confirms excellent generalization without overfitting.

Finding 2: Unexplored Biology Drives Predictive Power



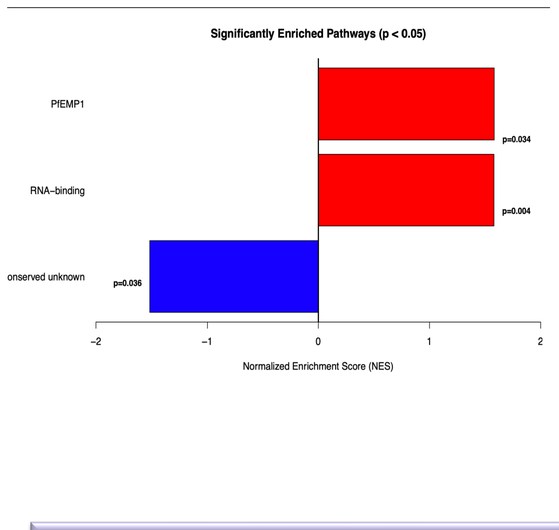
71.2% of predictive power →

Conserved unknown-function genes
17.7% → Genotype-specific expression
11.1% → Direct drug response

These anonymous genes whose roles remain uncharacterized emerged as the strongest predictors through unbiased ML. They represent high-priority drug target candidates: conserved, essential, and resistance-resistant.

Genotype patterns (17.7%) enable strain-specific therapy. Universal signatures (11.1%) capture core stress pathways.

Finding 3: Pathway Enrichment Reveals Mechanism



GSEA identified enriched pathways explaining efficacy:

Conserved *Plasmodium* Proteins (p=0.005)
→ Validates unknown genes as drug discovery priorities

RNA-Binding Proteins (p=0.020)
→ Post-transcriptional regulation = potential targets

PfEMP1 Virulence Factors (p=0.036)
→ Links virulence to drug susceptibility

CONCLUSION & IMPACT

TransQSAR-pf demonstrates that strategic integration of pathogen biology significantly enhances antimalarial drug prediction, achieving:

✅ **6.1% improvement** in predictive accuracy (R^2=0.762)
✅ **98.3% feature reduction** via biology-guided Boruta selection
✅ **Novel target identification:** 71.2% of predictive power from unknown-function genes
✅ **Mechanistic insights:** Links conserved stress pathways to compound efficacy

This framework represents a paradigm shift from structure-only to biology-informed drug discovery.