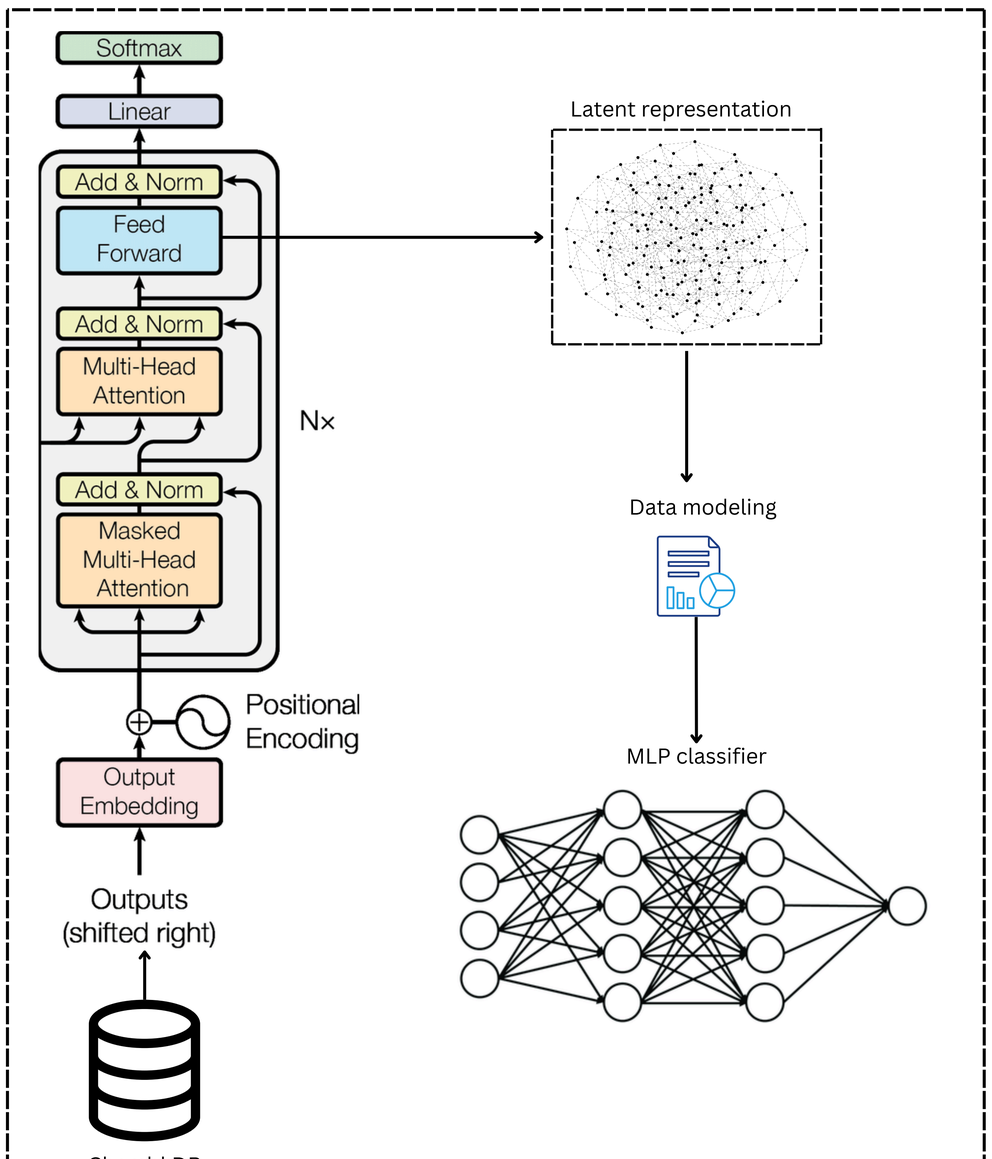
Latent Representations from Large Language Models for Ligand-Kinase Interaction Prediction in Drug Discovery

Support:

Leon S. C. Costa, Matheus M. P. da Silva, Fábio L. Custódio, Marisa F. Nicolás, Laurente E. Dardenne
Laboratório Nacional de Computação Científica- LNCC/MCTI - Brasil

Protein kinases are pivotal drug targets, yet screening the vast combinatorial landscape of >500 human enzymes against millions of ligands is experimentally prohibitive. We introduce a text-centric alternative that sidesteps the structural prerequisites of graph-based pipelines: ligands are expressed as SMILES strings embedded by IBM's SMI-TED, and kinases as amino-acid sequences embedded by Meta AI's ESM-2. Concatenating these n-dimensional vectors, a compact multilayer perceptron assigns interaction probabilities. By dispensing with 3-D structures and docking poses—requirements that limit current graph neural-network approaches [1], our method relies solely on readily available sequence and SMILES data, expanding target coverage and simplifying data preparation while preserving state-of-the-art accuracy. The subsequent figure 1 displays the distribution of training and validation metrics across epochs, visualized through violin plots. These plots compare the evolution of ten key performance indicators—loss, accuracy, precision, recall, F1, ROC_AUC, MCC, PR_AUC, Brier Score, and specificity—between the training and validation sets. In general, the training metrics show slightly higher values and greater variability compared to validation, suggesting a modest degree of overfitting but no severe degradation in generalization.

Methodology



Results

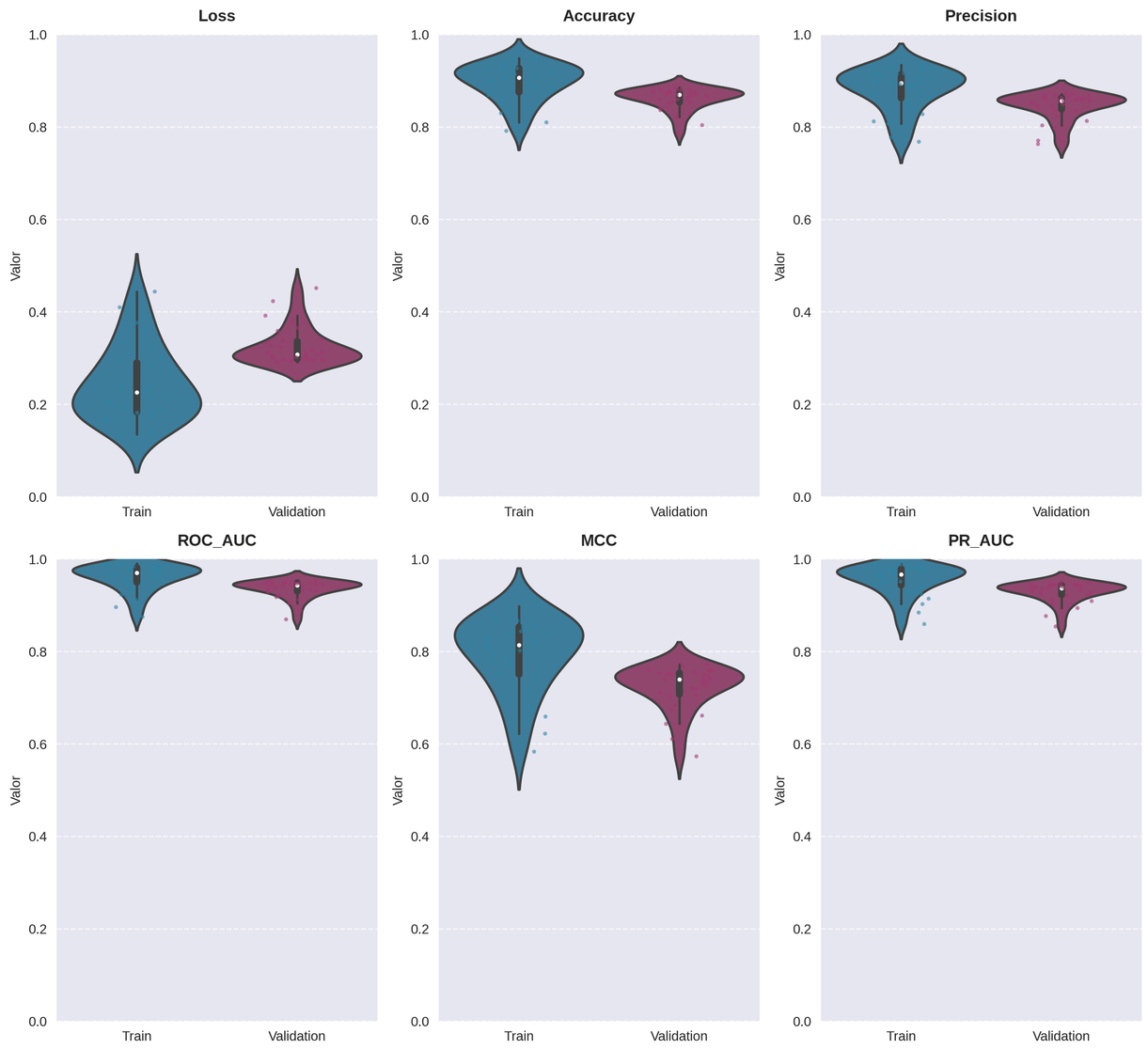


Figure 1 - Distribution of training and validation metrics by epoch. Loss decreases steadily during training and remains lower in validation, indicating stable learning. Accuracy, precision, recall, and F1 stabilize around 0.85–0.90 in both sets, demonstrating consistent performance. The ROC_AUC close to 0.9, reflecting strong predictive discrimination. MCC and Brier Score further support reliable classification and well-calibrated probability estimates. Overall, the model converges smoothly, with minimal divergence between train and validation metrics, confirming robust training and good generalization capability

The following figure 2 presents a comprehensive evaluation of the model's performance on the test set, using multiple diagnostic metrics. The confusion matrix reveals a strong classification performance, with 18,750 true negatives (46.0%) and 17,224 true positives (42.3%), while false positives and false negatives account for 6.7% and 5.0%, respectively. The ROC curve exhibits high discriminative power, with an AUC of 0.9516, indicating excellent separation between classes. Similarly, the precision-recall curve achieves an AUC of 0.9467, surpassing the baseline (0.883), which highlights robust performance even under imbalanced conditions. The prediction distribution shows clear separation between negative and positive class probabilities, with the decision threshold at 0.5 well-positioned between the two peaks. Key performance metrics confirm strong results: accuracy (0.883), precision (0.864), recall (0.893), F1-score (0.878), and ROC-AUC (0.952). The calibration curve indicates that the model is reasonably well-calibrated, particularly at higher predicted probabilities, though some deviation from perfect calibration is observed in intermediate ranges.

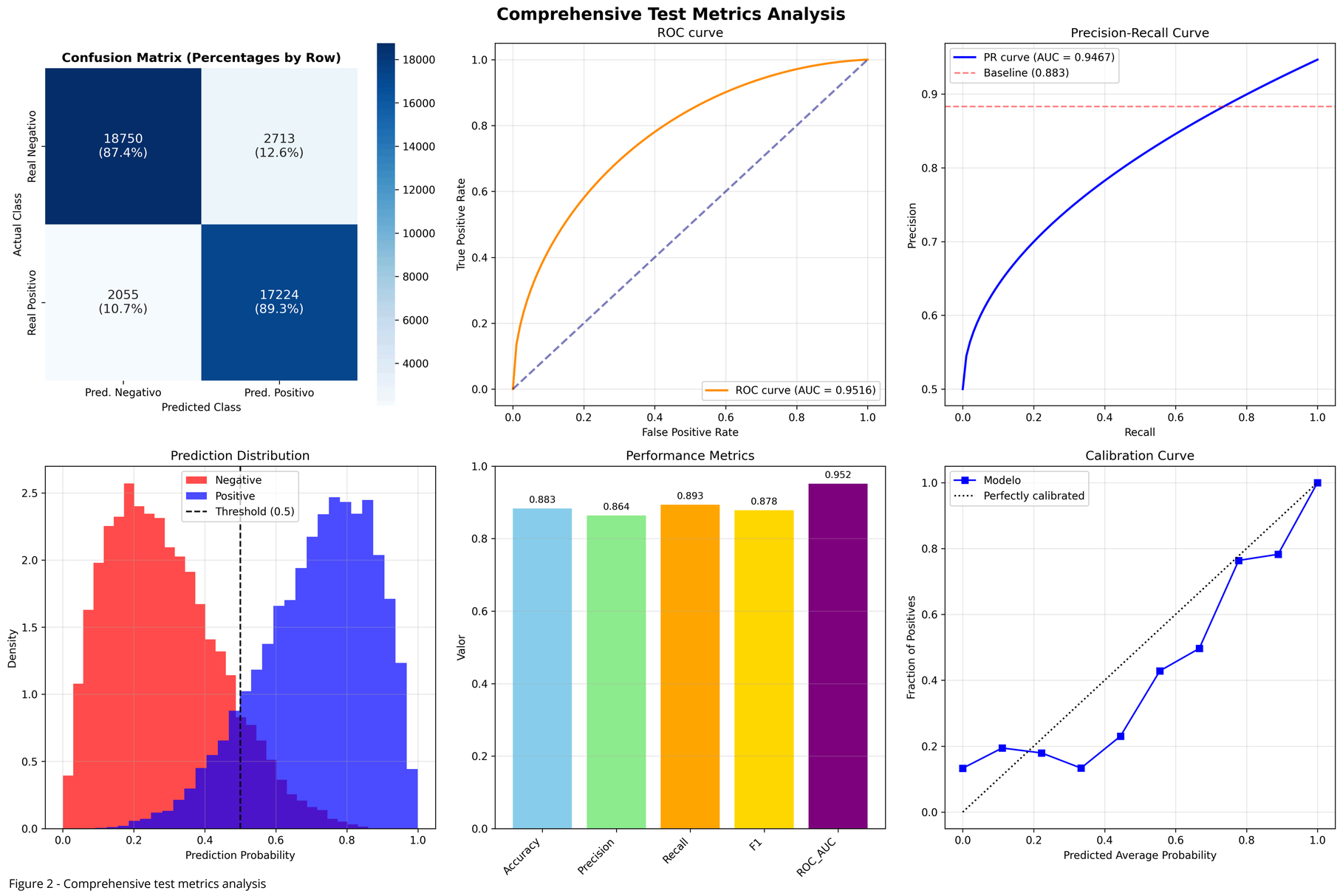


Figure 2 - Comprehensive test metrics analysis

Conclusion

Combining SMI-TED and ESM-2 embeddings in a compact multilayer perceptron yields a fully text-based screen that classifies kinase–ligand pairs as active or inactive with 86 % accuracy, ROC-AUC above 0.90 and MCC ≈ 0.73 on 42,717 unseen test examples, all without relying on 3-D structures or docking; this scalable approach therefore accelerates early-stage kinase drug discovery by rapidly screening vast chemical libraries against multiple potential targets , with minimal data preparation, to identify both primary targets and off-target interactions.

Reference:
[1] Ashish Vaswani, et al. (2023). NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. dl.acm.org/doi/10.5555/3295222.3295349.