Vol. 3, No. 2, 2025
DOI: 10.61975/gjset.v3i2.98

ORIGINAL ARTICLE

Fine-Tuning ChemBERTa for Predicting Activity of AXL Kinase Inhibitors in Oncogenic Target Modeling

Teuku Rizky Noviandy[1], Ghazi Mauer Idroes[2], Mohsina Patwekar[3] and Rinaldi Idroes[4,*]

[1]Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; [2]Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; [3]Department of Pharmacology, Luqman College of Pharmacy, Karnataka 585102, India; [4]School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received
25 July 2025

Accepted
29 September 2025

Available Online
11 October 2025

Keywords

ChemBERTa
QSAR modeling
AXL kinase
Fine-tuning
Drug discovery

Abstract

The development of selective kinase inhibitors remains a key objective in cancer drug discovery, where predictive computational models can significantly accelerate the identification of leads. In this study, we investigate the fine-tuning strategies of the transformer-based ChemBERTa model for quantitative structure–activity relationship (QSAR) modeling of AXL receptor tyrosine kinase inhibitors, an important therapeutic target implicated in tumor progression and metastasis. A dataset of AXL inhibitors was curated from the ChEMBL database. Three fine-tuning configurations, namely baseline, full fine-tune, and aggressive, were implemented to examine the influence of learning rate, weight decay, and the number of frozen transformer layers on model performance. Models were evaluated using accuracy, precision, recall, F1-score, and calibration metrics. Results showed that both the full fine-tune and aggressive configurations outperformed the baseline model, achieving higher precision and F1-scores while maintaining robust recall. The aggressive configuration achieved the most balanced performance, with improved calibration and the lowest expected calibration error, indicating reliable probabilistic predictions. Overall, this study highlights that controlled fine-tuning of ChemBERTa significantly enhances predictive performance and confidence estimation in QSAR modeling, offering valuable insights for optimizing transformer-based chemical language models in kinase-targeted drug discovery.

Introduction

The discovery and improvement of small-molecule inhibitors are key to modern cancer treatments [1,2]. Targeted therapies, which block specific kinases involved in tumor growth and spread, have become increasingly important because they are often more effective and less toxic than traditional chemotherapy [3,4]. Quantitative structure–activity relationship (QSAR) modeling has been a core method in computational drug discovery, enabling the prediction of biological activity from chemical structure and facilitating the virtual screening of large compound libraries [5–7]. However, traditional QSAR methods depend on manually designed molecular descriptors, which may not capture the complex, nonlinear patterns present in chemical data [8].

Recent progress in natural language processing (NLP) and transformer-based models has revolutionized molecular representation learning through chemical language models, such as ChemBERTa [9–11]. By treating SMILES strings as molecular “sentences,” ChemBERTa uses contextual embeddings pre-trained on large chemical datasets to capture detailed substructural and functional group patterns [12]. These deep learning–based representations have shown strong generalization and predictive performance in QSAR and molecular property prediction

tasks [13]. However, the best fine-tuning strategies for adapting pre-trained molecular transformers to specific biochemical targets, such as kinases, remain poorly understood.

The AXL receptor tyrosine kinase is a critical regulator of tumor progression, metastasis, and drug resistance in multiple cancer types [14–16]. Accurate prediction of AXL inhibitor activity is therefore essential for identifying novel therapeutic candidates. However, fine-tuning transformer-based models, such as ChemBERTa, presents several challenges: determining the appropriate training depth, managing overfitting, and balancing the benefits of transfer learning and task-specific adaptation [17]. Excessive freezing of model layers may limit representational flexibility, while overly aggressive fine-tuning can lead to instability and overfitting, particularly under data-limited conditions [18].

This study investigates effective fine-tuning strategies for ChemBERTa in the context of QSAR modeling of AXL kinase inhibitors, a class of small molecules with therapeutic potential in oncology. The goal is to understand how different approaches to adapting a pre-trained molecular transformer influence model performance. Three main configurations are evaluated. The baseline configuration employs partial fine-tuning, where only the final layers of the transformer are updated, enabling the model to adapt to target-specific features while maintaining general chemical representations. The full fine-tuning configuration involves unfreezing all layers and updating the entire network, aiming for maximum task-specific adaptation but with a higher risk of overfitting and training instability. The aggressive configuration explores a higher learning rate combined with moderate layer freezing to test whether rapid adaptation can yield faster convergence without sacrificing generalization.

This study aims to systematically evaluate the impact of various fine-tuning strategies on the performance of the ChemBERTa transformer model in predicting the bioactivity of AXL kinase inhibitors relevant to cancer therapy. By comparing baseline, full fine-tune, and aggressive configurations that differ in learning rate, regularization, and layer freezing depth, the research seeks to identify the best balance between knowledge transfer and task-specific adaptation in QSAR modeling. The study offers empirical insights into how the intensity of fine-tuning influences predictive accuracy, calibration reliability, and model interpretability. These results help clarify ChemBERTa's adaptability for kinase-focused drug discovery and offer practical guidance for building robust, data-efficient QSAR workflows in medicinal chemistry.

Materials and Methods

Experimental Workflow

The complete workflow for fine-tuning the ChemBERTa model on AXL kinase inhibitor classification is illustrated in Figure 1. The process begins with dataset preparation and preprocessing, followed by model architecture configuration and experimental setup. The workflow proceeds through model training and concludes with evaluation using standard performance metrics and visualization techniques.

Dataset Preparation

The dataset was obtained from the ChEMBL database for the Homo sapiens Tyrosine-protein kinase receptor UFO (AXL; Target ID: CHEMBL4895) [19]. Molecular structures were represented as Simplified Molecular Input Line Entry System (SMILES) strings with corresponding IC_{50} activity values. Compounds with $IC_{50} < 1000$ nM were labeled as active, while those with $IC_{50} \geq 1000$ nM were labeled as inactive [20]. All SMILES were canonicalized using RDKit to ensure structural consistency and remove duplicates [21]. The resulting dataset was divided into training (80%) and testing (20%) subsets using a stratified split to preserve class balance [22]. The training set contained 462 active and 293 inactive compounds, while the testing set included 116 active and 73 inactive compounds.

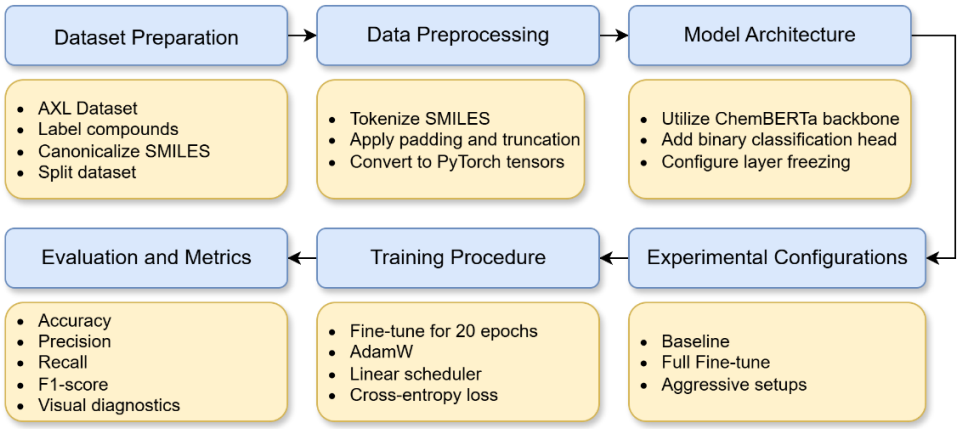

Figure 1. Overview of the ChemBERTa fine-tuning workflow for AXL kinase inhibitor prediction.

Data Preprocessing

Each canonical SMILES string was tokenized using the ChemBERTa tokenizer from the seyonec/ChemBERTa-zinc-base-v1 model. Tokenization was performed with automatic truncation and padding to a uniform sequence length, and tensors were generated for input IDs, attention masks, and labels. The tokenized datasets were subsequently converted into PyTorch TensorDataset objects and loaded into DataLoaders for efficient batch processing during training and validation.

Model Architecture

The ChemBERTa (RoBERTa-based) architecture pre-trained on large chemical corpora was used as the backbone for fine-tuning [23]. A classification head with two output neurons was appended for binary activity prediction (active/inactive). To evaluate the effect of transfer learning depth, different configurations froze varying numbers of lower transformer encoder layers. Layers were frozen by disabling gradient computation for selected encoder blocks using parameter name matching [24]. This allowed assessment of representational transfer versus full retraining.

Experimental Configurations

To systematically evaluate how fine-tuning intensity and regularization strategies affect predictive performance, three ChemBERTa configurations, namely baseline, full fine-tune, and aggressive, were implemented. Each configuration varied in the learning rate, weight decay, number of frozen transformer layers, and batch size, as summarized in Table 1.

Table 1. Summary of fine-tuning configurations used for ChemBERTa on AXL kinase inhibitor classification.

Configuration	Learning Rate	Weight Decay	Frozen Layers	Batch Size	Description
Baseline	3×10^{-5}	0.01	8	16	Partial fine-tuning of upper transformer layers.
Full fine-tune	2×10^{-5}	0.01	0	16	All transformer layers are unfrozen, and complete fine-tuning is performed.
Aggressive	5×10^{-5}	0	6	32	Higher learning rate, no weight decay, moderate freezing.

The baseline setup served as a partial fine-tuning control, where the lower eight transformer encoder layers were frozen to retain pre-trained chemical representations while allowing adaptation in the upper layers. This approach balances the reuse of learned molecular embeddings with limited task-specific flexibility. A moderate learning rate of 3×10^{-5} was chosen to ensure stable optimization given the reduced number of trainable parameters, while an L2 weight decay of 0.01 helped prevent overfitting by constraining parameter growth. This configuration provides a controlled benchmark for assessing the benefits of deeper fine-tuning.

The full fine-tune configuration unfroze all layers, allowing complete backpropagation through the entire ChemBERTa encoder. This enables the model to fully adapt its representational space to the downstream task, capturing subtle chemical relationships that may not have been emphasized during pre-training. A slightly reduced learning rate of 2×10^{-5} was used to stabilize training and mitigate the risk of catastrophic forgetting, given the larger number of trainable parameters. Weight decay of 0.01 and a batch size of 16 were kept consistent with the baseline to isolate the effect of unfreezing all layers and ensure a fair comparison of fine-tuning depth.

The aggressive configuration explored a higher learning rate of 5×10^{-5} combined with no weight decay to encourage faster convergence and more dynamic parameter updates. This setup was designed to test the limits of adaptation speed and learning flexibility under minimal regularization. To maintain some representational stability and avoid complete drift from pre-trained knowledge, six lower layers were frozen while the remaining layers adapted to task-specific chemical patterns. A larger batch size of 32 was employed to reduce gradient noise and stabilize updates under the higher learning rate, offering a practical trade-off between exploration and stability.

Training Procedure

All ChemBERTa models were fine-tuned using PyTorch and the Hugging Face Transformers framework under consistent experimental settings. Each configuration was trained for 20 epochs with the AdamW optimizer and a linear learning rate scheduler without warm-up. The cross-entropy loss function was used for binary classification of active versus inactive compounds. Mini-batch training was performed using the batch sizes specified for each configuration, and model parameters were updated after each batch via backpropagation. Depending on the setup, a defined number of lower transformer layers were frozen to preserve pre-trained chemical representations, while upper layers were fine-tuned to learn task-specific molecular features.

Evaluation and Metrics

Model performance was quantitatively assessed using four standard classification metrics, namely accuracy, precision, recall, and F1-score, as defined in Equations (1–4), respectively [25,26]. These metrics were computed on the independent test set using the scikit-learn library to ensure standardized and reproducible evaluation. Accuracy quantified the overall proportion of correct predictions, while precision and recall evaluated the model's ability to correctly identify active compounds. The F1-score, as the harmonic mean of precision and recall, provided a balanced measure of predictive performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

To further analyze classification behavior, several visual diagnostics were performed. Receiver Operator Characteristic (ROC) curve and precision–recall curve comparisons were generated to assess sensitivity–specificity trade-offs among configurations. Confusion matrices provided insights into prediction accuracy per class, while calibration (reliability) plots examined the agreement between predicted probabilities and observed outcomes. In addition, a molecular activity distribution plot based on kernel density estimation (KDE) visualized the distribution of predicted activity scores that separates active and inactive molecules. Collectively, these analyses enabled a comprehensive evaluation of each model's discriminative power, calibration, and interpretability [27].

Results and Discussion

The predictive outcomes of the three ChemBERTa fine-tuning configurations, baseline, full fine-tune, and aggressive, were evaluated using accuracy, precision, recall, and F1-score. The comparative results are summarized in Table 2, highlighting the influence of fine-tuning depth and regularization strategy on classification performance.

Table 2. Performance of ChemBERTa configurations on AXL kinase inhibitor classification.

Configuration	Accuracy	Precision	Recall	F1-score
Baseline	75.66	76.12	87.93	81.60
Full fine-tune	78.31	79.53	87.07	83.13
Aggressive	78.84	80.16	87.07	83.47

Accuracy represents the overall proportion of correctly classified compounds and serves as a general indicator of performance. Both the full fine-tune and aggressive models achieved higher accuracy than the baseline, suggesting that increased trainable depth allowed the model to better generalize across the chemical space of AXL inhibitors. Precision measures the proportion of predicted active compounds that were truly active, and its improvement in the Fine-tune and aggressive models (79.53% and 80.16%, respectively) reflects a reduction in false positives, which is important for minimizing experimental validation costs in drug discovery.

Recall quantifies the proportion of actual active compounds correctly identified. All configurations exhibited high recall values (above 87%), indicating that ChemBERTa effectively retained its sensitivity toward active molecules even when layers were frozen or hyperparameters adjusted. The F1-score integrates both precision and recall into a balanced measure of performance. The increase in F1-score from 81.60 in the baseline to 83.47 in the aggressive configuration demonstrates that deeper and less constrained fine-tuning improved the balance between identifying actives and avoiding misclassification of inactives. Overall, the results indicate that while recall remained consistently strong across all models, precision and F1-score benefited the most from broader fine-tuning, emphasizing the importance of adaptive learning in QSAR modeling with ChemBERTa.

The training progression of the three ChemBERTa configurations is illustrated in Figure 2, showing (a) the loss curve and (b) the accuracy curve across 20 epochs. These plots provide insights into model convergence behavior and stability during fine-tuning.

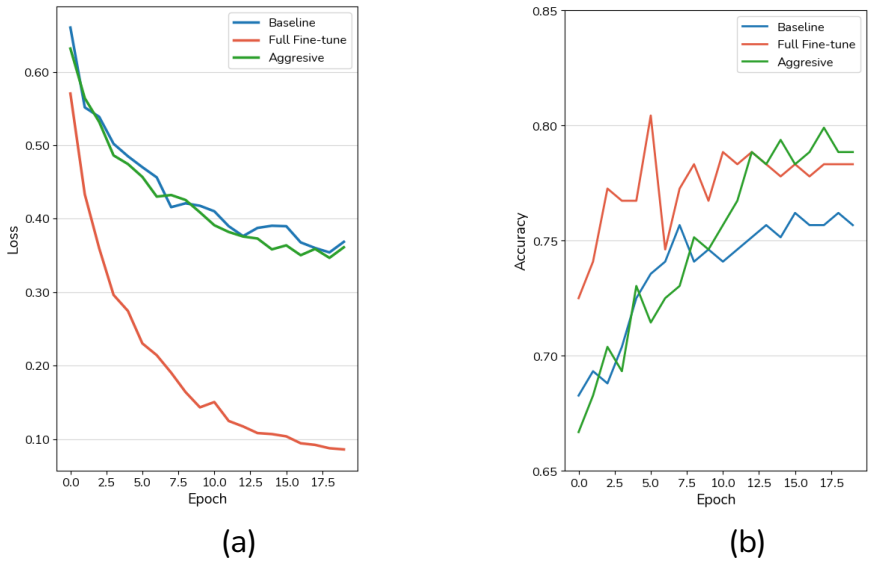

Figure 2. Training curves for ChemBERTa fine-tuning configurations: (a) loss per epoch and (b) accuracy per epoch.

In Figure 2a, the full fine-tune configuration demonstrated the fastest and most consistent reduction in loss, reaching a near-minimum by epoch 15, which indicates stable optimization when all layers were trainable. The aggressive configuration showed rapid early loss reduction but exhibited mild oscillations, suggesting that the higher learning rate introduced stronger gradient updates and minor fluctuations around the optimum. The baseline configuration maintained higher loss values throughout training, reflecting limited adaptability due to partial layer freezing.

Figure 2b shows that validation accuracy improved progressively for all configurations, with both full fine-tune and aggressive models achieving over 78% accuracy by the end of training. The baseline model's accuracy plateaued earlier, consistent with its reduced training flexibility. These patterns confirm that deeper fine-tuning and adjusted optimization parameters accelerated convergence and improved generalization for the AXL kinase QSAR task.

The classification performance of each ChemBERTa configuration was further examined through confusion matrices, as shown in Figure 3. These matrices illustrate the distribution of true and predicted labels for active and inactive compounds, allowing qualitative assessment of false positives and false negatives.

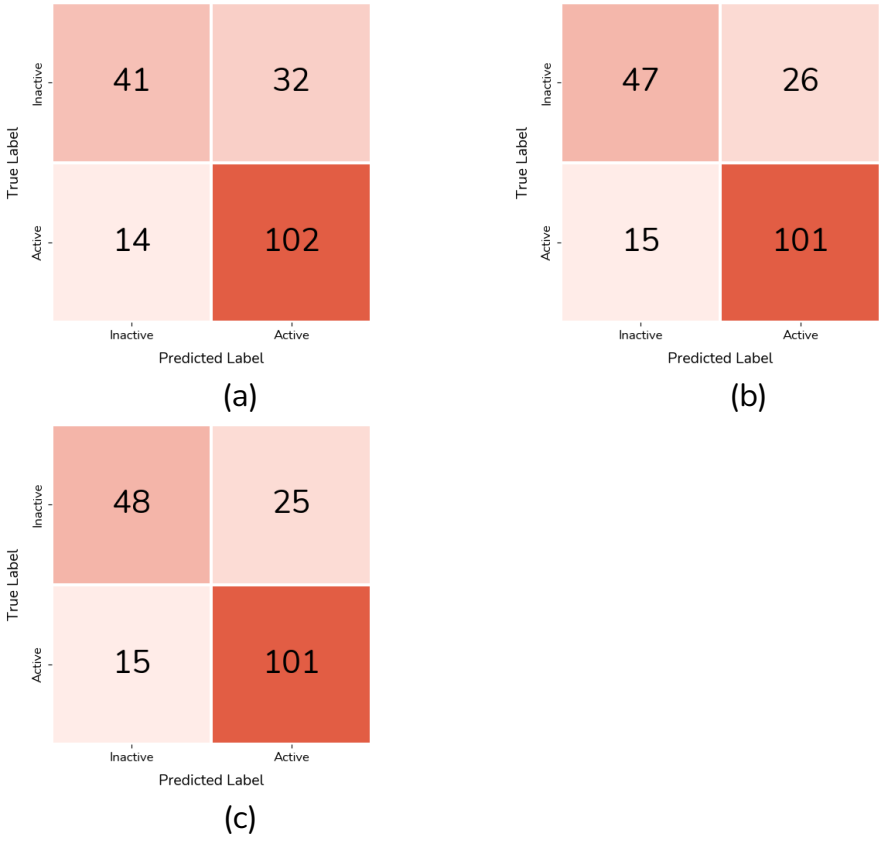

Figure 3. Confusion matrices for ChemBERTa configurations: (a) baseline, (b) full fine-tune, and (c) aggressive.

In Figure 3a, the baseline configuration correctly classified 102 active and 41 inactive compounds but misclassified 32 inactive molecules as active, indicating a tendency toward over-prediction of activity. The full fine-tune model (Figure 3b) improved both sensitivity and specificity, reducing false positives (26) while maintaining high true positive counts (101). This demonstrates that full-layer fine-tuning enhanced discrimination between active and inactive compounds.

The aggressive configuration (Figure 3c) achieved the most balanced confusion matrix, with 101 true actives and 48 true inactives, and the fewest misclassifications overall. This reflects a strong trade-off between precision and recall, consistent with its superior F1-score observed in Table 2. The relatively stable false negative count across all models suggests that ChemBERTa's underlying chemical representation effectively captures active compound features, while fine-tuning primarily improves precision by reducing overprediction of activity.

To evaluate the overall discriminative capability of each ChemBERTa configuration, ROC curves were generated, as shown in Figure 4. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) across all classification thresholds, providing a threshold-independent assessment of model performance.

All three configurations exhibited strong classification performance, with areas under the curve (AUC) ranging from 0.830 to 0.835. The full fine-tune model achieved the highest AUC value (0.835), indicating slightly superior discriminative ability compared with the baseline (0.830) and aggressive (0.833) models. The close proximity of these AUC values demonstrates that all fine-tuning strategies preserved ChemBERTa's inherent ability to separate active and inactive compounds effectively.

Notably, the steeper initial rise in the ROC curve for the full fine-tune model suggests improved early-stage sensitivity, where small increases in false positives yield larger gains in true positives. This reflects more effective internal feature adaptation, especially for borderline compounds near the activity threshold. Overall, the ROC analysis confirms that deeper fine-tuning slightly enhances classification robustness without sacrificing stability or calibration across thresholds.

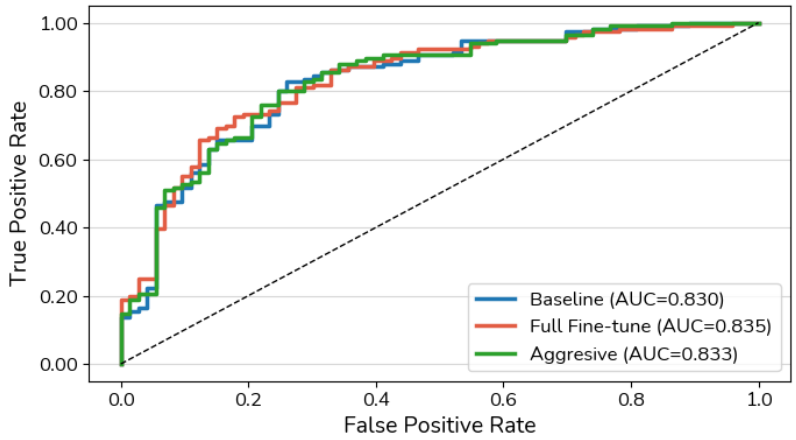

Figure 4. ROC curve comparison for ChemBERTa configurations: baseline, full fine-tune, and aggressive.

To further assess model performance on imbalanced classification and active compound detection, precision–recall curves were generated for all ChemBERTa configurations, as shown in Figure 5. The precision–recall curve illustrates the relationship between precision (positive predictive value) and recall (sensitivity) across various probability thresholds, offering a more informative evaluation than ROC analysis when the active class is less prevalent.

All three models displayed strong precision–recall trade-offs, with area under the precision–recall curve (AUC-PR) values between 0.874 and 0.884, confirming stable and reliable detection of active molecules. The full fine-tune configuration achieved the highest AUC-PR (0.884), indicating the most effective balance between minimizing false positives and maintaining high recall. The aggressive configuration achieved a comparable result (0.878), suggesting that higher learning rates did not compromise performance stability.

The consistently high AUC-PR values across all configurations demonstrate ChemBERTa's strong discriminative capacity for molecular bioactivity prediction. The minor yet consistent improvement from the baseline to the full fine-tune setup reflects that deeper fine-tuning enhances the model's ability to distinguish actives, particularly in decision regions near the activity threshold where predictive uncertainty is greatest.

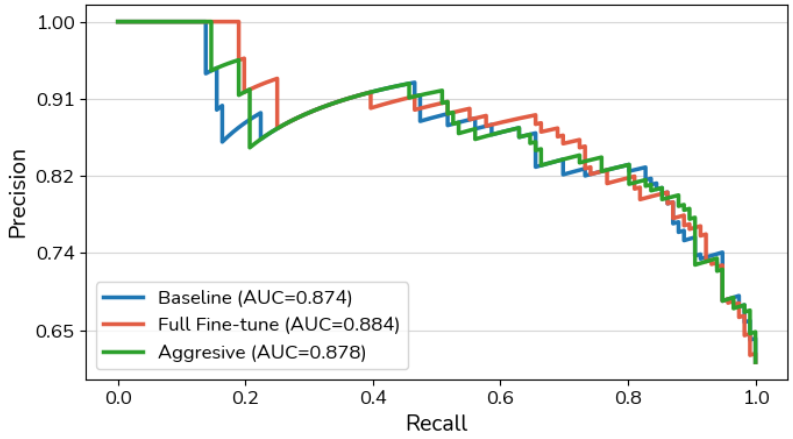

Figure 5. Precision–recall curve comparison for ChemBERTa configurations: baseline, full fine-tune, and aggressive.

To assess the reliability of predicted probabilities and determine how well the models' confidence values align with observed outcomes, calibration plots were generated as shown in Figure 6. These plots compare the predicted probability of activity with the true fraction of active compounds, providing a measure of how well the model's output probabilities reflect real-world likelihoods.

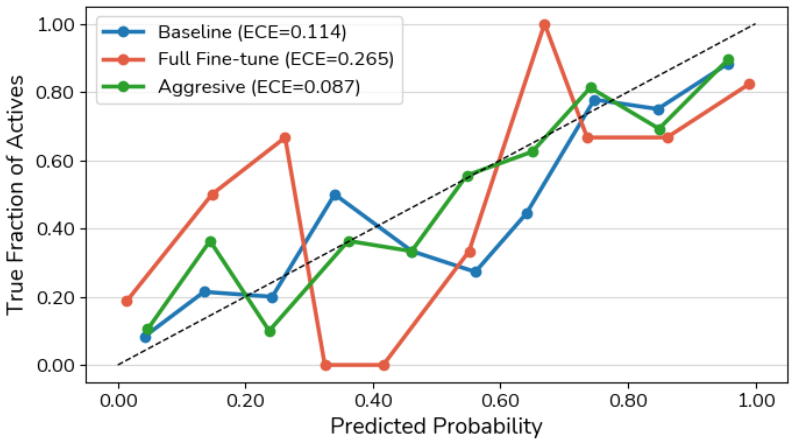

Figure 6. Calibration plots for ChemBERTa configurations: baseline, full fine-tune, and aggressive.

An ideally calibrated model would follow the diagonal reference line, where predicted probabilities perfectly match observed frequencies. The Expected Calibration Error (ECE) quantifies the deviation from this ideal relationship. As shown in Figure 6, the aggressive configuration achieved the lowest ECE (0.087), indicating the best calibration and most reliable probability estimates. The baseline model exhibited moderate calibration (ECE = 0.114), while the full fine-tune configuration showed overconfidence (ECE = 0.265), where predicted probabilities tended to overestimate true activity likelihoods.

These results suggest that while full-layer fine-tuning enhanced classification accuracy, it may have slightly reduced probabilistic reliability due to stronger adaptation and potential overfitting. In contrast, moderate regularization in the aggressive configuration preserved both

performance and well-calibrated uncertainty estimates, making it the most dependable for downstream decision-making in virtual screening applications.

To visualize how each ChemBERTa configuration distributed predicted probabilities across the active–inactive classification space, molecular activity distribution plots were generated using KDE, as shown in Figure 7. These plots illustrate the density of predicted probabilities for all compounds, with the decision boundary (0.5) marked as a reference threshold separating predicted actives from inactives.

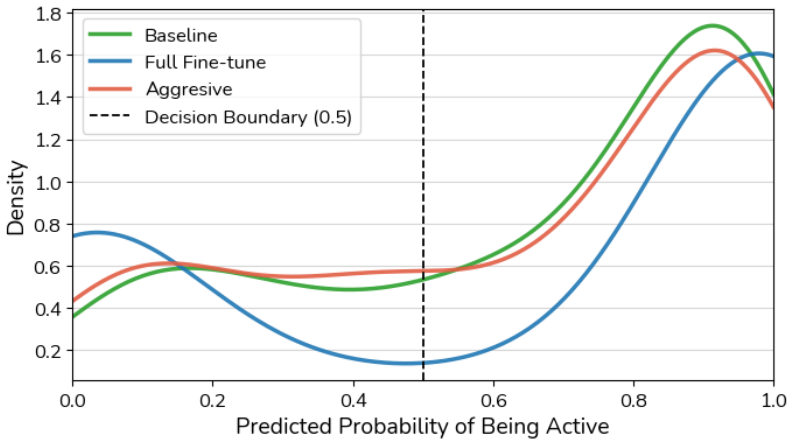

Figure 7. Molecular activity distribution plots for ChemBERTa configurations.

As shown in Figure 7, the baseline model displayed a flatter probability distribution with less distinct separation between active and inactive regions, suggesting weaker confidence in classification boundaries. The full fine-tune and aggressive configurations exhibited sharper bimodal distributions with higher densities near probabilities of 0 and 1, indicating stronger class separation and more decisive predictions. The aggressive model showed slightly more compact peaks, reflecting consistent confidence in distinguishing between active and inactive compounds.

The clear separation observed in the full fine-tune and aggressive models aligns with their improved accuracy and F1-scores reported earlier, confirming that deeper fine-tuning enhances the model's ability to internalize discriminative molecular features. The KDE profiles further indicate that while all models identify the general activity boundary, the degree of prediction sharpness correlates with training depth and learning rate intensity.

The comparative results across all configurations demonstrate that the extent of fine-tuning in ChemBERTa substantially influences predictive performance, calibration, and representational confidence for QSAR modeling of AXL kinase inhibitors. The full fine-tune and aggressive configurations consistently outperformed the baseline model in accuracy, precision, and F1-score, confirming that allowing deeper parameter updates enables more effective adaptation to the molecular characteristics of the target kinase. The aggressive configuration, in particular, balanced strong predictive power with well-calibrated probabilities, as evidenced by its lowest expected calibration error and sharpest activity distribution. These findings underscore that careful tuning of learning rate and regularization is essential for optimizing transformer-based chemical language models without compromising probabilistic reliability.

From an application perspective, these outcomes highlight the potential of ChemBERTa-based QSAR frameworks as reliable screening tools in early-stage drug discovery. Improved calibration and class discrimination directly translate to more confident prioritization of candidate molecules for synthesis and testing, thereby reducing false positives and experimental costs. The insights from this work can guide model selection and hyperparameter

strategies when deploying molecular transformers for specific therapeutic targets or under data-limited conditions.

However, several limitations should be noted. The dataset was limited to a single kinase target (AXL) and binary classification based on IC_{50} thresholds, which may not fully represent broader chemical diversity or multi-target selectivity. Furthermore, the study did not explore advanced fine-tuning techniques such as layer-wise learning rate decay, adapter-based tuning, or contrastive pretraining, which could yield more efficient parameter transfer. Future work should extend this analysis to larger, multi-target datasets and investigate transferability across related kinases. Incorporating uncertainty quantification, explainability methods, and active learning loops could further enhance interpretability and data efficiency, making ChemBERTa-based QSAR pipelines more robust and applicable to real-world medicinal chemistry workflows.

Conclusions

This study demonstrated that the extent of fine-tuning in ChemBERTa substantially affects its predictive accuracy, calibration, and generalization for QSAR modeling of AXL kinase inhibitors. Comparative analysis of baseline, full fine-tune, and aggressive configurations revealed that deeper fine-tuning improves precision, F1-score, and class discrimination by enabling greater task-specific adaptation, while moderate regularization preserves calibration reliability. Among the evaluated strategies, the aggressive configuration achieved the best balance between predictive power and probabilistic stability, indicating that controlled unfreezing combined with an optimized learning rate can enhance transformer-based molecular representations without overfitting. These findings highlight the potential of ChemBERTa as a robust framework for activity prediction and virtual screening in kinase-focused drug discovery and provide practical guidance for designing fine-tuning protocols that maximize both performance and interpretability in data-driven medicinal chemistry.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

[1] Zhong L, Li Y, Xiong L, Wang W, Wu M, Yuan T, et al. Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal Transduction and Targeted Therapy* 2021;6:201. https://doi.org/10.1038/s41392-021-00572-w.

[2] Liu G, Chen T, Zhang X, Ma X, Shi H. Small molecule inhibitors targeting the cancers. *MedComm* 2022;3. https://doi.org/10.1002/mco2.181.

[3] Min H-Y, Lee H-Y. Molecular targeted therapy for anticancer treatment. *Experimental & Molecular Medicine* 2022;54:1670–94. https://doi.org/10.1038/s12276-022-00864-3.

[4] Patwekar F, Patwekar M, Kamal MA. Synergizing phytonanotherapy and complementary medicine: Future horizons in cancer and diabetes care. *Global Translational Medicine* 2025;4:16. https://doi.org/10.36922/gtm.5840.

[5] Noviandy TR, Maulana A, Emran TB, Idroes GM, Idroes R. QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms. *Heca Journal of Applied Sciences* 2023;1:1–7. https://doi.org/10.60084/hjas.v1i1.12.

[6] Noviandy TR, Maulana A, Idroes GM, Maulydia NB, Patwekar M, Suhendra R, et al. Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery. *Malacca Pharmaceutics* 2023;1:48–54. https://doi.org/10.60084/mp.v1i2.60.

[7] Patwekar M, Patwekar F, Shaikh D, Fatema SR, Aher SJ, Sharma R. Receptor-based approaches and therapeutic targets in Alzheimer's disease along with role of AI in drug designing: Unraveling pathologies and advancing treatment strategies. *Applied Chemical Engineering* 2023;6. https://doi.org/10.24294/ace.v6i3.2338.

[8] Ganorkar SB, Heyden Y Vander. Recent trends in pharmaceutical analysis to foster modern drug discovery by comparative in-silico profiling of drugs and related substances. *TrAC Trends in Analytical Chemistry* 2022;157:116747. https://doi.org/10.1016/j.trac.2022.116747.

[9] Kenneth C, Imani A, Pardamean B. Leveraging ChemBERTa Robustness in Drug-Drug Interaction Classification via Molecular Decomposition. 2024 7th Int. Semin. Res. Inf. Technol. Intell. Syst., IEEE; 2024, p. 694–9. https://doi.org/10.1109/ISRITI64779.2024.10963372.

[10] Vatsyayan S. Leveraging Machine Learning and ChemBERTa for Efficient Identification of CB1 Receptor-Active Compounds. 2025 IEEE Conf. Artif. Intell., IEEE; 2025, p. 600–5. https://doi.org/10.1109/CAI64502.2025.00273.

[11] Patil A, Singh N, Patwekar M, Patwekar F, Patil A, Gupta JK, et al. AI-driven insights into the microbiota: Figuring out the mysterious world of the gut. *Intelligent Pharmacy* 2025;3:46–52. https://doi.org/10.1016/j.ipha.2024.08.003.

[12] Sadeghi S, Bui A, Forooghi A, Lu J, Ngom A. Can large language models understand molecules? *BMC Bioinformatics* 2024;25:225. https://doi.org/10.1186/s12859-024-05847-x.

[13] Mswahili ME, Jeong Y-S. Transformer-based models for chemical SMILES representation: A comprehensive literature review. *Heliyon* 2024;10:e39038. https://doi.org/10.1016/j.heliyon.2024.e39038.

[14] Goyette M-A, Côté J-F. AXL Receptor Tyrosine Kinase as a Promising Therapeutic Target Directing Multiple Aspects of Cancer Progression and Metastasis. *Cancers* 2022;14:466. https://doi.org/10.3390/cancers14030466.

[15] Engelsen AST, Lotsberg ML, Abou Khouzam R, Thiery J-P, Lorens JB, Chouaib S, et al. Dissecting the Role of AXL in Cancer Immune Escape and Resistance to Immune Checkpoint Inhibition. *Frontiers in Immunology* 2022;13. https://doi.org/10.3389/fimmu.2022.869676.

[16] Noviandy TR, Idroes GM, Hardi I. Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using XGBoost and Bayesian Optimization. *Journal of Soft Computing and Data Mining* 2024;5:46–56.

[17] Wang R, Ji Y, Li Y, Lee S-T. Applications of Transformers in Computational Chemistry: Recent Progress and Prospects. *The Journal of Physical Chemistry Letters* 2025;16:421–34. https://doi.org/10.1021/acs.jpclett.4c03128.

[18] Passi N, Raj M, Shelke NA. A Review on Transformer Models: Applications, Taxonomies, Open Issues and Challenges. 2024 4th Asian Conf. Innov. Technol., IEEE; 2024, p. 1–6. https://doi.org/10.1109/ASIANCON62057.2024.10838047.

[19] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Research* 2012;40:D1100–7. https://doi.org/10.1093/nar/gkr777.

[20] Noviandy TR, Idroes GM, Mohd Fauzi F, Idroes R. Application of Ensemble Machine Learning Methods for QSAR Classification of Leukotriene A4 Hydrolase Inhibitors in Drug Discovery. *Malacca Pharmaceutics* 2024;2:68–78. https://doi.org/10.60084/mp.v2i2.217.

[21] Bento AP, Hersey A, Félix E, Landrum G, Gaulton A, Atkinson F, et al. An open source chemical structure curation pipeline using RDKit. *Journal of Cheminformatics* 2020;12:51. https://doi.org/10.1186/s13321-020-00456-1.

[22] Noviandy TR, Idroes R. Interpretable Machine Learning QSAR Models for Classification and Screening of VEGFR-2 Inhibitors in Anticancer Drug Discovery. *Malacca Pharmaceutics* 2025;3:58–66. https://doi.org/10.60084/mp.v3i2.339.

[23] Chithrananda S, Grand G, Ramsundar B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. *ArXiv Preprint ArXiv:201009885* 2020.

[24] Goutam K, Balasubramanian S, Gera D, Sarma RR. LayerOut: Freezing Layers in Deep Neural Networks. *SN Computer Science* 2020;1:295. https://doi.org/10.1007/s42979-020-00312-x.

[25] Noviandy TR, Maulana A, Idroes GM, Suhendra R, Afidh RPF, Idroes R. An Explainable Multi-Model Stacked Classifier Approach for Predicting Hepatitis C Drug Candidates. *Sci* 2024;6:81. https://doi.org/10.3390/sci6040081.

[26] Noviandy TR, Maulana A, Irvanizam I, Idroes GM, Maulydia NB, Tallei TE, et al. Interpretable Machine Learning Approach to Predict Hepatitis C Virus NS5B Inhibitor Activity Using Voting-Based LightGBM and SHAP. *Intelligent Systems with Applications* 2025;25:200481. https://doi.org/10.1016/j.iswa.2025.200481.

[27] Tharwat A. Classification Assessment Methods. *Applied Computing and Informatics* 2021;17:168–92. https://doi.org/10.1016/j.aci.2018.08.003.