scientific reports

Check for updates

OPEN

ResisenseNet hybrid neural network model for predicting drug sensitivity and repurposing in breast Cancer

Anush Karampuri, Bharath Kumar Jakkula & Shyam Perugu✉

Breast cancer remains a leading cause of mortality among women worldwide, with drug resistance driven by transcription factors and mutations posing significant challenges. To address this, we present ResisenseNet, a predictive model for drug sensitivity and resistance. ResisenseNet integrates transcription factor expression, genomic markers, drugs, and molecular descriptors, employing a hybrid architecture of 1D-CNN + LSTM and DNN to effectively learn long-range and temporal patterns from amino acid sequences and transcription factor data. The model demonstrated exceptional predictive accuracy, achieving a validation accuracy of 0.9794 and a loss value of 0.042. Comprehensive validation included comparisons with state-of-the-art models and ablation studies, confirming the robustness of the developed architecture. ResisenseNet has been applied to repurpose existing anticancer drugs across 14 different cancers, with a focus on breast cancer. Among the malignancies studied, drugs targeting Low-grade Glioma (LGG) and Lung Adenocarcinoma (LUAD) showed increased sensitivity to breast cancer as per ResisenseNet's assessment. Further evaluation of the predicted sensitive drugs revealed that 14 had no prior history of anticancer activity against breast cancer. These drugs target key signaling pathways involved in breast cancer, presenting novel therapeutic opportunities. ResisenseNet addresses drug resistance by filtering ineffective compounds and enhancing chemotherapy for breast cancer. In vitro studies on sensitive drugs provide valuable insights into breast cancer prognosis, contributing to improved treatment strategies.

Keywords Drug resistance, Breast adenocarcinoma, Drug repurposing, DoRothEA, ResisenseNet

Breast cancer, characterized by its genetic variability, poses a significant threat to women, with drug resistance being a critical concern [1]. Factors such as transcription factors and genomic mutations intricately impact treatment outcomes, emphasizing the need for targeted intervention strategies[2]. Transcription factors and genomic markers play a crucial role in malignancy drug resistance. Dysregulation of transcription factors, such as STAT3, leads to aberrant activities, including the maintenance of cancer stem cell phenotypes and immune response modulation, favouring drug resistance and metastasis. Activation of STAT3 via growth factor receptors like EGF, IGF, PDGF, and VEGF triggers pathways such as JAK2, G-alpha0, and G-alpha1, promoting resistance to chemotherapeutic agents and radiotherapy. Tumor microenvironments, characterized by increased CD133 + cancer stem cells, contribute to chemotherapeutic resistance via STAT activation [3]. Mutations in the BCR-ABL1 Kinase Domain result in STAT3 hyperactivation and imatinib resistance in chronic myeloid leukaemia. CD44 + CD24-/low cells activate STAT3, inducing antiapoptotic changes in breast cancer cells treated with tamoxifen, while fibronectin, IL-6, and EGF-mediated STAT3 activation lead to trastuzumab resistance [3,4].

Additionally, HIF-1 alpha expression, associated with poor prognosis, correlates with drug resistance, particularly in HER + tumours with PTEN negativity [56]. HIF-1 alpha upregulates autophagy and maintains stemness, contributing to drug resistance against cytotoxic drugs such as doxorubicin, paclitaxel, gemcitabine, methotrexate, and docetaxel [5,7]. FOXO transcription factors mediate anticancer agent activity via PI3K inhibition and EGFR/HER2 inhibition[8,9], with low expression leading to resistance against paclitaxel and epirubicin [10]. Mutations, such as loss of tumor suppressor gene function and alterations in critical genes, impacting NF-KB involvement in MDR1 gene expression, lead to doxorubicin resistance in breast cancer through apoptosis inhibition and increased drug efflux [11]. For instance, mutations at R248Q of P53 lead to doxorubicin resistance, while

Department of Biotechnology, National Institute of Technology, Warangal 500604, India. ✉email: shyamperugu@nitw.ac.in

phosphorylation-induced loss of function of RB1 causes 5-FU resistance in breast, colon, and liver malignancies [12]. Deacetylation of SOX9 by histone deacetylase 5 (HDAC 5) results in tamoxifen resistance, and SOX9 upregulation induces resistance to trastuzumab and cisplatin via activation of PI3K/AKT/mTOR signalling pathways [13].

There is an urgent need to explore alternative drug molecules that can effectively replace drugs prone to resistance. Various studies have been conducted to repurpose drugs for this purpose. Kanghao Shao et al., 2022 introduced the regression-based machine learning model DTI-HETA, utilizing a novel Graphical Neural Network, HETA-DTI, designed for predicting drug-target interactions. This model leverages a heterogenous graph with attention mechanisms, integrating drug-drug and target-target similarity matrices with the DTI matrix to enhance prediction accuracy [14]. Ahmet Sureyya Rifaioglu et al., 2020 developed a prediction model that differs from conventional approaches by utilizing two-dimensional representations of structures instead of descriptors. Their model, trained on 704 target proteins, identified JAK proteins as novel targets for the drug cladribine through docking analysis [15]. Chang Sun et al., 2020 proposed a graph-based convolutional autoencoder and generative adversarial network, GANDTI, for drug-target interaction prediction. Their approach involved constructing heterogenous drug-drug and drug-target interaction networks and demonstrated efficacy through a case study on five molecules, showcasing GANDTI's ability to identify drug targets[16]. Qiao Liu et al., 2020 introduced the Deep CDR model, a hybrid model incorporating omics datasets and molecular descriptors for comprehensive prediction. Unlike traditional models, Deep CDR integrates convolutional neural networks and multiple subnetworks to understand complex relationships between omics expression profiles and molecular descriptors, facilitating precise drug-target interaction prediction[17].

In the evolving field of drug discovery, prior research has largely concentrated on regression and classification methodologies to predict drug-target interactions, thereby enhancing early-stage drug repositioning and repurposing efforts. Notably, the DeepDRK model developed by Yongcui Wang et al. in 2021 employs a machine learning framework that integrates multi-omics datasets to forecast drug responses in cancer treatment, analyzing over 20,000 combinations of cancer cell lines and drugs to identify promising candidates [18]. Similarly, Chen Cui et al. (2021) introduced GraphRepur, a graph neural network-based approach aimed at drug repositioning specifically for breast cancer, which combines drug exposure data with gene expression patterns and drug-drug interaction information to uncover novel applications for existing medications [19]. Furthermore, Priyanka et al. (2022) utilized a machine learning strategy to identify potential RET inhibitors for non-small cell lung cancer, encompassing model development, in silico validations, and lead identification[20], thereby underscoring the significant role of machine learning in advancing drug repurposing initiatives across various cancer types.

While these studies have laid a solid foundation and identified research gaps, there remains a pressing need to discern the sensitivity and resistance profiles of molecules before experimental procedures. In response to the pressing demand for innovative chemotherapeutic strategies to tackle drug resistance in malignancies, ResisenseNet, a deep learning-based predictive model was developed. The primary aspect that distinguishes the developed model from existing similar neural networks is its utilization of the DoRothEA (Discriminant Regulon Expression Analysis) database to train the model, along with its unique architecture that allows for the incorporation of protein-specific data to uncover patterns. Previous research using the DoRothEA database focused on developing a multi-omics approach to identify genetic factors contributing to resistance against the drug olaparib[21] and a machine learning and directed network optimization approach to unveil the regulatory patterns of the TP53 gene [22]. However, these studies were narrowly focused on understanding the regulatory functions of a single gene and its characterization.

The ResisenseNet model uniquely utilizes datasets from the DoRothEA repository, encompassing all available anticancer drugs alongside their 2D and 3D molecular descriptors, mutational markers, transcription factor expression data, and amino acid sequence patterns of transcription factors and target proteins. This comprehensive approach aims to elucidate the complex relationships within breast adenocarcinoma, facilitating the development of a predictive model for drug sensitivity and repurposing from other malignancies. In contrast to prior regression-based methodologies that primarily target drug-target interactions and binding affinities without exploring the molecular intricacies of drugs, transcription factors, and genomic markers, our classification framework enhances the identification of subtle patterns and allows for more effective categorization of drugs based on sensitivity profiles. This innovative strategy promises to refine early-stage drug discovery processes and guide personalized therapeutic interventions with greater precision and efficacy.

Results

As outlined in the methodology section, data retrieval, preprocessing, and model training were successfully completed. We employed a hybrid neural network consisting of two modules one to learn features from Amino acid sequences of Transcription factors and target proteins using a complex of 1D-CNN and LSTM, and the other module to learn the intricate features from transcription factors, genomic markers, and molecular descriptors using a DNN. The Hyperopt finetuning based hyperparameters were employed for the algorithms and were tabulated in Table 1. Please refer to Fig. 1 for model architecture.

The ablation studies conducted were represented in Fig. 2A, which delineates the performance of the model employing solely a 1D-CNN, which yielded loss and accuracy values of 0.088 and 0.792, representing the lowest metrics among the conducted experiments. Conversely, the integration of a 1D-CNN and an LSTM module, as illustrated in Fig. 2B, exhibited improved performance metrics, achieving loss and accuracy values of 0.059 and 0.815. To further augment the model's predictive capabilities, a DNN architecture was concatenated with the 1D-CNN, resulting in substantial enhancements, as depicted in Fig. 2C, with loss and accuracy values of 0.048 and 0.942. Finally, an additional experiment combined the output vector representation from the DNN with both the LSTM and 1D-CNN modules, attaining the pinnacle of validation metrics among all studies, with loss and accuracy values of 0.042 and 0.979, as represented in Fig. 2D (Fig. 3).

S.NO	Purpose	Algorithm	Hyperparameters
1	Target and Transcription Factor feature learning	1D-CNN (one dimensional - Convolutional Neural network)	Char_level: True, truncating: post, padding: post, maxlen (smiles): 120, maxlen (proteins): 1200, vocab_size (smiles): Length of word_index_smiles + 1, vocab_size (Proteins): Length of word_index_proteins + 1, embedding_dim: 128, Filters (conv_smiles) : 32, 64, 96, Filters (conv_protein) : 64, 98, 126, kernel_size (conv_smiles) : 4, 6, 8, kernel_size (conv_protein) : 4, 8, 12, optimizer: Adam, Learning Rate: 0.01, epochs: 100, batch_size: 25, loss: 'binary cross entropy', early_stopping_param: min_delta = 0.001, patience = 10, restore_best_weights = True
2	Target and Transcription Factor feature learning	LSTM (Long short-term memory), a RNN (Recurrent neural network) module	Layers: 2; Number of neurons: 128, 256; Learning rate: 0.002; Batch size: 64; Epochs: 75; Dropout Rate: 0.3; Bidirectional: True; Optimizer: Adam
3	Ligand, Genomic markers, and TF activity feature learning	DNN (Deep Neural Network)	Input layer: 3850, hidden layers : 6000, 3000, 500, output layer: 2, Activation Functions : 'relu', 'tanh', 'tanh', 'relu', 'tanh', Dropout Rate: 0.05, Optimizer: SGD (Stochastic Gradient Descent), Learning rate: 0.0001, Momentum: 0.9, Loss_Function: 'Binary Cross entropy', Metrics: 'accuracy', Epochs: 150, Batch size: 64, Validation Split: 0.30, Early Stopping Patience: 50, Restore Best Weights: True

Table 1. Hyperopt based optimized hyperparameters used to build the ResisenseNet model.

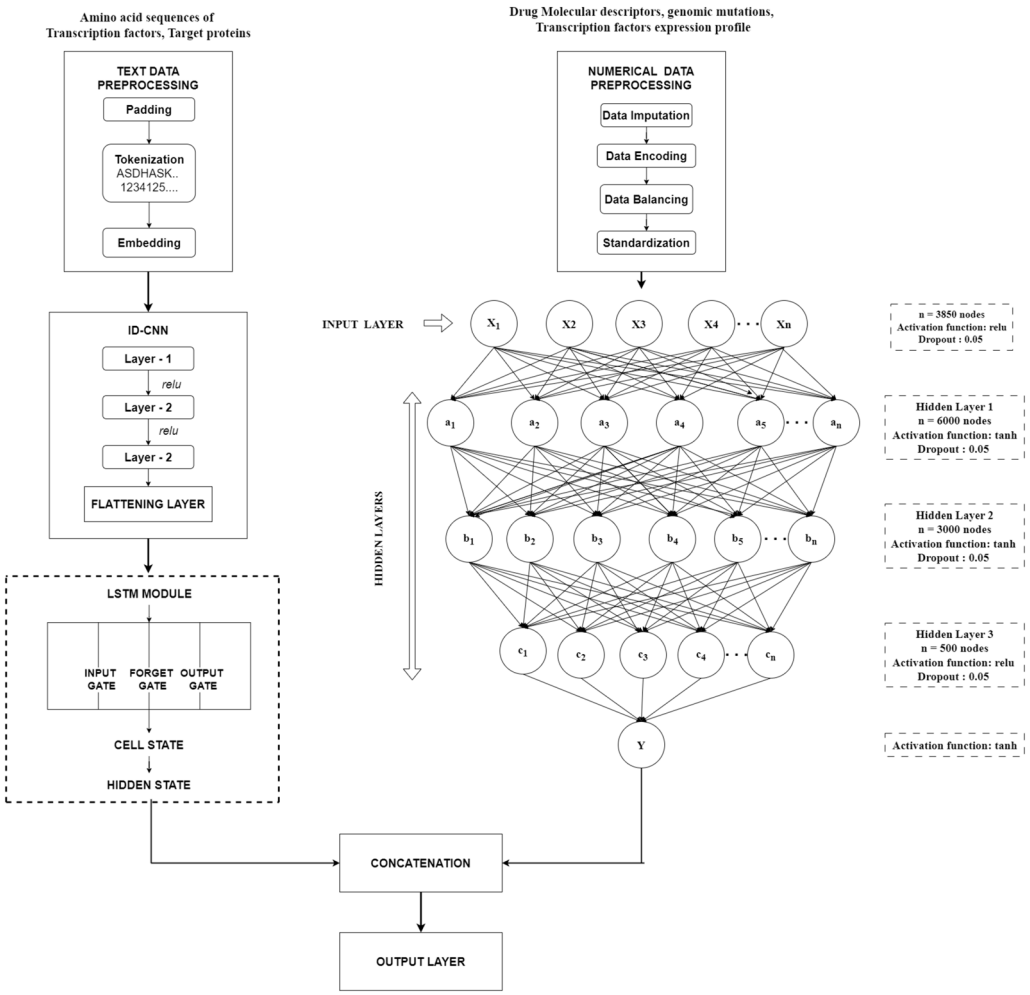

Fig. 1. Architecture of the ResisenseNet model. This model integrates a 1D-CNN with an LSTM (Long Short-Term Memory) module to capture long-range and temporal patterns from amino acid sequences, alongside a DNN (Deep Neural Network) module designed to analyse patterns from numerical datasets. The learned feature representation vectors from both modules are concatenated to form a final output layer capable of making binary classification predictions.

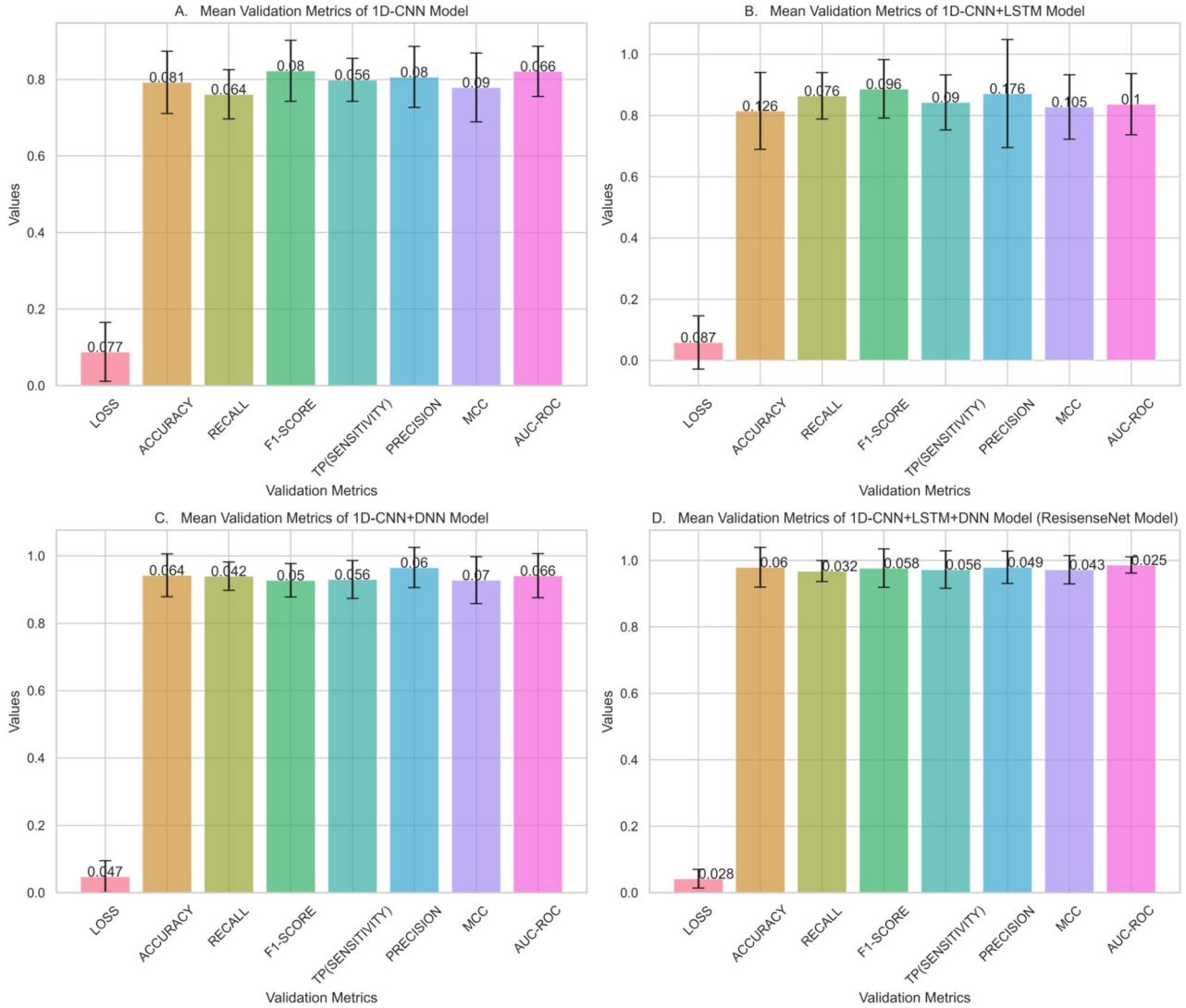

Fig. 2. Bar plots illustrating the results of the ResisenseNet model's ablation studies. Panel A depicts the model utilizing only the 1D-CNN module; Panel B shows the combination of the 1D-CNN and LSTM modules; Panel C presents the integration of the 1D-CNN and DNN modules, designed to capture patterns from both numerical and text-based data; and Panel D represents the final ResisenseNet model, incorporating all three modules: 1D-CNN, LSTM, and DNN.

The comprehensive validation metrics profiles for all experiments conducted during the ablation studies are presented in Supplementary Table S5. Supplementary Fig. S2 and S4 depict the trends in loss and accuracy throughout the epochs of training the ResisenseNet model. Figure 4 displays the confusion matrix, which illustrates the predictive performance of the ResisenseNet model on both the test and validation datasets. The results of the 10-fold cross-validation are detailed in Supplementary Fig. S3and Supplementary Table S10.

Thorough ablation studies have been meticulously conducted to finalize the model complexity. The predictive prowess was rigorously evaluated using test and validation sets, which were judiciously partitioned employing an initial test-train split. To ensure reproducibility, 12 random seed experiments were executed for both the test and validation sets. The model demonstrated commendable performance on the test set, yielding a loss value of 0.042 and an accuracy of 0.9794. It achieved a recall value of 0.968, an F1-score of 0.977, a sensitivity (True Positive Rate) of 0.9725, a precision (Positive Prediction Value) of 0.9795, a Matthews Correlation Coefficient of 0.9721, and an AUC-ROC Score of 0.9864, as derived from the confusion matrix, underscoring the model's exceptional fit. To further assess its predictive performance and generalizability, an internal validation set was employed, where the model exhibited a validation set accuracy of 0.9575 and a validation loss of 0.069. Additionally, it demonstrated a recall value of 0.948, an F1-score of 0.956, a sensitivity value of 0.9677, a precision value of 0.9219, a Matthews Correlation Coefficient of 0.9487, and an AUC-ROC Score of 0.9667. Please refer to the Fig. 3 and Supplementary Table S7. These findings highlight the promising generalizability capabilities of the ResisenseNet framework.

The baseline logistic classifier, trained on the same dataset as the ResisenseNet model, was evaluated for its predictive performance, yielding validation metrics with a loss value of 0.625 and an accuracy of 0.642.

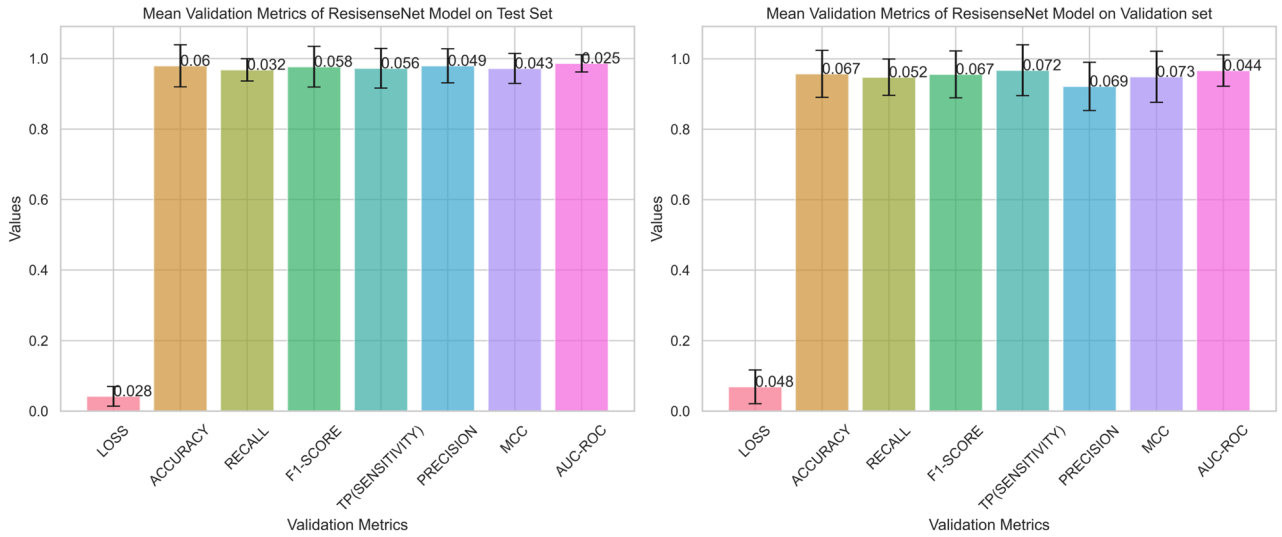

Fig. 3. Graphical representation of the validation metrics used to evaluate the developed ResisenseNet model on both the test set and validation set. The numerical values displayed above the bars in the plot indicate the standard deviation, which serves as error bars for the various random seeds tested.

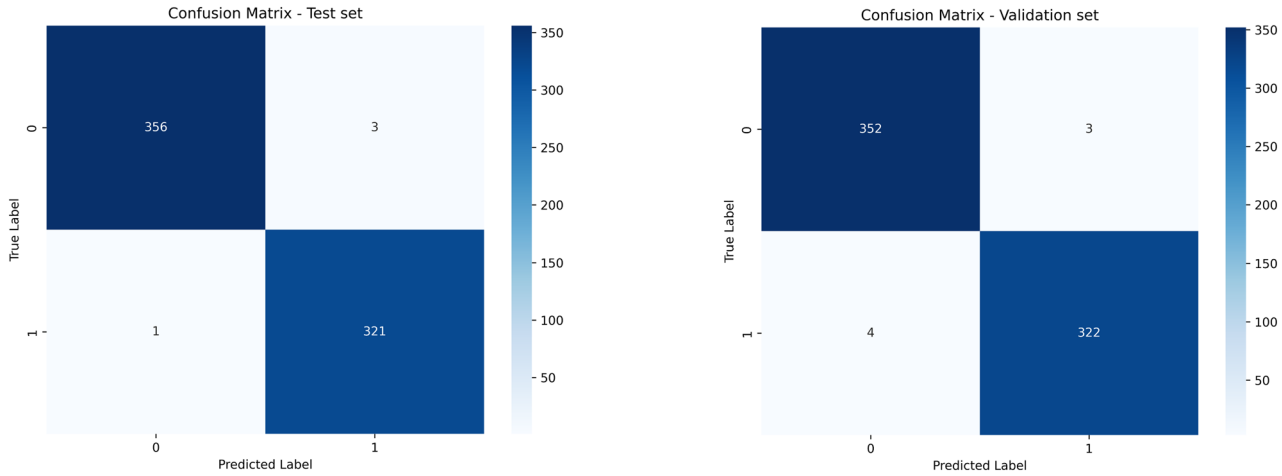

Fig. 4. Heatmap depicting the confusion matrix for the test and validation sets. In the matrix, the "True Label" represents the actual nature of the drugs, while the "Predicted Label" reflects the predicted outcome. This allows for the measurement of true positives, true negatives, false positives, and false negatives, which are the key components for calculating the validation metrics used to evaluate the model's performance. The value 0 represents sensitive data points, and 1 represents resistant data points.

Additional validation metrics are detailed in Supplementary Fig. S5 and S6, which also includes the hyperparameters employed for this model. The comparison between the baseline model and the developed ResisenseNet model aimed to assess the complexity and potential of the latter in capturing intricate features from proteins, transcription factor (TF) expression, genomic markers, and drug molecules. This evaluation serves as a critical benchmark for understanding the advancements achieved through the more sophisticated model.

To evaluate the performance of the developed model in out-of-distribution scenarios, we utilized test sets from different cancer types, specifically colorectal adenocarcinoma (COREAD) and lung adenocarcinoma (LUAD), which differ from the breast cancer data used for training in terms of transcription factor expression profiles, utilized drugs, molecular descriptors, genomic mutations, and various targets. This variation allows for a robust assessment of the model's generalizability potential. The validation metrics for the LUAD group yielded a loss value of 0.062 and an accuracy of 0.9014, while the COREAD group demonstrated a loss of 0.058 and an accuracy of 0.924. Additional validation metrics are illustrated in Supplementary Table S8 and Fig. 5. These results underscore the ResisenseNet model's generalizability and its stability in reproducing outcomes across diverse datasets.

A character level-CNN + DNN model was trained on the same dataset used to train the ResisenseNet model. Both models exhibited comparable results, with mean validation metrics from 12 random seeds and their standard deviations presented in Fig. 6 and Supplementary table S9. The SOTA model achieved loss and accuracy values of 0.0356 and 0.988, respectively, outperforming the ResisenseNet model, which recorded loss and accuracy

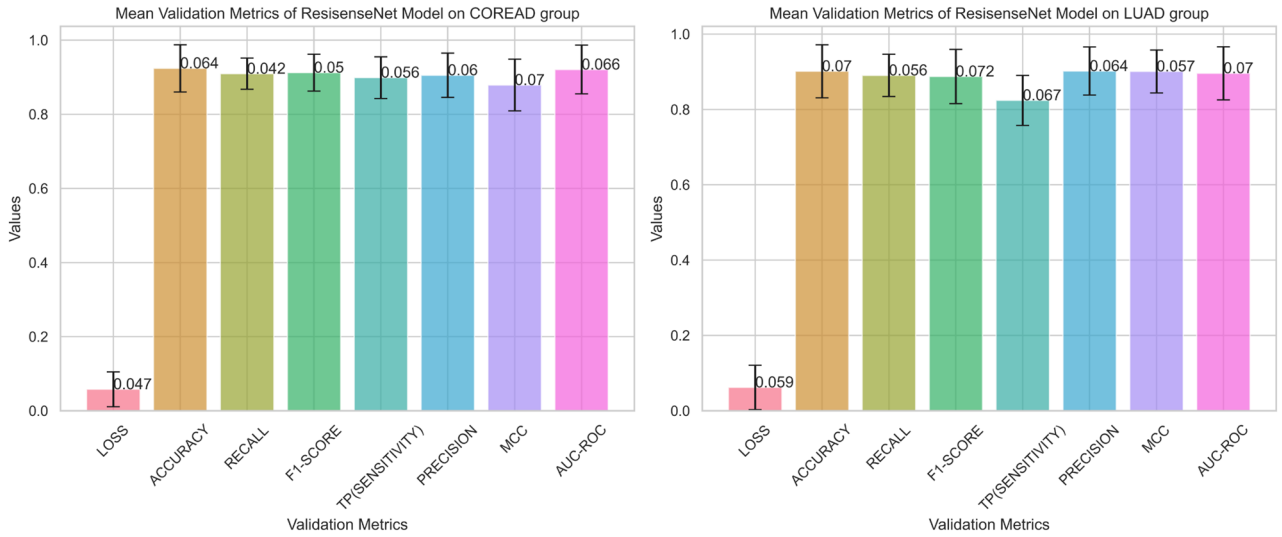

Fig. 5. Bar plot illustrating the validation metrics across different cancer groups (COREAD – colorectal adenocarcinoma and LUAD – lung adenocarcinoma) to assess the ResisenseNet model's performance in out-of-distribution scenarios, highlighting its generalizability and resilience to variations. The numerical values displayed above the bars in the plot indicate the standard deviation, which serves as error bars for the various random seeds tested.

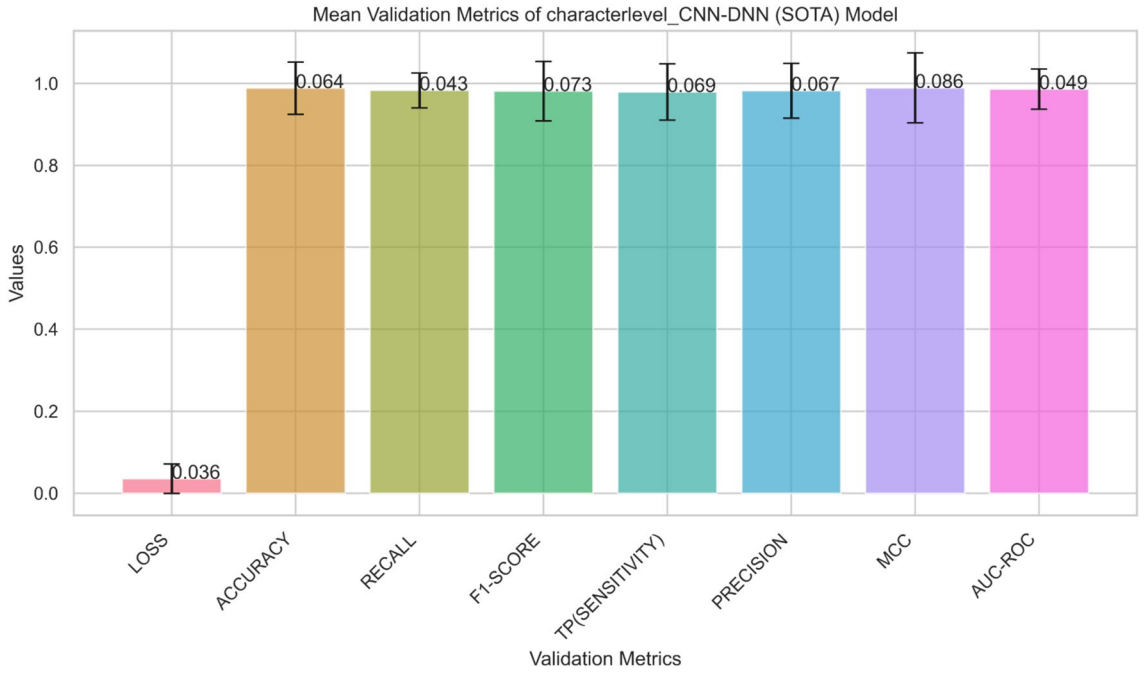

Fig. 6. Validation metrics of the state-of-the-art (SOTA) model used for comparison with the developed ResisenseNet model, highlighting performance differences and effectiveness across various evaluation criteria. The numerical values displayed above the bars in the plot indicate the standard deviation, which serves as error bars for the various random seeds tested.

values of 0.042 and 0.9794. Other validation metrics were similarly close, indicating no significant differences. A Wilcoxon signed-rank test was performed to assess statistical significance, revealing no notable differences in performance across random seeds. However, the SOTA model exhibited significantly larger standard deviations across all validation metrics, indicating instability in its predictions and reproducibility. In contrast, the ResisenseNet model demonstrated superior stability in prediction performance across various random seeds. This comparison provides a holistic view of the models' predictive capabilities.

The developed ResisenseNet model was used to screen the drugs which were currently being administered for various other cancers (14 different cancers, please refer to Supplementary Table S4) to identify the potential drugs which were sensitive or resistant in the presence of the interaction coefficients among genomic markers,

transcription factors, targets and drugs. Please refer to Supplementary Tables S1, S2, and S3. Among the 14 distinct malignancies examined, the drugs targeting Low-grade Glioma (LGG) and Lung Adenocarcinoma (LUAD) displayed heightened sensitivity toward breast cancer upon assessment through the developed model. Conversely, drugs specific to Colorectal Adenocarcinoma (COREAD) demonstrated a greater resistance against breast cancer. Following COREAD, Bladder Urothelial Carcinoma (BLCA), Lung Squamous Cell Carcinoma (LUSC), and Low-grade Glioma (LGG) exhibited similar trends of resistance against breast cancer. Please refer to Fig. 7.

In this study, we conducted a SHAP (SHapley Additive exPlanations) analysis to elucidate the contributions of various molecular descriptors to our predictive model. The SHAP values reveal the significance of each feature in influencing the model's output, enhancing our understanding of how different molecular characteristics affect drug behavior. The horizontal bar plot (Supplementary Fig. S7) illustrates the mean absolute SHAP values for the most impactful descriptors, with those related to molecular size, shape, and electrostatic properties showing the highest influence, each with a SHAP value of 3. These findings indicate that variations in these features significantly impact the predicted activity or efficacy of the drugs. Additionally, descriptors such as aromaticity and hydrophobicity also contributed notably, with SHAP values of 2.5 and 2, respectively, while descriptors like atomic connectivity and bond types, though less influential with SHAP values ranging from 1.5 to 1, still play a meaningful role in the model.

A comprehensive analysis of genomic markers was conducted to assess the efficacy of drugs in treating various malignancies, particularly focusing on breast cancer. This involved screening different drugs tailored to specific types of cancer. Each drug exhibited unique genomic markers that could account for its effectiveness or resistance across different cancer types. For instance, a drug might demonstrate sensitivity in breast cancer while being resistant to other cancers, or vice versa. Tables 2 and 4 detail the specific malignancies, corresponding drugs, and their respective targets, along with genomic markers indicating sensitivity or resistance in breast cancer. Meanwhile, Supplementary Fig. S6 provides a visual representation in the form of a heatmap plot, encompassing all considered malignancies. This plot highlights the varying responses of drugs across different cancer types, depicting both sensitivity and resistance patterns. Table 2 presents a comprehensive overview of sensitive drugs along with their corresponding mutations and targets across various cancer types. The identified sensitive drugs exhibit a diverse range of mechanisms targeting specific genomic alterations, thereby highlighting potential avenues for precision medicine in breast cancer treatment. For instance, ABT-263, identified as effective in COREAD, targets the BCL2 family proteins influenced by gain MET mutations, suggesting its potential to induce apoptosis in cancer cells harbouring these alterations. Similarly, Nutlin-3a, effective in GBM and SKCM, interacts with the MDM2 protein, offering a promising strategy for reactivating the p53 pathway in tumours with TP53 mutations. Furthermore, the identification of Dabrafenib as efficacious in THCA and BLCA, targeting BRAF mutations, underscores its relevance in MAPK pathway-driven cancers, highlighting the importance of personalized treatment strategies based on mutation profiles.

Drugs identified as sensitive to breast adenocarcinoma were further assessed for their historical involvement in anticancer research related to breast cancer, their FDA approval status, and whether they are novel compounds lacking prior research. To accomplish this, an extensive literature search was conducted using PubMed, DrugBank, and the FDA database, with the findings summarized in Table 3, which categorizes the drugs into three distinct sets: Set A, which comprises established and annotated FDA-approved drugs currently utilized for breast cancer treatment; Set B, which includes drugs identified through ResisenseNet predictions that have previously been investigated as potential anticancer agents against breast cancer, as well as those that enhance the efficacy

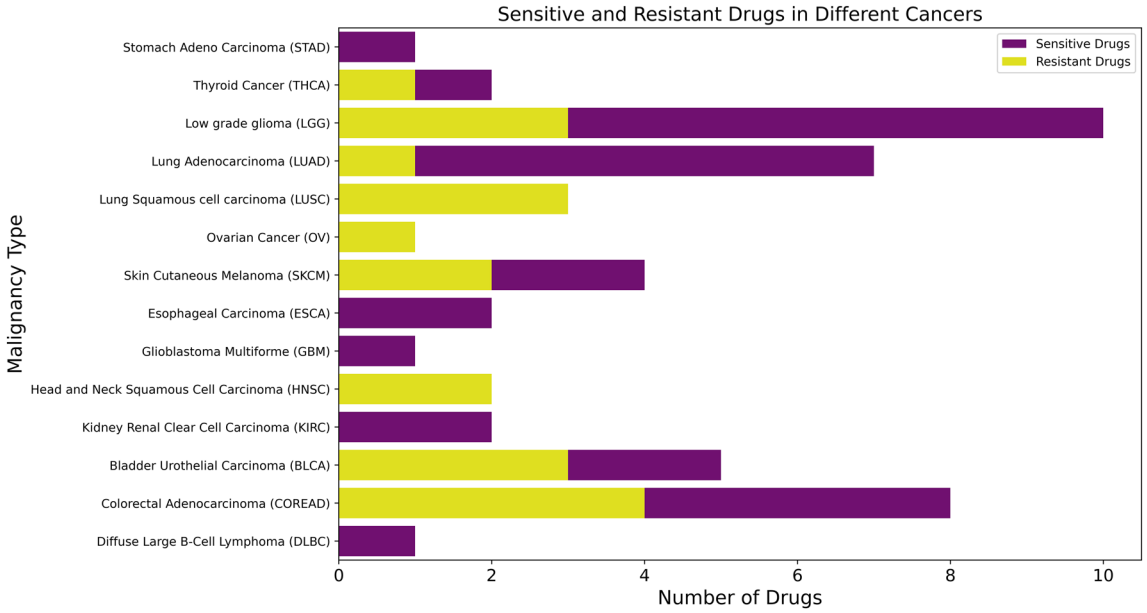

Fig. 7. Horizontal bar plot Representing the number of drugs from each among the 14 malignancies depicting resistance and sensitivity towards breast cancer based on the ResisenseNet model predictions.

CANCER TYPE	Drugs	MUTATIONS	Targets
COREAD	ABT-263	gain MET (cnaPANCAN129)	BCL2, BCL2L1, BCL2L2
	ABT-869	ATRX_mut	VEGFR and PDGFR family
	AR-42	CDK12_mut	HDAC
	AV-951	gain MYC (cnaPANCAN91)	VEGFR
	Bexarotene	LPHN2_mut	Retinioic acid X family agonist
	GW-2580	GNAS_mut	CSF1R (cFMS)
	NPK76-II-72-1	gain MYC (cnaPANCAN91)	PLK3
	rTRAIL	BRWD1_mut	TR10A (DR4), TR10B (DR5)
	XMD15-27	gain MYC (cnaPANCAN91)	CAMK2B, CLK2, DYRK1A, MAST1, STK39
DLBC	Nutlin-3a	TP53_mut	MDM2
ESCA	BX-795	gain CCND1, CTTN (cnaPANCAN59)	TBK1, PDPK1, IKK, AURKB, AURKC
	FK866	NOTCH1_mut	NAMPT
GBM	Nutlin-3a	TP53_mut	MDM2
LGG	AUY922	loss RPL22 (cnaPANCAN281)	HSP90
	AZD-2281	gain EGFR (cnaPANCAN124)	PARP1, PARP2
	BMS-345,541	loss RPL22 (cnaPANCAN281)	IKBKB
	CAL-101	loss RPL22 (cnaPANCAN281)	PI3Kdelta
	FK866	gain EGFR (cnaPANCAN124)	NAMPT
	LAQ824	loss RPL22 (cnaPANCAN281)	HDAC
	Olaparib	gain EGFR (cnaPANCAN124)	PARP1, PARP2
	PD-0332991	RB1_mut	CDK4, CDK6
	piperlongumine	gain EGFR (cnaPANCAN124)	Increases ROS levels
	Temozolomide	gain EGFR (cnaPANCAN124)	DNA Alkylating agent
	YK 4-279	gain EGFR (cnaPANCAN124)	RNA helicase A
	Zibotentan, ZD4054	gain EGFR (cnaPANCAN124)	Endothelin A Receptor
LUAD	Afatinib	EGFR_mut	ERBB2, EGFR
	Bosutinib	EGFR_mut	SRC, ABL, TEC
	EKB-569	EGFR_mut	EGFR
	FR-180,204	EGFR_mut	ERK
	Gefitinib	EGFR_mut	EGFR
	RDEA119	KRAS_mut	MAP2K1 (MEK1), MAP2K2 (MEK2)
	Trametinib	KRAS_mut	MAP2K1 (MEK1), MAP2K2 (MEK2)
OV	GSK690693	PIK3CA_mut	AKT
SKCM	Dabrafenib	BRAF_mut, NRAS_mut	BRAF
	Nutlin-3a	loss BNC2, CDKN2A, JAK2, PSIP1 (cnaPANCAN144), TP53_mut	MDM2
	OSI-906	ARID2_mut	IGF1R
	PLX4720	NRAS_mut	BRAF
	SB590885	NRAS_mut, BRAF_mut	BRAF
STAD	LY317615	BCOR_mut	PRKCB (PKCbeta)
	Dabrafenib	BRAF_mut	BRAF
	FR-180,204	BRAF_mut	ERK
	PLX4720	BRAF_mut	BRAF
THCA	Dabrafenib	BRAF_mut	BRAF
	FR-180,204	BRAF_mut	ERK
	PLX4720	BRAF_mut	BRAF
BLCA	Doxorubicin	TP53_mut	DNA Intercalating agent
	Gemcitabine	TP53_mut	DNA Replication
	IOX2	loss CHD9, CNOT1, CYLD, MMP2, NUP93 (cnaPANCAN117)	EGLN1
	KIN001-102	ARID1A_mut	AKT1
	MLN4924	ARID1A_mut	NEDD8-activating enzyme
	NSC-207,895	gain ASXL1 (cnaPANCAN363)	MDM4
	TPCA-1	ARID1A_mut	IKK

Table 2. Table illustrates the sensitive anticancer drugs identified for breast cancer following screening by the ResisenseNet model, alongside their associated malignancy types, genomic markers, and targeted mechanisms.

SET A - DRUGS	SET B - DRUGS	SET C	
		DRUGS	TARGETS
LUMINESPIB	PEVONEDISTAT39	**IDELALISIB**	PI3Kdelta
OLAPARIB	ENZASTAURIN40	**DAPORINAD**	NAMPT
PALBOCICIB	REFAMETINIB41	**DACINOSTAT**	HDAC
DOXORUBICIN	TRAMETINIB42	**YK 4-279**	RNA helicase A
	AFATINIB43	**ZIBOTENTAN**	Endothelin A Receptor
	BOSUTUNIB44	**PELITINIB**	EGFR
	PIPLARTINE	**FR-180,204**	ERK
	TEMOZOLOMIDE45	**GSK690693**	AKT
	REBEMADLIN	**DABRAFENIB**	BRAF
	BMS-345,54146	**LINSITINIB**	IGF1R
	BENZENEACETAMIDE	**PLX4720**	BRAF
	TIVOZANIB47, 48	**SB590885**	BRAF
	BEXAROTENE49, 50	**BENZOFURAZON**	MDM4
	VANDETANIB51, 52	**TPCA-1**	IKK
	NAVITOCLAX53, 54		

Table 3. Tabular representation of the history of sensitive drugs identified as anticancer agents for breast adenocarcinoma. Set A includes drugs that have already been recognized and annotated as anticancer agents for BRCA; Set B consists of drugs with prior research conducted on their efficacy as anticancer agents against BRCA; set C encompasses drugs that lack any previous research history and are not specifically annotated for breast cancer, identified as novel candidates through the predictions of the ResisenseNet model.

of other drugs when administered in combination and drugs currently in clinical trials that have not yet received complete FDA annotation as breast cancer-specific treatments, with references provided for each drug in Set B for cross-verification; and Set C, which contains drugs that are neither listed in DrugBank nor associated with any prior research on breast cancer, categorizing them as novel compounds identified through the repurposing studies conducted by the ResisenseNet model.

In contrast, Table 4 delineates resistant drugs against specific cancer types, shedding light on mutation patterns associated with drug ineffectiveness. The resistant drugs identified in this analysis exhibit distinct mutation profiles that confer resistance to their respective targets, posing challenges to therapeutic efficacy. For instance, Dabrafenib, deemed resistant in COREAD due to BRAF mutations, reflects the limitations of targeting the MAPK pathway in tumours with activated BRAF mutations. Similarly, GSK690693, ineffective in COREAD, showcases the impact of PTEN mutations on AKT signalling, leading to reduced sensitivity to AKT inhibitors. These findings underscore the intricate interplay between genomic alterations and drug response, emphasizing the need for tailored treatment approaches to overcome resistance mechanisms in breast cancer and other malignancies.

By leveraging this approach, we can efficiently pinpoint drugs that exhibit sensitivity or resistance in treating breast cancer, paving the way for drug repurposing endeavours. This is made possible through a model trained on extensive datasets containing activity profiles of transcription factors and genomic markers in patients, along with the drugs administered to them. By comprehensively understanding the interaction coefficients among these variables (Please refer to Supplementary Table S11), the ResisenseNet model is adept at predicting potential sensitive and resistant drugs for breast cancer.

Discussion

Garcia-Alonso et al., 2018 conducted an extensive investigation to explore the global impact of mutations and transcription factors on drug resistance. Their research revealed a multitude of transcription factors whose activity correlates with drug sensitivity and impedes tumour progression even post-drug administration[23]. The experimental data generated from this study has been made publicly available, with a detailed summary provided in Supplementary Tables S1, S2, and S3. This patient-specific experimental dataset served as the foundation for developing the ResisenseNet, a model designed for predicting drug sensitivity. While existing studies predominantly focus on drug-target interactions for drug repurposing and repositioning, our approach distinguishes itself by developing a model tailored to this unique dataset comprising experimentally validated data. This tailored approach enables more precise predictions, representing a novel advancement in addressing research gaps identified in the literature.

In developing the ResisenseNet architecture, comprehensive ablation studies revealed the unique contributions of each component, highlighting the model's capacity to capture diverse data aspects. The 1D-CNN excels in sequential feature extraction, while the LSTM effectively captures temporal dependencies, and the DNN enhances the model's ability to learn complex relationships. This synergistic integration is particularly advantageous for tasks requiring both spatial and temporal information. Rigorous validations, including 12 trials with a fixed random seed, underscore the robustness and reproducibility of the findings, mitigating concerns about random variation. Please refer to Fig. 2. The model's generalizability was further assessed through performance evaluations on test and internal validation sets, as well as in out-of-distribution scenarios involving diverse cancer datasets, demonstrating its resilience to variability and potential for transfer learning. Please refer to Figs. 3

CANCER TYPE	Drugs	MUTATIONS	TARGETS
COREAD	Dabrafenib	BRAF_mut	BRAF
	GSK690693	PTEN_mut	AKT
		CNOT1_mut	AKT
		MAP2K4_mut	AKT
		PTEN_mut	AKT
	PXD101, Belinostat	CLSPN_mut	HDAC
HNSC	EHT 1864	TP53_mut	Rac GTPases
	JQ1	FAT1_mut	BRD2, BRD3, BRD4
KIRC	AICAR	BAP1_mut	AAPK1 (AMPK) agonist
	Dabrafenib	NF2_mut	BRAF
LAML	5-Fluorouracil	TP53_mut	DNA antimetabolite
	AUY922	NRAS_mut	HSP90
	BAY 61-3606	NRAS_mut	SYK
	Bleomycin	NRAS_mut	DNA
	CAL-101	NRAS_mut	PI3Kdelta
	CGP-60,474	TP53_mut	CDK1,CDK2,CDK5,CDK7,CDK9
	CP466722	NRAS_mut	ATM
	Doxorubicin	TP53_mut, NRAS_mut	DNA intercalating
	Etoposide	TP53_mut, NRAS_mut	TOP2
	GSK-650,394	NRAS_mut	SGK3
	JNK-9 L	NRAS_mut	JNK
	KIN001-260	NRAS_mut	IKK
	NVP-TAE684	NRAS_mut	ALK
	OSI-027	NRAS_mut	MTORC1/2
	Paclitaxel	TP53_mut	Microtubules
	PI-103	NRAS_mut	PI3Ka, PRKDC (DNAPK)
	Pyrimethamine	NRAS_mut	Dihydrofolate reductase
	TG101348	NRAS_mut	JAK2
	TPCA-1	NRAS_mut	IKK
	XMD14-99	NRAS_mut	EPHB3, CAMK1
	Y-39,983	NRAS_mut	ROCK
	YM201636	NRAS_mut	FYV1
LUSC	Bicalutamide	MLL2_mut	ANDR (androgen receptor)
	KIN001-236	gain CCT5, TERT, TRIO (cnaPANCAN176)	TIE2
	PF-562,271	gain SOX2 (cnaPANCAN246)	FAK

Table 4. Table illustrates the resistant anticancer drugs identified for breast cancer following screening by the ResisenseNet model, alongside their associated malignancy types, genomic markers, and targeted mechanisms.

and 5. A comparative analysis with state-of-the-art methods revealed that the ResisenseNet not only performs competitively but also surpasses the selected SOTA model in stability assessments, as evidenced by significantly lower standard deviation values, thereby affirming its potential as a reliable and advanced solution in drug sensitivity predictions. Please refer to Fig. 6 and supplementary table S9.

The ResisenseNet model demonstrated strong performance, achieving a test set loss and accuracy of 0.042 and 0.979, respectively, with validation set metrics of 0.069 for loss and 0.9575 for accuracy. please refer to Fig. 3 and Supplementary Table S7. Hyperparameter fine-tuning using Hyperopt further validated the model's predictive accuracy and generalizability. Additionally, our analysis identified genomic markers associated with drug sensitivity and resistance, detailed in Tables 2 and 4, representing another novel aspect of our research.

The results presented in Table 3 have identified three sets of drugs, with Set A comprising four existing drugs already annotated by the FDA as effective for cancer treatment, validating the model's accuracy and reliability in identifying known breast cancer therapeutics. Set B consists of 15 drugs supported by evidence from PubMed-indexed research (in vivo, in vitro, and clinical trials), but not listed in DrugBank or the FDA database, representing emerging candidates and promising leads that corroborate the model's predictive analysis. While not entirely novel, these drugs suggest opportunities for repurposing existing medications for breast cancer treatment. Notably, Set C encompasses 14 drugs, including idelalisib, daporinad, dacinostat, YK 4-279, zibotentan, pelitinib, FR-180,204, GSK690693, dabrafenib, linsitinib, PLX4720, SB590885, benzofurazon, and TPCA-1, which have not been extensively researched for breast adenocarcinoma specifically, positioning them as potential novel candidates for breast cancer treatment that warrant further exploration.

The identified novel drug molecules targeting various pathways hold great promise as anti-cancer agents for breast cancer treatment. idelalisib, a PI3Kdelta inhibitor, can effectively target the PI3K/Akt/mTOR pathway, which is frequently dysregulated in breast cancer and plays a crucial role in tumor growth and survival[24]. daporinad, a NAMPT inhibitor, disrupts NAD + biosynthesis, leading to energy depletion and apoptosis in cancer cells[25]. dacinostat, a histone deacetylase (HDAC) inhibitor, modulates gene expression and induces cell cycle arrest and apoptosis[26]. zibotentan, an endothelin A receptor antagonist, can inhibit tumor angiogenesis and metastasis. pelitinib, an EGFR inhibitor, targets the epidermal growth factor receptor, which is overexpressed in many breast cancers. FR-180,204, an ERK inhibitor, blocks the MAPK pathway[27], while GSK690693, an AKT inhibitor, inhibits the PI3K/Akt signaling axis. Dabrafenib, PLX4720, and SB590885 are BRAF inhibitors that target the BRAF oncogene, which is mutated in a subset of breast cancers[28]. linsitinib, an IGF1R inhibitor, disrupts insulin-like growth factor signaling, which is important for tumor growth and survival[29]. benzofurazon, an MDM4 inhibitor, restores p53 function, leading to apoptosis in p53-wild-type tumors. TPCA-1, an IKK inhibitor, blocks the NF-κB pathway, which is involved in inflammation and tumor progression[30]. These targeted therapies, combined with a better understanding of breast cancer biology, hold great promise for improving treatment outcomes and reducing the burden of this devastating disease.

The SHAP analysis reveals that certain molecular descriptors are crucial for the model's predictions, offering valuable insights into the drug design process. High SHAP values for specific descriptors indicate their significance for the model's accuracy and reliability, allowing researchers to concentrate on modifying these molecular properties to improve drug efficacy. Additionally, the analysis enhances the model's interpretability by pinpointing which features most influence predictions, fostering transparency and trust in the model's behavior and ensuring alignment with established scientific principles. In summary, the SHAP analysis highlights the importance of key molecular descriptors in drug prediction models while providing a clear framework for understanding model decisions, facilitating informed decision-making in drug design and development, and ultimately aiding in the creation of more effective therapeutic agents.

While the 1D-CNN + LSTM module of the ResisenseNet model effectively learns features from the provided text-based amino acid sequences, a notable limitation arises from its inability to capture dynamic changes in protein topology [31, 32]. This is crucial as proteins contain diverse catalytic sites essential for molecule binding and inhibitory activity [32, 33]. These intricate patterns and features elude detection by the 1D-CNN model, necessitating Supplementary simulation data, such as 4D-descriptors data. 3D molecular descriptors, such as COMFA and COMSIA, offer a solution by quantitatively representing molecules in a three-dimensional latent space, accounting for electrostatic and steric interactions between atoms [34, 35]. Additionally, incorporating diverse omics datasets, as demonstrated by Qiao Liu et al. 2020, could provide a comprehensive understanding of the complex interplay between various omics datasets [17] and drug responsiveness in terms of sensitivity and resistance.

AI and ML are revolutionizing drug discovery, particularly in repositioning and repurposing drugs. Traditional methods are time-consuming, underscoring the importance of ML-based prediction models in the early stages [36, 37]. ResisenseNet, a cutting-edge drug sensitivity prediction model, exemplifies this by repurposing anticancer drugs for breast cancer treatment. Its adaptability, user-friendliness, and capacity to identify ineffective molecules early in development highlight its significance. However, while prediction models are invaluable in early drug development, rigorous in vitro and in vivo studies are indispensable for validation. Tables 2 and 4 elucidate drugs and their associated genomic mutations, crucial factors in assessing effectiveness.

Drugs listed in set C of Table 3 exhibit significant potential for selection in breast cancer therapy. Subsequent in silico validations, including pharmacophore modeling, molecular docking studies, and molecular dynamics simulations, can elucidate the stability and binding affinity of these identified molecules with their respective target proteins. Following these computational analyses, in vitro studies, along with in vivo cell line investigations and experiments utilizing animal models, can substantially enhance the likelihood of discovering novel drug candidates aimed at targeting breast adenocarcinoma.

Methods

The methodology for developing the ResisenseNet model encompasses several critical phases, beginning with data retrieval and preprocessing, where relevant datasets were meticulously collected and refined prior to model training. This is followed by model construction, optimization, and validation, which involved conducting a series of ablation studies and comparative analyses to evaluate model performance metrics and optimize hyperparameters. Upon finalizing the model architecture and assessing its predictive capabilities, a drug repurposing study was executed. This study aimed to identify sensitive drugs by repurposing existing anticancer agents across 14 different malignancies specifically within the context of the breast cancer-targeted ResisenseNet model. The results of this analysis were systematically tabulated and presented, highlighting the potential for novel drug candidates among those identified as sensitive.

Data collection and preprocessing

Data collection

The study used data from the DoRothEA (Discriminant Regulon Expression Analysis) repository, which includes transcription factor activities and their downstream targets in humans. This repository provides profiles of transcription factor (TF) activity, genomic markers (mutations), and their effects on drug sensitivity and resistance. Please refer to Supplementary Tables S1, S2, S3, S4. SMILES IDs and amino acid sequences of TFs and targets were sourced from the PubChem and NCBI repositories Respectively. Drugs from the dataset were downloaded and subjected to preprocessing, which included converting molecules into 3D structures, removing salts, and calculating molecular descriptors using the Padelpy Library in Python v3.12.0. This process generated a comprehensive representation of the molecules, yielding 1,444 2D descriptors and 421 3D descriptors,

encompassing various physiochemical, electronic, topological, and quantitative structure-activity relationship (QSAR) characteristics. After preprocessing, the dataset consisted of 4,540 data points related to breast adenocarcinoma, including 2,800 sensitive and 1,740 resistant samples. It incorporated features from the DoRothEA dataset as well as 1,865 molecular descriptors of the drugs. The consolidated dataset also included expression profiles for 127 transcription factors, data from 62 distinct drugs, and 65 different genomic mutation expression profiles. The distribution of these components is depicted in Supplementary Fig. S1. The dataset was then split into training, testing, and validation sets in a 70:15:15 ratio, resulting in 3158, 681, and 681 data points for each set, respectively, utilizing the train_test_split module from scikit-learn.

Data preprocessing

The comprehensive dataset underwent extensive preprocessing to ensure data integrity and readiness for analysis, which included encoding, outlier handling through kurtosis and skewness analysis, null value imputation using the simple imputer module, and SMOTE analysis. Additionally, power transformation-based data standardization was conducted using the scikit-learn library in Python version 3.12.0, while sequence-based data was excluded from these procedures.

Model development for drug sensitivity classification

The proposed ResisenseNet drug sensitivity prediction model consists of two main components, each responsible for analyzing distinct types of data. The first component focuses on protein and transcription factor (TF) features, utilizing a 1D-Convolutional Neural Network (CNN) and LSTM (Long Short-Term memory) model architecture. In contrast, the second component deals with numerical data, including small molecule properties, TF activity profiles, and genomic marker profiles, employing a Deep Neural Network (DNN) framework. The model architecture was illustrated in Fig. 1.

Ablation studies were conducted to analyze the predictive capability of the ResisenseNet model to elucidate how escalating complexity enhances the model's predictive efficacy with individual modules. Several variants were considered: a model with only 1D-CNN, a model combining 1D-CNN and LSTM, a model with 1D-CNN and DNN, and a complex model incorporating 1D-CNN, LSTM, and DNN. Various validation metrics were used to compare and evaluate the models' performance. Each experiment incorporated 12 random seed trials to evaluate reproducibility, with mean validation metrics graphically depicted alongside standard deviation error bars. Based on this analysis, the final architecture of the ResisenseNet model was determined to advance the research. Please refer to Fig. 2.

Protein and TF feature learning with LSTM and 1D-CNN complex

The ResisenseNet model processes textual amino acid and SMILES sequences through tokenization, where each amino acid is represented as a distinct token. These tokens are then transformed into dense vector representations in an embedding layer, capturing semantic relationships and enabling rich learning. The resulting vectors are input into a one-dimensional convolutional neural network (1D-CNN), where convolutional filters extract relevant patterns and motifs for classification. The output is flattened into a one-dimensional vector and passed to a long short-term memory (LSTM) layer, which uses input, forget, and output gates to manage information flow, retain relevant data, and learn long-range dependencies in the sequences. This combination of 1D-CNN and LSTM allows the model to extract both local and long-range patterns from transcription factors and target proteins, moving beyond traditional predictive modeling that focuses primarily on small molecules and omics datasets.

Small molecule, TF, and genomic marker feature learning with DNN

The second component processes numerical data, including small molecule properties, transcription factor activity profiles, and genomic marker profiles, using a deep neural network (DNN) architecture that features multiple dense layers and dropout layers to prevent overfitting. The model is compiled with binary cross-entropy loss and the Stochastic Gradient Descent (SGD) optimizer, incorporating a specified learning rate and momentum. The preprocessed numerical dataset serves as input to the DNN, where multiple neurons introduce non-linearity into the model. The DNN learns patterns from the numerical data through backpropagation, a supervised learning technique that adjusts the weights of the connections in the network based on the difference between the forecasted value and the actual value. The learning rate monitored during training and early stopping applied to further mitigate overfitting. Hyperopt, a Bayesian optimization technique, is utilized for hyperparameter optimization. Hyperopt is an advanced hyperparameter optimization technique that utilizes Bayesian optimization to iteratively refine the search for optimal hyperparameters. It models the model's performance as a probabilistic function of the parameters, starting with an initial set of hyperparameters and updating its understanding of the search space based on the results. This approach allows Hyperopt to focus subsequent evaluations on areas likely to yield better performance, in contrast to grid search methods, which exhaustively test all combinations of specified hyperparameters. Optimized hyperparameters are presented in Table 1.

After processing, the learned vector representations from both the 1D-CNN + LSTM and the DNN are concatenated to form a single vector, which captures both the sequential and numerical aspects of the data. This combined feature vector is then passed through a fully connected layer that processes the features and learns to make predictions regarding resistance and sensitivity.

ResisenseNet model validations

The prediction performance was assessed using 12 random seeds to ensure reproducibility. The average values and standard deviations for validation metrics, including accuracy, precision, sensitivity, Matthews correlation

coefficient, and AUC-ROC curves, were both tabulated and graphically represented. Binary cross-entropy was used as the loss function, quantifying the differences between predicted and actual values to guide model optimization. Accuracy measures the proportion of correct predictions among all predictions made, while recall (sensitivity) reflects the model's ability to correctly identify positive instances. Please refer to Eqs. (1) and (2) [38]. The F1 score, a harmonic mean of precision and recall, provides a balance between these two metrics, with precision measuring the accuracy of positive predictions. Please refer to Eqs. (3) and (4) [38]. The Matthews correlation coefficient (MCC) offers a balanced assessment of all four confusion matrix categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Please refer to Eq. (5) [38]. Ultimately, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) assesses the model's capability to differentiate between classes at different threshold levels[38].

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{1}$$

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives} \tag{2}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

$$Precision = \frac{True\ positives}{True\ positives + False\ positives} \tag{4}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

The validation metrics mentioned above served as standards for a comprehensive evaluation of the model. Initially, the performance of the developed ResisenseNet model was determined using test and validation datasets. The results, including the respective validation metrics, are illustrated in Fig. 3, while the average values and standard deviations of these metrics are presented in Supplementary Table S7. To evaluate the model's performance in out-of-sample scenarios, datasets from colorectal adenocarcinoma (COREAD) and lung adenocarcinoma (LUAD) were utilized, ensuring that the drugs, molecular descriptors, targets, and transcription factors differed from those used during training. The validation metrics for these results are shown in Fig. 5. Additionally, a baseline logistic classifier model was constructed to compare the performance of the ResisenseNet model, as no similar models currently exist. The validation metrics for this comparison are displayed in Supplementary Fig. S6. A 10-fold cross-validation was additionally conducted on the developed ResisenseNet model to assess its generalizability and predictive capabilities.

To compare the developed ResisenseNet model with state-of-the-art (SOTA) methods, a comprehensive literature review was conducted, revealing that recurrent neural networks combined with character-level convolutional neural networks are particularly effective. Since the ResisenseNet model incorporates an LSTM module, we trained a character-level CNN and DNN complex model using the same dataset for performance evaluation and comparison. Experiments with both the SOTA model and the ResisenseNet model were conducted to compare validation metrics and assess the new model's potential. The resulting validation metrics are presented in Fig. 6 and Supplementary Table S9. SHAP (Shapely Additive exPlainations) analysis was performed to assess the interpretability of the developed ResisenseNet model. The corresponding bar pot was represented in Supplementary Fig. S7.

Employing the resisenseNet model for predicting drug sensitivity across diverse malignancies

After establishing a breast cancer-specific sensitivity prediction model ResisenseNet, we applied a drug repurposing strategy using FDA-approved drugs from various cancer types to predict their efficacy against breast cancer. Screening 14 different cancer types alongside their respective drugs, we identified distinct patterns of effectiveness using the developed model. Fig. 7 displays the count of drugs associated with each of the 14 malignancies, specifically indicating their resistance or sensitivity towards breast cancer, as analyzed by the ResisenseNet model. We also observed genomic markers with unique attributes that potentially influence drug sensitivity. Notably, some markers showed efficacy in one cancer type while exhibiting resistance in breast cancer. These findings are summarized in Tables 2 and 3.

After the repurposing process, sensitive drugs were evaluated for their previous involvement in research as anticancer agents for breast adenocarcinoma. To conduct this investigation, we reviewed literature from PubMed, Drug Bank, and the US FDA drug repository for relevant studies and annotations related to breast cancer. This effort aimed to identify novel drugs that have not yet been reported but were predicted by the ResisenseNet model to be sensitive for breast cancer. The results of this analysis are summarized in Table 3.

Conclusion

This study introduced ResisenseNet, a hybrid neural network framework for predicting drug sensitivity and resistance. By leveraging the underutilized DoRothEA dataset, we integrated 1D-CNN, LSTM, and DNN modules to capture long-range and temporal patterns in amino acid sequences and transcription factor expression

data. Rigorous validation demonstrated the model's predictive accuracy, generalizability, and adaptability. In drug repurposing studies, ResisenseNet, tailored for breast cancer, identified 14 novel drug candidates across 14 cancer types, previously untested for breast cancer. Future molecular dynamics, along with in vitro and in vivo validations, are expected to confirm these candidates' potential against breast adenocarcinoma.

Data availability

We sourced transcription factors and genomic mutation expression data from the DoRothEA (Discriminant Regulon Expression Analysis) database, available at https://dorothea.opentargets.io/#/. Amino acid sequences for targets and transcription factors were obtained from the NCBI (National Centre for Biotechnology Information), while SMILES IDs and molecular descriptors for drugs were retrieved from ChemSPYDER and the PADELpy library in Python v3.12.0, respectively. For molecular descriptor calculation, please refer to the GitHub link: https://github.com/ecrl/padelpy.

Received: 4 June 2024; Accepted: 23 August 2024
Published online: 14 October 2024

References

1. Ye, F. et al. Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. *Mol. Cancer*. https://doi.org/10.1186/s12943-023-01805-y (2023).
2. Li, Z. Y., Zhu, Y. X., Chen, J. R., Chang, X. & Xie, Z. Z. The role of KLF transcription factor in the regulation of cancer progression. *Biomed. Pharmacotherapy*. https://doi.org/10.1016/j.biopha.2023.114661 (2023).
3. Garg, M. et al. The pleiotropic role of transcription factor STAT3 in oncogenesis and its targeting through natural products for cancer prevention and therapy. *Med. Res. Rev.***41**, 1291–1336 (2021).
4. Dorna, D. & Paluszczak, J. Targeting cancer stem cells as a strategy for reducing chemotherapy resistance in head and neck cancers. *J. Cancer Res. Clin. Oncol.***149**, 13417–13435 (2023).
5. McAleese, C. E., Choudhury, C., Butcher, N. J. & Minchin, R. F. Hypoxia-mediated drug resistance in breast cancers. *Cancer Lett.***502**, 189–199 (2021).
6. Tan, Q. et al. Up-regulation of autophagy is a mechanism of resistance to chemotherapy and can be inhibited by pantoprazole to increase drug sensitivity. *Cancer Chemother. Pharmacol.***79**, 959–969 (2017).
7. Chen, C. et al. Autophagy and doxorubicin resistance in cancer. *Anticancer Drugs*. **29**, 1–9 (2018).
8. Jiramongkol, Y. & Lam, E. W. F. FOXO transcription factor family in cancer and metastasis. *Cancer Metastasis Rev.***39**, 681–709 (2020).
9. Calissi, G., Lam, E. W. F. & Link, W. Therapeutic strategies targeting FOXO transcription factors. *Nat. Rev. Drug Discov*. **20**, 21–38 (2021).
10. Beretta, G. L., Corno, C., Zaffaroni, N. & Perego, P. Role of FoxO proteins in cellular response to antitumor agents. *Cancers (Basel)*. https://doi.org/10.3390/cancers11010090 (2019).
11. Abdin, S. M., Tolba, M. F., Zaher, D. M. & Omar, H. A. Nuclear factor-κB signaling inhibitors revert multidrug-resistance in breast cancer cells. *Chem. Biol. Interact*.https://doi.org/10.1016/j.cbi.2021.109450 *(2021)*.
12. Luo, Q. et al. ARID1A prevents squamous cell carcinoma initiation and chemoresistance by antagonizing pRb/E2F1/c-Myc-mediated cancer stemness. *Cell. Death Differ.***27**, 1981–1997 (2020).
13. Panda, M., Tripathi, S. K. & Biswal, B. K. SOX9: An Emerging driving factor from cancer progression to drug resistance. *Biochim. Biophys. Acta Rev. Cancer*. https://doi.org/10.1016/j.bbcan.2021.188517 (2021).
14. Shao, K. et al. DTI-HETA: Prediction of drug–target interactions based on GCN and GAT on heterogeneous graph. *Brief. Bioinform*. https://doi.org/10.1093/bib/bbac109 (2022).
15. Rifaioglu, A. S. et al. DEEPScreen: High performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.***11**, 2531–2557 (2020).
16. Sun, C., Xuan, P., Zhang, T. & Ye, Y. Graph convolutional autoencoder and generative adversarial network-based method for predicting drug-target interactions. *IEEE/ACM Trans. Comput. Biol. Bioinform*. **19**, 455–464 (2022).
17. Liu, Q., Hu, Z., Jiang, R. & Zhou, M. DeepCDR: A hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics*. **36**, I911–I918 (2020).
18. Wang, Y., Yang, Y., Chen, S. & Wang, J. Deepdrk: A deep learning framework for drug repurposing through kernel-based multi-omics integration. *Brief. Bioinform*. https://doi.org/10.1093/bib/bbab048 (2021).
19. Cui, C. et al. Drug repurposing against breast cancer by integrating drug-exposure expression profiles and drug-drug links based on graph neural network. *Bioinformatics*. **37**, 2930–2937 (2021).
20. Ramesh, P., Karuppasamy, R. & Veerappapillai, S. Machine learning driven drug repurposing strategy for identification of potential RET inhibitors against non-small cell lung cancer. *Med. Oncol.* https://doi.org/10.1007/s12032-022-01924-4 (2023).
21. Lim, W-J., Kim, H. M., Oh, Y. & Pyo, J. Multiomics approach to understanding olaparib resistance and predicting drug response (2023). https://doi.org/10.1101/2023.04.04.535542
22. Triantafyllidis CP *et al.* A machine learning and directed network optimization approach to uncover TP53 regulatory patterns. *iScience* (2023). 10.1016/j.isci.2023.108291
23. Garcia-Alonso, L. et al. Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.***78**, 769–780 (2018).
24. Cerma, K. et al. Targeting PI3K/AKT/mTOR pathway in breast cancer: from biology to clinical challenges. *Biomedicines*. https://doi.org/10.3390/biomedicines11010109 (2023).
25. Mogol, A. N. *et al.* NAD + Metabolism generates a metabolic vulnerability in endocrine-resistant Metastatic breast tumors in females. Endocrinology (United States) (2023). https://doi.org/10.1210/endocr/bqad073
26. Mehmood, S. A. et al. Recent advancement of HDAC inhibitors against breast cancer. *Med. Oncol.*https://doi.org/10.1007/s12032-023-02058-x *(2023)*.
27. Li, X. et al. Bruceine A: Suppressing metastasis via MEK/ERK pathway and invoking mitochondrial apoptosis in triple-negative breast cancer. *Biomed. Pharmacotherapy*. https://doi.org/10.1016/j.biopha.2023.115784 (2023).
28. Katz, S. et al. A Topical BRAF inhibitor (LUT-014) for treatment of radiodermatitis among women with breast cancer. *JAAD Int.***15**, 62–68 (2024).
29. Chandran, S., Harmey, J., Toomey, S. & Harmey Phd, J. Inhibition of the IGF signalling pathway in MDA-MB-231 triple-negative Inhibition of the IGF signalling pathway in MDA-MB-231 triple-negative breast cancer cells breast cancer cells AUTHOR(S). (2012).
30. Pavitra, E. et al. The role of NF-κB in breast cancer initiation, growth, metastasis, and resistance to chemotherapy. *Biomed. Pharmacotherapy*. https://doi.org/10.1016/j.biopha.2023.114822 (2023).
31. Noé, F., De Fabritiis, G. & Clementi, C. Machine learning for protein folding and dynamics. (2019).

32. Audagnotto, M. et al. Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble. *Sci. Rep.*https://doi.org/10.1038/s41598-022-13714-z *(2022).*
33. Tsuchiya, Y. & Tomii, K. Neural networks for protein structure and function prediction and dynamic analysis. *Biophys. Rev.***12**, 569–573 (2020).
34. Kim, J. H. & Jeong, J. H. Structure-activity relationship studies based on 3D-QSAR CoMFA/CoMSIA for thieno-pyrimidine derivatives as triple negative breast cancer inhibitors. *Molecules*. https://doi.org/10.3390/molecules27227974 (2022).
35. El Rhabori, S. et al. Design, 3D-QSAR, Molecular docking, ADMET, molecular dynamics and MM-PBSA simulations for new anti-breast cancer agents. *Chem. Phys. Impact*. https://doi.org/10.1016/j.chphi.2023.100455 (2024).
36. Karampuri, A. & Perugu, S. A breast cancer-specific combinational QSAR model development using machine learning and deep learning approaches. *Front. Bioinf.*https://doi.org/10.3389/fbinf.2023.1328262 *(2023).*
37. Karampuri, A., Kundur, S. & Perugu, S. Exploratory drug discovery in breast cancer patients: A multimodal deep learning approach to identify novel drug candidates targeting RTK signaling. *Comput. Biol. Med.*https://doi.org/10.1016/j.compbiomed.2024.108433 *(2024).*
38. Jiao, Y. & Du, P. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant. Biology*. **4**, 320–330 (2016).
39. Chen, Y. et al. Nedd8-activating enzyme inhibitor MLN4924 (Pevonedistat), Inhibits miR-1303 to suppress human breast cancer cell proliferation via targeting p27Kip1. *Exp. Cell. Res.*https://doi.org/10.1016/j.yexcr.2020.112038 *(2020).*
40. Mina, L. et al. A phase II study of oral enzastaurin in patients with metastatic breast cancer previously treated with an anthracycline and a taxane containing regimen. *Invest. New. Drugs*. **27**, 565–570 (2009).
41. O'shea, J. et al. A preclinical evaluation of the MEK inhibitor refametinib in HER2-positive breast cancer cell lines including those with acquired resistance to trastuzumab or lapatinib. (2017).
42. Lim, B. et al. ONC201 and an MEK inhibitor trametinib synergistically inhibit the growth of triple-negative breast cancer cells. Biomedicines. (2021). https://doi.org/10.3390/biomedicines9101410
43. Hickish, T. *et al.* Afatinib alone and in combination with vinorelbine or paclitaxel, in patients with HER2-positive breast cancer who failed or progressed on Prior Trastuzumab and/or Lapatinib (LUX-Breast 2): An open-label, Multicenter, Phase II Trial. *Breast Cancer Res. Treat.* **192**, 593–602 (2022).
44. Roy, T. *et al.* A phase I trial of palbociclib and bosutinib with fulvestrant in patients with metastatic hormone receptor positive and HER2 negative (HR + HER2-) breast cancer refractory to an aromatase inhibitor and a CDK4/6 inhibitor. *Contemp. Clin. Trials Commun.* https://doi.org/10.1016/j.conctc.2023.101110 (2023).
45. Xu, J. *et al.* Phase II trial of veliparib and temozolomide in metastatic breast cancer patients with and without BRCA1/2 mutations. *Breast Cancer Res. Treat.* **189**, 641–651 (2021).
46. Lokesh Battula, V. *et al.* IKK inhibition by BMS-345541 suppresses breast tumorigenesis and metastases by targeting GD2 + cancer stem cells. *Oncotarget* **8**(23), 36936 (2017).
47. Mayer, E. L. *et al.* A Phase i dose-escalation study of the VEGFR inhibitor tivozanib hydrochloride with weekly paclitaxel in metastatic breast cancer. *Breast Cancer Res. Treat.* **140**, 331–339 (2013).
48. Kumar, N., Raza, M. & Sehrawat, S. Intuitive repositioning of an anti-depressant drug in combination with tivozanib: Precision medicine for breast cancer therapy. *Mol. Cell. Biochem.* **476**, 4177–4189 (2021).
49. Chen, L., Long, C., Nguyen, J., Kumar, D. & Lee, J. Discovering alkylamide derivatives of bexarotene as new therapeutic agents against triple-negative breast cancer. *Bioorg. Med. Chem. Lett.* **28**, 420–424 (2018).
50. Thomas, P. S. *et al.* Phase I dose escalation study of topical bexarotene in women at high risk for breast cancer. *Cancer Prev. Res.* **16**, 47–55 (2023).
51. Li, L. *et al.* Vandetanib (ZD6474) Induces antiangiogenesis through mTOR–HIF-1 alpha–VEGF signaling axis in breast cancer cells. *Onco Targets Ther.* **11**, 8543–8553 (2018).
52. Tam, S. et al. The ixabepilone and vandetanib combination shows synergistic activity in docetaxel-resistant MDA-MB-231 breast cancer cells. *Pharmacol. Rep.***74**, 998–1010 (2022).
53. Lee, E. Y. *et al.* Human breast cancer cells display different sensitivities to ABT-263 based on the level of survivin. *Toxicol. In Vitro* **46**, 229–236 (2018).
54. Lee, A. *et al.* Synergism of a novel MCL-1 downregulator, acriflavine, with navitoclax (ABT-263) in triplenegative breast cancer, lung adenocarcinoma and glioblastoma multiforme. *Int. J. Oncol.* https://doi.org/10.3892/ijo.2021.5292 (2022).

Acknowledgements

We express our gratitude to the National Institute of Technology, Warangal for their valuable support in our research work.

Author contributions

AK: Conceptualization; Methodology; Data curation; Validation; Visualization; Writing & review original draft. BJ: Conceptualization; Methodology; Writing & review original draftSP: Conceptualization; Formal analysis; Project administration; Resources; Software; Supervision; Review original draft.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-71076-0.

Correspondence and requests for materials should be addressed to S.P.