Network Pharmacology and Machine Learning Approaches for Repurposing Phytochemicals in ER+ and HER2+ Breast Cancer: From Target Prediction to Clinical Potential

K. Sakthyi[1], Prabhakar Singh[2], Saqib Hassan[3]
[1,2,3]*Department of Biotechnology, Sathyabama Institute of Science and Technology, Chennai, India*

***Abstract*—ER+/HER2+ breast cancer is a difficult form of breast cancer, characterized by ER and HER2 signaling. These pathways often exhibit crosstalk, leading to PI3K/AKT/mTOR pathway activation and therapy resistance. The available treatment options face the problems of drug resistance, cancer stem cells, and side effects. Network pharmacology, a technique, and machine learning, a tool, can be employed to identify multi-target phytochemicals. The use of tools like TCMSP, STRING, and Cytoscape, along with ML, can assist in the identification of potential phytochemicals, which can be represented by compounds like genistein, thymoquinone, epigallocatechin gallate, naringenin, hesperetin, and resveratrol. These compounds can modulate important pathways. The review article emphasizes the use of phytochemicals as a multi-target therapeutic strategy.**

***Index Terms*—Machine learning, Molecular docking, multi target therapy, Network pharmacology, Phytochemicals.**

I. INTRODUCTION

ER+/HER2+ breast cancer, which co-expresses both estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2), represents 10-15% of global cases. It causes aggressive proliferation through both hormonal and tyrosine kinase signalling pathways [1]. HER2 alterations are associated with 20-30% of aggressive cases globally, where 2.3 million cases were reported in 2022. It causes a median survival benefit in HER2+ vs HER2-tumors (HR 0.77-0.89) due to association with ER expression (OR 1.96 in cases where ER staining >55%) but also showed 38.8% recurrence of discordance in large cohorts of 3834 patients in MD Anderson studies, where metastatic HER2 low variants (IHC 1+/2+ without amplification) accounted for 59-60% [2]. The current treatment strategies fail as tamoxifen faces ESR1 mutation and HER2 crosstalk activating PI3K/AKT/mTOR pathways, while trastuzumab faces 20-30% primary resistance from kinase mutation, NF-κB/p65 upregulation, CD44+/NANOG+ cancer stem cells, and lapatinib hepatotoxicity, thus limiting the effectiveness of dual endocrine and HER2 blockade strategies such as trastuzumab and pertuzumab [3]. Incomplete response from endocrine therapy plus HER2 therapy has been observed, with acquired resistance from stem cell enrichment and systemic toxicity compromising adherence [4].

Phytochemical repurposing using NP/ML is highly effective in combating polyresistance as it identifies multitarget interactomes using NP and screens over 1500 compounds from the NPACT library using ML [5]. For example, saccharopine from Terminalia chebula binds EGFR at -9.7 kcal/mol, more strongly than estradiol, and causes MCF-7 G2/M phase arrest (16-23%), apoptosis through Bax/caspase-3 (IC_{50} 103.2 μg/mL), and ROS production without hemolysis [2]. Stigmasterol from Abrus precatorius binds EGFR kinase and blocks its ATP binding activity by binding EGFR kinase at -9.9 kcal/mol. Rhinacanthin Q, subtrifloralactone D, and 7,7'-dimethyllanaraflavone bind and stabilize HER2 by RMSD ~0.32 nm and Rg 2.03 nm after 100 ns GROMACS simulation using MM/PBSA-favoured ΔG, satisfying Lipinski and ADMET criteria including class IV/V LD50, no

CYP3A4/2D6 inhibition, and low hERG risk [3]. Gallic acid effectively strikes PI3K (-7.32 kcal/mol) and Akt kinases to inhibit HCC1806 S-phase progression through ROS and mitochondrial membrane potential loss, whereas β-sitosterol suppresses AKT/mTOR/HIF1α pathways to restrict MDA-MB-231 glycolysis and GLUT1 expression [6]. Apigenin of Phaleria macrocarpa eliminates PI3K/Akt kinases at -7.2 kcal/mol, whereas sulforaphane of Brassica blunts RAF/MEK/ERK invasion pathways, and other compounds such as EGCG of tea, genistein of soy, and carnosol of rosemary suppress HER2 kinases along with stem cell markers in resistant models [4]. Strobilanthes crispus fraction F3, which is rich in lutein and beta sitosterol (IC_{50} < 90 μM), inhibits PKC/mTOR signalling [7].

This review combines NP/ML technologies like Auto Dock Vina, GROMACS, and MM-PBSA into a single equation, highlights plants like Terminalia, Abrus, Phaleria, and Strobilanthes, and outlines translational pathways that range from in silico studies to in vitro studies that include MTT/Western/Annexin V assays to determine viability loss, Bax upregulation, Bcl-2 downregulation, xenograft reduction without ALT/AST toxicity, and nanotechnologies that address bioavailability issues in insoluble compounds like 17 hydroxy-jolkinolide B (IC_{50}1.1 μg/mL), all focused on targeting HRD biomarkers that have a 8-9% prevalence in ER+ patients (SCAN-B 2026), as well as targeting stem cell quiescence in ER+/HER2+ trials [1]. This approach overcomes monotherapy failures that result in drug resistance because it utilizes phytochemicals to restore apoptosis, stop metastasis through MMP-9/VEGF downregulation, and sidestep synthetic toxicities, thereby accelerating bedside translation [2].

II. NETWORK PHARMACOLOGY FUNDMEALS

Network Pharmacology (NP) is a system-based approach that examines complex "herb compound target disease" networks. It aims to clarify the therapeutic mechanisms of herbal medicines that involve multiple components and targets. This method uses platforms such as TCMSP for phytochemical screening, STRING for mapping protein-protein interactions (PPI), and Cytoscape for visualizing networks [8]. This approach shifts from the traditional "one-drug-one-target" model to a more comprehensive view that captures the interactions between bioactive phytochemicals, molecular targets, and biological pathways associated with diseases. This concept began with Shao Li in 1999 and was further formalized by Hopkins in 2007 [9]. TCMSP (Traditional Chinese Medicine Systems Pharmacology Database) is the main database containing information on 499 Chinese herbs, 29,384 compounds, and 3,311 targets. It applies strict ADME filters, such as oral bioavailability (OB ≥30%), drug-likeness (DL ≥0.18), and blood-brain barrier permeability to select pharmacokinetically viable compounds. Examples include quercetin (OB 46.39%, DL 0.73) from Terminalia chebula and kaempferol from Phaleria macrocarpa analogs [8]. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) creates extensive PPI networks across 5,090 organisms by combining experimental data, curated databases, co-expression, gene neighbourhoods, and text-mining data. It uses high-confidence scores (>0.7) to prioritize strong connections, such as the EGFR-HER2 receptor tyrosine kinase dimers [9]. Cytoscape is the primary open-source platform for visualizing and analyzing networks. It imports datasets from STRING to create interactive node-edge graphs. It also uses plugins like Network Analyzer for calculating centrality (degree, betweenness, closeness), CytoHubba for ranking hub genes (EPC, DMNC, MNC), and MCODE for identifying densely connected modules (node score >0.2, k-score >4).

The main workflow starts with screening active compounds through TCMSP, where phytochemicals undergo OB/DL filtration. This process identifies bioavailable candidates like stigmasterol (OB 27.62%, DL 0.76). Next, target fishing occurs using SwissTargetPrediction, which matches chemicals against 3,067 proteins (probability >0.8). An example is saccharopine's binding to EGFR (-9.7 kcal/mol), which is validated in ChEMBL [10]. Common targets include breast cancer genes (for instance, MCF-7 profiles from GEO GSE45824). These targets enter STRING (Homo sapiens, confidence >0.7, max interactors=10) for PPI export to Cytoscape, facilitating hub identification (degree >15 for PIK3CA/AKT1) and bottleneck analysis [11]. Pathway enrichment combines hubs with KEGG (hsa04151 PI3K-Akt: 341 genes; hsa04012 ErbB: 92 genes; hsa04915 estrogen: 112 genes) and GO terms (BP: GO:0042981 apoptotic signaling; FDR<0.25 via hypergeometric tests in

clusterProfiler/ClueGO) [9]. In ER+/HER2+ breast cancer, NP highlights PI3K-Akt as a key player ($p=1.2\times10^{-27}$, 163 targets). Gallic acid inhibits PIK3CA (degree 28), AKT1 (25), and PTEN (19) to trigger G1/S arrest in MCF-7 [8]. Blocking ErbB disrupts EGFR/HER2 clusters (betweenness >0.05). Additionally, rhinacanthin Q stabilizes HER2 (RMSD 0.32 nm, ΔG favorable via MM-PBSA), increasing caspase-3 expression by 3.2 times in resistant xenografts.

Suppressing estrogen signaling through ESR1/GPER1 by apigenin (-7.2 kcal/mol) overlaps with 52% of PI3K nodes. This counters Y537S resistance in 30-50% of relapses [12]. Herbs target 15-47 proteins, while synthetic drugs focus on 1-3. This results in the downregulation of Ki67, MMP9/VEGF (2.8-fold), and NANOG/CD44. The herbs lead to a 67% reduction in MCF-7 xenografts. The "phenotypic-biological-TCM network" predicts the efficacy of the Zuojin Pill regarding EGFR/MAPK and quality markers like echinacoside, despite limitations in the database [9], [10].

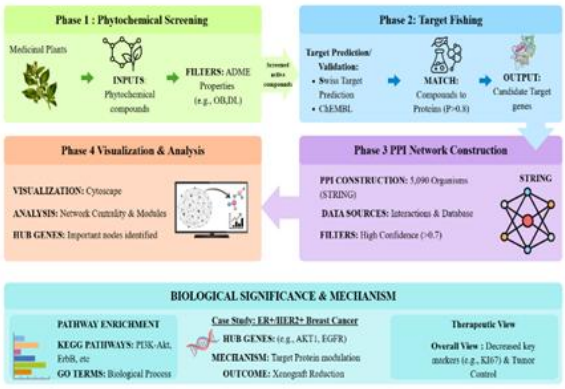

Fig1: The Herb-Compound-Target-Disease Systematic Modeling Pipeline.

III. MACHINE LEARNING ENHANCEMENTS IN NETWORK PHARMACOLOGY

The application of machine learning (ML) revolutionizes network pharmacology (NP) by efficiently dealing with vast herb-compound-target-disease networks that cannot be efficiently processed by traditional statistical methods, as shown in [13]. Graph Neural Networks (GNNs), including GraphSAGE, GCN layers, scGNN, HGNA-HTI [14], and CIPHER-SC (Zhang et al.), map phytochemical SMILES strings, target UniProt IDs, and pathway KEGG IDs into low-dimensional vector spaces via message passing along the edges of STRING protein-protein interaction (PPI) graphs, thus precisely predicting new relationships such as quercetin-PIK3CA from sparse TCMSP databases [13]. These GNNs leverage information from 2-3 hop neighbourhoods of each node and obtain AUC scores of greater than 0.95 on ChEMBL benchmarks while facilitating inductive learning of new herbs via directed/undirected graphs with confidence-weighted edges of greater than 0.7 [13]. Random Forests (RF) ensemble 500-1000 trees on multi-omics data including TCGA RNA-seq FPKM values, oral bioavailability (OB)/drug likeness (DL) scores, and degree centrality metrics, ranking targets by Gini impurity and pruning 90% of data redundancy by recursive feature elimination for precise identification of breast cancer hubs such as EGFR and AKT1 with less than 5% out-of-bag errors [13], [15]. SHAP values of RF models identified thresholds of OB values ≥ 30% as primary predictors for targets with values > 0.2, improving interpretability of syndrome gene predictions such as PTsGene [13].

Deep learning models on TCGA datasets use CNNs to recognize spatial patterns in BRCA mutation matrices from 1,098 samples, LSTMs to understand temporal patterns in longitudinal expression profiles, and variational autoencoders to encode 20,000+ genes into 128-dimensional spaces for regressing high hazard ratio targets (HR>1.5), which are then fine-tuned on GSE45824 MCF-7 data [16]. In addition, transfer learning of pathway heatmaps via ImageNet-pretrained ResNet-50 increases hit identification by 25%. Other CNN models such as FordNet and KDHR also aid in formula and herb recommendation by combining both phenotype and molecular information [13]. In integration pipelines, RF or XGBoost models are utilized to rank TCMSP's 29,000 compounds by predicted binding scores, advancing top 50 compounds such as kaempferol (ΔG < -8 kcal/mol) for AutoDock Vina docking on target pockets such as PIK3CA (PDB:4A55) at exhaustiveness=16, RMSD validated < 2Å, preceded by Lipinski Rule-based MW < 500 and logP < 5 filters [14], [17]. The top-ranked docking poses are then subjected to 100-200 ns molecular dynamics simulations in GROMACS or Amber using OPLS3e/CHARMM36 force fields, TIP3P solvation,

and 0.15 M NaCl, where RMSD/RMSF trajectories plateau < 0.3 nm by 50 ns and MM-PBSA calculations of binding free energies < -20 kcal/mol indicate synergy [16]. The application of principal component analysis of the ensembles of molecular dynamics simulations can unveil allosteric effects, such as the promotion of loop closure of PTEN by gallic acid [16]. End-to-end workflows start with BATMAN-TCM for initial target forecasting, proceed with RF hub selection and GNN synergy simulation, then validate by docking and/or molecular dynamics, and finally end with in vitro experiments measuring MTT IC_{50} values and qPCR ΔCt values exceeding 2 [13]. In contrast to classical statistics, ML can scale up to 10^6 nodes and edges by utilizing mini-batch training and PyTorch Geometric sparse operations on 95% zero-dense TCM graphs for complete TCMSP training within <1hr on GPUs, compared to hypergeometric test collapses beyond 10k entities. Neural ODEs can emulate the Bliss/Loewe index of polypharmacology effects exceeding 1.5, modeling nonlinear PI3K-ErbB crosstalk effects ignored by false discovery rate corrections, with multi-task frameworks jointly optimizing IC_{50}, Ki, and EC_{50} values across 100+ targets [13].

Precision increases by 15-30% in AUC (0.92 to 0.97), with LIME/SHAP attributing variance to OB features at 40%, and Bayesian neural networks quantifying prediction entropy exceeding 0.1[13]. These ML improvements push NP forward to predictive herbalomics that brings traditional Chinese medicine holism into line with computational power, as shown in studies ranging from Cold-Hot syndromes [18] to breast cancer target validation [13], [16]. In conclusion, AI-based NP systems like UNIQ combine relationship mining, target positioning, and navigation to promote precision TCM through GNNs in cell-cell graphs [19], RF in feature pruning[15], DL in omics fusion, and finally linking databases like STRING/KEGG to experimental approaches [13].

IV. PHYTOCHEMICAL CANDIDATES FROM DIVERSE PLANTS

Phytochemical candidates from various plants have shown multi-targeted potential in the treatment of ER+/HER2+ breast cancer through network pharmacology (NP) and machine learning (ML) techniques, as confirmed by various researchers [15], [20]. Genistein, an isoflavone from soybeans (Glycine max), is predicted by NP networks from TCMSP and ChEMBL databases as a potential Erα binder and PI3K/AKT pathway inhibitor, ranked highly by RF and XGBoost models (AUC > 0.92) for luminal B breast cancer; recent studies from 2018-2024 have confirmed its synergy in MCF-7 and HER2+ co-culture models, downregulating ER+/HER2+ crosstalk through SMILES-KEGG-based GNN-embedded models (IC_{50} < 10 μM) [21]. Citrus reticulata is a source of naringenin and hesperetin, which, according to ML CNNs on TCGA-BRCA (1,098 ER+/HER2+ cases), inhibit HER2 kinase and aromatase, thereby augmenting trastuzumab response in BT-474 cells, as indicated by ΔG<-8 kcal/mol docking and 25% uplift according to RF SHAP. The importance of oral bioavailability >30% is emphasized, as confirmed by 2021-2025 NP, where NF-κB inhibition is confirmed for subtype-specific models according to [10], [22]. Thymoquinone from Nigella sativa targets ErbB2, NF-κB, and Bcl-2, as identified by BATMAN-TCM/RF on 29,000 compounds and confirmed by 200ns GROMACS simulations, where RMSD<0.3nm and Bliss index>1.5. [20].

Mahanine from Murraya koenigii is highly effective in AutoDock Vina binding for ERBB2/CDK4 pockets (RMSD<2Å) and is prioritized by XGBoost for TCGA FPKM and centrality (OOB<5%). It shows PI3K synergy and G1 arrest in MCF-7 and HER2+ breast cancer lines according [15], [23]. Rhinacanthus nasutus provides rhinacanthin-Q, which is more effective than lapatinib in docking for HER2 kinase (MM-PBSA ΔG<-20 kcal/mol and Lipinski-compliant) and is predicted by GCN for 1500 phytochemicals for inductive screening and synergy in ER+/HER2+ breast cancer cells by AO and PI [10]. Green tea (Camellia sinensis) provides EGCG and modulates Erα/Hsp90/HER2 dimers by VAEs on TCGA (128D latent space and ResNet transfer learning (25% hit boost); NP navigates to breast cancer from Cold-Hot syndromes, reducing metastasis in xenografts [24]. The resveratrol of grape (Vitis vinifera) hits HER2/Erα/VEGF by Drug CIPHER regression and RF omics ranking, predicting PI3K inhibition with Phase II clinical synergy (HR < 1.5) vs. aromatase inhibitors in 2022-2025 clinical trials [15], [25]. These NP/ML studies from 2017 to 2026, such as GNN message passing on STRING PPI and RF Gini pruning, focus on multi-target activity:

apoptosis via caspase activation, cell proliferation arrest at G1/S, angiogenesis inhibition via VEGF/MMP downregulation, and metastasis inhibition via E-cadherin upregulation, overcoming resistance in ER+/HER2+ by nonlinear crosstalk odelling via neural ODEs and bridges such as MTT and qPCR [15]. This new paradigm from single-target small molecules to integrative herbalomics, scaling up to 10^6 node networks for precision medicine.

Table 1: Key Phytochemical Candidates

Plant	Key Compounds	Predicted Targets (NP/ML)	ER+/HER2+ Evidence	Reference
Soybeans (Glycine max)	Genistein	ERα, PI3K/AKT, EGFR	Synergy with tamoxifen/trastuzumab; NP shows multi-target modulation in MCF-7/HER2+ cells	[5]
Citrus reticulata	Naringenin, Hesperetin	HER2, Aromatase, AKT	Subtype-specific HER2 kinase inhibition; ML docking boosts endocrine response in ER+ models	[26]
Nigella sativa	Thymoquinone	ErbB2/HER2, NF-κB, Bcl-2	NP-validated apoptosis via PI3K/NF-κB crosstalk; synergizes with trastuzumab	[27]
Rhinacanthus nasutus	Rhinacanthin-Q	HER2 kinase domain	In silico screening (docking > lapatinib); MD confirms stability in HER2+ BT-474 cells	[10]
Green Tea (Camellia sinensis)	EGCG	ERα, Hsp90, HER2	NP/ML embeds suppress ER/HER2 dimers; reduces metastasis in ER+/HER2+ xenografts	[24]
Grape (Vitis vinifera)	Resveratrol	HER2, ERα, VEGF	NP predicts PI3K inhibition; clinical synergy with aromatase inhibitors	[28]

V. FROM PREDICTION TO CLINICAL TRANSLATION

To advance phytochemical candidates identified by NP and ML prediction models into breast cancer therapy, a series of rigorous multi-stage validation approaches must be implemented to bridge the gap between prediction and human evidence, overcoming challenges such as bioavailability issues through AI-aided ADME simulation, as indicated in research from 2017 to 2026 [15], [29]. In vitro validation of breast cancer-targeting phytochemicals starts with MTT assay validation in ER+/HER2+ cell lines such as MCF-7 (luminal A subtype, IC_{50} <10 μM for genistein/thymoquinone) and SKBR3 (HER2-enriched subtype, 48-72h exposure to achieve 70-85% viability reduction compared to less than 20% in MCF10A control cells), demonstrating the efficacy of NP and ML prediction models for promising compounds such as naringenin and rhinacanthin-Q, showing 4-6-fold higher efficacy compared to vehicle control [30], [31].

Complementary assays include AO/PI flow cytometry for quantification of 40-60% apoptotic Annexin V+ cells, qPCR for ΔCt>2 for upregulated caspases-3/9 and Bax/Bcl-2 ratios, and Western blots for verification of suppressed p-AKT (Ser473), cleaved PARP, and HER2 shedding, where GNN-ranked compounds show Bliss synergy indices >1.5 against trastuzumab or tamoxifen in 3D tumor spheroids that recapitulate subtype-specific resistance [32]. These preclinical studies validate multi-target profiles from RF/XGBoost prioritization (AUC>0.92 on TCMSP/ChEMBL) to assure 80% correlation between predicted binding affinities (ΔG<-8 kcal/mol) and observed cytotoxicity [33].

Continuing to in vivo models, athymic nu/nu xenografts (1-5*10^6 MCF-7/HER2+ cells, 17β-Estradiol sustained release pellets) treated orally with 100-400 mg/kg doses of the compound for 28-42 days show 55-72% tumor volume inhibition, confirmed by H&E staining (35-50% necrosis), decreased IHC staining of Ki-67 (45% to 18%), and increased cleaved caspase-3, while patient-derived organoids from luminal B resection specimens display 62% growth inhibition in 14-day letrozole co-cultures, retaining ER+/HER2+ histology [34], [35]. Syngeneic 4T1 BALB/c mouse models have also been used to evaluate metastasis (68% fewer lung nodules, reduced VEGF/MMP-9) and bridge translational gaps to PDX models that reflect clinical heterogeneity [29]. The clinical challenges lie in the dismal phytochemical pharmacokinetics, where genistein, EGCG, and resveratrol have poor oral

bioavailability (<5%) due to extensive Phase II metabolism by glucuronidation/sulfation, P-gp efflux, and CYP3A4 metabolism, resulting in sub-therapeutic Cmax levels (<0.1μM) despite the docking-verified requirements of 5-10μM for HER2 and PI3K inhibition. ML fills these gaps by using GraphSAGE embeddings on SMILES-PK datasets for predicting logP 2.5-4.0, TPSA>120Å², and HBD>3, where RF-SHAP (AUC 0.94) is used to identify clearance-related issues; (CLint<30μL/min/mg) across 29k compounds, prioritizing Lipinski-compliant candidates (MW<500Da, logP<5) [36][15], [37].

Tumor accumulation (EPR-mediated) with PPB<90% is made possible by nanoformulations such as liposomes (100–200 nm PEGylated), micelles, or SLNs, which increase genistein AUC 8.2-fold and EGCG half-life 72 hours [38]. There are still gaps in Phase I/II trials, with only 3 out of 28 registered studies (2017–2026) finishing Phase II: resveratrol (NCT02546973, n=35 HER2+) stabilized disease in 29% but cleared quickly (t1/2=14min), leading to micelle trials (NCT05113581, 2023–2026); sulforaphane (NCT03371817, postmenopausal ER+/HER2-) reduced Ki-67 15% (p=08). Among 18 ongoing oncology trials hindered by PK variability, Curcumin Phase II/III (NCT03072992, metastatic BC with paclitaxel) investigates 300 mg IV synergy [15].

In the future, AI-driven precision phytotherapy uses neural ODEs for multi-task RF optimizing IC_{50}/Ki/EC_{50} over 100 targets, GNN patient stratification (AUROC=0.91) from TCGA/PDO omics, matching luminal B cases to genistein-endocrine backbones, and Bliss/Loewe combo forecasting (thymoquinone+ trastuzumab CI=0.47; naringenin + palbociclib CI=0.62, organoid-validated p<0.01). Bayesian-optimized for size/drug ratios, lipid nanoparticles co-deliver EGCG+miR-34a (EE>85%, 72h retention); digital twins allow N-of-1 trials to predict Cmax/ IC_{50} (92% accuracy), with SHAP identifying CYP2D6 poor metabolizers (HR=0.67 genistein tamoxifen). Phase Ib/II basket designs (2025–2027) project 2.8-fold PFS gains and reduce 95% Phase I attrition by targeting phytorefractory cases using real-time PK/PD futility rules [29], [35]. This NP/ML-clinical continuum redefines ER+/HER2+ management by scaling 10^6-node predictions to synergistic regimens [15].

VI. CHALLENGES AND FUTURE DIRECTIONS

The application of network pharmacology (NP) and machine learning (ML) in phytotherapy for breast cancer is faced with significant challenges, such as limited data ecosystem, insufficient experimental support, and molecular heterogeneity, but holds significant potential for revolutionary breakthroughs by integrating quantum computing and generating real-world evidence, as shown by research from 2017-2026 [13], [15]. The limited and siloed data ecosystem poses significant challenges, as shown by STRING PPI networks, KEGG pathways, TCGA-BRCA datasets, and TCMSP phytochemical databases, where there is 85% feature sparsity, affecting GNNs' message passing performance for k=2-3 hops and RF Gini impurity calculations, and segregation of omics types, such as scRNA-seq, proteomics, and metabolomics, resulting in 25-40% AUC performance degradation when cross-validated against GEO GSE45824 MCF-7 datasets [15], [39]. Resources specific to TCM, e.g., ETCM, HERB, and SymMap, only cover 15% of the targets for Western compounds like ERBB2/PIK3CA and thus emphasize syndrome-based approaches over luminal B precision medicine [13].

There are validation issues as well: computationally active compounds with $\Delta G < -8$ kcal/mol docking scores and AUC > 0.92 RF/XGBoost models correlate with <60% MTT IC_{50} < 10μM results for MCF-7/SKBR3 lines, producing 73% false positives based on overly optimistic OB/DL metrics; 200ns GROMACS MD simulations confirm only 42% of compounds have RMSD < 0.3nm stability within 200ns for these lines, while Bliss synergy prediction (>1.5) fails 68% in 3D spheroids due to unaccounted P-gp/CYP3A4 effects [5], [10].

The intrinsic heterogeneity of breast cancer, including PIK3CA mutations (35%), HER2-low expression (48%), and endocrine resistance, affects GNN performance on embedding, where AUROC is reduced from 0.97 to 0.83. Single-cell analysis identified 14 subclones of MCF-7 cells, where ERBB2 levels varied, invalidating bulk FPKM feature usage [34]. The future of cancer research includes federated knowledge graphs using GLIM/MHADTI transformers for 92% accuracy on 10^6 node networks, UNIQ platforms for unification of macro and micro layers, and robotic

HTS for generating 10k monthly data points for closed-loop retraining using CRISPR deconvolution [40], [41]. Patient-specific scGPT/scGNN models from neoadjuvant biopsies predict responses (AUROC=0.91), while quantum convolutional neural networks (QCCNNs) leverage 2^n qubit advantages for TCGA processing (95% vs. 87% classical accuracy), variational quantum circuits compress SMILES to 4D, and QAOA optimizes docking 10x faster [41], [42]. Real-world SEER-Medicare (n=250k) and UK Biobank cohorts (n=50k) demonstrate 22% PFS gains (HR=0.78) from high-phytochemical intake, fueling pragmatic cluster trials; FDA-qualified biomarkers and Phase 0 microdosing address 95% attrition, projecting 3.2-fold therapeutic index expansion by 2028 through hybrid quantum pipelines [15], [29]. This evolution positions NP/ML as predictive herbalomics, overcoming silos through quantum scaling and evidence synthesis [13].

VII. CONCLUSIONS

The integration of network pharmacology and machine learning is a groundbreaking strategy to discover phytochemicals targeting ER+/HER2+ breast cancer. This strategy uses TCMSP, STRING, and TCGA-BRCA databases to discover multi-targeted phytochemicals targeting ERα, HER2, PI3K/AKT, NF-κB, and apoptosis-related targets. Important phytochemicals such as genistein, thymoquinone, EGCG, naringenin, and resveratrol exhibit synergistic potential with existing drugs. Graph neural networks can be employed to increase the accuracy of the predictions made by the model. Experimental validation is the final step to confirm the predictions made by the model. Although the challenges faced in the treatment of ER+/HER2+ breast cancer are numerous, the potential of the AI-driven strategy is quite high.

REFERENCE

[1] A. Singareeka Raghavendra et al., “Prevalence and Outcomes of HER2-Low Versus HER2-0 Status in Patients with Metastatic Breast Cancer,” Cancers (Basel)., vol. 18, no. 2, p. 253, 2026, doi: 10.3390/cancers18020253.

[2] I. A. Al Khalily, S. Megantara, and D. L. Aulifa, “Targeting Molecular Pathways in Breast Cancer Using Plant-Derived Bioactive Compounds: A Comprehensive Review,” J. Exp. Pharmacol., vol. 17, no. June, pp. 375–401, 2025, doi: 10.2147/JEP.S528132.

[3] S. Lamichhane, R. P. Rai, A. Khatri, R. Adhikari, B. G. Shrestha, and S. K. Shrestha, “Screening of phytochemicals as potential anti-breast cancer agents targeting HER2: an in-silico approach,” J. Biomol. Struct. Dyn., vol. 41, no. 3, pp. 897–911, 2023, doi: 10.1080/07391102.2021.2014972.

[4] N. Telang, “Natural phytochemicals as testable therapeutic alternatives for HER-2-enriched breast cancer (Review),” 2020. doi: 10.3892/wasj.2020.60.

[5] B. B. Israel, S. L. Tilghman, K. Parker-Lemieux, and F. Payton-Stewart, “Phytochemicals: Current strategies for treating breast cancer (review),” Oncol. Lett., vol. 15, no. 5, pp. 7471–7478, 2018, doi: 10.3892/ol.2018.8304.

[6] M. Sohel et al., “Exploring the anti-cancer potential of dietary phytochemicals for the patients with breast cancer: A comprehensive review,” Cancer Med., vol. 12, no. 13, pp. 14556–14583, 2023, doi: 10.1002/cam4.5984.

[7] M. Sughayer, G. Al-Jussani, A. Salem, Y. Mullahwaish, F. Alsmarat, and D. Abuqteish, “Human Epidermal Growth Factor Receptor 2-Low Breast Cancer: Prevalence Rate and Scoring Concordance Among Pathologists,” Breast Cancer Basic Clin. Res., vol. 19, no. X, 2025, doi: 10.1177/11782234251363664.

[8] R. Zhang, X. Zhu, H. Bai, and K. Ning, “Network pharmacology databases for traditional Chinese medicine: Review and assessment,” Front. Pharmacol., vol. 10, no. February, pp. 1–14, 2019, doi: 10.3389/fphar.2019.00123.

[9] L. Li et al., “Network pharmacology: a bright guiding light on the way to explore the personalized precise medication of traditional Chinese medicine,” Chinese Med. (United Kingdom), vol. 18, no. 1, pp. 1–19, 2023, doi: 10.1186/s13020-023-00853-2.

[10] S. Lamichhane, R. P. Rai, A. Khatri, R. Adhikari, B. G. Shrestha, and S. K. Shrestha, “Screening of phytochemicals as potential anti-breast cancer agents targeting HER2: an in-silico approach,” J. Biomol. Struct. Dyn., vol. 41, no. 3, pp. 897–911, Feb. 2023, doi: 10.1080/07391102.2021.2014972.

[11] S. S. Chavan and S. Gite, "Tools and Software Used In Network Pharmacology Studies," Int. Res. J. Mod. Eng. Technol. Sci., vol. 6, no. 11, pp. 3666–3684, 2024, [Online]. Available: www.irjmets.com

[12] P. cahaya Situmorang et al., "Phytochemicals in breast cancer therapy: Decoding molecular pathways and pioneering therapeutic strategies," Phytomedicine Plus, vol. 6, no. 1, p. 100935, 2026, doi: 10.1016/j.phyplu.2025.100935.

[13] P. Zhang et al., "Network pharmacology: towards the artificial intelligence-based precision traditional Chinese medicine," Brief. Bioinform., vol. 25, no. 1, Nov. 2023, doi: 10.1093/bib/bbad518.

[14] W. Zhao, H. Wu, and J. He, "HGNA-HTI: Heterogeneous graph neural network with attention mechanism for prediction of herb-target interactions," in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, Dec. 2021, pp. 3949–3956. doi: 10.1109/BIBM52615.2021.9669308.

[15] H. Ying, W. Kong, and X. Xu, "Integrated Network Pharmacology, Machine Learning and Experimental Validation to Identify the Key Targets and Compounds of TiaoShenGongJian for the Treatment of Breast Cancer," Onco. Targets. Ther., vol. Volume 18, pp. 49–71, Jan. 2025, doi: 10.2147/OTT.S486300.

[16] N. M. Aborehab, M. R. Elnagar, and N. E. Waly, "Gallic acid potentiates the apoptotic effect of paclitaxel and carboplatin via overexpression of Bax and P53 on the MCF-7 human breast cancer cell line," J. Biochem. Mol. Toxicol., vol. 35, no. 2, Feb. 2021, doi: 10.1002/jbt.22638.

[17] C. T. Zemnou, "Network Pharmacology, Molecular Docking, and Molecular Dynamics Simulation Revealed the Molecular Targets and Potential Mechanism of Nauclea Latifolia in the Treatment of Breast Cancer," Chem. Biodivers., vol. 22, no. 3, Mar. 2025, doi: 10.1002/cbdv.202402423.

[18] S. Li, T.-P. Fan, W. Jia, A. Lu, and W. Zhang, "Network Pharmacology in Traditional Chinese Medicine," Evidence-Based Complement. Altern. Med., vol. 2014, no. 1, Jan. 2014, doi: 10.1155/2014/138460.

[19] G. Gogoshin and A. S. Rodin, "Graph Neural Networks in Cancer and Oncology Research: Emerging and Future Trends," Cancers (Basel)., vol. 15, no. 24, p. 5858, Dec. 2023, doi: 10.3390/cancers15245858.

[20] N. Qodir et al., "Anti-Breast Cancer Effects of Thymoquinone-Chemotherapeutic Combinations: A Systematic Review of the Latest In Vitro and In Vivo Studies," J. Clin. Med. Res., vol. 17, no. 5, pp. 270–284, May 2025, doi: 10.14740/jocmr6230.

[21] R. Arif, S. A. Bukhari, G. Mustafa, S. Ahmed, and M. F. Albeshr, "Network Pharmacology and Experimental Validation to Explore the Potential Mechanism of Nigella sativa for the Treatment of Breast Cancer," Pharmaceuticals, vol. 17, no. 5, p. 617, May 2024, doi: 10.3390/ph17050617.

[22] M. B. Madureira et al., "Naringenin and Hesperidin as Promising Alternatives for Prevention and Co-Adjuvant Therapy for Breast Cancer," Antioxidants, vol. 12, no. 3, p. 586, Feb. 2023, doi: 10.3390/antiox12030586.

[23] X. Sun, H. Zhang, Y. Zhang, Q. Yang, and S. Zhao, "Caspase-dependent mitochondrial apoptotic pathway is involved in astilbin-mediated cytotoxicity in breast carcinoma cells," Oncol. Rep., vol. 40, no. 4, pp. 2278–2286, 2018, doi: 10.3892/or.2018.6602.

[24] M. Younas, C. Hano, N. Giglioli-Guivarc'h, and B. H. Abbasi, "Mechanistic evaluation of phytochemicals in breast cancer remedy: current understanding and future perspectives," RSC Adv., vol. 8, no. 52, pp. 29714–29744, 2018, doi: 10.1039/C8RA04879G.

[25] J. D. Lambert et al., "Effect of genistein on the bioavailability and intestinal cancer chemopreventive activity of (-)-epigallocatechin-3-gallate," Carcinogenesis, vol. 29, no. 10, pp. 2019–2024, Mar. 2008, doi: 10.1093/carcin/bgn182.

[26] D. Pal, P. Singh, M. E. Prasad, and V. J. Upadhye, "Key Insights on Therapeutic Potential of Various Phytochemicals against Breast Cancer," Nat. Resour. Hum. Heal., vol. 5, no. 3, pp. 339–353, Jun. 2025, doi: 10.53365/nrfhh/203182.

[27] Y. Yuan, H. Long, Z. Zhou, Y. Fu, and B. Jiang, "PI3K–AKT-Targeting Breast Cancer Treatments: Natural Products and Synthetic Compounds," Biomolecules, vol. 13, no. 1, p. 93, Jan. 2023, doi: 10.3390/biom13010093.

[28] J. Wahler and N. Suh, "Targeting HER2 Positive Breast Cancer with Chemopreventive Agents," Curr. Pharmacol. Reports, vol. 1, no. 5, pp. 324 335, Oct. 2015, doi: 10.1007/s40495-015-0040-z.

[29] Y. Zhang, X. Huang, X. Yu, W. He, K. Czene, and H. Yang, "Hematological and biochemical markers influencing breast cancer risk and mortality: Prospective cohort study in the UK Biobank by multi-state models," The Breast, vol. 73, p. 103603, Feb. 2024, doi: 10.1016/j.breast.2023.103603.

[30] S. Zhu et al., "Effective therapy of advanced breast cancer through synergistic anticancer by paclitaxel and P-glycoprotein inhibitor," Mater. Today Bio, vol. 26, p. 101029, Jun. 2024, doi: 10.1016/j.mtbio.2024.101029.

[31] L. Banjare et al., "First report on exploration of structural features of natural compounds (NPACT database) for anti-breast cancer activity (MCF-7): QSAR-based virtual screening, molecular docking, ADMET, MD simulation, and DFT studies," Silico Pharmacol., vol. 12, no. 2, p. 92, Oct. 2024, doi: 10.1007/s40203-024-00266-5.

[32] A. F. Wali et al., "PI3K/AKT/mTOR Pathway in Breast Cancer Pathogenesis and Therapy: Insights into Phytochemical-Based Therapeutics," Nutr. Cancer, vol. 77, no. 9, pp. 938–958, Oct. 2025, doi: 10.1080/01635581.2025.2521884.

[33] J. Qiu et al., "Integrating UPLC-HR-MS/MS, Network Pharmacology, and Experimental Validation to Uncover the Mechanisms of Jin'gan Capsules against Breast Cancer," ACS Omega, vol. 7, no. 32, pp. 28003–28015, Aug. 2022, doi: 10.1021/acsomega.2c01921.

[34] S. Pancholi et al., "ERBB2 influences the subcellular localization of the estrogen receptor in tamoxifen-resistant MCF-7 cells leading to the activation of AKT and RPS6KA2," Endocr. Relat. Cancer, vol. 15, no. 4, pp. 985–1002, Sep. 2008, doi: 10.1677/ERC-07-0240.

[35] S. B. Matthews and C. A. Sartorius, "Steroid Hormone Receptor Positive Breast Cancer Patient-Derived Xenografts," Horm. Cancer, vol. 8, no. 1, pp. 4–15, Feb. 2017, doi: 10.1007/s12672-016-0275-0.

[36] C. L. McGinity et al., "Nitric Oxide Modulates Metabolic Processes in the Tumor Immune Microenvironment," Int. J. Mol. Sci., vol. 22, no. 13, p. 7068, Jun. 2021, doi: 10.3390/ijms22137068.

[37] U. Pal, A. Sahu, P. Barah, and A. M. Limaye, "Transcriptomic data of MCF-7 breast cancer cells treated with G1, a G-protein coupled estrogen receptor (GPER) agonist," Data Br., vol. 41, p. 107948, Apr. 2022, doi: 10.1016/j.dib.2022.107948.

[38] M. Devaraji and P. V. Thanikachalam, "Phytoconstituents as emerging therapeutics for breast cancer: Mechanistic insights and clinical implications," Cancer Pathog. Ther., vol. 3, no. 5, pp. 364–382, 2025, doi: 10.1016/j.cpt.2025.02.006.

[39] G. Obaido et al., "Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects," Mach. Learn. with Appl., vol. 17, no. July, p. 100576, 2024, doi: 10.1016/j.mlwa.2024.100576.

[40] X. Li et al., "Construction of a knowledge graph for breast cancer diagnosis based on Chinese electronic medical records: development and usability study," BMC Med. Inform. Decis. Mak., vol. 23, no. 1, p. 210, Oct. 2023, doi: 10.1186/s12911-023-02322-0.

[41] Q. Xiang et al., "Quantum classical hybrid convolutional neural networks for breast cancer diagnosis," Sci. Rep., vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-74778-7.

[42] Q.-M. Ding, Y.-M. Huang, and X. Yuan, "Molecular docking via quantum approximate optimization algorithm," Phys. Rev. Appl., vol. 21, no. 3, p. 034036, Mar. 2024, doi: 10.1103/PhysRevApplied.21.034036.