Machine Learning-Driven Insights into Nanoparticle Delivery Efficiency for Cancer Therapy

Abstract ID: 2058174

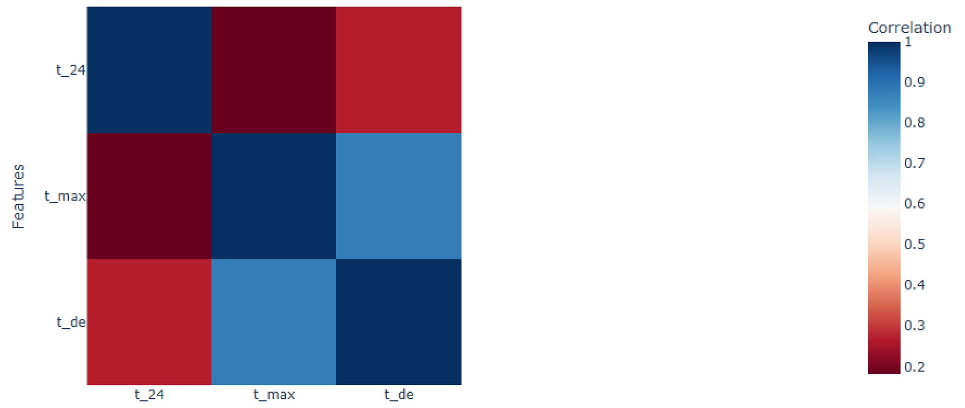
Vandana Soni, Arpana Purohit , Utkarsh Soni , Sushil K. Kashaw
Department of Pharmaceutical Sciences
Dr. Harisingh Gour University, (A+ Central University; Sagar (M. P.)-470003

INTRODUCTION

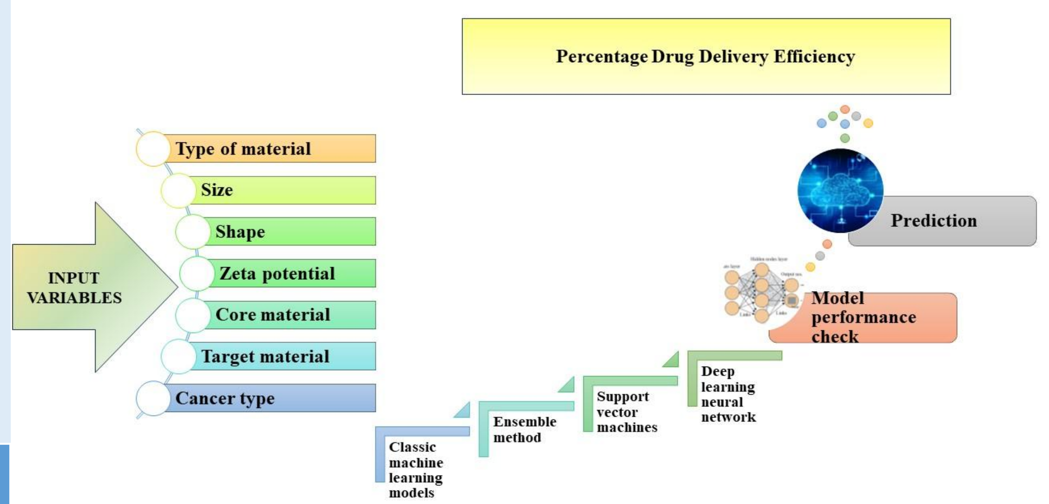
Cancer nanomedicine offers a promising approach to enhancing drug delivery efficiency, yet the unpredictable in vivo behaviour remains a challenge. Factors such as nanoparticle type, material, size, shape, surface modifications, zeta potential, targeting strategies, and tumor microenvironment dynamics significantly impact delivery efficiency. A systematic analysis of these factors is crucial for optimizing nanocarrier design and improving therapeutic outcomes. This study employs machine learning to identify key determinants influencing nanoparticle performance in preclinical cancer models. A dataset comprising 533 entries on nanoparticle physicochemical properties and delivery efficiencies in preclinical cancer models. Rigorous data preprocessing, feature engineering, five-fold cross-validation, and independent dataset validation ensured model accuracy and robustness. The approach successfully highlighted the most influential factors in nanoparticle-based drug delivery, providing valuable insights for optimizing nanocarrier design.

Names	mse_t_max	r2_t_max	mse_t_24	r2_t_24	mse_t_de	r2_t_de
LinearRegression	5.061454	0.564632	14.23478	0.2486	2.282802	0.945659
Ridge	3.741134	0.678201	6.868127	0.637458	1.878809	0.955276
Lasso	3.656982	0.68544	4.366838	0.769491	1.975023	0.952986
LassoLar	3.642928	0.686648	4.314398	0.772259	1.975024	0.952986
ElasticNet	3.642927	0.686649	4.251616	0.775573	1.405718	0.966538
OrthogonalMatchingPursuit	5.85667	0.49623	6.796353	0.641246	1.994605	0.95252
BayesianRidge	4.951655	0.574076	4.657329	0.754157	1.931722	0.954017
ARDRegression	5.40238	0.535307	6.886352	0.636496	1.828661	0.95647
TweedieRegressor Variant 1	5.04545	0.566009	9.487198	0.499207	16.51294	0.60692
TweedieRegressor Variant 2	6.057383	0.478966	11.16823	0.410472	12.99274	0.690716
SGDRegressor	1.88E+27	-1.62E+26	6.85E+24	-3.62E+23	9.33E+25	-2.22E+24
PassiveAggressiveRegressor Variant 1	25.53511	-1.19644	28.95988	-0.52868	15.03926	0.642
PassiveAggressiveRegressor Variant 2	24.63854	-1.11932	9.084854	0.520445	11.50002	0.72625
RANSACRegressor	9.138615	0.213929	5.364463	0.71683	1.54251	0.963282
HuberRegressor(max_iter=1000)	3.394345	0.708031	4.973444	0.737471	1.615779	0.961537
TheilSenRegressor	4.742527	0.592065	5.79454	0.694128	1.978523	0.952903
QuantileRegressor	13.9219	-0.19751	24.25246	-0.2802	48.93982	-0.16498
KernelRidge	4.790948	0.5879	7.532457	0.60239	3.207777	0.923641
SVR	13.76526	-0.18404	23.72592	-0.2524	48.78225	-0.16123
NuSVR	12.91596	-0.11098	22.18922	-0.17128	45.84772	-0.09137
LinearSVR	13.16835	-0.13269	4.787662	0.747278	6.883586	0.836141
KNeighborsRegressor	10.30419	0.113671	18.78729	0.00829	37.38134	0.110162
GaussianProcessRegressor	16.32951	-0.40461	25.01569	-0.32048	55.02881	-0.30992
PLSRegression	5.589033	0.519251	6.506162	0.656564	5.6801	0.864789
CCA	4.354568	0.625436	12.37772	0.346627	2.436724	0.941995
PLSCanonical	21.3076	-0.8328	27.93585	-0.47463	24.55669	0.415444
GradientBoostingRegressor	4.203703	0.638413	4.764448	0.748503	2.377055	0.943416
DecisionTreeRegressor	23.23981	-0.99901	5.176936	0.726729	5.037744	0.88008
HistGradientBoostingRegressor	4.38692	0.622653	3.194215	0.83139	6.11654	0.8544
RandomForestRegressor	4.704446	0.59534	2.321254	0.87747	1.901598	0.954734
ExtraTreesRegressor	5.843384	0.497373	1.441854	0.92389	2.114463	0.949667
IsolationForest	12.78224	-0.09948	23.20352	-0.22483	48.90175	-0.16407
AdaBoostRegressor	4.079479	0.649098	3.512662	0.81458	2.509278	0.940268
BaggingRegressor	2.718246	0.766186	5.076725	0.732019	3.792603	0.90972

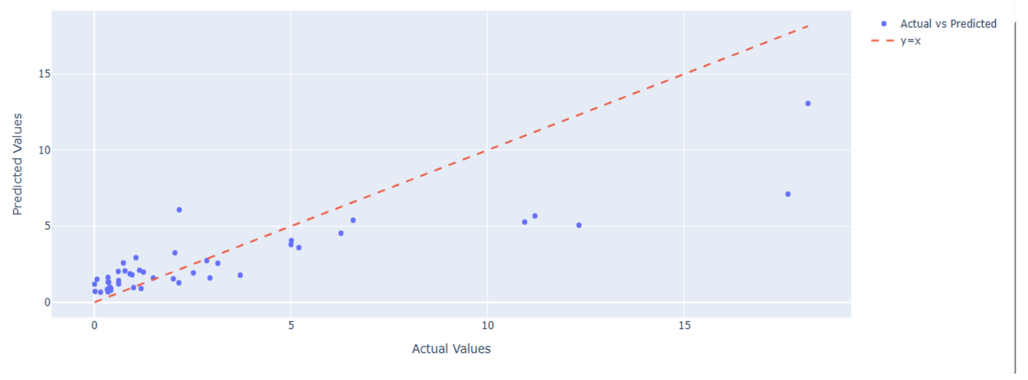
Correlation Heatmap of Delivery Metrics



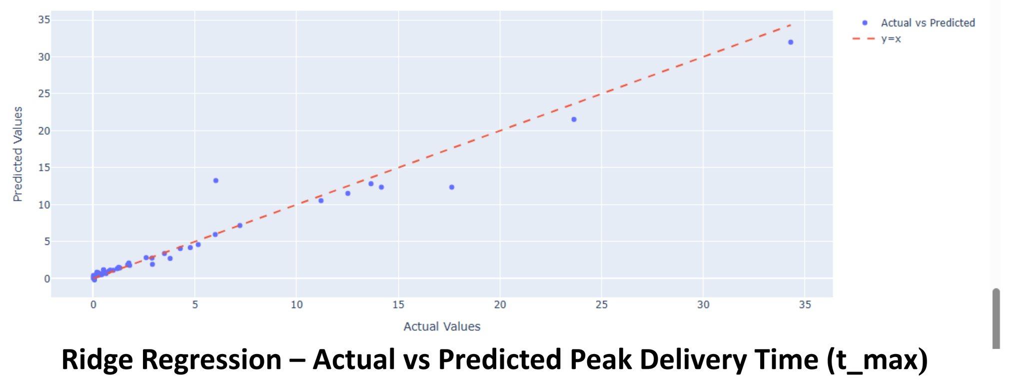
WORKING SCHEME



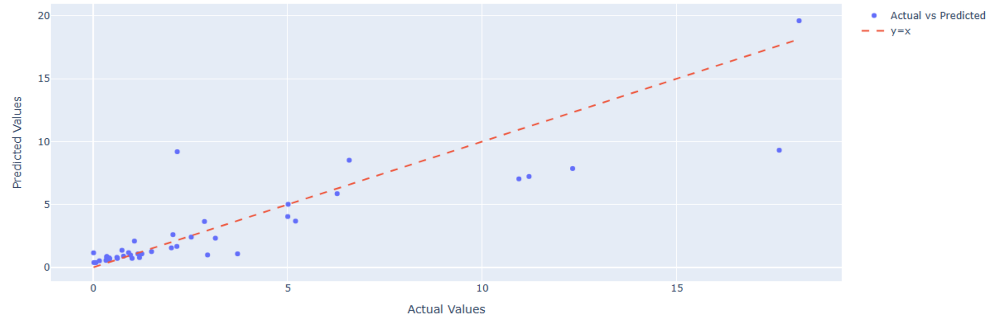
Ridge Regression– Actual vs Predicted Delivery Efficiency (t_24)



Ridge Regression – Actual vs Predicted Total Delivery Efficiency (t_de)



Ridge Regression – Actual vs Predicted Peak Delivery Time (t_max)

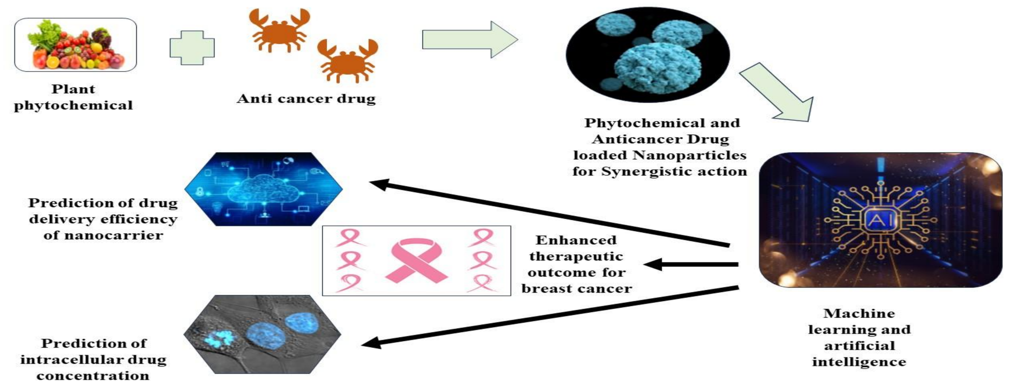


RESULTS AND CONCLUSION

Ridge Regression demonstrated strong predictive accuracy (MSE = 1.87, R^2 = 0.95), while ElasticNet and Huber Regressor achieved comparable accuracy (R^2 = 0.96 each). Ensemble models such as Gradient Boosting and Random Forest exhibited excellent prediction capabilities for 24-hour delivery efficiency (R^2 > 0.87). The analysis identified critical factors governing nanoparticle delivery efficiency, laying the foundation for rational nanocarrier design. These findings contribute to addressing challenges in predicting nanoparticle behavior and optimizing drug delivery systems for cancer therapy.

This study demonstrates the efficacy of machine learning in identifying key determinants of nanoparticle delivery efficiency. The insights gained can facilitate the design of optimized nanocarriers, improving drug delivery strategies in cancer therapy. Future research should explore integrating deep learning approaches and experimental validation to further refine predictive models.

FUTURE PERSPECTIVE



References- 1) Lin, Zhoumeng, et al. "Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches." International journal of nanomedicine (2022): 1365-1379.