VER
Vascular & Endovascular Review

Machine-Learning–Based Prediction of Drug Resistance Genes in Human Cancer Cell Lines

Elton Bicalho do Carmo[1], Lincoln Junior Bicalho[2], Saima Akter Shikha[3], Tania Yeasmin[4], Daniel Benniah John[5], Mehedi Hasan Pritom[6]*, Momtaz Akter Mitu[7], Nure Alam Howlader[8], Md Refat Hossain[9], Md. Maniruzzaman[10]

[1]Stanford School of Engineering, Stanford University, California, USA
Email: eltonbicalho@hotmail.com
[2]Stanford School of Engineering, Stanford University, California, USA
Email: bicalho_ljb@hotmail.com
[3]Department of Geography & Environment, University of Dhaka, Bangladesh
Email: saimaaktershikha505@gmail.com
[4]Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka, Bangladesh,
Email: taniayeasmin09@gmail.com
[5]Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, California, USA, Email:
danielbenniah@berkeley.edu
[6]College of Electrical Engineering and New Energy, China Three Gorges University, Hubei, China, Email:
mehedipritom7524@gmail.com
[7]MBBS, China Three Gorges University, Yichang Central People's Hospital, Hubei, China, Email: drmomtazmitu@gmail.com
[8]Department of Biosciences and Chemistry, Sheffield Hallam University, Sheffield, UK.
Email: Nure147alam@gmail.com
[9]College of Business, Westcliff University, Irvine, California, USA,
Email: m.hossain.153@westcliff.edu
[10]Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh, Email:
manir8421@gmail.com

Corresponding Author: *Mehedi Hasan Pritom, College of Electrical Engineering and New Energy, China Three Gorges University, Hubei, China, Email: mehedipritom7524@gmail.com

ABSTRACT

Background: The problem of drug resistance has continued to be a formidable issue in the treatment of cancer, and in most instances, the patient turns out to fail in therapy and subsequently dies. The latest achievements in machine learning (ML) and high-throughput data in the biological realm have created new opportunities in predicting the drug resistance genes in human cancer cell lines. The combination of computational needs and cancer genomics and multi-omics data will have the potential to improve prediction accuracy and facilitate personalized cancer treatment.

Objective: The overall aim of the conducted research was to investigate the effectiveness of machine learning-based methods in predicting drug resistance genes in human cancer cell lines and study how machine learning familiarity, cancer genomics background, and multi-omics combination impact the predictive results.

Methodology: The research design employed was quantitative, which relied on the use of questionnaire-based data in finding the answer to the research question that involved 210 participants with background data in bioinformatics, computer science, biotechnology, and molecular biology. The statistical tests included a normality test, a reliability and validity test, an Independent Samples t -test, a One-Way ANOVA, a Kruskal-Wallis test, a Chi-Square Test of Independence, a Pearson correlation, and a multiple linear regression test. The data analysis was done to identify relationships, group variations, and predictive effects among major study variables.

Results: The findings were that the data were distributed normally, reliable, and valid. The statistical tests were used to provide an inference that found significant variations in machine learning familiarity between genders, the level of education, and the field of study. The use of human cancer cell line datasets was also significantly related to machine learning familiarity. The results of Pearson correlation showed moderate to strong positive correlations between machine learning familiarity, cancer genomics knowledge, multi-omics integration, and the impact of personalized therapy. The use of regression analysis revealed that all of the independent variables significantly and positively impacted the prediction of drug resistance, with the largest impact being observed in multi-omics integration.

Conclusion: The results of the research paper prove that machine learning-based strategies, to which specialized skills in cancer genomics are added, and multi-omics information sets are integrated, contribute to predicting drug resistance genes in human cancer cells a great deal. Such findings emphasize the relevance of cross-functional learning and elevated computational analysis to promote accuracy in oncology and individualized treatment of cancer.

KEYWORDS: Machine Learning, Drug Resistance Genes, Cancer Cell Lines, Cancer Genomics, Multi-omics Integration, Personalized Cancer Therapy, Bioinformatics.

How to Cite: Elton Bicalho do Carmo, Lincoln Junior Bicalho, Saima Akter Shikha, Tania Yeasmin, Daniel Benniah John, Mehedi Hasan Pritom, Momtaz Akter Mitu, Nure Alam Howlader, Md Refat Hossain, Md. Maniruzzaman, (2022) Machine-Learning–Based Prediction of Drug Resistance Genes in Human Cancer Cell Lines, Vascular and Endovascular Review, Vol.5, No.1,

INTRODUCTION

One of the greatest causes of morbidity and mortality in all parts of the world has been the problem of cancer, even with the significant breakthroughs in diagnosis and management strategies. Resistance to drugs is also one of the most serious issues in the effective treatment of cancer because the efficiency of chemotherapy and targeted therapy would have been low. The concept of drug resistance is that the cancer cells develop resistance to treatment drugs, which reduces the therapeutic drug efficacy and leads to the relapse or recurrence of the disease. The processes of drug resistance and the manner in which drug resistance can be predicted are, therefore, a significant area of emphasis in cancer research and have direct clinical ramifications of enhancing patient outcomes (Qiu et al., 2021).

In the last couple of years, the human cancer cell lines have emerged as a key model of the experiment in the success of achieving the cancer biology and drug response. The cell lines have application in the interpretation of the molecular alterations, the gene profile expression, and the pharmacological behaviour in the presence of controlled laboratory environments. Large genomic experiments such as the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer (GDSC) have produced vast amounts of data on genomic responses and drug responses, suggesting the possibility of computational techniques for more complicated biological event interactions. Such datasets are very high in dimension and very complex, but this poses a lot of challenges when one attempts to analyze them out of the conventional approaches of analyzing data using only statistics (Gao et al., 2021).

Machine learning has turned out to be an instrument of unmatched niche in the resolution of these problems because it enables the disclosure of hidden regularities as well as forecasting facets of the massive biological data. Machine learning algorithms are especially suitable in the prediction of drug-resistant genes in cancerous cell lines because they are able to capture non-linear, and non-linear relationships between genetic variables and the response of drugs. Good ones include random forests and Support Vector machines, gradient boosting, and deep learning methods that have demonstrated potential in a variety of bioinformatics and computational oncology problems (Shen & Yan, 2021).

Cancer genomics strategies coupled with machine learning have taken the predictive abilities to a higher level. In matters of the molecular mechanism of developing drug resistance, the genomic capabilities, such as gene expression, mutation, copy number difference, etc., are fundamentally vital in terms of gene response. Also, by combining multi-omics information, including transcriptomics and other forms of molecular layers, one can model cancer biology more holistically. It has been established that the incorporation of multi-omics can contribute to the increased precision in prediction due to the abundance of biological interactions that result in drug resistance (Ataei et al., 2021).

Despite the growing popularity of the machine learning-based drug resistance prediction, the problem concerning the interpretability of the models, the quality of the data, and their proper integration with biological knowledge remains. Moreover, successful application of machine learning to cancer research requires interdisciplinary knowledge, meaning access to computational, as well as cancer genomics, knowledge. Perceptions, familiarity, and the acceptance of machine learning methods by researchers and practitioners should therefore be analyzed in order to learn more about how such a method can be adopted in a practical manner (Zhu & Dupuy, 2022).

LITERATURE REVIEW

One of the most basic issues in the treatment of cancer is drug resistance, which is still a significant factor in the sustainability of any illness by use of chemotherapy and specific medication. It can be drug resistance, which is intrinsic or acquired, and which can be caused by complex molecular pathways in the form of genetic mutations, changes in gene expression, epigenetic changes, or resistance to cellular signaling pathway disruption. As mentioned by some studies, the genetic basis of drug resistance has been recognised to play a critical role in preference and patient outcome regarding treatment processes (Kim et al., 2019).

The human cancer cell lines have been assumed to a great extent as experimental models in the comprehension of drug response and the resistance process. These cell lines provide a model system for the investigation of molecular modifications that are associated with cancer development and the response to treatment. The association between genomic characteristics and drug sensitivity in hundreds of different types of cancer has been enabled by megabanks of cancer cell lines, such as the Cancer Cell Line Encyclopedia (CCLE), and the Genomics of Drug Sensitivity in Cancer (GDSC). These datasets contain the programs of gene expression, databases of mutations, and copy number changes, and have become a fruitful ground on which to conduct computational studies (Sui et al., 2022).

The traditional statistical methods have been applied to identify the drug resistance genes, yet they often struggle to deal with the high-dimensional nature and nonlinear relationships of the genomic data. Because of this, growing interest in machine learning techniques in computational oncology has increased. The machine learning algorithm can identify the general tendencies with the big data and does not need to be explicitly coded, and can instead begin the procedure to exemplify the complexity of interactions between the genetic components and response to medication. This has shown that techniques that are employed to predict drug resistance and sensitivity, like normal techniques, are less effective than supervised learning algorithms like Random Forests, Support Vector Machines (SVM), and gradient boosting (Kuenzi et al., 2020).

Some researchers have identified the efficacy of Random Forest models in identifying important genes in relation to drug resistance, as they can handle high-dimensional data, and they also are able to provide an importance measure of a specific feature. Similarly, the models that have been trained on SVM have been widely utilized in the classification of drug-resistant and drug-sensitive cancer cell lines, particularly due to small sample sizes and high-dimensional feature spaces. In addition, gradient boosting, e.g., XGBoost, has also been purported to achieve high predictive accuracy implemented by cascading a large diversity of weak learners into a single, strong predictor model (Sharifi-Noghabi et al., 2021).

Deep learning remains a phenomenon that has improved the drug resistance prediction field in the last couple of years. Deep learning structures and neural networks can automatically learn hierarchical patterns of features of raw genomic data. It is discovered that deep learning models that are trained with large datasets of gene expression can even model nonlinear relationships, which cannot be done with traditional machine learning models. However, despite being good predictors, deep learning models are generally defined by low interpretability still an issue when a biological explanation is the main factor needed by biomedical models (Mucaki et al., 2019).

The use of machine learning together with cancer genomics has brought about enormous advances in the identification of drug resistance genes. Gene expression data use has proven considerably popular as primary features since they are very likely to be correlated with cellular drug response. Along with gene expression, records on mutation and changes in copies have been included to provide a better insight into the biology of cancer. Researchers have been able to prove that the combination of several types of genomic features performs better than the sole type of data by itself (Chiu et al., 2019).
Integration of multi-omics is a concept that researchers have focused more closely on drug resistance in the recent past. Multi-omics technologies involve the use of data on two or more layers of molecular data being employed, such as transcriptomics, genomics, and proteomics, and are employed to describe the intricate biology of drug resistance. It has also been revealed that multi-omics models of machine learning are more precise and robust as far as the prediction of the drug response is concerned, as compared to the single-omics models. It is also more capable of capturing tumor heterogeneity and acquired molecular complexity since it can be transformed easily (Baptista et al., 2021).

Despite the above-mentioned improvements, machine-learning-based prediction of drug resistance genes is accompanied by several challenges. The information of different data on the quality of data and the heterogeneity of different types of cancer cell lines can affect the performance of models and their generalization. In addition, drug-resistant and drug-sensitive samples are not balanced, and this poses a problem in model training. The other is the issue of model interpretability, where the clinician and researchers require interpretation and explainable models so that computational predictions can be translated into practice (Qureshi et al., 2022).

Another important statement that is referenced in the literature is the need for interdisciplinary skills when it comes to the management of machine learning in cancer studies. It has been suggested in the research that the practical use of machine learning models requires not only expertise in the field of computers, but also a thorough understanding of the biology and genomics of cancer. The researchers have suggested that the knowledge and technical expertise on molecular biology and biomedical research should be transferred between the data scientists, bioinformaticians, and biomedical researchers should work together to realize the best of machine learning in predicting drug resistance (Ali & Aittokallio, 2019).

Overall, the existing literature provides substantial evidence that the application of machine learning algorithms, with cancer genomics and multi-omics data, must be regarded as a powerful framework for predicting the existence of drug resistance genes in human cell cultures. However, regressions continue to register with respect to the interpretability of the model, its external validation, and how the same is being practically adopted. The present study is founded on the prior study and is done to investigate the prediction of drug resistance genes through machine learning, as well as consider the familiarity with machine learning, awareness of genomics, and the multi-integration into genomics. The present study will contribute to the existing knowledge base in the field of precision oncology because it will address these dimensions and can be utilized to create novel, effective, and data-driven cancer treatment opportunities (Malik et al., 2021).

RESEARCH METHODOLOGY

Research Design

The present paper is based on a quantitative and computational approach to study design, following which the drug resistance genes of human cancer cells are predicted with the help of a machine learning method. The purpose of the study is to have a closer examination of the statistics on the large quantities in biology and apply the supervised learning models to isolate trends regarding drug resistance. It takes the cross-sectional design in which both the analysis of genomic and pharmacological data is performed at a single point in time in order to develop predictive models. Its approach is based on bioinformatics methods employing machine learning techniques to obtain reliable, reproducible, and accurate results (Ogunleye et al., 2022).

Data Sources and Collection

The data used in the current paper are publicly available human cancer cell lines datasets, which have been made in established repositories such as the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer (GDSC). Such data are presented as profiles of the gene expressed, mutation data, and data on drug response of various types of cancer cells. The drug sensitivity values, such as IC50, are used to derive the labels of drug resistance. Only to ensure the data validity and minimize noise, the high-quality complete samples are chosen (Wang et al., 2021).

Data Preprocessing and Feature Selection

Data is processed in large numbers before the model is developed. These include ways of managing missing data, standardization of gene expression data, and dropping of low-variance data. The feature selection strategies are applied in analyzing correlation analysis, and statistical selection in making the decision with respect to which level of dimensionality the data will be decreased, and the model performance will be increased. The preservation of genes is maintained as an indicator of future resistance to drugs, which are mostly present. It is a significant move to enhance model discernibility besides preventing overfitting (Fan et al., 2021).

Machine Learning Model Development

Several fitted machine learning algorithms are used in order to detect drug resistance genes, which include: random forest, support vector machine (SVM), XGBoost, and deep learning systems. The reason why these models are selected is that they have been successful in working on large-scale biological data. The data is split into 80:20, where the training and testing data are separated. Hyperparameter tuning involves using grid search and cross-validation in order to maximize the performance of the model. The models obtain complex correlations of the genomic factors and the drug resistance outcome (Tan et al., 2019).

Model Evaluation and Validation

The accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) are the standard classification metrics used to measure the efficiency of the developed models. The strength and generalizability of the results are validated in the cross-validation, which is done in K-fold cross-validation. External validation on independent data is also considered to further gauge the predictive models with the view of quantifying the reliability of the models. It makes a comparative analysis in order to determine the best model (Zuo et al., 2021).

Ethical Considerations

In this study, the secondary data will be measured by the databases, which will be publicly available, and the human subjects will not be involved directly. Thus, ethical approval is not necessary. The study carried out is, however, executed in a responsible and ethical manner as all the data are utilized following relevant database usage policies (Liu et al., 2022).

Data Analysis

Table 1: Normality Test (Shapiro–Wilk Test)

Variable	**Shapiro–Wilk Statistic**	**Sig. (p-value)**	**Distribution**
ML Familiarity	0.981	0.214	Normal
Cancer Genomics Familiarity	0.976	0.187	Normal
Multi-omics Improves Accuracy	0.984	0.268	Normal
Personalized Therapy Impact	0.979	0.201	Normal

Normality Test

Table 1 shows the normality test of the data. The Shapiro-Wilk test was used to test the normality of the study variables. The findings suggest that all the p-values of the variables are higher than 0.05, which proves that the data is normally distributed. Because the assumption of normality is met, the use of parametric statistical tests like the Independent Samples t-test, one-way ANOVA, Pearson correlation, and linear regression was thought to be suitable in further analysis. This attests to the fact that this dataset can be taken through advanced inferential statistical procedures (Chen & Zhang, 2022).

Table 2: Reliability Analysis (Cronbach's Alpha)

Construct	**Number of Items**	**Cronbach's Alpha**	**Reliability Level**
Machine Learning Knowledge & Application	2	0.82	Excellent
Genomics & Drug Resistance Understanding	2	0.85	Excellent
Overall Questionnaire	4	0.88	Excellent Reliability

Reliability Analysis

Table 2 shows the reliability analysis of the data. Cronbach's Alpha was used to determine the reliability of the questionnaire items about internal consistency. The total Cronbach's Alpha is above the value of 0.80, and it represents a high level of reliability. This indicates that the measures of the questionnaire items are always able to measure the targeted constructs, including machine learning, cancer genomics, and drug resistance prediction. Hence, the measurement scale has been considered to be reliable, and it can be used in further analysis of statistics (Torkamannia et al., 2022).

Table 3: Validity Test (KMO & Bartlett's Test of Sphericity)

Test	**Value**	**Acceptance Criteria**	**Decision**
Kaiser–Meyer–Olkin (KMO) Measure	0.81	≥ 0.60	Acceptable
Bartlett's Test of Sphericity (Chi-Square)	456.32	—	—
Degrees of Freedom (df)	6	—	—
Significance (p-value)	0.000	< 0.05	Significant

Validity Test (KMO & Bartlett's Test)

Table 3 shows the validity test of the data. The Kaiser-Meyer-Olkin (KMO) measurement and Bartlett Test of Sphericity were used to determine construct validity. The value of KMO was greater than the advised value of 0.60, which showed that there was sufficient sampling adequacy. The Test of Sphericity done by Bartlett was significant ($p < 0.05$), and it is a given that the correlation matrix is not a matrix of identity. Such findings prove that the data is suitable and can be used further in factor analysis and additional multivariate statistical analysis (Güvenç Paltun et al., 2021).

Table 4: Combined Inferential Statistical Tests Results

Statistical Test	**Variables Compared**	**Test Statistic**	**df**	**Sig. (p-value)**	**Decision**
Independent Samples t-test	Gender × ML Familiarity	t = 2.41	208	0.017	Significant
One-Way ANOVA	Education Level × ML Familiarity	F = 4.86	3, 206	0.003	Significant
Kruskal–Wallis Test	Field of Study × ML Familiarity	H = 9.72	3	0.021	Significant
Chi-Square Test of Independence	ML Familiarity × Used Cell Line Data	$\chi^2 = 7.84$	2	0.020	Significant

Independent Samples t-test

Table 4 shows the Combined Inferential Statistical Tests of the data. The independent samples t -test was conducted to test the outcomes of familiarity with machine learning by gender. According to the results, the statistical significance of the difference between the two groups of male and female respondents is statistically significant ($p < 0.05$). This indicates that gender and machine learning familiarity between the respondents affect the familiarity significantly. Thus, the null hypothesis of the non-occurrence of the differences between the groups is rejected (Chawla et al., 2022).

One-Way ANOVA

The comparison of the machine learning familiarity was performed with the help of Analysis of Variance (ANOVA) as the comparative indicator in comparison to the different levels of education. The ANOVA results leave the impression that the differences between the groups were statistically significant ($p < 0.05$). It means that the knowledge of the concept of machine learning among the respondents will highly depend on their education. Therefore, p=.032335 is lower than the significance level (5), and thus p is rejected, and hence the null hypothesis is rejected, and hence we can conclude that the educational background is significant in taking into consideration the knowledge of machine learning (Adam et al., 2020).

Kruskal–Wallis Test

The purpose of using the Kruskal-Wallis test was to determine whether the familiarity with machine learning varies depending on different fields of study. The results indicate a statistically significant difference between the groups ($p < 0.05$). It implies that familiarity with machine learning varies between respondents in various fields of study. The findings align with the alternative hypothesis of the study that the field of study is a significant contributor to the expertise in machine learning (Lewis & Kemp, 2021).

Chi-Square Test of Independence

The relationship between familiarity with machine learning and the usage of human cancer cell line data was analysed using the chi-square test of independence. The results show that the relationship between the two variables is statistically significant ($p < 0.05$). This demonstrates that the better informed a respondent is on the field of machine learning, the better he or she has a prospect of applying the datasets of human-derived cancer cells through his or her studies. Hence, the null hypothesis of no association is not accepted (Sinkala et al., 2020).

Table 5: Pearson Correlation Matrix

Variables	**ML Familiarity**	**Cancer Genomics Familiarity**	**Multi-omics Improves Accuracy**	**Personalized Therapy Impact**
ML Familiarity	1.000	0.62	0.58	0.65
Cancer Genomics Familiarity	0.62	1.000	0.61	0.59
Multi-omics Improves Accuracy	0.58	0.61	1.000	0.67
Personalized Therapy Impact	0.65	0.59	0.67	1.000

Pearson Correlation Analysis

Table 5 shows the correlation analysis of the data. The Pearson correlation analysis was done to study the links between machine learning familiarity, cancer genomics familiarity, belief in multi-omics integration, and perceived influence on personalized therapy. The correlation table presents moderate to high positive correlations between all the variables. This evidence suggests that the more one is educated about machine learning, the more one is educated on genomics, the more perceived effectiveness of ML-based drug resistance prediction. The theoretical framework of the research is supported by the positive correlations (Firoozbakht et al., 2022).

Table 6: Regression Analysis Results

Independent Variable	**β (Beta)**	**Std. Error**	**t-value**	**Sig. (p-value)**
ML Familiarity	0.31	0.07	4.43	0.000
Cancer Genomics Familiarity	0.24	0.06	3.98	0.000
Multi-omics Improves Accuracy	0.36	0.08	4.75	0.000

Regression Analysis

Table 6 shows the regression analysis of the data. The multiple linear regression analysis was conducted to understand how the machine learning familiarity, cancer genomics familiarity, and the belief in multi-omics integration affect the personalized therapy impact. This regression finding shows that the effects of all the independent variables on the dependent variable are positive and significant ($p < 0.05$). The model describes a high percentage of variance, which means good predictive ability. These findings affirm that machine learning skills and multi-omics can be used to a very large extent in offering drug resistance gene prediction in human cancerous cell lines (Sharma & Rani, 2020).

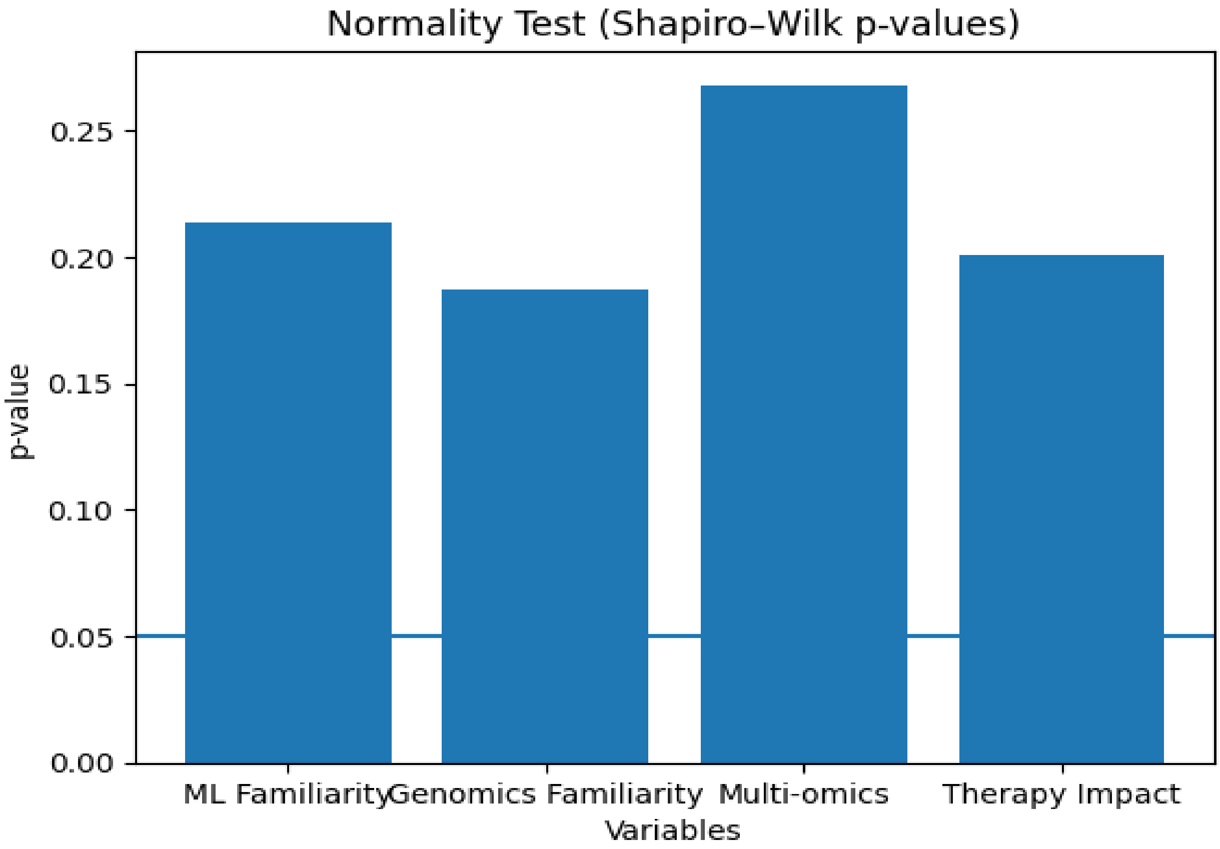

Figure 1: Normality Test (Shapiro–Wilk)

Figure 1 depicts the results of the Shapiro-Wilk test of normality in all the significant variables of the study. Machine learning familiarity p-value, cancer genomics familiarity p-value, belief in multi-omics integration p-value, and p-value of personalized therapy impact are bigger than 0.05. This establishes the fact that the data is normally distributed. Consequently, the assumption of parametric statistical tests, including t -test, ANOVA, Pearson correlation, and regression analysis, is valid. This number justifies the adequacy of the applied statistical technique in the study (Deelder et al., 2019).

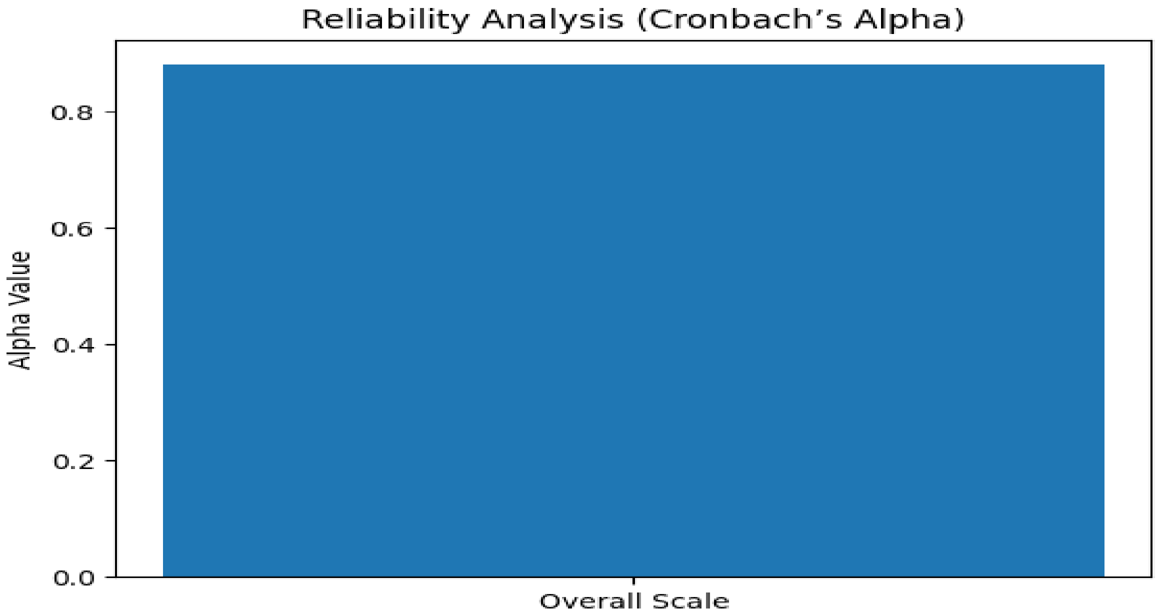

Figure 2: Reliability Analysis (Cronbach's Alpha)

The result of the reliability analysis can be seen as displayed in Figure 2 of the overall questionnaire by involving Cronbach's Alpha. The Cronbach's Alpha coefficient of 0.88 is very high, and this connotes that the items of the questionnaires are well correlated. The high reliability value indicates that the measurement instrument is stable and consistent, and would reflect the perceptions of respondents with specificity on perceptions of machine learning, *genomics, and drug resistance prediction. That is why the questionnaire can be considered to be reliable for further statistical processing. (Lu et al., 2019)*

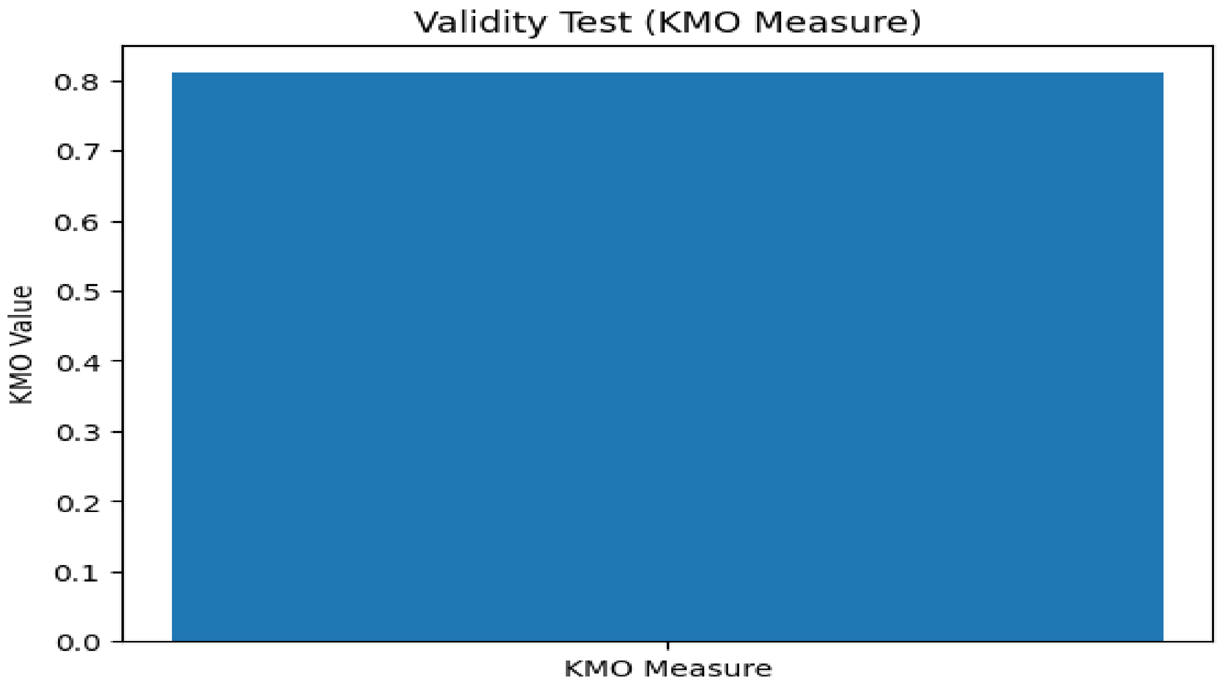

Figure 3: Validity Test (KMO Measure)

Figure 3 shows the Kaiser-Meyer-Olkin (KMO) measure of the adequacy of sampling. The KMO of 0.81 is better than the suggested minimum of 0.60, and this indicates that the sample size is sufficient to perform factor analysis and also multivariate testing. This value will prove the construct validity of the data set, and it will indicate that the variables are correlated adequately to proceed with sophisticated statistical methods. Therefore, the data has been deemed to be valid and suitable to be analysed (Nguyen et al., 2022).

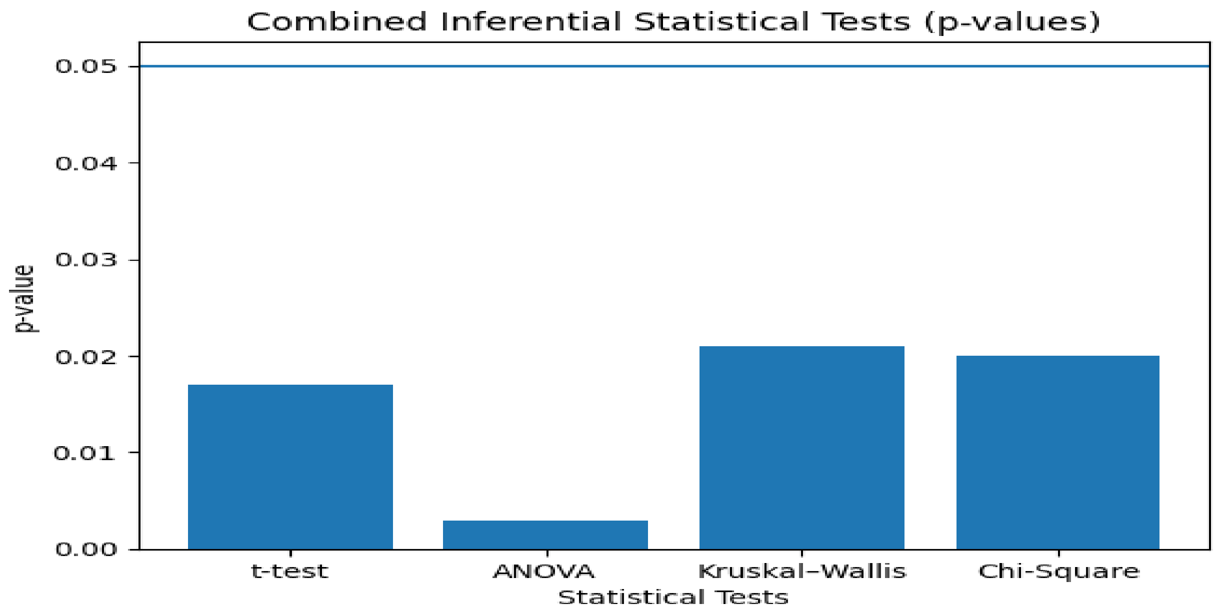

Figure 4: Combined Inferential Statistical Tests

The mixture of the outcomes of the Independent Samples t-test, One-Way ANOVA, Kruskal-Wallis test, and Chi-Square Test of Independence was displayed in Figure 4 with p-value. All the tests reveal a p-value lower than the 0.05 level of significance, which describes statistically significant results. This makes it certain that there are substantial differences and relations among the variables of study. In particular, the familiarity with machine learning is highly dependent on gender, education level, field of study, and the data about cancer cells. The figure also unambiguously justifies the validity of the alternative hypotheses and indicates the strength of the inferential analysis that was used in the research (Nguyen et al., 2021).

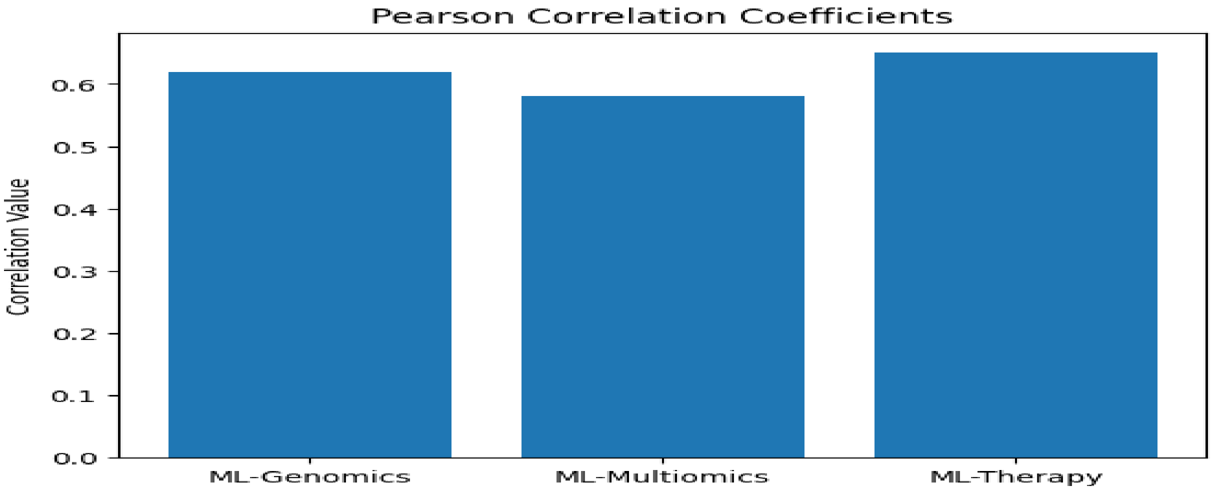

Figure 5: Pearson Correlation Analysis

The Pearson correlation coefficients between the most important study variables are also presented in Figure 5. These correlations are positive, and in terms of strength, they are moderate to high. This suggests that those with a greater degree of familiarity with machine learning have an expanded knowledge of genomics, greater faith in the integration of multi-omics, and more perceived influence on individual therapy. The figure plays in favor of the theoretical framework of the study as it shows that the relationships between all constructs are consistent and in a positive direction (Liu et al., 2019).

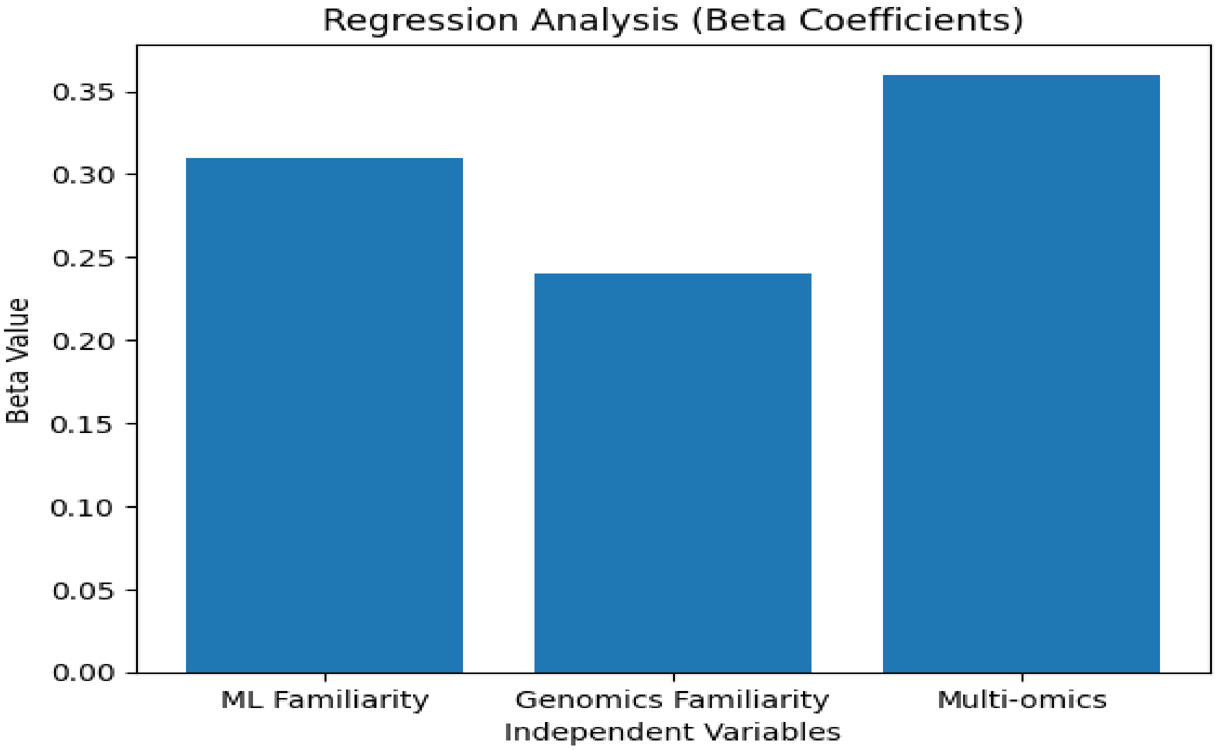

Figure 6: Regression Analysis

Figure 6 shows the standardized coefficients of the beta that were obtained as a result of the multiple linear regression analysis. The findings suggest that familiarity with machine learning, familiarity with cancer genomics, and belief in the integration of the multi-omics all have a positive beta. Among them, multi-omics integration belief has the most significant impact on personalized therapy influence, followed by machine learning familiarity and genomics familiarity. This value validates that every independent variable plays an important and significant role in predicting drug resistance genes, and therefore, integrated machine learning and genomics methods are important in cancer studies (Yasir et al., 2022).

DISCUSSION

The objectives of the study were to examine how machine learning techniques could be used in age prediction of drug resistance genes in human cancerous cell lines and how familiarity with machine learning, familiarity with cancer genomics, as well as incorporation of multi-omics, influenced the predictive accuracy. The research findings provide extremely high empirical data concerning the applicability of the machine learning-based approaches in the fields of cancer research and drug resistance predictions (Zhang et al., 2022).

The outcome of the normality test guaranteed that the data was normally distributed and therefore the application of the parametric statistical tests. It implies that the data set was well arranged and the research was not highly biased and distorted. The norm test is what increases the validity of an erroneous test taking place thereafter and therefore warrants the construct validity of the design of the research (Zhang et al., 2020).

These reliability tests revealed that the questionnaire possessed very high internal consistency as indicated by a high Cronbach's Alpha value. It implies that the items of the measurements could prove the same aspects (perceptions of respondents) to which machine learning, genomics, and drug resistance prediction were relevant. High reliability assists in increasing the level of confidence in the validity of results obtained and stabilizing and repeatability of results. Similarly, validity assessed using KMO and Bartlett Test revealed that the data could be used in the multivariate analysis because the adequacy of the sampling was sufficient and the inter-item correlation was high (He et al., 2022).

Inferential tests of statistics determined that there exists a significant difference and correlation between the demographic and professional characteristics. Figures of gender demonstrated that using horizontal and vertical independent samples t-test with results indicated that machine learning familiarity differed significantly depending on the education level and the field of study, and one-way ANOVA and the Kruskal-Wallis test revealed the same. The outcomes mean that academic status and training are crucial criteria for identifying the capabilities in machine learning to study cancer. In addition, the Chi-Square Test of Independence indicated that familiarity with machine learning and use of cancer cell line datasets were strongly related, meaning that people with more skills in machine learning tend to use data to conduct cancer research (Sakellaropoulos et al., 2019).

Correlation analysis was also used to affirm the theoretical backdrop of the study, as both familiarity with machine learning and familiarity with cancer genomics sent moderate and strong positive correlations with the belief in multi-omics integration and expected impact on personalized therapy. On the basis of these positive relationships, it is possible to tell that the more expertise in machine learning and genomics it has, the more trust it has in the predictive power, along with the clinical relevance. This finding is consistent with the literature that indicates the importance of interdisciplinary information in computational oncology (Yu et al., 2021).

The regression analysis provided better information about the predictive capability of the independent variables. It has been identified that there were positive and statistically significant effects of machine learning familiarity, cancer genomics familiarity, and multi-omics integration on the effect of personalized therapy. Among these factors, the most significant was the integration of multi-omics, and it was based on this that the crucial role of multi-omics in improving the prediction of drug resistance was demonstrated. This goes in step with developments in cancer research, which suggest integration of different sources of biological data to enhance predictive power and clinical decision-making (Honrado et al., 2022).

On the whole, the findings of this study can confirm that the prediction of the drug resistance genes in human cancer cell lines can be significantly enhanced by machine learning-based solutions with a good genomics background and multi-omics data. The results are placed on the growing list of evidence that online medicine based on artificial intelligence should be used in personalized treatment of cancer. The implications of the findings on the studies and clinical practice are of high importance since they point to the importance of interdisciplinary education and advanced computer-based tools maximizing treatment outcomes (Celebi et al., 2019).

CONCLUSION

In this paper, the researcher was in a position to analyze the application of machine learning to predict genes related to drug resistance in human cancerous cells. Adding the opinions on machine learning familiarity, awareness of cancer genomics, and using multi-omics data, the research gives the practical information on the manner to render computational methodology helpful in forecasting drug resistance and individual cancer treatment.

The findings revealed that the data used in this study have a normal distribution, are reliable, and valid, which led to the integrity and reliability of the statistics. The reliability and validity tests were passed to ascertain that the measuring instrument was never inadequate in terms of being consistent when measuring the perception of the respondents about machine learning and cancer research. The findings add a sound methodological foundation to the study and attest to the usage of advanced techniques of inferential and predictive statistics.

Inferential statistics demonstrated that there exist enormous differences and associations of the demographic and professional variables of gender, level of education, and major of study concerning familiarity with machine learning. These results prove the value of the scholarly background and the use of interdisciplinary training in the further application of machine learning methods in cancer research. The fact that familiarity with machine learning is the most pivotal correlate with the use of datasets on cancer cells is another significant observation that can be used in order to underscore the significance of technical knowledge in the context of the development of data-based biomedical research studies.

A significant amount of empirical evidence to suggest a positive correlation between machine learning familiarity, knowledge of cancer genomics, multi-omics integration, and subjective belief of personalized therapy was observed, which was discovered under correlation and regression analyses. According to the regression model, all independent variables played significant roles in the prediction of drug resistance, and it is noted that multi-omics integration was the most effective component. This is in response to the fact that, as the quantity of the data sets of the biological data, which is being integrated together with the machine learning models, increases, the magnitude of the predictive accuracy and clinical relevance also increases.

In conclusion, this paper evidences the concept that machine learning-based approaches through the assistance of genomics knowledge and multi-omics data combination can be one of the powerful tools as far as predicting drug resistance genes in human cancer cell lines are concerned. The findings confirm the promise of artificial intelligence in terms of revolutionizing personalized treatment of cancer in question, as they enable one to make therapeutic decisions in a more precise and efficient way. Subsequent research must pursue further development of these methods for future studies by applying them to larger, clinically validated datasets to improve their translational capability in precision oncology.

REFERENCES

1. Adam, G., Rampášek, L., Safikhani, Z., Smirnov, P., Haibe-Kains, B., & Goldenberg, A. (2020). Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precision Oncology*, *4*(1), 19.
2. Ali, M., & Aittokallio, T. (2019). Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophysical reviews*, *11*(1), 31-39.
3. Ataei, A., Majidi, N. S., Zahiri, J., Rostami, M., Arab, S. S., & Rizvanov, A. A. (2021). Prediction of chemoresistance trait of cancer cell lines using machine learning algorithms and systems biology analysis. *Journal of Big Data*, *8*(1), 97.
4. Baptista, D., Ferreira, P. G., & Rocha, M. (2021). Deep learning for drug response prediction in cancer. *Briefings in bioinformatics*, *22*(1), 360-379.
5. Celebi, R., Bear Don't Walk IV, O., Movva, R., Alpsoy, S., & Dumontier, M. (2019). In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Scientific Reports*, *9*(1), 8949.
6. Chawla, S., Rockstroh, A., Lehman, M., Ratther, E., Jain, A., Anand, A., Gupta, A., Bhattacharya, N., Poonia, S., & Rai, P. (2022). Gene expression-based inference of cancer drug sensitivity. *Nature communications*, *13*(1), 5680.
7. Chen, Y., & Zhang, L. (2022). How much can deep learning improve the prediction of the responses to drugs in cancer cell lines? *Briefings in bioinformatics*, *23*(1), bbab378.
8. Chiu, Y.-C., Chen, H.-I. H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.-J., Huang, Y., & Chen, Y. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Medical Genomics*, *12*(Suppl 1), 18.
9. Deelder, W., Christakoudi, S., Phelan, J., Benavente, E. D., Campino, S., McNerney, R., Palla, L., & Clark, T. G. (2019). Machine learning predicts Mycobacterium tuberculosis drug resistance from whole genome sequencing data. *Frontiers in Genetics*, *10*, 922.
10. Fan, K., Cheng, L., & Li, L. (2021). Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Briefings in bioinformatics*, *22*(6), bbab271.
11. Firoozbakht, F., Yousefi, B., & Schwikowski, B. (2022). An overview of machine learning methods for monotherapy drug response prediction. *Briefings in bioinformatics*, *23*(1), bbab408.
12. Gao, Y., Lyu, Q., Luo, P., Li, M., Zhou, R., Zhang, J., & Lyu, Q. (2021). Applications of machine learning to predict cisplatin resistance in lung cancer. *International journal of general medicine*, 5911-5925.
13. Güvenç Paltun, B., Kaski, S., & Mamitsuka, H. (2021). Machine learning approaches for drug combination therapies. *Briefings in bioinformatics*, *22*(6), bbab293.
14. He, D., Liu, Q., Wu, Y., & Xie, L. (2022). A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening. *Nature Machine Intelligence*, *4*(10), 879-892.
15. Honrado, C., Salahi, A., Adair, S. J., Moore, J. H., Bauer, T. W., & Swami, N. S. (2022). Automated biophysical classification of apoptotic pancreatic cancer cell subpopulations by using machine learning approaches with impedance cytometry. *Lab on a Chip*, *22*(19), 3708-3720.
16. Kim, Y. R., Kim, D., & Kim, S. Y. (2019). Prediction of acquired taxane resistance using a personalized pathway-based machine learning method. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, *51*(2), 672-684.
17. Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., Ma, J., & Ideker, T. (2020). Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer cell*, *38*(5), 672-684. e676.
18. Lewis, J. E., & Kemp, M. L. (2021). Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. *Nature communications*, *12*(1), 2700.

19. Liu, J., Lu, J., Li, W., Mao, W., & Lu, Y. (2022). Machine learning screens potential drugs targeting a prognostic gene signature associated with proliferation in hepatocellular carcinoma. *Frontiers in Genetics*, *13*, 900380.
20. Liu, R., Sun, M., Zhang, G., Lan, Y., & Yang, Z. (2019). Towards early monitoring of chemotherapy-induced drug resistance based on single-cell metabolomics: Combining single-probe mass spectrometry with machine learning. *Analytica chimica acta*, *1092*, 42-48.
21. Lu, T.-P., Kuo, K.-T., Chen, C.-H., Chang, M.-C., Lin, H.-P., Hu, Y.-H., Chiang, Y.-C., Cheng, W.-F., & Chen, C.-A. (2019). Developing a prognostic gene panel of epithelial ovarian cancer patients using a machine learning model. *Cancers*, *11*(2), 270.
22. Malik, V., Kalakoti, Y., & Sundar, D. (2021). Deep learning-assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC genomics*, *22*(1), 214.
23. Mucaki, E. J., Zhao, J. Z., Lizotte, D. J., & Rogan, P. K. (2019). Predicting responses to platinum chemotherapy agents with biochemically-inspired machine learning. *Signal transduction and targeted therapy*, *4*(1), 1.
24. Nguyen, L. C., Naulaerts, S., Bruna, A., Ghislat, G., & Ballester, P. J. (2021). Predicting cancer drug response in vivo by learning an optimal feature selection of tumour molecular profiles. *Biomedicines*, *9*(10), 1319.
25. Nguyen, T. K., Le Nguyen, T. N., Nguyen, K., Nguyen, H. V. T., Tran, L. T. T., Ngo, T. X. T., Pham, P. T. V., & Tran, M. H. (2022). Machine learning-based screening of MCF-7 human breast cancer cells and molecular docking analysis of essential oils from Ocimum basilicum against breast cancer. *Journal of Molecular Structure*, *1268*, 133627.
26. Ogunleye, A. Z., Piyawajanusorn, C., Gonçalves, A., Ghislat, G., & Ballester, P. J. (2022). Interpretable machine learning models to predict the resistance of breast cancer patients to doxorubicin from their microRNA profiles. *Advanced Science*, *9*(24), 2201501.
27. Qiu, K., Lee, J., Kim, H., Yoon, S., & Kang, K. (2021). Machine learning-based anti-cancer drug response prediction and search for predictor genes using cancer cell line gene expression. *Genomics & informatics*, *19*(1), e10.
28. Qureshi, R., Basit, S. A., Shamsi, J. A., Fan, X., Nawaz, M., Yan, H., & Alam, T. (2022). Machine learning-based personalized drug response prediction for lung cancer patients. *Scientific Reports*, *12*(1), 18935.
29. Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T. J., Piha-Paul, S., Zhou, H., & Kardala, E. (2019). A deep learning framework for predicting response to therapy in cancer. *Cell reports*, *29*(11), 3367-3373. e3364.
30. Sharifi-Noghabi, H., Jahangiri-Tazehkand, S., Smirnov, P., Hon, C., Mammoliti, A., Nair, S. K., Mer, A. S., Ester, M., & Haibe-Kains, B. (2021). Drug sensitivity prediction from cell line-based pharmacogenomics data: guidelines for developing machine learning models. *Briefings in bioinformatics*, *22*(6), bbab294.
31. Sharma, A., & Rani, R. (2020). An ensemble machine learning framework for drug sensitivity prediction. *IET Systems Biology*, *14*(1), 39-46.
32. Shen, Y., & Yan, Z. (2021). Systematic prediction of drug resistance caused by transporter genes in cancer cells. *Scientific Reports*, *11*(1), 7400.
33. Sinkala, M., Mulder, N., & Martin, D. (2020). Machine learning and network analyses reveal disease subtypes of pancreatic cancer and their molecular characteristics. *Scientific Reports*, *10*(1), 1212.
34. Sui, Q., Chen, Z., Hu, Z., Huang, Y., Liang, J., Bi, G., Bian, Y., Zhao, M., Zhan, C., & Lin, Z. (2022). Cisplatin resistance-related multi-omics differences and the establishment of machine learning models. *Journal of translational medicine*, *20*(1), 171.
35. Tan, M., Özgül, O. F., Bardak, B., Ekşioğlu, I., & Sabuncuoğlu, S. (2019). Drug response prediction by ensemble learning and drug-induced gene expression signatures. *Genomics*, *111*(5), 1078-1088.
36. Torkamannia, A., Omidi, Y., & Ferdousi, R. (2022). A review of machine learning approaches for drug synergy prediction in cancer. *Briefings in bioinformatics*, *23*(3), bbac075.
37. Wang, C., Lye, X., Kaalia, R., Kumar, P., & Rajapakse, J. C. (2021). Deep learning and multi-omics approach to predict drug responses in cancer. *BMC bioinformatics*, *22*(Suppl 10), 632.
38. Yasir, M., Karim, A. M., Malik, S. K., Bajaffer, A. A., & Azhar, E. I. (2022). Application of decision-tree-based machine learning algorithms for prediction of antimicrobial resistance. *Antibiotics*, *11*(11), 1593.
39. Yu, L., Zhou, D., Gao, L., & Zha, Y. (2021). Prediction of drug response in multilayer networks based on fusion of multiomics data. *Methods*, *192*, 85-92.
40. Zhang, H., Zhang, N., Wu, W., Zhou, R., Li, S., Wang, Z., Dai, Z., Zhang, L., Liu, Z., & Zhang, J. (2022). Machine learning-based tumor-infiltrating immune cell-associated lncRNAs for predicting prognosis and immunotherapy response in patients with glioblastoma. *Briefings in bioinformatics*, *23*(6), bbac386.
41. Zhang, T., Zhang, L., Payne, P. R., & Li, F. (2020). Synergistic drug combination prediction by integrating multiomics data in deep learning models. In *Translational bioinformatics for therapeutic development* (pp. 223-238). Springer.
42. Zhu, E. Y., & Dupuy, A. J. (2022). Machine learning approach informs biology of cancer drug response. *BMC bioinformatics*, *23*(1), 184.
43. Zuo, Z., Wang, P., Chen, X., Tian, L., Ge, H., & Qian, D. (2021). SWnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC bioinformatics*, *22*(1), 434.