

RESEARCH ARTICLE

Editorial Process: Submission:18/01/2026 Acceptance:13/10/2026 Published:24/10/2026

Machine Learning-Based Single-Cell RNA-seq Analysis in Drug Response: A Hybrid Multi-Stage Gene Selection and Soft Voting Approach

Mahmood Khalsan^{1,2*}, Alia H. Al-Khafaji¹, Fawaz Al-Alloosh¹

Abstract

Objective: Oral squamous cell carcinoma (OSCC) is the most common type of oral cancer, and cisplatin is widely used as a first-line chemotherapeutic agent. However, resistance to chemotherapy remains a significant challenge in the treatment of OSCC. The development of chemoresistance is driven by various molecular mechanisms and is further complicated by interactions within the tumor microenvironment (TME). Recent advances in high-throughput sequencing technologies, such as single-cell RNA sequencing (scRNA-seq), have enable the analysis of gene expression at the single-cell level. This approach provides deeper insights into complex intratumoral heterogeneity and facilitates the development of more targeted and effective therapies. Nevertheless, the large volume of data generated by scRNA-seq presents challenges for traditional statistical methods, a limitation that has been effectively addressed through advancements in machine learning. **Methods:** In this study, we developed a hybrid multi-stage gene selection framework combined with a soft voting classifier to predict cancer drug response at the single-cell level by integrating large-scale bulk cell line data. A total of 100 significant genes were selected as features for training the developed soft voting classifier. **Result:** The results demonstrated the effectiveness of the proposed machine learning model, achieving an accuracy of 97.9%, precision of 98.9%, recall of 96.9%, and an F1-score of 97.9% in predicting cisplatin response in patient-derived OSCC Cells at the single-cell level. **Conclusion:** This study highlights the effectiveness of combining single-cell RNA sequencing with machine learning to predict cisplatin response in oral squamous cell carcinoma. The proposed hybrid multi-stage gene selection and soft voting classifier successfully captured intratumoral heterogeneity at the single-cell level. Using 100 selected genes, the model achieved high predictive performance, demonstrating strong accuracy and robustness. These findings emphasize the value of machine learning in analyzing high-dimensional scRNA-seq data. Overall, this approach provides a promising foundation for precision medicine and improved therapeutic decision-making in OSCC.

Keywords: Single-Cell RNA-Seq, Cisplatin Sensitivity and Resistance, OSCC, Machine learning, Gene Selection

Asian Pac J Cancer Prev, 27 (10), 3709-3715

Introduction

The seventh most prevalent form of cancer worldwide is head and neck cancer, which arises from the anatomic sites that compose the upper aero digestive tract [1, 2]. 90% of head and neck cancer is accounted for squamous cell carcinoma, that includes oral squamous cell carcinoma (OSCC) [2]. OSCC, a solid tumor, is the most prevalent form of cancer found in the oral cavity [3]. It develops in the oral mucosa where it affects the epithelial cells in the oral cavity due to accumulation of several genetic mutations within the cells. OSCC results in disfiguration and functional impairments, such as issues with speech and taste, that impact the quality of life of patients [4].

In recent years, advancements in next-generation sequencing (NGS) technologies have emerged as powerful

tools in cancer research. However, most of the traditional methods, including RNA sequencing, consider tumor cells as single units, which analyze bulk RNA samples and represent an average of gene expression across many cells, despite the intra-tumoral heterogeneity [5, 6]. On the other hand, single-cell profiling methods, such as single-cell RNA sequencing (scRNA-seq), enable a more detailed understanding of intra-tumoral heterogeneity which allow the identification of gene expression profiles of individual cells within the tumor. By analyzing individual cells and their complex network of mechanisms within the tumor microenvironment (TME), insights into the complex cellular landscape, which contribute to tumor progression, immune suppression, and drug resistance is provided [5, 7].

Single-cell profiling methods may contribute to

¹Scientific Department, Warith International Cancer Institute, Karbala, 56001, Iraq. ²College of Science, Information Technology, The University of Warith Al-anbiyaa, Karbala, 56001, Iraq. *For Correspondence: mahmood.jasim@uowa.edu.iq.

developing more effective anticancer treatments, especially for cancer types exhibiting chemo resistance. Chemo resistance significantly affects the efficacy of cisplatin (cis-diaminedichloroplatinum) in OSCC treatment, a chemotherapeutic agent widely used as a first-line drug for OSCC. Resistance to cisplatin occurs through various mechanisms, including molecular and epigenetic changes, hypoxia and immunosuppression within the TME and disruptions in programmed cell death (PCD) [8]. Understanding the heterogeneity of the tumor and at a single-cell level can provide valuable insights into how these mechanisms operate, enabling the development of more targeted and effective therapies to overcome chemoresistance and improve treatment outcomes for OSCC patients [9, 10].

To analyze the massive data generated by scRNA-seq, traditional linear statistical methods are challenging to be used [7]. Existing feature selection methods in healthcare often fall short due to the complexity of medical data. They struggle with high-dimensional feature sets and may not effectively account for irrelevant or redundant features, leading to suboptimal classification performance [10]. Machine learning methods, particularly deep learning, are increasingly being applied for many fields to analyze scRNA-seq data. Deep learning tools use neural networks with many layers to automatically detect complex patterns in large datasets [11, 12]. These tools are particularly effective for handling high-dimensional data that poses significant challenges in terms of storage, processing, and analysis due to its complexity and volume. Specifically, deep learning can help uncover relationships between gene expression profiles and tumor characteristics, such as cancer subtypes, tumor heterogeneity, and resistance to treatments. For instance, deep learning models can classify cancer cells based on their gene expression patterns, allowing for a better understanding of how different subpopulations within a tumor behave [5, 12, 13].

This paper presents the following key contributions

1. Development of a hybrid multi-stage gene selection framework to address the high dimensionality of gene expression data, improve classifier performance, reduce overfitting, and enhance computational efficiency by minimizing training time.
2. Design of a soft voting classifier for predicting cisplatin sensitivity or resistance, enabling more informed selection of effective chemotherapy treatments.
3. Establishment of a state-of-the-art automated pipeline that integrates data pre-processing, gene selection, and classification into a unified framework for efficient and streamlined analysis.

Materials and Methods

Recent advancements in hybrid machine learning models, such as the MP+CNN+BiLSTM framework demonstrated highlight the potential of integrating multiple architectures to enhance prediction accuracy, which could be adapted for analyzing gene expression data in personalized medicine [13]. Based on that the proposed model begins with pre-processing, followed by

the development of a Hybrid Multi-Stage Gene Selection and Soft Voting Classifier (HMSGSSVC) model as shown in Figure 1. This approach aims to identify the most significant genes associated with drug sensitivity and resistance. After selecting the relevant genes, a soft voting classifier is applied to train the model and make predictions based on these selected genes.

The proposed model can be divided into three stages as follows

Pre-processing

The data was pre-processed to align with formats compatible with AI systems. This process includes handling missing and eliminating duplicate entries to ensure data integrity and enhance its suitability for analysis.

Hybrid Multi-Stage Gene Selection (HMSGSS)

The utility of combining feature selection, a strategy that can be adapted for analyzing gene expression data in personalized medicine [14]. Two advanced feature selection methods, Minimum Redundancy Maximum Relevance (MRMR) and Chi2-squared, were applied to select top 100 genes that highly impact the drug response. Thereafter, fusion gene selection method was applied to select the genes that was selected by the two techniques. Genes that have been selected by two different techniques are more important than others.

1. Minimum Redundancy Maximum Relevance

MRMR is a feature selection method designed to identify discriminative features for training classifier models. Initially developed for classifying DNA microarray data, it selects features based on their high relevance to the target variable and low redundancy with other features. Features are ranked using the minimal-redundancy-maximal-relevance criteria [15]. Relevance is calculated utilizing the F-statistic for continuous features or mutual information for discrete features, while redundancy is measured using Pearson correlation for continuous features or mutual information for discrete features.

2. Chi2-squared Test

The Chi-squared test is a statistical method commonly employed to assess the independence between two variables. In the context of feature selection, this technique calculates the Chi-squared statistic for each feature relative to the target variable (class label) [16]. Features are subsequently ranked based on their Chi-squared scores, and those with the highest scores are selected for further analysis [17].

Integrating these two techniques enhance the overall process. Chi-squared selected features are statistically independent and highly relevant to the target variable but does not address redundancy between features. MRMR complements this by selecting features that balance high relevance with low redundancy, ensuring that the chosen features provide unique and non-overlapping information. By applying Chi-squared as an initial filter to eliminate irrelevant features and then refining the selection with

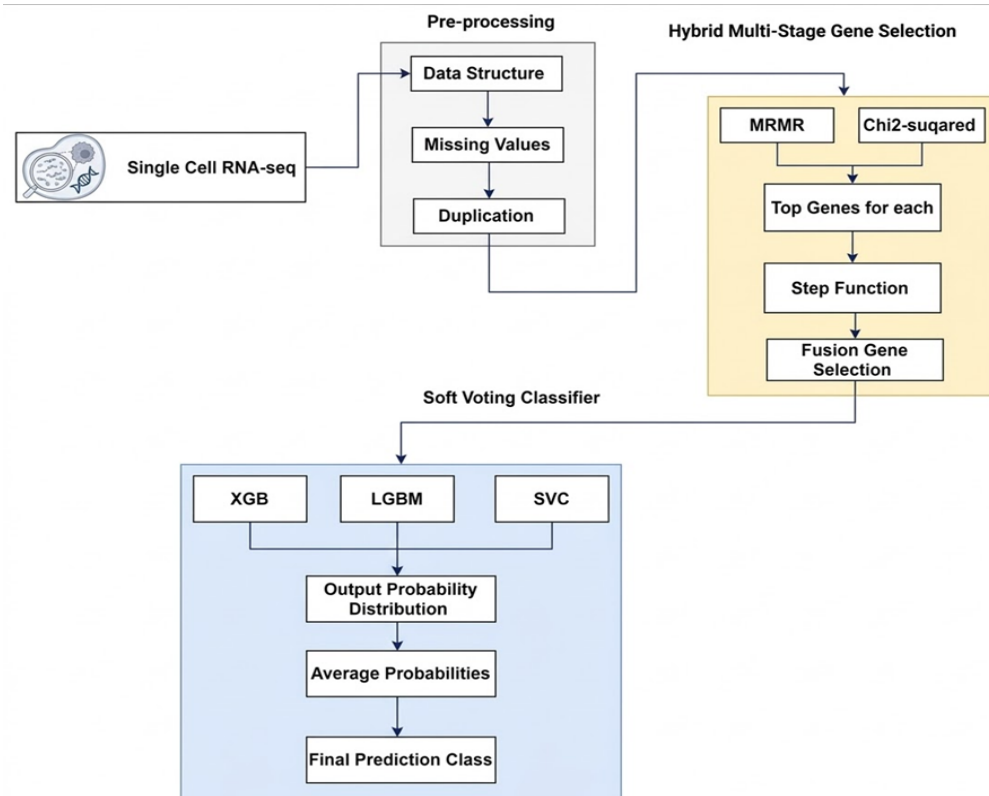


Figure 1. The Proposed Model Framework

MRMR, the resulting feature set is both informative and optimized for modelling, improving predictive performance and computational efficiency.

Soft Voting Classifier

Three different classifier approaches were used to obtain the predicted class probability for each class in each classifier. These probabilities were then averaged across the classifiers to obtain a single probability for each class. Subsequently, the classes were ranked based on their average probabilities in descending order. On the other hand, the class with the highest average probability was chosen as the final predicted class that will be compared to the actual class. This method is known as the voting classifier that is used to enhance the overall accuracy and robustness of the model.

1. eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGB) is a powerful technique that can be used for both classification and regression tasks. It is highly efficient and scalable with classification which predict accurately with structured data. Additionally, it can reduce overfitting by regularization. XGB is commonly employed in competitive ML due to its speed and accuracy [18, 19].

2. Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is an advanced algorithm designed for speed and accuracy, particularly with huge datasets. It uses a histogram-based method to enhance training time and memory usage. It provides high performance with large datasets without

sacrificing accuracy [20].

3. Support Vector Classifier

Support Vector Classifier (SVC) is commonly employed for classification and regression purpose, with its key strength being in classification due to their high performance. SVC attempts to find an optimal hyperplane in an n-dimensional space to separate data points into different classes effectively [21]. However, SVC has limitations that impact its performance in some scenarios. For example, SVC struggles with large datasets compared to smaller ones and underperforms when datasets contain noisy data, such as overlapping target classes [22]. Using soft voting for three different ML approaches, XGBoost, LightGBM, and SVC, has major contribution for high dimensional datasets like gene expression data. SVC uses optimizing decision boundaries in high dimensional spaces, while the other two techniques effectively handle feature interactions and prioritize relevant features through their adaptive nature. This results in a robust and accurate classifier, well-suited for complex tasks such as disease classification and biomarker discovery.

Cross-Validation

Cross-validation (CV) is a statistical method widely employed in ML approaches to reduce overfitting across different classifier models. The k-fold cross-validation method train a model on multiple subsets of the dataset rather than one single training set [23]. The dataset is partitioned into k folds, with the model trained iteratively on each fold. This process promotes generalization, indicating a more reliable and robust model. Furthermore,

it provides a clearer evaluation of the algorithm's predictive performance.

Evaluation performance

There are four evaluation performance measurement methods used to evaluate the performance of any proposed model. The evaluation performance methods are explained as follows:

1. Accuracy (AC) in machine learning refers to the proportion of correctly predicted instances to the total number of instances. Mathematically, it is calculated as below [24]:

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

2. Precision (Pre) is defined as the ratio of correctly predicted positive outcomes to the total predicted positive outcomes [24]:

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

3. Recall is described as the proportion of correctly retrieved instances to the total number of relevant instances [25]:

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

4. F1-score (F1) is the harmonic mean of precision and recall, ranging from 0 (the worst) to 1 (the best) [25]. It combines precision and recall in a single metric that evaluate a classifier's performance.:

$$F1 = 2 * \frac{Pre * Rec}{Pre + Rec} \quad (4)$$

Results

Dataset

In this study, we used a high-throughput single-cell RNA-seq dataset to profile the transcriptomes of cisplatin-sensitive, resistant, and drug-treated cells derived from OSCC samples that were isolated from both primary and metastatic sites. This data was collected from GEO (GSE117872) free public datasets. The datasets were collected by Yiqun Elaine Cao from Singapore [26-28].

The datasets include 1048 samples and 22744 genes.

Split data

The study employed stratified 5-fold cross-validation to preserve class distribution across folds, with splitting performed at the patient level. Integration of bulk RNA-seq and scRNA-seq datasets was conducted exclusively within the training portion of each fold, ensuring that no information from the validation set was used during the integration process. Furthermore, hyperparameter tuning was carried out an inner cross-validation loop, thereby minimizing the risk of optimistic bias.

Significant Gene Selection

One hundred significant genes were identified using the proposed HMSGs method and used as input for training the classifier models. The results demonstrate that this gene selection approach substantially enhances classifier performance. Specifically, the accuracy of the developed voting classifier increased to 97.9% when using the HMSGs method, compared to 95.9% with MRMR and 94.6% with the Chi-squared method applied individually. In contrast, traditional machine learning approaches without any gene selection exhibited considerably lower accuracies, ranging from 49.6% to 87.7%, as summarized in Table 1.

Machine Learning Performance

The proposed model achieved highest accuracy compared to five classifier techniques and using multi-stage gene selection method accomplished higher performance compared to single gene selection method. Table 1 presents a comparison of classification performance with and without the application of gene selection methods. The highest accuracy achieved without gene selection among the five traditional classifiers is 87.7%, while the proposed soft voting classifier achieved a superior accuracy of 93%. Additionally, the table compares the performance of the proposed classifier when individual gene selection methods were used versus when the proposed multi-stage gene selection method was applied. The results indicate that the multi-stage gene selection method outperforms the individual feature gene approaches. Furthermore, employing gene selection significantly reduces the high dimensionality of the data,

Table 1. Comparing the Proposed Model Outcomes to Other Classifier Techniques

N- Genes	Gene Selection Method	Classifier	Accuracy	Precision	Recall	F1-score
22744	No	DT	49.6	57	51	51
		GNB	47.9	55.4	42.8	45
		KNN	78.4	80.6	81.6	79.7
		SVM	87.7	88.6	89.6	88.4
		MLP	81.6	88.6	85.8	82.6
22744	No	Voting Classifier	93	92	93	92.5
100	MRMR	Voting Classifier	95.9	96	95.9	96
100	Chi2-squared	Voting Classifier	94.6	94.6	94.6	94.6
168	Union (MRMR +Chi2)	Voting Classifier	95.4	95.2	95.6	95.3
100	Hybrid Multi-Stage Gene Selection	Voting Classifier	97.9	98.9	96.9	97.9

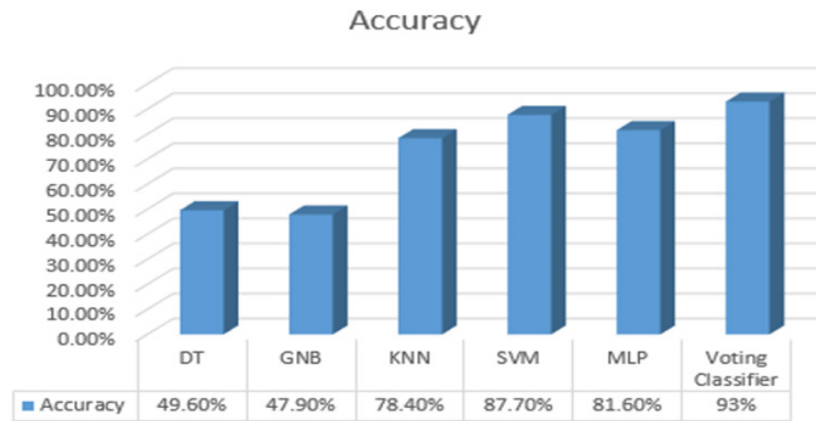


Figure 2. Comparing Classifier Accuracy Score without Using Gene Selection Method

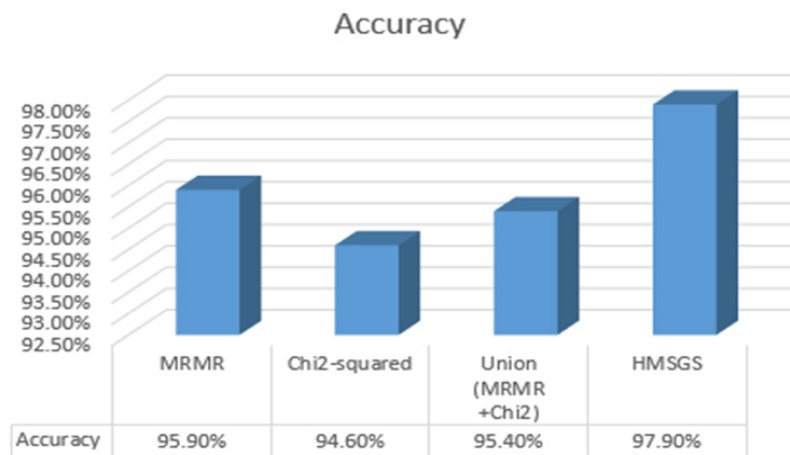


Figure 3. Comparing the Developed Voting Classifier Model Using Different Gene Selection Techniques against the Developed MSGSS

which in turn minimizes overfitting and decreases the time required for classifier training.

In summary, Figures 2 and 3 highlight the impact of gene selection on classifier performance, comparing results with and without gene selection and showing the individual contributions of each method. The findings confirm that the proposed HMSGS method, combined with the voting classifier, outperforms all other approaches, achieving a maximum accuracy of 97.9%.

Discussion

Cisplatin is the most common used chemotherapy agent as a first-line medication in the treatment of OSCC patients. However, chemotherapy resistance is common in OSCC patients, making the treatment with cisplatin not always as effective as expected. Therefore, applying high-throughput sequencing technologies may enhance our understanding of the intra-tumoral environment and help in the development of more effective chemotherapeutic agents [29].

However, the complexity inherent in the data generated by high-throughput sequencing technologies poses a significant challenge when employing traditional statistical methods. These conventional approaches

often struggle to handle the vast amount of information, particularly when dealing with large, multidimensional datasets that include intricate biological patterns. This limitation has driven the development of machine learning techniques, which are better equipped to process and analyze such complex data efficiently.

Machine learning algorithms, with their ability to uncover hidden patterns and model non-linear relationships, offer a promising alternative to traditional statistical methods, enabling more accurate and robust insights from high-throughput sequencing data. We developed a Hybrid Multi-Stage Gene Selection and Soft Voting Approach classifier and trained it on known cell types to accurately classify Single cell RNA-seq profiles of primary, metastatic, sensitive, drug-resistant cells. The proposed model consistently performed as the top-ranked classifier across five classifier approaches performance metrics. Additionally, the proposed model showed better results compared to other published work. By using Deep Transfer Learning models, the authors achieved 85 % and 92% in accuracy and F1-score, respectively, [30] with the same dataset that was used in our model (GSE117872). Our model did achieve 97.9% for accuracy and f1-score. Another study by [31] used a Deep Transfer Learning (DTL) model and Relief to select the top 200 genes that

were used as identifiers for training the DTL model. The results from this model showed an accuracy of 60%, while our model had an accuracy of 97%. To the best of our knowledge, very limited numbers of studies have used machine learning to analyze single-cell RNA-seq data for drug response.

In conclusion, this study has developed a Multi-Stage Gene Selection method to identify the most significant genes associated with cisplatin-sensitive, resistant, and drug-treated cells in OSCC patients. Additionally, an advanced soft voting classifier was designed to predict drug sensitivity or resistance rely on the selected genes. Importantly, the model demonstrates the potential to predict a patient's drug response before treatment, using the identified gene signatures, paving the way for personalized therapeutic strategies. The developed HMSGs did select 100 top genes that highly associated with drug response. These genes were used as identifiers to train soft voting classifier method. The findings of this study demonstrated that the proposed model (HMSGSSVC) outperformed the other existing classifier models with an accuracy of 97.9 %, precision of 98.9 % precision, recall of 96.9 % and f1-score of 97.9 %. Although the proposed model did achieve promising results, the study has some limitations that may be addressed in future research. The model was evaluated using only one dataset; in the future, it could be tested on a variety of datasets to ensure its robustness. Additionally, the model currently utilizes only one type of data gene expression profiles. It can be extended to incorporate clinical, radiological, and other data types to assess its adaptability and ensure accurate, comprehensive results.

Author Contribution Statement

All authors contributed equally in this study.

Acknowledgements

Funding Statement

This research was conducted without external financial support and was fully funded by the author.

Ethical Declaration

This study did not require ethical approval from an institutional review board or ethics committee, as it exclusively utilized freely available, de-identified public datasets. All data were accessed and analyzed in accordance with the original data providers' ethical guidelines and terms of use.

Data Availability Statement

The data analyzed in this study are publicly available from established open-access repositories. No new data were generated during the course of this research.

Conflict of Interest

The author declares that there are no commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Dunn LA, Ho AL, Pfister DG. Head and Neck Cancer: A Review. *JAMA*. 2026;335(6):531-41. <https://doi.org/10.1001/jama.2025.21733>.
- Mody MD, Rocco JW, Yom SS, Haddad RI, Saba NF. Head and neck cancer. *Lancet*. 2021;398(10318):2289-99. [https://doi.org/10.1016/s0140-6736\(21\)01550-6](https://doi.org/10.1016/s0140-6736(21)01550-6).
- Tan Y, Wang Z, Xu M, Li B, Huang Z, Qin S, et al. Oral squamous cell carcinomas: state of the field and emerging directions. *Int J Oral Sci*. 2023;15(1):44. <https://doi.org/10.1038/s41368-023-00249-w>.
- Gharat SA, Momin M, Bhavsar C. Oral squamous cell carcinoma: current treatment strategies and nanotechnology-based approaches for prevention and therapy. *Crit Rev Ther Drug Carrier Syst*. 2016;33(4):363-400. <https://doi.org/10.1615/CritRevTherDrugCarrierSyst.2016016272>.
- Wu Z, Lawrence PJ, Ma A, Zhu J, Xu D, Ma Q. Single-cell Techniques and Deep Learning in Predicting Drug Response. *Trends Pharmacol Sci*. 2020;41(12):1050-65. <https://doi.org/10.1016/j.tips.2020.10.004>.
- Zhu B, Wang Y, Ku LT, van Dijk D, Zhang L, Hafler DA, et al. scNAT: A deep learning method for integrating paired single-cell RNA and t cell receptor sequencing profiles. *Genome Biol*. 2023;24(1):292. <https://doi.org/10.1186/s13059-023-03129-y>.
- Tirosh I, Suvà ML. Deciphering human tumor biology by single-cell expression profiling. *Annu Rev Cancer Biol*. 2019;3:151-66.
- Cheng Y, Li S, Gao L, Zhi K, Ren W. The Molecular Basis and Therapeutic Aspects of Cisplatin Resistance in Oral Squamous Cell Carcinoma. *Front Oncol*. 2021;11:761379. <https://doi.org/10.3389/fonc.2021.761379>.
- Kim KT, Lee HW, Lee HO, Kim SC, Seo YJ, Chung W, et al. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol*. 2015;16(1):127. <https://doi.org/10.1186/s13059-015-0692-3>.
- Gu Z, Heng Y, Fan R, Luo J, Ju L. Single-cell RNA sequencing reveals cellular and molecular heterogeneity in extensive-stage small cell lung cancer with different chemotherapy responses. *Cancer Cell Int*. 2025;25(1):157. <https://doi.org/10.1186/s12935-025-03785-z>.
- Abadi F, Ghanimi HMA, Alarni AD, Soliman NF, El-Shafai W, Abbas A, et al. Chimp optimization algorithm based feature selection with machine learning for medical data classification. *Comput Syst Sci Eng*. 2023;47:2791-814. <https://doi.org/10.32604/csse.2023.038762>.
- Abadi F, Ghanimi HMA, Sadeeq MAM, Alkhayyat A, Kareem ZH, Mahmood SN, et al. Hybrid deep learning enabled load prediction for energy storage systems. *Comput Mater Contin*. 2023;75(2):3359-74. <https://doi.org/https://doi.org/10.32604/cmc.2023.034221>.
- Erfanian N, Heydari AA, Feriz AM, Iañez P, Derakhshani A, Ghasemigol M, et al. Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed Pharmacother*. 2023;165:115077. <https://doi.org/10.1016/j.biopha.2023.115077>.
- Qi G, Wang L, Nan M, Fu Y, Ren Q, Zhang Y, et al. scXDR: Drug response prediction across single-cell datasets via heterogeneous network transfer learning. *Commun Biol*. 2026;9(1):139. <https://doi.org/10.1038/s42003-025-09418-5>.
- Ghanimi HM, Sengan S, Sadu VB, Kaur P, Kaushik M, Alroobaea R, et al. An open-source MP+ CNN+ BiLSTM model-based hybrid model for recognizing sign language on smartphones. *Int J Syst Assur Eng Manag*. 2024;15(8):3794-

- 806.
16. Ghanimi HMA, Alagusundari N, Banda J, Sengan S, Somasekar J, Thomas J, et al. Linear discriminant analysis-based deep learning algorithms for numerical character handwriting recognition. *J Auton Intell.* 2024;7:1621. <https://doi.org/10.32629/jai.v7i5.1621>.
 17. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics.* 2017;18(1):9. <https://doi.org/10.1186/s12859-016-1423-9>.
 18. Zhai Y, Song W, Liu X, Liu L, Zhao X. A chi-square statistics based feature selection method in text classification. In 2018 IEEE 9th International conference on software engineering and service science (ICSESS) 2018 Nov 23 (pp. 160-163). IEEE.
 19. Ray S, Alshouli K, Roy A, AlGhamdi A, Agrawal DP. Chi-squared based feature selection for stroke prediction using AzureML. In 2020 Intermountain Engineering, Technology and Computing (IETC) 2020 Oct 2 (pp. 1-6). IEEE.
 20. Chen M, Liu Q, Chen S, Liu Y, Zhang CH, Liu R. XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access.* 2019;7:13149-58. <https://doi.org/10.1109/ACCESS.2019.2893448>.
 21. Almgoshi H, Mithra C, Ratnam ASK, Shanmugam S, Saravanan V, Marapelli B. Data visualisation models for analytics use artificial intelligence to predict diabetes in women. *J Mach Comput.* 2025;5(1):551-60. <https://doi.org/10.53759/7669/jmc202505043>.
 22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. 2017.
 23. Maharjan A. Machine Learning Approach for Predicting Cancer Using Gene Expression (2020) (UNLV Theses, Dissertations & Capstones No. 3922).
 24. Farooq MA, Corcoran P, Rotariu C, Shariff W. Object detection in thermal spectrum for advanced driver-assistance systems (adas). *IEEE Access.* 2021;9:156465-81. <https://doi.org/10.1109/ACCESS.2021.3129150>.
 25. Anwar Ul Hassan CH, Khan MS, Shah MA. Comparison of machine learning algorithms in data classification. *International Conference on Automation and Computing (ICAC).* 2018:1-6. <https://doi.org/10.23919/IconAC.2018.8748995>.
 26. Khalsan M, Mu M, Al-Shamery E, Ajit S, Machado L, Opopu Agyeman M. A novel fuzzy classifier model for cancer classification using gene expression data. *IEEE Access.* 2023;11:115161-78. <https://doi.org/10.1109/ACCESS.2023.3325381>.
 27. Khalsan M, Mu M, Al-Shamery ES, Machado L, Agyeman MO, Ajit S. Enhancing Cancer Classification Through the Development of a Fuzzy Gene Selection-Wrapper Plus Method. In 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET) 2023 Sep 12 (pp. 39-44). IEEE.
 28. Sharma A, Cao EY, Kumar V, Zhang X, Leong HS, Wong AML, et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat Commun.* 2018;9(1):4931. <https://doi.org/10.1038/s41467-018-07261-3>.
 29. Ravasio A, Myaing MZ, Chia S, Arora A, Sathe A, Cao EY, et al. Single-cell analysis of EphA clustering phenotypes to probe cancer cell heterogeneity. *Commun Biol.* 2020;3(1):429. <https://doi.org/10.1038/s42003-020-01136-4>.
 30. Suphavitai C, Chia S, Sharma A, Tu L, Da Silva RP, Mongia A, et al. Predicting heterogeneity in clone-specific therapeutic vulnerabilities using single-cell transcriptomic signatures. *Genome Med.* 2021;13(1):189. <https://doi.org/10.1186/s13073-021-01000-y>.
 31. Ghosh S. Cisplatin: the first metal based anticancer drug. *Bioorg Chem.* 2019;88:102925. <https://doi.org/10.1016/j.bioorg.2019.102925>.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.