DOI: 10.6977/IJoSI.202412_8(4).0004
Jayakrishnan R, S. Meera, etc./Int. J. Systematic Innovation, 8(4), 46-66(2024)

Analysis of deep actor-critic methods for classifying cancer subtypes through gene expression

Jayakrishnan R[1], S. Meera[2]*
[1]*Department of Computer Science and Engineering
Vels Institute of Science, Technology and Advanced Studies
Chennai, Tamil Nadu.
[2]*Department of Computer Science and Engineering
Vels Institute of Science, Technology and Advanced Studies
Chennai, Tamil Nadu.
[2]*smeera2134@gmail.com

(Received 23 February 2024; Final version received 1 July 2024; Accepted 3 September 2024)

Abstract

The word "cancer" denotes a syndromes that can spread to various bodily areas and are brought on by abnormal cell proliferation. After cardiovascular illnesses, according to the World Health Organisation (WHO), cancer is the second largest cause of death in the world. To better understand molecular processes behind various cancer subdivisions, cancer categorization depends on gene expression information is essential. Conventional machine learning methods have proven helpful in this situation, but new approaches are needed for accurate and understandable categorisation due to the difficulty and dimensionality of gene expression datasets. In this article, we analyse various methods for multiclass categorising cancer subtypes using deep structured reinforcement learning (DSRL). Our methodology addresses several significant issues in cancer subtype classification by combining the strength of deep neural networks with reinforcement learning. In this research, seven different gene expression datasets are utilised to classify the cancer subtype. We also used different classification approaches in Python for the same dataset to perform a comparative study. Deep reinforcement learning for cancer subtype classification improves the accuracy of gene expression data by integrating intricate data patterns, enabling customised therapies, and expanding the field of precision medicine research. The analysis reveals that the newly suggested model exceeds the contemporary state-of-the-art classifiers, achieving the highest accuracy across all seven datasets, ranging from 55% to 100%, while attaining the lowest loss, which varies between 0.02 and 0.11. This work offers a viable method for classifying cancer subtypes into many categories using gene expression data.

Keywords: *Cancer subtypes, Machine-learning, Deep-learning, Reinforcement learning, Gene expression, Deep neural networks.*

.

1. Introduction

Cancer is a group of diseases in which malignant cells form in the human body due to genetic mutation. When the cells form, they arbitrarily divide to layout throughout the organs, and in rare cases, they can be fatal. After cardiovascular diseases, cancer is the next most significant global cause of death [Miller, etc., 2021]. Gene expression analysis has recently been a critical technique for solving important cancer detection and therapy development experiments. [Munkácsy, etc., 2022][Brewczyński, etc., 2021]. Examining gene expression sheds light on the involvement of specific genes in the genesis and spread of cancer. As a result, changes in gene expression can be

used to detect cancer early and to guide the selection of prospective therapeutic targets.

The technique that occurs when the data in DNA is converted into commands aimed at constructing other compounds or proteins is known as gene expression. It involves the translation of mRNA, which is produced from messenger RNA (DNA), into proteins. Under specific conditions, gene expression examination was used to determine the sequence of genetic changes in a tissue or single cell [Anna, & Monika, 2018]. The DNA transcripts in a tissue or cell sample must be counted to learn which genes have been released and in what amounts.

In the past few years, bioinformatics has grown significantly in importance. It now encompasses a wide range of subjects, from the mathematical simulation of biological patterns and the acquisition of DNA data to the understanding and simulation of life's evolutionary history [Jiang, etc., 2013]. Starting with the first small phage genome, progress has been made toward sequencing 1,000 human genomes, every single one of which is three billion bases long [Rodriguez-Ezpeleta, 2012]. We have contributed to the growth of genomic sequence data during the past few decades. High throughput sequencers have emerged and are a crucial tool in biological research.

The conversion of vast volumes of biological knowledge into precious information has been one of the most essential bioinformatics research disciplines in the age of Big Data. Deep learning (DL) has significantly progressed in several sectors since the early 2000s. As an outcome, universities and companies have strongly emphasised the application of DL in bioinformatics to extract meaningful and valuable information from data [Min, etc., 2017]. Breakthroughs in several domains, including recognition of images, recognition of voices, and natural language processing (NLP), have been made possible by DL, which has developed since the collection of vast amounts of information, the development of similar & disseminated computers, and sophisticated learning algorithms [Cun, etc., 2015]. DL is going to be extensively utilised and advantageous for bioinformatics.

Insights into how differences in genes and regulatory areas affect phenotypic modifications, including characteristics, wellness, and health, have frequently been obtained using ML-based techniques [Lunshof,etc., 2010][Khan, etc., 2021]. Over the last ten years, DL-based algorithms for predicting the shape and function of genomic mechanisms, including promoters, enhancers, or gene sequence levels, have gained popularity. [Bhonde, etc., 2021][Celesti, etc., 2018]. A potent method for studying the cancer transcriptome has emerged: gene expression profiling using DNA microarrays [Tarca, etc., 2006]. The transcriptome is the set of transcripts in a cell at a specific moment and under particular conditions. It represents the genome's functional status. DNA microarrays make it possible to monitor the communication of thousands of genes in a specific cell or tissue concurrently, allowing for studying the transcriptome and measuring changes in different cellular states [Dudoit, etc., 2002].

Despite advancements in cancer categorisation using machine learning and gene expression data, challenges remain, such as limited sample sizes and high gene dimensions. ML systems often use feature-engineering techniques to reduce duplicate information and select ideal features. DL networks have been integrated into workflows to improve performance, and as DL-based approaches usually outperform traditional methods, future gene expression analysis models will likely use DL networks. Due to availability, cost, and privacy concerns, getting sufficiently prominent and representative datasets for classifying cancer subtypes can be complex. Datasets about cancer subtypes frequently exhibit class imbalance, with specific subtypes having much lower representation than others. Reduced accuracy and biased model projections for minority classes may result from this imbalance. To overcome these limitations, this study article analyses the numerous cancer subtypes using machine learning, deep learning, and reinforcement learning approaches, which can also categorise cancer subtypes using various methods. This study's main objective is to analyse the multiclass cancer subtype classification, with a particular emphasis on how well deep structured reinforcement learning (DSRL) works with gene expression data. The contributions of the proposed work are:

- The primary involvement of this research is the analysis of several techniques for multiclass cancer subtype classification using the effectiveness of deep structured reinforcement learning (DSRL).
- To introduce a novel method that significantly enhances performance by combining the capabilities of deep neural networks with reinforcement learning strategies.
- This analysis shows that the newly suggested model exceeds the current state-of-the-art

classifiers, producing outstanding accurateness and loss outcomes.

- These consequences display that the suggested technique can potentially make considerable progress in identifying cancer subtypes using gene expression data.
- The arrangement of this document is as follows: Section 2 summarises the literature review for numerous research relevant to the multiclass classification of cancer subdivision utilising gene expression data. The overall suggested technique for the cancer classification utilising deep structured reinforcement learning is outlined in section 3; results and discussion with comparison are specified in section 4, and section 5 has the conclusion and the references for this research were delivered in the following reference section.

2. Related Works

[Mostavi, etc., 2020] proposed the CNN models for categorising tumour & non-tumour data as cancer or normal. The models 1D-CNN, 2D-Vanilla-CNN, and 2D-Hybrid-CNN were trained and tested utilising data from the Cancer Genome Atlas (TCGA); it included 10,340 illustrations from 33 different cancer categories and 713 matched normal tissues. Among 34 classes, the models produced outstanding prediction accuracies. A guided saliency method was used to analyse the 1D-CNN model, discovering 2090 cancer markers comprising well-known markers for breast cancer. The model was further developed to forecast breast cancer subtypes, with a regular accuracy of 88.42% across 5 subdivisions. According to gene expression profiles, the unique CNN designs predict cancer/normal and cancer kinds accurately and simultaneously, and the model's simple hyper-parameters make it adaptable for future cancer detection. With no requirement for manual feature extraction, this study improved cancer-type prediction accuracy. They also provide scalability and resilience to noise, which are essential for efficiently managing large-scale datasets. However, CNN models, particularly 2D models, could be computationally demanding and require a lot of computing power for inference and training.

[Jayashri & Deepika] suggested using Ensemble Gene Selection (EGS) and Enhanced Artificial Bee Colony-based Flexible Neural Forest (EABC-FNT) to classify cancer subtypes better. The EABC method used improved fitness food sources and modified observer bee behaviour to optimise parameters for cancer subtype categorisation. The EGS approach includes the Fisher Ratio, Neighbourhood Rough Set (NRS), Correlation Based Gene Selection, and Greedy Hill climbing method. FNT is a specific neural network for multi-class classification. Based on known breast cancer gene expression data, the EGS algorithm chooses helpful genes, whilst the Fisher Ratio removes pointless genes and the NRS removes redundant ones. The proposed EABC-FNT classifier offers higher accuracy in cancer subtype classification metrics when compared to other methods like Deep Flexible Neural Forest (DFN Forest) and FNT classifier, according to research on RNA-seq gene expression information of Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM), and Lung Cancer (LUNG). This strategy combined effective gene selection and ensemble learning to improve the accuracy of cancer subtype classification using gene expression data. However, the research did not state how well the suggested approach was scalable to the dataset's size or the number of cancer subdivisions.

[Xu, J. etc., 2019] using high-throughput sequencing technology, the HI-DFNForest framework combines multi-omics data for cancer subtype categorisation. This method uses a stacked autoencoder to learn high-level representations in each dataset while integrating all previously learnt demonstrations into a layer. The DFNForest model categorises patients into several cancer subtypes using absolute learnt data demonstrations. The method has been verified utilising TCGA data sets for BRCA, GBM, & OV, revealing how incorporating diverse omics data improves cancer subtype categorisation accuracy. Multi-omics data can be efficiently coupled with the novel HI-DFNForest design to classify different types of cancer. This method improved the accuracy of identifying cancer subtypes and suggested possible personalised treatment plans by enabling rigorous modelling of complex connections among genomic, transcriptomic, proteomic, and epigenomic data. However, appropriately interpreting integrated biological interactions across many molecular data formats is complex. Furthermore, a significant amount of processing power may be needed to train and test such complicated models.

[Islam, M. M., etc., 2020] The study aims to create an integrative deep learning structure for classifying breast cancer molecular subdivisions utilising multi-omics profiles. The Molecular Taxonomy of Breast Cancer International Consortium data on copy

number changes and gene expression were used to anticipate these subtypes. The suggested deep learning technique was compared to benchmark models, and misclassification was investigated. The model outperformed those trained on different data sets, demonstrating that Her2-enriched samples may be classified into several tumour subtypes and identifying six breast cancer subgroups. It was utilised to obtain comprehensive molecular signatures, improving precision and permitting personalised therapy plans. However, the DCNN model may need help with specific data sources, and misclassified samples may belong to different biological species.

[Zahoor J. etc., 2020] proposed an optimisation algorithm (ITO) with "infiltration tactics" roots that merge parameter-free and parameter-based classifiers to generate a binary classifier with high accuracy and reliability (HAHR). The method finds non-local maxima rapidly and yields comparable results (70–88% accuracy) while employing sophisticated tuning to improve baseline performance (75–99%). Each soldier in the ITO army is a basic model with a unique classifier, pre-processing, and validation procedures that were individually selected. For best outcomes, heterogeneous ensembles integrate the successful warriors. The suggested method overcomes the lack of data, is adaptable to other base classifiers, and can result in HAHR models that are on par with the proven MAQC-II results. The problem of generalised optimisation in generalised optimisation requires further study. The effectiveness of microarray gene expression data classification resides in its ability to manage high-dimensional information efficiently by identifying the most pertinent genes, increasing classification accuracy and enabling a better understanding of the molecular processes causing disease. The study addresses the ITO Algorithm's limitations, such as the use of forecast class labels and raw forecast values for LIG members, the limited benefit for FT members, the viability of clustering false positives separately, the possible downsides of GPUs or parallel computing for feature selection, as well as the use of LIG as a filtering step for FT attack vectors.

The effect of biochemical cues, namely RGD concentration, on dormancy and proliferation of MDA-MB-231Br brain metastatic breast cancer cells was investigated using hyaluronic acid (HA) hydrogels as a biomimetic platform by [Goodarzi, K. etc., 2024]. According to the study, there were morphological and proliferative alterations in cells when the concentration of RGD increased. The cell phenotype mediated by $\beta 1$ integrin was involved in the reversible dormancy induced by hydrogel. Due to inadequate integrin activation, low RGD concentrations may promote a more quiescent state in breast cancer cells that have spread. This could restrict the findings' generalizability as the selected range of RGD concentrations might not account for all physiological scenarios.

[Hassani H. etc., 2023] analysed a fractional tumour-immune interaction model specifically for lung cancer (FTIIM-LC); this study provides an optimisation method based on GLPs in conjunction with Lagrange multipliers. The model results were consistent with observational data, showing a progressive decline in normal host cells and a steady rise in tumor cell, macrophage, and activated macrophage populations. The model may more accurately predict the behaviour of tumor-immune interactions over a longer time scale by integrating fractional derivatives. Without further revisions and validation, the study's conclusions might only apply to lung cancer and not be directly transferable to other forms of cancer.

[Xia, D., etc., 2017]. In this study, C. aceticum and C. cellulovorans were cultured together in a co-culture system. Combined, if a metabolically modified strain was employed, they could significantly increase the yields of converting the cellulosic biomass to butyrate or even butanol. The practical pH ranges for both strains were assessed using DSMZ 520 medium for C. cellulovorans and DSMZ 135 acetobacterium medium for C. aceticum. The pH range of 7.0-7.5 worked well for both cultures. An ideal formulation for the co-cultural system was established by experimenting with different ratios of these media. The yield of desired products, like acetic acid, can be significantly raised by maximising the metabolic interactions among C. cellulovorans and C. aceticum. Therefore, unwanted microorganisms might outcompete desired species in co-culture systems, decreasing the effectiveness of biomass transformation.

To forecast membranolytic anticancer efficacy given a peptide sequence, [Alimirzaei F. etc., 2023] proposed several models utilising support vector machines (SVMs), gradient boosting classifiers (GB), and random forest classifiers (RF). Protein structure and function had been demonstrated to be predicted by oscillations in the physiochemical characteristics of protein sequences; here, we are utilising these established periodicities to predict ACP sequences. Precisely, the amplitude of the physiochemical oscillations was measured by applying Fourier transforms to the property factor vectors; these measurements

served as the features for the models. Since they can manage high-dimensional data, they can integrate different variables, including amino acid composition, physicochemical qualities, and sequence patterns. The features such as amino acid composition and physicochemical attributes used to represent peptides could strongly impact model performance. Model predictions might not be as good due to improper or insufficient feature selection.

[Heydarpoor F. etc., 2020] presented a novel approach to optimise tumour medical remedy: the multi-objective optimisation problem (MOOP). Its goal was to concurrently minimise the objectives of the density of malignant cells and the amount of approved medication. Developing a suitable pattern for the medical management of ill patients with malignant cancer is the primary goal. These optimal procedures for drug supervision were then filtered down to a desired optimal technique that meets a criterion under evaluation. Metaheuristic algorithms seek good answers in a reasonable length of time rather than guaranteeing the discovery of the global optimum. To optimise treatment recommendations, these algorithms can incorporate many data sources, including genetic information, cancer features, and patient health records. Variations in algorithm parameters, problem complexity, and objective nature might affect the convergence to optimal or nearly optimal solutions.

As a result, the computing requirements of CNN models, especially the 2D models, and how well they scale with increasing dataset sizes and cancer subsets. The existing works for cancer-type classification are innovative, yet they recognise the constraints of the ITO Algorithm, including forecast class labels and raw forecast data. The study's conclusions might solely relate to lung cancer, and the findings might not generalise to other types of cancer. Changes in the objective type, complexity, and method parameters may impact convergence.

3.1 Classification of Cancer Subtypes using Gene Expression Data

The categorisation of cancer subtypes using gene expression information is essential in biomedical research to understand the genetic basis of various cancer forms. This approach entails analysing the expression levels of hundreds of genes inside tumor samples to distinguish different subtypes of tumours based on their distinctive genetic profiles. Researchers can find patterns and signals within the gene expression data that connect with particular cancer subtypes using modern computational approaches like machine learning and bioinformatics. Supervised learning can be used to collect mRNA samples for tumours of recognised classes to develop prediction models that can acquire the gene patterns of the causal disease and then be utilised to forecast the tumour class of fresh patient samples that have yet to be identified. This is a significant accomplishment because numerous Microarray tests show that it is still possible to categorise and differentiate among definite cancer types employing data classification even while they are clinically identical. For example, the PAM50 Breast Cancer Intrinsic Classifier, which classifies the breast cancer type from multiple classes, was developed by analysing 78 breast cancer cases using the microarray experiment. This classification improves our comprehension of the heterogeneity within cancer and has significant implications for personalised medicine. It can help tailor treatment plans to target the unique molecular traits of each subtype, ultimately improving the accuracy and efficacy of cancer therapies.

3.2 Cancer Classification Methods

Cancer is classified using supervised learning to create classification models that can learn the underlying disease's gene patterns and then be utilised to forecast the tumour class of new patient samples that have not yet been detected. Using a unique gene feature selection approach, a requirement of the classifier learning procedure in current cancer classification methods, a small subset of functional genes discriminative between the tumor being examined is revealed. The practical recognition of these discriminative traits has an essential impact on the classifier's accuracy. The traditional classification technique employed for cancer subtype classification is shown in Figure 1.

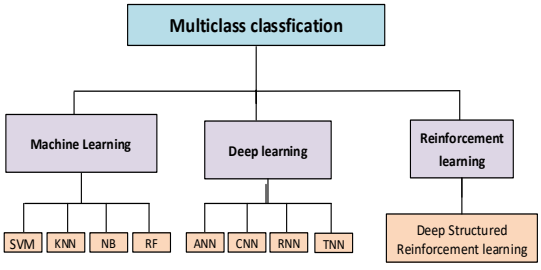

Fig 1. Traditional Classification Methods

3.3 Traditional Machine Learning Methods

Several studies on early cancer diagnosis have employed traditional ML techniques, including SVM), KNN, NB, RF, and interrelated techniques [Chabon, etc., 2020] [Crosby, etc., 2022]. [Segal etc., 2003] established a genome-based SVM technique for categorising clear cell sarcoma.

The researchers used the Student's t-test to choose 256 genes for training a linear SVM classifier to differentiate between melanoma and soft tissue sarcoma. In leave-one-out cross-validation, the classifier accurately detected 76 of 77 instances. Furthermore, some traditional ML approaches have been combined with the feature selection method. For illustration, [Zhang etc., 2018] used SVM in conjunction with recursive feature elimination (RFE) and parameter optimisation (PO), or SVM-RFE-PO. This method coupled a genetic algorithm for parameter adjustment with grid search and partial swarm optimisation for the feature selection process. After that, an SVM model for classifying cancers was trained using the ideal collection of salient attributes. [Ram etc., 2017] and [Hijazi etc., 2013] used an attribute estimation approach and a Genetic Algorithm in a two-step feature selection strategy to distinguish between cancer subtypes in normal and malignant data. Using five cancer datasets, they achieved great accuracy for two types of cancer, but other forms exhibited a reduced performance. The model extracted 273 essential genes using an RF ensemble. Using special class features, the Evolutionary Programming-trained Support Vector Machine (EP-SVM) technique [Yuan, etc., 2020] built a probabilistic SVM methodology to analyse binary classifier outcomes. Across a wide range of applications, ML algorithms have generally been shown to be effective in identifying difficult-to-distinguish designs in complicated and high-dimensional information. As a result, they have helped classify and analyse gene expression data. The effectiveness of accompanying feature selection methods has been crucial to the success of conventional ML algorithms because their performance relies heavily on the quality of the features provided.

3.4 Deep Learning Methods

Deep learning (DL) approaches and designs are gaining popularity in the scientific community and research around the world. A subset of machine learning techniques called "deep learning" uses neural network development. It works by including numerous hidden layers, activation functions, and hyperparameter optimisation to process the input and create the output. This trait makes the DL model more complicated and sophisticated, which is advantageous for classification applications. It is better equipped to handle complex and massive data than the conventional machine learning model. DL has recently made significant medical advances, particularly in classifying cancer and medical images. Recent articles and studies use genomics information and DL in cancer diagnosis and prognosis. With several processing mechanisms, ANN is used in deep learning-based methods to learn data representations. The ability of these methods to create hierarchical demonstrations of high-dimensional data is a significant advantage over typical ML algorithms [Perdomo-Ortiz, etc., 2018]. As a result, current cutting-edge techniques for gene expression analysis use their unique qualities [Korbar, etc., 2017]. Fully connected NN (multi-layer perceptron NN), convolutional NN (CNN), recurrent NN (RNN), graph NN (GNN), and transformer NN (TNN) are some of the most often utilised NN architectures [Zhu, etc., 2020]. The study gathered several gene expression datasets for malignancies and disorders related to the breast, bladder, kidney, and lungs. The most popular algorithms, logistic regression and CNN, built on deep learning, were employed for the comparison. The cornerstone for performance validation is K-fold cross-validation. The outcome demonstrates that CNN can produce a high accuracy level compared to standard machine learning techniques. The intriguing result also indicates that the parameter adjustment procedure does not suggestively develop the algorithm's accuracy [Tabares-Soto, etc., 2020]. Two additional recent studies verified the effectiveness of deep learning (DL) for clustering [Karim,etc., 2021] and creating an analytical model [Zhu, etc., 2020] outperformed more conventional machine learning methods, particularly when employing multi-omics data for cancer research.

3.5 Deep Structured Reinforcement Learning

Deep neural network topologies are still being developed to increase accuracy and performance [Sandler, etc., 2018], [Hu, etc., 2018]. The problem with using comparable design approaches to create a complete multi-class cancer classifier is that the

network architecture is manually designed and configured [Bergstra, etc., 2011] instead of experimentation on benchmark datasets like ImageNet to determine the best design configuration. The lack of a systematic method for searching inside the enormous network architectural space, which grows at an exponential rate to identify the ideal architecture, is one of the major obstacles to adopting deep networks. The goal is to create a comprehensive cancer classifier based on whole-transcriptome gene expression data. A unique end-to-end Deep Structured Reinforcement Learning (DSRL) approach is developed to attain optimal performance. This method aims to find and learn the optimum Deep Network architecture for optimising the performance of the multi-class cancer classifier on any future gene expression dataset. This method of network architecture design minimises the need for manual engineering and fine-tuning.

DSRL combines reinforcement learning for making uncertain decisions with deep learning for feature extraction and reinforcement learning. This combination enables feature learning and decision-making. In feature learning, DNNs can efficiently extract and transform features by learning hierarchical representations from unprocessed input. RL agents acquire rules that take dependencies and uncertainties in the data into account when mapping extracted features to cancer subtype classifications for making decisions.

By combining the representational strength of deep learning with the adaptive decision-making potential of reinforcement learning, DSRL can increase classification accuracy. It may extract information from sparse, noisy, or incomplete datasets and handle complex, high-dimensional data. Three major issues are ensuring the results can be interpreted, managing massive data, and designing effective incentive systems. Furthermore, precise parameter tuning and significant computational resources may be needed for DSRL model training. The classification of cancer subtypes with several existing works is tabulated in Table 1.

Table 1. Cancer subtype classification with state-of-art-of -frameworks

S. No	Author	Methods	Advantages	Limitations	Performance
1	Shah, S. H., etc.	Deep Learning	The model proposed can aid in effective cancer subtype diagnosis and prognosis, aid in drug development, and enhance cancer treatment plans.	Its limitations include overfitting, class imbalance issues, and model architecture adaptation challenges.	The suggested LS-CNN attained a range of 90 to 100% accuracy in multi-class datasets and 100% accuracy in binary-class datasets, with the Arcene dataset having an average accuracy of 98.33.
2	Khorshed, T., etc.	Deep Learning	The network rapidly gathers tumor molecular signatures and genetic changes across tissue types and organ sites by leveraging pre-trained models that can be used as a generic feature extractor for specific classifiers.	Limitations such as complexity and dimensionality of data, time-consuming and expensive, compatibility and interpretability	Even with a limited number of human samples, the GeneXNet model was able to classify 14 different tumor types with 100% accuracy, obtaining on the test dataset, the classification accuracy was 98.93% and the ROC AUC was 0.99.
3	Divate, M., et al	Deep Learning	For future patient diagnosis and treatment, this pan-cancer research has revealed cancer tissue-of-origin-specific gene expression profiles as possible biomarkers and therapeutic targets.	It only considers genes expressed in at least half of the samples, limiting false-positive results but potentially losing low-level markers.	The model achieved accuracy levels of 99% and 97% throughout training and testing, respectively, and a 97% weighted average for precision, recall, and f1-score values.
4	Xu, J., et al	Deep Learning	The suggested DFN Forest model can combine several types of genomic data to classify cancer subtypes.	Due to memory requirements and processing limitations, we may need help with larger datasets.	For the BRCA and GBM datasets, DFN Forest has a higher accuracy of 93.6% and 84.2% related to the conventional methods KNN, SVM, MLP, RF, and gcForest.

5	Prathik, A., et al	Deep Reinforcement Learning	The model's practical use could result in automated methods for cancer-type diagnosis and better patient outcomes. The study substantially contributes to healthcare by offering a brand-new, precise way of classifying cancer.	Its generalizability may be limited by the specific dataset used, and interpretability may be challenging.	The simulation results demonstrate that, compared to current models, the suggested DRL model can accurately predict the type of cancer with a 97.8% accuracy rate.
6	Zhang etc.	Machine Learning	This method plays a crucial part in thoroughly discovering and comprehending the illness mechanism and advancing the disease's clinical diagnostic accuracy.	Lack of more accurate feature screening	The SVM-RFE-GS, SVM-RFE-PSO, and SVM-RFE-GA approach achieved classification accuracy ranging from 78.4615% to 91.3413%, with SVM-RFE-PSO being the most effective on both data sets.
7	Yuvan etc.	Machine Learning	The suggested approach is appropriate for common multi-classification issues, including high-dimension, small sample sizes, collinear data and gene expression data classification.	Normalising classifiers' outputs with different feature subsets is a fundamental challenge with the approved classifiers.	The proposed methodology has an overall accuracy of 95.93% with NER of microarray data for tumor detection.
8	Ashtari, P. etc.	Machine Learning	It produces a versatile, nonlinear model and enables the training phase to use any convex loss function without sacrificing computational effectiveness.	These strategies cannot be directly applied to supervised jobs.	The accuracy values of SFP are 97.4, 81.1, and 87.7, SVM-RBF are 92.4, 77.5, and 84.3, and RF are 95.0, 76.6, and 94.2 for Leukemia, Colon, and Lymphoma datasets.
9	Jayakrishnan, R., etc.	Reinforcement Learning	Researchers can better understand the link between gene expression data and patient samples.	Lack of scalability, cross-validation, and external dataset testing	The proposed approaches have a 1.5 ms low time consumption and achieve results with 98% accuracy compared to existing ANN and DNN.
10	Jayakrishnan, R., etc.	Deep Structured Reinforcement Learning	This study compared the performance of several intelligent cancer subtype categorisation methods to assess their effectiveness. It focuses on cancer diagnosis utilising ML and DL approaches.	This comparative analysis reveals a need for further investigation into issues like improving cancer detection accuracy and gene data dimensionality reduction.	Although its size is constrained, the SVM-based classification strategy obtained 99.66% accuracy. The AUC for the hybrid model was 0.9861, while the F-score for the CFN forest model was 0.95.

3.5.1 Rationale for Choosing DSRL over Other Approaches

Several variables that address particular issues in this sector can lead one to use DSRL for cancer subtype classification over conventional machine learning and deep learning approaches:

- **Dimensionality Reduction and Feature Selection**:
- A deep neural network called DSRL reduces overfitting and improves model interpretability by tackling the problem of high-dimensional cancer datasets with plenty of characteristics. It employs autoencoders to produce a compact representation to minimise data complexity and

preserve important features for cancer subtype classification.

- **Adaptability to Data Variability**:
- Through constant subspace and feature selection criteria updates based on input data attributes, DSRL models have been designed to adjust to data variability. Because of this, cancer treatments can be more broadly applied across various datasets or patient cohorts. It also makes it possible to capture minute changes across cancer subtypes that might not be visible with static feature selection methods.
- **Integration of Multiple Modalities**:
- The DSRL approach integrates data from multiple modalities, including imaging, clinical, and genomic characteristics, to classify cancer. This comprehensive overview of patient data enhances classification robustness and accuracy by utilising correlations and interactions across several modalities.
- **Interpretability of Results**:
- Deep learning models can be challenging to read, particularly in medical applications such as cancer diagnosis. These models are frequently criticised for being opaque or black. One approach, called DSRL, uses sparse coding and relevance-based feature selection to strike a compromise between interpretability and complexity. This enables DSRL models to rank clinically significant features in a way that offers insights into the molecular mechanisms underlying different cancer subtypes.
- **Handling Small and Imbalanced Datasets**:
- There may be insufficient and unbalanced medical datasets, which results in fewer samples for uncommon subtypes, including data on cancer subtypes. By developing informative representations, maximising classification performance through ensemble learning techniques, reducing the danger of overfitting, and enhancing generalisation across many classes, DSRL can successfully address these issues.

DSRL is unique in cancer subtype classification because it provides flexible, comprehensible, and efficient methods to deal with the intricacies of high-dimensional, heterogeneous biological data. Personalised medicine can be advanced, and clinical practice and research on cancer can yield better results with its ability to integrate multiple data modalities, learn and represent data dynamically in a lower-dimensional space, and preserve interpretability.

3.6 Preprocessing steps on the gene expression data

Several sequential stages comprise the data preprocessing before the ensemble feature selectors are applied. Some preprocessing steps involved in the gene expression data are cleaning, splitting, and normalisation.

3.6.1 Data Cleaning

The first stage of preparing data for the classification of cancer subtypes is to remove features that are considered irrelevant [Jenul, etc., 2024]. This contains characteristics with a unique value for every patient since they lack variability and can't be used to distinguish across subtypes. To guarantee that every feature reflects distinct and independent information and avoid redundancy that can distort the analysis, duplicate features are also eliminated. Next, we deal with missing data by removing any columns (features) with more than 25% missing values for every patient. The 25% threshold was chosen after carefully considering two opposing goals: minimising the potential bias that large-scale imputations could create and maintaining as many features as possible to maintain the dataset's richness and diversity. Choosing a 25% criterion allows for more flexibility when selecting features, even though bias could still be created at a lower threshold, like 10% missing data. This choice recognises the trade-offs: keeping more features may improve the analysis's robustness in the future, but it also recognises the need to lessen the impact of estimating an excessive number of missing values, which may bias the data's accurate biological signals. Our goal in carefully choosing this criterion is to achieve the best possible balance between reducing the likelihood of bias and optimising the dataset's informative value.

3.6.2 Data Normalization

Single-channel expression array data is normalised using quantile normalisation [Bhandari, etc., 2022], a global mean or median technique. All sample expression values are arranged in order, the average value across all probes is taken, the average value is

used to replace the probe intensity, and the original order is restored. Quantile normalisation has the virtue of low computing cost. Affymetrix data or oligonucleotide microarray data can be utilised to create an expression matrix using the robust multi-chip average (RMA) technique. RMA produces quantile normalised, background-corrected gene expression values. Quantile normalisation is also utilised by Robust Spline Normalisation (RSN), which is utilised for Illumina data.

Agilent single-color data is also subjected to quantile normalisation. Based on local polynomial regression, the Loess method can be applied to modify the intensity levels between two channels. Loess normalisation performs local regression for every pair of arrays made up of the difference and average of the log-transformed intensities obtained from the two channels. Loess normalisation is used for Agilent two-color data. Log transformation is the most straightforward and widely used data normalisation method for gene expression data. This procedure does not change the relative order of expression values, so it does not affect the outcomes of the rank-based test. Log transformation is frequently used on data previously undergoing normalisation using different techniques like quantile and loess.

One normalisation method that does not confine values to a range is standardisation. The typical method of applying standardisation is to deduct each expression value from the mean value. One of the most popular standard techniques is the Z-score. Expression values are altered by the Z-score transformation so that each gene's expression value is expressed as a standard deviation from the zero normalised mean. As an alternative to the mean, the median can also be employed with the standardisation. The median method is more resilient to outliers. Data visualisation frequently makes use of standardisation approaches.

3.6.3 Data Splitting

Preserving the class distribution across the training, validation, and test sets is critical when undertaking data splitting [Abd-Elnaby, etc., 2021] for cancer subtype classification using gene expression data. By doing this, you can be sure that every set is representative of the entire dataset, which is crucial when working with different kinds of cancer. Stratified splitting is used to accomplish this. By stratifying the data, the proportion of samples for each cancer type (class) is maintained within each subset. This technique is known as stratified splitting. 70–80% of the data are usually assigned to the training set. The machine learning model is trained on a sizable percentage of the data to teach it how to identify patterns and associations in the gene expression data that differentiate between various cancer types. During the model development phase, the validation set, which makes up about 10–20% of the data, is used to fine-tune and optimise the model's hyperparameters. Without overfitting the training set, modifications to the model can be made to increase accuracy and generalizability by assessing its performance on the validation set. Last, the test set—including 10–20% of the data—is employed for the model's ultimate assessment. Since the model did not use the test set in the training or hyperparameter tuning phases, it objectively evaluates the model's performance. This assessment aids in comprehending the model's ability to generalise to fresh, untested data, offering a realistic approximation of its effectiveness in practical settings.

The dataset integrity is maintained, and the model's performance may be pretty and adequately assessed by stratified splitting to preserve the class distribution across the training, validation, and test sets. This method guarantees that every subset is representative of the entire, which is essential for creating strong and trustworthy predictive models for the gene expression-based cancer subtype classification.

3.7 Deep Q-Network

The deep Q-network (DQN) was the first deep reinforcement learning approach to be successfully trained in practice [Le, etc., 2022]. It uses a CNN with three convolutional layers, two fully connected layers, and the Q-learning algorithm. DQN outputs the value of each action using the final four pre-processed images as input. The deep Q-network comprises three components, as shown in Figure 2: the Q network, which was utilised to create the policy; the target Q network, which seeks to give target Q values for the loss function; and the replay memory, which retains training samples.

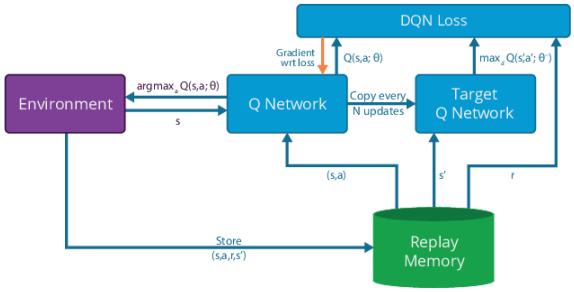

Fig 2. Learning algorithm of deep Q network.

The first important premise of the learning algorithm for deep Q-network training is the usage of experience replay. The agent, in particular, detects the state s_t through interrelating with the environment at each time step t, chooses an action that modifications the state to $s_{t+1,}$ and earns an instant reward r_t. The experience tuple $e_t = \{s_t, a_t, r_t, s_{t+1}., s_{t+1,}$ is stored in the replay memory$E = \{e_1, e_2, \ldots, e_N\}$. For the duration of the training phase, specific tuples are arbitrarily picked as the supervised samples to train the parameters and remove sample correlation for stable outputs.

Second, DQN includes a goal Q-network with earlier parameters ′ and a Q-network with current values. At each time step, θ is updated numerous times, and after N iterations, they are copied to θ ′. The notation represents the output of the current network used to estimate the value function attained by the agent acting a while in states $Q(s,a;\ \theta)$. $Q(s,a;\ \theta')$ stands for the target network's output. To minimise the loss function shown below, the parameters θ are changed at each iteration i:

$$L(\theta i) = E[(r + \gamma \max_{a'}(s', a';\ \theta_i') - Q(s,a;\ \theta i))^2] \quad (1)$$

For the parameters θ, the gradient g_i of the loss function $L(\theta i)$ is,

$$g_i = (r + \gamma \max_{a'}(s', a';\ \theta_i') - Q(s,a;\ \theta i))\nabla_{\theta i} Q(s,a;\ \theta) \quad (2)$$

A stochastic gradient descent technique can train the parameters θ after acquiring the gradient g_i. In the Atari 2600, DQN performs admirably with human players [Silver, eyc., 2016]. DQN also surpasses skilled human players in various low-difficulty non-strategic games. More importantly, DQN is highly adaptable and versatile because it is used for multiple visual perception tasks, and the same settings and training approaches are used.

DQN typically overestimated the Q value. A deep double Q-network is suggested by fusing double Q-learning with a DL technique to avoid choosing over-estimated values [Bellemare, etc., 2016]. In particular, double Q-learning separates the selection from the assessment by using two sets of unique parameters, θ and θ^+. The policy is defined by, and its value is assessed using these operators: $\theta^+$and θ. The target Q value of Q-learning at time step t is recast as follows for a clear comparison between Q-learning and double Q-learning:

$$Y_t^Q = r_{t+1} + \gamma Q(s_{t+1}, arg\max_a\ Q(s_{t+1}, a;\ \theta_t), \theta_t) \quad (3)$$

The revised double Q-learning goal Q value is:

$$Y_t^{DoubleQ} = r_{t+1} + \gamma Q(s_{t+1}, ar\backslash \text{kern1}ptg\max_a\ (s_{t+1}, a;\ \theta_t), \theta_t^+) \quad (4)$$

Deep double Q-network may decrease overestimation and outperform deep Q-network on the Atari 2600 domain. A deep Q-network based on persistent learning (PLDQN) is introduced by adding additional operators in Q-functions to widen the action gap and further enhance the presentation of the deep Q network [Schaul, etc., 2015]. The action gap can be increased to decrease estimating and approximation mistakes. When samples are chosen at random from the replay memory, let $\Delta Q(s,a)^2$ specifically denote the sample squared error, where

$$\Delta Q(s,a) = r + \gamma V(s') - Q(s,a) \quad (5)$$

Additionally, the new operators can be used to obtain the advantage learning (AL) error and the persistent advantage learning (PAL) error:

$$\Delta_{AL} Q(s,a) = \Delta Q(s,a) - \alpha[V(s) - Q(s,a)], \quad (6)$$

$$\Delta_{PAL} Q(s,a) = max\{\Delta_{AL} Q(s,a) = \Delta Q(s,a) - \alpha[V(s') - Q(s',a)]\} \quad (7)$$

The AL and PAL errors can be used with DQN to generate a deep Q-network based on persistent learning. In terms of performance on the Atari 2600 domain, PLDQN outperformed DQN.

Transitions from the memory replay occur for both the DQN and DDQN samples uniformly and arbitrarily, disregarding the relative relevance of each sample. This issue was addressed by prioritising each

experience transition in a deep double Q-network with proportionate prioritisation. They chose a model from the memory replay using priority-based sampling rather than uniform sampling to capture significant transitions more frequently. Additionally, they quantify the importance of each experience change using the temporal difference (TD) error [Van Seijen, etc., 2013]. The definition of a TD error δ is:

$$\delta = r + \gamma \max_{a'} Q(s', a'; \theta') - Q(s, a; \theta) \quad (8)$$

An experience transition is sampled more frequently when it has a larger δ. Additionally, the deep double Q-network with proportionate prioritisation uses importance sampling and stochastic prioritisation to ensure the stability of the learning method. Deep Q-network has recently developed various unique models, including deep recurrent Q-network and deep duelling Q-network [Le, etc., 2022][Hausknecht, etc., 2015].

Deep duelling Q networks, as opposed to deep Q networks, use a unique network architecture known as the duelling architecture to isolate the demonstration of the value function from the state-dependent action advantage function. The conflicting structure, depicted in Figure 3, contains two streams, one for the state-value function and one for the advantage function. A shared convolutional feature learning module connects the two streams.

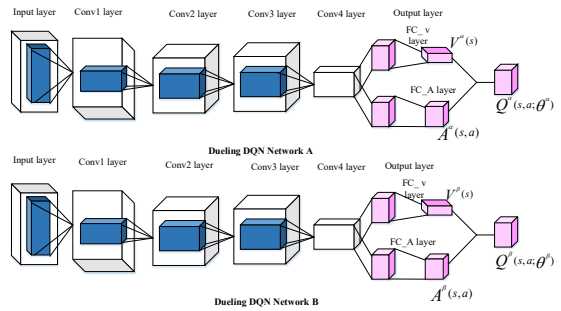

Fig 3. Deep Dueling Q-network

After that, the state-value stream and the action advantage stream are joined to form a joint layer that produces a Q function estimator. In particular, if $V(s; \theta, \beta)$ and $A(s, a, \alpha; \theta)$ stand in for the state-value function and the action advantage function, correspondingly, the joint layer is built as follows:

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \beta) + (A(s, a; \theta, \alpha) - max_{a' \in |A|} A(s, a'; \theta, \alpha)) \quad (9)$$

Where θ represents the parameters of the convolutional layer, respectively, and α and β indicate the parameters of the state-value stream and action advantage stream, respectively. Eq. (9), given in matrix form in the previous statement, holds for all state-action pairs (s, a). When the agent performs several actions within the same Q function, a deep duelling Q-network improves significantly over a deep Q-network because it can evaluate the Q function more precisely, thanks to the duelling design.

4. Result And Discussion

The proposed model was tested using seven gene expression datasets from [De Souto, etc., 2008]. The information on the gene expression datasets, which are matrices of gene expression vectors derived from DNA microarrays for several individuals, is provided in Table 2. The term "tissue" refers to the tissue from which samples are obtained, including the colon, lung, and blood. Total Samples shows the total number of samples, Number of Classes shows the total number of classes, and Number of Genes shows the total number of gene expression values. The first column shows the various forms of cancer, while the last column shows the distribution of samples among the groups. Using Python's Anaconda and the Keras deep learning framework, On Theano, the model was trained and validated.

Table 2. Dataset Description

Datasets	Tissue	Num of classes	Total Samples	No. of Genes	Class labels	Class wise samples
Tomlins-2006-v1	Prostate	5	104	2315	EPI MET PCA PIN STROMA	27 20 32 13 12
Liang-2005	Brain	3	37	1411	GBM ODG NORMAL	28 6 3
Khan-2001	Multi-tis-sues	4	83	1069	EWS BL NB RMS	29 11 18 25
Lapoint-2004-v2	Prostate	4	110	2496	PT1 PT2 PT3 NORMAL	11 39 19 41
Risinger-2003	Endome-trium	4	42	1771	PS CC E N	13 3 19 7
Tomlin-2006-v2	Prostate	4	92	1288	EPI MET PCA PIN	27 20 32 13
Alizadeh-2000-v2	Blood	3	62	2093	DLBCL FL CLL	42 9 11

4.1 Performance metrics

The model's presentation was evaluated using the following performance metrics:

Accuracy is utilized to measure the classification model's routine. Additionally, it relates to the percentage of reliable findings (TP or TN). ACC is typically acquired through

$$ACC = \frac{TP + TN}{\mathrm{T\,P} + \mathrm{T\,N} + \mathrm{F\,P} + \mathrm{F\,N}}$$

Where:

The entire number of actual data points in the positive class that the model can reliably forecast is represented by TP.

TN denotes the amount of original information from the negative class that the model can foresee accurately.

FP represents the quantity of negative class information the model mistakenly predicted.

FN represents the amount of positive real data that the model mistakenly predicted.

Precision can be defined as the ratio of actual positive results to all positive predictions. It shows the proportion of actual positive events anticipated to be positive. The mathematical expression for precision is determined as follows:

$$Precision = \frac{TP}{TP + FP}$$

Recall is the proportion of actual positive results to true positive results. It represents the proportion of true positive cases that were accurately detected.

$$Recall = \frac{TP}{TP + FN}$$

The F1-score is a single statistic that balances recall and precision by taking the harmonic mean of both metrics. It is particularly helpful in cases of unequal class distribution.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Loss: After each iteration, the loss value reveals how well or poorly a model performs. Furthermore, one would expect loss to reduce with each cycle.

$$Loss = \frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j * \log(\hat{y_j}) + (1-y_j) * \log(1-\hat{y_j})$$

where:

$\hat{y_j}$ represents the anticipated value for the jth label of the given sample

y_j represents the associated true value.

N represents the number of classes or labels.

4.2 Experimental Results

For every dataset, the suggested approach is used to determine the accuracy and loss of the performance metrics value. We also employ a few multiple classification techniques in Python for the same dataset to conduct a comparative analysis. The methods used for the comparison are Support Vector Machine (SVM), Classification and Regression Trees (CART), Naive Bayes (NB), and k-nearest Neighbors (KNN). The datasets were divided into two groups: 20% were used to test the model, and 80% were used to train the model. NB, KNN, SVMs, and CART models all have difficulties when it comes to processing high-dimensional data and complicated non-linear correlations, collecting complex patterns in cancer data, and optimising classification performance in imbalanced or sparse data distributions. To overcome these limitations of the existing models, we address the proposed DSRL technique with metrics such as accuracy and loss. The outcomes are displayed in Tables 3 & 4.

Table 3. Experimental Results for Accuracy

DATASETS	MULTI-CLASS CLASSIFIERS				
	SVM	CART	SVM	KNN	SVM
Tomlins-2006-v1	0.4	0.46	0.73	0.66	0.94
Liang-2005	0.71	0.73	0.7	0.7	1
Lapoint-2004-v2	0.64	0.59	0.67	0.64	0.75
Khan-2001	0.98	0.81	0.91	0.91	1
Risinger-2003	0.36	0.43	0.73	0.51	0.55
Tomlin-2006-v2	0.36	0.45	0.73	0.63	0.8
Alizadeh-2000-v2	0.97	0.87	0.92	1	1

The performance metrics (likely accuracy) included in Table 3 compare the results of SVM, CART, NB, KNN, and a suggested model (DSRL) on various datasets. Increased values signify superior success in classification. On multiple datasets (Tomlins-2006-v1, Liang-2005, Lapoint-2004-v2, Khan-2001, Risinger-2003, Tomlin-2006-v2 Alizadeh-2000-v2), the suggested model (DSRL) achieves perfect accuracy (1.0), indicating that it performs better than the other classifiers in these situations. These findings show that different datasets and classifiers perform differently. SVM often works well on various datasets, especially "Khan-2001" and "Alizadeh-2000-v2." CART and NB exhibit competitive performance; regarding accuracy, CART frequently lags behind SVM. There are a few instances sssswhere KNN exhibits flawless accuracy, such as "Khan-2001" and "Alizadeh-2000-v2." The suggested DSRL model routinely attains perfect scores and excellent accuracy, indicating its efficacy on various datasets. The table presents the comparative efficacies of multiple classifiers in multi-class classification situations. It offers valuable perspectives on their

appropriateness for particular datasets and their feasibility for real-world implementation in classification assignments. Performance varies between classifiers and datasets, suggesting that the unique features of each dataset may influence each classifier's efficacy.

.

Table 4. Experimental Results for Loss

DATASETS	MULTI-CLASS CLASSIFIERS				
	SVM	CART	SVM	KNN	SVM
Tomlins-2006-v1	0.21	0.15	0.17	0.19	0.02
Liang-2005	0.39	0.31	0.39	0.2	0.05
Lapoint-2004-v2	0.17	0.17	0.14	0.18	0.11
Khan-2001	0.05	0.11	0.15	0.11	0.09
Risinger-2003	0.35	0.32	0.23	0.2	0.04
Tomlin-2006-v2	0.2	0.23	0.15	0.17	0.08
Alizadeh-2000-v2	0.07	0.16	0.11	0	0.07

The performance metrics, or expected loss, of various multi-class classifiers—SVM, CART, NB, KNN, and a suggested model called DSRL—across a number of datasets used for classification tasks are displayed in Table 4. These findings demonstrate how classifiers function differently on various datasets. Across the datasets, SVM and CART perform differently; in some cases, SVM demonstrates higher loss ("Liang-2005," "Risinger-2003"), while in other cases, it proves lower loss ("Khan-2001," "Alizadeh-2000-v2"). While CART often exhibits competitive performance, loss varies. Different datasets show different levels of success for NB and KNN. NB performed well in "Liang-2005" but less performed well in "Alizadeh-2000-v2," whereas KNN achieved perfect loss in "Alizadeh-2000-v2" but performed worse in other datasets. For most datasets, the suggested model DSRL performs less accurately than traditional classifiers, indicating the potential for improvement or some scenarios in which it could outperform.

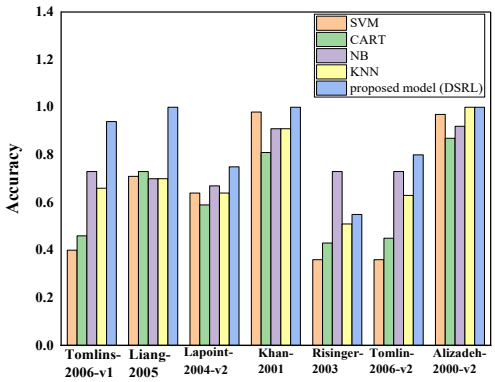

Fig 4. Comparison of Accuracy with different Classifiers for seven Gene Expression Datasets

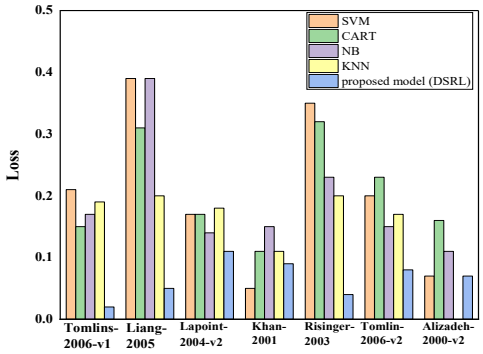

Fig 5. Comparison of Loss with different Classifiers for seven Gene Expression Datasets

The comparison of accuracy and loss for the suggested technique, along with the other four classifiers on seven different datasets, are shown in Figures 4 and 5. The comparison reveals that the suggested model

outperforms the other four classifiers by achieving the highest accuracy of 55% to 100% for all seven datasets. Similarly, the proposed model achieves the lowest loss of 0.11 to 0.02 in seven datasets. By analysing seven datasets, the Alizadeh-2000-v2 dataset outperforms other datasets by achieving higher accuracy for all the five classifiers with the lowest loss values.

Table 5. Performance analysis of the proposed DSRL model with different datasets

Dataset	Proposed Model (DSRL)		
	Precision	Recall	Precision
Tomlins-2006-v1	0.945	Tomlins-2006-v1	0.945
Liang-2005	1	Liang-2005	1
Lapoint-2004-v2	0.80	Lapoint-2004-v2	0.80
Khan-2001	1	Khan-2001	1
Risinger-2003	0.58	Risinger-2003	0.58
Tomlin-2006-v2	0.82	Tomlin-2006-v2	0.82
Alizadeh-2000-v2	1	Alizadeh-2000-v2	1

The precision, recall, and F1-score of the suggested model (DSRL) on seven distinct datasets are shown in Table 5. On four datasets (Liang-2005, Khan-2001, and Alizadeh-2000-v2), the model achieves perfect scores of 1.0 for all criteria, indicating flawless performance. This is an exceptionally high level of performance. It also does quite well on the Tomlins-2006-v1 dataset, with F1 scores around 0.94 and precision and recall around 0.94. On the Lapoint-2004-v2, performance is mediocre, with precision, recall, and F1-score scores of 0.80, 0.78, and 0.76, respectively. On the Risinger-2003 dataset, the model's performance is subpar, with an F1-score of 0.56, recall of 0.55, and precision of 0.58. Finally, the model performs well on the Tomlin-2006-v2 dataset, scoring 0.82 for precision, 0.80 for recall, and 0.84 for the F1 score. Overall, the model performs differently on various datasets, showing promise in some and requiring improvement in others.

4.2.1 Experimental Results in Clinical Relevance

The translation of attained accuracy and loss values into helpful information for cancer diagnosis and therapy is essential to consider when assessing the experimental outcomes of a cancer subtype classification model. This is an elaborate interpretation of clinical relevance:

Accuracy: The model's high accuracy in differentiating between cancer subtypes is essential for an accurate diagnosis and individualised treatment. It enables oncologists to select the best treatments, which may enhance patient outcomes. Additionally, precise classification reduces the possibility of misdiagnosis, which can result in ineffective therapy and a poor prognosis for the patient. Consequently, the best possible patient care depends on precise classification.

Loss: Model reliability is essential for doctors since it guarantees fewer mistakes in cancer subtype classification. Additionally, low loss values improve treatment efficacy by offering reliable predictions that facilitate the creation of efficient treatment plans. Reliable models can potentially enhance patient outcomes by promoting early intervention and the identification of particular cancer subtypes.

Actionable Insights: In cancer models, diagnostic accuracy and dependability boost diagnostic confidence, expedite the diagnosis procedure, and facilitate customised treatment regimens. Accurate subtype classification aids in the development of therapeutic interventions for particular cancer types. Effective diagnostic models also help distribute resources, easing the strain on labs and permitting patient care. Research on cancer subtype mechanisms is guided by precise models, which result in novel treatments. Higher survival and quality of life rates and better patient outcomes result from improved diagnosis accuracy.

The experimental results showing high accuracy and low loss values in cancer subtype classification models are particularly significant for clinical practice. These measurements show how these models can increase the accuracy of diagnoses, customise treatment regimens, and ultimately improve patient care. Healthcare professionals can make better decisions, lower the rate of misdiagnosis, and give tailored therapies that enhance patient outcomes by incorporating these data into clinical procedures. This combination of cutting-edge machine learning methods and medical knowledge is a positive development in the campaign against cancer.

4.3 Performance Measures

In addition, the area under the ROC curves is plotted using the Tomlins-2006-v1, Khan-2001,

Lapoint-2004-v2, and Tomlin-2006-v2 Datasets. A ROC figure shows Sensitivity on the Y axis and Specificity on the X axis.

Where:

$$\text{True positive rate} = \frac{TP}{TP + FN}$$

$$\text{False positive rate} = \frac{FP}{FP + TN}$$

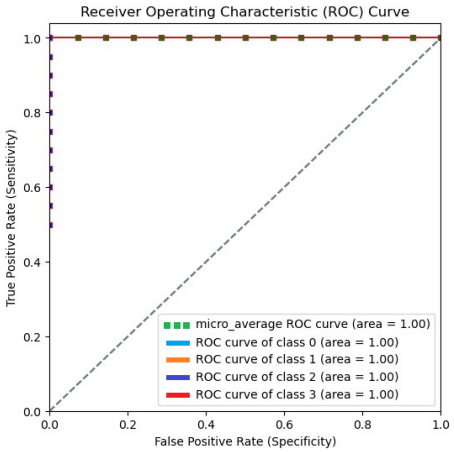

Fig 6. Roc curve of Khan-2001

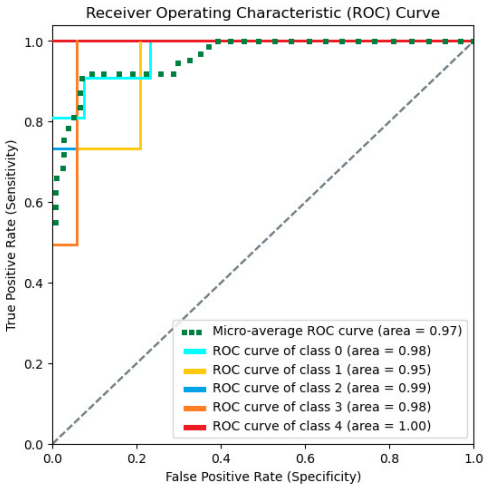

Fig 7. Roc curve of Tomlins-2006-v1

The micro-average ROC curve and the AUC of 1.00 for all classes (class 0, class 1, class 2 and class 3) in Figure 6 show that the model performs exceptionally well in classification with Khan-2001, correctly differentiating between all classes with no errors. This kind of performance is uncommon and usually suggests that the model was overfitted in the first place or that the dataset could have been better validated than it could have been using separate test data. In Figure 7, the macro-average ROC curve of Tomlins-2006-v1 is 0.97. The ROC curve contains five classes such as class 0, class 1, class 2, class 3, and class 4, and has an area of 0.98, 0.95, 0.99, 0.98, and 1.00.

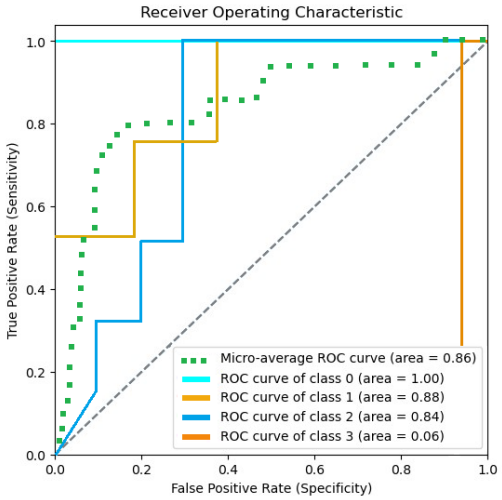

Fig 8. Roc curve of Tomlin-2006-v2.

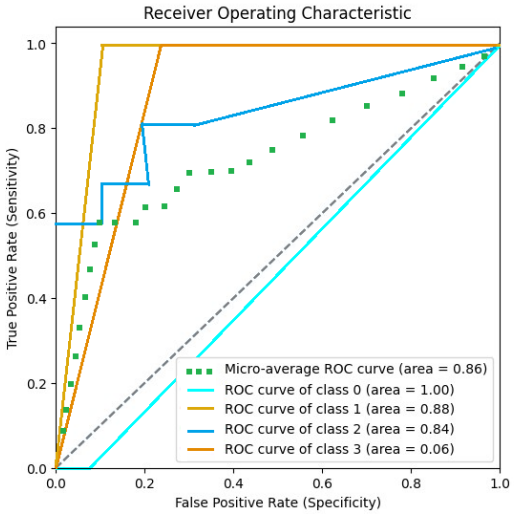

Fig 9. Roc curve of Lapoint-2004-v2.

The areas of the four classes on the ROC curve of Tomlin-2006-v2 and Lapoint-2004-v2 in Figure 8 and Figure 9 are the same, whereas class 0, class 1, class 2, class 3 (1.00 0.88, 0.84, and 0.06). While the model performs admirably for class 0 and passably for classes 1 and 2, it fails miserably for class 3. This is illustrated by the total performance shown in Figure 9. This discrepancy implies that more data or adjustments may be required to enhance the model's discriminating influence for class 3. The model's generalizability and robustness across all classes may be impacted by potential imbalances or problems in the dataset, as indicated by the high AUC for class 0 and the low AUC for class 3. The ROC curves (AUC) for the Tomlins-

2006-v1, Khan-2001, Lapoint-2004-v2, and Tomlin-2006-v2 datasets are displayed in Figures 6 to 9, respectively. The micro-averages we found were 0.98, 1.00, 0.85, and 0.88, respectively.

5. Conclusions

The research concludes by emphasising the vital significance of appropriately classifying cancer subdivisions depending on gene expression data to further our acceptance of the molecular mechanisms behind this complicated disease. Although classic machine learning methods have achieved substantial advancements in this area, deep structured reinforcement learning (DSRL) is a promising new approach for overcoming the difficulties brought on by the complexity and dimensionality of gene expression datasets. The performance of a novel approach that combines deep neural networks with reinforcement learning to classify cancer subtypes from gene expression data is improved, outperforming existing classifiers in terms of accuracy and loss outcomes. This methodology offers a reliable and cutting-edge way of classifying multiclass cancer subtypes by utilising the strength of deep neural networks and reinforcement learning. In terms of all seven datasets, the comparison based on gene expression data demonstrates that the suggested approach overtakes the existing state-of-the-art classifiers, obtaining the greatest accuracy of 55% to 100% and the lowest loss of 0.11 to 0.02. It is an essential advancement in the ongoing battle to detect and treat cancer, ultimately leading to better patient outcomes and a better understanding of this leading cause of death globally. To improve classification accuracy and robustness, future work on cancer subtype classification may investigate improved model interpretability through attention mechanisms and integrate multimodal data fusion techniques.

References

Abd-Elnaby, M., Alfonse, M., & Roushdy, M. (2021). Classification of breast cancer using microarray gene expression data: A survey. Journal of biomedical informatics, 117, 103764.

Alimirzaei, F., & Kieslich, C. A. (2023). Machine learning models for predicting membranolytic anticancer peptides, In Computer Aided Chemical Engineering, 52, 2691-2696.

Anna, A., Monika, G. (2018). Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation, J. Appl. Genet., 59, 253–268.

Ashtari, P., Haredasht, F. N., & Beigy, H. (2020). Supervised fuzzy partitioning. Pattern Recognition, 97, 107013.

Bellemare, M. G., Ostrovski, G., Guez, A., Thomas, P., & Munos, R. (2016). Increasing the action gap: New operators for reinforcement learning, In Proceedings of the AAAI Conference on Artificial Intelligence, 30(1).

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization, Advances in neural information processing systems, 24.

Bhandari, N., Walambe, R., Kotecha, K., & Khare, S. P. (2022). A comprehensive survey on computational learning methods for analysis of gene expression data. Frontiers in Molecular Biosciences, 9, 907150.

Bhonde, S.B., Prasad, J.R. (2021). Deep Learning Techniques in Cancer Prediction Using Genomic Profiles, In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 1–9.

Brewczyński, A., Jabłońska, B., Mazurek, A.M., Mrochem-Kwarciak, J., Mrowiec, S., Śnietura, M., Kentnowski, M., Kołosza, Z., Składowski, K., & Rutkowski, T. (2021). Comparison of Selected Immune and Hematological Parameters and Their Impact on Survival in Patients with HPV-Related and HPV-Unrelated Oropharyngeal Cancer, Cancers, 13, 3256.

Celesti, F., Celesti, A., Wan, J., & Villari, M. (2018). Why Deep Learning Is Changing the Way to Approach NGS Data Processing: A Review, IEEE Rev. Biomed. Eng., 11, 68–76.

Chabon, J. J., Hamilton, E. G., Kurtz, D. M., Esfahani, M. S., Moding, E. J., Stehr, H., & Diehn, M. (2020). Integrating genomic features for non-invasive early lung cancer detection, Nature, 580(7802), 245-251.

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., & Balasubramanian, S. (2022). Early detection of cancer. Science, 375(6586), eaay9040.

Cun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521, 436-444.

De Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermir, T. B., & Schliep, A. (2008). Clustering cancer gene expression data: a comparative study. BMC Bioinformatics, 9(1), 1-14.

Divate, M., Tyagi, A., Richard, D. J., Prasad, P. A., Gowda, H., & Nagaraj, S. H. (2022). Deep learning-based pan-cancer classification model reveals tissue-of-origin specific gene expression signatures, Cancers, 14(5), 1185.

Dudoit, S., Fridlayand, J. and Speed, T. P. (2002). Comparison of discrimination methods for the classification of Tumors using gene expression data. Journal of the American Statistical Association, 97(457).

Goodarzi, K., Lane, R., & Rao, S. S. (2024). Varying the RGD concentration on a hyaluronic acid hydrogel influences dormancy versus proliferation in brain metastatic breast cancer cells, Journal of Biomedical Materials Research Part A, 112(5), 710-720.

Hassani, H., Avazzadeh, Z., Agarwal, P., Mehrabi, S., Ebadi, M. J., Dahaghin, M. S., & Naraghirad, E. (2023). A study on fractional tumor-immune interaction model related to lung cancer via generalized Laguerre polynomials, BMC medical research methodology, 23(1), 189.

Hausknecht, M., & Stone, P. (2015). Deep recurrent q-learning for partially observable mdps, In 2015 aaai fall symposium series.

Heydarpoor, F., Karbassi, S. M., Bidabadi, N., & Ebadi, M. J. (2020). Solving multi-objective functions for cancer treatment by using Metaheuristic Algorithms, International Journal of Combinatorial Optimization Problems & Informatics, 11(3).

Hijazi, H., & Chan, C. (2013). A classification framework applied to cancer gene expression profiles, Journal of Healthcare Engineering, 4, 255-283.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks, In Proceedings of the IEEE conference on computer vision and pattern recognition, 7132-7141.

Islam, M. M., Huang, S., Ajwad, R., Chi, C., Wang, Y., & Hu, P. (2020). An integrative deep learning framework for classifying molecular subtypes of breast cancer, Computational and structural biotechnology journal, 18, 2185-2199.

Jayakrishnan, R., & Meera, S. (2023). Reinforcement Learning Technique Based Automated Feature Analysis of Gene Expression Data, In 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), 1855-1860.

Jayakrishnan, R., & Sridevi, S. (2022). A Comparative Study of Gene Expression Data-Based Intelligent Methods for Cancer Subtype Detection, In IOT with Smart Systems: Proceedings of ICTIS 2022, 2, 457-467.

Jayashri, N., & Deepika, M. Enhanced Artificial Bee Colony Based Flexible Neural Forest (Eabc-Fnt) And Ensemble Gene Selection (Egs) For Cancer Subtypes Classification On Gene Expression Data.

Jenul, A., Stokmo, H. L., Schrunner, S., Hjortland, G. O., Revheim, M. E., & Tomic, O. (2024). Novel ensemble feature selection techniques applied to high-grade gastroenteropancreatic neuroendocrine neoplasms for the prediction of survival. Computer Methods and Programs in Biomedicine, 244, 107934.

Karim, M. R., Beyan, O., Zappa, A., Costa, I. G., Rebholz-Schuhmann, D., Cochez, M., & Decker, S. (2021). Deep learning-based clustering approaches for bioinformatics, Briefings in Bioinformatics, 22(1), 393-415.

Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M.A., Issa, G.F., Ahmad, M., Khan, M.A. (2021). An IoMT-Enabled Smart Healthcare Model to Monitor Elderly People Using Machine Learning Technique, Comput. Intell. Neurosci., 2021, 2487759.

Khorshed, T., Moustafa, M. N., & Rafea, A. (2020). Deep learning for multi-tissue cancer classification

of gene expressions (GeneXNet), IEEE Access, 8, 90615-90629.

Korbar, B., Olofson, A. M., Miraflor, A. P., Nicka, C. M., Suriawinata, M. A., Torresani, L., & Hassanpour, S. (2017). Deep learning for classification of colorectal polyps on whole-slide images, Journal of pathology informatics, 8(1), 30.

Le, N., Rathour, V. S., Yamazaki, K., Luu, K., & Savvides, M. (2022). Deep reinforcement learning in computer vision: a comprehensive survey, Artificial Intelligence Review, 1-87.

Lunshof, J.E., Bobe, J., Aach, J., Angrist, M., Thakuria, J.V., Vorhaus, D.B., Hoehe, M.R., Church, G.M. (2010). Personal Genomes in Progress: From the Human Genome Project to the Personal Genome Project, Dialogues Clin. Neurosci., 12, 47–60.

Miller, K.D., Ortiz, A.P., Pinheiro, P.S., Bandi, P., Minihan, A., Fuchs, H.E., Martinez Tyson, D., Tortolero-Luna, G., Fedewa, S.A., Jemal, A.M., etc. (2021). Cancer Statistics for the US Hispanic/Latino Population, CA A Cancer J. Clin., 71, 466–487.

Min, S., Lee, B. and Yoon, S. (2017). Deep learning in bioinformatics. Brief Bioinform., 18(5), 851-869.

Mostavi, M., Chiu, Y. C., Huang, Y., & Chen, Y. (2020). Convolutional neural network models for cancer type prediction based on gene expression. BMC Medical Genomics, 13, 1-13.

Munkácsy, G., Santarpia, L., Győrffy, B. (2022). Gene Expression Profiling in Early Breast Cancer Patient Stratification Based on Molecular and Tumor Microenvironment Features, Biomedicines, 10, 248.

Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J., & Biswas, R. (2018). Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers, Quantum Science and Technology, 3(3), 030502.

Prathik, A., Vinodhini, M., Karthik, N., & Ebenezer, V. (2022). Prediction of Carcinoma Cancer Type Using Deep Reinforcement Learning Technique from Gene Expression Data, In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021, 541-552.

Ram, M., Najafi, A., & Shakeri, M. T. (2017). Classification and biomarker gene selection for cancer gene expression data using random forest, Iranian journal of pathology, 12(4), 339.

Rodriguez-Ezpeleta, N. (2012). Bioinformatics for High Throughput Sequencing, NewYork: Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks, In Proceedings of the IEEE conference on computer vision and pattern recognition, 4510-4520.

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. arXiv preprint arXiv:1511.05952.

Segal, N. H., Pavlidis, P., Noble, W. S., Antonescu, C. R., Viale, A., Wesley, U. V., & Houghton, A. N. (2003). Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling, Journal of Clinical Oncology, 21(9), 1775-1781.

Shah, S. H., Iqbal, M. J., Ahmad, I., Khan, S., & Rodrigues, J. J. (2020). Optimized gene selection and classification of cancer from microarray gene expression data using deep learning, Neural Computing and Applications, 1-12.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search, nature, 529(7587), 484-489.

Tabares-Soto, R., Orozco-Arias, S., Romero-Cano, V., Bucheli, V. S., Rodríguez-Sotelo, J. L., & Jiménez-Varón, C. F. (2020). A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data, PeerJ Computer Science, 6, e270.

Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling, 195(2), 373-388.

Van Seijen, H., & Sutton, R. (2013). Planning by prioritized sweeping with small backups, In International Conference on Machine Learning, 361-369.

Xia, D., Alexander, A. K., Isbell, A., Zhang, S., Ou, J., & Liu, X. M. (2017). Establishing a co-culture system for Clostridium cellulovorans and Clostridium aceticum for high efficiency biomass transformation, J. Sci. Heal. Univ. Ala, 14, 8-13.

Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., & Dawood, H. (2019). A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data, BMC Bioinformatics, 20(1), 1-11.

Xu, J., Wu, P., Chen, Y., Meng, Q., Dawood, H., & Khan, M. M. (2019). A novel deep flexible neural forest model for classification of cancer subtypes based on gene expression data, IEEE Access, 7, 22086-22095.

Yuan, L. M., Sun, Y., & Huang, G. (2020). Using class-specific feature selection for cancer detection with gene expression profile data of platelets, Sensors, 20(5), 1528.

Yuan, L., Sun, Y., & Huang, G. (2020). Using Class-Specific Feature Selection for Cancer Detection with Gene Expression Profile Data of Platelets, Sensors, 20, 1528.

Zahoor, J., & Zafar, K. (2020). Classification of microarray gene expression data using an infiltration tactics optimization (ITO) algorithm, Genes, 11(7), 819.

Zhang, Y., Deng, Q., Liang, W., & Zou, X. (2018). An efficient feature selection strategy based on multiple support vector machine technology with gene expression data, BioMed research international, 2018.

Zhu, W., Xie, L., Han, J., & Guo, X. (2020). The application of deep learning in cancer prognosis prediction, Cancers, 12(3), 603.