

# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





# Transformer Based Gene Expression Signature Prediction for Early Cancer Detection and Classification

S.T.P.L. Varaprasad<sup>1</sup>, B. Navya<sup>2</sup>, J. Rishitha<sup>3</sup>, V. Gowtham<sup>4</sup>, G. Lokesh<sup>5</sup>, M. khagapathi<sup>6</sup>

Assistant Professor, Department of CSE (AI&ML), Nadimpalli Satyanarayana Raju Institute of Technology (NSRIT),  
Vishakhapatnam, India<sup>1</sup>

Students, Department of CSE (AL&ML), Nadimpalli Satyanarayana Raju Institute of Technology (NSRIT),  
Vishakhapatnam, India<sup>2,3,4,5,6</sup>

**ABSTRACT:** The process of identifying cancer at an early stage followed by its proper identification method leads to enhanced patient survival rates and enables doctors to create suitable treatment strategies. Next-generation sequencing technologies have developed through their technological advancements to produce extensive gene expression data which scientists use to study the molecular patterns associated with different cancer types and cancer progress stages. The combination of high-dimensional gene expression data and limited sample sizes proves to be a major obstacle for both traditional machine learning techniques and deep learning systems. The traditional models need human experts to choose their features but they fail to detect sophisticated gene interactions that scientists require to discover dependable cancer detection biomarkers which work in early stages. We developed a Transformer-Based Gene Expression Signature Prediction model which enables early detection and classification of cancer cases in this research. The framework employs a transformer architecture which contains a multi-head self-attention mechanism to learn gene relationships while it extracts essential gene expression patterns without needing prior feature selection. The system uses genomic data analysis techniques to connect distant genomic information which enables better biomarker detection and more accurate early stage cancer type identification. Our research demonstrates that the transformer-based approach outperforms traditional machine learning and deep learning methods in all evaluation metrics which include accuracy and precision and recall and F1-score. The model decreases misclassification rates while it shows the gene expression patterns that relate to cancer. The research results demonstrate that transformer-based models present a strong method for analyzing gene expression data which enables better early cancer detection results and better support for decision-making purposes

**KEYWORDS:** Gene Expression, Transformer Model, Deep Learning, Cancer Detection, Cancer Classification, Biomarker Identification, Early Cancer Detection, Genomic Data Analysis.

## I. INTRODUCTION

Genes serve as the basic genetic transmission elements which hold the genetic material that living organisms inherit. The cell's gene expression processes control its biological functions while overseeing its various cellular activities. Genetic disorders develop when mutations or any other abnormal changes occur in genes because these alterations disrupt normal gene expression patterns. Cancer represents one of the most important genetic disorders because it develops when normal genes called proto-oncogenes undergo mutations that create oncogenes which cause cells to multiply uncontrollably and develop tumors. Global health reports indicate that cancer remains one of the primary global death causes while the disease will experience substantial growth throughout the next fifty years. The early detection process together with the precise cancer type classification process establishes a foundation which leads to enhanced treatment results and longer life expectancy for patients. Next-generation sequencing (NGS) technological advancements enable researchers to create extensive gene expression datasets which deliver crucial data about cancer molecular operating systems. The datasets show expression patterns for thousands of genes which researchers study to discover crucial biomarkers that demonstrate cancer development and progression. Gene expression data present a challenging situation because they contain excessive dimensionality which makes it difficult to analyze their data through ordinary techniques. Traditional machine learning techniques often struggle to handle such data effectively because they require extensive preprocessing and feature selection. Deep learning methods have improved complex



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

biomedical dataset analysis but researchers still find it difficult to capture long-range gene relationships. The recent development of transformer-based deep learning models has attracted research interest because these models demonstrate the capability to understand the relationships between different contextual elements.

### 1.1 Classification using Conventional Machine Learning

Cancer classification through gene expression data uses traditional machine learning (ML) methods. The classification algorithms are executed after the feature selection methods identify the relevant gene features from the available gene data. The common ML algorithms available for cancer prediction include Naïve Bayes and k-Nearest Neighbor (KNN) and Random Forest (RF) and Logistic Regression (LR) and Support Vector Machine (SVM). The models use gene expression data to determine whether samples show cancerous characteristics or they show different types of cancer. Multiple studies have used these algorithms to classify cancer subtypes across breast cancer and lung cancer research datasets. SVM has emerged as the method that provides better classification performance through its precise predictions which outperform other available methods. The conventional ML methods face multiple challenges when they attempt to process genomic data through their established procedures. The gene expression datasets contain multiple dimensions because they store information about thousands of genes while having only a few available samples. The classification process becomes more difficult because of this imbalance, which increases the likelihood of overfitting and decreases the chances of successful generalization. Many ML algorithms need users to provide their data in an organized format while they must conduct extensive data cleaning processes. The systems face challenges because they need to handle multiple classes and deal with extremely unbalanced training pairs. The traditional ML approaches enable researchers to obtain valuable information, but these methods fail to complete the genetic connections that link cancer development through gene interaction.

### 1.2 Deep Learning and Gene Expression Analysis

Researchers have started using deep learning (DL) methods to analyze biomedical and genomic data because artificial intelligence technology has progressed. Deep learning models learn complex patterns from raw data because they can process data without needing manual feature engineering. In recent years, Convolutional Neural Networks (CNNs) have shown promising results in analyzing gene expression data for cancer detection and classification. Multiple research studies have used CNN architectures to train models that identify cancerous and normal samples by studying patterns present in extensive gene expression datasets. Researchers developed one-dimensional and two-dimensional CNN models which they used to analyze gene expression profiles and discover potential biomarker genes that link to specific cancer types. Deep learning methods use gene expression data to create image-like data that neural networks can process for predicting patient survival and disease progression. Some frameworks use CNN-based methods while other frameworks use attention mechanisms to find out which genes matter most for cancer subtype classification. The techniques help researchers find biologically important genes that play a role in disease progression. Many existing methods need complex preprocessing procedures along with extra biological data to function properly which makes them unsuitable for use with various datasets.

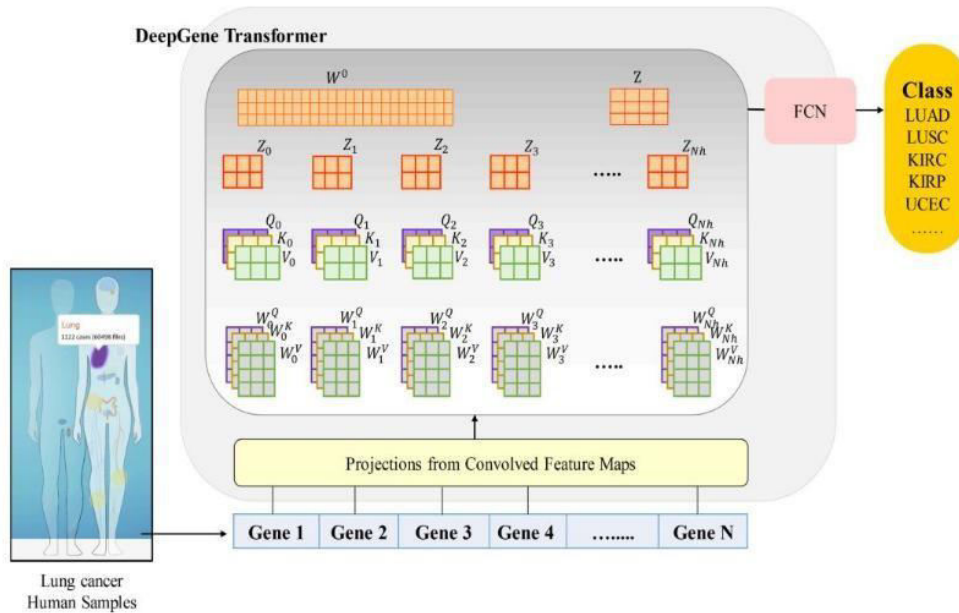
### 1.3 Proposed Approach

Despite making progress in cancer classification through deep learning models, existing methods still depend on feature selection methods which need to decrease gene expression data dimensions before they can run their models. Feature selection provides benefits by removing redundant genes which cuts down on processing needs, but it also creates multiple problems. The techniques require heavy computational power, and they risk losing essential cancer research data because they remove genetic material which helps understand cancer progression. The project introduces the Transformer-Based Gene Expression Signature Prediction model which serves to identify and classify cancers in their initial stage of development. Transformer systems employ self-attention technology which enables the system to recognize how multiple genes interact with each other. Transformers enable direct learning of useful patterns from high-dimensional data because they do not need extensive feature selection processes which traditional systems need. The proposed approach uses a transformer network to analyze gene expression data which detects critical gene patterns that relate to cancer development. The system can extract valuable gene expression data through its gene interaction modeling system which boosts its ability to correctly classify cases. The end-to-end framework works to improve early cancer detection capabilities while it helps discover biomarkers which enable precise medical treatment and clinical decision support.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



**Figure 1: Graphical Representation of the Proposed Transformer-Based Framework for Early Cancer Detection and Classification**

Traditional gene expression classification methods treat all features with equal importance without assessing how specific genes affect cancer subtype identification. The attention mechanism enables the model to concentrate on specific gene groups which hold the highest importance for classification work and produces scores that indicate the importance of each gene. The model identifies important genes from all available genes which helps to make predictions more accurate and easier to understand.

We present a complete transformer-based system which uses gene expression data to detect cancer at an early stage and determine cancer subtypes in our research. The multi-head selfattention module enables simultaneous learning of complex genomic relationships between thousands of genes and multiple patient samples and different cancer subtypes. The model achieves better performance on both binary classification and multiclass classification tasks because attention heads work together to improve its performance on biased datasets.

This research produces the following main research outcomes:

The Transformer-based Gene Expression Signature Prediction system uses a multi-head selfattention system together with one-dimensional (1D) convolution layers to handle highdimensional gene expression data. This system delivers strong data representations without the need of initial feature evaluation.

Our research shows that a multi-head self-attention layer which has the right configuration can function as a substitute for traditional 2D convolutions. This new approach requires less computational resources which makes it possible to handle extensive genomic data sets.

This framework represents an early attempt to build a complete end-to-end model for cancer subtype classification through direct learning from raw gene expression data which enables early disease detection and biomarker discovery.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### II. MATERIAL AND METHODS

#### 2.1 METHODOLOGY

The proposed system employs a Transformer-based deep learning framework to predict gene expression signatures for early cancer detection and classification. The methodology effectively processes high-dimensional genomic data while it detects complex gene interactions to enhance prediction results. The workflow starts with the process of collecting data and preparing it for analysis. Researchers gather gene expression datasets from genomic databases that are accessible to the public which include expression data for thousands of genes across different samples. The raw data undergo preprocessing steps which include normalization and missing value handling and noise reduction to achieve data quality and consistent results. After preprocessing the dataset gets divided into two parts which serve as training data and testing data for model creation and assessment. The process includes the creation of feature representations and the generation of embedding representations. The system represents each gene expression profile as a high-dimensional feature vector which contains the molecular characteristics of a sample. The system converts these vectors into embedded representations which the transformer architecture can process with high efficiency. The system operates through its core component which uses a Transformer model that implements a multi-head self-attention system. The system uses attention weights to allow the architecture to identify important gene characteristics which enable it to build remote connections and recognize complex genetic patterns. The attention mechanism allows the model to discover key gene expression patterns which lead to cancer development and disease advancement. During the training phase the model learns to separate samples into two groups: cancer and non-cancer and then further divides cancer samples into their respective subtypes. The system evaluates model performance through various metrics which include accuracy and precision and recall and F1-score.

#### 2.2 Tumor Gene Expression Dataset

The researchers used The Cancer Genome Atlas TCGA which is a cancer genomics project that supplies extensive molecular data and clinical information for cancer research purposes as their data source. The TCGA database provides multiple biological data types which include mRNA expression data and medical imaging results and DNA methylation patterns and copy number variation data and somatic mutation information and protein expression data. TCGA stores genomic and molecular data from more than 20000 primary tumor specimens which come with corresponding normal tissue samples that represent 33 distinct cancer types.

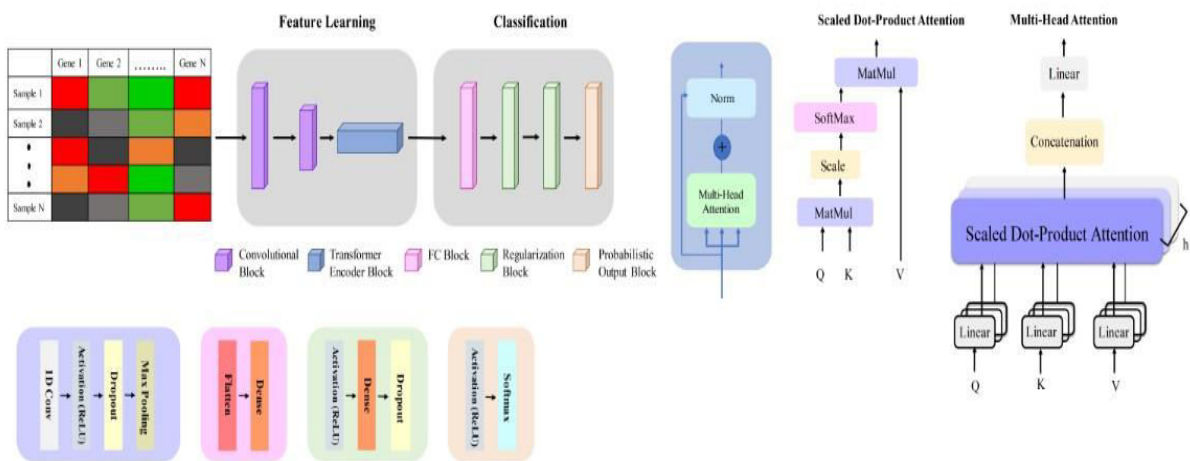


Figure2. Proposed Transformer-based Gene Expression Classification Architecture.

The genome characterization procedure produces various data types which consist of clinical data and molecular study metadata. The research team analyzed RNA-sequencing RNA-seq expression data from lung cancer datasets which included lung adenocarcinoma LUAD and lung squamous cell carcinoma LUSC. The researchers used these datasets to create a model which could identify and classify lung cancer subtypes through analysis of gene expression patterns.



## International Journal of Innovative Research in Computer and Communication Engineering (IJRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Researchers measured gene expression levels through the transformation of  $\log_2$  FPKM 1 which uses FPKM Fragments Per Kilobase of transcript per Million mapped reads as a standard method to normalize RNA-seq data. The transformation process establishes a stable data variance while decreasing the impact of high expression values which makes the data suitable for machine learning applications.

The dataset includes cancer tissue samples and normal tissue samples for each lung cancer subtype. The supplementary material presents a distribution diagram of these samples which can be found in Figure S1.

### 2.3 Data Preprocessing

Data preprocessing serves as a necessary process which prepares input data for successful machine learning and deep learning applications. The researchers executed multiple preprocessing procedures to establish clean and organized gene expression data which they required for their model training.

The raw dataset underwent initial processing which eliminated all categorical rows and columns that did not serve any purpose for model development. The feature matrix was constructed by excluding all rows which contained gene identifiers and all columns which contained categorical data about different class labels. The attributes which were eliminated from the dataset enabled the researchers to obtain class labels for each sample which they used as target variables in their classification work.

The research study develops a universal classification system which learns from complete gene expression data instead of observing specific cancer-related biomarkers for research purposes. The research team conducted preprocessing without using manual biomarker filtering methods.

The research team conducted an analysis of the dataset to identify any missing data and existing discrepancies. The research team took proper steps to handle any missing or invalid data in order to maintain the quality of the dataset. The research team organized and cleaned the data before they divided the complete dataset into three separate parts.

Training set – used to train the model

Validation set – used to tune model parameters and prevent overfitting

Testing set – used to evaluate the final performance of the model

The preprocessing pipeline transforms data into its required format which allows users to input data into the deep learning system.

## III. RESULTS

### 3.1 Implementation and Training Setup

The researchers used Python to create a deep learning system which utilized Scikit-learn and Keras libraries while TensorFlow operated as the core processing engine. The training procedure used GPU resources to speed up its operations which resulted in better model outcomes. The testing process used a workstation which contained an Intel Xeon i7-8700K processor running at 3.70 GHz and 128 GB of RAM and an NVIDIA GTX 1080 Ti GPU. The system specifications of this setup made it possible to process the complex gene expression data which required high dimensional analysis. The training process involved assessing different hyperparameter settings to find optimal system settings for the proposed architectural design. The study tested both general deep learning hyperparameters and specific parameters which control the attention mechanism.

The study investigated the following parameters:

Learning rate, Network depth, Dropout rate, Number of attention heads, Attention key dimension, Attention value dimension.

The researchers tested multiple parameter combinations through random search and crossvalidation methods to determine which model configuration produced the best results.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The system reached its highest performance level after completing 95 training epochs. The training process used an early stopping method which halted operations when five consecutive epochs showed no improvement in validation accuracy.

### I. Evaluation Metrics

The proposed model evaluation required multiple performance metrics for its effectiveness assessment. The selected metrics demonstrate effectiveness in handling imbalanced datasets which frequently occur in biomedical research. The following evaluation metrics were used:

1. Precision-Precision measures how many of the predicted positive samples are actually correct.
2. Recall-The model successfully identified actual positive samples according to the recall measurement.
3. F1 Score-The F1 score combines precision and recall through its harmonic mean to establish an equitable evaluation of classification effectiveness.
4. Matthews Correlation Coefficient (MCC)-The Matthews Correlation Coefficient evaluates the quality of binary classifications and is especially useful for imbalanced datasets. The MCC value ranges between -1 and +1: +1 indicates perfect prediction, 0 indicates random prediction and -1 indicates complete disagreement between predictions and observations

The research team conducted five training sessions for the model. They used the model's average accuracy from all training sessions to report their findings.

### 3.2 Data Splitting Strategies

The research implemented distinct data partitioning methods to handle both binary and multiclass classification tasks.

#### Binary Classification

The dataset for the binary classification task which examined lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) was divided into three parts.

70% Training set, 20% Validation set and 10% Testing set. The system learned from most data while keeping different datasets ready for both validation and assessment purposes.

#### Multiclass Classification

The research evaluated three data partitioning methods for multiclass classification which used normal tissue LUSC and LUAD as classifications.

60% Training – 20% Validation – 20% Testing, 70% Training – 20% Validation – 10% Testing

And 80% Training – 10% Validation – 10% Testing

The experimental results demonstrated that increasing training data size boosts model performance, and the 80–10–10 split resulted in optimal outcomes.

### 3.3 Comparison with Traditional Machine Learning Methods

Gene expression datasets show high dimensionality because they include thousands of gene features but only provide a small number of patient samples. The data imbalance causes traditional data mining techniques to experience overfitting problems.

Machine learning methods solve this problem through their combination of feature selection techniques with dimensionality reduction techniques which they apply before performing classification. The research study evaluated the proposed model against traditional machine learning algorithms which used feature selection methods as their base.

Feature Selection Methods, Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), Mutual Information (MI), Classification Algorithms, Logistic Regression (LR) Random Forest (RF), Support Vector Machine (SVM) . The researchers used a 5-repeated 3fold cross-validation method to test the effectiveness of each testing procedure. The experimental results showed that the proposed CNN-Transformer system achieved better results than all conventional machine learning methods. The model achieved: The system reached 98% testing accuracy when using a 4-head attention setup. The system reached 99% testing accuracy when using an 8-head attention setup.

The best results from standard methods occurred when PCA combined with SVM to reach an Area Under Curve (AUC) value of 0.93.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

### 3.4 Results and Performance Analysis

The proposed architecture achieved strong performance in multiclass classification tests because it successfully handled all tested data partitioning methods. The best results were obtained when 80% of the data was used for training, indicating that the model benefits from larger training datasets. The box plots of cross-validation scores and MCC values showed that the proposed architecture achieved class-specific MCC scores which exceeded traditional machine learning methods by approximately 10 percent. The team tracked validation accuracy and loss changes during the training process. The results showed that validation accuracy steadily improved throughout training, indicating that the model successfully learned meaningful patterns from the gene expression data. The accuracy curves maintained their stability throughout cross-validation folds, which showed that the model predicted better than random guessing and achieved consistent results.

### IV. RESULTS SCREENSHOTS

Register page:

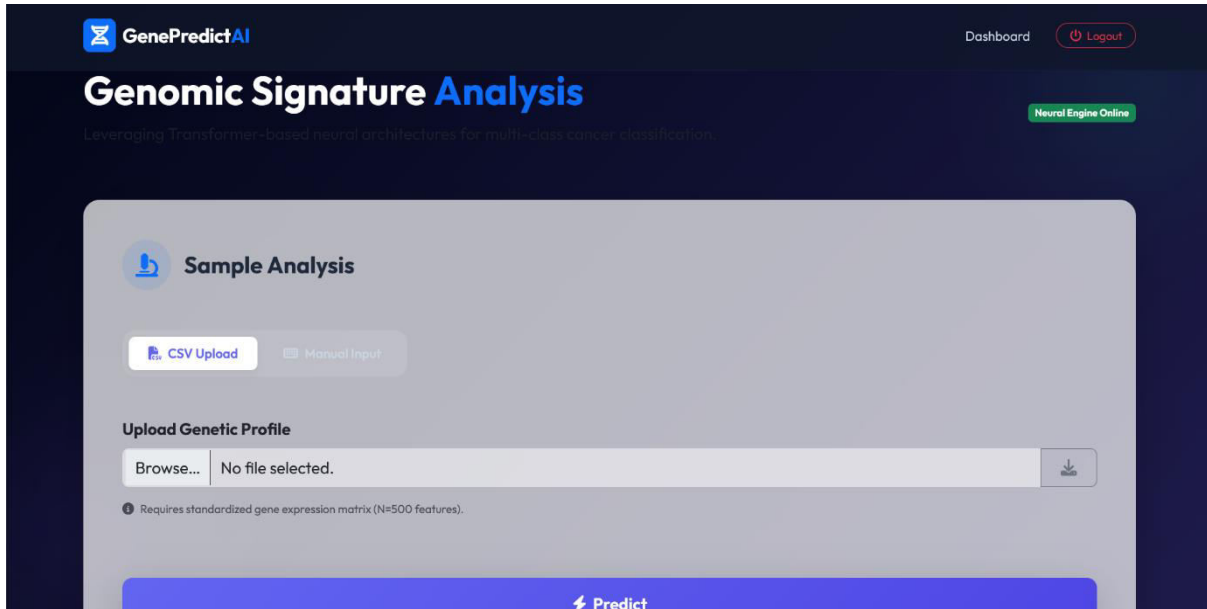
Login page:



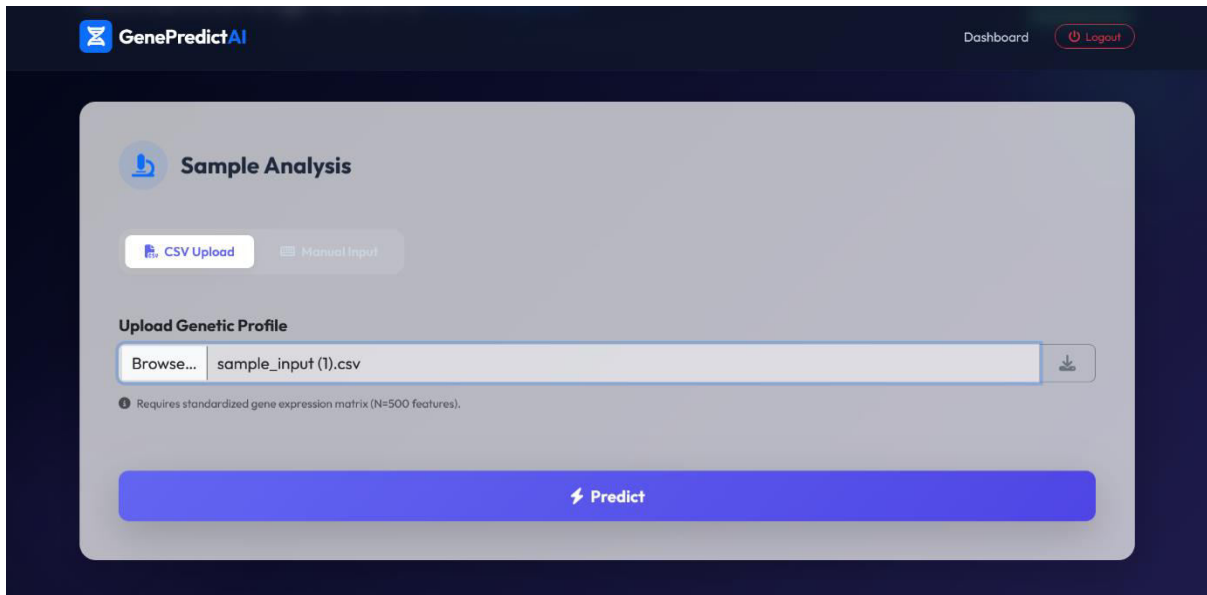
# International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Home page:



Preprocessed Data:

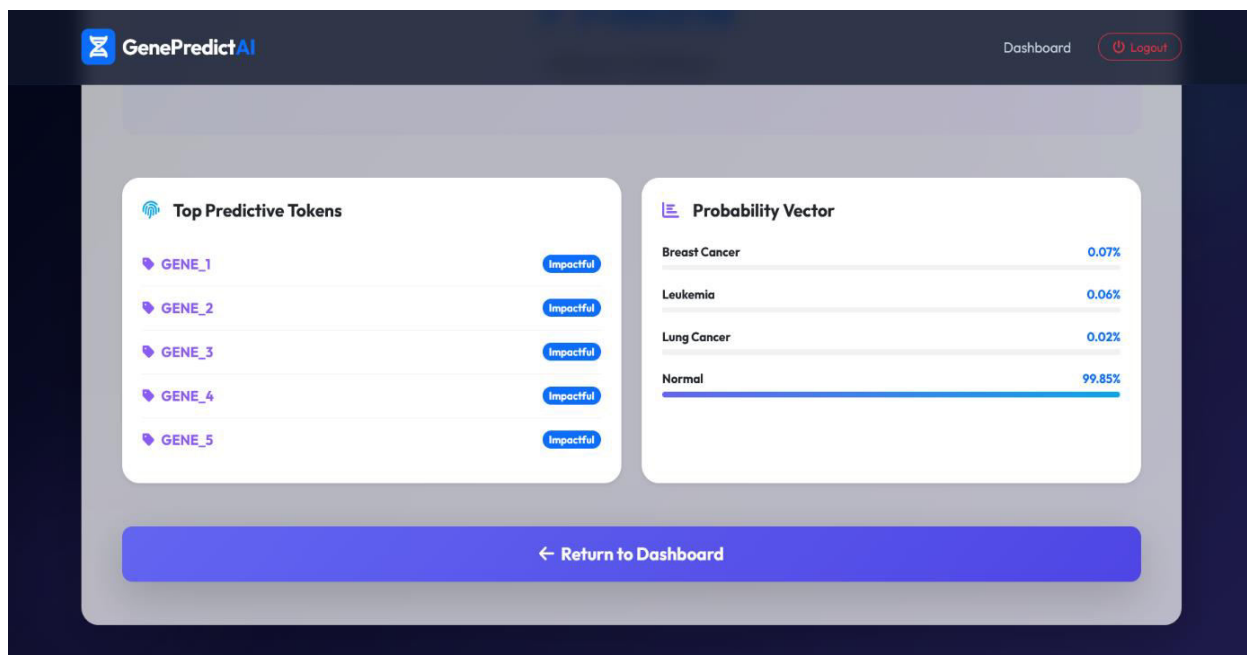
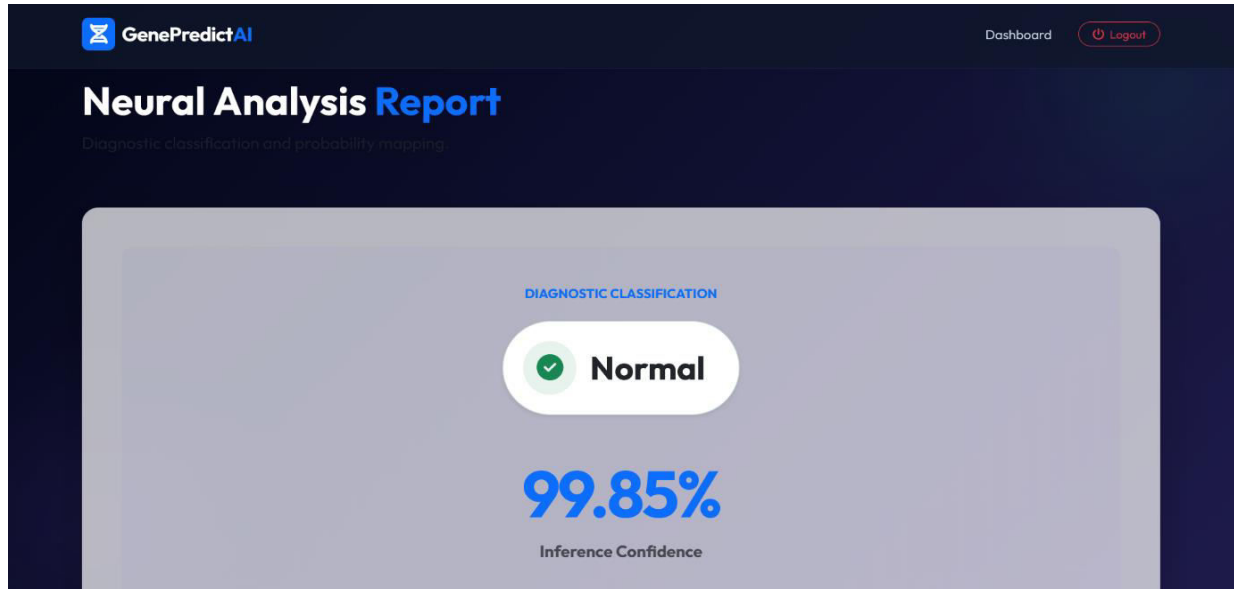




# International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Output page:



## IV. DISCUSSION

Deep learning models need attention mechanisms because they serve essential functions in natural language processing and machine translation tasks. The study developed a multi-head self-attention (MHSA) mechanism for use with a 1D CNN architecture to enhance cancer subtype classification accuracy based on gene expression data.

The traditional method requires researchers to choose specific genes which they believe connect to cancer development. The method fails to identify critical biological information which exists within the unselected genes. The proposed framework implements an end-to-end learning system which processes entire gene expression datasets without performing any feature selection.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The model uses convolutional layers together with self-attention mechanisms to study both local gene connections and worldwide gene patterns. The attention mechanism uses importance weights to evaluate specific genes which enables the model to identify essential features necessary for cancer subtype classification.

The multi-head attention structure enables different attention heads to discover separate gene relationships which creates a more comprehensive and biologically relevant data representation.

The DeepGene Transformer framework which we developed shows strong abilities to enhance cancer subtype prediction results through its analysis of complex transcriptomic data.

### V. CONCLUSION

Lung cancer remains one of the most serious health challenges worldwide, with a very high incidence and mortality rate. The precise identification of lung cancer subtypes through early detection is crucial for advancing diagnostic methods and treatment strategies and enhancing patient survival rates. Cancer subtype prediction in traditional machine learning models requires researchers to use feature selection methods as their primary approach for reducing gene expression data dimensionality which leads to classification. The methods used for this task face difficulties in finding critical biological data because they only perform effectively with particular datasets that contain evenly distributed data.

The research presents a Transformer-based gene expression classification framework which combines 1D Convolutional Neural Networks (1D CNN) and Multi-Head Self-Attention (MHSA) technology. The proposed architecture processes high-dimensional gene expression data in an end-to-end manner, eliminating the need for manual feature selection. The convolutional layers extract vital local structures from gene expression vectors, whereas the attention mechanism determines how genes interact with each other, helping the model to identify the critical features that lead to accurate cancer subtype classification.

The experimental results showed that the proposed model achieved high accuracy rates for classifying both binary and multiclass cancer subtypes. The hybrid CNN-Transformer architecture showed better performance compared to traditional machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machines combined with feature selection methods. The results show that attention mechanisms work well with convolutional networks to process biological datasets which have high dimensionality.

The provided framework shows significant capabilities because it can enhance cancer subtype classification through its evaluation of gene expression data. Through its attention mechanism the model reveals how gene expression patterns connect to patient samples which enables researchers to comprehend how different genes function in cancer development.

The framework needs extension because it should enable classification of additional cancer types and analysis of extensive multi-omics datasets. The predictive performance will improve when researchers combine gene expression data with histopathological images and clinical parameters because this approach will create more accurate cancer diagnosis and prognosis studies.

### REFERENCES

1. World Health Organization, *WHO report on cancer: Setting priorities, investing wisely and providing care for all*, 2020.
2. Kwang-Hyun Yu and Michael Snyder, "Omics profiling in precision oncology," *Molecular & Cellular Proteomics*, vol. 15, no. 8, pp. 2525–2536, 2016.
3. Michael Snyder, *Genomics and Personalized Medicine: What Everyone Needs to Know*. Oxford University Press, 2016.
4. William D. Travis, Elisabetta Brambilla, and Gregory J. Riely, "New pathologic classification of lung cancer: relevance for clinical practice and clinical trials," *Journal of Clinical Oncology*, vol. 31, no. 8, pp. 992–1001, 2013.



## International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. Stephan C. Schuster, "Next-generation sequencing transforms today's biology," *Nature Methods*, vol. 5, no. 1, pp. 16–18, 2008.
6. Douglas Hanahan and Robert A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.
7. Andrea Sboner et al., "Molecular sampling of prostate cancer: a dilemma for predicting disease progression," *BMC Medical Genomics*, vol. 3, no. 1, 2010.
8. Rafael M. Luque-Baena et al., "Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data," *Theoretical Biology and Medical Modelling*, vol. 11, no. 1, 2014.
9. Arun Natarajan and T. Ravi, "A survey on gene feature selection using microarray data for cancer classification," *International Journal of Computer Science & Communication*, vol. 5, no. 1, pp. 126–129, 2014.
10. Nguyen T. Duc et al., "3D deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, 2020.
11. Sandeep Dubey et al., "Anomalous event recognition in videos based on joint learning of motion and appearance," *Applied Sciences*, vol. 11, no. 3, 2021.
12. Martin Popel et al., "Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals," *Nature Communications*, vol. 11, no. 1, 2020.
13. Andre Esteva et al., "Deep learning-enabled medical computer vision," *npj Digital Medicine*, vol. 4, no. 1, 2021.
14. Guangyu Liu, *Beyond predictive modeling: new computational aspects for deep learning based biological applications*, Massachusetts Institute of Technology, 2020.
15. J. Wu and C. Hicks, "Breast cancer type classification using machine learning," *Journal of Personalized Medicine*, vol. 11, no. 2, 2021.
16. Fang Hu et al., "Gene expression classification of lung adenocarcinoma into molecular subtypes," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 4, pp. 1187–1197, 2019.
17. S. Tarek et al., "Gene expression based cancer classification," *Egyptian Informatics Journal*, vol. 18, no. 3, pp. 151–159, 2017.
18. M. Mostavi et al., "Convolutional neural network models for cancer type prediction based on gene expression," *BMC Medical Genomics*, vol. 13, 2020.
19. Keras-vis contributors, "Keras-vis visualization toolkit," GitHub repository, 2017.
20. G. López-García et al., "Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data," *PLOS ONE*, vol. 15, no. 3, 2020.
21. F. Gao et al., "DeepCC: a novel deep learning-based framework for cancer molecular subtype classification," *Oncogenesis*, vol. 8, no. 9, 2019.
22. Amir Beykikhoshk et al., "DeepTRIAGE: interpretable biomarker scores using attention mechanisms for classification of breast cancer subtypes," *BMC Medical Genomics*, vol. 13, 2020.
23. M. Bicakci et al., "Metabolic imaging based sub-classification of lung cancer," *IEEE Access*, vol. 8, pp. 218470–218476, 2020.
24. J.-H. Chiang and S.-H. Ho, "A combination of rough-based feature selection and RBF neural network for classification using gene expression data," *IEEE Transactions on Nanobioscience*, vol. 7, no. 1, pp. 91–99, 2008.
25. Shangqian Liu et al., "Feature selection of gene expression data for cancer classification using double RBF-kernels," *BMC Bioinformatics*, vol. 19, 2018.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



SJIF Scientific Journal Impact Factor



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING



9940 572 462



6381 907 438



ijircce@gmail.com



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details