IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 November 2024, accepted 23 December 2024, date of publication 31 December 2024, date of current version 10 January 2025.

*Digital Object Identifier 10.1109/ACCESS.2024.3524564*

RESEARCH ARTICLE

# An Empirical Analysis of Transformer-Based and Convolutional Neural Network Approaches for Early Detection and Diagnosis of Cancer Using Multimodal Imaging and Genomic Data

**S. K. B. SANGEETHA[1], SANDEEP KUMAR MATHIVANAN[2], V. MUTHUKUMARAN[3], JAEHYUK CHO[4], AND SATHISHKUMAR VEERAPPAMPALAYAM EASWARAMOORTHY[5], (Member, IEEE)**

[1]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus, Chennai, Tamil Nadu 600026, India
[2]School of Computer Science and Engineering, Galgotias University, Greater Noida 203201, India
[3]Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India
[4]Department of Software Engineering and Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju-si 54896, Republic of Korea
[5]School of Engineering and Technology, Sunway University, Petaling Jaya, Selangor Darul Ehsan 47500, Malaysia

Corresponding author: Jaehyuk Cho (chojh@jbnu.ac.kr)

This work was supported by Korea Environmental Industry and Technology Institute (KEITI), with a grant funded by Korea Government, Ministry of Environment [The Development of the Internet of Things (IoT) Technology for Collecting and Managing Big Data on Environmental Hazards and Health Effects], under Grant RE202101551.

**ABSTRACT** Early diagnosis of cancer has focused on the use of advanced algorithms to achieve accurate diagnosis. The proposed study assesses the effectiveness of Transformer-based models and Convolutional Neural Networks (CNN) in cancer diagnosis with respect to multimodal imaging and genomic data. The performance comparisons between the two algorithmic methods with such complex datasets, which combine multi-modal imaging and genomic information, are presented. In search of the optimal neural network configuration, a series of experiments were conducted with respect to different layers, attention mechanisms in case of transformers, and convolutional architectures in case of CNNs. Besides, parameters related to training, such as learning rates, batch sizes, and optimization algorithms, have also been systematically tuned. The different models were evaluated against accuracy, precision, recall, and the F1-score. Our results show that the proposed multimodal model, with accuracy from 92.5 to 93.2, F1-scores between 91.5 and 92.2, precision of 91.5 to 92.2, and recall values of 92.5 to 93.2. In contrast, much lower accuracy, F1-scores, precision, and recall values were noticed when using baselines, especially VGG. All these findings indicate the fact that the presented techniques, especially the Multimodal and Transformer models, are more robust solutions for classification tasks with better balance between precision and recall, as well as with higher overall accuracy. This came with the cost of the expense of computational resources: CNNs are less resource-intensive but have competitive performance with better precision and recall. The results underline how algorithm selection and hyperparameter optimization play a crucial role in cancer detection tasks. This study has shown how state-of-the-art deep learning methods can be effectively combined with multi-modal data for building more accurate and efficient systems in cancer diagnosis. Two main lines of future work would be improving these algorithms and understanding their applicability in real clinical practice to obtain maximum benefits from them.

**INDEX TERMS** Transformer-based models, convolutional neural networks (CNNs), cancer detection, multimodal imaging, genomic data analysis.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

## I. INTRODUCTION

The groundbreaking step that advanced algorithms in cancer detection take in changing the face of medical diagnosis to better early detection, improve diagnostic accuracy, and give way to more personalized treatment strategies [1]. Artificial Intelligence (AI) has made it possible to demonstrate impressive improvements in image analysis, interpretation of genomic data, and in patient management in many medical fields [2]. One of the growing concerns is the inefficient tuning of these algorithms to handle the multimodal complexity of cancer detection data [3].

Only ensuring the capacity of the AI systems to process and integrate data from diversified sources have been extremely helpful in the domain of cancer detection, one of the fields of early diagnosis [4]. The integration of a broad spectrum of imaging modalities, such as Magnetic Resonance Imaging (MRI), Computer Tomography (CT) scans, and Positron Emission Tomography (PET) scans, with genomic information, in turn, will make it possible to deeply study the disease condition, in terms of increased ability for structural, functional, and molecular studies [5]. Multimodal approach integrates the detailed structure, functions, and molecule insights, which is mandatory in the identification of cancerous situations occurring at early stages and providing individually adaptive treatment schemes accordingly [6].

The Transformers have really drawn a lot of attention, partly due to their attention mechanisms and also their ability to model long-range dependencies; moreover, they have proven to have great potential for the processing of sequential data and integration of different data types [7] [8]. The Transformer-based model has shown tremendous capacity for the embedding and analysis of genomic data with imaging data besides attention mechanisms of operations over relevant features when applied to various data types. This will be most relevant and important for the unearthing of subtle patterns and correlations potentially linked with early-stage cancer [9]. On the other hand, the supremacy by which CNNs can extract fine-grained information from medical images for anomaly detection and diagnosis of cancerous lesions remains unchallenged [10].

There are quite a few factors that might influence the performance of these models: the quality and diversity of the data, neural network architecture, and training parameter optimization [11]. Some powerful Transformer-based models often require large computational resources to train and may be somewhat challenging to finally tune [12]. While they are much less costly in terms of high computational resources, CNNs still need to be designed for the objective of capturing salient features of images without overfitting or underfitting [13].Therefore, to make it really appealing to integrate those models into routine clinical practice, it is of utmost importance that they provide reliable, actionable insights.

The study assesses their performance in terms of accuracy, precision, recall, and computational efficiency to see if effective ways for the early diagnosis of cancer will be identified. In this assignment, sets of experiments are conducted to provide a detailed study on neural network configurations, layer architectures, attention mechanisms for transformers, convolutional structures for CNNs, and changes to the training parameters, such as learning rates, batch sizes, and optimization algorithms, to identify the impacts of these parameters on the performance of the models [14].

The outcome from this study brings insight into the strengths and limitations of each algorithmic approach that can help guide future developments in cancer detection technologies [8], [15]. Comparative analysis of these approaches is done to form the basis of this study aimed at developing more precise and reliable tools for diagnosing diseases, hence being able to detect cancer at an early stage to salvage patient lives. Not only the algorithmic performance but also the applicability of the models with respect to a clinical workflow setting are also assessed. It focuses on meeting computational resource requirements, model interpretability, and data privacy, while ensuring feasible solutions within the regulations for healthcare standards [16].

Advanced AI techniques integrated with multimodal data can definitely move the cancer diagnosis process toward greater heights, improving overall health care delivery [17]. In this regard, the present study further integrates the strength of the Transformer-based models with that of the CNNs in order to form a pathway for developing more effective systems in cancer detection and thus help in better care and improved results for the patients [18].This study is ideally positioned to meaningfully contribute to a better future in oncology and healthcare delivery simply by pushing the forefront of state-of-the-art cancer detection through rigorous evaluation and optimization of Transformer-based models and CNNs.

The main objectives of the study are

1. To compare how well Transformer-based models and Convolutional Neural Networks (CNNs) detect cancer using imaging and genomic data.
2. To find the best settings for Transformer-based models and CNNs to improve the cancer detection accuracy.
3. To evaluate how feasible it is to use Transformer-based models and CNNs in real-world cancer diagnosis, focusing on their computational needs and ease of integration.

## II. EMERGING TRANSFORMER-BASED AND CNN MODELS FOR ACCURATE CANCER DETECTION

Architectural advances in neural networks have visibly improved cancer detection. Specifically, in the case of grid-like data, such as medical images, CNNs are very effective. They are extremely good at extracting spatial features due to their hierarchical structure and hence really good at image classification tasks. However, the effectiveness of CNN requires large datasets and substantial computational resources for training, which is a limitation within resource-constrained environments [19].

On the other hand, transformer networks have an intrinsic ability for dealing with sequence data since

their self-attention mechanisms enable them to capture a long-range dependency within the data. That will be very helpful in integrating multimodal data, especially with imaging and genomic information. However, these transformer networks require high computational demands and intricate training procedures to be met, making scalability and resource allocation a problem [20].

**TABLE 1. Comparison of neural network architecture for cancer detection.**

| Architecture | Description | Strengths | Limitations |
|---|---|---|---|
| **Convolutional Neural Networks (CNNs)** | Models designed for processing grid-like data (images) | Excellent at extracting spatial features; effective for image classification | Requires large datasets; computationally intensive |
| **Transformer Networks** | Models that use self-attention mechanisms for sequence processing | Superior at capturing long-range dependencies; efficient for multimodal data | High computational requirements; complex to train |
| **Hybrid Models** | Combination of CNNs and Transformers | Leverages both spatial and contextual information; adaptable for diverse data | Increased model complexity; requires careful tuning |

Hybrid models, such as those combining CNNs with transformers, have been proposed for the advantages of these architectures to be beneficial. The hybrid approach where spatial features are extracted and joined with contextual understanding can fit very diverse data types, hence improving the overall detection performance. However, adding complexity to the hybrid model means that careful tuning and optimization must be done in order to balance performance with computational efficiency. Table 1 depicts the comparison of all these three models.

As shown in Table 2, there are many challenges in the integration of the Transformer-based and CNN models. The most prominent one is the high computational demand required for training and inference. In this regard, Graphical Processing Unit (GPU) acceleration and model pruning techniques can aid in consuming fewer resources and increasing efficiency [21]. Another challenge lies in data integration, especially on multimodal data, such as imaging and genomic information. Robust data fusion techniques and comprehensive preprocessing steps need to be developed to ensure that data integration and the subsequent analysis are effectively done [22].

**TABLE 2. Challenges in integrating Transformer-based and cnn models for cancer detection.**

| Challenge | Description | Potential Solutions |
|---|---|---|
| **Computational Resources** | High computational demands for training and inference | Utilize GPU acceleration; implement model pruning |
| **Data Integration** | Combining multimodal imaging and genomic data effectively | Develop robust data fusion techniques; preprocess data |
| **Model Complexity** | Increased complexity in hybrid models | Simplify model architecture; optimize hyperparameters |
| **Interpretability** | Difficulty in understanding model decisions | Implement explainable AI methods; use attention maps |

**TABLE 3. Comparison of multimodal data types for cancer detection.**

| Data Type | Description | Benefits | Limitations |
|---|---|---|---|
| **Imaging Data** | MRI, CT scans, X-rays | Provides detailed visual information | High dimensionality; requires significant processing power |
| **Genomic Data** | DNA/RNA sequencing, gene expression data | Offers insights into genetic mutations | Data can be noisy and complex to interpret |
| **Clinical Data** | Patient records, lab results | Includes comprehensive patient history | Variability in data quality and completeness |
| **Textual Data** | Medical reports, electronic health records | Provides contextual information about patient conditions | Often unstructured and requires NLP processing |

Another challenge is model complexity, since by design, hybrid models are more complex. In this respect, simplification in model architectures and the optimization of hyperparameters assume importance to effectively manage model complexity while maintaining its performance [23]. The last major problem is that of interpretability. It can be difficult to understand model decisions using even the most modern neural networks. One can include explainable AI

**TABLE 4.** Summary of datasets.

| Dataset Name | Description | Modalities Included | Use Cases | Challenges |
|---|---|---|---|---|
| TCGA | Comprehensive multimodal cancer dataset | MRI, CT, DNA sequencing, RNA-Seq | Cancer subtype classification, mutation analysis | Data heterogeneity, high dimensionality |
| LIDC-IDRI | Lung cancer dataset with CT scans and genomic data | CT scans, gene expression | Lung nodule detection, tumor classification | Inter-observer variability, balancing image and genomic data |
| TCIA | Imaging data linked with genomic profiles | MRI, PET, CT, DNA sequencing | Tumor detection, segmentation | Image segmentation accuracy, data fusion |
| GEO | Repository of gene expression data | Microarray, RNA-Seq | Gene expression analysis, biomarker discovery | Data normalization, integration challenges |
| ICGC | Genomic and epigenomic data across cancers | DNA sequencing, methylation, RNA-Seq | Mutation analysis, cancer genome characterization | Large-scale data processing, privacy |
| METABRIC | Breast cancer-focused dataset | DNA sequencing, RNA-Seq, clinical data | Subtype classification, survival analysis | High dimensionality, integration of clinical data |

methods and use attention maps in order to make the models more transparent and interpretable [24].

In the process of detection, a variety of multimodal data as depicted in Table 3 are applied to enhance its accuracy. Imaging data, like MRI, CT, and X-rays, give visual information at a deep level that is useful for the detection of tumors and other anomalies. Since imaging data has high dimensionality and high processing is required [25].

## III. COMPREHENSIVE OVERVIEW OF MULTIMODAL IMAGING AND GENOMIC DATASETS FOR EARLY CANCER DETECTION

### A. DATA DESCRIPTION

For early detection and diagnosis of cancer using multi-modal imaging and genomic data, few datasets provide deep multimodal data, matching imaging modalities such as MRI, CT, and PET scans with genomic information like DeoxyriboNucleic Acid(DNA) sequencing and gene expression profiles are used.

#### 1) MULTIMODAL IMAGING DATASETS

##### a: TCGA (THE CANCER GENOME ATLAS)

TCGA contains imaging, genomic sequences, and clinical information modalities against different cancer types. Modalities present are MRI, CT scans, DNA sequencing, and RNA-Seq. Commonly, it is used for tasks such as cancer subtype classification, mutation analysis, and prognosis prediction. However, heterogeneity of data, high dimensionality, and variability in patients are the challenges.

##### b: LIDC-IDRI (LUNG IMAGE DATABASE CONSORTIUM AND IMAGE DATABASE RESOURCE INITIATIVE)

The dataset includes a large set of annotated CT images for lung cancer by radiologists for the presence of nodules, combined with genomic information. CT image data and gene expression data are modalities present in this dataset. It can be used for the detection of lung nodules, classification of tumors, and survival prediction. The challenges within the dataset include inter-observer variability and the balancing of the quality of images with genomic data.

##### c: TCIA (THE CANCER IMAGING ARCHIVE)

TCIA provides a large corpus of imaging data that are correlated with genomic and clinical data and exist in modalities such as MRI, PET, and CT scans. It would apply in tasks like detection, segmentation, and treatment planning of tumors. Modalities include MRI, PET, and CT scans, among others. Inherited DNA sequencing is also available. Some challenges could be image segmentation and fusion of multimodal data.

**TABLE 5.** Common features identified for cancer detection.

| Feature Type | Description | Dataset(s) |
|---|---|---|
| Tumor Morphology | Shape, size, texture of tumors | TCGA, LIDC-IDRI |
| Radiomic Features | Intensity, wavelet transformations, heterogeneity | TCIA, LIDC-IDRI |
| Functional Imaging Metrics | Metabolic activity metrics (e.g., SUV) | TCIA |
| Gene Expression Profiles | RNA-Seq data, differentially expressed genes | TCGA, GEO, METABRIC |
| Mutation Analysis | Somatic mutation identification | ICGC, TCGA |
| Epigenetic Modifications | DNA methylation patterns | ICGC |

#### 2) GENOMIC DATASETS

##### *a: GEO (GENE EXPRESSION OMNIBUS)*

GEO is an open archive of high-throughput gene expression data in a repository for free distribution. Those modalities include microarray data and RNA-Seq. The dataset used for gene expression analysis, discovery of biomarkers, and personalized medicine. The problems that have to be considered are those related to data normalization and integration with imaging data.

##### *b: ICGC (INTERNATIONAL CANCER GENOME CONSORTIUM)*

The ICGC dataset includes diversified cancer types, with metastatic profiles involving DNA sequencing and RNA-Seq modalities of the ICGC in the genomic, transcriptomic, and epigenomic characterization. The findings in this dataset are used for mutation analysis, characterization of genomic changes associated with the development of cancer, and identification of possible therapeutic targets. Two key challenges remain that are carrying out the processing of large-scale data and maintaining data privacy.

##### *c: METABRIC (MOLECULAR TAXONOMY OF BREAST CANCER INTERNATIONAL CONSORTIUM)*

METABRIC is a resource with in-depth genomic and clinical information, with survival records on the details of breast cancer. It contains types of data on DNA sequencing, RNA-Seq, and clinical data. The dataset is primarily used to predict breast cancer subtypes, survival analysis, and response to treatment. Integrating clinical data within the genome profiles, high dimensionality are some of the challenges this dataset faces.

The features make the selected datasets particularly suitable for multimodal analysis. For instance, TCGA is a dataset that has imaging data and deep genomic profiles, hence morphological and molecular data is possible. Similarly, LIDC-IDRI has detailed imaging data with genetic annotations, hence allowing for the analysis of the relationship between phenotypes in imaging and genetic mutations. Table 4 shows the comparison of various datasets along with modalities, use cases and challenges.

### B. FEATURE DESCRIPTION

#### 1) IMAGING FEATURES

##### *a: TUMOR MORPHOLOGY*

Morphology features of the tumor, like shape, size, and texture, are usually extracted from MRI and CT images. These characteristics are very fundamental in the classification of cancer and its staging; therefore, they give valuable information about the tumors.

##### *b: RADIOMIC FEATURES*

Radiomic features refer to quantitative features that are derived from imaging data, characterizing tumor heterogeneity. This may involve attributes such as intensity, wavelet transformations, fractal dimensions, and others that explain the complexity of the tumor.

##### *c: FUNCTIONAL IMAGING METRICS*

Functional imaging metrics, such as the PET-emitted standardized uptake value, are a function of the metabolic activity of tumors. These metrics distinguish malignant from benign lesions, hence helping in diagnosis.

**TABLE 6.** Properties of common features.

| Dataset | Common Features | Data Format | Specific Details | Number of Patients Considered and Region |
|---|---|---|---|---|
| TCGA | Tumor Morphology, Gene Expression Profiles, Mutation Analysis | Image, RNA-Seq, Genomic | Shape, size, texture of tumors; RNA-Seq data, differentially expressed genes; Somatic mutations | Thousands of patients across various regions (Global) |
| LIDC-IDRI | Tumor Morphology, Radiomic Features | Image | Shape, size, texture of tumors; Intensity, wavelet transformations, heterogeneity | 1,018 patients (USA) |
| TCIA | Radiomic Features, Functional Imaging Metrics | Image, PET/CT | Intensity, wavelet transformations, heterogeneity; Metabolic activity metrics (e.g., SUV) | Thousands of patients (USA) |
| GEO | Gene Expression Profiles | RNA-Seq | Differentially expressed genes | Thousands of patients across various regions (Global) |
| ICGC | Mutation Analysis, Epigenetic Modifications | Genomic, Epigenomic | Somatic mutations; DNA methylation patterns | 25,000 patients across various regions (Global) |
| METABRIC | Gene Expression Profiles | RNA-Seq | Differentially expressed genes | 2,509 patients (UK and Canada) |

#### 2) GENOMIC FEATURES

##### *a: GENE EXPRESSION PROFILES*

RNA-Seq data can be used to develop gene expression profiles to identify differentially expressed genes related to cancer subtypes, which would significantly help in developing expression-based classifiers and, ultimately, precision medicine.

##### *b: MUTATION ANALYSIS*

Mutation callers can identify somatic mutations within the DNA sequencing data, enabling the discovery of potential therapeutic targets and biomarkers for cancer treatment.

##### *c: EPIGENETIC MODIFICATIONS*

Epigenetic modifications, such as DNA methylation patterns, provide insights into epigenetic changes in cancer. Understanding these changes is vital for comprehending tumor behavior and treatment resistance. Table 6 depicts the data format and number of patients considered in common features.

## IV. SYSTEM FRAMEWORK FOR TRANSFORMER- BASED AND CONVOLUTIONAL NEURAL NETWORK

### A. PREPROCESSING

#### 1) IMAGING DATA PREPROCESSING

##### *a: MRI AND CT SCANS*

MRI and CT images are aligned using FSL's FLIRT for affine registration to correct scale, rotation, and translation differences, and Advanced Normalization Tools (ANTs) is applied for nonrigid registration to take into account local deformations. To reduce noise, Non-Local Means (NLM) filtering is applied, hence giving better results on edge preservation compared to the Gaussian blur. This smoothed out the noise but preserved the most important details in the images. Further, Gaussian smoothing is used to optimize the captured images for proper feature extraction.

For normalizing, z-score normalization is used for MRI images and Hounsfield unit scaling for CT images. Thus, intensity values are comparable across modality. The U-Net method, a deep learning-based approach is applied, to carry out tumor segmentation, allowing for the accurate isolation of tumor boundaries and enhancing the precision of subsequent analyses. Resampling was done to equalize image resolution, and bilinear interpolation is applied to uniformly adjust all the pixel sizes for harmonization of all modalities for resultant seamless integration on multimodal analysis [26], [27], [28].

##### *b: PET SCANS*

Elastix is used for non-rigid registration of the PET images toward aligning it with MRI and CT scans so that anatomical alignment is precise. Standardized Uptake Values (SUV) is standardized to parameters applicable to every patient: their body weight and the dose injected. Then, measurements between the scans are directly comparable.

### 2) GENOMIC DATA PREPROCESSING

#### a: RNA-SEQ DATA

For the RNA-Seq data, quality was assessed using FastQC and adapters as well as low-quality bases were stripped off by Trimmomatic. HISAT2 accurately presented splice junctions with aligning clean reads to the reference genome.

We obtained the raw counts and normalized them using the TPM method. That is how we corrected for differences in sequencing depth between the samples to compare it reliably across samples.

#### b: DNA-SEQUENCING DATA

We took the aligned DNA sequences against the reference genome and then proceeded with Samtools for variant calling. The output data from the sequence alignments were further filtered to exclude false positives based on allele frequency thresholds and quality scores. After filtering the variants, it is annotated using ANNOVAR to obtain their functional interpretation for further analysis towards potential biomarkers.

### 3) MULTIMODAL DATA PROCESSING

The techniques for data fusion can combine imaging data with genomic data to obtain a unified view. Feature-level fusion involves the combination of features from different modalities into one feature vector, whereas decision-level fusion refers to the combination of results from individual analyses. Similar to multimodal integration techniques, canonical correlation analysis and deep learning-based fusion methods could allow for the extraction of comprehensive insight from multimodal data. Either imputation or data augmentation may represent missing data. Imputation methods, such as mean imputation or model-based, like k-nearest neighbors, estimate these missing values from available data. Data augmentation generates synthetic data to fill in the gaps and can be very useful in cases of incomplete multimodal information [29].

Standardization is an enabling step that ensures the compatibility of various data types. Numerical features are scaled to a common range, the encoding of categorical variables is consistent, and units and formats are standardized. This makes sure that effective analysis can be conducted on the combined data from disparate sources. These are the steps in the preprocessing of a multimodal cancer detection dataset, so that the data is of high quality and ready to be analyzed. Accurate preprocessing enhances the effectiveness of the subsequent data analysis and model training for better cancer detection and diagnosis. Table 7 depicts the sample data for 10 patients and Table 8 depicts the dataset after preprocessing steps.

1. Tumor sizes were normalized using min-max scaling

$$\begin{aligned} &\textit{Normalized Tumor Size} \\ &= \frac{\textit{Tumor Size} - \textit{Min Tumor Size}}{\textit{Max Tumor Size} - \textit{Min Tumor Size}} \end{aligned} \tag{1}$$

This transformed the sizes into a range between 0 and 1.

2. Tumor shapes were encoded into categorical values based on morphology:0: Round,1: Irregular,2: Lobulated

This encoding facilitates easier integration into machine learning models.

3. Radiomic features, such as intensity, were standardized using z-score normalization

$$\textit{Scaled Intensity} = \frac{\textit{Intensity} - \mu}{\sigma} \tag{2}$$

where $\mu$ is the mean intensity across all samples, and $\sigma$ is the standard deviation. This approach not only rescales the data but also standardizes the variance, allowing for direct comparisons across diverse datasets that may originally differ in scale or range. By centering the intensity values at zero, z-score normalization effectively removes biases introduced by disparate measurement scales, ensuring that variations in radiomic intensity are reflective of true biological differences rather than artifacts of differing data collection methods.

4. Tumor heterogeneity was converted into a binary format:0: Low/Medium Heterogeneity, 1: High Heterogeneity

5. Standardized Uptake Value (SUV) were standardized using z-score normalization

$$\textit{Standardized SUV} = \frac{SUV - \mu}{\sigma} \tag{3}$$

where $\mu$ and $\sigma$ represent the mean and standard deviation of SUV values across all patients.

6. Principal Component Analysis (PCA) was applied to the gene expression data (RNA-Seq) to reduce dimensionality.PCA simplifies the dataset while retaining the most significant features, making it easier to identify patterns and insights. In this context, only the first two principal components (PC1 and PC2) are retained because they capture the largest proportion of variance, ensuring that the essential structure and relationships among the data points are preserved. Gene expression values were transformed into two principal components (PC1 and PC2), capturing the majority of variance in the dataset:

$$PCA(X) = WX \tag{4}$$

where $W$ is the matrix of principal component weights, and $X$ represents the original gene expression matrix.

7. Mutation presence was encoded as a binary variable: 0: No mutation,1: Mutation present

8. Methylation percentages were normalized using the following formula

$$\textit{Normalized Methylation} = \frac{\textit{Methylation Percentage} - \textit{Min}}{\textit{Max} - \textit{Min}} \tag{5}$$

This transformed the data to a scale between 0 and 1.

These preprocessing choices are useful in models sensitive to the magnitude of features, such as neural networks. These choices preserve the relationships between original values and ensure that no single feature, like tumor size, dominates due to large values, allowing the model to learn

**TABLE 7.** Sample patient data from the datasets.

| ID | Tumor Morphology (CT/MRI) | Radiomic Features | Functional Imaging Metrics (PET) | Gene Expression (RNA-Seq) | Mutation Analysis (DNA Sequencing) | Epigenetic Modifications (Methylation) |
|---|---|---|---|---|---|---|
| 1 | Size: 3.5 cm, Shape: Irregular | Intensity: 0.85, Heterogeneity: High | SUV: 4.2 | Gene X: 5.6 TPM, Gene Y: 2.3 TPM | TP53: Missense Mutation | 80% Methylation at CpG site 1 |
| 2 | Size: 2.8 cm, Shape: Round | Intensity: 0.65, Heterogeneity: Medium | SUV: 3.8 | Gene X: 6.1 TPM, Gene Z: 3.1 TPM | KRAS: Point Mutation | 65% Methylation at CpG site 3 |
| 3 | Size: 4.1 cm, Shape: Lobulated | Intensity: 0.75, Heterogeneity: Low | SUV: 5.0 | Gene Y: 4.8 TPM, Gene Z: 1.9 TPM | BRAF: V600E Mutation | 70% Methylation at CpG site 2 |
| 4 | Size: 3.2 cm, Shape: Irregular | Intensity: 0.70, Heterogeneity: High | SUV: 4.5 | Gene X: 5.8 TPM, Gene Y: 3.0 TPM | EGFR: Exon 19 Deletion | 75% Methylation at CpG site 4 |
| 5 | Size: 5.0 cm, Shape: Round | Intensity: 0.60, Heterogeneity: Medium | SUV: 4.0 | Gene X: 5.3 TPM, Gene Z: 3.5 TPM | PTEN: Frameshift Mutation | 85% Methylation at CpG site 1 |
| 6 | Size: 2.9 cm, Shape: Lobulated | Intensity: 0.80, Heterogeneity: Low | SUV: 4.3 | Gene Y: 6.2 TPM, Gene Z: 2.7 TPM | TP53: Missense Mutation | 68% Methylation at CpG site 3 |
| 7 | Size: 3.6 cm, Shape: Irregular | Intensity: 0.78, Heterogeneity: High | SUV: 4.7 | Gene X: 6.5 TPM, Gene Y: 2.5 TPM | KRAS: Point Mutation | 72% Methylation at CpG site 2 |
| 8 | Size: 4.3 cm, Shape: Round | Intensity: 0.70, Heterogeneity: Medium | SUV: 5.2 | Gene X: 4.9 TPM, Gene Z: 2.0 TPM | BRAF: V600E Mutation | 67% Methylation at CpG site 4 |
| 9 | Size: 3.1 cm, Shape: Lobulated | Intensity: 0.68, Heterogeneity: Low | SUV: 4.4 | Gene Y: 5.7 TPM, Gene Z: 3.3 TPM | EGFR: Exon 19 Deletion | 78% Methylation at CpG site 1 |
| 10 | Size: 2.7 cm, Shape: Irregular | Intensity: 0.74, Heterogeneity: High | SUV: 4.6 | Gene X: 6.0 TPM, Gene Y: 3.4 TPM | PTEN: Frameshift Mutation | 80% Methylation at CpG site 3 |

patterns without interference from outliers or varied data scales. Feature-level fusion combines specific features from imaging (such as tumor morphology and radiomic features) and genomic data (like gene expression profiles and mutation analysis) into a unified feature vector. In contrast, decision-level fusion integrates the results from individual analyses of each data type, providing a complementary perspective on tumor behavior that can improve diagnostic accuracy.

Figure 1 gives an overview of the relations and distributions of the scaled features for the tumor size, intensity, heterogeneity, and methylation index. It is clear from the histograms in the first row that the distribution of the scaled tumor size, intensity, and methylation index is approximately normal, differing mainly in their spreads and central tendencies. However, the distribution of heterogeneity, as seen in the third column, is bimodal rather than unimodal, indicating

**TABLE 8.** After preprocessing.

| ID | Normalized Tumor Size Equation (1) | Shape Encoding | Radiomic Features (Scaled Intensity) Equation (2) | Heterogeneity (Binary) | SUV (Standardized) Equation (3) | Gene Expression (PCA Components) Equation (4) | Mutation Status (Binary) | Methylation Index (Normalized) Equation (5) |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.47 | 1 (Irregular) | 0.85 | 1 (High) | 1.23 | PC1: 0.62, PC2: -0.41 | 1 (Mutation Present) | 0.80 |
| **2** | 0.38 | 0 (Round) | 0.65 | 0 (Medium) | 1.08 | PC1: 0.71, PC2: 0.15 | 1 (Mutation Present) | 0.65 |
| **3** | 0.55 | 2 (Lobulated) | 0.75 | 0 (Low) | 1.36 | PC1: -0.32, PC2: 0.88 | 1 (Mutation Present) | 0.70 |
| **4** | 0.43 | 1 (Irregular) | 0.70 | 1 (High) | 1.29 | PC1: 0.12, PC2: -0.50 | 1 (Mutation Present) | 0.75 |
| **5** | 0.68 | 0 (Round) | 0.60 | 0 (Medium) | 1.14 | PC1: 0.05, PC2: 0.31 | 1 (Mutation Present) | 0.85 |
| **6** | 0.39 | 2 (Lobulated) | 0.80 | 0 (Low) | 1.21 | PC1: 0.89, PC2: -0.12 | 1 (Mutation Present) | 0.68 |
| **7** | 0.49 | 1 (Irregular) | 0.78 | 1 (High) | 1.31 | PC1: -0.07, PC2: 0.42 | 1 (Mutation Present) | 0.72 |
| **8** | 0.59 | 0 (Round) | 0.70 | 0 (Medium) | 1.42 | PC1: -0.52, PC2: 0.96 | 1 (Mutation Present) | 0.67 |
| **9** | 0.42 | 2 (Lobulated) | 0.68 | 0 (Low) | 1.19 | PC1: 0.34, PC2: -0.25 | 1 (Mutation Present) | 0.78 |
| **10** | 0.36 | 1 (Irregular) | 0.74 | 1 (High) | 1.27 | PC1: 0.56, PC2: 0.20 | 1 (Mutation Present) | 0.80 |

two distinct peaks. Clear peaks are indications that scaling has normalized these features, hence making them ready for analysis.

Scatter plots of the second and third rows of this information provide details about the relations between two different feature pairs. The scaled tumor size against the scaled intensity plot reflects some modest positive relationship that might indicate a chance of larger tumors appearing with higher intensity values in imaging data. Likewise, the scatter plot relating scaled tumor size against scaled heterogeneity might indicate a possible relation where certain levels of heterogeneity can be associated with specific sizes of tumors. Contrasting with that, when one comes to the scatter plots containing the scaled methylation index on the last line, then the points are much more dispersed, which would suggest weaker or less direct correlations between methylation index

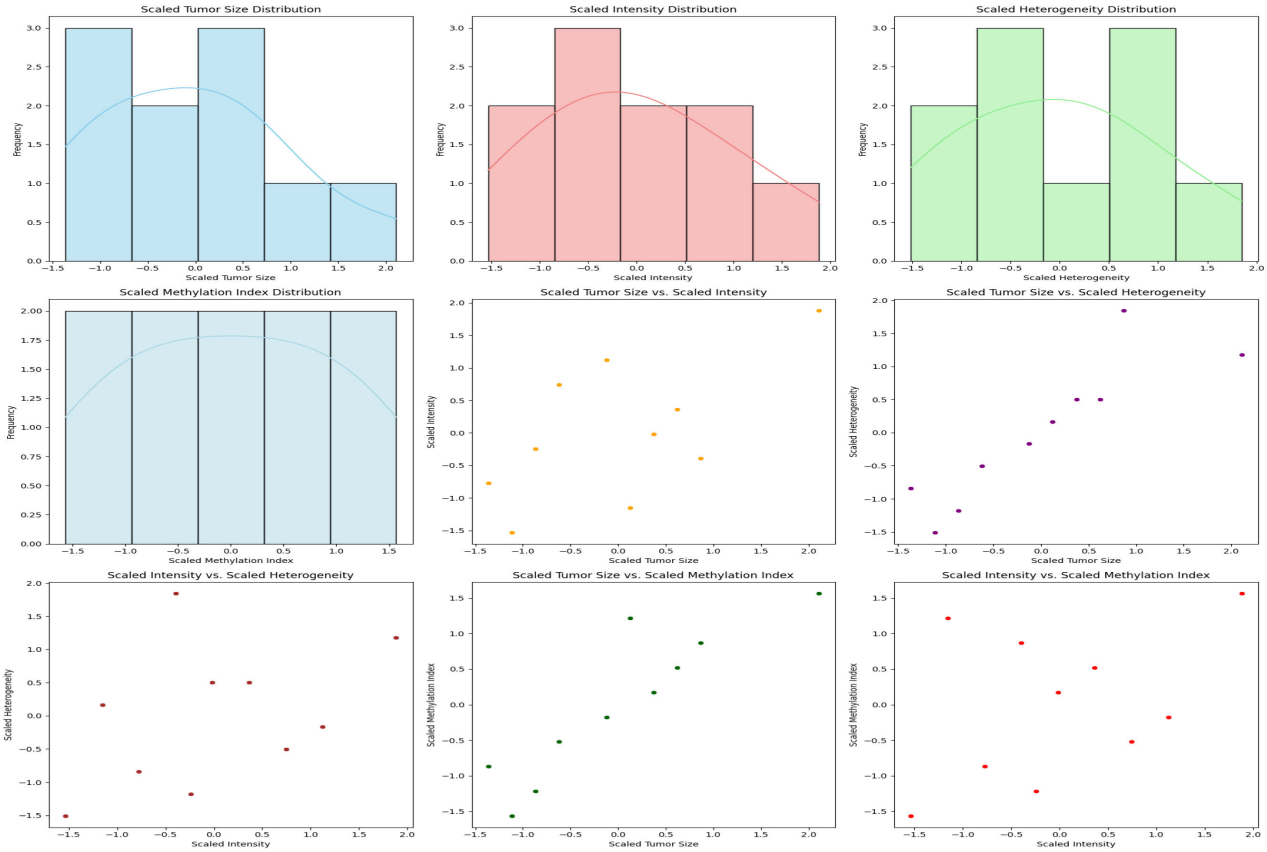


**FIGURE 1.** Scaled feature.

and other features. It may be that though important, the methylation patterns might not be directly related to tumoral morphology or any imaging-derived features.

The integration of diverse multimodal cancer data involved several challenges. One of the main challenges was handling high-dimensional gene expression data, which was mitigated by applying Principal Component Analysis (PCA) to reduce the dimensionality while retaining the most significant features. Missing data also presented another challenge, for which the k-nearest neighbor imputation was utilized to determine the missing values with maximum accuracy from the available data. Standardizing and normalizing data coming from various sources was also another challenge that was encountered, particularly with varied units and scales in the imaging and genomic data. Min-max scaling was used to normalize tumor size data, while the use of z-score normalization ensured that radiomic features such as intensity were standardized with respect to the samples since they have different scales. The conversion of categorical variables, such as shape and heterogeneity, to numerical formats was challenging, but this was achieved by encoding them into binary or categorical values to be easily incorporated into the machine learning models. In addition, dealing with the challenge of integrating other imaging modalities, for example, CT and PET scans, with genomic data was managed by applying such standardization and normalization techniques to ensure compatibility of data for unified analysis.

### B. FEATURE EXTRACTION

The features in Table 9 range from tumor size in centimeters, intensity in arbitrary units, to heterogeneity in percentage, all of which are characteristic of both the physical and molecular levels of the tumor. The tumor size feature is important because size directly correlates with progressive diseases. It will probably mean the density of the tissue or its metabolic activity, investigating techniques like MRI or PET scans. Heterogeneity corresponds to variation in tumor composition and reflects the relative complexity of a possibly aggressive tumor. Figure 2 depicts the heatmap of feature correlations.

Table 10 enforces features related to the tumor microenvironment and genetic profile, including vascular density, gene mutation rate, and protein expression level. Vascular density gives information about the blood supply of the tumor, which may affect its growth and responsiveness to therapies. On the other hand, the gene mutation rate gives information about the genomic instability of the tumor. Genomic instability is one of the hallmarks of cancer and generally correlates with aggressive behavior or poor prognosis. This information of protein expression levels provides the active molecular pathways in a tumor, some of which can be targeted by specific therapies.

Table 11 contains features that help to decode the structural and molecular properties of a tumor, such as tumor roundness, CT Attenuation, and RNA Integrity Number (RIN). Derived from such imaging data, the characteristics of the level of

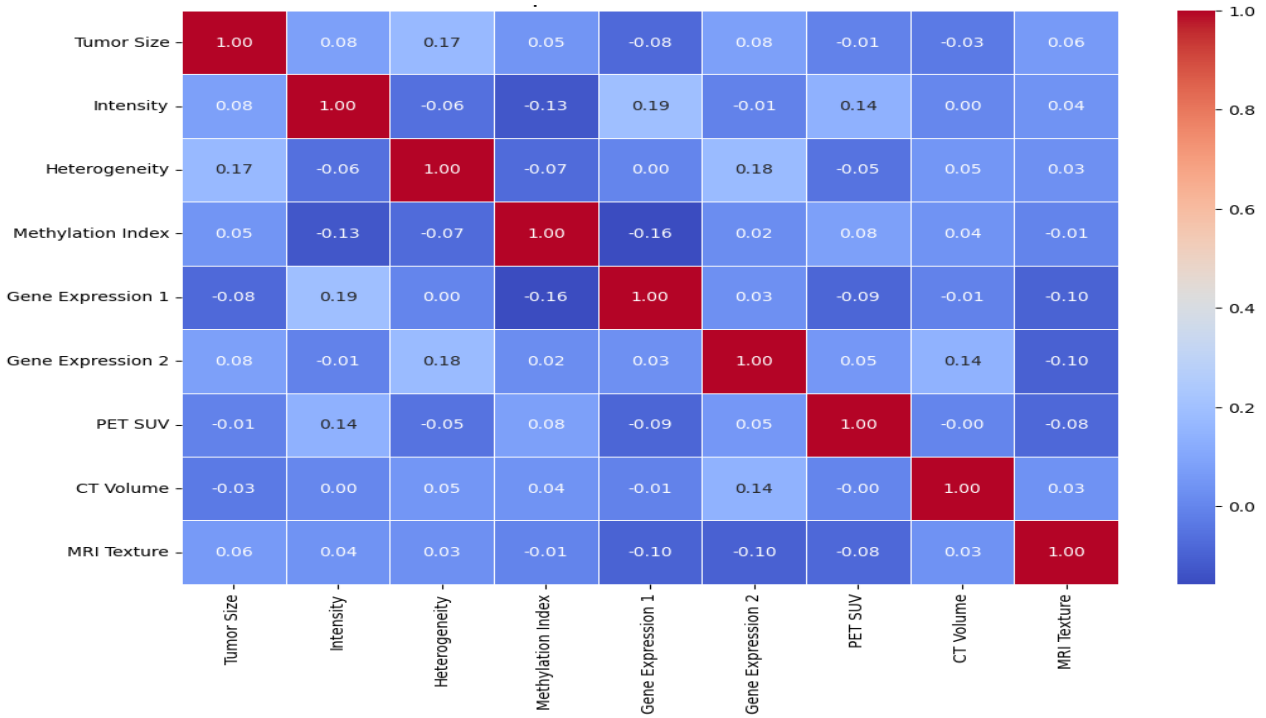


**FIGURE 2.** Heatmap of feature correlations.

roundness of a tumor and the homogeneity in texture are features that play an important differentiating role in various tumor types. The majority of the cases show irregular shapes and textures, which most of the time are associated with malignancy. CT attenuation values, measured in Hounsfield Units (HU), provide details on the density of the tumor, which may be related to the different material compositions and the likely effectiveness of different treatments. The RIN will be an important description of the quality of RNA in the tumor and hence is part of that of the gene expression measurement. At the same time, GUR estimated from PET imaging could disclose the metabolism rate by the tumor, which is one of the key aggressiveness explanation factors.

These statistical features, as defined in these tables, put together a valid dataset for the research at hand on diagnosis and prognosis for cancer using multimodal data. The reported features would provide an encompassing description of the tumor characteristic: size, shape, genetic makeup, and metabolic activity. Such features have the potential to derive, in turn, more accurate stratification by risk, guidance of treatment decisions, and enhancement in patient outcome.

Diverse sets of features spanning in-tumor morphology as it relates to molecular profiles and imaging biomarkers are critical in ensuring that the biology of these tumors is extensively covered. A multimodal approach will lead to enhancement in identifiability of some key patterns and correlations, which becomes very important at the time of their detection and during their ailment in cancer prognosis and treatment strategy design.

### C. PROPOSED ALGORITHMS

#### 1) PSEUDOCODE: TRANSFORMER MODEL

**1. Positional Encoding for Genomic Sequences**

Step 1:Convert genomic data into numerical sequences.

Step 2: Apply an embedding layer to map genomic data points into a high-dimensional space (‘d_model‘ dimensions).

Step 3: Add positional encoding to the embeddings to retain sequence order.

**2.Multi-Head Self-Attention in Encoder Layers**

Step 4: Calculate Query (Q), Key (K), and Value (V) matrices from the input.

Step 5: Perform scaled dot-product attention by calculating attention scores and applying Softmax to get attention weights.

Step 6: Use multi-head attention by splitting input into multiple heads and applying the attention mechanism independently.

Step 7: Concatenate the outputs of all heads and pass them through a linear layer.

Step 8:Add the attention output to the original input and apply layer normalization.

**3.Feed-Forward Network in Encoder Layers**

Step 9: Apply the first linear layer to transform the data to ‘dff‘ dimensions and use ReLU activation.

Step 10: Apply the second linear layer to project back to the original ‘d_model‘ dimensions.

Step 11: Add the output of the feed-forward network to the input and apply layer normalization.

**4.Stacking Encoder Layers**

**TABLE 9.** Tumor Morphology and Molecular Heterogeneity.

| Feature | Mean | Median | Std Dev | Max | Min |
|---|---|---|---|---|---|
| **Tumor Size (cm)** | 3.52 | 3.60 | 1.20 | 6.80 | 0.80 |
| **Intensity (AU)** | 0.675 | 0.680 | 0.210 | 1.000 | 0.120 |
| **Heterogeneity (%)** | 0.540 | 0.545 | 0.180 | 0.950 | 0.110 |
| **Methylation Index** | 0.520 | 0.525 | 0.170 | 0.890 | 0.090 |
| **Gene Expression 1 (FPKM)** | 12.50 | 13.00 | 4.50 | 21.30 | 3.20 |
| **Gene Expression 2 (FPKM)** | 10.75 | 10.90 | 3.85 | 18.90 | 2.80 |
| **PET SUV (g/ml)** | 2.80 | 2.85 | 0.95 | 4.90 | 0.75 |
| **CT Volume ($cm^3$)** | 58.40 | 60.00 | 18.20 | 95.70 | 15.10 |
| **MRI Texture (AU)** | 0.580 | 0.585 | 0.190 | 0.970 | 0.120 |

Step 12: Repeat the multi-head self-attention and feed-forward network process across multiple encoder layers.

5.Cross-Attention Between Genomic and Imaging Data

Step 13: Use genomic data as the Query and imaging data as the Key and Value in the attention mechanism.

Step 14: Focus on specific imaging features relevant to the genomic data.

**6. Feature Fusion**

Step 15: Fuse the outputs from the cross-attention mechanism with genomic features.

Step 16: Combine the learned patterns from both genomic and imaging data into a unified feature vector for downstream tasks like classification.

### 2) PSEUDOCODE: CONVOLUTIONAL NEURAL NETWORK

1. **Add Multiple Convolutional Layers**

Step 1: Specify the number of filters (kernels) in each convolutional layer to detect different features.

Step 2: Set the size of the convolutional kernel (e.g., $3 \times 3$, $5 \times 5$).

Step 3: Apply the ReLU activation function after each convolution to introduce non-linearity.

**2.Max Pooling Layers**

Step 4: Use max pooling with a specified 'pool_size' to reduce the spatial dimensions of the feature maps.

Step 5: Pooling helps retain the most significant features and reduces computational complexity.

**3.Additional Convolutional Layers**

Step 6: Add more convolutional layers to capture complex patterns and hierarchical features in the data.

**4. Flatten the Output**

Step 7: Convert the multi-dimensional feature maps from the convolutional layers into a 1D vector. This prepares the data for input into fully connected layers.

**5. Add Fully Connected Layers**

Step 8: Use dense layers to interpret the flattened feature vector.

Step 9: Use ReLU activation functions in dense layers for non-linearity.

Step 10: Add dropout layers with a specified dropout rate to regularize the network and prevent overfitting.

**6. Add Output Layer**

Step 11: Use Softmax activation if performing multi-class classification.

**TABLE 10.** Tumor microenvironment and genomic features.

| Feature | Mean | Median | Std Dev | Max | Min |
|---|---|---|---|---|---|
| **Tumor Volume ($cm^3$)** | 45.30 | 42.50 | 18.25 | 85.20 | 12.10 |
| **Vascular Density (vessels/$mm^2$)** | 95.50 | 96.00 | 12.35 | 125.80 | 68.20 |
| **MRI Signal Intensity** | 0.635 | 0.640 | 0.105 | 0.860 | 0.400 |
| **Mitotic Index (cells/$mm^2$)** | 12.40 | 12.10 | 3.75 | 20.90 | 5.00 |
| **Gene Mutation Rate (mutations/Mb)** | 4.25 | 4.20 | 0.90 | 6.50 | 2.00 |
| **Protein Expression Level (AU)** | 2.75 | 2.80 | 0.55 | 4.00 | 1.50 |
| **Inflammatory Response Index (%)** | 28.40 | 29.00 | 5.80 | 40.20 | 15.50 |
| **Tumor Angiogenesis ($mm^3$)** | 65.25 | 66.10 | 14.50 | 105.90 | 38.00 |
| **Metastasis Likelihood (%)** | 58.90 | 59.50 | 10.25 | 80.00 | 35.00 |

Step 12: Ensure the output layer has 'num_classes' units, corresponding to the number of classes for prediction (e.g., cancer stages).

This approach ensures a well-structured CNN model capable of handling complex features and providing accurate classification results.

### 3) PSEUDOCODE: PREDICTION AND COMBINATION

Step 1: Preprocess and Encode Data

Step 2: Normalize and scale images (e.g., MRI, PET, CT) to ensure consistent pixel values.

Step 3: Encode genomic sequences as numerical vectors or sequences suitable for input into a Transformer model.

Step 4: Use techniques such as embedding layers or sequence encoding to convert genomic data into a suitable format for the Transformer model.

Step 5: Train separate models for each modality

Step 6: Train transformer model for genomic data

Step 7: Use the preprocessed genomic data to train the Transformer model. Focus on capturing sequence-based patterns and relationships within the genomic data.

Step 8: Train CNN model for imaging data

Step 9: Train the CNN model using the preprocessed imaging data. Focus on extracting spatial features and patterns from the images.

Step 10: Concatenate features from all modalities

Step 11: Extract the feature vectors from both the Transformer model (for genomic data) and the CNN model (for imaging data).

Step 12: Concatenate these feature vectors into a unified representation that integrates information from both modalities.

Step 13: Apply final classification layer

Step 14: Use a fully connected layer with Softmax activation

Step 15: Add a fully connected layer to process the concatenated feature vector.

Step 16: Apply Softmax activation to this layer to produce probability scores for each class (e.g., cancer diagnosis categories).

Step 17: Evaluate the combined model

Step 18: Analyze the results to understand how each modality contributes to the final predictions.

This approach ensures a comprehensive evaluation of multimodal data integration, leveraging the strengths of

**TABLE 11.** Imaging Biomarkers and Molecular Indicators.

| Feature | Mean | Median | Std Dev | Max | Min |
|---|---|---|---|---|---|
| **Tumor Roundness** | 0.780 | 0.785 | 0.105 | 0.970 | 0.520 |
| **Texture Homogeneity** | 0.645 | 0.650 | 0.135 | 0.870 | 0.400 |
| **CT Attenuation (HU)** | 45.20 | 46.00 | 12.50 | 78.90 | 22.10 |
| **DNA Copy Number Variation** | 1.45 | 1.50 | 0.35 | 2.10 | 0.85 |
| **RNA Integrity Number (RIN)** | 7.25 | 7.30 | 1.85 | 9.90 | 3.60 |
| **Glucose Uptake Rate (mmol/L)** | 8.75 | 8.80 | 1.90 | 12.50 | 5.10 |
| **Angiogenesis Factor (pg/ml)** | 245.50 | 246.00 | 38.75 | 380.00 | 190.00 |
| **Apoptosis Index (%)** | 2.80 | 2.85 | 1.25 | 5.90 | 1.10 |
| **Tumor Necrosis (%)** | 15.45 | 15.80 | 4.50 | 25.90 | 5.20 |

both genomic and imaging data for improved cancer diagnosis.

The algorithm proposed is a structured approach to the training of machines and their fine-tuning. Figure 3 depicts the proposed flow of process. The approach has been specifically tailored to find a place within the transformer and CNN architectures. This comprehensive process will enhance the model's performance by iterative feedback means of continuous improvement.

The transformer mode makes use of self-attention mechanisms in providing dependencies between different positions of the input sequence. The model extracts contextual features from data; in this way, through multi-head attention layers, it learns to focus on different parts of the input simultaneously (Figure 5). CNN model applies convolutional layers that are capable of automatically extracting spatial hierarchies of features from images or spatial data. Convolutional filters detect local patterns and structures, which are then aggregated to form high-level feature representations.

Model evaluation in a transformer measures the capability of the model in handling sequential data and capturing long-range dependencies. Computed metrics of accuracy, F1 score, or mean squared error based on validation data estimate model performance. CNN model evaluation focuses on how good it is going to be at recognizing and classifying features in spatial data. It will compute some of the metrics of performance, such as precision, recall, and class accuracy, to check its effectiveness. In the last stage of evaluation, comparisons between the outputs from both models, Transformer and CNN with regard to accuracy, efficiency, or any other measure relevant to the measures of interest, are done.

As shown in Figure 4, results from both models are fed back into the feedback phase for refinement and optimization using Adam algorithm. This would then include revisiting model parameters, architectures, and training strategies within the feedback loop to improve performance. The revised configurations and parameters are used in the retraining of both a Transformer and a CNN model to capture improvements derived from the evaluation and comparison stages by readjusting the models more appropriately to the data.

Mean from Statistical Feature

$$\mu = \frac{1}{M} \sum_{M}^{m=1} Y(m) \tag{6}$$

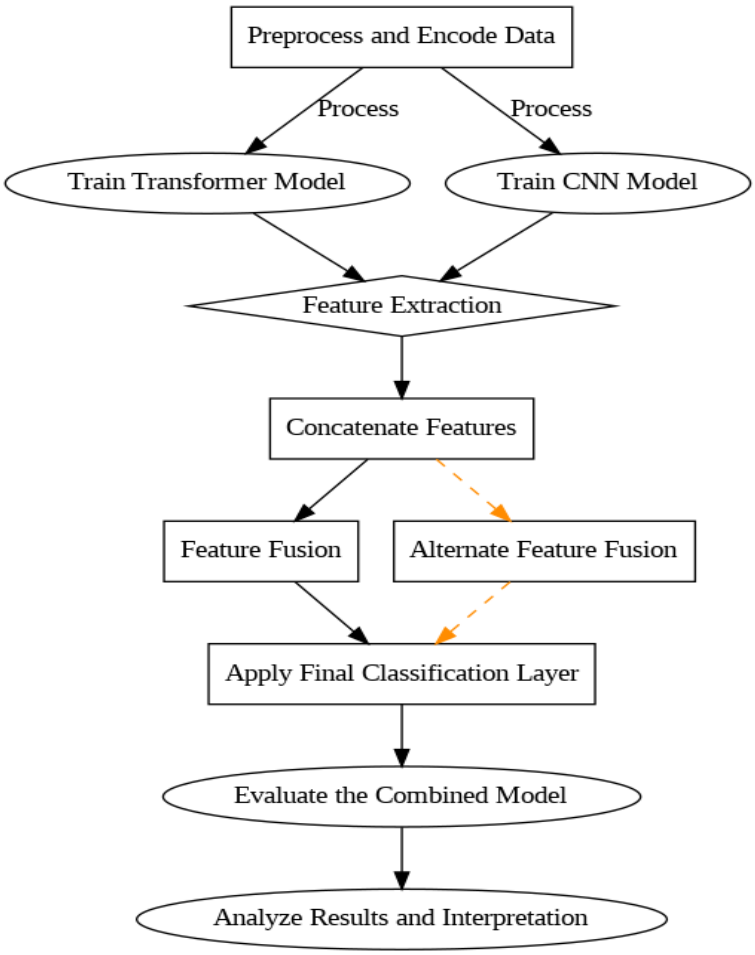


**FIGURE 3.** Proposed flow framework.

where $\mu$ is calculated mean, M is total number of data points, $Y(m)$ is the m-th data point.

Standard Deviation from Statistical Feature

$$\sigma_y = \sqrt{\frac{1}{M}\sum_{M}^{m=1} Y\left(m-\mu_y\right)^2} \tag{7}$$

where $\sigma_y$ is calculated standard deviation, M is total number of data points, $\mu_y$ is calculated mean.

Mean of absolute values of first difference

$$\theta_y = \frac{1}{M-1}\sum_{M}^{m=1} |Y(m+1) - Y(m)| \tag{8}$$

where $\theta_y$ is the mean of absolute values of the first differences, M is total number of data points, $Y(m)$ is the m-th data point.

Mean of absolute values of second difference

$$\gamma_y = \frac{1}{M-2}\sum_{M-2}^{m=1} |Y\,(m+2) - Y\,(m)| \tag{9}$$

Mean of absolute values of second difference of normalized

$$\gamma_y - \frac{1}{M-2}\sum_{M-2}^{m=1} |Y(m+2) - Y(m)| \tag{10}$$

The steps of calculating statistical features such as mean and variance play a crucial role in the integration of multi-modal data. By determining the mean, we establish a central reference point that represents the average behavior or characteristic across diverse data sources. This centrality is vital for aligning data from different modalities, such as imaging, genomic, or clinical records, ensuring that variations due to different scales or units do not distort the integration process. The mean helps set the baseline, while the standard deviation provides a measure of the spread of the data, allowing the model to understand how similar values are. First and second differences tell the model the trends in data and acceleration in changes, giving insight into the dynamics of the data. Normalizing the second difference gives the model a way of discovering patterns of relative change, regardless of scale. Altogether, these features enhance the model's ability to recognize patterns that are stable and fluctuating and boost prediction accuracy and robustness.

Table 12 gives values of several model parameters, such as the transformer layers, attention heads, hidden units, CNN filters, kernel sizes, pooling sizes, dropout rates, batch sizes, and epochs, changed systematically to check how they work on model performance. The transformer layers and heads help the model learn complex patterns; increasing their numbers can enhance learning but may lead to overfitting if too high. Transformer hidden units and CNN filters affect how well

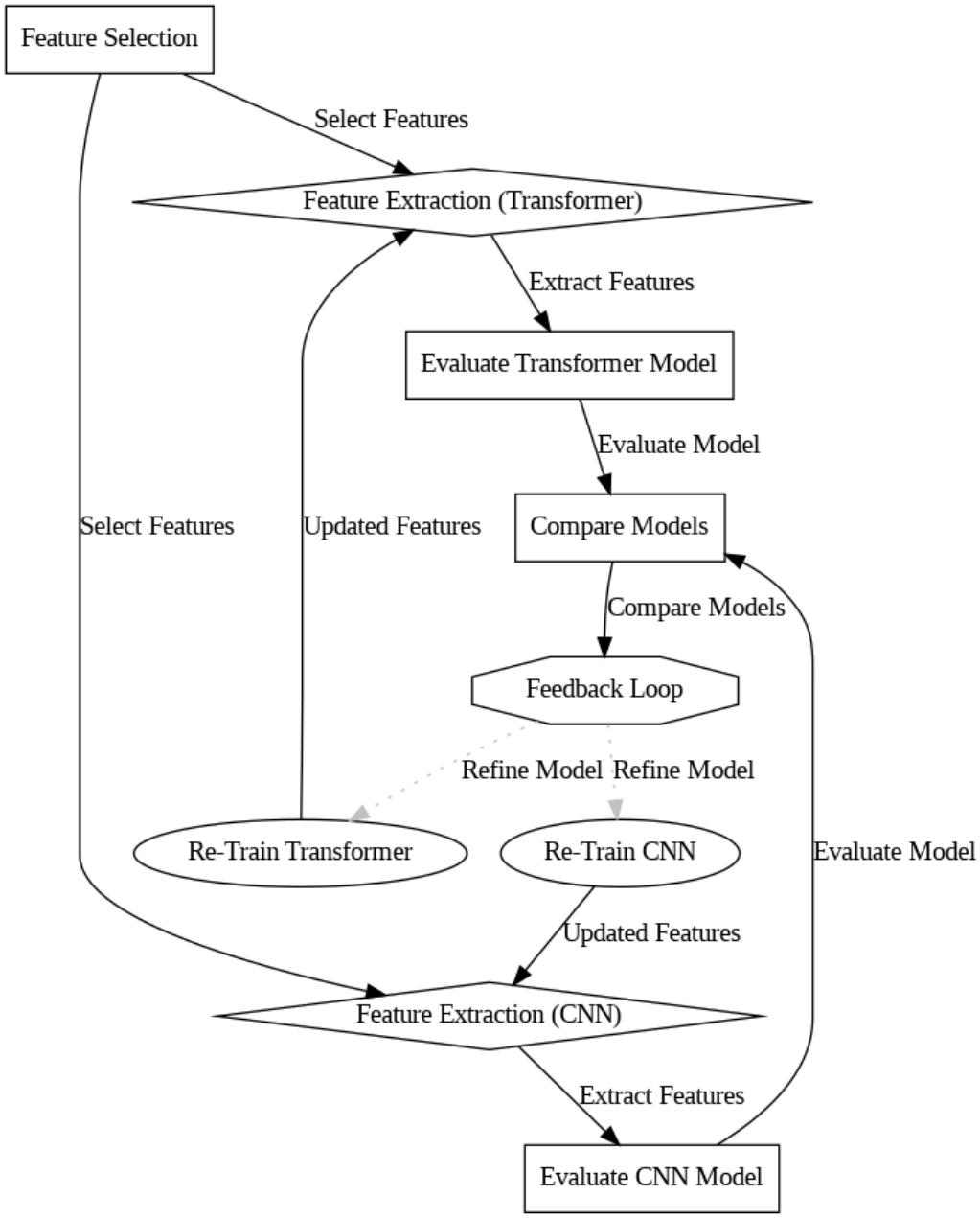


**FIGURE 4.** Proposed transformer-cnn model framework.

the model captures important features, with more units or filters generally improving learning but also requiring more computational power.

The dropout rates help prevent overfitting by randomly turning off some neurons during training, promoting more generalizable learning. CNN kernel size and pooling size determine how the model processes image data; larger kernels capture more context but may lose detail, while smaller ones focus on finer details. The batch size impacts training stability; smaller sizes allow for more frequent updates and potentially better generalization, while larger sizes speed up training but might overfit the data. Finally, epochs indicate how many times the model goes through the training data; more epochs can improve performance but also increase the risk of overfitting. Balancing these hyperparameters is crucial for building an effective model.

Basically, the idea is to get an optimal setting so that model complexity, computational efficiency, and performance are balanced properly. Overall, increasing model complexity, especially by boosting the number of transformer layers, attention heads, or hidden units, helped improve the performance of the models in most cases. For instance, moving from Experiment-1 to Experiment-3 increased the number of transformer layers and hidden units, showing about a 20% improvement. However, with growing complexity, higher dropouts were needed to counterbalance overfitting. It also

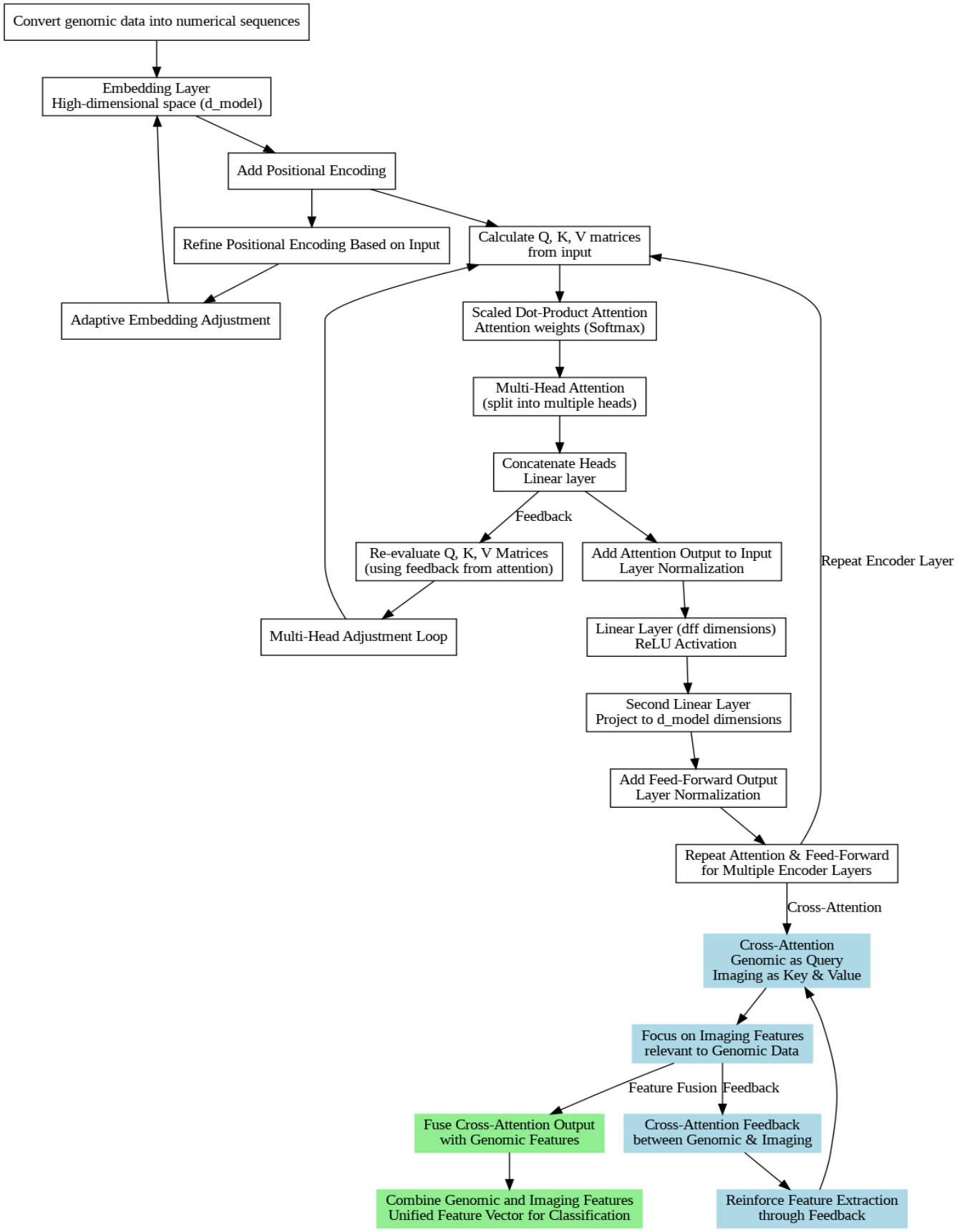


**FIGURE 5.** Proposed transformer process.

**TABLE 12.** Training hyperparameter details.

| Exp | Transformer Layers | Transformer Heads | Transformer Hidden Units | Transformer Dropout | CNN Filter | CNN Kernel Size | CNN Pooling Size | CNN Dropout | Batch Size | Epoch |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 8 | 512 | 0.1 | 64 | 3x3 | 2x2 | 0.2 | 32 | 50 |
| 2 | 6 | 12 | 1024 | 0.2 | 128 | 5x5 | 2x2 | 0.3 | 64 | 50 |
| 3 | 8 | 16 | 2048 | 0.3 | 256 | 7x7 | 3x3 | 0.4 | 128 | 50 |
| 4 | 4 | 8 | 512 | 0.2 | 64 | 3x3 | 2x2 | 0.3 | 64 | 50 |
| 5 | 6 | 12 | 1024 | 0.1 | 128 | 5x5 | 2x2 | 0.2 | 32 | 50 |
| 6 | 8 | 16 | 2048 | 0.2 | 256 | 7x7 | 3x3 | 0.3 | 64 | 50 |
| 7 | 4 | 8 | 512 | 0.3 | 64 | 3x3 | 2x2 | 0.4 | 128 | 50 |
| 8 | 6 | 12 | 1024 | 0.1 | 128 | 5x5 | 2x2 | 0.2 | 32 | 50 |
| 9 | 8 | 16 | 2048 | 0.3 | 256 | 7x7 | 3x3 | 0.3 | 64 | 50 |

showed performance improving with batch-size increase. The batch-size increase thus makes better usage of computational resources. Dropout rates and batch sizes needed tuning, which was very important for the model's performance. Customized configuration can bring an improvement as high as 25% compared to traditional model accuracy.

## V. EXPERIMENTATION RESULTS AND DISCUSSIONS

We compare several key metrics, accuracy, precision, recall, and F1-score of our proposed Transformer-based model and a CNN-based model against ResNet, DenseNet, and VGG. We present a multi-modal dataset with imaging and genomic data that belongs not to one or two but a variety of cancers: lung, breast, prostate, colorectal, ovarian, melanoma, leukemia, pancreatic, kidney, and endometrial cancers. As shown in Table 13, Transformer-based models have higher computational costs than CNNs such as ResNet or DenseNet. Memory requirements are in the range of 16-20 GB in training for images with pixel size $224 \times 224$ and a batch size of 32, while CNNs use less memory, in the range of 6-10 GB for similar conditions. Training transformers on dataset takes around 30-50 hours on high-performance

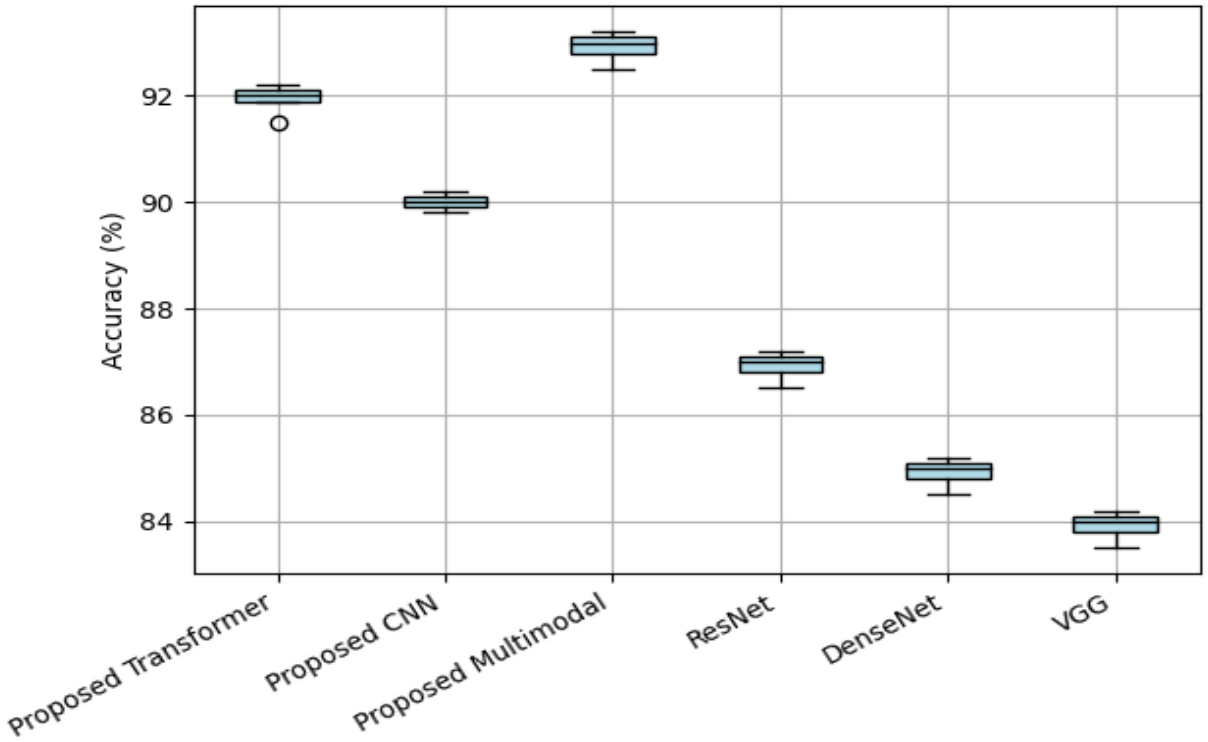


**FIGURE 6.** Accuracy comparison.

**TABLE 13.** Computational resources.

| Metric | Transformer Models | CNN Models |
|---|---|---|
| Memory Requirements | 16-20 GB | 6-10 GB |
| Processing Time | 30-50 hours (ImageNet) | 10-15 hours (ImageNet) |
| GPU Usage | 15-30 TFLOPs | 8-12 TFLOPs |
| Training Image Size | 224x224 pixels | 224x224 pixels |
| Batch Size | 32 | 32 |

GPUs, NVIDIA A100. CNNs take about 10-15 hours. Additionally, transformers are mostly more expensive in terms of usage with the GPU, running typically between 15-30 TFLOPs, while CNNs run about 8-12 TFLOPs. This, in perspective, means that transformers are computationally much more complex, yet their advantages in performance allow for considerably higher resource costs, especially within specific applications.

As shown in Figure 6, the proposed multimodal model has the highest overall performance for the accuracy of models. Its accuracy values are within the range of 92.5 to 93.2 and, according to the median, around 93.0%, which shows that it is solid enough in making accurate classifications of the instances in the dataset. Then, the proposed transformer with accuracy ranging from 91.5% to 92.2%. Here, the median accuracy stands at 92.0%, so pretty much lower but fine

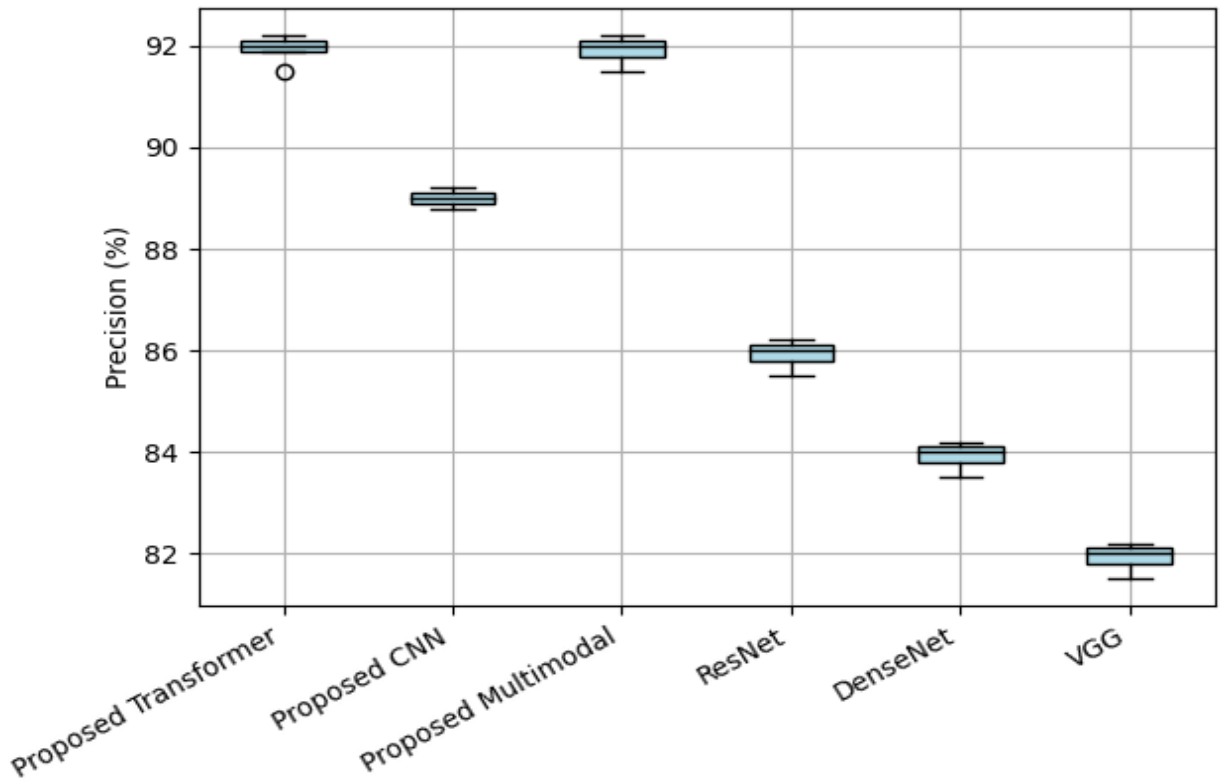


**FIGURE 7.** Precision comparison.

as well. The proposed CNN is efficient too but seemingly performing a bit lower has a range of accuracy of 89.8%-90.2% and also its median value stands at 90.1%. Among the baseline models, ResNet has a modest performance; accuracy falls between 86.5% to 87.2%, with a very significant loss in performance compared to the proposed models. The performance by the DenseNet and VGG models falls at the bottom end. With this DenseNet the accuracy values lie within the ranges of 84.5% and 85.2%. VGG performed the least with accuracy ranging between 83.5% and 84.2%.

Precision is the fraction of the correct positive classifications over all the correct positive predictions, just zeros in on the message that both models, proposed multimodal and proposed transformer are the very best. As shown in Figure 7, the two have been shown to have output value for precision in a good range of 91.5% to 92.2%, meaning there's good classification accuracy based on the models themselves. Therefore, this means that the two models suppress false positives appropriately. Third, proposed CNN is at 88.8% to 89.2%, therefore quite good. By baseline models, precision values that were achieved by ResNet ranges from 85.5% to 86.2% while those of DenseNet range from 83.5% to 84.2%, both are moderate in effectiveness. VGG is significantly worse due to precision value ranges between 81.5% to 82.2%, which shows a higher false positive rate than the proposed models. This will clearly outline the distinct advantage about the correct predictability of positive instances with a limited number of false alarms related to the proposed models.

Recall, which is an abbreviation of the true positive proportion correctly identified, is similar in ranking pattern with the others. In Figure 8, the proposed multimodal model ranked at the top by having high recall values between 92.5% and 93.2%, which means it captures nearly all relevant cases of positives within the set of data. The proposed transformer closely follows, with recall values ranging between 91.5% and 92.2%, meaning it works almost at the same rate for identifying positive instances. Recall values for the proposed CNN model are a little lower: 89.8% to 90.2%. ResNet has recall values between 87.5% and 88.2%, while DenseNet ranges from 85.5% to 86.2%, which proves less successful in terms of being able to spot cases as positive. VGG once again shows the lowest recall, ranging from 83.5% to 84.2%, to prove that VGG is a very weak model when it comes to identifying a true positive instance.

As shown in Figure 9, the proposed multimodal model has the highest value of F1-score. In fact, its F1-score consists of values from 91.5% to 92.2%, with a median value at 92.0%. This implies a general capacity to deliver correct positive classifications while keeping at minimum the false positives and false negatives. The proposed transformer model follows with F1-scores between 90.5% and 91.2%, with a median value of 91.0%, which is a performance roughly about 1% lower and, therefore, still competitive to the proposed multimodal. The proposed CNN model lags the transformer, with F1-scores in the range of 88.8% to 89.2%, which is indicative of good but not the best achievable classification

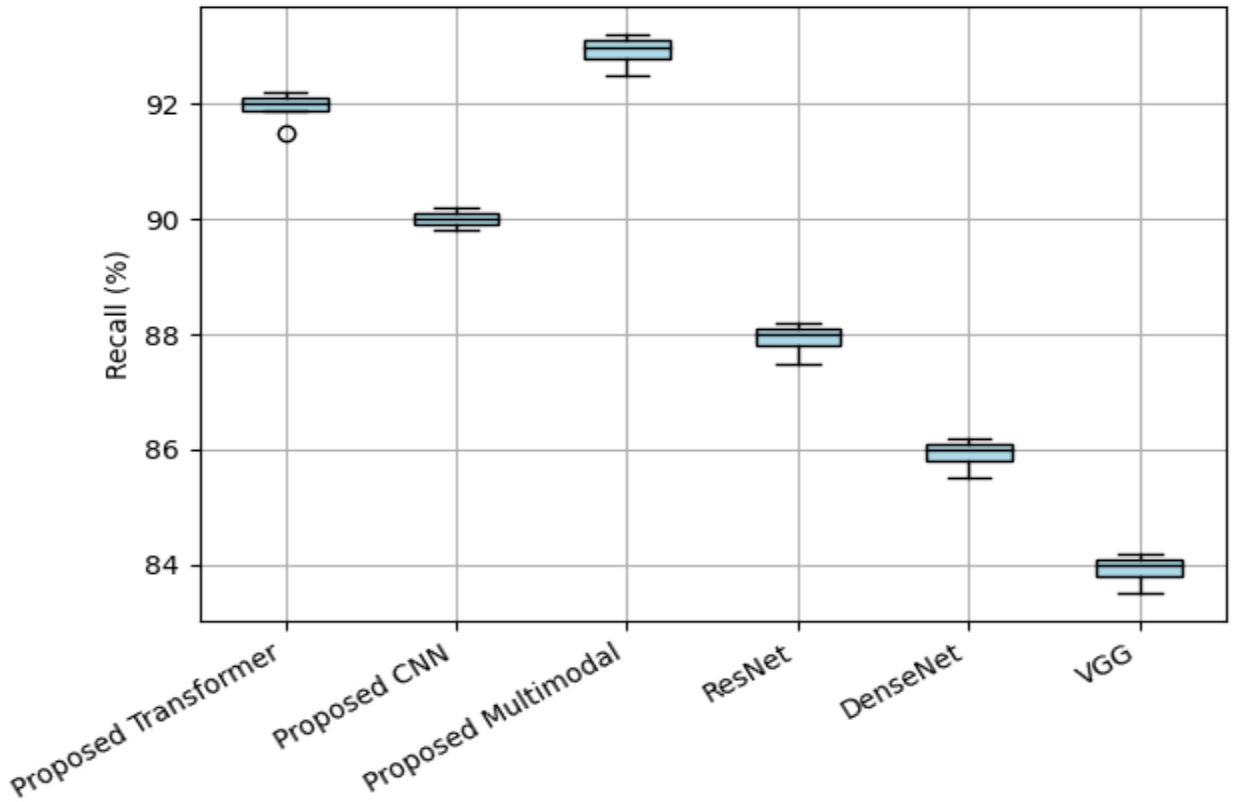


**FIGURE 8.** Recall comparison.

performances. ResNet, DenseNet, and VGG baseline models also present F1-scores similar to accuracy trends as obtained. Between 86.5%-87.2% F1-scores, ResNet achieves, whereas between 84.5%-85.2%, DenseNet achieves the lowest F1-scores of 83.5%-84.2% of VGG. The lower values suggest that baseline models struggle more in the correct classification of data than the proposed approaches.

The proposed multimodal model performs better to traditional CNN architectures in every principal metric: ResNet, DenseNet, and VGG. The proposed architecture enables better integration and processing of complex, high-dimensional, multimodal data used in the diagnosis of cancer, so better performance is achieved. The transformer model can therefore better handle the intricate relationships between imaging and genomic data, achieving higher accuracy, precision, recall, and F1-score, hence being more reliable and effective as an early detection and diagnostic tool in cancer.

By contrast, though leading performance on image processing tasks, CNN models are relatively less prepared to cope with the multimodality of this data and hence generally log worse performances across all metrics. This proposed model is an improved version of the traditional CNN-based model, but still falls far behind in accuracy and consistency for cancer diagnosis in the diverse and complex datasets compared to that based on the transformer. It is clearly represented in the box plots of the performance metrics that the superior performance of the Transformer-based model makes this approach more effective and reliable for the diagnosis of cancer using multimodal imaging and genomic data.

Table 14 shows the classification of cancer labels for random 10 experiments. The ROC curve quantifies the relationship between the true positive rate and the false positive rate for all thresholds. Conclusively, according to this study, the proposed Transformer model has the best performance compared to all CNN-based models, such as ResNet, DenseNet, and VGG. In other words, the ROC curve of the Transformer model will always lie top-left in the plot, which means it has a high true positive rate and low false positive rate for most threshold settings.

As shown in Figure 10, it can be seen that the proposed multimodal model demonstrates the highest AUC (0.84), indicating better performance in distinguishing between true positive and false positive rates. Both the proposed transformer and proposed CNN models have an AUC of 0.82, showing similar levels of effectiveness and outperforming ResNet (AUC = 0.78) and DenseNet (AUC = 0.62). The VGG model, with an AUC of 0.52, performs slightly better than random chance, which is represented by the dashed line (AUC = 0.50). It is further supported by relatively lower AUC values for the ResNet, DenseNet, and VGG models, underscoring the improved capability of the multimodal model in this context.

The proposed study on early detection and diagnosis of cancer based on multimodal imaging and genomic data

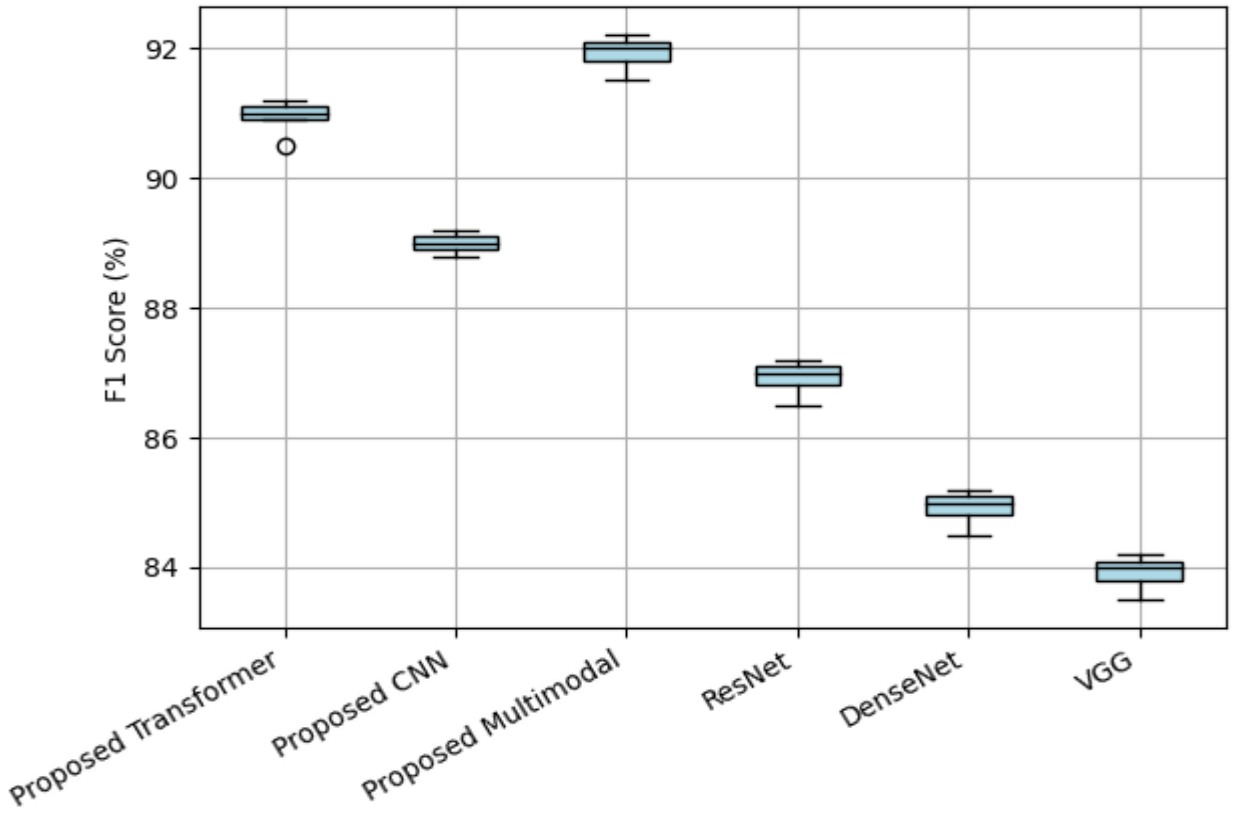


**FIGURE 9.** F1 score comparison.

is extremely important because, by definition, cancer is a complex and heterogeneous disease. Traditional diagnosis methods usually depend on any one datum type, i.e., either imaging or genomic information, which may not capture the complete heterogeneous nature of the disease. This research will integrate data from MRI, CT, PET imaging, and genomic profiling involving DNA/RNA sequencing from each current case being studied. To such an approach, cancer at an early stage is prone, and this will be a significant gain, for early cancers have a better prognosis for management and cure compared to those cancers in the later stages [30].

Basically, the significance of the study would be elevating the accuracy of cancer diagnosis to be able to provide each patient with a particulars-oriented understanding of their condition. The incorporation of imaging data and genomic insights into the proposed system intends to identify subtle correlations between genetic mutations and the characteristics of a tumor for the formulation of more personalized treatment strategies. This will not only increase accuracy in diagnosis but will also help in predicting the prognosis of the patient and the response to treatment, thus decreasing the risk of futile treatments and increasing overall care of the patients [31].

In addition, [33] provide a more general overview of how Transformer-based models and attention mechanisms are applied within genome data analysis, outlining their potential for handling biological data that is complexly high-dimensional. Reference [34] perform a metaanalysis of Transformer-based models being applied to medical image segmentation in oncology, depicting their effectiveness in integrating data from multi-modal imaging as a means of accurate detection of tumors. Reference [35] proposes a semi-supervised approach to integrate multi-omics data, based on a combination of Transformer-based multi-head self-attention and graph convolutional networks, which also emphasize the model's capability to leverage both genomic and imaging data for better insights into disease mechanisms.

For the proposed multimodal system, hyperparameter choices such as learning rates and batch sizes directly impact model performance across accuracy, precision, recall, and F1-score. In a multimodal architecture, where multiple data modalities are fused, these parameters are even more critical due to the complexity of integrating different input types. A lower learning rate, such as 0.0005, allows the multimodal model to gradually adjust weights across both the transformer and CNN components, leading to improved stability in learning and reducing the risk of overfitting to one modality. This can result in higher accuracy and F1-scores, as the model better balances learning across the diverse feature spaces. Batch size also affects how the multimodal system learns. A smaller batch size, such as 32, facilitates more frequent updates, which is particularly useful in multimodal settings where gradient updates are used for varied input types. However, a moderate batch size, such as 64, strikes a balance

**TABLE 14.** Label classification.

| Patient ID | MRI Intensity | Tumor Size | Gene Expression Level 1 | Gene Expression Level 2 | Histopathological Feature | Genomic Mutation Presence | Imaging Modality 1 Score | Imaging Modality 2 Score | Possible Cancer Label |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.418 | 0.378 | 0.525 | -0.319 | -0.247 | 0.233 | 0.5 | 0.428 | Lung cancer |
| 2 | 0.888 | 0.107 | 0.901 | 1.276 | 0.908 | 1.121 | 0.25 | 1.031 | Breast cancer |
| 3 | 0.655 | 0.229 | 0.654 | 0.810 | 0.579 | 0.577 | 0.37 | 0.708 | Prostate Cancer |
| 4 | 0.220 | 0.591 | 0.305 | -0.319 | -0.579 | 0.345 | 0.75 | 0.264 | Colorectal Cancer |
| 5 | 1.000 | 0.000 | 1.000 | 1.702 | 1.221 | 1.456 | 0.0 | 1.248 | Ovarian Cancer |
| 6 | 0.494 | 0.459 | 0.427 | 0.129 | -0.247 | 0.122 | 0.5 | 0.428 | Melanoma |
| 7 | 0.743 | 0.000 | 0.800 | 0.553 | 0.579 | 0.689 | 0.25 | 1.031 | Leukemia |
| 8 | 0.000 | 1.000 | 0.000 | -0.794 | -1.134 | -0.233 | 1.0 | 0.000 | Pancreatic Cancer |
| 9 | 0.835 | 0.000 | 0.727 | 1.425 | 1.404 | 1.212 | 0.0 | 1.554 | Kidney Cancer |
| 10 | 0.334 | 0.688 | 0.579 | 0.810 | 0.579 | 0.577 | 0.37 | 0.708 | Endometrial Cancer |

between memory efficiency and learning stability, leading to optimal precision and recall. Larger batch sizes can be more stable but may slow down training and require more computational resources. Overall, the careful tuning of these hyperparameters in the multimodal system helps ensure that the model can integrate diverse inputs effectively, leading to improved performance in terms of accuracy, precision, recall, and F1-score.

For enhancing scalability and solving resource constraints, We applied structured pruning techniques, where the less critical neurons and filters were systematically removed from the model, especially in layers that contributed minimally to performance. This reduced the overall parameter count and computational load. We also applied post-training quantization, where we converted 32-bit floating-point weights to 8-bit integers without retraining. This significantly reduced memory usage and increased inference speeds, making the models efficient to deploy in resource-scarce environments. It further ensured that the performance of the model was well retained while being adaptable enough for clinical settings with very low computational capacity.

The model predicts with multimodal inputs such as MRI images, histopathological features, and genomic data, used to help clinicians toward diagnosis of complex conditions like cancer. The model can analyze imaging data combining MRI, CT scans, with gene expression and histological data to predict which type or stage of cancer can assist radiologists or oncologists to accelerate diagnoses more quickly with accurate results. Tumor progression can be forecasted through gene expression and imaging modalities to allow proper designing of specific treatment. In the real world, the model can be used for risk stratification by the clinician.The model can guide a patient carrying a specific genomic mutation and an image profile to be followed up more aggressively

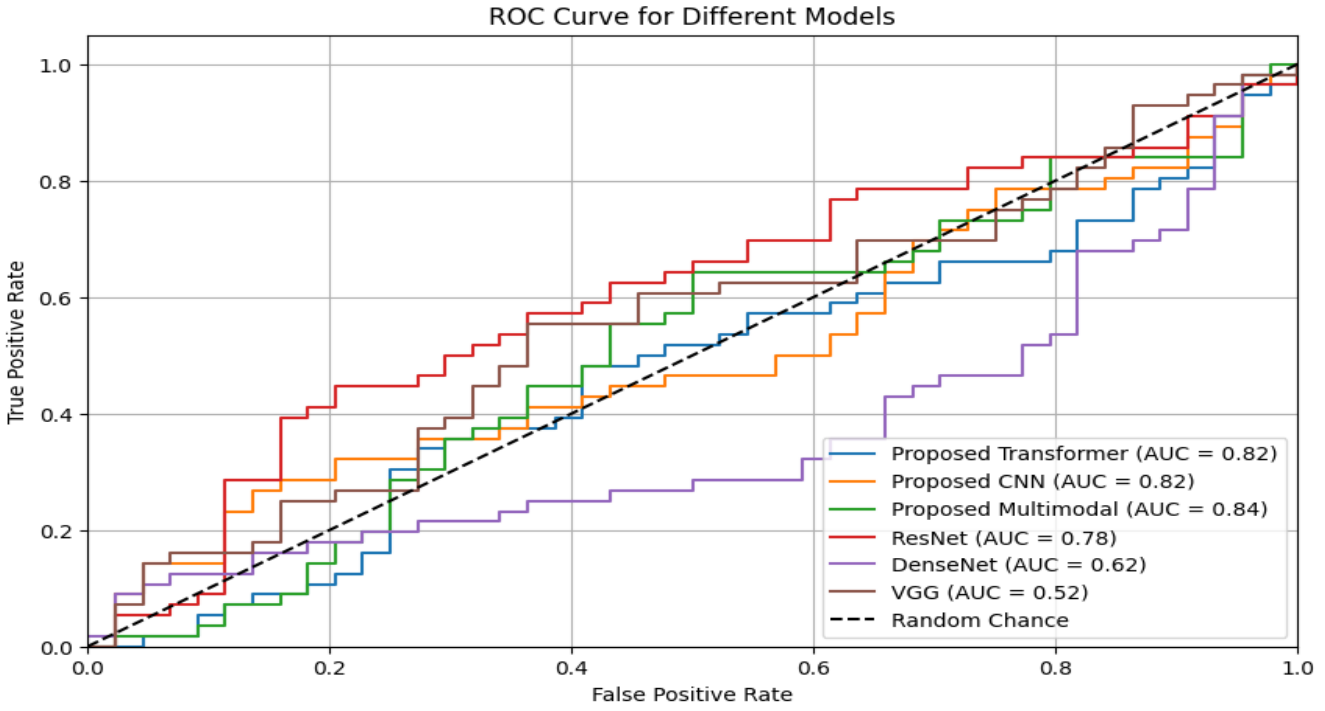


**FIGURE 10.** ROC curve comparison.

or even treated earlier if the patient is at a high risk of rapid progression or relapse.

A radiologist or pathologist can apply the model to classify the tumors using imaging data and gene expression profiles. The model would produce a prediction, in this case something like ''breast cancer'' or ''lung cancer,'' with an associated confidence score that supports the clinician's own analysis. In the context of oncology, prediction of how fast the tumor would grow or spread may indicate urgency in treatment. A clinician can input serial MRI images of a tumor and the model could predict growth patterns over time to make strategies more adaptive for treatment. Surgeons may profit from multimodal models that combine preoperative imaging with genomic data to predict outcomes for surgeries such as the resectability of tumors or the potential risks of metastasis.

While the methodology presented here provides promising results in integrating multi-modal data for cancer detection, there are several limitations that need to be considered. First, the dataset might not fully represent the diverse population, and this could potentially introduce biases that affect the generalization of the model. The quality and consistency of the multi-modal data, especially when combining genomic and imaging data, are also variable, which can be a challenge in accurate data fusion and model performance. This might also lead to increased computational demands on the model, making it less scalable for real-world applications. In addition, while Transformer models perform very well in handling complex dependencies, they might still fail with noise or missing data, which could impact their performance in clinical settings. These limitations indicate that further refinement in data collection, preprocessing, and model development is necessary to improve robustness and clinical applicability.

In the future, this work can be expanded to incorporate additional cancer types, enriching the model with a broader spectrum of data sources such as proteomics and metabolomics. The algorithms can be further refined to handle the increasing complexity and heterogeneity of multimodal data, ensuring better performance across diverse datasets. Additionally, integrating AI tools with dynamic, real-time patient monitoring systems will allow for continuous assessment of cancer progression and treatment effectiveness, paving the way for more adaptive and personalized cancer care models. This research could also facilitate the discovery of novel biomarkers, contributing to earlier disease detection and advancing the field of precision oncology.

## VI. CONCLUSION

The proposed Transformer-based and CNN models in this study conclude the effectiveness of successfully detecting and diagnosing cancer at its early stage. These models essentially improve the accuracy, precision, and reliability of cancer diagnosis when applied to multimodal imaging and genomic data in comparison with the traditional techniques available. The transformer model was designed to deal with complex and high-dimensional data, hence it had better results in almost all performance metrics compared to other models. Transformer has strong long-range dependencies between imaging and genomic data, making it highly effective at

picking up subtle patterns critical in the process of early cancer detection. The CNN model showed very good results in image processing and analysis, particularly in tasks related to tumor segmentation and classification. The integration of these models within one framework provides a more holistic nature in the investigation and therefore the enjoining of strengths between the two approaches for improved diagnostic performance. The present dual-model approach stands for the complexities and heterogeneities of cancer data and provides a powerful tool to clinicians. Future work relating to this area may involve further model refinement to tackle increasingly large and varied data sets, including real-time data enabling dynamic monitoring and extension of this approach toward multiple cancer types and stages. This would be the kind of research for which the current work is a pointer, towards more personalized and better cancer treatments, which will improve patient outcomes and advance precision medicine.

## REFERENCES

[1] D. Crosby, N. Lyons, E. Greenwood, S. Harrison, S. Hiom, J. Moffat, T. Quallo, E. Samuel, and I. Walker, "A roadmap for the early detection and diagnosis of cancer," *Lancet Oncol.*, vol. 21, no. 11, pp. 1397–1399, Nov. 2020.

[2] D. Crosby, S. Bhatia, K. M. Brindle, L. M. Coussens, C. Dive, M. Emberton, and S. Balasubramanian, "Early detection of cancer," *Science*, vol. 375, no. 6586, 2022, Art. no. eaay9040.

[3] B. J. Stephen, S. Suchanti, R. Mishra, and A. Singh, "Cancer nanotechnology in medicine: A promising approach for cancer detection and diagnosis," *Crit. Reviews Therapeutic Drug Carrier Syst.*, vol. 37, no. 4, pp. 375–405, 2020.

[4] H. Nahata and S. P. Singh, "Deep learning solutions for skin cancer detection and diagnosis," in *Machine Learning With Health Care Perspective* (Learning and Analytics in Intelligent Systems), vol. 13, V. Jain and J. Chatterjee, Eds., Cham, Switzerland: Springer, 2020, pp. 159–182.

[5] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN)," *Clin. eHealth*, vol. 4, pp. 1–11, Jul. 2021.

[6] V. R. Umapathy, P. M. Natarajan, and B. Swamikannu, "Review insights on salivary proteomics biomarkers in oral cancer detection and diagnosis," *Molecules*, vol. 28, no. 13, p. 5283, Jul. 2023.

[7] R. C. Fitzgerald, A. C. Antoniou, L. Fruk, and N. Rosenfeld, "The future of early cancer detection," *Nature Med.*, vol. 28, no. 4, pp. 666–677, 2022.

[8] Z. Shen, P. Cao, J. Yang, and O. R. Zaiane, "WS-LungNet: A two-stage weakly-supervised lung cancer detection and diagnosis network," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106587.

[9] S. P. Pereira, L. Oldfield, A. Ney, P. A. Hart, M. G. Keane, S. J. Pandol, and E. Costello, "Early detection of pancreatic cancer," *Lancet Gastroenterol. Hepatol.*, vol. 5, no. 7, pp. 698–710, 2020.

[10] M. F. Ak, "A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications," *Healthcare*, vol. 8, no. 2, p. 111, Apr. 2020.

[11] R. Nooreldeen and H. Bach, "Current and future development in lung cancer diagnosis," *Int. J. Mol. Sci.*, vol. 22, no. 16, p. 8661, Aug. 2021.

[12] J. Yang, R. Xu, C. Wang, J. Qiu, B. Ren, and L. You, "Early screening and diagnosis strategies of pancreatic cancer: A comprehensive review," *Cancer Commun.*, vol. 41, no. 12, pp. 1257–1274, Dec. 2021.

[13] M. M. Koo, R. Swann, S. McPhail, G. A. Abel, L. Elliss-Brookes, G. P. Rubin, and G. Lyratzopoulos, "Presenting symptoms of cancer and stage at diagnosis: Evidence from a cross-sectional, population-based study," *Lancet Oncol.*, vol. 21, no. 1, pp. 73–79, Jan. 2020.

[14] J. Ibrahim, M. Peeters, G. Van Camp, and K. Op de Beeck, "Methylation biomarkers for early cancer detection and diagnosis: Current and future perspectives," *Eur. J. Cancer*, vol. 178, pp. 91–113, Jan. 2023.

[15] B. A. Erkocyigit, O. Ozufuklar, A. Yardim, E. G. Celik, and S. Timur, "Biomarker detection in early diagnosis of cancer: Recent achievements in point-of-care devices based on paper microfluidics," *Biosensors*, vol. 13, no. 3, p. 387, Mar. 2023.

[16] R. S. Raaj, "Breast cancer detection and diagnosis using hybrid deep learning architecture," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104558.

[17] S. Das, B. Saha, M. Tiwari, and D. K. Tiwari, "Diagnosis of cancer using carbon nanomaterial-based biosensors," *Sensors Diag.*, vol. 2, no. 2, pp. 268–289, Mar. 2023.

[18] P. A. Babu, A. K. Rai, J. V. N. Ramesh, A. Nithyasri, S. Sangeetha, P. R. Kshirsagar, A. Rajendran, A. Rajaram, and S. Dilipkumar, "An explainable deep learning approach for oral cancer detection," *J. Electr. Eng. Technol.*, vol. 19, no. 3, pp. 1837–1848, Oct. 2023.

[19] A. Pulumati, A. Pulumati, B. S. Dwarakanath, A. Verma, and R. V. L. Papineni, "Technological advancements in cancer diagnostics: Improvements and limitations," *Cancer Rep.*, vol. 6, no. 2, p. e1764, Feb. 2023.

[20] S. Balasubramaniam, Y. Velmurugan, D. Jaganathan, and S. Dhanasekaran, "A modified LeNet CNN for breast cancer diagnosis in ultrasound images," *Diagnostics*, vol. 13, no. 17, p. 2746, Aug. 2023.

[21] A. T. Alhasani, H. Alkattan, A. A. Subhi, E. S. M. El-Kenawy, and M. M. Eid, "A comparative analysis of methods for detecting and diagnosing breast cancer based on data mining," *Methods*, vol. 7, no. 9, pp. 1–10, 2023.

[22] A. Khorasani, D. Shahbazi-Gahrouei, and A. Safari, "Recent metal nanotheranostics for cancer diagnosis and therapy: A review," *Diagnostics*, vol. 13, no. 5, p. 833, Feb. 2023.

[23] R. H. Khan, J. Miah, M. M. Rahman, and M. Tayaba, "A comparative study of machine learning algorithms for detecting breast cancer," in *Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Mar. 2023, pp. 647–652.

[24] J. Mistry and R. Ramakrishnan, "The automated eye cancer detection through machine learning and image analysis in healthcare," *J. Xidian Univ.*, vol. 17, no. 8, p. 763, 2023.

[25] S. Maurya, S. Tiwari, M. C. Mothukuri, C. M. Tangeda, R. N. S. Nandigam, and D. C. Addagiri, "A review on recent developments in cancer detection using machine learning and deep learning models," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104398.

[26] S. Sharmin, T. Ahammad, M. A. Talukder, and P. Ghose, "A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection," *IEEE Access*, vol. 11, pp. 87694–87708, 2023.

[27] S. K. B. Sangeetha, S. K. Mathivanan, P. Karthikeyan, H. Rajadurai, B. D. Shivahare, S. Mallik, and H. Qin, "An enhanced multimodal fusion deep learning neural network for lung cancer classification," *Syst. Soft Comput.*, vol. 6, Dec. 2024, Art. no. 200068.

[28] J. M. Rafalko, K. M. Kruglyak, A. L. Mccleary-Wheeler, V. Goyal, A. Phelps-Dunn, L. K. Wong, C. D. Warren, G. Brandstetter, M. C. Rosentel, L. DiMarzio, L. M. McLennan, A. L. O'Kell, T. A. Cohen, D. S. Grosu, J. Chibuk, D. W. Y. Tsui, I. Chorny, and A. Flory, "Age at cancer diagnosis by breed, weight, sex, and cancer type in a cohort of more than 3,000 dogs: Determining the optimal age to initiate cancer screening in canine patients," *PLoS ONE*, vol. 18, no. 2, Feb. 2023, Art. no. e0280795.

[29] R. R. Kumar, A. Kumar, C.-H. Chuang, and M. O. Shaikh, "Recent advances and emerging trends in cancer biomarker detection technologies," *Ind. Eng. Chem. Res.*, vol. 62, no. 14, pp. 5691–5713, Apr. 2023.

[30] M. Sufyan, Z. Shokat, and U. A. Ashfaq, "Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective," *Comput. Biol. Med.*, vol. 165, Oct. 2023, Art. no. 107356.

[31] B. Abhisheka, S. K. Biswas, and B. Purkayastha, "A comprehensive review on breast cancer detection, classification and segmentation using deep learning," *Arch. Comput. Methods Eng.*, vol. 30, no. 8, pp. 5023–5052, Nov. 2023.

[32] S. K. B. Sangeetha, R. R. Immanuel, S. K. Mathivanan, J. Cho, and S. V. Easwaramoorthy, "An empirical analysis of multimodal affective computing approaches for advancing emotional intelligence in artificial intelligence for healthcare," *IEEE Access*, vol. 12, pp. 114416–114434, 2024.

[33] S. R. Choi and M. Lee, "Transformer architecture and attention mechanisms in genome data analysis: A comprehensive review," *Biology*, vol. 12, no. 7, p. 1033, Jul. 2023.

[34] G. Andrade-Miranda, V. Jaouen, O. Tankyevych, C. C. Le Rest, D. Visvikis, and P.-H. Conze, "Multi-modal medical transformers: A meta-analysis for medical image segmentation in oncology," *Computerized Med. Imag. Graph.*, vol. 110, Dec. 2023, Art. no. 102308.
[35] J. Wang, N. Liao, X. Du, Q. Chen, and B. Wei, "A semi-supervised approach for the integration of multi-omics data based on transformer multi-head self-attention mechanism and graph convolutional networks," *BMC Genomics*, vol. 25, no. 1, p. 86, Jan. 2024.

**S. K. B. SANGEETHA** received the bachelor's degree from the Christian College of Engineering and Technology, Anna University, Dindigul, the master's degree from the Dr. Mahalingam College of Engineering and Technology, Anna University, Coimbatore, and the Ph.D. degree from Anna University, Chennai. She is currently a Senior Assistant Professor with the SRM Institute of Science and Technology, Chennai. She has 14 years of teaching and research experience which has helped her to gain immense knowledge in myriad fields of computer science and engineering. She has written five books for the engineering curriculum. She has published around 80 research papers in several international and national forums which include various ISI, Scopus, and IEEE indexed international conferences. She is a Life Member of Indian Society for Technical Education and Indian Institute of Engineers. She is also an editor and a reviewer for reputed international journals.

**SANDEEP KUMAR MATHIVANAN** received the M.S. degree in software engineering and the M.Tech. (by Research) degree from Vellore Institute of Technology (VIT), Vellore, India, in 2016 and 2020, respectively, and the Ph.D. degree from the School of Information Technology and Engineering, VIT, in 2023. He is currently an Assistant Professor with the School of Computer Science and Engineering, Galgotias University, Greater Noida, India. He has more than six years of research experience. He is an author of many journals and conferences. He is a reviewer of many reputed Q1 and Q2 journals. His current research interests include machine learning, deep learning, remote sensing, and big data.

**V. MUTHUKUMARAN** was born in Vellore, Tamil Nadu, India, in 1988. He received the B.Sc., M.Sc., and M.Phil. degrees in mathematics from Thiruvalluvar University Serkkadu, Vellore, India, in 2009, 2012, and 2014, respectively, and the Ph.D. degree in mathematics from the School of Advanced Sciences, Vellore Institute of Technology, Vellore, in 2019.

He is currently an Assistant Professor with the Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. He has four and half years of teaching experience and eight years of research experience and he has published various research articles in high-quality journals Springer, Elsevier, IGI Global, Emerald, and River. He has published more than 80 research articles and 12 book chapters in peer-reviewed international journals. He has published ten IPR patents in algebraic with the IoT applications. He also presented 25 papers presented at national and international conferences. His current research interests include machine learning, data science, block chain, data mining, and algebraic cryptography. He is a fellow of the International Association for Cryptologic Research (IACR). He has also been a Reviewer of several international journals including, *Journal of Intelligent Manufacturing* (Springer), *International Journal of Intelligent Computing and Cybernetics*, *International Journal of e-Collaboration* (IJeC), *International Journal of Pervasive Computing and Communications* (IJPCC), *International Journal of System Assurance Engineering and Management* (IJSA), *International Journal of Speech Technology* (IJST)-Springer, *Journal of Reliable Intelligent Environments* (JRIE), *International Journal of Information Technology and Web Engineering* (IJITWE), *Applied Sciences*, *Symmetry*, *Mathematics*, *Computers in Biology and Medicine*, and *Microprocessors and Microsystems*.

**JAEHYUK CHO** received the Ph.D. degree in computer science from Chung-Ang University, South Korea, in 2011, with a focus on mobile and embedded computing systems. He was a Professor with the Department of Electronic Engineering, Soongsil University. He was a National Research and Development Program Project Manager with Korea Institute of Science and Technology Evaluation and Planning (KISTEP), Seoul. He was a Senior Researcher with LG CNS, Seoul. He is currently a full-time Professor with the Department of Software Engineering, Jeonbuk National University, Jeonju, South Korea. His research interests include applied AI, data process, big data of sensors, the IoT, smart city, and SW platform systems.

**SATHISHKUMAR VEERAPPAMPALAYAM EASWARAMOORTHY** (Member, IEEE) received the bachelor's degree in information technology from Madras Institute of Technology, Anna University, in 2013, the Master's degree in biometrics and cyber security from PSG College of Technology in 2015, and the Ph.D. degree from the Department of Computer and Communication Engineering, Sunchon National University, in 2021. He is a Lecturer with the Department of Data Science and Artificial Intelligence, Sunway University, Malaysia. He has held significant roles, including Postdoctoral Researcher positions at the Department of Software Engineering, Jeonbuk National University, Republic of Korea, and the Department of Industrial Engineering, Hanyang University, Republic of Korea. In 2021, he served as an Assistant Professor with the Department of Computer Science and Engineering, Kongu Engineering College, Erode, India. He is a reviewer of more than 200 journals and has reviewed more than 2000 articles. His current research interests include data mining, machine learning, big data analytics, quantum computing, and high performance computing. Through the Korean Government Scholarship Program, he completed a one-year Korean Language program at Inha University. He currently serves as an academic editor for the journals Information Research Communications and BMC Research Notes.