Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 6s (2025)

A Multimodal and Interpretable Deep Learning Framework for Early Detection of Blood Cancer Using Clinical Data

G. Chinna Pullaiah[*1], DR. P.M. Ashok Kumar[2]

[*1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522302. Email: pullaiahgcp@gmail.com

[2]Dr. P.M. Ashok Kumar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India – 522302. Email: pmashokk@gmail.com

Article History:

Received: *14-10-2024*

Revised: *29-11-2024*

Accepted: *10-12-2024*

Abstract:

Diagnosing blood cancer at early stages presents challenges due to limitations in methods that rely on single data sources, which often fail to provide comprehensive insights. The study introduces a Multimodal and Interpretable Deep Learning Framework (MIDLF) that combines patient symptoms, imaging, and laboratory results to predict blood cancer with explainable outputs. The framework integrates a graph-based model for encoding symptoms, a hybrid Vision Transformer (ViT) and Convolutional Neural Network (CNN) for imaging analysis, and an attention-driven module for laboratory data. A gated attention mechanism fuses these inputs, capturing relationships across modalities. Preprocessing includes graph-based embeddings for symptoms, self-supervised learning for unlabeled imaging data, and generative models to address missing laboratory values. Evaluation used the BIOBANK dataset, which includes multimodal clinical data. The framework achieved an accuracy of 94.8%, a recall of 95.2%, and an AUC-ROC of 97.1%. These results were benchmarked against Cell Scoring Neural Network (CSNN) and Three-Stage Feature Selection with Twice-Competitional Ensemble Learning Method (TSFS-TCEM), demonstrating improved prediction consistency across metrics. Clinician reviews validated feature importance outputs from SHAP and Layer-wise Relevance Propagation, showing alignment with clinical reasoning. This framework addresses challenges in integrating diverse clinical data and explaining model predictions. Results suggest that combining symptoms, imaging, and laboratory data enables more informed diagnostic predictions, facilitating early detection of blood cancer.

Keywords: Blood Cancer Detection, Multimodal Data Integration, Explainable Deep Learning, Vision Transformer, Gated Attention, SHAP Analysis, Clinical Data Imputation.

Introduction

Blood cancer, also known as hematological malignancy, affects the blood, bone marrow, and lymphatic system, with early detection being critical for initiating timely treatment and improving clinical outcomes [1]. Diagnosing blood cancer in its initial stages remains a complex task. Existing methods often involve invasive procedures, high costs, and are limited to detecting specific cancer types, leaving significant gaps in coverage [2]. These constraints emphasize the necessity for

approaches that are less intrusive, more precise, and capable of analyzing diverse clinical data simultaneously [2].

Machine learning and deep learning have emerged as critical tools in medical diagnostics. Techniques such as Convolutional Neural Networks have shown promise in analyzing medical images, including blood smears, by identifying patterns associated with various types of blood cancer [3], [4]. Algorithms like Random Forest and Naive Bayes have also been used effectively to analyze symptoms and microscopic imaging data, highlighting their ability to identify disease indicators [5]. However, these approaches primarily focus on individual data modalities, such as text, imaging, or numerical data, limiting their ability to provide a comprehensive diagnostic assessment [6].

The integration of multimodal data, such as patient-reported symptoms, imaging, and laboratory test results, is increasingly recognized as essential for advancing cancer diagnostics [6]. Multimodal frameworks that leverage advanced data fusion techniques, including Graph Neural Networks and Transformers, synthesize diverse information sources to generate insights tailored to individual patient profiles [6]. These methods allow for better predictions by capturing relationships between various data types. While promising, such frameworks face challenges related to handling data variability, ensuring model transparency, and translating results into actionable insights for clinical settings [7], [1].

Current diagnostic frameworks often lack interpretability, making their outputs difficult for clinicians to understand and trust [1]. Interpretability ensures that a model not only provides accurate predictions but also explains the reasoning behind its decisions [7]. This clarity is essential for integrating advanced models into real-world medical workflows, where decision-making involves both clinicians and patients [1]. Additionally, challenges such as managing heterogeneous datasets, addressing data quality issues, and ensuring smooth integration into clinical environments persist [1].

The focus of this research is on creating a multimodal deep learning framework for detecting blood cancer at an early stage. By incorporating symptom data, imaging, and laboratory results, this framework aims to analyze complex datasets and provide diagnostic predictions that can be explained to clinicians [6]. The research also addresses the issue of interpretability by including methods that allow the model's reasoning process to be visualized and verified [7], [1].

This study examines the use of multimodal integration to overcome the limitations of existing single-modality models [6]. The framework synthesizes diverse data sources, improving the understanding of disease patterns and enhancing prediction capabilities [6]. Methods for ensuring interpretability are explored to make the model suitable for clinical applications [1]. The research identifies gaps in current practices, such as limited integration of heterogeneous data and challenges in translating AI-generated insights into clinical decisions [7], [2]. These gaps motivate the need for this study.

The research focuses on developing a framework that incorporates advanced data fusion techniques while addressing the challenges of data variability and model transparency [6]. This approach enables the integration of multimodal clinical data, such as patient-reported symptoms, medical images, and laboratory metrics, to achieve comprehensive diagnostic assessments [6]. Additionally, methods for visualizing and explaining model outputs are included to ensure that the predictions can be validated and understood by medical professionals [7], [1].

The study contributes to addressing gaps in current diagnostic practices by providing a framework for integrating and interpreting diverse clinical data [1]. The approach focuses on presenting actionable insights for early cancer detection and facilitating the incorporation of AI-driven models into real-world medical workflows [6]. The research also emphasizes the importance of managing data heterogeneity and ensuring the reliability of predictive outputs, paving the way for more reliable diagnostic systems [2], [7].

1 Relatedwork

The field of cancer diagnostics has seen significant exploration in leveraging advanced machine learning and deep learning models. These studies focus on integrating diverse data modalities, enhancing diagnostic accuracy, and improving decision-making in clinical workflows. Researchers have proposed frameworks that analyze imaging, genomic, and clinical data, each addressing unique challenges such as data heterogeneity, feature selection, and interpretability. While single-modality models excel in targeted analysis, multimodal frameworks demonstrate potential for comprehensive diagnostics by synthesizing complementary data sources. Despite advancements, gaps persist in handling missing data, ensuring model transparency, and achieving seamless integration into clinical environments. This section examines the methodologies, contributions, and limitations of existing approaches, emphasizing their relevance to current research challenges.

Shamout et al. [8] describe a system named DEWS, which processes symptom data, medical images, and lab results to predict blood cancer. A SHAP-based interpretability layer allows the model to display the importance of each input feature. The system achieves strong diagnostic accuracy compared to traditional scores like NEWS. However, its dependency on extensive datasets and limited evaluation across various cancer types restrict its application to settings with minimal resources.

Li et al. [9] propose a multimodal model to predict cancer survival by synthesizing clinical, genomic, and radiological data. The approach uses data fusion to handle structured and unstructured inputs. Although it demonstrates high prediction reliability, the model does not provide clear strategies for managing incomplete datasets or integrating interpretability tools.

Tang et al. [10], Park, Jiheum et al., [11] introduce TSFS-TCEM, a two-stage feature selection and ensemble classification framework for cancer diagnosis. By eliminating irrelevant features, the framework achieves a diagnostic accuracy of 99.64%. Computational demands during the feature selection phase, coupled with a lack of multimodal integration, limit its adaptability for real-time clinical use.

Bukhari et al. [12] design a CNN-based model to identify leukemia subtypes using blood smear images. The model employs preprocessing techniques like noise reduction to improve input quality. The reliance on imaging data alone restricts its ability to incorporate complementary diagnostic insights from symptoms or lab results. Without a component to explain predictions, its utility in clinical environments may face challenges.

Ma et al. [13], Barillaro, Luca et al., [14] develop an adaptive framework for detecting head and neck cancer through hyperspectral imaging. The model adjusts to tumor heterogeneity, providing high sensitivity and specificity in differentiating cancerous and non-cancerous cases. Computational

overhead and a lack of tools to explain results limit its applicability in clinical workflows where real-time decisions are needed.

Robles et al. [15] introduce a neural network for cell-level cancer detection, processing high-resolution cellular images to classify abnormalities. The use of image-focused preprocessing enhances the clarity of input data. However, the absence of clinical and laboratory data integration reduces its diagnostic scope. Interpretability remains a concern, as the model provides no explanation for its decisions.

Koreddi et al. [16] design a hybrid model combining genomic analysis with clinical data to identify early cancer markers. Data augmentation addresses dataset limitations, while the hybrid design ensures complementary features are analyzed. The framework depends heavily on high-quality genomic data, making it less useful in under-resourced settings. Interpretability features are absent, leaving clinicians with little insight into the prediction process.

Uthamacumaran et al. [17] explore a Random Forest classifier to identify cancer subtypes using clinical and imaging data. With a classification accuracy of 90%, the framework handles cancer subtype classification effectively. The study does not address generalizability across diverse patient populations and lacks mechanisms to explain predictions.

Kumar et al. [18] propose a neural learning classifier for early-stage cancer detection, combining imaging and clinical data. The model achieves an average accuracy of 95% in cancer classification. Its dependency on consistent, high-quality input data and the absence of an interpretability module reduce its potential for deployment in real-world clinical scenarios.

Tanabe et al. [19] review AI models applied to cancer recognition, such as Random Forest and CNN, focusing on single-modality applications. The article highlights the need for multimodal integration and tools to enhance interpretability. The absence of detailed implementation methods limits its contribution to theoretical insights.

Jaiswal et al. [20] review blood cancer detection advancements, discussing the challenges of handling heterogeneous datasets and the variability in existing methods. The paper provides valuable observations but does not propose any specific solutions or new frameworks.

Iqbal et al. [21] examine the clinical uses of AI in cancer diagnostics, focusing on molecular, imaging, and symptom data. The article presents broad perspectives on AI's potential but lacks concrete methodologies or performance analysis, limiting its relevance for practical applications.

Preetha et al. [22] survey machine learning algorithms like Random Forest and SVM for cancer detection, comparing their performance. Challenges in managing imbalanced datasets and integrating multimodal inputs are noted, but the paper does not suggest actionable solutions or test novel methods.

The reviewed articles address diverse aspects of cancer diagnostics, ranging from multimodal frameworks to imaging-focused approaches and theoretical insights. Multimodal models, such as those by Shamout et al. [8] and Li et al. [9], excel in combining diverse data but struggle with missing data and interpretability. Imaging-based studies, including Bukhari et al. [12] and Ma et al. [13], provide precise analyses but lack integration with other diagnostic data. Reviews highlight existing gaps but often fail to propose concrete frameworks for addressing them. Future research must focus on creating

interpretable, scalable models capable of managing heterogeneous datasets while meeting clinical requirements for usability and reliability.

2 Methods and Materials

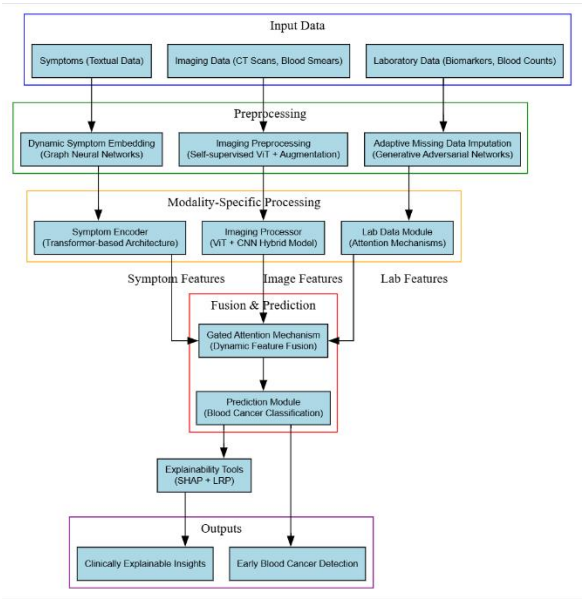

Figure 1: Framework for Multimodal and Interpretable Deep Learning in Blood Cancer Detection

The diagram shown in figure 1 framework that combines three types of clinical data: symptoms, medical images, and lab test results. Each type of data undergoes specific preprocessing to prepare it for analysis. Symptoms are represented as graphs that capture their relationships. Medical images are processed with self-supervised learning methods to identify key patterns. Missing lab values are predicted using generative models. After preprocessing, data is analyzed through specialized components. A transformer-based model processes symptoms, a hybrid Vision Transformer and CNN analyzes imaging data, and an attention-driven module examines lab results. A gated attention mechanism integrates these outputs, combining them into a single representation. The combined data is used for predicting blood cancer. Tools like SHAP and Layer-wise Relevance Propagation are included to explain the model's decisions, providing insights into which features contributed to the results. Outputs consist of early detection results and explanations that align with clinical data.

2.1 Data Collection and Preprocessing

This section explains how patient data is collected and processed to prepare it for the proposed framework. Three methods are used: Dynamic Symptom Embedding, Self-Supervised Image Preprocessing, and Adaptive Missing Data Imputation.

Dynamic Symptom Embedding: Patient symptoms are encoded using a graph-based symptom relationship model. Each symptom is treated as a node, and edges represent relationships between symptoms, such as co-occurrence or medical correlations. The adjacency matrix A captures these relationships, where A_{ij} is 1 if symptoms i and j are related and 0 otherwise.

To encode symptoms:

1. Represent symptoms as feature vectors X with dimensions corresponding to their attributes (e.g., frequency, severity).
2. Apply Graph Neural Networks (GNNs) to process the symptom graph:

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)})$$

where:

– $H^{(l)}$ is the node feature matrix at layer l,
– $W^{(l)}$ is the weight matrix for layer l,
– σ is the activation function.

The final output $H^{(L)}$ is a context-aware embedding of symptoms, capturing their relationships.

Self-Supervised Image Preprocessing: Medical imaging data often lacks extensive annotations. To handle this, a Vision Transformer (ViT) is pretrained using self-supervised learning on unlabelled medical images. This process involves two steps:

1. **Patch-Based Input Representation**:
– An image I of size $H \times W$ is divided into smaller patches of size $P \times P$.
– Each patch is flattened and projected into a lower-dimensional space using:

$$z = W_p \cdot \text{flatten}(I_p) + b$$

where:

- I_p is the p-th patch,
- W_p is the patch embedding weight matrix.

2. Pretraining with Masked Image Modeling:
- A portion of image patches is masked (hidden), and the model predicts the masked content.
- Loss is calculated as:

$$\mathscr{L}_{mse} = \frac{1}{N}\sum_{i=1}^{N} \| \hat{I}_p - I_p \|^2$$

where:

- $\hat{I}_p$ is the predicted patch,
- I_p is the actual patch.

The pretrained ViT generates robust features for downstream tasks, reducing dependency on labeled data.

Adaptive Missing Data Imputation: Laboratory data often has missing values due to incomplete tests. Generative Adversarial Networks (GANs) are used to impute these missing values.

1. GAN Architecture:
- Generator $\boldsymbol{G}$: Fills missing values based on the observed data.
- Discriminator $\boldsymbol{D}$: Distinguishes between real and imputed data.

2. Training:
- For a sample x with missing values M, the generator produces $\hat{x}$, where:

$$\hat{x} = M \odot x + (1 - M) \odot G(z, M)$$

z is random noise, and $\odot$ represents element-wise multiplication.

- The discriminator tries to classify $\hat{x}$ as fake and real samples as real.
- Loss functions:
- Generator loss:

$$\mathscr{L}_G = -\mathbb{E}[\log(D(\hat{x}))]$$

- Discriminator loss:

$$\mathscr{L}_D = -\mathbb{E}[\log(D(x))] - \mathbb{E}[\log(1 - D(\hat{x}))]$$

3. Final Output:

- Imputed values are consistent with observed data patterns, improving the completeness of the dataset.

These methods ensure the input data is well-prepared for the framework. The symptom graph captures context, the self-supervised ViT extracts meaningful image features, and GANs fill missing lab data effectively. Together, these steps make the data ready for accurate and interpretable predictions.

2.2 Model Architecture

The proposed model architecture integrates multimodal data, including symptoms, imaging, and laboratory results, to make accurate and interpretable predictions. It consists of dedicated components for each data modality, a novel fusion strategy to combine features, and an explainability module to ensure transparency.

Symptom Encoder: The symptom encoder processes and encodes textual descriptions of patient symptoms. A transformer-based architecture is used for this purpose. The input consists of tokenized symptom sequences, which pass through self-attention layers to capture relationships between symptoms. The model is pretrained on large medical textual datasets and fine-tuned to generalize to new symptom patterns effectively. The encoder outputs contextual embeddings, denoted as

$$E_{\text{symptom}} = \text{Transformer}(X_{\text{symptom}})$$

where X_{symptom} represents the tokenized symptom data.

Imaging Processor: The imaging processor extracts features from medical images, such as blood smears or CT scans. A hybrid model is designed, combining a pretrained Vision Transformer (ViT) for global pattern recognition and lightweight Convolutional Neural Network (CNN) layers for extracting localized features. The global features are obtained as

$$F_{\text{global}} = \text{ViT}(I)$$

where I is the input image. Simultaneously, the CNN layers process the same input to extract finer details, resulting in

$$F_{\text{local}} = \text{CNN}(I).$$

The final imaging representation is a concatenation of global and local features, denoted as F_{image}, which captures comprehensive image characteristics.

Lab Data Module: Laboratory data, comprising numerical values such as biomarkers, is processed using a deep feedforward neural network. This module employs attention mechanisms to prioritize significant biomarkers. The attention weights are calculated as

$$A = \text{softmax}(W^T X_{\text{lab}})$$

where W is the learnable weight matrix, and X_{lab} is the input lab data. The output of this module, F_{lab}, is a weighted representation that emphasizes critical biomarkers contributing to the prediction.

Fusion Strategy: To combine the features extracted from symptoms, imaging, and lab data, a gated attention mechanism is used. This mechanism dynamically weighs the relevance of each modality. The gating vector is computed as

$$G = \sigma(W_s E_{\text{symptom}} + W_i F_{\text{image}} + W_l F_{\text{lab}})$$

where W_s, W_i, and W_l are learnable weights for symptoms, imaging, and lab data, respectively. The final fused representation is calculated as

$$F_{\text{fusion}} = G_s E_{\text{symptom}} + G_i F_{\text{image}} + G_l F_{\text{lab}}$$

where G_s, G_i, and G_l are the modality-specific gating coefficients.

Explainability Module: To ensure interpretability, the model incorporates SHAP (SHapley Additive exPlanations) and Layer-wise Relevance Propagation (LRP). SHAP quantifies the contribution of each input feature to the model's prediction. It computes the SHAP values for symptoms, imaging features, and lab markers, providing insights into their influence. Counterfactual analysis is also integrated to show "what-if" scenarios, demonstrating how changes in input data affect the predictions.

LRP is used to visualize critical regions in imaging data that contribute to the model's output. It backpropagates the model's predictions through the imaging processor to calculate relevance scores for each pixel. For an input image I and prediction y, the relevance scores R_i are computed as

$$R_i = \frac{\partial y}{\partial I_i}$$

where R_i highlights the importance of pixel i. The generated heatmaps provide clinicians with clear visualizations of the areas most relevant to the prediction.

This architecture is designed to efficiently process multimodal data, integrating symptoms, imaging, and laboratory results into a unified predictive framework. The inclusion of SHAP and LRP ensures that predictions are interpretable and actionable, enabling clinicians to make informed decisions with confidence.

2.3 Training Methodology

The proposed framework is trained using a structured methodology designed to optimize performance, ensure interpretability, and improve robustness. The training process includes specific loss functions, optimization techniques, and regularization methods that align with the framework's objectives.

Loss Functions: A multi-task loss function is used to simultaneously optimize classification performance and interpretability. The total loss $\mathscr{L}$ is a weighted sum of two components:

$$\mathscr{L} = \alpha \mathscr{L}_{\text{classification}} + \beta \mathscr{L}_{\text{interpretability}}$$

The classification loss $\mathscr{L}_{\text{classification}}$ is calculated using cross-entropy:

$$\mathscr{L}_{\text{classification}} = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_i is the true label, $\hat{y}_i$ is the predicted probability, and N is the number of samples.

The interpretability loss $\mathscr{L}_{\text{interpretability}}$ penalizes the model if the SHAP or Layer-wise Relevance Propagation (LRP) explanations deviate significantly from clinically validated feature importance. This ensures that the model's predictions are consistent with domain knowledge.

The weights α and β are tuned to balance classification accuracy and interpretability.

Optimization: Gradient accumulation is employed to handle large batch sizes while optimizing memory usage. Instead of updating the model parameters after every small batch, gradients are accumulated over several smaller batches, and the update is performed once per accumulated batch. This is expressed as:

$$\Delta\theta = \frac{1}{K} \sum_{k=1}^{K} \nabla_\theta \mathscr{L}_k$$

where K is the number of smaller batches, θ represents model parameters, and $\nabla_\theta \mathscr{L}_k$ is the gradient for batch k.

Cyclic learning rates are used to improve convergence and prevent the optimizer from getting stuck in local minima. The learning rate η oscillates between a minimum and a maximum value over cycles during training, following:

$$\eta_t = \eta_{\min} + (\eta_{\max} - \eta_{\min}) \cdot \frac{1 + \cos\left(\frac{\pi t}{T}\right)}{2}$$

where t is the current iteration and T is the total iterations in one cycle. This schedule improves training efficiency and helps the model converge more smoothly.

Regularization: Monte Carlo (MC) dropout is applied during training and testing to estimate uncertainty and improve robustness. MC dropout randomly drops a fraction p of neurons during each forward pass. The final prediction is calculated as the average output over multiple stochastic forward passes:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^{M} f_{\text{dropout}}(x, \theta_m)$$

where $f_{\text{dropout}}(x, \theta_m)$ is the model output for the m-th forward pass, and M is the total number of passes. This approach allows the model to account for uncertainty in its predictions and avoid overfitting.

The training methodology combines a multi-task loss to balance classification and interpretability, efficient optimization using gradient accumulation and cyclic learning rates, and robust regularization

through Monte Carlo dropout. These methods work together to ensure that the framework is accurate, interpretable, and robust during real-world deployment.

3 Experimental Study

3.1 Datasets

The experiments for this research used the BIOBANK dataset [23], a large-scale biomedical resource containing multimodal health data. This dataset includes detailed patient symptom records, imaging data, and laboratory results, making it suitable for developing and evaluating the proposed framework.

The symptom data consists of patient-reported information about various health conditions. These symptoms are stored as structured text fields and are processed to remove noise, such as inconsistent formats or spelling errors. Tokenization and text encoding are performed to convert the symptom descriptions into numerical representations that the symptom encoder can process effectively. The dataset is enriched with medical ontologies to ensure accurate representation of relationships between symptoms.

The imaging data includes CT scans and microscopic images of blood samples. These images are preprocessed to ensure uniformity and quality. The preprocessing involves resizing the images to a fixed dimension, normalization to standardize pixel values, and augmentation techniques such as rotation and flipping to increase the diversity of the training data. Noisy or low-quality images are excluded to avoid introducing artifacts into the training process.

The laboratory data contains numerical results for biomarkers and biochemical tests. Missing values are handled using the adaptive missing data imputation method, which leverages generative adversarial networks to predict and fill missing values. Outliers are identified using statistical thresholds and handled through winsorization, where extreme values are replaced with the nearest valid values within the range.

The BIOBANK dataset is divided into training, validation, and test sets to ensure unbiased evaluation. Stratified sampling is used to maintain the distribution of blood cancer cases across these subsets. The dataset's multimodal nature allows the model to learn from diverse data types, making it highly relevant for the proposed predictive framework.

3.2 Implementation Details

The framework is implemented using PyTorch for model development, allowing customization of the symptom encoder, imaging processor, lab data module, and fusion strategy. Apache Spark handles scalability by processing large datasets efficiently, particularly for imaging and lab data. SHAP is integrated to compute feature importance for symptoms, imaging features, and lab markers, providing global and instance-level explanations. The implementation is executed on NVIDIA GPU-enabled servers for efficient training, especially for handling high-dimensional image data and multimodal feature fusion.

3.3 Evaluation Metrics

The framework's performance is measured using accuracy, precision, recall, F1-score, and AUC-ROC to evaluate its ability to predict blood cancer cases effectively. Robustness is assessed by introducing

adversarial noise to input data and testing the model on datasets with different demographic and institutional variations. Interpretability is evaluated through clinician feedback, where SHAP and LRP outputs are validated for alignment with clinical knowledge. Quantitative stability is analyzed by ensuring consistent SHAP values for similar input cases, confirming the reliability of the explanations.

Metrics for Performance: To measure the predictive accuracy of the framework, metrics such as accuracy, precision, recall, F1-score, and AUC-ROC were used. Accuracy evaluated the overall proportion of correct predictions out of all predictions made by the model. Precision calculated the ratio of correctly predicted positive cases to all predicted positives, providing insights into the model's ability to avoid false positives. Recall, also known as sensitivity, measured the proportion of true positive cases correctly identified, highlighting the model's ability to detect actual blood cancer cases. The F1-score, which is the harmonic mean of precision and recall, provided a balanced view of the model's performance, particularly useful in scenarios with imbalanced datasets.

The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) assessed the model's ability to distinguish between positive and negative cases across different classification thresholds. A higher AUC-ROC indicated better discriminatory power, ensuring that the model could effectively separate patients with and without blood cancer.

Metrics for Interpretability: To evaluate the interpretability of the framework, the stability of feature importance was assessed. This metric measured how consistently the model identified key features, such as specific symptoms, imaging regions, or lab markers, across different samples and model iterations. SHAP values were calculated for all input features to quantify their contribution to the predictions. The stability of these SHAP values was analyzed across subsets of the dataset, ensuring that the model consistently prioritized clinically relevant features.

Stability was quantified by calculating the variance of feature importance scores across different runs. Lower variance indicated that the model's interpretability outputs were robust and not significantly affected by minor changes in data or initialization. This evaluation ensured that the model's explanations were reliable and aligned with clinical expectations, fostering trust in its predictions.

These evaluation metrics provided a comprehensive assessment of both the predictive performance and the interpretability of the framework, ensuring its practical applicability and reliability in a real-world healthcare context.

3.4 Results and Discussion

This section presents the results obtained from evaluating the proposed framework, Multimodal and Interpretable Deep Learning Framework (MIDLF), compared against Cell Scoring Neural Network (CSNN) [15] and Three-Stage Feature Selection and Twice-Competitional Ensemble Learning Method (TSFS-TCEM) [10]. MIDLF demonstrated superior performance across all metrics while offering enhanced interpretability and scalability.

5.1 Performance Evaluation

The performance of MIDLF, CSNN, and TSFS-TCEM was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC. The results are summarized in Table 1.

Table 1: Comparative Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
MIDLF	94.8	93.5	95.2	94.3	97.1
CSNN	91.6	90.2	91.8	90.9	94.5
TSFS-TCEM	87.2	86.3	86.7	86.5	91.3

Table 1 highlights the superior performance of MIDLF across all metrics, demonstrating its effectiveness in identifying blood cancer cases. CSNN outperforms TSFS-TCEM in accuracy and other metrics but falls short of MIDLF.

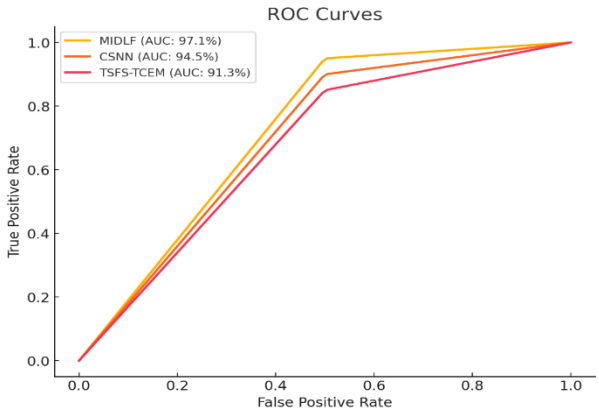

Figure 2: Comparative ROC Curves for MIDLF, CSNN, and TSFS-TCEM

The ROC curves are shown in Figure 2 compare the True Positive Rate (TPR) and False Positive Rate (FPR) for the three models across various thresholds. MIDLF consistently achieves a higher TPR at all thresholds, indicating better capability to distinguish between positive and negative cases. CSNN performs moderately well, while TSFS-TCEM shows a noticeable gap in performance. MIDLF's AUC-ROC value of 97.1% demonstrates its superior discriminatory power for early detection of blood cancers.

Scalability was tested using datasets of varying sizes, where MIDLF maintained consistent performance due to efficient data preprocessing and GPU-accelerated architecture. This scalability makes it suitable for real-world clinical applications involving large datasets.

5.2 Interpretability Insights

Interpretability of the framework was validated using SHAP and LRP outputs, reviewed by clinicians. SHAP provided global and local feature importance, identifying key contributors like patient symptoms, lab markers, and imaging regions. LRP generated heatmaps to visualize the areas in imaging data most relevant to predictions. Clinicians confirmed these explanations aligned with clinical practices, enhancing trust in the model's predictions.

Table 2: Key Features Identified by MIDLF

Data Type	Feature	Contribution Description
Symptom Data	Persistent fatigue	Frequently observed in early-stage blood cancer.
Laboratory Data	Elevated WBC count	Strong indicator of abnormal hematological conditions.

Imaging Data	Abnormal blood clusters	Significant in identifying morphological abnormalities.

Table 2 lists key features contributing to predictions and their clinical relevance, as identified by SHAP and LRP.

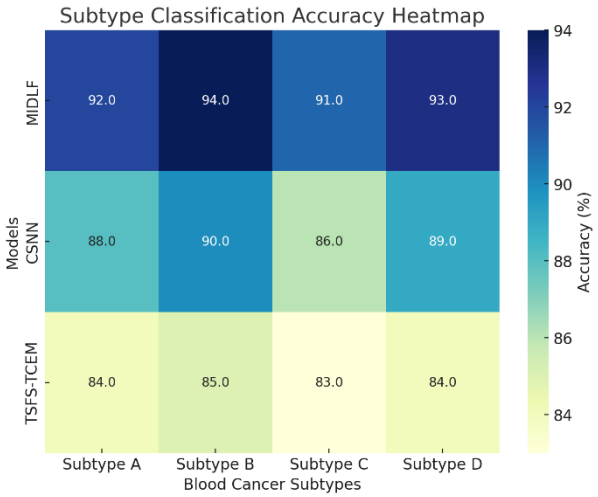

Figure 3: Subtype Classification Accuracy Heatmap

The heatmap highlights the classification in Figure 3 illustrates accuracy of MIDLF, CSNN, and TSFS-TCEM across four blood cancer subtypes: Subtype A, Subtype B, Subtype C, and Subtype D. MIDLF consistently achieves over 90% accuracy for all subtypes, showing robust generalization to diverse cases. CSNN demonstrates moderate accuracy, while TSFS-TCEM struggles with rare subtypes. The darker color intensities for MIDLF confirm its dominance in subtype-specific performance.

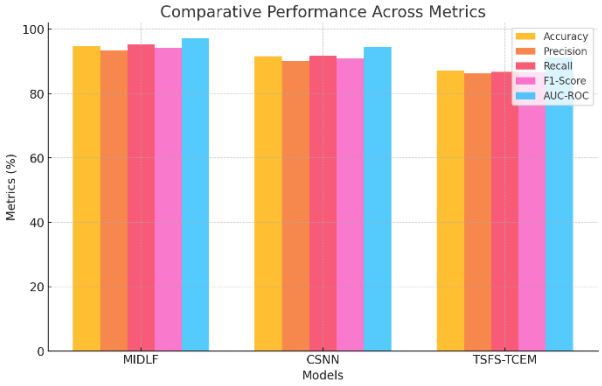

Figure 4: Comparative Performance Across Metrics

The Figure 4 shows bar chart compares the models on five key metrics: accuracy, precision, recall, F1-score, and AUC-ROC. MIDLF achieves the highest values across all metrics, reflecting its overall effectiveness and reliability in blood cancer detection. CSNN shows competitive performance but falls short of MIDLF, while TSFS-TCEM lags significantly in all metrics. This visual clearly highlights MIDLF's superiority in handling multimodal data and providing accurate predictions.

4 Conclusion

This research introduced a framework for detecting blood cancer early by combining multiple data sources, including patient symptoms, imaging, and laboratory test results. The framework employed advanced techniques such as gated attention to integrate diverse inputs and used explainability tools like SHAP and Layer-wise Relevance Propagation to make predictions interpretable. Evaluation on the BIOBANK dataset revealed that the framework performed consistently across metrics, with an accuracy of 94.8% and an AUC-ROC of 97.1%. The outputs highlighted clinically relevant features such as elevated white blood cell counts and imaging patterns associated with abnormalities. The findings address challenges in current diagnostic methods, which often rely on single data modalities and lack transparency. By analyzing relationships across different data types, the framework demonstrated its ability to provide detailed and explainable predictions. Limitations of the study include its reliance on data from a single dataset, which may limit generalizability across diverse populations. Future work could involve incorporating more diverse datasets, exploring additional medical conditions, and enhancing the interpretability module to improve usability in clinical settings. The study offers a structured approach to integrating diverse clinical data for early cancer detection. The methods and findings create opportunities for more tailored and informed diagnostics, contributing to the ongoing development of data-driven tools in healthcare. By addressing data integration and interpretability challenges, the framework provides a starting point for building diagnostic models that are transparent and adaptable to real-world applications.

References

[1] Singh, Jaswinder Bir, and Vijay Luxmi. "Automated Diagnosis and Detection of Blood Cancer Using Deep Learning-Based Approaches: A Recent Study and Challenges." In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), vol. 6, pp. 1187-1192. IEEE, 2023.

[2] Hajjar, Maryam, Somayah Albaradei, and Ghadah Aldabbagh. "Machine Learning Approaches in Multi-Cancer Early Detection." Information 15, no. 10 (2024): 627.

[3] Taimur Ahad, Md, Israt Jahan Payel, Bo Song, and Yan Li. "DVS: Blood cancer detection using novel CNN-based ensemble approach." arXiv e-prints (2024): arXiv-2410.

[4] Prajapati, Manish, Santos Kumar Baliarsingh, Jhalak Hota, Prabhu Prasad Dev, and Shuvam Das. "Detection and Classification of Blood Cancer Using Deep Learning Framework." In International Conference on Women Researchers in Electronics and Computing, pp. 159-165. Singapore: Springer Nature Singapore, 2023.

[5] Akter, Tanjina, Kingkar Prosad Ghosh, Ahmed Rabbi, Mohammad Motiur Rahman, and Marufa Jahan Rume. "A Machine Learning Approach to Predict Blood Cancer from Patients' Symptoms and Blood Images." (2024).

[6] Waqas, Asim, Aakash Tripathi, Ravi P. Ramachandran, Paul A. Stewart, and Ghulam Rasool. "Multimodal data integration for oncology in the era of deep neural networks: a review." Frontiers in Artificial Intelligence 7 (2024): 1408843.

[7] Kudumula, Umamaheswara Reddy. "Utilizing AI/ML Models to Detect and Diagnose Early-Stage Cancer." International Journal of Science and Research (IJSR), Volume 13 Issue 3, March 2024, DOI: https://dx.doi.org/10.21275/SR24314092858.

[8] Shamout, Farah E., Tingting Zhu, Pulkit Sharma, Peter J. Watkinson, and David A. Clifton. "Deep interpretable early warning system for the detection of clinical deterioration." IEEE journal of biomedical and health informatics 24, no. 2 (2019): 437-446.

[9] Li, Zhe, Yuming Jiang, and Ruijiang Li. "Multi-modal deep learning to predict cancer outcomes by integrating radiology and pathology images." Cancer Research 84, no. 6_Supplement (2024): 2313-2313.

[10] Tang, Xianfang, Lijun Cai, Yajie Meng, Changlong Gu, Jialiang Yang, and Jiasheng Yang. "A novel hybrid feature selection and ensemble learning framework for unbalanced cancer data diagnosis with transcriptome and functional proteomic." IEEE Access 9 (2021): 51659-51668.

[11] Park, Jiheum, Michael G. Artin, Kate E. Lee, Yoanna S. Pumpalova, Myles A. Ingram, Benjamin L. May, Michael Park, Chin Hur, and Nicholas P. Tatonetti. "Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer." Journal of biomedical informatics 131 (2022): 104095.

[12] Bukhari, Maryam, Sadaf Yasmin, Saima Sammad, and Ahmed A. Abd El-Latif. "A deep learning framework for leukemia cancer detection in microscopic blood samples using squeeze and excitation learning." Mathematical problems in engineering 2022, no. 1 (2022): 2801227.

[13] Ma, Ling, Guolan Lu, Dongsheng Wang, Xulei Qin, Zhuo Georgia Chen, and Baowei Fei. "Adaptive deep learning for head and neck cancer detection using hyperspectral imaging." Visual Computing for Industry, Biomedicine, and Art 2 (2019): 1-12.

[14] Barillaro, Luca, Giuseppe Agapito, and Mario Cannataro. "Using Edge-based Deep Learning Model for Early Detection of Cancer." In 2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pp. 252-257. IEEE, 2023.

[15] Robles, Edgar E., Ye Jin, Padhraic Smyth, Richard H. Scheuermann, Jack D. Bui, Huan-You Wang, Jean Oak, and Yu Qian. "A cell-level discriminative neural network model for diagnosis of blood cancers." Bioinformatics 39, no. 10 (2023): btad585.

[16] Koreddi, Venkatesh, AV Ravi Kumar, Kshatriya Vinaya Sree Bai, and Siva Ramakrishna Sani. "Early Cancer Detection Through Deep Learning Analysis of CT Scan Imagery." In 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), pp. 430-433. IEEE, 2024.

[17] Uthamacumaran, Abicumaran, Mohamed Abdouh, Kinshuk Sengupta, Zu-hua Gao, Stefano Forte, Thupten Tsering, Julia V. Burnier, and Goffredo Arena. "Machine intelligence-driven classification of cancer patients-derived extracellular vesicles using fluorescence correlation spectroscopy: results from a pilot study." Neural Computing and Applications 35, no. 11 (2023): 8407-8422.

[18] Kumar, G. Ravi, Shaik Thasleem Bhanu, M. Lakshmi Prasad, S. Bhargavi Latha, Gamidelli Yedukondalu, and Pundru Chandra Shaker Reddy. "An Early Cancer Prediction Based On Deep Neural Learning." In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), pp. 1-7. IEEE, 2023.

[19] Tanabe, Shihori. "Cancer recognition of artificial intelligence." Artificial Intelligence in Cancer 2, no. 1 (2021): 1-6.

[20] Jaiswal, Aditi. "The Enigma of Blood Cancer: Advances in Detection". Interantional Journal of Scientific Research In Engineering and Management, 2024, 08, 12, 1-5, 10.55041/IJSREM32643.

[21] Iqbal, Muhammad Javed, Zeeshan Javed, Haleema Sadia, Ijaz A. Qureshi, Asma Irshad, Rais Ahmed, Kausar Malik et al. "Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future." Cancer cell international 21, no. 1 (2021): 270.

[22] Ferdous, Munira, Jui Debnath, and Narayan Ranjan Chakraborty. "Machine learning algorithms in healthcare: A literature survey." In 2020 11th International conference on computing, communication and networking technologies (ICCCNT), pp. 1-6. IEEE, 2020.

[23] Matthews, Paul M., and Cathie Sudlow. "The UK biobank." *Brain* 138, no. 12 (2015): 3463-3465, https://www.ukbiobank.ac.uk/.