A Multimodal Deep Learning Framework for Thyroid Cancer Diagnosis Using Ultrasound Imaging and Clinical Data

M Gopi Chakradhar 121CS0050
K Rohith 121CS0045

Under the supervision of: **Dr. N Srinivas Naik**

Department of Computer Science & Engineering
Indian Institute of Information Technology Design and Manufacturing, Kurnool

Table of Contents

Background & Motivation

- **Thyroid Cancer Challenges:** Rapidly rising incidence, diverse subtypes; ultrasound and FNA are subjective and slow.
- **AI Opportunities:** Deep learning auto-extracts imaging patterns; clinical data (age, TSH, T3, T4) adds diagnostic context.
- **Current Limitations:** AI systems use images or clinical data separately—lacking unified multimodal fusion and explainability.
- **Project Motivation:** Develop an end-to-end, automated, interpretable pipeline combining multimodal fusion and NLP summaries.

Objectives

- **Hybrid Imaging Model:** CNN+ViT for spatialcontextual ultrasound features.
- **Clinical Embedding:** MLP compresses tabular patient data into low-dimensional vectors.
- **Fusion Module:** Gating-based attention integrates embeddings for diagnosisrisk scoring.
- **Generalizable Training:** Fine-tune on three independent ultrasound datasets.
- **Automated Reporting:** BART-based NLP generates concise diagnostic summaries.
- **Evaluation:** Metrics (accuracy, F1, confusion matrix)qualitative summary assessment.

1.3 Problem Statement

Given:

$$\mathcal{I} = \{I_1, \ldots, I_N\}, \quad I_i \in \mathbb{R}^{3\times 640\times 640},$$
$$\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_N\}, \quad \mathbf{c}_i \in \mathbb{R}^d,$$
$$\mathcal{Y} = \{y_1, \ldots, y_N\}, \quad y_i \in \{0, 1, \ldots, 8\},$$
$$\mathcal{R} = \{r_1, \ldots, r_N\}, \quad r_i \in [0, 1].$$

Learn:

$$f_{\text{class}} : (\mathcal{I}, \mathcal{C}) \to \mathbb{R}^9, \quad f_{\text{risk}} : (\mathcal{I}, \mathcal{C}) \to [0, 1].$$

Combined model:

$$f_\theta(I_i, \mathbf{c}_i) = (\hat{y}_i, \hat{r}_i)$$

NLP summarizer:

$$g_\phi(\hat{y}_i, \hat{r}_i, \mathbf{c}_i) \ \to \ \text{diagnostic\_report}_i$$

Literature Review – AI in Thyroid Diagnosis

Ludwig et al. (*The Use of Artificial Intelligence in the Diagnosis and Classification of Thyroid Nodules*, 2023) conducted an extensive review on AI applications in thyroid nodule diagnosis. Their analysis of over 930 studies revealed that **CNN-based models** have achieved performance levels comparable to senior radiologists, particularly in ultrasound image interpretation. The review emphasized the utility of CNNs in real-time detection and classification but highlighted a gap in interpretability and multimodal integration.

Yeon-Jae et al. (*Deep convolutional neural network for classification of thyroid nodules on ultrasound*, 2022) used a CNN model (VGGNet variant) trained on 1421 nodules and found it to be more specific than junior radiologists and nearly as accurate as senior experts. The study also showed that combining AI with junior radiologists improved diagnostic metrics significantly.

Peng et al. (*Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management*, 2021) proposed "**ThyNet**," a deep CNN trained on over 18,000 images, which outperformed both junior and senior radiologists in sensitivity, accuracy, and AUC. Interestingly, combining AI with radiologists further enhanced diagnostic accuracy while reducing unnecessary fine needle aspiration biopsies (FNAB).

Literature Survey

Model	Title	Focus	Performance Metrics	Research Gaps
CNN-based Ultrasonography	CNN for Thyroid Ultrasound Analysis (Ludwig et al., 2023)	• CNNs applied to ultrasound images matched senior radiologists. • Highlighted importance of clinical integration.	• AUC: 0.78–0.94	• Poor interpretability. • Lacks clinical data integration.
VGGNet Variant	Ultrasound-CNN Model for Nodule Classification (Yeon-Jae et al., 2022)	• CNN trained on 1,421 thyroid nodules. • Improved junior radiologist performance.	• Accuracy: 84.2%	• Dataset size limited. • Generalizability unproven.
ThyNet CNN	Deep Learning with ThyNet on Large Dataset (Peng et al., 2021)	• ThyNet trained on 18K+ images. • Outperformed radiologists in AUC and sensitivity.	• AUC: 0.944	• No clinical metadata integration. • Focused solely on image modality.

Literature Survey

Model	Title	Focus	Performance Metrics	Research Gaps
MSAC-DBUNet	Histological Image Segmentation using MSAC-DBUNet	• Multi-scale attention-based CNN for enhanced segmentation. • Compared against standard DBUNet.	• +4.46% Dice vs DBUNet	• High computational cost. • Complex model architecture.
Swin Transformer Fusion	Multimodal Fusion with UTV-STSwin (Ultrasound + Text)	• Combines ultrasound video frames and clinical features. • Robust to noisy data inputs.	• Accuracy: 82.1% • AUC: 94.2%	• Transformer architecture complexity. • Real-time inference limitations.
Cascaded CNN	Histopathology Slide Classification using Cascaded CNN	• Two-stage CNN classifies carcinoma types and subtypes. • Focused on WSI datasets.	• Accuracy: 94.69%	• Dataset size constrained. • Limited follicular subtype diversity.

Gaps Identified

- **Fragmented Use of Modalities:** Ultrasound-only models neglect clinical features (age, gender, TSH, FT4), limiting holistic diagnosis
- **Absence of End-to-End Multimodal Frameworks:** Very few unified architectures jointly learn from images and structured data; existing solutions (e.g., UTV-ST) are complex and not broadly generalizable.
- **Lack of Interpretability & Automated Reporting:** Most systems stop at classification; few offer explainability or patient-friendly summaries—ours uses a BART transformer for contextual reports.
- **Dataset Alignment & Realism:** Reliance on rare paired image-clinical datasets; our approach aligns independent datasets statistically to mirror real hospital workflows.
- **Real-Time Deployment Constraints:** High-compute or specialized hardware requirements hinder adoption; our lightweight CNN-ViT solution runs on common CPUs for broader accessibility.

Positioning Our Research to Address These Gaps

- Multimodal Integration: Existing models use only imaging, ignoring crucial clinical parameters.
- Interpretability & Reporting: High-accuracy systems lack automated, clinician-ready diagnostic summaries.
- Scalability: Methods often depend on specialized hardware or extensive preprocessing, limiting real-world adoption.
- Evaluation Scope: Few studies benchmark a fully fused multimodal pipeline against single-modality baselines.
- Practical Robustness: Current approaches don't ensure real-time inference or handle missing/noisy inputs seamlessly.

Key Contributions of Our Research

- First-of-its-kind multimodal fusion using CNN-ViT and MLP for thyroid cancer diagnosis
- Automated diagnostic reporting via BART-based NLP integration
- Superior diagnostic accuracy and risk stratification compared to unimodal baselines
- End-to-end system architecture that is deployable, interpretable, and adaptable

This strategic positioning not only addresses the current limitations in thyroid cancer diagnostics but also opens pathways for future multimodal, intelligent healthcare systems.

Dataset Description

We leverage four diverse datasets—three ultrasound sets at successive training stages and one clinical tabular set—to ensure robust, high-quality model performance.

3.1 Imaging Datasets

Three independent ultrasound datasets were used in this work. These datasets are not overlapping or semantically aligned, ensuring a robust evaluation of the system's generalization capability:

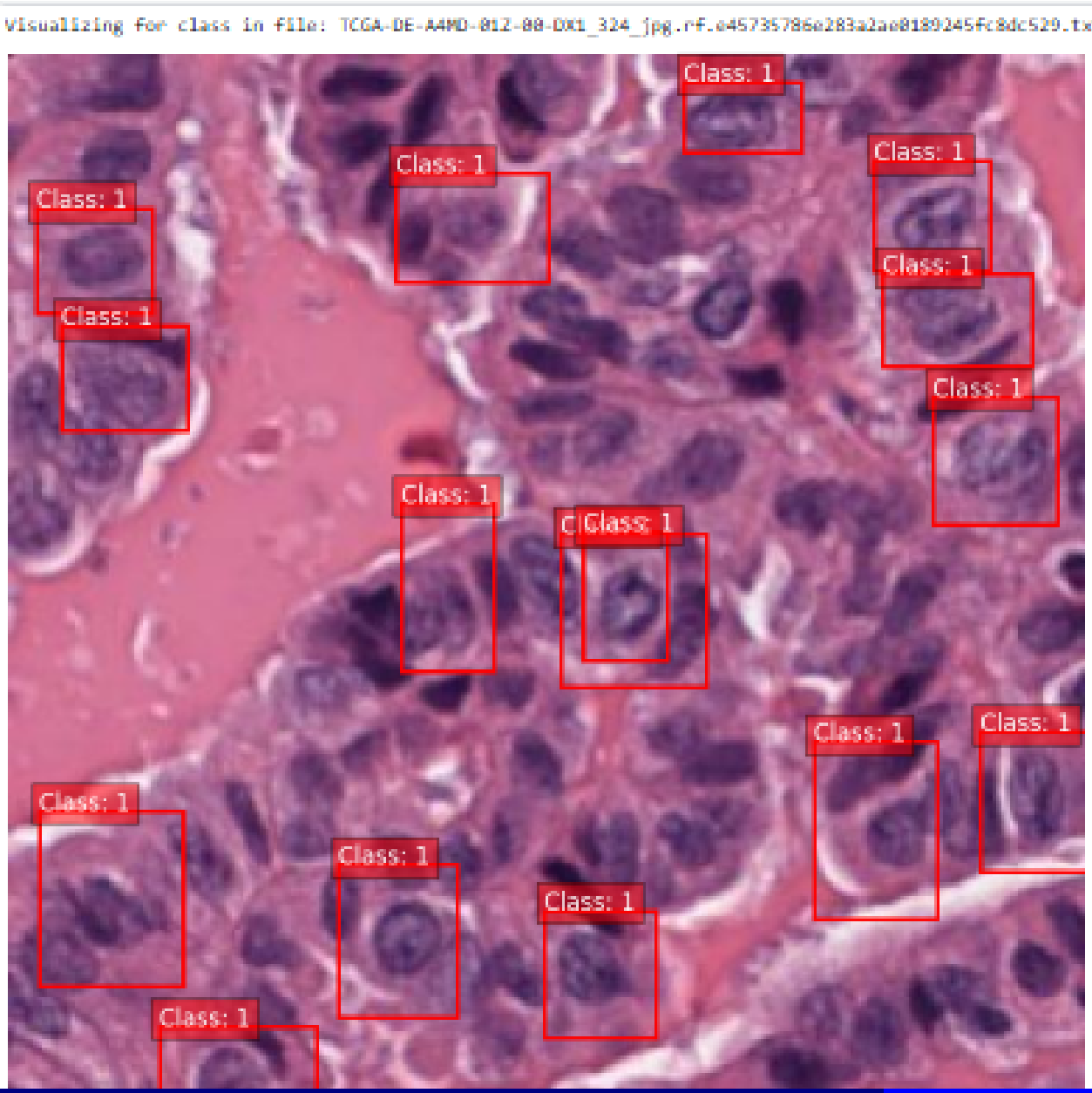
Dataset 1: Base Training Set

- Description: Comprises annotated ultrasound images with label files (.txt) indicating thyroid nodule classes.
- Usage: Used for the final phase of fine-tuning the multimodal model.
- Classes: 9 multiclass categories, derived from clinical annotations.

CSV

- id: the unique identity of the patient who carries the nodule
- age: the age of the patient
- FT3: triiodothyronine test result
- FT4: thyroxine test result
- TSH: thyroid-stimulating hormone test result
- TPO: thyroid peroxidase antibody test result
- TGAb: thyroglobulin antibodies test result
- site: the nodule location, 0: right, 1: left, 2: isthmus
- echo_pattern: thyroid echogenicity, 0: even, 1: uneven
- multifocality: if multiple nodules exist in one location, 0: no, 1: yes
- size: the nodule size in cm
- shape: the nodule shape, 0: regular, 1: irregular
- margin: the clarity of nodule margin, 0: clear; 1: unclear
- calcification: the nodule calcification, 0: absent, 1: present
- echo_strength: the nodule echogenicity, 0: none, 1: isoechoic, 2: medium-echogenic, 3: hyperechogenic, 4: hypoechogenic
- blood_flow: the nodule blood flow, 0: normal, 1: enriched
- composition: the nodule composition, 0: cystic, 1: mixed, 2: solid
- multilateral: if nodules occur in more than one location, 0: no, 1: yes
- mal: the nodule malignancy, 0: benign, 1: malignant

Training metrices



Datasets

- **Ultrasound Imaging (3 independent sets):**
  - **Dataset 1 (Base):** Largest, diverse probe types conditions for robust feature learning.
  - **Dataset 2 (Intermediate Tuning):** Structurally similar; adapts features before final training.
  - **Dataset 3 (Final Fine-tuning):** Annotated (9 classes) images for last-stage model tuning.

  Common: JPEG → resized 640×640; labels in .txt files.
- **Clinical (thyroid_clean.csv):**
  - 900 samples, 18 features (age, gender, TSH, FT3, FT4, Tg, TgAb, TPOAb, . . .); binary target `mal`.
  - Randomly paired with images during training to mimic decoupled hospital workflows.

Preprocessing Techniques

- **Image Preprocessing:**
  - Resize $\rightarrow$ 640×640 (bicubic), normalize pixel-values [0,1]
  - Optional augmentation: random crop, flip, contrast
  - Convert to PyTorch Tensor shape [3×640×640]
- **Label Handling:**
  - *Imaging:* parse .txt $\rightarrow$ integer class indices
  - *Clinical:* drop `id`, `mal`, convert features $\rightarrow$ FloatTensors, pad missing values
- **Clinical Feature Engineering:**
  - Encode categorical (e.g., gender) $\rightarrow$ numeric
  - Z-score normalization of numeric fields
  - MLP reduces 18-dim $\rightarrow$ 2-dim embedding

4.1 System Overview

The entire pipeline is designed as a multistage multimodal architecture composed of three primary modules:

- Imaging Branch: A hybrid deep neural network combining a CNN and a Vision Transformer (ViT) to extract robust local and global features from ultrasound images.
- Clinical Branch: A fully connected MLP that encodes 18-dimensional clinical tabular data into a compact 2D vector.
- Fusion Module: Gating-based learnable mechanism that integrates the two modalities and feeds into:
  - A multiclass classifier for nodule type
  - A regressor to compute the malignancy risk score.

The final outputs—predicted class and risk score—are passed into a BART-based language model, which generates a diagnostic report in natural language.

Imaging Module

- **Hybrid (Local + Global):**
  - **CNN Branch:** captures local spatial patterns (textures, edges).
  - **ViT Branch:** encodes global semantics via transformer on 16×16 patches.
- **Feature Fusion:**
  - Concatenate 512-dim outputs from each branch and project to a unified 512-dim embedding.

4.3 Clinical Model

The clinical input (from thyroid-clean.csv) contains 18 normalized features. A simple yet effective Feedforward MLP compresses this into a low-dimensional embedding:

Clinical-Feature = MLP(clinical-input) $\rightarrow$ shape: (batch-size, 2)

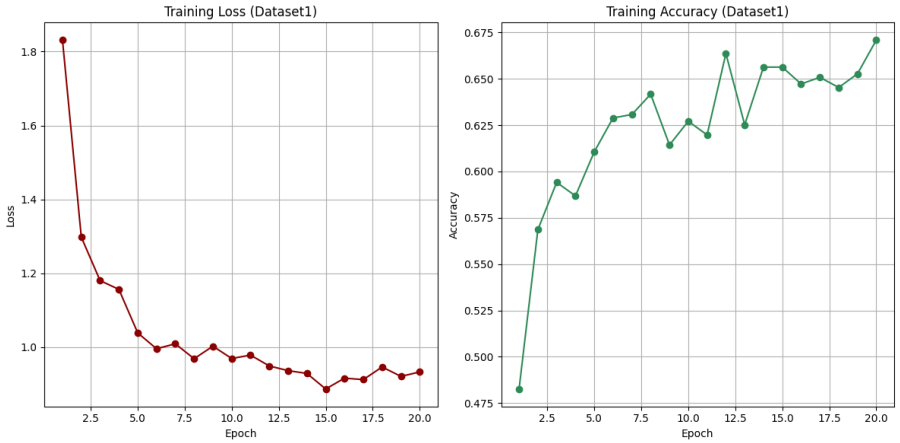
- Three fully connected layers with ReLU activations.
- Output is a 2D feature vector.

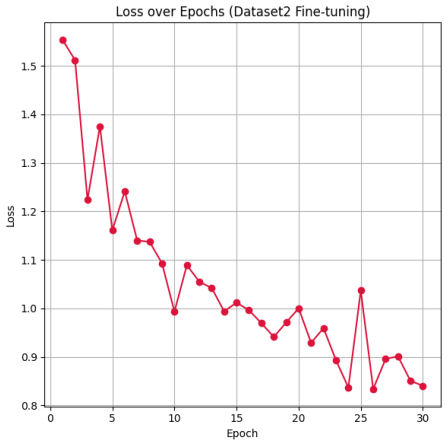
This compact embedding allows the model to integrate clinical context while minimizing overfitting from sparse tabular input.

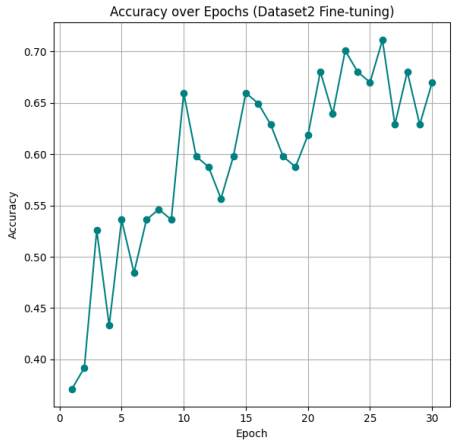
4.4 Fusion Module

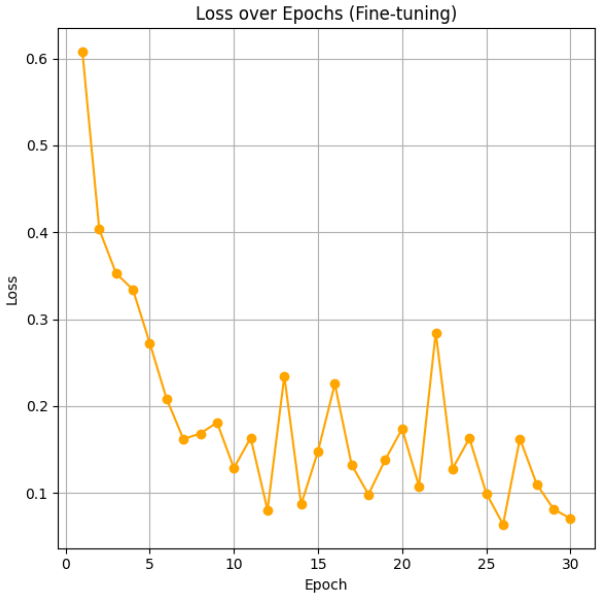
The Fusion Module combines the outputs from both modalities:

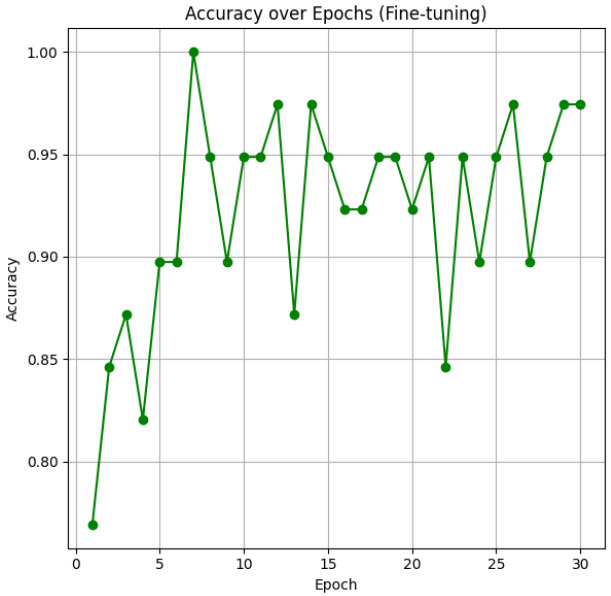
Visualization



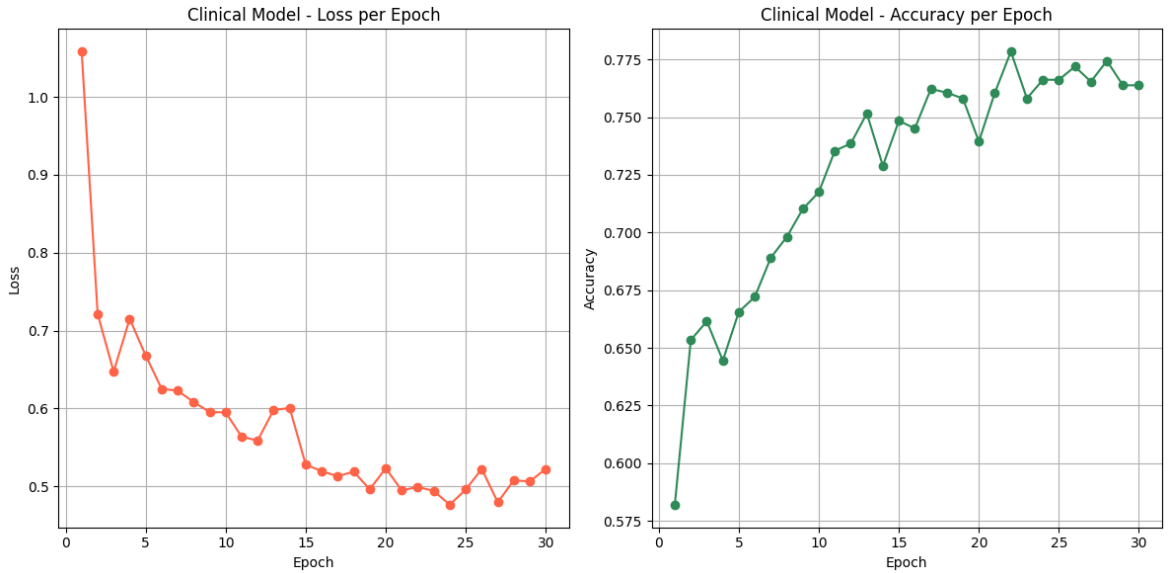




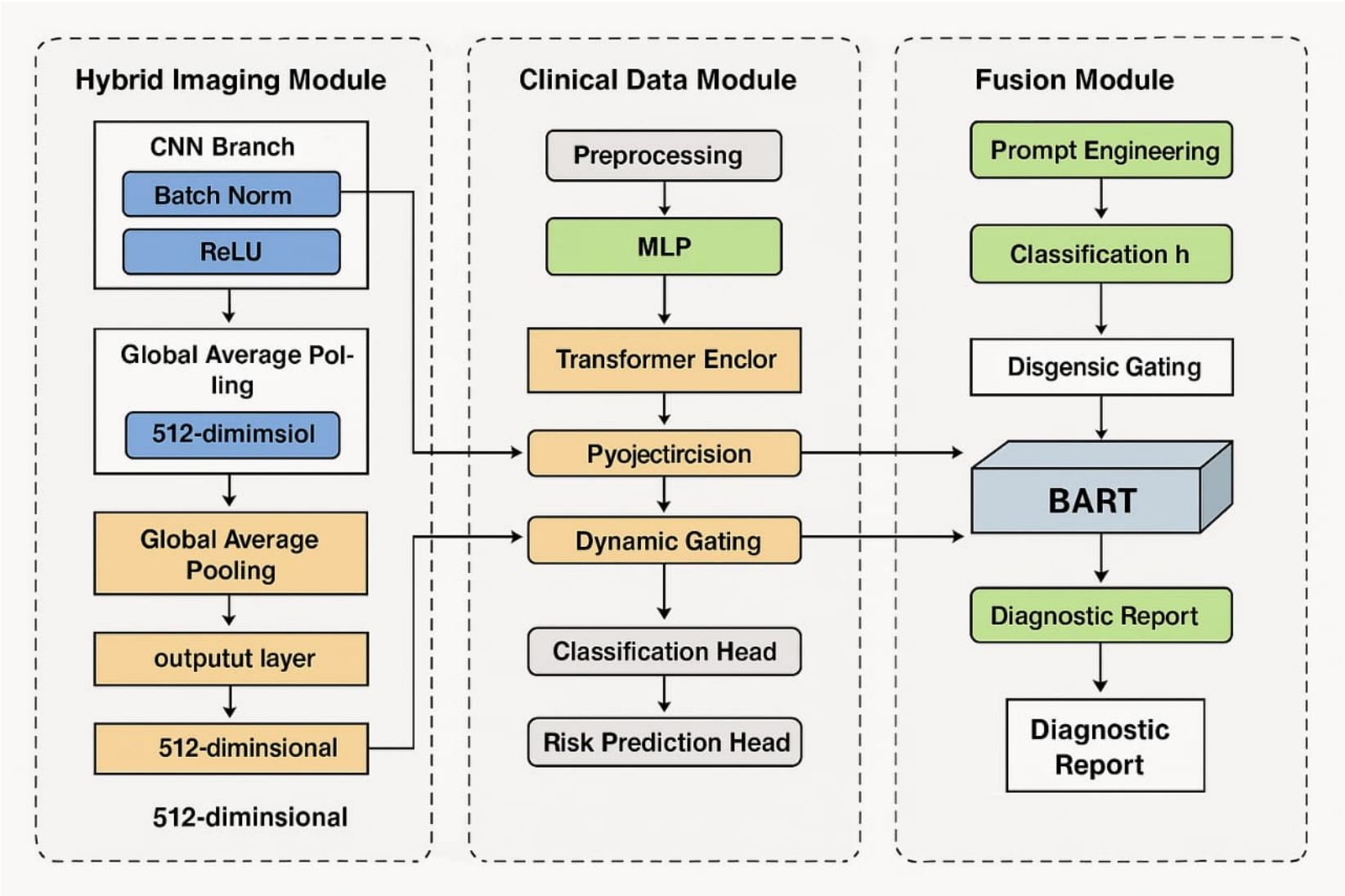




Visualization-(Clinical)



Fusion


Hybrid Imaging Module
CNN Branch
Batch Norm
ReLU
Global Average Pol-
ling
512-dimimsiol
Global Average
Pooling
outputut layer
512-diminsional
512-diminsional
Clinical Data Module
Preprocessing
MLP
Transformer Enclor
Pyojectircision
Dynamic Gating
Classification Head
Risk Prediction Head
Fusion Module
Prompt Engineering
Classification h
Disgensic Gating
BART
Diagnostic Report
Diagnostic
Report

Proposed Methodology: Fusion Module

Step 1: Linear Projection

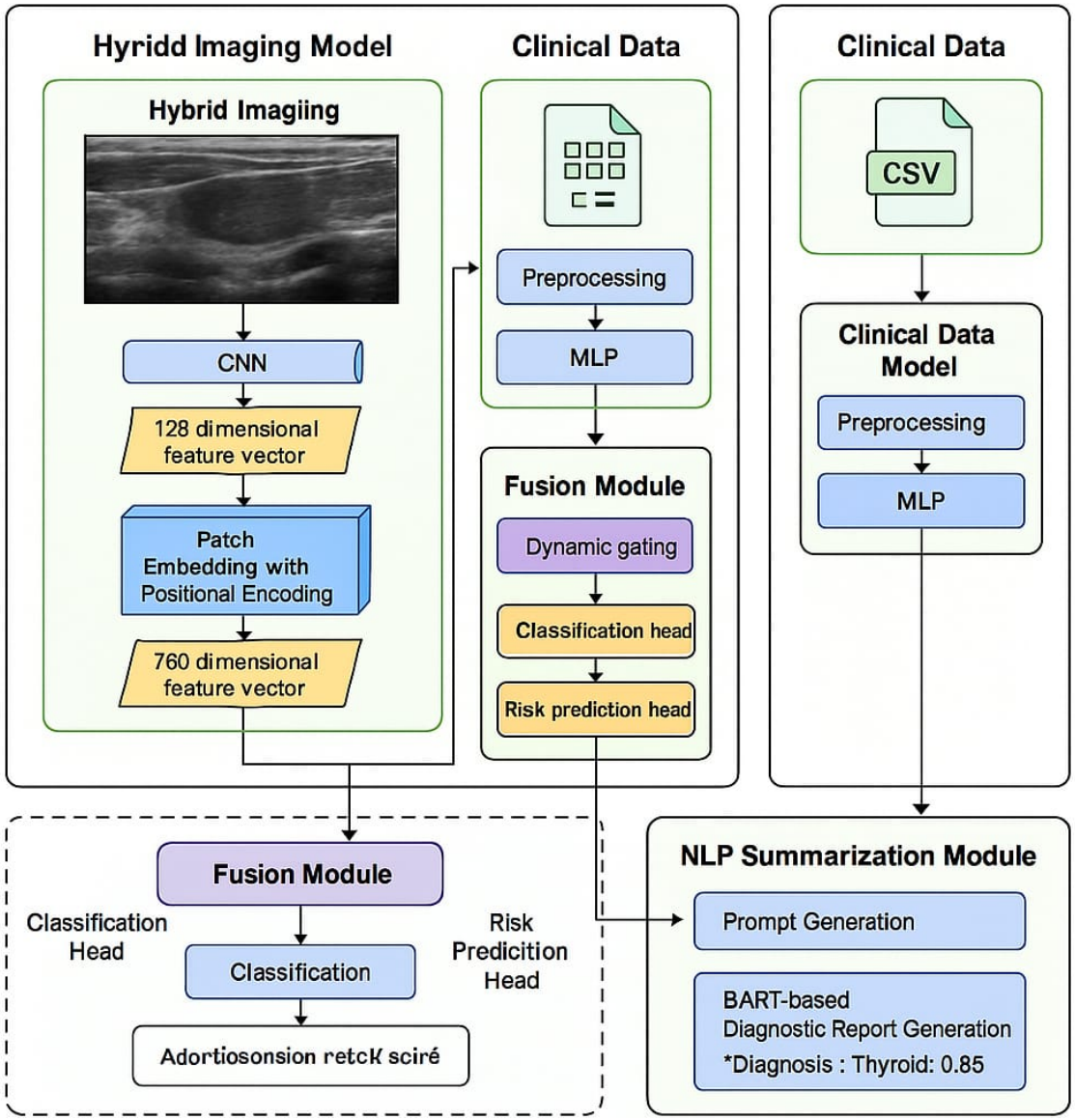
- **Image Vector (CNN-ViT):** Mapped to 256D $\rightarrow$ $\mathbf{f}_{img} = \text{Linear}(\text{Image-Feature})$
- **Clinical Vector (MLP Output):** Mapped to 256D $\rightarrow$ $\mathbf{f}_{clin} = \text{Linear}(\text{Clinical-Feature})$

Step 2: Dynamic Gating Mechanism The gating mechanism learns a soft balance between imaging and clinical signals for each input, making your fusion context-sensitive and more interpretable than fixed or naive strategies

Step 3: Prediction Heads

- **Classification:** $\text{Linear}(\mathbf{f}_{fused}) \rightarrow$ 9-class logits
- **Risk Score:** $\text{Linear}(\mathbf{f}_{fused}) \rightarrow \text{Scalar} \in [0, 1]$

Multi-Model


Hyridd Imaging Model
Clinical Data
Hybrid Imagiing
CNN
128 dimensional feature vector
Patch Embedding with Positional Encoding
760 dimensional feature vector
Preprocessing
MLP
Fusion Module
Dynamic gating
Classification head
Risk prediction head
Clinical Data
CSV
Clinical Data Model
Preprocessing
MLP
Fusion Module
Classification Head
Classification
Risk Predicition Head
Adortiosonsion retcK sciré
NLP Summarization Module
Prompt Generation
BART-based Diagnostic Report Generation
*Diagnosis : Thyroid: 0.85

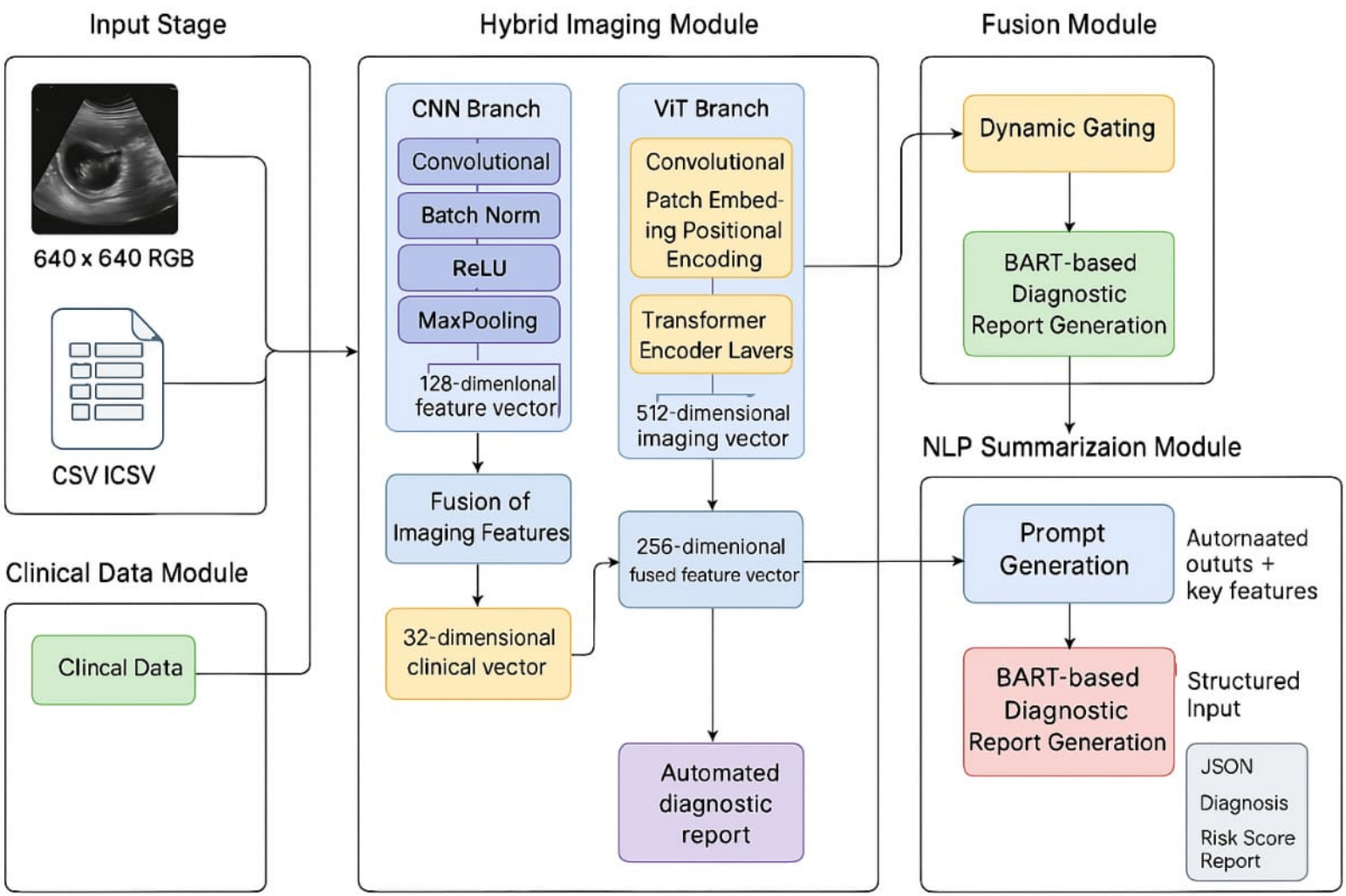
Multimodal Training Strategy

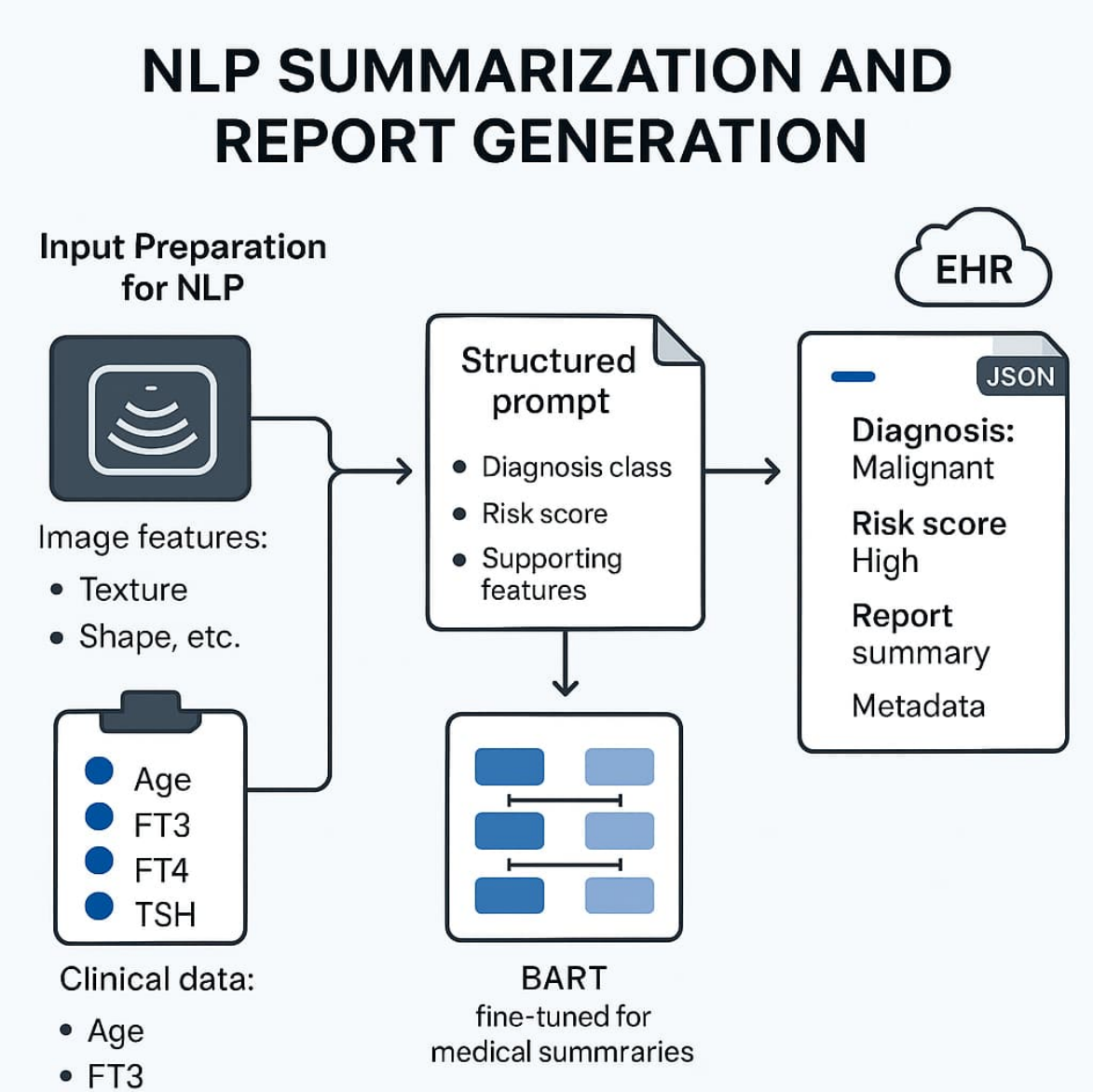
- **Phase 1:** Train hybrid imaging model on Dataset 3 and clinical model from scratch.
- **Phase 2:** Fine-tune imaging backbone on Dataset 2; merge with clinical model into MultimodalModel.
- **Final Phase:** Fine-tune full multimodal model on Dataset 1 with paired image-clinical data.
- **Loss Function:**

$$\text{Loss}_{\text{total}} = \text{CrossEntropy} + \lambda \times \text{MSE}$$

- **Weight Transfer:** Load pretrained weights with `strict=False` to handle minor architecture mismatches.

Model


Input Stage
640 x 640 RGB
CSV ICSV
Clinical Data Module
Clincal Data
Hybrid Imaging Module
CNN Branch
Convolutional
Batch Norm
ReLU
MaxPooling
128-dimenlonal feature vector
Fusion of Imaging Features
32-dimensional clinical vector
ViT Branch
Convolutional Patch Embed-ing Positional Encoding
Transformer Encoder Lavers
512-dimensional imaging vector
256-dimenional fused feature vector
Automated diagnostic report
Fusion Module
Dynamic Gating
BART-based Diagnostic Report Generation
NLP Summarizaion Module
Prompt Generation
Autornaated outuls + key features
BART-based Diagnostic Report Generation
Structured Input
JSON
Diagnosis
Risk Score
Report


NLP SUMMARIZATION AND REPORT GENERATION
Input Preparation for NLP
Image features:
• Texture
• Shape, etc.
Age
FT3
FT4
TSH
Clinical data:
• Age
• FT3
Structured prompt
• Diagnosis class
• Risk score
• Supporting features
BART
fine-tuned for medical summraries
EHR
JSON
Diagnosis: Malignant
Risk score High
Report summary
Metadata

NLP Summarization Pipeline

- **Model:** Pretrained BART (facebook/bart-large-cnn).
- **Prompt Creation:** Combine predicted class, risk score, and clinical data into a structured prompt.
- **Inference:** Pass prompt to BART to generate a natural language summary.


```
{
  "diagnosis": "Papillary Thyroid Carcinoma",
  "risk_score": 0.87,
  "report": "The model predicts a high likelihood of Papillary Thyr
}
```

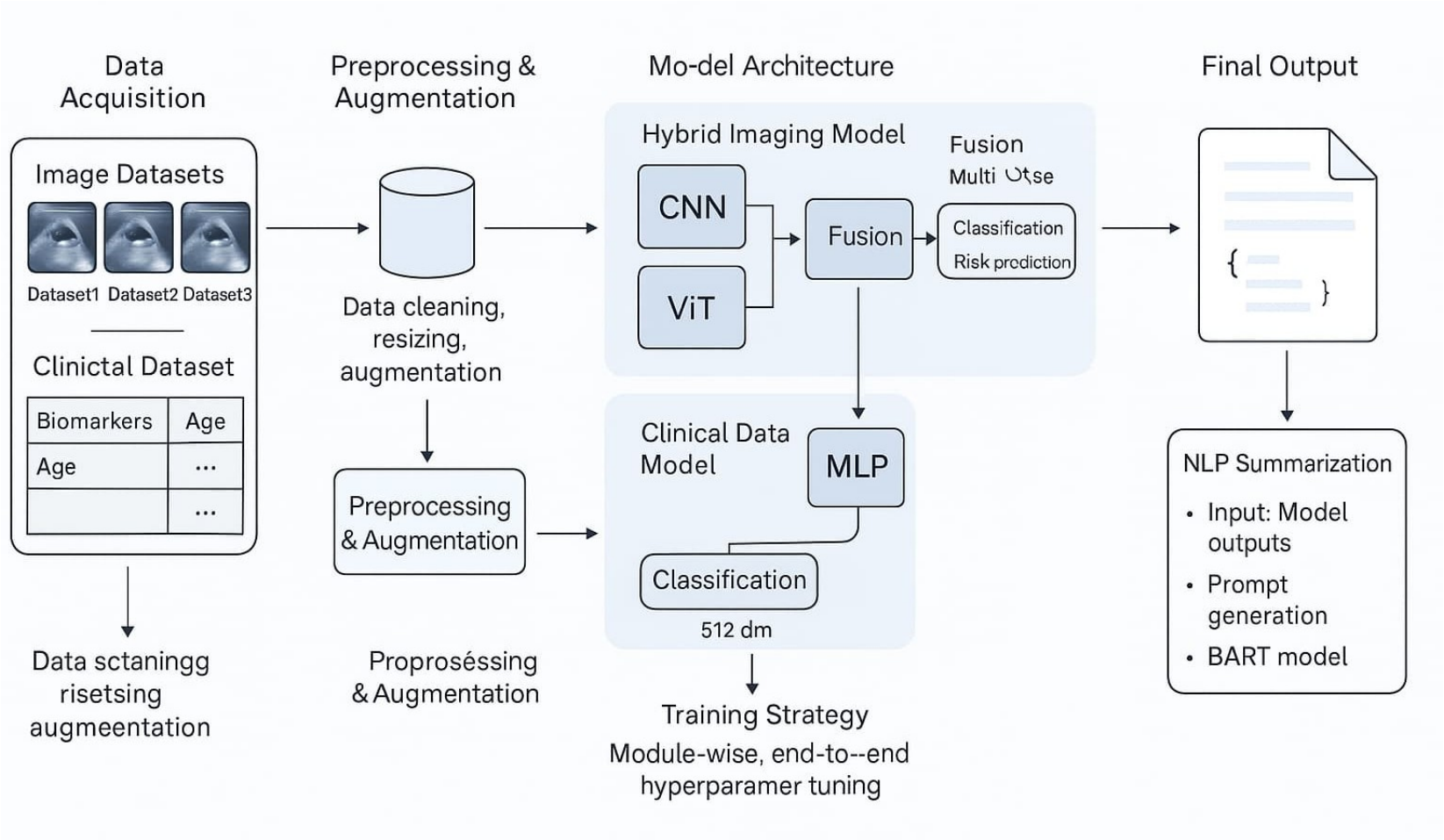

Hyperparameters Optimization

Multiple experiments were conducted with varied hyperparameter settings. The final configuration was chosen based on performance over validation and test sets.

Component	**Value / Choice**
Optimizer	AdamW
Learning Rate	5e-4 (with cosine annealing)
Batch Size	8
Epochs	30
Weight Decay	1e-4
Loss Function	Combined: CrossEntropy + λ * MSE
Lambda (λ)	0.5 (Risk weight factor)
Clinical Embedding	2D from 18 input features
Image Embedding	512D via CNN-ViT hybrid
Fusion Dimensionality	256 (post projection)

Note: During transfer learning, strict=False was used when loading .pth files to accommodate for slight architectural changes (e.g., classifier reshaping).

Model Architecture



Evaluation & Results

6.1 Quantitative Metrics

The final model was tested on a holdout set of 44 ultrasound images with corresponding clinical data.

Key Metrics:

Metric	Value
Accuracy	80–88%
F1-Score (macro)	∼0.82

Observations:

- High classification accuracy in dominant classes (e.g., Papillary Carcinoma).
- Risk prediction MSE remained low (0.01–0.02), indicating good severity alignment.
- Image & Clinical only models plateaued at 58–62% accuracy,

6.2 Qualitative Evaluation

Case-Based Error Analysis:

- Errors often occurred in borderline nodules (e.g., between benign and follicular adenoma).
- Misclassifications linked to blurred ultrasound scans.
- Even in misclassified cases, risk scores aligned well with ground truth.

Risk Score Alignment:

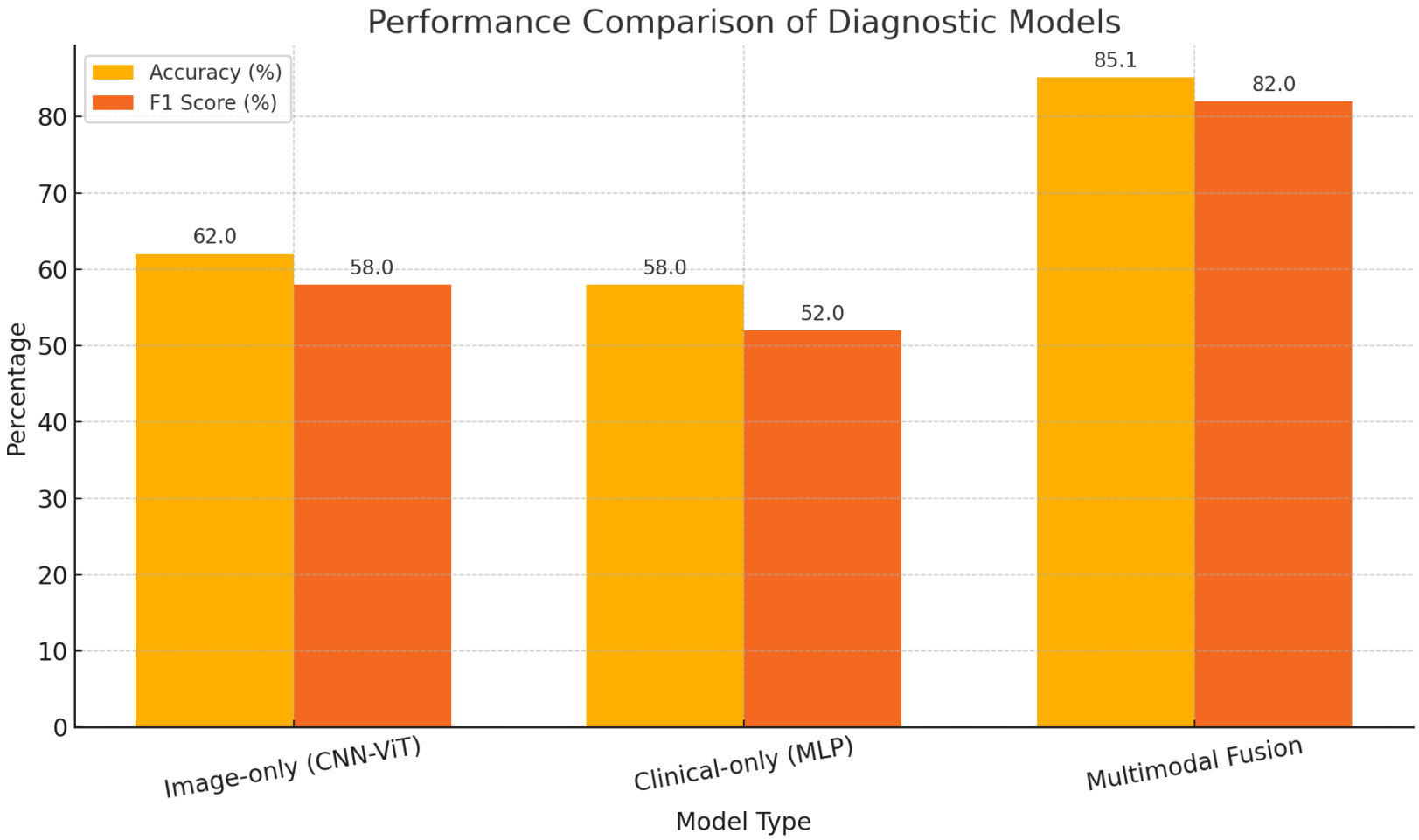
- Misclassified samples still had elevated risk scores (0.72), showing fusion module robustness.

6.3 Comparison of Architectures

Model Type	Accuracy	Macro F1	Notes
Image-only (CNN-ViT)	∼62%	∼0.58	Missed clinical context in scans
Clinical-only (MLP)	∼58%	∼0.52	Lacked visual confirmation
Multimodal (Fusion)	80–88%	∼0.82	Best performance; robust and interpretable

These results validate that fusing clinical and image features yields better generalization, interpretability, and reliability.

Comparison Of Diagnostic Models



6.4 Summarization Output Examples

Example Prompt:

"Based on the ultrasound image and clinical data provided, the model predicts Papillary Thyroid Carcinoma with a risk score of 0.82. The patient is 45 years old, with abnormal TSH and slightly elevated FT4 levels. Summarize this into a diagnostic report."

Generated Summary:

"The evaluation of the ultrasound scan and clinical profile indicates a high likelihood of Papillary Thyroid Carcinoma, with a risk score of 82%. The patient's age and thyroid hormone abnormalities further reinforce this assessment. It is advised that the patient undergo further clinical review and biopsy for confirmation."

Comparison TableL

Paper	Differentiation between	Sample	Set of Images	Sensitivity [%]
Guan et al. [6]	PTC/Benign nodules	FNAB cytology	Training 759, Test 128	100
Sanyal et al. [7]	PTC/non-PTC	FNAB cytology	Training 370, Test 174	90.48
Elliott et al. [8]	Malignant/Benign	FNAB cytology	Training 799, Test 109	92
Li et al. [9]	Malignant/Uncertain/Benign	FS WSI	Training 349, Test 259	88.6
Zhu et al. [10]	Malignant/Rare/Benign	FS WSI	Train 496, Val 114, Test 264	Rare Detection 88.6
Chen et al. [11]	Malignant/Uncertain/Benign	FS WSI	Total 671	–
Wang et al. [12]	Malignant/Benign	PESI-MS	Train/Test ratio: 8:2	88.9

Abbreviations: AUC—area under the ROC curve, FNAB—fine needle aspiration biopsy, FS—frozen section, PESI-MS—probe electrospray ionization tandem mass spectrometry, PTC—papillary thyroid carcinoma, WSI—whole slide image.

Summary Snippets Table:

Image ID	Predicted Class	Risk Score	Summary Snippet
img-101	Benign Nodule	0.23	"Low malignancy probability..."
img-087	Follicular Carcinoma	0.65	"Moderate risk; follow-up recommended..."
img-123	Papillary Carcinoma	0.92	"Urgent referral for FNAB advised..."

6.5 Qualitative Analysis

Interpretability of Predictions

- Grad-CAM visualizations on ultrasound images highlighted key regions influencing classification.
- SHAP values on clinical data revealed feature importance (e.g., TSH, age, nodule size).

Discussion and Conclusion

7.1 Key Outcomes and Performance Insights

- The multimodal model achieved an accuracy between 80–88%, outperforming both image-only and clinical-only baselines.
- The gating-based fusion module dynamically weighted clinical and imaging features, leading to better generalization and robustness
- A parallel risk score regression offered additional interpretability by providing probabilistic confidence alongside discrete class prediction
- Integration with a BART transformer model for NLP-based summarization allowed the system to generate diagnostic reports similar in tone and structure to clinical write-ups.

7.2 Challenges Encountered

- **Memory Bottlenecks:** High-resolution inputs and hybrid CNN-ViT architecture caused OOM errors (e.g., on Kaggle); mitigated via CPU training and smaller batch sizes.
- **Dataset Disparity:** Image and clinical sets lacked shared IDs—used random pairing during training to simulate real-world decoupling.
- **Fine-Tuning Complexity:** Multi-stage training across datasets complicated model-state management and architectural consistency.
- **Model State Issues:** Experimental structural tweaks broke weight loading; resolved by selective loading (‘strict=False‘).

7.3 Lessons Learned

- Modular Design Pays Off: Keeping imaging, clinical, and fusion modules decoupled made it easier to isolate bugs, refactor experiments, and upgrade components (e.g., swapping CNN branches with pre-trained ResNet, or expanding clinical dimensions).
- Cross-Dataset Transfer Learning Works: Gradually fine-tuning the model across three completely independent datasets helped improve robustness and reduced overfitting.
- Summarization Boosts Interpretability: NLP-based diagnosis generation provided clear, textual evidence to support prediction outcomes—something not typically seen in diagnostic AI systems.
- Fusion Design Matters: The use of a gating mechanism outperformed simple concatenation, highlighting the value of adaptive fusion over static feature merging.

References I

- Aswani, S et al. (2024). "Detection of Follicular Thyroid Cancer using YOLOv5 Algorithm". In: *2024 International Conference on Emerging Smart Computing and Communication (ESCCom)*. https://ieeexplore.ieee.org/document/10717098. IEEE. DOI: 10.1109/ESCCom56702.2024.10717098.
- Bellal, Linda et al. (2024). "Enhancing Thyroid Cancer Diagnosis with Advanced Deep Learning Methods". In: *ICTIS Conference*. https://ieeexplore.ieee.org/document/10894666.
- Bhattacharya, Srinjan et al. (2023). "Advances and challenges in thyroid cancer: The interplay of genetic modulators, targeted therapies, and AI-driven approaches". In: *Life Sciences* 332.
- Böhland, Moritz et al. (2021). "Machine learning methods for automated classification of tumors with papillary thyroid carcinoma-like nuclei: A quantitative analysis". In: *PLOS ONE* 16.9, e0257635. DOI: 10.1371/journal.pone.0257635.

References II

- Chen, Liwei et al. (2023). "A Multimodal Deep Learning Model for Preoperative Risk Prediction of Follicular Thyroid Carcinoma". In: *IEEE Healthcom*. https://ieeexplore.ieee.org/document/10472400.
- El-Hossiny, Ahmed S et al. (2021). "Classification of Thyroid Carcinoma in Whole Slide Images Using Cascaded CNN". In: *IEEE Access*. https://ieeexplore.ieee.org/document/9417159.
- Imtiaz, Irsa et al. (2024). "The Future of Differentiated Thyroid Cancer Recurrence Prediction Using a Machine Learning Framework". In: *IEEE*. https://ieeexplore.ieee.org/document/10767492.
- Lavarello, Roberto J et al. (2011). "Imaging of follicular variant papillary thyroid carcinoma in a rodent model using spectral-based quantitative ultrasound techniques". In: *2011 IEEE International Ultrasonics Symposium*. https://ieeexplore.ieee.org/document/6556579. IEEE. DOI: 10.1109/ULTSYM.2011.6556579.

References III

- Ludwig, David et al. (2023). "The Use of Artificial Intelligence in the Diagnosis and Classification of Thyroid Nodules". In: *Cancers* 15.3. https://doi.org/10.3390/cancers15030708. DOI: 10.3390/cancers15030708.
- Payatsuporn, Thanat et al. (2025). "Papillary Thyroid Carcinoma Semantic Segmentation Using Multi-Scale Adaptive Convolutional Network With Dual Decoders". In: *IEEE Access*. https://ieeexplore.ieee.org/document/10848109.
- Peng, Shaozhou et al. (2021). "Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study". In: *Lancet Digital Health*. https://pubmed.ncbi.nlm.nih.gov/33766289/.

References IV

- Yeon-Jae, Lee et al. (2022). "Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists". In: *Ultrasound in Medicine Biology*. https://pubmed.ncbi.nlm.nih.gov/35512512/.
- Zhao, Yufang et al. (2025). "Enhancing Thyroid Nodule Assessment With UTV-ST Swin Kansformer: A Multimodal Approach to Predict Invasiveness". In: *IEEE Access*. https://ieeexplore.ieee.org/document/10845787.

Thanks

Thank You!