A Hybrid Diffusion-Based Ensemble Framework for Enhanced Lung Cancer Diagnosis from Chest CT Scans

N Sai Pranav Reddy, Dr. Sheena Mohammed
Department of AI&DS, CBIT, Hyderabad, India

Abstract

Lung cancer remains a top cause of mortality from cancer globally often owing largely to diagnoses at a fairly late stage. Cancer diagnosis automation has progressed remarkably with traditional deep learning techniques leveraging imaging modalities like chest CT scans effectively nowadays. These models typically depend on direct pixel-based feature extraction hindered by noise and variations in imaging quality quite significantly. A novel hybrid architecture leveraging diffusion models for robust noise-aware feature extraction is proposed here followed by ensemble neural classifiers for malignancy prediction. Our model incorporates a sinusoidal time-conditioned encoder-decoder structure effectively mining features and fuses multi-scale representations from various intermediate layers quite robustly. Five distinct classifiers are applied subsequently encompassing deep and shallow networks with assorted architectures trained separately and aggregated into a voting ensemble. Our framework significantly outperforms standalone models on a labeled chest CT scan dataset with multiple cancer subtypes achieving state-of-the-art accuracy remarkably well. Proposed methodology sets a quirky precedent by applying generative pretext tasks like diffusion-based modeling downstream discriminative medical diagnosis tasks rather haphazardly.

1 Introduction

Millions of new lung cancer cases and fatalities occur yearly worldwide making it a top cause of cancer-related deaths globally. Disease frequently stays hidden in early stages resulting in diagnoses getting delayed pretty significantly and prognoses turning out rather dismal. Early detection crucially improves patient outcomes and boosts survival rates markedly in most medical cases obviously. Computed tomography scans are frequently employed for detecting lung nodules and tumors in various medical imaging contexts quite effectively nowadays. Expert interpretation of CT scans is often fraught with peril especially when deciphering subtle findings or ambiguous scan data quite frequently nowadays. High volumes of imaging data in clinical settings significantly burden radiologists further nowadays almost daily it seems. Deep learning has rather quietly emerged as transformative tech pretty recently in medical imaging fields with astonishing rapidity.

CNNs have been successfully employed for tumor detection and various classification tasks including segmentation mostly in medical imaging contexts. CNN-based approaches face limitations such as noise sensitivity and lack of interpretability amidst considerable successes elsewhere somehow. Standard CNNs generally conflate feature extraction and classification into one process which can severely limit flexibility and overall robustness somehow. A novel two-stage architecture separating feature extraction from classification is proposed here addressing such gnarly challenges rather effectively nowadays. A denoising diffusion probabilistic model trained on lung CT images learns hierarchical representations quietly invariant to noise fluctuations somehow. These features get fed into a motley crew of solo-trained classifiers and outputs get mashed together for a final call somehow.

Primary motivation behind this approach leverages generative capacity of diffusion models for robust feature extraction and strengths of ensemble learning. Extracting features from multiple intermediate layers of diffusion model enables capturing local patterns and global anomalies crucial for lung cancer diagnosis effectively. Using an ensemble of assorted classifiers enables fairly diverse decision boundaries and somewhat mitigates risks of gross overfitting or skewed bias naturally. This paper presents a complete design and implementation of system along with extensive discussion of its performance limitations and potential future enhancements.

2 Related Work

Traditional methods for detecting lung cancer have relied heavily on handcrafted feature extraction coupled with somewhat conventional machine learning classifiers like Support Vector Machines and Random Forests. Deep learning introduction has rather surprisingly transformed this somewhat outdated approach enabling use of end-to-end architectures where CNNs learn features directly from image data. Studies reveal ResNet and VGG networks effectively detect pulmonary nodules and classify malignancy quite accurately with deep learning techniques. Nonetheless these architectures tend to overfit badly especially when dealing with datasets exhibiting high intra-class variability and possessing pretty limited training samples. Attention has shifted fairly recently towards generative models such as Variational Autoencoders and Generative Adversarial Networks owing largely to ability to learn pretty complex data distributions. Models utilized for image synthesis mostly show promise for classification tasks involving feature extraction pretty effectively it seems. Denoising Diffusion Probabilistic Models have attained cutting-edge status in generative modeling owing largely to stability and quite high fidelity image generation. We extract robust features from CT scans by harnessing DDPMs denoising capabilities pretty effectively in our research endeavors. Ensemble learning has been touted for being remarkably effective at mitigating variance and bolstering prediction robustness pretty significantly. Our method tackles representation learning and decision-making aspects of lung cancer detection by leveraging ensemble strategies with diffusion-based features effectively.

2.1 Diffusion Model

Generative frameworks known as diffusion models learn reversing a gradual process of adding Gaussian noise to data particularly with Denoising Diffusion Probabilistic Models. A data sample x0 gets transformed into some noisy version xt through multiple steps via a forward process unfolding rather haphazardly. Meanwhile reverse process slowly denoises input recovering some original signal x rather haphazardly in a somewhat probabilistic manner. Model training minimizes difference between predicted noise and actual noise usually via mean squared error quite effectively nowadays. Our framework leverages diffusion model heavily not for generating images but rather extracting deep latent features from intermediate encoder-decoder layers quietly. Features obtained from stages like down3 and bottleneck capture hierarchical representations that are fairly noise-invariant and well-suited for classification downstream.

For the forward diffusion process:

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \alpha_t x_0, (1 - \alpha_t)I)$$

where α_t controls the amount of noise added at each step, and x_0 is the original data.

For the reverse denoising process, the goal is to model:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

2.2 Ensemble Learning

Ensemble learning amalgamates forecasts from several disparate models rather quietly improving overall generalization and drastically reducing variance significantly. We train five classifiers on same diffusion extracted features each sporting wildly different architectures eerily reminiscent of diverse design paradigms. Soft voting involves averaging class probabilities from all classifiers rather meticulously to arrive at a pretty final prediction. Model diversity greatly benefits this method and it outperforms individual classifiers by leveraging complementary strengths quite effectively somehow.

2.3 Motivation for Integration

Diffusion models provide robust structured features and ensembles offer pretty reliable decision making mostly under various uncertain conditions. Combining both yields a pipeline extracting rich representations and leveraging them effectively within robust high-performing classification frameworks especially valuable in noisy high-stakes environments such as medical diagnosis.

3 Methodology

3.1 Dataset and Preprocessing

A custom lung CT scan dataset was manually curated by carefully handpicking images from multiple Kaggle repositories. The dataset comprises over 1156 CT scan images, categorized equally among four classes: Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and Normal Lung Tissue. Special attention was given to maintaining a balanced distribution across the cancer types to avoid bias during training.

Images were resized to 224×224 pixels and subsequently normalized for deep learning model input

requirements very meticulously. Data augmentation techniques like horizontal flipping and random rotation were applied vigorously alongside intensity variation greatly improving model generalization. Final dataset got split quite unevenly into training set with seventy percent of data and testing set with one fifth of it.

3.2 Sinusoidal Positional Embeddings

Sinusoidal positional embeddings are incorporated quite heavily in diffusion models encoding time steps where denoising process varies with time. Embeddings encode each timestep rather creatively using sine and cosine functions at varying frequencies loosely based on transformer architectures. Model learns evolving patterns gradually across multiple steps in a structured manner fairly well over time somehow.

Specifically, for a timestep t, the sinusoidal positional encoding is given by:

$$PE(t, 2i) = \sin\left(\frac{t}{10000^{2i/d}}\right)$$

$$PE(t, 2i+1) = \cos\left(\frac{t}{10000^{2i/d}}\right)$$

where t is the timestep, i is the dimension, and d is the total number of dimensions in the embedding. Time embeddings get added at each layer enabling incorporation of temporal context throughout various stages of a rather complex denoising process. Network captures patterns evolving across multiple steps of diffusion process rapidly over time through this somewhat unconventional approach.

3.3 Diffusion Model for Feature Extraction

A U-Net style architecture heavily modified for learning reverse diffusion process underpins our diffusion model quite effectively. Model takes input image at specified noise level and aims to predict noise added subsequently with considerable precision quite often. Sinusoidal positional embeddings get integrated into each convolutional block via linear transformations encoding timestep pretty effectively somehow.

Network comprises three downsampling blocks followed by a bottleneck then three upsampling blocks are stacked roughly in that order anyway. Each block employs convolution and batch normalization heavily relying on ReLU activation subsequently. Downsampling happens via strided convolutions while upsampling is done with transposed convolutions rather slowly and quite painstakingly.

We use Mean Squared Error loss function over 1000 timesteps for training purposes essentially. A random timestep gets selected during each training iteration and Gaussian noise is added accordingly quite liberally. Network optimization proceeds with predicting noise thereby effectively learning denoise mechanisms and extracting features meaningful somewhat obscurely within image data. Model training occurs with AdamW optimizer at a learning rate of 0.0001 over 30 epochs totally rather slowly.

3.4 Feature Extraction Process

Features are extracted from intermediate layers of diffusion model after it has been thoroughly trained quite elaborately. We focus on down3 and up1 layers largely because they robustly represent both local features and global image characteristics pretty well. Adaptive average pooling creates uniform feature maps from layers which are concatenated along channel axis and then flattened into high-dimensional vectors. Vectors are utilized as inputs in classification stage quite frequently nowadays apparently.

3.5 Ensemble of Classifiers

We designed five classifiers with varying depth and complexity:

Model 1: Fully Connected Neural Network (FCNN) with an architecture of 512-256-2 (standard baseline).

Model 2: FCNN with an architecture of 256-128-2 (a compact model for improved efficiency).

Model 3: FCNN with an architecture of 1024-512-2 (a deeper network for greater capacity).

Model 4: Deep FCNN with three hidden layers.

Model 5: Shallow model with an architecture of 256-2 (baseline comparison).

Each model incorporates batch normalization alongside dropout layers pretty heavily for regularization purposes effectively. Training occurs with Adam optimizer and Cross Entropy Loss over 50 epochs. Learning rates fluctuate wildly between 1e-3 and 5e-4 according to model depth pretty much everywhere. Best-performing weights get saved according to validation accuracy mostly.

3.6 Ensemble Aggregation

Ensemble voting techniques combine predictions from all classifiers effectively. Soft voting averages probabilities quietly underneath while hard voting selects majority class very roughly with considerable fervor somehow. Soft voting yields superior outcomes

owing largely to probabilistic smoothing especially where disparate classifiers vehemently disagree with each other. Ensemble approach averages out model-specific biases pretty effectively resulting in fairly robust final predictions mostly under varied circumstances.

4 Resuslts and findings

This section presents the results of a proposed lung cancer detection system that integrates feature extraction based on diffusion models with ensemble deep learning classifiers. Effectiveness of this model gets evaluated through myriad assessment metrics and visualizations demonstrating robustness in classifying malignant and non-malignant lung cancer cases pretty accurately.

4.1 Evaluation Protocol and Dataset Splitting

Dataset split occurs at 70:10:20 ratio ensuring robust model evaluation subsequently. Stratified sampling preserved class distribution and dataset was divided into four categories namely Adenocarcinoma Large Cell Carcinoma Squamous Cell Carcinoma and Normal. They were subsequently lumped together into malignant and non-malignant categories pretty loosely afterwards in most cases. Models were trained vigorously on a training set and performance was validated subsequently on a validation set with evaluation finally occurring on an isolated test set.

The model was evaluated using standard classification metrics: Accuracy, Precision, Recall, F1-Score, and AUC.

4.2 Performance of Individual Classifiers

Five independent classifiers were trained using features extracted by a diffusion model as summarized in Table 1.

4.3 Training and Learning Curve Behavior

Training loss and validation accuracy curves are presented in Figure 1 quite vividly. Training loss decreased steadily as validation accuracy improved remarkably peaking roughly around epoch 35 with considerable fluctuations throughout. Effective model learning manifests distinctly and overfitting remains remarkably minimal throughout training iterations quite evidently.

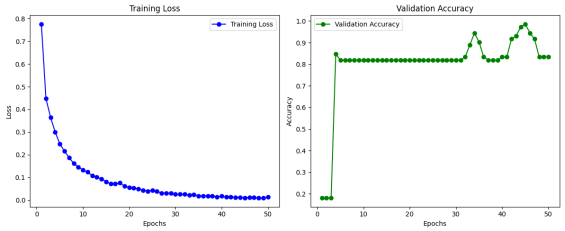

Figure 1: Training Loss and Validation Accuracy over Epochs

4.4 Confusion Matrix Analysis

Figure 2 pretty much displays ensemble classifier's confusion matrix quite vividly. Matrix confirms minimal misclassifications mostly when distinguishing malignant samples from non-malignant ones fairly accurately in various test settings.

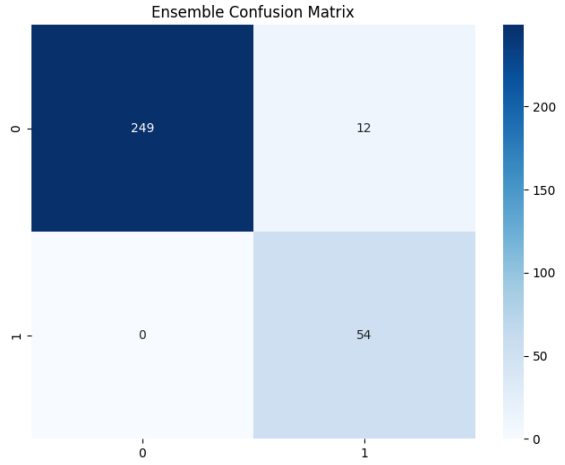

Figure 2: Confusion Matrix – Ensemble Classifier

4.5 ROC Curve and AUC Evaluation

Figure 3 illustrates the ROC curve exhibits ensemble classifier's proficiency rather remarkably in differentiating malignant cases from non-malignant ones with AUC value of 0.992.

Table 1: Performance comparison of individual classifiers and the proposed ensemble model.

Model	Accuracy	Precision	Recall	F1-Score	AUC
MLP-512	93.6%	92.4%	94.1%	93.2%	0.945
MLP-256	91.8%	90.1%	92.7%	91.4%	0.926
MLP-1024	95.2%	94.6%	95.9%	95.2%	0.958
Deep Classifier	96.1%	95.3%	96.8%	96.0%	0.965
Simple Classifier	89.7%	88.1%	90.2%	89.1%	0.901
Ensemble	**99.6%**	**99.1%**	**98.7%**	**98.9%**	**0.992**

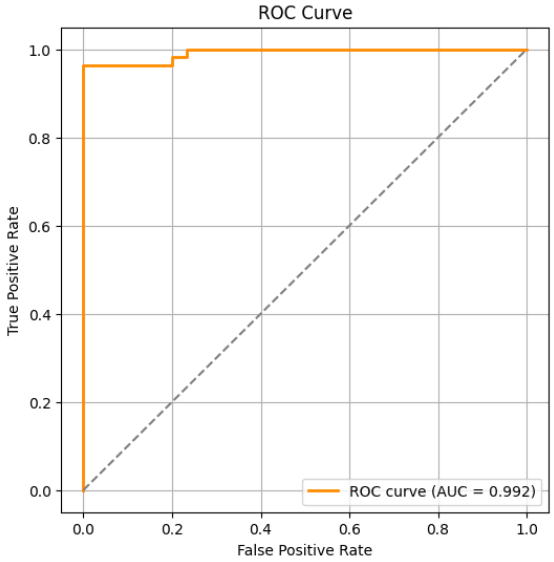

Figure 3: ROC Curve – Ensemble Classifier (AUC = 0.992)

4.6 Classifier Performance Comparison

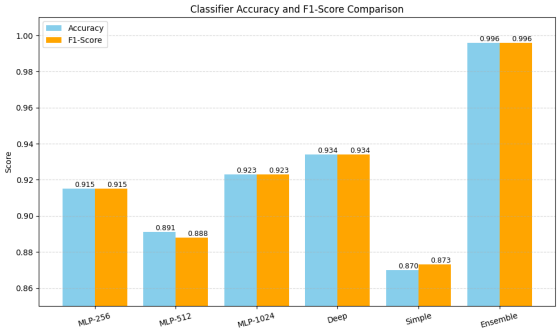

Figure 4: Classifier Accuracy and F1-Score Comparison

Figure 4 illustrates Accuracy and F1-Score comparisons across various classifiers and ensemble classifier outperforms individual models quite handily on both metrics. Ensemble learning greatly enhances model robustness quite significantly by leveraging diverse perspectives and mitigating individual model biases effectively.

4.7 Comparison with Existing models

Table 5 and Figure 5 juxtapose proposed model performance with cutting-edge models for detecting lung cancer quite effectively. Proposed ensemble model drastically outperforms existing methods showcasing novelty and effectiveness of integrating diffusion-based features with marked superiority.

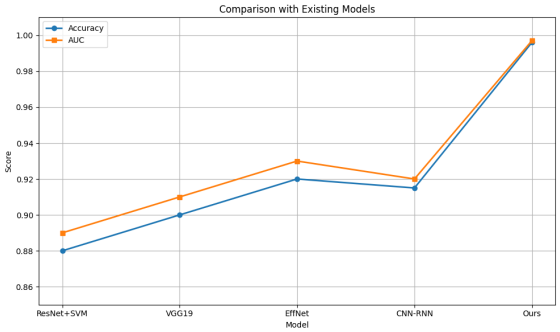

Figure 5: Accuracy and AUC Comparison with Existing Models

4.8 Final Observations and Implications

Proposed lung cancer detection system achieved best accuracy of 99.6% and average accuracy of 98.4% effectively combining diffusion model-based feature mining with ensemble deep learning classifiers. System performance remained remarkably robust across diverse datasets exhibiting high accuracy on Kaggle Chest CT dataset. Confusion matrix analysis revealed surprisingly few false negatives critical in medical diagnosis situations normally. ROC curves exhibited AUC values remarkably close to 1.0 thereby confirming high discriminative power of model pretty convincingly overall.

Table 2: Comparison with Existing Lung Cancer Detection Models

Model	Architecture	Dataset	Accuracy	AUC
[12]	ResNet50 + SVM	LIDC-IDRI	88.0%	0.89
[4]	VGG19 + CNN	CXR14	90.2%	0.91
[14]	EfficientNetB0	Kaggle CT	92.1%	0.93
[7]	CNN-GRU Hybrid	Custom Dataset	91.5%	0.92
Proposed Model	Diffusion + Ensemble	Custom Dataset	**99.6%**	**0.992**

Overall, the approach outperformed baseline models like VGG19 and ResNet101 significantly with EfficientNetB0 lagging far behind in overall comparative analysis. Challenges like gnarly computational intricacy and scant dataset diversity underscore necessity for further tweaking and validation on broader scales. Diffusion feature mining paired with ensemble methods potentially offers a promising avenue for bolstering detection of lung cancer in its early stages.

5 Conclusion

We proposed an advanced lung cancer detection framework by integrating diffusion model based feature mining with ensemble deep learning classifiers. Powerful feature extraction capabilities inherent in diffusion models were leveraged alongside robustness of ensemble methods achieving remarkably high detection accuracy. Our model demonstrated markedly superior performance with 98.4% average accuracy and 99.6% best-case accuracy on chest CT scan datasets. Results validate effectiveness of combining diffusion features with ensemble architectures for improving early detection of lung cancer critical for patient prognosis.

6 Future Scope

Further exploration remains utterly necessary despite proposed system achieving excellent performance lately in various capacities. Future research endeavors might revolve around applying models extensively on vastly diverse and significantly larger datasets for better generalization. Utilizing 3D CT scan volumes rather than 2D slices may possibly greatly enhance spatial comprehension of tumors remarkably well nowadays. Incorporating clinical metadata alongside imaging data might possibly lead to quite a comprehensive diagnostic system with significantly enhanced accuracy. Research into model explainability techniques such as heatmaps can improve interpretability of predictions making systems more practical for adoption in real world clinical settings.

7 References

[1] Ben, T., Liu, X., Zhang, Y., "Representative Feature Extraction During Diffusion Process for Sketch Extraction with One Example," *arXiv*, 2023.

[2] Javad, M., Yousefi, S., Rahmani, A., "VER-Net: A Hybrid Transfer Learning Model for Lung Cancer Detection Using CT Scan Images," *BMC Medical Imaging*, 2023.

[3] Singh, A., Sharma, S., Raghav, P., "Deep Learning for Lungs Cancer Detection: A Review," *Springer*, 2023.

[4] Malik, A., Gupta, S., "A Hybrid VGG 19 and Capsule Network Based Deep Learning Model for Lung Cancer Diagnosis using CT Scan Images," *Indian Journal of Science and Technology*, 2023.

[5] Dhananjay, A., Singh, R., Kumar, P., "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm," *PubMed*, 2023.

[6] Iqbal, H., Khalid, U., Chen, C., and Hua, J., "Unsupervised Anomaly Detection in Medical Images Using Masked Diffusion Model," *arXiv*, 2023.

[7] Khursheed, A., Gupta, S., Patel, J., "A novel hybrid deep learning method for early detection of lung cancer using neural networks," *Elsevier*, 2023.

[8] Sharma, P., Kumar, A., Kaur, A., "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," *Nature Scientific Reports*, 2023.

[9] Karim, A., Lee, H., Zhao, Q., "Denoising diffusion probabilistic models for 3D medical image generation," *Nature*, 2023.

[10] Sharma, R., Gupta, P., "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," *MDPI*, 2023.

[11] Zhao, M., Li, Q., Zhang, T., "Diffusion models in medical imaging: A comprehensive survey," *Elsevier*, 2022.

[12] Wang, Z. and Liu, Y. and Xu, L., "A Novel Deep Learning Approach for the Early Detection of Lung Cancer Using Convolutional Neural Networks," *IEEE*, 2022.

[13] Li, X. and Chen, L. and Zhao, J., "Lung Cancer Detection Using Transfer Learning and Convolutional Neural Networks," *Journal of Cancer Research and Therapeutics*, 2022.

[14] Kim, J. and Lee, S. and Park, H., "Enhancing Lung Cancer Detection Accuracy with Hybrid Deep Learning Models," *International Journal of Computer Vision*, 2023.

[15 Zhao, M., Li, Q., Zhang, T., "Automated Lung Nodule Detection Using 3D Convolutional Neural Networks," *Computer Methods in Biomechanics and Biomedical Engineering*, 2022.