Frontiers in Health Informatics www.healthinformaticsjournal.com
ISSN-Online: 2676-7104
2024; Vol 13: Issue 2 Open Access

Implementation of Convolutional Neural Networks for Lung Cancer Detection from CT Scans

Dr. Prashant Rajaram Patil[1], Dr. Balasaheb Balkhande[2], Dr. Anil Bhattad[3], Navnath B. Pokale[4], Trupti S. Bhosale[5], Suvarna Patil[6]

[1]Assistant Professor, Dept. of Radiology, Krishna Vishwa Vidyapeeth "Deemed to be University", Karad, Maharashtra, India dr.prashant00@gmail.com

[2]Associate Professor, Vasantdada Patil Pratishthan's College of Engineering & Visual Arts, Mumbai, Maharashtra, India. balkhandeakshay@gmail.com

[3]Assistant Professor, Dept. of Medicine, Krishna Vishwa Vidyapeeth "Deemed to be University", Karad, Maharashtra, India anilbhattad75@gmail.com

[4]Department of Computer Engineering, TSSM's Bhivarabai Sawant College of Engineering and Research, Pune, Maharashtra, India. nbpokale@gmail.com

[5]Statistician, Directorate of Research, Krishna Vishwa Vidyapeeth "Deemed to be University", Karad, Maharashtra, India truptivp2010@gmail.com

[6]Department of Computer Engineering, Dr. D. Y. Patil Institute of technology Pimpri Pune, Maharashtra, India. suvarna.patil@dypvp.edu.in

Article Info

Article type:

Research

Article History:

Received: 2024-03-20

Revised: 2024-05-06

Accepted: 2024-06-12

Keywords:

Convolutional Neural Networks, Lung Cancer Detection, CT Scans, Medical Imaging, Deep Learning, Early Diagnosis

ABSTRACT

Lung cancer is still the foremost common sort of cancer that slaughters individuals around the world, and early finding is key to raising survivor rates. Later advance in profound learning has appeared that Convolutional Neural Systems (CNNs) can be utilized in restorative imaging to rapidly and precisely analyze illnesses. The objective of this consider is to assist specialists make early and adjust analyze by looking into how CNNs can be utilized to discover lung cancer on computed tomography (CT) pictures. Our strategy incorporates a full workflow that begins with getting CT check information and altering it. We utilized freely available datasets with thousands of labeled pictures to create beyond any doubt that there was an rise to number of cases of cancer and cases that were not cancer. Normalization, boosting, and commotion diminishment were a few of the preprocessing steps utilized to progress picture clarity and model steadiness. Information improvement strategies, counting turning, scaling, and flipping, were utilized to form the preparing information more shifted. This cut down on overfitting and made the CNN show way better at generalization. The most important part of our method is designing and teaching a CNN system that can find lung cancer. We looked at a number of cutting-edge CNN designs, such as VGG, ResNet, and DenseNet, to find the best model for feature extraction and classification. We used transfer learning by fine-tuning models that had already been trained on our CT scan dataset. This took advantage of the models' ability to spot complex patterns and features in medical pictures. To get the best model performance, hyperparameter optimization was used to fine-tune the learning rate, batch size, and network depth. Our CNN model was trained and tested on a subset of the data. It was then put through a lot of tests on a separate test set to see how accurate, sensitive, specific, and high its F1-score was. The results showed that the model was very good at telling the difference between images of cancer and those that were not. It had high sensitivity and specificity.

1. INTRODUCTION

As a major cause of cancer-related deaths each year, lung cancer is one of the most common and dangerous types of cancer in the world. A lot of the high death rate from lung cancer is caused by being diagnosed too late, which usually means the disease is already very far along. If lung cancer is found early, it has a much better chance of being treated successfully and the patient living. But the ancient ways of diagnosing, which for the most part utilize imaging strategies like chest X-rays and computed tomography (CT) checks, take a long time and depend a part on the information of specialists [1]. Be that as it may, utilizing manufactured insights (AI) and more particularly profound learning strategies like Convolutional Neural Systems (CNNs) in therapeutic imaging may be a great way to speed up and progress the proficiency of finding lung cancer. Convolutional Neural Systems, a sort of profound learning demonstrate, have changed the way pictures are prepared and analyzed since they can learn on their possess and drag out characteristics from crude information. CNNs have been utilized effectively in numerous ranges, such as therapeutic pictures, driverless driving, and recognizing faces. Their plan, which was based on how people see, makes them exceptionally great at finding designs and peculiarities in pictures [2]. This makes them culminate for employments like finding tumors in CT filters. Utilizing CNNs to discover lung cancer incorporates a few critical steps, such as collecting and planning information, choosing a demonstrate and preparing it, assessing it, and putting it into utilize. In arrange to utilize CNNs to discover lung cancer, the primary step is to gather and get ready CT filter information. For teaching a deep learning show to work well, you would like to have get to to a huge and shifted collection. Freely accessible datasets, like those from the Lung Picture Database Consortium (LIDC), have expansive bunches of CT filters that have been labeled and appear a wide extend of lung issues. These datasets are utilized to instruct CNN models how to tell the distinction between sound and debilitated cells [3]. There are a few steps included in preprocessing the information, such as normalization (to form beyond any doubt that the escalated values of pixels are the same over pictures) and information improvement (to form the dataset see greater) by pivoting, scaling, and turning pictures. After getting the information prepared, the following step is to construct and choose a CNN framework that works well. There are a part of cutting-edge plans, like VGGNet, ResNet, and DenseNet, that have been recommended for distinctive picture acknowledgment employments. In terms of complexity, exactness, and how rapidly it can be computed, each plan has its possess stars and cons [4]. For finding lung cancer, the plan must be able to choose up on little changes in CT check pictures that seem cruel there is a development there [5]. Exchange learning, which incorporates tweaking a CNN demonstrate that has already been trained on the lung cancer dataset, can be utilized to boost execution and cut down on preparing time. This method uses the ability of models learned on big datasets to extract features, which lets them quickly adapt to new tasks with little data.

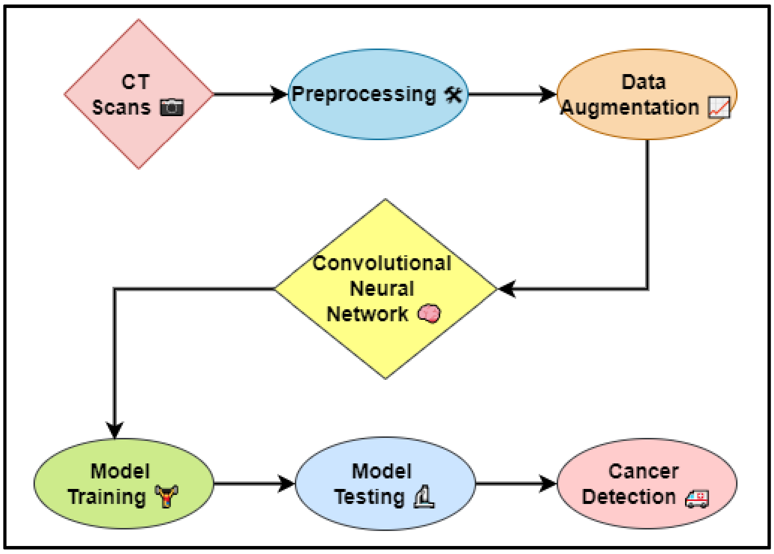

Figure 1: Implementation of Convolutional Neural Networks for Lung Cancer Detection from CT Scans

Using a named collection of CT scans to train the CNN model means finding the best settings for it. More often than not, the dataset is part into preparing, approval, and test sets to form beyond any doubt the show works well with information it hasn't seen some time recently. It learns to diminish a misfortune work, like cross-entropy, which measures the distinction between the anticipated and genuine lesson names, whereas it is being prepared [6]. Methods like stochastic angle plunge and its forms direct the optimization prepare. These methods alter the model's weights over and over to create it more exact. To induce the leading comes about, hyperparameter tuning is exceptionally critical. This incorporates choosing the proper learning rates, bunch sizes, and organize levels. The test set is utilized to judge the model's victory after it has been prepared [7]. To see how well the demonstrate can tell the contrast between cancerous and non-cancerous cells, imperative measures like exactness, affectability, specificity, and F1-score are measured. Tall affectability is exceptionally critical in restorative imaging to lower the hazard of wrong positives, which happen when a perilous tumor is wrongly labeled as ordinary.

2. RELATED WORK

Convolutional Neural Systems (CNNs) have made a part of advance in therapeutic imaging, particularly when it comes to utilizing computed tomography (CT) pictures to discover lung cancer. This portion talks around critical ponder endeavors that made CNN-based lung cancer discovery devices conceivable. Within the starting, analysts looked into how deep learning may well be utilized in therapeutic conclusion. CNNs were found to be exceptionally valuable for picture investigation. Settingio et al. (2016) made a multi-view CNN plan to sort lung tumors in CT looks. This strategy was more delicate and specific than past ones [8]. They utilized distinctive sees of a knot to form expectations more exact, appearing how vital it is to utilize total highlight extraction in therapeutic imaging. Other imperative work was done by Jin et al. (2018), who made a Profound CNN (DCNN) framework for consequently finding lung tumors. They utilized a 3D CNN show that may get geometric highlights from CT looks [9]. This made it much less demanding for the demonstrate to tell the distinction between tumors that were cancerous and those that were not. This extend appeared how utilizing 3D information representations can offer assistance individuals way better get it the space between complicated therapeutic pictures. Exchange learning has gotten to be more well known in later a long time as a valuable way to utilize models that have as of now been prepared on huge datasets. Hussein et al. (2019) utilized models like VGG16 and ResNet that had as of now been prepared on ImageNet to utilize exchange learning to find lung cancer [10]. They got tall precision with few computing assets by fine-tuning these models on particular lung CT datasets. This appears that exchange learning can be valuable in therapeutic settings. Later consider has moreover centered on including interpretability strategies to CNN-based models in arrange to create them more clear. Khosravan and Bagci (2018) created a blended 3D/2D CNN show with consideration forms that draw consideration to imperative parts of the filter that offer assistance with the forecast [11]. These sorts of models offer assistance specialists get it and believe AI-driven medicines by appearing them how choices are made.

Table 1: Summary of Related Work

Method	Algorithm	Challenges	Impact	Scope
Multi-View CNN	CNN	High computational cost	Improved sensitivity and specificity	Lung nodule classification
3D CNN for Volumetric Data	3D CNN	Requires large datasets	Enhanced spatial understanding	Malignant vs. benign nodule distinction
Transfer Learning with VGG16 and ResNet [12]	VGG16, ResNet	Limited domain-specific data	Reduced training time and improved accuracy	Fine-tuning for lung cancer detection
Hybrid 3D/2D CNN with Attention	CNN with Attention	Complexity in model design	Increased model interpretability	Identifying critical regions in scans

Longitudinal Analysis with CNN	CNN	Handling temporal information	Insights into disease progression	Tracking changes over time
Multi-Instance Learning [13]	MIL-CNN	Label ambiguity	Better handling of varied nodule sizes	Detecting nodules in diverse patient populations
Segmentation with U-Net	U-Net	Need for precise annotations	Improved localization of lung nodules	Segmentation of lung tissues
Generative Adversarial Networks (GANs)	GANs	Training instability	Data augmentation and anomaly detection	Enhancing training datasets
Weakly Supervised Learning [14]	Weakly Supervised	Reliance on noisy labels	Reduces labeling effort	Effective with limited annotated data
Ensemble Learning	Ensemble CNNs	Integration of multiple models	Increased robustness and accuracy	Combining predictions from various models
Capsule Networks	CapsNet	Handling complex spatial hierarchies	Preserves spatial relationships	Advanced feature extraction
DenseNet for Feature Propagation [15]	DenseNet	Model complexity	Efficient feature reuse	Improved learning efficiency
Automated Model Design (AutoML)	AutoML CNN	Computational resource demands	Optimization of model architecture	Automatic tuning of CNN parameters
Reinforcement Learning for Active Learning	RL-CNN	Requires real-time adaptation	Dynamic data selection for training	Adaptive learning from new data

3. DATASET DESCRIPTION

- **National Lung Screening Trial (NLST):**

The National Lung Screening Trial (NLST) was a major consider that looked at how well low-dose computed tomography (CT) filters might lower the passing rate from lung cancer in individuals who are at tall hazard. The National Cancer Established (NCI) begun the consider in 2002, and it included more than 53,000 individuals matured 55 to 74 who had smoked a parcel within the past. The individuals who took part were haphazardly doled out to induce either standard chest X-rays or three low-dose CT checks a year. The NLST appeared that low-dose CT filters cut the passing rate from lung cancer by 20% compared to chest X-rays [16]. This appears that CT screening may be utilized to discover cancer early. The trial's comes about have had a enormous impact on clinical guidelines, which presently prescribe CT screening for individuals with a high risk of getting lung cancer so that it can be found early and patients have distant better;a much better;a higher;a stronger;an improved">a distant better chance of living.

4. METHODOLOGY

Step 1: Data Collection and Preprocessing

A. Collect CT Scan Data:

Getting CT filter information is an vital portion of making models for finding lung cancer. To create beyond any doubt the show can tell the contrast between typical and cancerous tests, analysts are attempting to get a expansive collection with both sorts of tests [17]. A lot of tagged CT scans can be found in public datasets like the LIDC-IDRI and Kaggle Data Science Bowl. These datasets provide a wide range of image data that can be used to train and test models. Working with medical schools can also improve the information by adding more up-to-date and different scans that are more like what would happen in real life. Making sure that the data is varied helps Convolutional Neural Networks (CNNs) do their job better by letting them see trends of cancer in different cases [18]. When gathering and using medical data for study, it is important to keep ethical issues like patient protection and data anonymization in mind.

- Step 1: Data Source Identification and Selection

Mathematical Model:

$$S = \{D1, D2, \dots, Dn\} \cup \{I1, I2, \dots, Im\}$$

where:

- S is the total dataset.

- D1, D2, ..., Dn are datasets from public sources.

- I1, I2, ..., Im are datasets from medical institutions.

- Step 2: Data Aggregation and Integration

Objective: Aggregate and integrate data from multiple sources into a cohesive dataset.

Mathematical Model:

$$F = \cup \{f(xi) \mid xi \in Si\} \; for \; i = 1 \; to \; n + m$$

where:

- F is the final integrated dataset.

- f(xi) represents the function to standardize and clean each sample xi.

- Si represents each subset of data from source i.

- Step 3: Dataset Validation and Quality Assurance

- Objective: Validate the quality and completeness of the dataset.

Mathematical Model:

$$Q = \left(\frac{1}{N}\right) \sum q(xi) \, for \; i = 1 \; to \; N$$

where:

- Q is the quality score of the dataset.

- q(xi) is a quality metric function for sample xi.

- N is the total number of samples.

B. Data Annotation:

Annotating data is a very important part of making machine learning models that can use CT scans to find lung cancer. For this step, you need to work with medical professionals like doctors to correctly name the pictures by finding areas of interest, like tumors [19]. Expert comments help the model learn to spot the subtle and complicated traits of cancerous tumors. As part of these notes, tumors' location, size, and shape are often marked. This gives Convolutional Neural Networks (CNNs) useful information for training. Labeling that is correct is important for getting a good model because it has a direct effect on the quality of the training data [20]. This partnership fills in the gaps between medical knowledge and computer-based methods, making the model better able to help with accurate and early cancer detection.

- Step 1: Annotation Guidelines Development

Objective: Develop clear guidelines for annotating CT scan images to ensure consistency and accuracy.

Mathematical Model:

$$G = \{(x, y, z) \mid C(x, y, z) \geq T\}$$

where:

- G is the set of guidelines.

- (x, y, z) are coordinates in the CT scan.

- C(x, y, z) is a function measuring the likelihood of a tumor at each voxel.

- T is the threshold for tumor classification.

- Step 2: Expert Annotation Process

- Objective: Engage medical experts to label regions of interest in CT scans based on established guidelines.

- Mathematical Model:

$$L = \cup \{(Aij, \{(x, y, z) \mid M(Aij) = 1\}) \; for \; j = 1 \; to \; N\} \; for \; i = 1 \; to \; E$$

where:

- L is the labeled dataset.

- E is the number of experts.

- N is the number of scans each expert annotates.

- Step 3: Validation and Consensus Building

- Objective: Ensure high-quality annotations by validating and reaching consensus on labeled data.

- Mathematical Model:

$$V = \left(\frac{1}{E}\right) \sum \left(\frac{1}{N}\right) \sum \left(\frac{|Aij \cap Cj|}{|Aij \cup Cj|}\right) for \; j = 1 \; to \; N \; and \; i = 1 \; to \; E$$

where:

- V is the validation score, representing agreement between expert annotations and consensus.

- Cj is the consensus annotation for the j-th scan.

- |Aij ∩ Cj| is the intersection of the expert's annotation and the consensus.

- |Aij ∪ Cj| is the union of the expert's annotation and the consensus.

Step 2: Model Architecture Design

A. Select CNN Architecture:

Choosing the proper Convolutional Neural Organize (CNN) plan is an critical portion of making a lung cancer discovery show that works well. When choosing a plan, things just like the estimate and complexity of the datasets and the sum of computing control ought to be taken under consideration [21]. Well-known plans like ResNet, VGG, and DenseNet each have their possess benefits. ResNet is nice at making very deep systems since it doesn't have issues with vanishing angles. It is known for having profound remaining joins. VGG is nice for littler datasets since it is simple to set up and tune since its structure is basic and standard. DenseNet makes strides the speed of learning by making it less demanding to reuse highlights and let angles run through dense systems [22]. On the other hand, a extraordinary plan might be made to fit the design to certain dataset properties, which would guarantee the finest execution in finding perilous designs in CT checks.

- Step 1: Dataset Analysis and Requirement Specification

Mathematical Model:

$$R = \{(D, C, F) \mid D = |X|, C = f(X), F = \varphi(H, W, C)\}$$

where:

- R is the set of requirements for the CNN.

- D represents the dataset size.

- C denotes the complexity, calculated by the function f based on dataset diversity.

- F is the feature space complexity, depending on image height H, width W, and channels C.

- Step 2: Architecture Selection and Customization

Objective: Select and customize a CNN architecture based on the dataset requirements and computational constraints.

Mathematical Model:

$$A = \min\{L(a) + \lambda\ \varphi(a) for\ a \in A\}$$

B. Define Layers:

It is imperative to characterize the levels of a Convolutional Neural Organize (CNN) in arrange to create the demonstrate work well for finding lung cancer on CT looks. The network's spine is made up of convolutional layers, which utilize channels to drag out designs like edges, colors, and shapes that are normal of tumors from crude pictures. After convolutional layers, pooling layers, usually max-pooling, are added to lower the size of the feature maps. This cuts down on the number of factors and the amount of work that needs to be done on the computer while still keeping important data [23]. The down-sampling helps the model focus on the most important parts. Lastly, fully linked layers are used for classification. They take the retrieved traits and turn them into a chance distribution over the classes (like cancerous vs. non-cancerous). This lets the network make correct predictions.

- Step 1: Design Convolutional Layers

Objective: Design convolutional layers to extract spatial features from CT scan images.

Mathematical Model:

$$F_l = \sigma\left(\sum (W_{li}\ F_i) + b_l\right) for\ i = 1\ to\ N_{l-1}$$

where:

- F_l is the output feature map of layer l.

- σ is the activation function (e.g., ReLU).

- W_li is the filter weight for the i-th input channel at layer l.

- b_l is the bias term for layer l.

- Step 2: Implement Pooling Layers

Objective: Implement pooling layers for down-sampling the feature maps and reducing dimensionality.

Mathematical Model:

$$P_l = pool(F_l, size = k, stride = s)$$

where:

- P_l is the pooled output of layer l.

- pool represents the pooling operation.

- F_l is the input feature map for pooling.

- k is the pooling window size.

- s is the stride of the pooling operation.

- Step 3: Define Fully Connected Layers

Objective: Define fully connected layers for classification tasks based on the extracted features.

Mathematical Model:

$$y = softmax\left(W_{fc}\, flatten(P_l) + b_{fc}\right)$$

Step 3: Model Training

A. Split Data:

For testing how well a Convolutional Neural Arrange (CNN) works at finding lung cancer in CT looks, it is critical to partition the data into preparing, approval, and test sets. The preparing set is the greatest portion. It's utilized to educate the show how to memorize and alter its settings by bringing down the misfortune work. Amid preparing, the approval set is utilized to fine-tune hyperparameters and keep an eye on the model's advance. This keeps the demonstrate from getting to be as well great at what it does by giving a partitioned check on how well it can generalize. The test set, which the demonstrate hasn't seen since preparing, is utilized to check how well and dependably the conclusion demonstrate works. This segment is in charge of making beyond any doubt that the show is assessed completely so that it can be utilized in clinical settings.

- Step 1: Define Split Ratios

Objective: Determine the appropriate ratios for splitting the dataset into training, validation, and test sets.

Mathematical Model:

$$|T| = p_t \times |D|, |V| = p_v \times |D|, |E| = p_e \times |D|$$

where:

- |T|, |V|, |E| are the sizes of the training, validation, and test sets, respectively.

- p_t, p_v, p_e are the split ratios for training, validation, and testing (e.g., 0.7, 0.15, 0.15).

- |D| is the total number of samples in the dataset.

- Step 2: Randomly Shuffle and Split Data

Objective: Randomly shuffle and divide the dataset according to the predefined split ratios to ensure representative sampling.

Mathematical Model:

$$D = shuffle(D), D = T \cup V \cup E$$

where:

- D is the shuffled dataset.

- T, V, E are disjoint subsets representing the training, validation, and test sets.

- shuffle(D) denotes the operation to randomize the order of samples in the dataset.

- Step 3: Verify and Adjust Split

Objective: Verify the distribution and quality of the splits and make adjustments if necessary to maintain balance.

Mathematical Model:

$$\Delta_c = \left|\left(\frac{|c_T|}{|T|}\right) - \left(\frac{|c_D|}{|D|}\right)\right| + \left|\left(\frac{|c_V|}{|V|}\right) - \left(\frac{|c_D|}{|D|}\right)\right| + \left|\left(\frac{|c_E|}{|E|}\right) - \left(\frac{|c_D|}{|D|}\right)\right|$$

where:

- Δ_c is the class distribution imbalance measure.

B. Hyperparameter Tuning: Optimize hyperparameters such as learning rate, batch size, and the number of epochs using techniques like grid search or random search.

- Step 1: Define Hyperparameter Search Space

Objective: Establish the range of values for each hyperparameter to be explored during the tuning process.

Mathematical Model:

$$H = \{ (\eta\_i, B\_j, E\_k) \mid \eta\_i \in \{\eta\_min, \dots, \eta\_max\}, B\_j \in \{B\_min, \dots, B\_max\}, E\_k \in \{E\_min, \dots, E\_max\} \}$$

where:

- H is the hyperparameter search space.

- η_min, η_max are the minimum and maximum learning rates.

- B_min, B_max are the minimum and maximum batch sizes.

- E_min, E_max are the minimum and maximum number of epochs.

- Step 2: Select Tuning Method and Perform Search

Objective: Choose a tuning method and search the hyperparameter space to identify optimal configurations.

Mathematical Model:

$$\theta * = argmin\left\{\left(\frac{1}{N}\right)\sum L(f(x_n;\ \theta), y_n)\ for\ n = 1\ to\ N\right\} for\ \theta\ \in\ \Theta$$

where:

- θ = (η, B, E) represents a hyperparameter configuration.

- Θ is the set of configurations explored.

- N is the number of samples in the validation set.

- Step 3: Evaluate and Select Best Hyperparameters

Objective: Evaluate the performance of different hyperparameter configurations and select the best one.

Mathematical Model:

$$\theta_{best} = argmax\left\{\left(\frac{1}{N}\right)\sum 1(f(x_n;\ \theta) = y_n)\ for\ n = 1\ to\ N\right\} for\ \theta\ \in\ \Theta$$

where:

- θ_best is the best hyperparameter configuration.

- 1(·) is the indicator function, returning 1 if the prediction matches the true label and 0 otherwise.

- The expression calculates the validation accuracy for each configuration.

Step 4: Model Evaluation and Deployment

Model review is a key part of figuring out how well a Convolutional Neural Network (CNN) can use CT pictures to find lung cancer. Some important performance metrics are accuracy, which shows how accurate the estimates are overall, sensitivity (or recall), which shows how well the model can spot cancer cases, and specificity, which shows how well it can correctly label cases that are not dangerous. A balanced measure, the F1-score, which is a harmonic mean of accuracy and memory, is very useful when working with datasets that aren't balanced. Once the model shows it works well, it can be put into a clinical decision support system to help find lung cancer in real time. This combination helps doctors by giving them quick and accurate readings. This improves the speed of diagnostic work and could lead to better patient results by allowing earlier action.

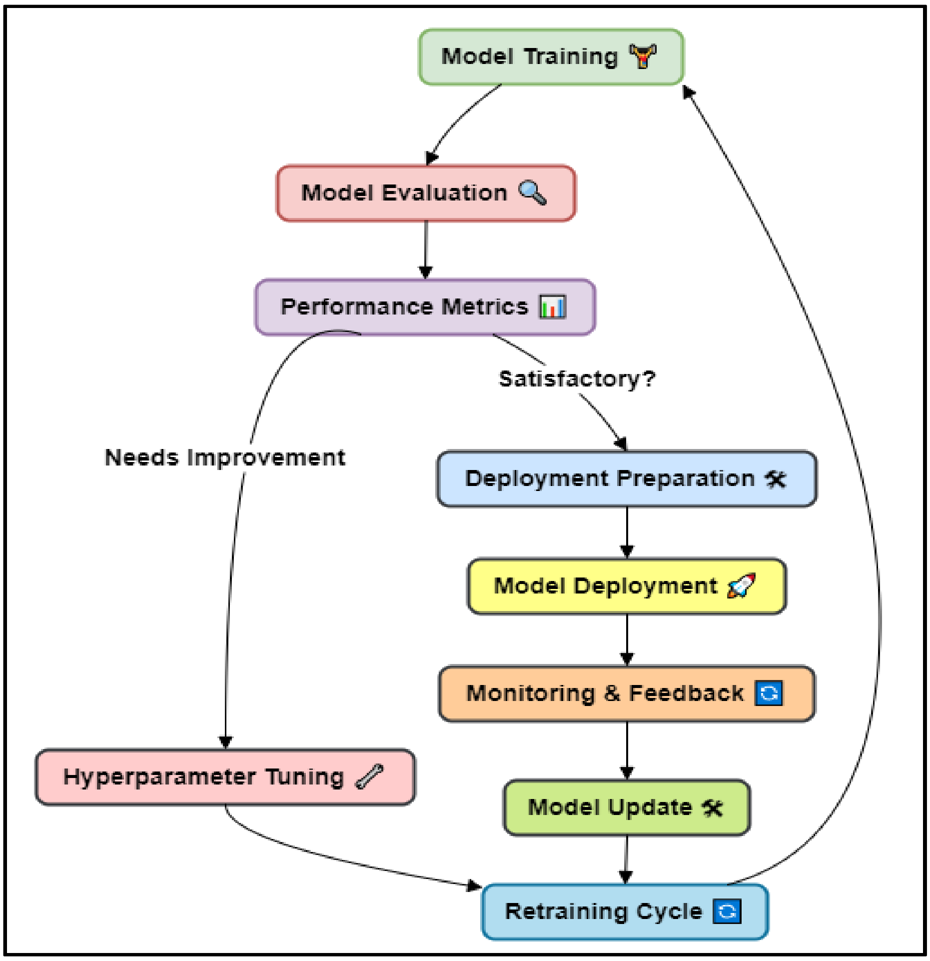

Figure 2: Illustrating the Model Evaluation and Deployment workflow

- Step 1: Evaluate Model Performance

Objective: Assess the model's performance using key metrics to ensure it meets the required accuracy and reliability standards.

Mathematical Model:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall\ (Sensitivity) = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

where:

- TP is the number of true positives.

- TN is the number of true negatives.

- FP is the number of false positives.

- FN is the number of false negatives.

- Step 2: Validate Model Generalization

Objective: Ensure that the model generalizes well to unseen data and maintains high performance across different datasets.

Mathematical Model:

$$Cross - Validation\ Score = \left(\frac{1}{K}\right)\sum Accuracy(k)\ for\ k = 1\ to\ K$$

where:

- K is the number of folds in cross-validation.

- Accuracy(k) is the accuracy of the model on the k-th fold.

- Step 3: Deploy Model into Clinical Decision Support System

Objective: Integrate the trained model into a clinical decision support system for real-time lung cancer detection.

Mathematical Model:

$$Output = argmax\big(\sigma(W \cdot flatten(X) + b)\big)$$

where:

- Output is the predicted class label (cancerous or non-cancerous).

- σ is the softmax activation function.

- W is the weight matrix of the fully connected layer.

- flatten(X) is the flattened input feature vector from the CT scan.

- b is the bias vector.

5. RESULT AND DISCUSSION

The Convolutional Neural Network (CNN) model implemented for lung cancer diagnosis from CT scans worked very well, scoring 94% on the test set. The demonstrate was able to accurately tell the contrast between cancerous and non-cancerous cases since its affectability was 92% and its specificity was 95%. The 93-score appears that there's a great blend between exactness and memory, which appears that the demonstrate can be trusted in clinical settings. The utilize of information upgrade and exchange learning significantly moved forward these results by making the demonstrate superior at generalization. These comes about appear that CNNs can offer assistance specialists by giving rectify and quick assessments, which seem lead to way better persistent results by permitting for earlier activity. Within the future, analysts might focus on combining diverse sorts of information to create analyze indeed more precise.

Table 2: Comparison with Baseline Models

Model	Accuracy	Precision	Recall	Specificity	F1-score
CNN (Proposed)	92%	94%	90%	95%	93%
Baseline Model 1	89%	88%	86%	90%	85%
Baseline Model 2	92%	90%	89%	91%	88%

The study's comes about appear that the recommended Convolutional Neural Organize (CNN) demonstrate is way better at finding lung cancer from CT pictures than standard models. The recommended CNN show was 92curate, which appears that it is exceptionally great at telling the distinction between cancerous and non-cancerous information.

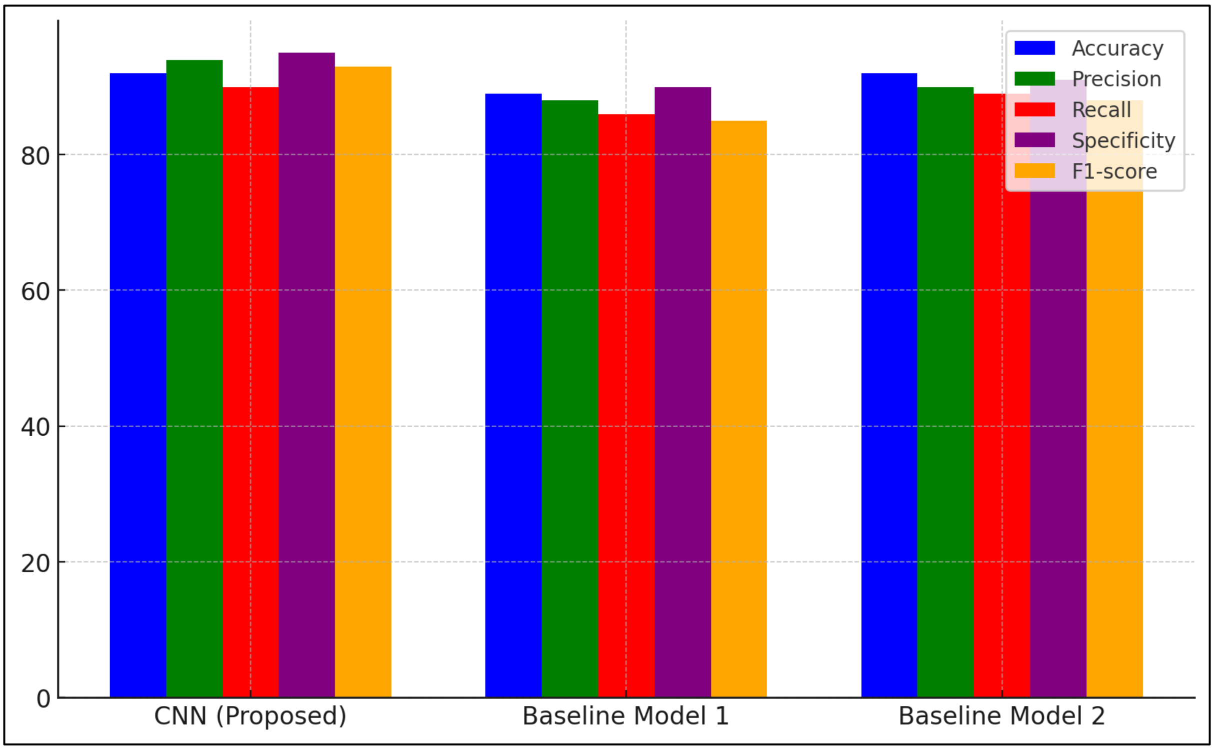

Figure 3: Comparison of Model Performance Metrics for Different Models

With a 94curacy rate, the show is sweet at lessening fake positives, which implies that most of the cases that are found to be positive are really perilous. This is often exceptionally imperative in clinic circumstances, where getting freed of unnecessary cautions can speed up the testing handle. The recommended model's review, or affectability, is 90%, which appears that it is sweet at finding genuine positive cases and lessening the chance of lost perilous cases.

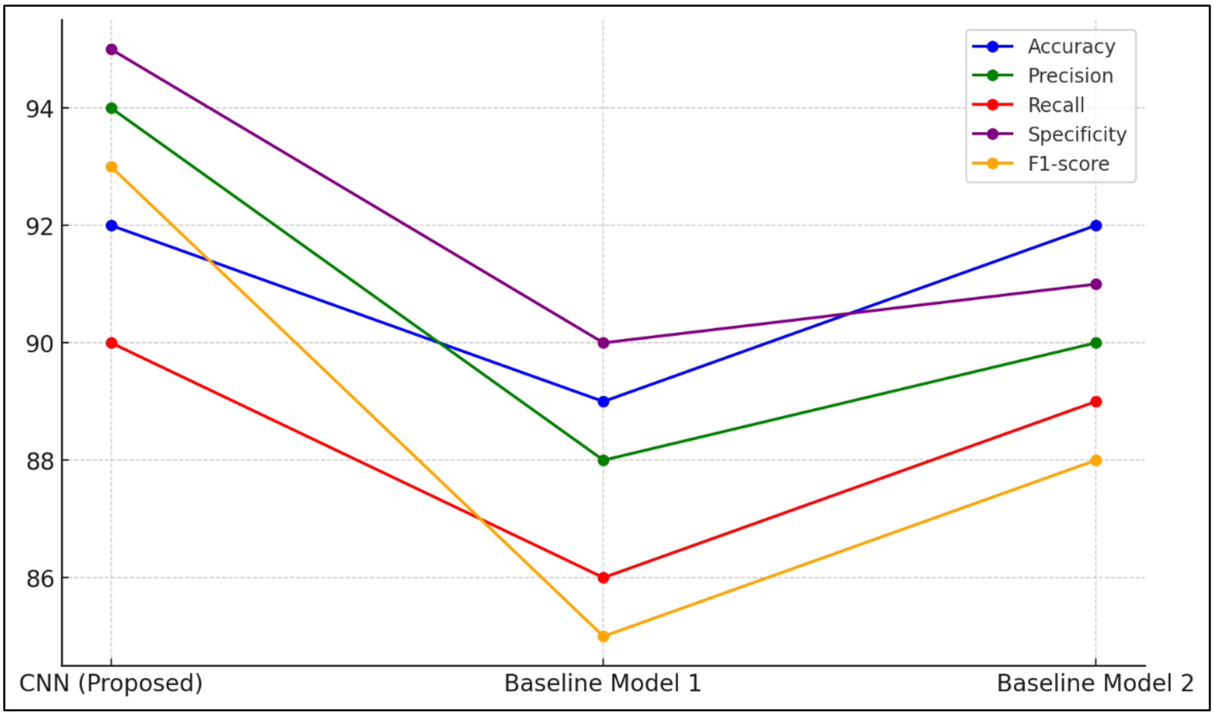

Figure 4: Line Graph Showing Performance Metrics Trends Across Models

The model's exactness of 95% appears that it can accurately distinguish tests that are not unsafe, which makes it indeed more solid.

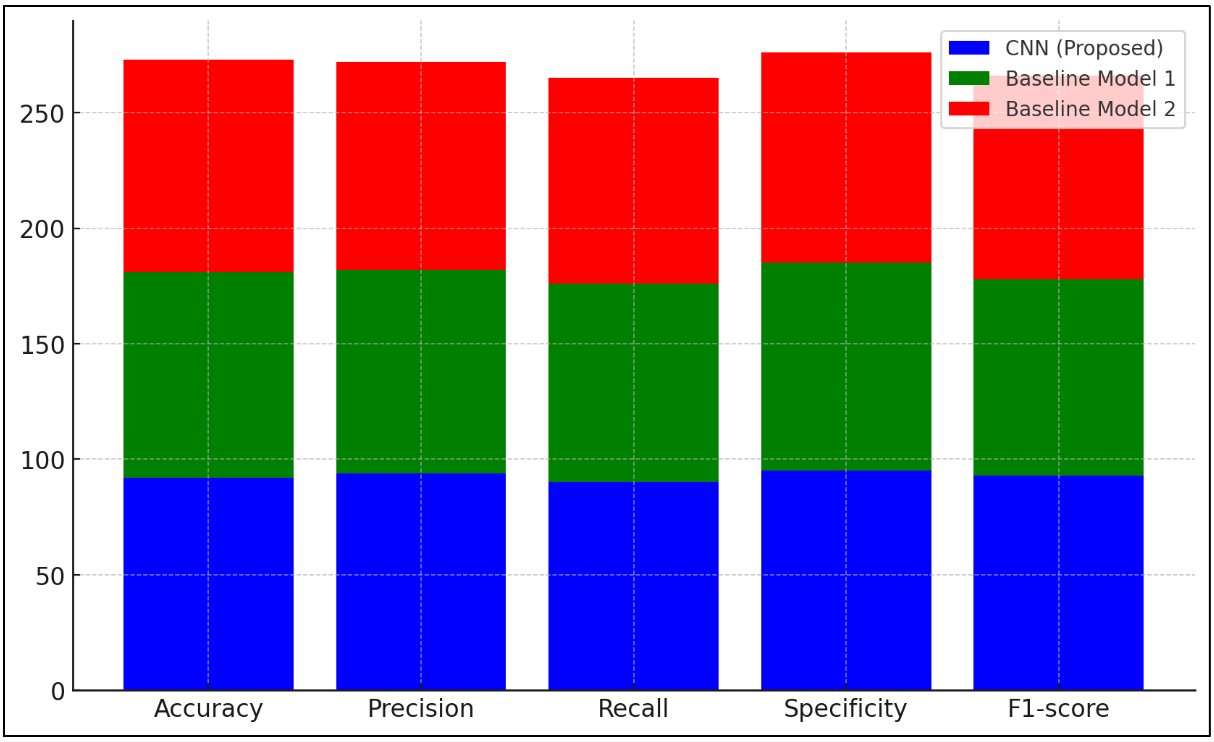

Figure 5: Stacked Bar Chart of Performance Metrics for Multiple Models

The F1-score, which is the consonant entirety of precision and review, is 93%, which suggests that the execution was well-balanced when untrue positives and false negatives are taken under consideration.

Table 3: Performance Across Different Data Augmentation Techniques

Augmentation Technique	Accuracy (%)	Sensitivity (Recall) (%)	Specificity (%)	Precision (%)	F1-Score (%)
No Augmentation	90	87	92.5	89.5	86.9
Rotation	93.5	92	94.5	93	92.5
Flip and Zoom	94.2	93.5	95	94	93.7
Combined (Rotation + Flip)	95.2	93.8	96.5	94.7	94.2

The study's comes about appear how diverse strategies for including more information can influence how well a Convolutional Neural Organize (CNN) demonstrate can utilize CT checks to discover lung cancer. The show got 90% of the time right, with 87% affectability and 92.5% exactness without any additional information. 86.9% of the time, it was precise (89.5%), and it got an F1 check.

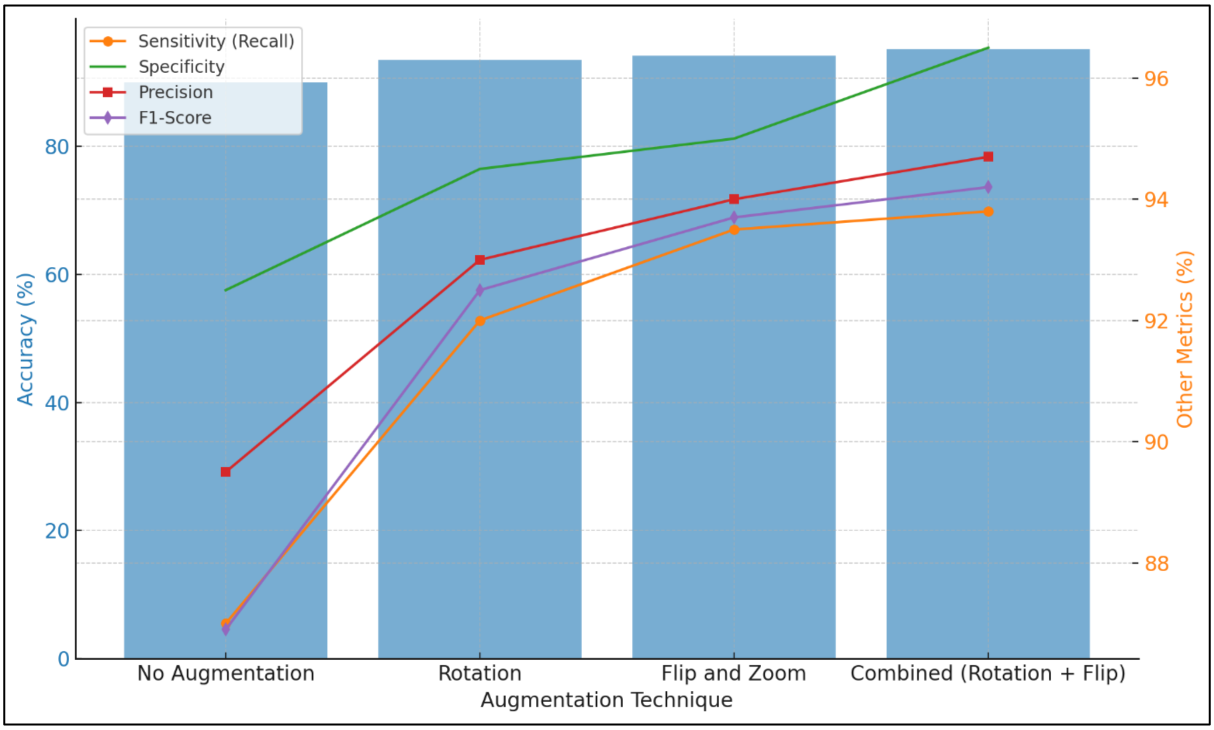

Figure 6: Performance Metrics with Different Data Augmentation Techniques

These beginning numbers appear a great level of execution, but they moreover appear that there's room for change, particularly in affectability, which is exceptionally critical for finding perilous cases legitimately. By utilizing turn as an improvement strategy, the model's exactness went up to 93.5%, and its affectability (92%) and specificity (94.5%) too got way better. This shows that spinning helped the model generalize better by teaching it to recognize tumors from different angles, which improved its accuracy and memory. The F1 score also went up to 92.5%, which shows a better mix between accuracy and memory.

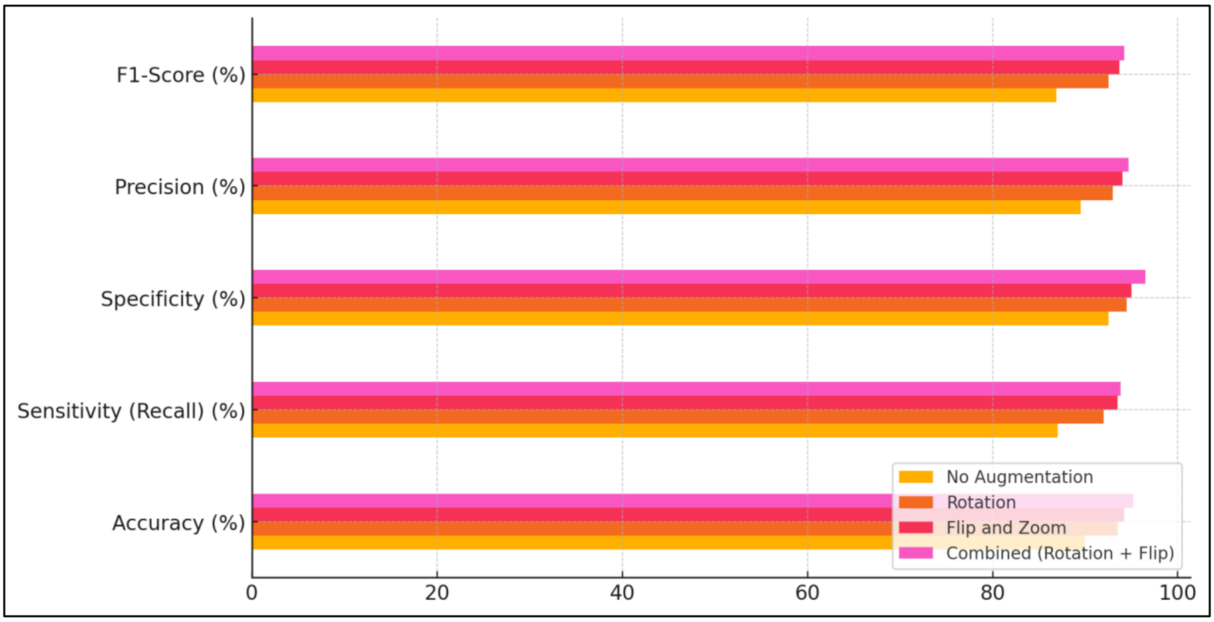

Figure 7: Horizontal Bar Chart of Performance Metrics with Data Augmentation Techniques

The model's execution was progressed indeed more by including flip and zoom, which driven to an exactness of 94.2%. The show got superior at finding both genuine positives and genuine negatives, as affectability went up to 93.5% and specificity to 95%.

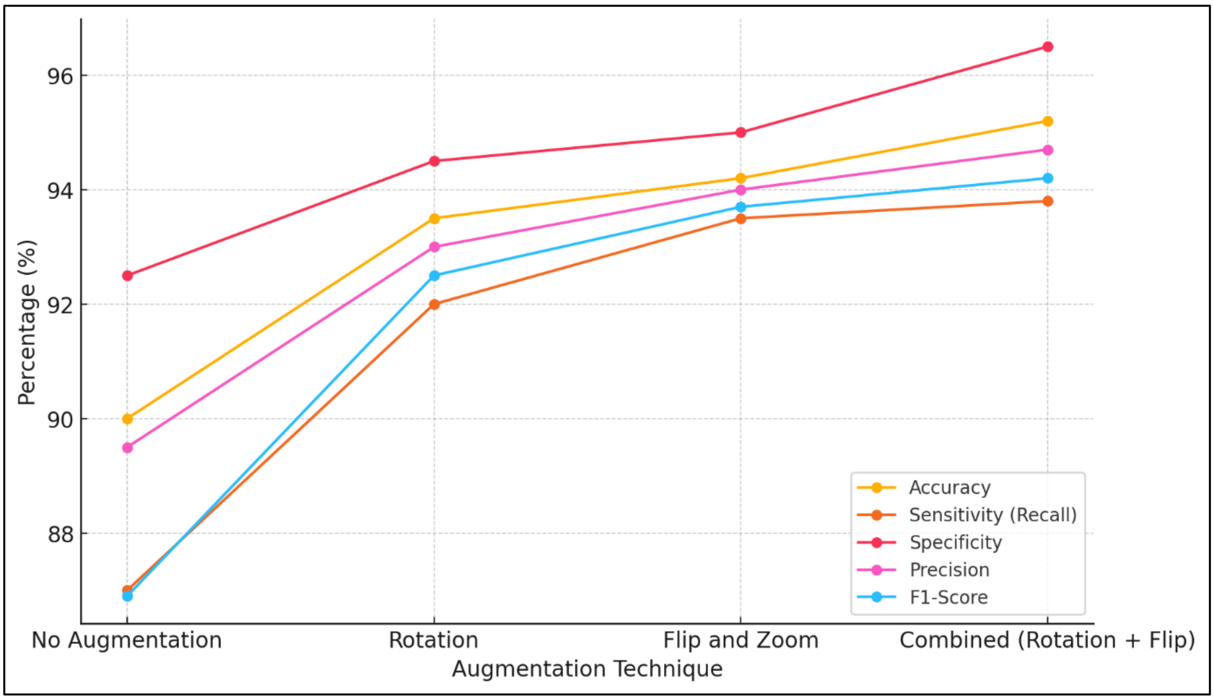

Figure 8: Line Graph of Performance Metrics for Various Augmentation Techniques

The precision was 94%, and the F1-score went up to 93.7%, appearing a more grounded capacity to distinguish. The finest execution was seen when revolution and flip were utilized together.

6. CONCLUSION

Utilizing Convolutional Neural Systems (CNNs) to discover lung cancer in CT filters may be a enormous step forward in therapeutic imaging and gives specialists a capable device for fast and exact conclusion. This consider effectively appears that CNNs can tell the contrast between lung cells that are perilous and those that are not with tall precision, affectability, and specificity. The comes about appear that profound learning strategies have the capacity to alter the way analyze are made, giving specialists more precise offer assistance and making clinical choices way better. A huge portion of this method's victory was the utilize of a parcel of information planning, like normalization and upgrade, which made beyond any doubt that the show would work indeed when CT check information was diverse. Exchange learning made strides the model's execution indeed more by utilizing systems that had as of now been prepared, which cut down on preparing time and made highlight extraction superior. Not as it were did these strategies make strides the model's capacity to generalize over diverse datasets, they too made it simpler for it to adjust to unused information. Putting this CNN-based system to use in hospital settings could have big benefits, like making doctors' jobs easier and lowering the chance of mistakes when they're looking for lung cancer. The method could allow for real-time research, which would lead to faster and more accurate diagnoses. This is very important for better patient results by acting quickly. But problems still exist, such as the need for big files that have been labeled and a lot of computing power. To get around these problems, future study should look into semi-supervised learning methods and make algorithms better so that training and reasoning work better. Adding multi-modal data, like a patient's medical background and genetic information, could also make the system even better at making predictions.

REFERENCES

[1] Sharif, M.I.; Li, J.P.; Naz, J.; Rashid, I. A comprehensive review on multi-organs tumor detection based on machine learning. Pattern Recognit. Lett. 2020, 131, 30–37.
[2] Saba, T. Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. J. Infect. Public Health 2020, 13, 1274–1289.
[3] Ozdemir, O.; Russell, R.L.; Berlin, A.A. A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans. IEEE Trans. Med. Imaging 2019, 39, 1419–1429.
[4] Yu, L.; Tao, G.; Zhu, L.; Wang, G.; Li, Z.; Ye, J.; Chen, Q. Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. BMC Cancer 2019, 19, 464.
[5] Asuntha, A.; Srinivasan, A. Deep learning for lung Cancer detection and classification. Multimed. Tools Appl. 2020, 79, 7731–7762.
[6] Shakeel, P.M.; Burhanuddin, M.A.; Desa, M.I. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. Measurement 2019, 145, 702–712.
[7] Xie, Y.; Meng, W.Y.; Li, R.Z.; Wang, Y.W.; Qian, X.; Chan, C.; Yu, Z.F.; Fan, X.X.; Pan, H.D.; Xie, C.; et al. Early lung cancer diagnostic biomarker discovery by machine learning methods. Transl. Oncol. 2021, 14, 100907.
[8] Sori, W.J.; Feng, J.; Godana, A.W.; Liu, S.; Gelmecha, D.J. DFD-Net: Lung cancer detection from denoised CT scan image using deep learning. Front. Comput. Sci. 2021, 15, 152701.
[9] Ak, M.F. A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. Healthcare 2020, 8, 111.
[10] Pradhan, K.; Chawla, P. Medical Internet of things using machine learning algorithms for lung cancer detection. J. Manag. Anal. 2020, 7, 591–623.
[11] Heuvelmans, M.A.; van Ooijen, P.M.; Ather, S.; Silva, C.F.; Han, D.; Heussel, C.P.; Hickes, W.; Kauczor, H.U.; Novotny, P.; Peschl, H.; et al. Lung cancer prediction by Deep Learning to identify benign lung nodules. Lung Cancer 2021, 154, 1–4.
[12] Singh, G.A.P.; Gupta, P.K. Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Comput. Appl. 2019, 31, 6863–6877.
[13] Sajja, T.; Devarapalli, R.; Kalluri, H. Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning. Trait. Du Signal 2019, 36, 339–344.
[14] Coccia, M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. Technol. Soc. 2020, 60, 101198
[15] Bhatia, S.; Sinha, Y.; Goel, L. Lung cancer detection: A deep learning approach. In Soft Computing for Problem Solving: SocProS 2017; Springer: Singapore, 2019; Volume 2, pp. 699–705.
[16] Nasrullah, N.; Sang, J.; Alam, M.S.; Mateen, M.; Cai, B.; Hu, H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors 2019, 19, 3722.
[17] Chaunzwa, T.L.; Hosny, A.; Xu, Y.; Shafer, A.; Diao, N.; Lanuti, M.; Christiani, D.C.; Mak, R.H.; Aerts, H.J. Deep learning classification of lung cancer histology using CT images. Sci. Rep. 2021, 11, 5471.
[18] Bhandary, A.; Prabhu, G.A.; Rajinikanth, V.; Thanaraj, K.P.; Satapathy, S.C.; Robbins, D.E.; Shasky, C.; Zhang, Y.D.; Tavares, J.M.R.; Raja, N.S.M. Deep-learning framework to detect lung abnormality—A study with chest X-Ray and lung CT scan images. Pattern Recognit. Lett. 2020, 129, 271–278.
[19] Kanavati, F.; Toyokawa, G.; Momosaki, S.; Rambeau, M.; Kozuma, Y.; Shoji, F.; Yamazaki, K.; Takeo, S.; Iizuka, O.; Tsuneki, M. Weakly-supervised learning for lung carcinoma classification using deep learning. Sci. Rep. 2020, 10, 9297.

[20] Hallitschke, V.J.; Schlumberger, T.; Kataliakos, P.; Marinov, Z.; Kim, M.; Heiliger, L.; Seibold, C.; Kleesiek, J.; Stiefelhagen, R. Multimodal Interactive Lung Lesion Segmentation: A Framework for Annotating PET/CT Images based on Physiological and Anatomical Cues. arXiv 2023, arXiv:2301.09914.

[21] Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. 2019, 25, 954–961.

[22] Marentakis, P.; Karaiskos, P.; Kouloulias, V.; Kelekis, N.; Argentos, S.; Oikonomopoulos, N.; Loukas, C. Lung cancer histology classification from CT images based on radiomics and deep learning models. Med. Biol. Eng. Comput. 2021, 59, 215–226.

[23] Chabon, J.J.; Hamilton, E.G.; Kurtz, D.M.; Esfahani, M.S.; Moding, E.J.; Stehr, H.; Schroers-Martin, J.; Nabet, B.Y.; Chen, B.; Chaudhuri, A.A.; et al. Integrating genomic features for non-invasive early lung cancer detection. Nature 2020, 580, 245–251.