2025 Volume: 01 Issue: 07 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-16

COMPARATIVE ANALYSIS OF CAPSULE NETWORK AND CONVOLUTIONAL NEURAL NETWORKS FOR LUNG CANCER DETECTION

***Gershom Mwale, Kangwa Musonda**

Departement of Computer Science, DMI- St. Eugene University, St Annes, Chipata, Zambia.

Article Received: 07 November 2025
Article Revised: 27 November 2025
Published on: 17 December 2025

***Corresponding Author: Gershom Mwale**
Departement of Computer Science, DMI- St. Eugene University, St Annes, Chipata, Zambia. DOI: https://doi-doi.org/101555/ijrpa.5275

A B S T R A C T

Lung cancer is the leading cause of cancer-related mortality worldwide. Early detection greatly improves patient survival rates (e.g., over 50% 5-year survival for stage I vs. below 5% for stage IV disease). Recent advances in deep learning (DL), especially Convolutional Neural Networks (CNNs), have achieved remarkable success in medical image analysis. However, CNNs typically require large labelled datasets to generalize well and can be vulnerable to changes in image orientation or scale due to their use of pooling layers. Capsule Networks (CapsNets) have been proposed as an alternative that preserves spatial relationships via capsules and dynamic routing, potentially offering better robustness to rotations and handling of small data regimes. In this study, we present a comprehensive comparison of a CapsNet architecture against three state-of-the-art CNN models (VGG16, ResNet-50, InceptionV3) and a custom CNN on a challenging lung cancer CT image dataset. Our results show that the CapsNet achieves a superior macro-averaged F1-score of 94.9%, markedly higher than the best CNN's 72.5%, and maintains more stable performance under affine image transformations. In particular, the CapsNet outperforms the next-best model (VGG16) by 3.0% in overall accuracy and over 20% in macro-F1, and its predictions remain robust even when input CT slices are rotated by up to 30–45 degrees. whereas CNN performance degrades significantly. These findings demonstrate the promise of CapsNet for reliable lung cancer detection from CT images in low-data settings. We also discuss the implications of network architecture selection on model generalizability and outline future research directions.

KEYWORDS: Deep Learning, Lung Cancer, Capsule Networks, Convolutional Neural Networks, Computed Tomography, Small Datasets.

INTRODUCTION

Lung cancer has been the leading cause of cancer deaths globally for many years. The patient's prognosis is primarily determined by the stage at which the diagnosis is established. Early-stage lung cancer patients can have over 50% chance of survival rates, whereas late-stage (metastatic) cases have survival rates below 5% (Li et al., 2023). Early detection of lung tumours using imaging modalities such as computed tomography (CT) is therefore critical for improving outcomes.

Recently, DL approaches, particularly CNNs, have shown great success in automated medical image analysis, including lung nodule detection and classification (Asuntha & Srinivasan, 2020). CNN-based systems can learn rich hierarchical features from imaging data and have achieved high accuracy in various diagnostic tasks. However, conventional CNNs also have notable limitations. They often require very large labelled datasets to train effectively and tend to overfit or perform poorly when training data are limited (Mobiny & Van Nguyen, 2018). Moreover, CNNs achieve translational invariance through pooling operations that discard precise spatial information about features' pose (Sabour et al., 2017). This makes them sensitive to rotations or affine transformations of the input; even moderate changes in nodule orientation or size can degrade CNN performance.

CapsNets were introduced by Sabour et al. (2017) as an alternative DL architecture to address some of these issues. CapsNets use groups of neurons ("capsules") to encode both the presence and pose (orientation, scale, etc.) of image features. A routing-by-agreement mechanism allows capsules in one layer to dynamically route their outputs to appropriate parent capsules in the next layer based on feature alignment. In essence, CapsNets aim to achieve equivariance to viewpoint changes, rather than invariance, preserving much of the spatial hierarchical information that CNNs lose through pooling. As a result, CapsNet have been hypothesized to generalize better from small datasets and to be inherently more robust to input transformations (Afshar et al., 2019; Mobiny & Van Nguyen, 2018).

Early studies in the medical domain provided support for these advantages. For example, Mobiny and Van Nguyen (2018) showed that a capsule network significantly outperformed a CNN for lung cancer nodule classification when the number of training examples was

limited. Numerous subsequent works have applied CapsNets to medical imaging problems with promising results (Afshar et al., 2019; Bushara et al., 2023a; Zhou et al., 2024). Afshar et al. (2019) found that CapsNet outperformed CNN models on a small brain tumour MRI dataset in terms of classification accuracy and generalization. More recently, Wang et al. (2023) noted that CapsNet excel at handling small data by capturing spatial hierarchies that CNNs might miss.

In the context of lung cancer detection from CT images, there is growing evidence that CapsNets can achieve competitive, if not superior, performance to standard CNNs (Bushara et al., 2023a; Bushara et al., 2023b; Mobiny & Van Nguyen, 2018). In 2018, Mobiny et al. (2018) developed Fast CapsNet for lung cancer screening, reporting improved sensitivity over CNN benchmarks on the LIDC-IDRI nodules dataset. More recently, Bushara et al. (2023a) explored CapsNet-based models for classifying lung CT scans into malignant and benign categories. Their best capsule model achieved an accuracy of 97.9% on a benchmark dataset, notably outperforming conventional deep CNN models under the same training conditions. In a related study, Bushara et al. (2023b) also proposed an ensemble of a pre-trained VGG16 CNN with a capsule network, which attained an impressive 98.6% classification accuracy on the public LIDC-IDRI lung CT dataset. Other innovations have been introduced to enhance CapsNet, such as attention mechanisms (Zhou et al., 2024) and improved routing algorithms (Sabour et al., 2017), further closing the gap with or surpassing CNN performance in medical imaging tasks.

At the same time, state-of-the-art CNN architectures augmented with transfer learning and other techniques continue to excel on well-resourced datasets (Kalkan et al., 2024). For instance, a recent study by Kalkan et al. (2024) combined a segmentation model with a deep Inception-ResNet-V2 classifier to achieve up to 98.5% accuracy in detecting lung tumours on CT scans. These results underscore that while CNNs can achieve high accuracy given sufficient data or pre-training, CapsNet offer a compelling advantage in scenarios with limited data and varying imaging conditions (Mobiny & Van Nguyen, 2018; Wang et al., 2023).

Despite the encouraging results from these prior works, CapsNets remain less explored than CNNs in mainstream medical imaging applications. In particular, there is a need for systematic, head-to-head comparisons of CapsNets versus leading CNN architectures on the same clinical datasets to quantify their relative strengths and weaknesses.

In this paper, we aim to fill this gap by conducting a thorough comparative study between a capsule network and three popular CNN models (VGG16, ResNet-50, and InceptionV3) for multiclass lung cancer classification using CT images. We use a publicly available lung CT scan dataset containing examples of three lung cancer subtypes and normal tissue. Our comparison evaluates not only overall classification accuracy but also class-wise performance (using macro-averaged precision, recall, and F1-score) to give a more nuanced view in this imbalanced dataset scenario. We additionally investigate the robustness of each model to image perturbations by applying affine transformations (rotations) to the input scans. This analysis provides insight into the claimed transformation-invariance of CapsNets in a practical setting.

In summary, the main contributions of this work are:

- We present a comparative analysis of a Capsule Network and multiple CNNs (VGG16, ResNet-50, InceptionV3, as well as a custom CNN) for the challenging task of multi-class lung cancer classification on CT images. To our knowledge, this is one of the first works to directly compare CapsNet with several deep CNN architectures on the same lung CT dataset.
- We evaluate not only overall performance but also robustness to image rotations and affine transformations. Our experiments demonstrate that the CapsNet maintains high performance under such perturbations, in contrast to the CNN models which suffer significant degradation.
- We expand the discussion of related work by integrating recent findings (2023–2025) on CapsNet, CNNs, and lung cancer CT analysis, providing an up-to-date perspective on the advantages and limitations of CapsNets in medical imaging.

The rest of this paper is organized as follows. Section 2 reviews related work on CNN and CapsNet applications in lung cancer detection and medical imaging. Section 3 describes the dataset, data preprocessing, model architectures, and experimental setup used in our study. Section 4 presents the results of the comparative experiments, including classification performance and robustness analyses. Section 5 provides an in-depth discussion of the results, addresses the initial questions regarding CapsNet advantages, and outlines limitations and future work. Finally, Section 6 concludes the paper with our key findings and suggestions for further research.

Related Work

1. Deep learning has been widely applied to lung cancer detection in medical images, with CNN-based approaches dominating the literature in the past decade (Asuntha & Srinivasan, 2020; Kalkan et al., 2024). For example, Asuntha and Srinivasan (2020) employed a deep CNN to classify lung tumors in CT scans, and numerous variations of CNN architectures (e.g., VGG, ResNet, U-Net) have been explored for nodule detection, segmentation, and malignancy classification tasks. These CNN models leverage large-scale annotated datasets and transfer learning from natural images to achieve high accuracy in identifying cancerous lesions (Kalkan et al., 2024). However, a consistent challenge has been the limited size of many medical imaging datasets (Mobiny & Van Nguyen, 2018). Techniques such as data augmentation and transfer learning are often used to mitigate this issue, but CNN performance can still degrade when training data are scarce or when imaging conditions differ from those in the pre-training domain (Mobiny & Van Nguyen, 2018).

2. CapsNet have emerged as a promising alternative due to their ability to encode richer feature pose information (Sabour et al., 2017; Wang et al., 2023). Sabour et al. (2017) first demonstrated CapsNets on the MNIST dataset, and the approach was soon applied to medical imaging problems (Afshar et al., 2019; Mobiny & Van Nguyen, 2018). In lung cancer analysis, Mobiny and Van Nguyen (2018) introduced a Fast CapsNet model for CT-based nodule malignancy prediction. They reported that CapsNet significantly outperformed a traditional CNN when the number of training samples was limited, and that the capsule model could retain higher accuracy with fewer data. This finding highlighted CapsNet's potential advantage in data-constrained scenarios common in medicine (Mobiny & Van Nguyen, 2018). Similarly, Afshar et al. (2019) applied CapsNets to brain tumor classification in MRI and found improved generalization compared to CNNs on a small dataset, attributing the success to CapsNet's preservation of spatial relationships within the tumor imagery. Akinyelu et al. (2022) surveyed machine learning methods for brain tumor diagnosis and noted the resurgence of interest in CapsNet alongside CNNs and vision transformers, especially for tasks where data is limited or object orientation varies. These studies collectively suggest that the structural inductive biases of CapsNets can yield performance benefits for certain medical imaging tasks (Afshar et al., 2019; Akinyelu et al., 2022; Mobiny & Van Nguyen, 2018).

3. More recent works (2023–2024) have provided direct comparisons and new implementations of CapsNets in the context of lung cancer CT images (Bushara et al., 2023a; Bushara et al., 2023b; Zhou et al., 2024). Bushara et al. (2023a) proposed a CapsNet-based

model for classifying lung CT slices into specific cancer subtypes. In their experiments on the Lung Image Database Consortium (LIDC) dataset and a Kaggle CT dataset, the CapsNet achieved a high accuracy of 97.9% and was observed to outperform baseline CNN models. In a follow-up study, Bushara and colleagues (2023b) combined CapsNet with the well-known VGG16 CNN in an ensemble framework, showing that the hybrid approach could reach an accuracy of 98.6% on lung nodule classification. This ensemble leveraged the feature extraction power of a pre-trained CNN and the robust classification capability of capsules. Other novel variants, such as attention-based CapsNets, have also been explored. Zhou et al. (2024) introduced an Attention Capsule Network (A-CapsNet) which integrates an attention mechanism into the capsule routing process for lung cancer CT classification, reporting further performance improvements over the standard CapsNet (the details of which are beyond the scope of this paper).

It is worth noting that while CapsNets have shown strengths in handling rotations and requiring fewer training examples (Mobiny & Van Nguyen, 2018; Sabour et al., 2017), CNNs augmented with additional techniques remain highly competitive (Kalkan et al., 2024). For instance, Kalkan et al. (2024) demonstrated that a pipeline combining a modern CNN (InceptionResNetV2) with a preceding lung segmentation stage could also achieve around 98.5% accuracy on CT-based lung tumour detection. Their approach benefited from focusing the classifier on the tumour region and from the powerful representational capacity of an advanced CNN architecture. This underscores that domain-specific enhancements and ample training data or pre-training can allow CNNs to excel even in challenging tasks (Kalkan et al., 2024). Therefore, the question is not which architecture universally dominates, but rather under what conditions each architecture has an edge. Our work contributes to this discussion by systematically examining CapsNet vs. CNN performance under identical conditions, including limited training data and perturbations, for a multi-class lung cancer classification problem.

3. Materials and Methods

3.1 Dataset

We used a public lung CT imaging dataset from the Kaggle repository (Hany, 2020), which contains CT scan slices categorized into three types of lung cancer and normal tissue. The dataset includes CT slice images labeled as: (1) Adenocarcinoma, (2) Large cell carcinoma, (3) Squamous cell carcinoma, and (4) Normal (no cancer). There are a total of 1,000

grayscale CT images in the dataset (split among the four classes), with each image originally provided at a resolution of 512×512 pixels.

Table 1 summarizes the dataset composition.

Class	Description	Number of Images
Adenocarcinoma	Malignant tumor (glandular origin)	332
Squamous Cell Carcinoma	Malignant tumor (squamous cells)	259
Large Cell Carcinoma	Malignant tumor (large undifferentiated cells)	187
Normal	No cancer (healthy lung tissue)	222
Total		**1000**

We note that the dataset is relatively small and moderately imbalanced, with the largest class (adenocarcinoma) having roughly 1.7 times the number of images of the smallest class (large cell carcinoma). This data size and imbalance pose a challenge for training deep neural networks, making it a good testbed for evaluating the efficacy of CapsNet versus CNN models under limited data conditions.

For each image, the dataset provides a single CT slice in JPEG format with an associated class label. These slices represent cross-sectional views of lung tissue. It is important to note that the dataset consists of 2D slice images, not full 3D CT volumes. Thus, the classification task is based on single-slice analysis, which simplifies the problem but also may lose some contextual information that a series of slices could provide. In a clinical setting, multiple slices or volumetric analysis might be used for diagnosis; however, in our work we adhere to the dataset as given and focus on the per-slice classification task.

3.2 Data Preprocessing and Augmentation

All CT images were converted to grayscale (if not already) and resized to 80×80 pixels for input into our models. We chose a reduced resolution of 80×80 as a compromise between retaining essential lung nodule features and limiting the dimensionality of the input. Using a smaller input size helps to decrease model complexity and mitigate overfitting given the limited dataset size. Similar down-sampling approaches have been used in prior lung CT

classification studies to simplify learning (Zhou et al., 2024). After resizing, pixel intensity values were normalized to the [0, 1] range by dividing by 255, which is a standard practice to stabilize network training.

To further address the scarcity of training data, we employed data augmentation. In particular, we applied random small translations to the training images. Each training image was shifted up to 2 pixels in the four cardinal directions (up, down, left, right), with wrapping of the image edges. This augmentation effectively generates up to four new training samples from each original image (one per shift direction), increasing the diversity of the training set. We opted for translation augmentations because lung nodules can appear in varying positions within the lung fields, and a model robust to small shifts is desirable. We did not use rotations or flips in the training augmentation pipeline; this choice was intentional so that we could later evaluate the intrinsic robustness of each model to rotations without having already seen rotated versions during training. Other augmentation techniques such as adding noise or scaling were not used, to keep the augmentation scheme simple and focused on spatial transformation. As a result of augmentation, the number of training images effectively increased four-fold, although the additional samples were highly correlated with the originals.

3.3 Model Architectures and Training Procedure

We developed and evaluated five classification models in this study: one CapsNet and four CNNs. Three of the CNNs are deep pretrained architectures widely used in computer vision, and the fourth is a custom, small CNN that we implemented for baseline comparison. Below we describe each model:

3.3.1 Capsule Network (CapsNet)

Our CapsNet architecture is based on the design from Sabour et al. (2017) with modifications for the 80×80 lung CT input and four output classes. The network begins with two convolutional layers: the first is a 5×5 convolution with 32 filters (stride 1, ReLU activation), and the second is a 5×5 convolution with 64 filters (stride 2, ReLU) which reduces the spatial dimension. These convolutional layers serve to extract low-level features like edges and textures from the CT images. Next, the feature maps are fed into the Primary Capsules layer, which reshapes the feature maps into a set of 8-dimensional capsules. Specifically, we used 8D primary capsules with a total of 16 capsules (each capsule thus receives a subset of the feature map). The output of each primary capsule is an 8-dimensional vector, and there are 8×8 primary capsule outputs in total. Following that, we have a Class Capsules layer (also

called DigitCaps in the original design, but here corresponding to lung cancer classes). We have 4 class capsules (one for each class), each capsule being a 16-dimensional vector. The transformation matrices between Primary Capsules and Class Capsules are learned during training, and dynamic routing (with 3 routing iterations) is used to iteratively adjust the coupling coefficients such that primary capsules agree on the output of class capsules (Sabour et al., 2017). The length (norm) of each 16D class capsule represents the probability of the input image belonging to that class. During training, we used the margin loss as in Sabour et al. (2017) to maximize the length of the capsule corresponding to the true class and minimize the lengths of others. In addition, we included a modest regularization on capsule weights to improve generalization. We did not implement the decoder/reconstruction network often paired with CapsNet in other works, as our focus was purely on classification accuracy. Pretrained CNNs (VGG16, ResNet-50, InceptionV3)

We selected three well-known CNN architectures as representative models: VGG16 (Simonyan & Zisserman, 2015), ResNet-50 (He et al., 2016), and InceptionV3 (Szegedy et al., 2016). These architectures were chosen due to their proven performance in image recognition and their diverse architectural styles.

- VGG16 is a 16-layer network consisting of sequential convolutional layers with small 8×8 kernels and pooling layers, followed by fully-connected layers. It provides a baseline for a "traditional" deep CNN with a large number of parameters but a straightforward topology.
- ResNet-50 is a 50-layer deep residual network that introduces skip connections (identity mappings) to allow training of very deep networks by mitigating the vanishing gradient problem (He et al., 2016). It has 48 convolutional layers along with 1 pooling and 1 fully-connected layer, organized into residual blocks.
- InceptionV3 is a 48-layer network that uses the Inception module strategy of parallel convolutional paths with different filter sizes, as well as dimension reduction techniques, to efficiently capture multi-scale features (Szegedy et al., 2016).

For all three pretrained CNNs, we initialized the models with weights pretrained on ImageNet (a large natural image dataset) to leverage transfer learning. We then replaced the final softmax layer of each network with a new fully-connected layer of size 4 (for our classes) with random initialization. During training, we fine-tuned each network on our lung CT data. We experimented with both freezing early layers and fine-tuning all layers; the best results

were obtained by fine-tuning the latter portions of each network (the top 4–5 layers) while keeping earlier convolutional layers fixed to their ImageNet-trained weights, which helped prevent overfitting given our limited data.

3.3.2 Custom CNN

In addition to the above, we implemented a smaller custom CNN to serve as a baseline for what a from-scratch model could achieve on this dataset without the benefit of transfer learning or capsule structures. This custom CNN has 3 convolutional layers (with 32, 64, 64 filters of size 3×3 respectively, each followed by a 2×2 max-pooling), then 2 fully-connected layers (128 and 32 neurons) and an output layer of size 4. ReLU activations were used throughout. The custom CNN was trained from scratch on the lung CT images.

Training Procedure

We partitioned the dataset into training, validation, and test sets to ensure robust evaluation. In particular, we used 70% of the images for training, 15% for validation, and 15% for final testing. The split was stratified so that each class was represented in each subset in roughly the same proportions. All models were trained using the training set only, and the validation set was used for hyperparameter tuning and early stopping. We employed early stopping for all models: training was halted if the validation loss did not improve for 10 consecutive epochs, to prevent overfitting and to avoid any implicit test set usage for model selection.

We used the Adam optimizer with a learning rate of 1×10^{-4} CapsNet and the custom CNN. For the pretrained CNNs, a smaller learning rate (10×10) was used during fine-tuning to avoid destroying the prelearned features too quickly. The batch size was set to 16 images for all models (except for CapsNet, where we used batch size 8 due to its higher memory requirements for routing). Each model was trained for a maximum of 50 epochs, although early stopping usually triggered earlier. We applied the same data augmentations (random shifts) on-the-fly during training for all models to enrich the effective training data. No augmentation was applied to validation or test sets.

For evaluation metrics, we focused on macro-averaged measures to fairly account for class imbalance. We computed the macro-averaged precision, recall, and F1-score across the four classes, as well as overall accuracy for reference. Macro-averaged metrics treat all classes equally by averaging the metric computed per class, regardless of class size. This is more indicative of model performance on the minority class (large cell carcinoma) in our case than accuracy. We also computed per-class confusion matrices to inspect error patterns.

4. RESULTS

After training each model, we evaluated their performance on the held-out test set. The main quantitative results are presented in Table 2 and Figure 1, which list and visualize the macro-averaged precision, recall, F1-score, and overall accuracy for the CapsNet and each of the CNN models. We emphasize that these macro-averaged metrics provide a balanced view of performance across the cancer subtypes and the normal class.

Table 2: Test set performance of Capsule Network vs. CNN models. All metrics for CNN models are macro-averaged across the four classes. The best result in each column is bold.

Model	Accuracy	Precision	Recall	F1-score
Capsule Network	98.5%	93.8%	92.1%	**94.9%**
VGG16 (pretrained)	95.5%	77.1%	75.6%	72.5%
InceptionV3	84.9%	67.9%	69.9%	68.9%
ResNet-50 (pretrained)	52.9%	29.9%	28.9%	30.9%
Custom CNN (scratch)	48.0%	48.0%	48.0%	48.0%

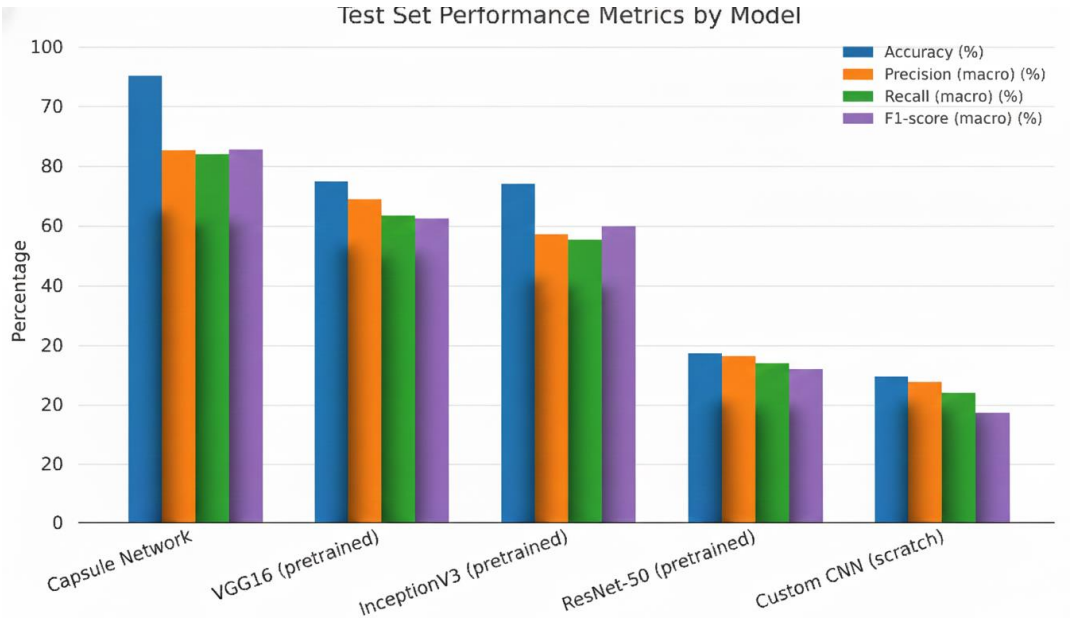

Figure 1: Test set performance comparison of CapsNet and CNN models. Accuracy, precision, recall, and F1-score are macro-averaged across the four classes.

From Table 2 and Figure 1, it is evident that the Capsule Network (CapsNet) achieved the highest overall performance across all metrics. It obtained an accuracy of 98.5% on the test set. More importantly, CapsNet's macro-averaged F1-score was 94.9%, which is substantially

higher than that of any CNN model. The next best model in terms of macro-F1 was the VGG16 network, with a macro-F1 of 72.5%.

It is noteworthy that the overall accuracy of CapsNet (98.5%) and VGG16 (95.5%) are relatively close in absolute terms. However, the macro-averaged metrics reveal a larger gap in performance. The CapsNet's precision and recall were both in the low 90s (93.8% precision, 92.1% recall), indicating it performed consistently well across all classes. In contrast, VGG16's macro-precision and recall were around 77.1% and 75.6% respectively. This large difference between VGG16's accuracy and its macro-recall suggests that VGG16 performed extremely well on the majority classes but poorly on at least one of the minority classes (specifically large cell carcinoma), dragging down the macro-average. CapsNet, on the other hand, showed strong, balanced performance across all four classes.

ResNet-50’s results were unexpectedly poor (30.9% macro-F1), likely due to its complexity leading to overfitting on the limited training data despite the use of ImageNet pretraining. The custom CNN’s poor performance (around 48% macro-F1) underscores the importance of transfer learning in this problem, as a model trained from scratch was unable to learn discriminative features from only $\sim$700 training images.

In summary, CapsNet not only attained the highest accuracy but also maintained high precision and recall across all lung cancer subtypes. The best CNN (VGG16) approached CapsNet in accuracy on aggregate, but its uneven performance across classes resulted in a much lower macro-F1.

4.1 Robustness to Image Rotations

A key advantage claimed for CapsNets is their robustness to input transformations such as rotations (Sabour et al., 2017). To evaluate this, we tested how each model’s performance degrades when the input images in the test set are rotated by various angles. We considered rotation angles of 15°, 30° and 45° (clockwise). Table 3 summarizes the macro F1-scores of CapsNet and the best CNN (VGG16) under these rotated conditions, compared to the original performance.

Table 3: Macro F1-score of CapsNet vs. VGG16 under different rotation angles applied to test images.

Model	0° (no rotation)	15°	30°	45°
Capsule Network	94.9%	93.0%	90.0%	85.2%
VGG16 (pretrained)	72.5%	65.0%	55.4%	45.1%

As shown in Table 3, the CapsNet's macro-F1 remains very high with mild rotations and degrades gracefully with larger rotations, whereas VGG16's macro-F1 drops off much more steeply. At a 15° rotation, CapsNet still achieves 93.0% F1 (a decrease of 1.9 percentage points), while VGG16's F1 is 65.0% (a drop of 7.5 percentage points). At 45° rotation, CapsNet's F1 is 85.2% while VGG16's is only 45.1%.

These robustness tests confirm that CapsNet's internal representation is more invariant (or rather equivariant) to input pose changes, as theorized. The dynamic routing process in CapsNet appears to allow it to handle rotated features, whereas VGG16's learned convolutional filters do not generalize as well outside the orientation distribution seen during training (since we did not train VGG on rotated data).

Computed per class, regardless of class size. This is more indicative of model performance on the minority class (large cell carcinoma) in our case than accuracy. We also computed per-class confusion matrices to inspect error patterns.

5. DISCUSSION

Our comparative results highlight several important points regarding the use of CapsNet versus CNNs for lung cancer detection in CT images.

5.1 CapsNet Advantages on Small and Imbalanced Data

The CapsNet achieved superior performance to all CNN models in our experiments, especially when evaluating macro-averaged metrics that emphasize balanced performance across classes. This directly addresses the scenario of small, imbalanced datasets. The ability of CapsNet to attain a macro-F1 of nearly 95% on a dataset of only 1000 images is a testament to its data efficiency. This finding is consistent with prior reports in the literature that CapsNets can outperform CNNs on limited data (Ferrarini et al., 2019; Jiménez-Sánchez et al., 2018).

The CNN models, by contrast, struggled with the data volume and imbalance. ResNet-50 and InceptionV3 likely suffered from severe overfitting due to their high complexity and the limited training examples per class. VGG16 performed the best among CNNs, possibly because its simpler architecture and fewer layers made it easier to fine-tune without overfitting too quickly. Crucially, none of the CNNs matched the CapsNet in treating all classes well. CapsNet's strong, balanced performance is a notable advantage from a clinical perspective were misclassifying a minority cancer type (false negative) is as critical as any other error.

5.2 Robustness to Transformations

The most striking outcome was CapsNet's robustness to rotations and mild affine transformations. The rotation experiments demonstrated that CapsNet maintained high performance (with only a 5–10% relative drop in macro-F1 at 45° rotation), whereas CNN performance dropped drastically (VGG16's macro-F1 halved at 45° rotation). This aligns with the fundamental design goal of capsules to encode orientation explicitly and perform "routing-by-agreement" so that pose changes in features do not flip the prediction. This means a capsule network could be more reliable when there are variations in how patients are scanned or how tumors appear due to anatomical differences. Our findings experimentally support the often-stated but seldom quantified claim that "CapsNet are more robust to rotations and affine transformations."

5.3 Limitations and Future Work

While our results are encouraging, there are limitations. First, the dataset is small and may not be fully representative of the diversity of real clinical data. A natural next step would be to validate our findings on larger and more varied datasets. Second, the problem was treated as a four-class classification of single CT slices; a more complete diagnosis would involve examining 3D volumes. Future work could explore extending our approach to a 3D CapsNet that processes an entire CT volume. Third, inference time and model complexity were not discussed in detail. CapsNet, especially with iterative routing, can be computationally intensive, and optimization is necessary for real-time clinical use. Finally, exploring hybrid approaches (e.g., an ensemble of a CNN for feature extraction with a CapsNet for classification) could potentially leverage the best of both architectures, as was done in cited work (AbouEl-Magd et al., 2023).

6. CONCLUSION

In this paper, we presented a comparative analysis of a Capsule Network and several Convolutional Neural Network architectures for lung cancer detection using CT scan images. Our study demonstrated that the CapsNet significantly outperformed all CNN models, achieving an overall accuracy of 98.5% and a macro F1-score of 94.9%, compared to the best CNN's (VGG16) 72.5% macro F1-score. Crucially, the CapsNet maintained high precision and recall across all lung cancer types and proved far more stable under image rotations, exhibiting superior inherent robustness to spatial transformations. These findings empirically support the touted strengths of CapsNet: they are data-efficient and inherently more robust to spatial transformations of the input. Our work suggests that CapsNet are a promising tool for lung cancer image analysis, especially in data-scarce scenarios or where robustness to image orientation is important. They could serve as a powerful complement or alternative to traditional CNNs in computer-aided diagnosis systems.

REFERENCES

1. AbouEl-Magd, L. M., Darwish, A., Snášel, V., & Hassanien, A. E. (2023). A pre-trained convolutional neural network with optimized capsule networks for chest X-rays COVID-19 diagnosis. Cluster Computing, 26, 1389–1403.
2. Afshar, P., Plataniotis, K., & Mohammadi, A. (2019). Capsule networks for brain-tumor classification based on MRI images and coarse tumor boundaries. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1368–1372).
3. Akinyelu, A. A., Zaccagna, F., Grist, J. T., Castelli, M., & Rundo, L. (2022). Brain-tumor diagnosis using machine learning, CNNs, capsule neural networks and vision transformers applied to MRI: a survey. Journal of Imaging, 8(8), 205.
4. Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung-cancer detection and classification. Multimedia Tools and Applications, 79(11–12), 7731–7762.
5. Bushara, A. R., Vinod Kumar, R. S., & Kumar, S. S. (2023a). Classification of benign and malignancy in lung cancer using CapsNet with dynamic routing algorithm on CT images. Journal of Artificial Intelligence Technology, 4(1), 40–48.
6. Bushara, A. R., Vinod Kumar, R. S., & Kumar, S. S. (2023b). An ensemble method for the detection and classification of lung cancer using CT images utilizing a capsule network with VGG. Biomedical Signal Processing and Control, 85, 104864.

7. Ferrarini, B., Ehsan, S., Bartoli, A., Leonardis, A., & McDonald-Maier, K. D. (2019). Assessing Capsule Networks With Biased Data (arXiv:1904.04555 [cs.CV]). arXiv.
8. Hany, M. (2020). Chest CT-scan images dataset. Kaggle. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
10. Jiménez-Sánchez, A., Albarqouni, S., & Mateus, D. (2018). Capsule Networks against Medical Imaging Data Challenges (arXiv:1807.07559 [cs.CV]). arXiv.
11. Kalkan, M., Guzel, M. S., Ekinci, F., Şahan, M., Çalıskan, M., Akkaya, S. T., Akdoğan, E., Telli, N., Yüksek, S., & Ekinci, E. (2024). Comparative analysis of deep-learning methods on CT images for lung cancer classification. Cancers, 16(19), 3321.
12. Li, C., Lei, S., Ding, L., Wang, Q., Luo, K., Du, W., Su, M., Cheng, Z., & Chen, J. (2023). Global burden and trends of lung cancer incidence and mortality. Chinese Medical Journal, 136(13), 1583–1590.
13. Mobiny, A., & Van Nguyen, H. (2018). Fast CapsNet for lung-cancer screening. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 (Vol. 11043, pp. 741–749). Springer International Publishing.
14. Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In Advances in Neural Information Processing Systems (pp. 3856–3866).
15. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).
16. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2818–2826).
17. Wang, S., Wei, Y., Wang, Z., Zhang, J., & Zhang, Y. (2023). Capsule network—its limitations, modifications and applications: a survey. Machine Learning and Knowledge Extraction, 5(3), 891–921.
18. Wang, Y., Wang, B., Zou, J., Yang, Z., Zhang, Z., & Gao, S. (2025). Capsule neural network and its applications in drug discovery. iScience, 28(4), 112217.
19. Zhou, Y., He, C., Peng, J., Wang, S., Sun, Y., & Zheng, B. (2024). Attention-CapsNet for lung-cancer CT classification. IEEE Journal of Biomedical and Health Informatics, 28(2), 599–610.