NTCIR-18: Proceedings of the 18th NTCIR Conference on Evaluation ofInformation Access Technologies, June 10-13, 2025, Tokyo, Japan
DOI: https://doi.org/10.20736/0002002069

NITKC at the NTCIR-18 RadNLP shared task: Using Graph-RAG in a lung cancer staging method with Natural Language Processing for Radiology

Aoi Kondo
National Institute of Technology, Kurume College
Japan
k-104_2024@kurume.kosen-ac.jp

Tan You Quan Bernon
Temasek Polytechnic
Singapore
tc-138_2024@kurume.kosen-ac.jp

Tsubasa Oka
National Institute of Technology, Kurume College
Japan
a4108to@kurume.kosen-ac.jp

Hiroaki Koga
National Institute of Technology, Kurume College
Japan
h-koga@kurume-nct.ac.jp

Mikio Oda
National Institute of Technology, Kurume College
Japan
oda@kurume-nct.ac.jp

Abstract

The NITKC team participated in the RadNLP Shared task of TNM classification from lung cancer radiology reports written in English, using an LLM-based approach. LLM accuracy varies depending on training methods and the number of parameters. We aimed to solve this task using open-source LLMs with fewer parameters than closed-source, proprietary LLMs and made improvements accordingly. Open-source LLMs have less prior knowledge than closed-source LLMs, putting them at a disadvantage for TNM classification. To address this, we used Graph-RAG to improve accuracy and address issues by representing domain knowledge for unfamiliar tasks as a graph and incorporating it as knowledge into the LLM. This method uses a graph database to represent domain knowledge for TNM classification in a graph structure. It dynamically incorporates the graph information into LLM prompts, compensating for the knowledge gaps in open-source LLMs and enabling more accurate inference. Additionally, to enhance performance, we trained BioBERT and MedBERT on a dataset labeled with lung cancer progression stages and utilized these inference results concurrently. As a result, we achieved a joint accuracy of 0.2963 in the TNM classification task. This demonstrates that our approach effectively mitigates the limitations of open-source LLMs in TNM classification.

Keywords

Medical Natural Language Processing, Large Language Model, Graph-RAG, Radiology Reports, Cancer Staging

Team Name

NITKC

Subtasks

Sub task (English track) and Main task (English track)

1 Introduction

We participated in the NTCIR-18 RadNLP 2024 shared task [14]. This is a task for automatically staging lung cancer from radiology reports. The radiology reports are written based on medical images such as CT and MRI by doctors. However, the paper [8] reported that the radiology reports do not always provide the stage. Therefore, the shared task aimed to automatically stage lung cancer based on the radiology reports.

In this shared task, there were two tasks which are related to Medical Natural Language Processing (Medical NLP). The first task is called "main task". It is to predict TNM classification which defines the severity of lung cancer. The Second task is called "sub task". It is a task to predict the eight labels related to lung cancer from each sentence of the radiology report. The datasets in this task are lung cancer reports which are post-treatment. They are made from the images Radiopaedia[1] for this shared task. Therefore, they do not include personal information.

Our method used Graph-RAG and incorporated the sub task and main task labels. RAG framework is a method to provide the LLM with additional information. The Graph-RAG algorithm is a part of the RAG framework and is efficiently designed for graph structure. Therefore, we used sub task results and provided the LLMs with structured definitions of the TMN classification via Graph-RAG.

In the experiments, we used LLM and defined an original graph in the main task. This graph incorporated results from the sub task. Finally, we ranked eighth out of 12 in the sub task and 11th out of 16 in the main task on the RadNLP 2024 shard task leaderboard.

We discuss our method and the effect of the Graph-RAG in Medical NLP via a comparison to another approach to add information to the LLM. Therefore, this paper makes the following contributions:

- Proposed a method for predicting the TMN classification from a radiology report using Graph-RAG (main task).
- Proposed a method for predicting the eight labels related to lung cancer from each sentences of the radiology reports (sub task).
- Conformed and discussed the effect of the Graph-RAG in the Medical NLP.

[1] https://radiopaedia.org/

2 Related Work

This paper uses many technologies related to NLP and references a lot of papers. This section describes the summary of the technology and the use case in our method.

Transformer

Transformer [11] is a neural network architecture. It is based on a self-attention mechanism that allows for efficient modeling of long-range dependencies without recurrent structures, unlike previous architectures such as RNN and LSTM.

In natural language processing, there are many successful models which are based on the Transformer architecture. In the field of medical NLP, these models are adopted in many studies. For example, in MedNLP-SC [12] in NTCIR-17 which was previously the shared task, many teams used transformer-based architecture. In this paper, we used BERT-based architecture and GPT architectures for solving sub task and main task, both of which are based on the Transformer architecture.

BERT

BioBERT [3] is a model trained by the corpus of papers in the field of biomedical via BERT architecture. BERT [1] is a neural network architecture using Transformer's encoder. The BERT was trained on a large corpus by using the next sentence prediction algorithm, masking strategy. MedBERT [10] is also BERT-based architecture trained by the corpus which is created from the electronic healthcare records.

The BERT is effective for sentence-level classification of the free-text radiology reports [6]. The BERT-based model has also proven effective in the field of Medical NLP. For example, previous research [5] in MedNLP-SC uses a BioBERT-base model in the experiment. In our approach, we used the MedBERT and the BioBERT for solving sub task which is a sentence-level classification problem.

GPT

GPT [7] is an architecture that uses a Transformer's decoder. LLM (Large Language Model) such as ChatGPT[2], Gemini[3], and Claude[4] has a strong generalization performance. They train very large datasets via the GPT architecture. There are two types of LLM, closed-source LLM and open-source LLM. The closed-source LLM is proprietary. It has a very strong performance but does not public information about the parameters. On the other hand, the open-source LLM is freely available and has publicly accessible parameters. We use a model part of the open-source LLM.

In NLP, The GPT is a strong method for many types of the task. The LLMs have expanded into domains that involve complex tasks, including the medical field. The GPT is valid in the field of Medical NLP. For example, previous research [2] in NTCIR-17 used GPT-3.5-Turbo[5].

We used an LLM to predict the TMN classification labels required in the main task. In this approach, we used sub task results and gave the LLM information including the definition of the TMN classification.

RAG

RAG [4] is a retrieval-based framework designed to enhance the LLM performance. The RAG allows external information to be provided to the LLM, improving prediction accuracy. There are many types of the RAG algorithms. We use the Graph-RAG algorithm. It is a graph-based algorithm and strong to field using a graph approach.

In the recent NLP field, the RAG is based on the LLM technology. It is used for developing the LLM which is able to think of reason and retrieval in the Internet information. In the medical NLP, RAG is also conformed effective. A study [9] indicates that the accuracy is growing to use RAG in medical question-answering tasks.

In our method, we adopted the Graph-RAG in the main task. The main task involves classification according to the TMN system. The TMN classification has many labels and such labels reference one another, which makes the graph-based approach appropriate.

3 Proposed Methods

3.1 Sub task

The sub task involves document segmentation to predict eight spans containing useful information for lung cancer treatment. In NLP terminology, this sub task is a multi-label binary sentence classification. The eight labels are "Pleural", "Lymphadenopathy", "Satellite", "Extension", "Measure", "Inclusion", "Atelectasis", "Distant". Their definitions involve the TMN system. In this task, we adopted three different types of models using the BERT architecture.

First, we created a MedBERT model that trained six target labels via fine-tuning. In the fine-tuning, we trained all layers of the MedBERT including the input layer, attention layer, and output layer.

Second, in the two target labels, their classification tasks are more difficult than the label classification of others. Therefore, we trained separate models using the BioBERT for some labels, and three models in total for the sub task.

3.2 Main task

The main task is a multi-label document classification to correctly determine T, N, and M categories for each radiology report. The TMN classification categories are further divided into more detailed subcategories:: "T0", "Tis", "T1mi", "T1a", "T1b", "T1c", "T2a", "T2b", "T3", "T4" and "M0", "M1a", "M1b", "M1c" and "N0", "N1", "N2", "N3". We follow the TMN classification defined by the Japan Lung Cancer Society (JLCS)[6].

Figure 1 provides an overview of our method for the main task. We used sub task results and provided the TMN classification definitions to LLM using Graph-RAG. The final output is generated by the LLM.

[2]https://chatgpt.com/
[3]https://gemini.google.com/
[4]https://claude.ai/
[5]https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/

[6]https://www.haigan.gr.jp/publication/guideline/examination/2022/1/0/220100000000.html

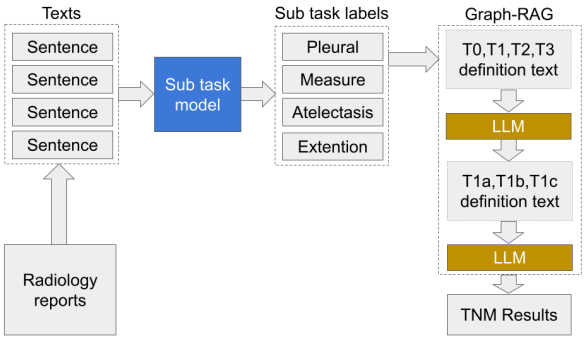

Figure 1: An overview of our method for the main task

Table 1: Our experimental setup for the sub task

Model name	A	B	C
Base model	MedBERT	BioBERT	BioBERT
Label(s)	Pleural, Lymphadenopathy, Satellite, Extension, Measure, Inclusion	Atelectasis	Distant
Batch Size	4	4	4
Epochs	10	10	10

4 Experiments

4.1 Sub task

We used the MedBERT[7] and the BioBERT[8] for predicting. The MedBERT is used to train model for the labels: "Pleural", "Lymphadenopathy", "Satellite", "Extension", "Measure", and "Inclusion". We call this model "A". The BioBERT is used to train models for the labels: "Atelectasis", and "Distant". We call these models "B" or "C". The batch size is set to four and the number of the epochs is 10 for both models. Table 1 summarizes the experimental setting.

In the training, we used the pipeline provided by HuggingFace Transformers [13]. Our fine-tuning method is defined by AutoModel, we only changed the batch size and the number of epochs, other parameters followed the default values of the TrainingArguments class of the HuggingFace[9].

For evaluation, the task uses $F_{2.0}$ score. The final score is computed sentence-wise average over the seven labels from Measure to Distant. The $F_{2.0}$ score is defined as

$$F_{2.0} = 5\frac{precision \cdot recall}{4 \cdot (precision + recall)} \quad (1)$$

4.2 Main task

For the main task, we used the Graph-RAG, which requires a predefined graph. This graph is shown in Figure 2. It follows the medical knowledge and the TMN definition by ours. In the experiments, we used the Neo4j[10] which is a graph database management system.

[7]https://huggingface.co/Charangan/MedBERT
[8]https://huggingface.co/dmis-lab/biobert-v1.1
[9]https://huggingface.co/docs/transformers/v4.50.0/en/main_classes/trainer#transformers.TrainingArguments
[10]https://neo4j.com/

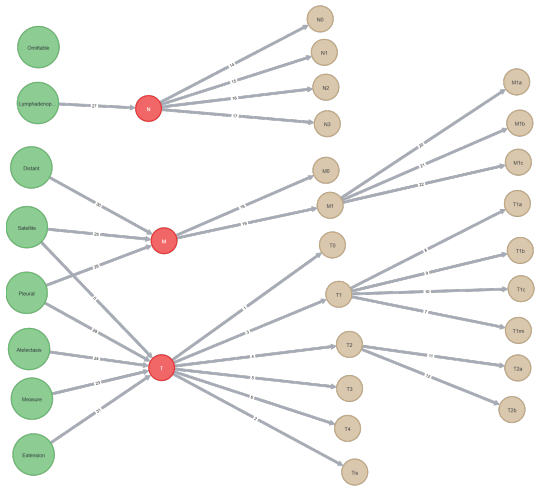

Figure 2: The graph structure

We used the Lamarck-14B-v0.7[11] LLM model. This model is a merged model combining several other foundation models.

The prompts used in our experiments are listed in the appendix section. In our approach, we provided the LLM with the TMN classification definition via the Graph-RAG. The definitions provided by the JLCS are written in Japanese and translated into English for use in the model prompts.

For evaluation, the task used an accuracy score. The final score is computed report-wise average over the several categories, the evaluation used two metrics: fine and coarse scores. They are fine scores and coarse scores. The fine score is the proportion of radiology reports with accurate predictions for all the T, N, and M factors. The coarse score ignores distinctions between detailed subcategories. The accuracy score is defined as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

5 Results

5.1 Sub task

In the sub task, we obtained the scores shown in Table 2. We used three models "A", "B", and "C". Model "A" provides scores for six labels, while models "B", and "C" each provide a score for one label. In the final submission score, the Lymphadenopathy $F_{2.0}$ score was a taken from the model "B" score, and the Pleural $F_{2.0}$ score was a token from the model "C" score. As a result, we ranked eighth out of 12 on the leaderboard of the RadNLP 2024 Shared task.

5.2 Main task

In the main task, we obtained the scores shown in Table 3. Our final submission result which is listed on the leaderboard of the RadNLP 2024 Shared Task, corresponds to the joint accuracy in the fine score evaluation. We ranked 11th out of 16 on the leaderboard.

[11]https://huggingface.co/sometimesanotion/Lamarck-14B-v0.7

Table 2: $F_{2.0}$ scores of the sub task

Model name	A	B	C	Result
Inclusion	0.947	––	––	0.947
Measure	0.735	––	––	0.735
Extension	0.741	––	––	0.741
Atelectasis	0.877	––	––	0.877
Satellite	0.635	––	––	0.635
Lymphadenopathy	––	0.927	––	0.927
Pleural	––	––	0.905	0.905
Distant	0.786	––	––	0.786

Table 3: Accuracy scores of the main task

Evaluation type	Fine	Coarse
Joint accuracy	0.296	0.482
T accuracy	0.457	0.642
N accuracy	0.864	0.864
M accuracy	0.778	0.815

Table 4: Comparison between Graph-RAG approach and Long-Context approach in validation dataset

Approach	**Graph-RAG**		Long-Context	
Evaluation type	Fine	Coarse	Fine	Coarse
Joint accuracy	0.500	0.667	0.273	0.527
T accuracy	0.611	0.796	0.473	0.746
N accuracy	0.907	0.907	0.764	0.764
M accuracy	0.852	0.889	0.782	0.837

6 Discussions

We discuss a valid effectiveness of our proposed method. In the main task, we adopted the Graph-RAG using the graph which includes information on the TMN classification and the relationship between the TMN system and labels of the sub task. In the results, we got nearly 0.3 accuracy score. However, it does not indicate the validity of the Graph-RAG in the field of Medical NLP. Therefore, we confirm that the Graph-RAG approach is better than another approach which is called Long-Context in this task.

The Long-Context approach and the RAG approach are too similar. The Graph-RAG provides additional information following the graph but the Long-Context approach is a method providing the LLM with all information at the same time.

Table 4 shows the comparison between the Graph-RAG and the Long-Context in the validation datasets. Validation datasets are a part of the datasets. The Table 4 shows that the Graph-RAG score is higher than the Long-Context approach. Therefore, our Graph-RAG is a better than normal Long-Context approach. We think that the Graph-RAG is valid in predicting the structured system similar to the TMN system.

7 Conclusions

In this paper, we proposed methods for the RadNLP 2024 shared task. The sub task was a sentence-level classification task involving the TMN system, defined by eight labels. We employed BERT-based models, BioBERT and MedBERT, fine-tuned by the HuggingFace pipeline. The main task was a report-level classification problem targeting lung cancer staging according to the TMN system. To address this, we propose a novel approach using the Graph-RAG for predicting these labels. In the Graph-RAG approach, we successfully integrated sub task results into the main task via a graph structure.

In the main task, we obtained the score predicted by the TMN classification using our method. In the sub task, we obtained the score predicted by the eight labels using our method. In the leaderboard of the RadNLP 2024 Shared Task, we ranked eighth out of 12 in the sub task and 11th out of 16 in the main task. In the discussion, we discuss a comparison between the Graph-RAG and the Long-Context. It indicated that our Graph-RAG approach is better than the normal Long-Context approach in this task.

In future work, we plan to further explore the method enhanced to the LLM. The Graph-RAG approach also can provide reasons for the LLMs. The medical AI is required to provide reasons to the users. Therefore, we will explore the method for predicting the TMN system and provide the reasons why LLM predicts one label.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186.

[2] Smilla Fox, Martin Preiß, Florian Borchert, Aadil Rasheed, and Matthieu-P Schapranow. 2023. HPIDHC at NTCIR-17 MedNLP-SC:Data Augmentation and Ensemble Learning for Multilingual Adverse Drug Event Detection. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. 185–192.

[3] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (Feb. 2020), 1234–1240.

[4] Patrick Lewis, Ethan Perez, Aleksandara Piktus, F Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, M Lewis, Wen-Tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Neural Inf Process Syst* abs/2005.11401 (May 2020), 9459–9474.

[5] Hsiao-Chuan Liu, Vidhya Nataraj, Chia-Tung Tsai, Wen-Hsuan Liao, Tzu-Yu Liu, Mike Tian-Jian Jiang, and Min-Yuh Day. 2023. IMNTPU at the NTCIR 17 Real MedNLP Task: Multi Model Approach to Adverse Drug Event Detection from Social Media. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. 193–199.

[6] Daiki Nishigaki, Yuki Suzuki, Tomohiro Wataya, Kosuke Kita, Kazuki Yamagata, Junya Sato, Shoji Kido, and Noriyuki Tomiyama. 2023. BERT-based transfer learning in sentence-level anatomic classification of free-text radiology reports. *Radiol. Artif. Intell.* 5, 2 (March 2023), e220097.

[7] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. *OpenAI* (June 2018).

[8] Raphael Sexauer, Thomas Weikert, Kevin Mader, Andreas Wicki, Sabine Schädelin, Bram Stieltjes, Jens Bremerich, Gregor Sommer, and Alexander W Sauter. 2018. Towards more structure: Comparing TNM staging completeness and processing time of text-based reports versus fully segmented and annotated PET/CT data of non-small-cell lung cancer. *Contrast Media Mol. Imaging* 2018 (Nov. 2018), 5693058.

[9] Jiwoong Sohn, Yein Park, Chanwoong Yoon, Sihyeon Park, Hyeon Hwang, Mujeen Sung, Hyunjae Kim, and Jaewoo Kang. 2024. Rationale-Guided retrieval augmented generation for medical question answering. *arXiv [cs.CL]* (Oct. 2024).

[10] Charangan Vasantharajan, Kyaw Zin Tun, Ho Thi-Nga, Sparsh Jain, Tong Rong, and Chng Eng Siong. 2022. MedBERT: A pre-trained language model for biomedical named entity recognition. In *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł Ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I Guyon, U Von Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett (Eds.),

Vol. 30. Curran Associates, Inc.

[12] Shoko Wakamiya, Lis Kanashiro Pereira, Lisa Raithel, Hui-Syuan Yeh, Peitao Han, Seiji Shimizu, Tomohiro Nishiyama, Gabriel Herman Bernardim Andrade, Noriki Nishida, Hiroki Teranishi, Narumi Tokunaga, Philippe Thomas, Roland Roller, Pierre Zweigenbaum, Yuji Matsumoto, Akiko Aizawa, Sebastian Möller, Cyril Grouin, Thomas Lavergne, Aurélie Névéol, Patrick Paroubek, Shuntaro Yada, and Eiji Aramaki. 2023. NTCIR-17 MedNLP-SCSocial Media Adverse Drug Event Detection: Subtask Overview. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. none.

[13] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Stroudsburg, PA, USA, 38–45.

[14] Yuta Nakamura, Koji Fujimoto, Jonas Kluckert, Michael Krauthammer, Jun Kanzawa, Akira Katayama, Tomohiro Kikuchi, Ryo Kurokawa, Wataru Gonoi, Peitao Han, Kiyoto Hashimoto, Yuki Tashiro, Shouhei Hanaoka, Shuntaro Yada, and Eiji Aramaki. 2025. NTCIR-18 RadNLP 2024 Overview: Dataset and Solutions for Automated Lung Cancer Staging. In *In Proceedings of the NTCIR-18 Conference*.

A Appendix

In the Appendix section, it shows the prompt for solving the main task in our method. We use LangChain[12] and define separate of the prompts in the experiments. The main prompts section shows the original prompts that we give to the LungChain. The TMN classification definition prompts show the TMN classification definition that we give to the LLM via the LungChain.

A.1 Main prompts

The Listing 1 shows the original prompts of the main task. In the values of the prompts, it is replaced with the TMN classification definition following the graph via the Graph-RAG by the LungChain.

There are some values in the main prompts:

- The "limitation_list" in the prompt is a value for teaching a range of the answers to the LLM. For example, in the T label classification phase, "limitation_list" is replaced by "T0, T1, T2, T3, T4, Tis" by the LungChain.
- The "graph_text" in the prompt is replaced by the TMN classification via the Graph-RAG by the LungCain.
- The "question" is also replaced the radiology reports for input.

Listing 1: Prompt of the main task

```
Task: Please answer the question referring
    to the preprocess below.

Instructions:
- Please only answer and do not repeat
    questions or task statements.
- Please output only one of "{
    limitation_list}" in your answer.
- Please do not add any information other
    than what is provided.
- Do not include copies of questions or
    explanations in your output.

preprocess:{graph_text}

input:{question}

question:
Of the classifications defined in preprocess
    , what category does the data entered as
     input fall under?
Please output only the answer. Please do not
     output unnecessary information.
You are only outputting one answer out of "{
    limitation_list}".
Answers such as "None" are not accepted.

answer:
```

[12] https://www.langchain.com/

A.2 TMN classification definition prompts

This section shows the definition of the TMN classification given to the LLM via the Graph-RAG. The "T" label and the "M" label have more detailed categories. So when LLM predicts the label of the existing detail label, the LungChain is provides more detailed definition following the Graph structure. The Listing 2 shows the definition of the "T" classification and the Listing 3, the Listing 4 shows more detailed definition of the "T1" label and "T2" label. And also, the Listing 5 shows the definition of the "M" classification, and the Listing 6 shows more detailed definition of the "M1" label. the Listing 7 shows the definition of the "N" classification.

Listing 2: Definition of the T classification

```
T0: No primary tumor
Tis: carcinoma in situ: in the case of lung
    field type, the diameter of the full
    component is 0 cm and the diameter of
    the entire lesion is <= 3 cm.
T1: Tumor diameter <= 3 cm, covered by lung
    or visceral pleura, no bronchoscopic
    evidence of central invasion beyond the
    lobar bronchus (i.e., not extending into
     the main bronchus)
T2: The diameter of the bronchus is >3 cm
    and <=5 cm, or the diameter of the
    bronchus is <=3 cm, but either of the
    following is present:
- Involvement of the main bronchus but not
    the tracheal bifurcation
- Involvement of the pleura on the visceral
    side
- Partial or total unilateral atelectasis or
     obstructive pneumonia extending to the
    pulmonary hilum.
T3: Full diameter > 5 cm and <= 7 cm, or
    full diameter <= 5 cm but one of the
    following:
- Direct involvement of the lateral pleura,
    chest wall (including superior sulcus
    tumor), diaphragmatic nerve, or
    pericardium
- Discontinuous paraneoplastic nodules in
    the same lobe
T4: Full component >7 cm in diameter, or
    involvement of the diaphragm,
    mediastinum, heart, great vessels,
    trachea, recurrent nerve, esophagus,
    vertebral body, or tracheal bifurcation
    of any size, or a nodule of a secondary
    tumor in a different lung lobe on the
    same side.
```

Listing 3: Definition of the T1 classification

```
T1mi: Minimally invasive adenocarcinoma:
    Partially enhancing type, diameter of
    the enhancing component <= 0.5 cm and
    total lesion diameter <= 3 cm
T1a: Diameter of the substantial component
    <= 1 cm and not equivalent to Tis or
    T1mi
T1b: Diameter of the full component > 1 cm
    and <= 2 cm
T1c: Diameter of the full component > 2 cm
    and <= 3 cm
```

Listing 4: Definition of the T2 classification

```
T2a: diameter of the substantial component >
     3 cm and <= 4 cm
T2b: diameter of the full component > 4 cm
    and <= 5 cm
```

Listing 5: Definition of the M classification

```
N0: No regional lymph node metastasis
N1: Metastasis to ipsilateral peribronchial
    and/or ipsilateral hilar or
    intrapulmonary lymph nodes, including
    direct invasion of primary tumor
N2: Metastasis to ipsilateral mediastinal
    and/or subbronchial lymph nodes
N3: Metastases to the contralateral
    mediastinum, contralateral pulmonary
    hilum, ipsilateral or contralateral
    anterior scalene muscle, or
    supraclavicular fossa lymph nodes
```

Listing 6: Definition of the M1 classification

```
M1a: Paraneoplastic nodule in contralateral
    lung, pleural or pericardial nodule,
    malignant pleural effusion (ipsilateral
    or contralateral), malignant pericardial
     effusion
M1b: Single distant metastasis to one organ
    other than lung
M1c: Multiple distant metastases to one or
    more organs other than lung
```

Listing 7: Definition of the N classification

```
M0: No distant metastasis
M1: Distant metastasis
```