

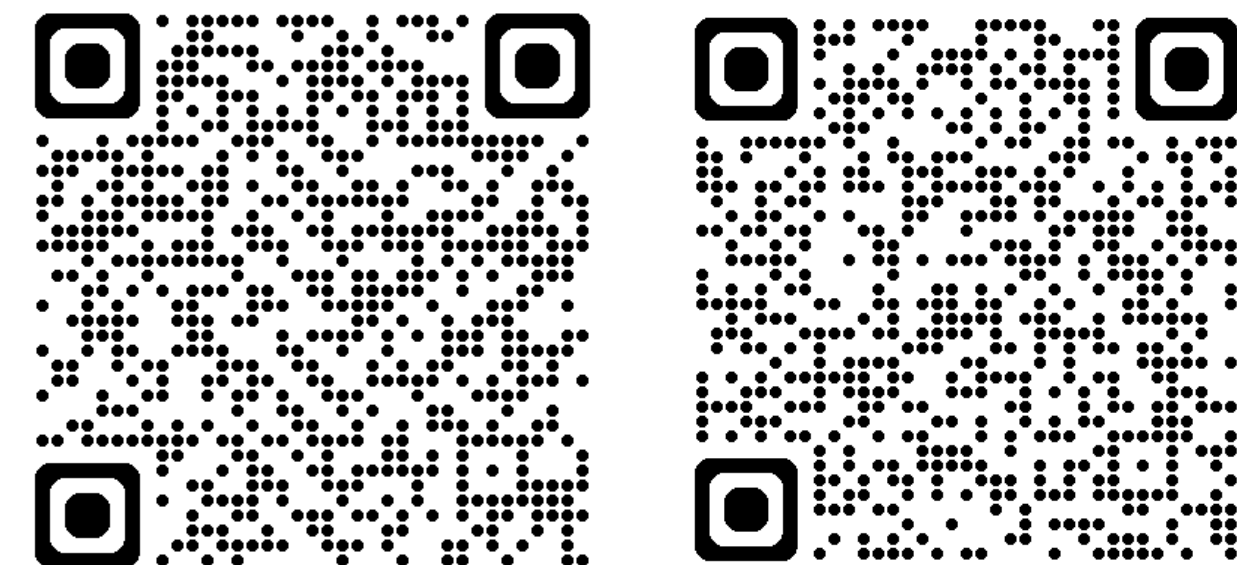
Exploring Large Language Model (LLM) for TNM Categorizing and Re-categorizing Nasopharyngeal Carcinoma (NPC) from Structured Text Reports

Junru Zhong¹, John Tsz-ho Chow², Kwok-yan Li³, Ho-sang Leung², Qi-yong Ai⁴, Ann King¹, Lun M Wong¹

¹Dept. of Imaging and Interventional Radiology, The Chinese University of Hong Kong, ²Dept. of Imaging and Interventional Radiology, Prince of Wales Hospital, Hospital Authority New-Territory

East Cluster, ³Department of Radiology, Princess Margaret Hospital, Hospital Authority Kowloon West Cluster, ⁴Dept. Diagnostics Radiology, The University of Hong Kong

Contact: jrzhong@link.cuhk.edu.hk (J Zhong); lun.m.wong@cuhk.edu.hk (L M Wong)



CUHK DIIR

ePoster

Copies of this poster obtained through QR, AR and/or text key codes are for personal use only and may not be reproduced without written permission of the authors.

BACKGROUND

- Radiology reports** of NPC MRI is an important step for their management.
- Patients are concisely characterized through assigning a TNM category
- These categories are defined rigorously based on international guidelines UICC/AJCC, and is important metrics to describe any research cohort.
- However, these guideline improves and revises over time, requiring updates for existing data, which would be laborious for large datasets.
- Large language models (LLM)** evolve quickly with emerging reasoning capability that showed promises for complex task.
- We are therefore interested to know, if the original text report has sufficient details reported, can LLM follow new guideline and perform re-categorization?

Research Question

- Can LLM perform **TNM categorisation** and **re-categorise** according to the guidelines based on structured report of NPC patients?

METHODS

Patients and Data

- 379 NPCs with pre-treatment structured MRI reports for LLM categorising according to the 7th edition UICC/AJCC guidelines. Use the original categories as reference.
- A subset of 100 reports for recategorise to 8th edition UICC/AJCC guidelines. Use the categories by senior radiologists as the reference.

Task setting

- LLMs categorise TNM stages from structured head-and-neck MRI reports.

LLM prompting

- System prompt:** defined task background, instruction, and output format.
- Initial categorising message:** 7th edition UICC/AJCC guidelines, followed by the structured head-and-neck MRI reports.
- Recategorising message:** difference between 7th and 8th UICC/AJCC guidelines, followed by the structured head-and-neck MRI reports.

LLM selection

- General purpose models:** Gemini 2.5 Flash, Gemini 2.5 Pro, Claude 3.5 Haiku, Claude 3.5 Sonnet, DeepSeek V3 0324, DeepSeek R1, GPT 4o Mini, GPT 4o, Gemma 3
- Medically trained model:** MedGemma, HuatuoGPT-o1

Performance evaluation

- Reference standards:** Initial 7th staging categories recommendation from the health records (blinded to the LLM), 8th re-categorisation was performed by a senior radiologist.
- Human performance:** Two radiologist trainees performed recategorising. Their results were compared with LLMs.
- Accuracy calculation:** Accuracy was calculated for each category (T & N) between LLM predicted and reference values. **For intermediate categories by human radiologists (both junior and senior), we followed the clinical principle to identify the lower stage as the patients' category (e.g., "T1 OR T3" → T1).**

RESULTS

- General-purpose LLMs accuracies for initial (T=40.06%-85.75%, N=33.65%-84.17%) and recategorize (T=39.29%-85.00%, N=23.81%-82.00%).
- Medical-specified LLMs accuracies for initial (T=40.37%-49.08%, N=31.40%-46.70%) and recategorise (T=31.00%-36.00%, N=38.00%-44.00%).
- Best initial 7th categorisation and 8th recategorisation in average accuracy of T and N categories were Claude 3.5 Sonnet (T=84.70%, N=84.17%) and DeepSeek V3 (T=84.00%, N=81.00%).
- Human readers performance on recategorising: T=89%-93%, N=88%-90%.

CONCLUSION

- Large (>200B) general-purpose LLMs showed potential for both categorisation and recategorisation tasks, but not for small size models.
- Medical-specified LLMs had no advantage over other models.
- Human readers performance remains superior for recategorisation, while LLM suffered from hallucinations.

Example case 1

#1123, T2w Fat Suppressed, no contrast. **Reference categories: 7th – T4, N2; 8th – T3, N2.**

Original Report (keywords for recategorisation are highlighted):

PRIMARY TUMOUR:

There is a very large primary nasopharyngeal carcinoma, the bulk of which lies deep to the nasopharynx. The tumour is involving predominantly the right side of the nasopharynx with the most bulky component in the region of the fossa. There is invasion into the right side of the nasal cavity (direct anterior extension as well as via the sphenopalatine foramen). There is extensive inferior extension into the right side of the oropharynx to involve the right palatine tonsil and right side of the soft palate. Medially, the palatine component extends to the midline and also involves the uvula; while laterally, it wraps around the posterior aspect of the superior alveolar process of the maxilla and extends into the buccal fat space, as well as inferiorly to the region of the retromolar trigone.

There is extensive right parapharyngeal region tumour invasion to involve the parapharyngeal fat space, medial and lateral pterygoid muscles. A bulky component also extends posteriorly into the retropharyngeal soft tissues (just crosses the midline to the left side also). The posterolateral component of the tumour is continuous with the retropharyngeal metastatic lymph nodes which is displacing the right internal carotid artery, which appears encased in several sections. There is a large right otomastoid effusion.

There is extensive bony invasion at the following sites:

- Skull base to **involve the right pterygoid bone**, body and right wing of the sphenoid, full length of the right side of the clivus down to the foramen magnum and right occipital condyle and right petrous apex. There is invasion of the right foramen rotundum, ovale, and lacerum and the vidian canal. There is involvement of the right pterygopalatine fossa, right pterygomaxillary fissure, right sphenopalatine foramen, right pterygopalatine canal, and there is early invasion of the right inferior orbital fissure. Invasion is also present in the lower skull base to involve the right hypoglossal nerve canal and jugular fossa (the right sigmoid and transverse venous sinus may be compressed/involved).
- There is suspicious early invasion through the posterior aspect of the right maxillary sinus and there is tumour invasion into the posterior aspect of the right alveolar process.

The tumour invades the floor of the sphenoid sinus. There is also suspicious early involvement of the posterior aspect of the right maxillary sinus and the tumour abuts the right posterior ethmoid air cells. Note is also made of unusual mucosal thickening along the more anterior medial wall of the right maxillary sinus where tumour cannot be excluded. There is non-specific but suspicious enhancement along both cavernous sinuses and dura of the right middle cranial fossa. The tumour extends into the right infratemporal fossa. No gross invasion of the orbits. The hypopharynx is unremarkable.

NODAL METASTASES:

Right side of the neck:

There are metastatic nodes in the following sites: retropharyngeal (which measure up to 3.3 cm and extend down to the level of the lower oropharynx); upper and mid internal jugular chain (levels II, III), where there are multiple matted nodes measuring up to 2.7 cm in maximum dimension. There are small indeterminate submandibular and lower internal jugular nodes. No definite supraclavicular lymphadenopathy.

Left side of the neck:

There is a small metastatic node in the left upper internal jugular chain posterior to the vein (level II), which although not enlarged, shows extracapsular spread.

There are multiple further small non-enlarged nodes in the left side of the neck, involving the retropharyngeal, upper and mid internal jugular chain (levels II, III) and posterior triangle (level Vb). These nodes are indeterminate but should be regarded with suspicion.

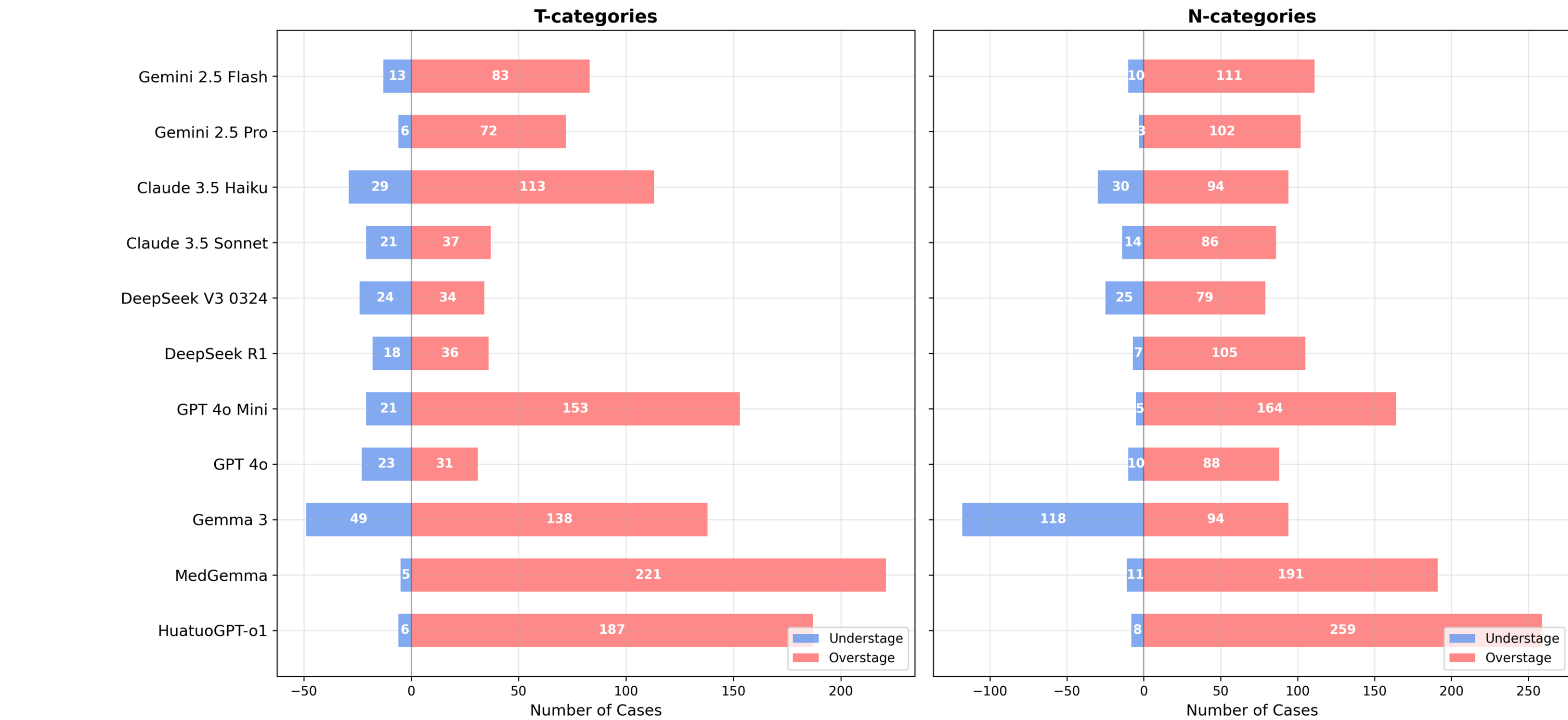
DISTANT METASTASES:

No distant metastases or second primary tumours are identified in the head and neck, upper thorax or bones.

LLM (GPT4o) categories:

Initial categorise (7th) – ☒ T4, ☒ N2; recategorise (8th) – ☒ T4, ☒ N2

- LLM (GPT4o) reasons for recategorise ☒ T4: **Tumor with extensive bony invasion, involvement of cranial nerves, and extension to infratemporal fossa.**
[☒ **Incorrect, involvement of pterygoid bone is down-rated as T3 in 8th (unless there were lateral infiltration beyond it). LLM hallucinated on cranial nerves, which did not appeared in the report.**]
- LLM (GPT4o) reasons for recategorise ☒ N2: **Bilateral lymph node involvement, all <= 6cm, above caudal border of cricoid cartilage.**
[☒ **Correct, bilateral nodes <= 6cm remains above the caudal border of the cricoid cartilage remained N2.**]



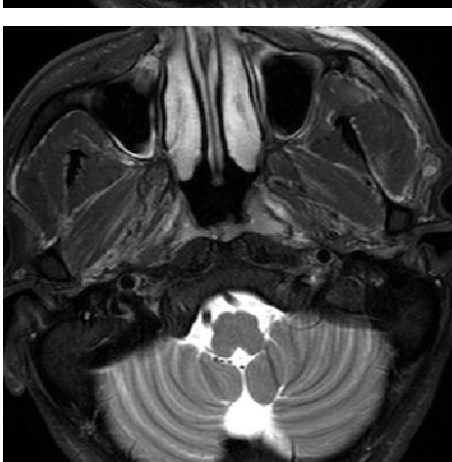
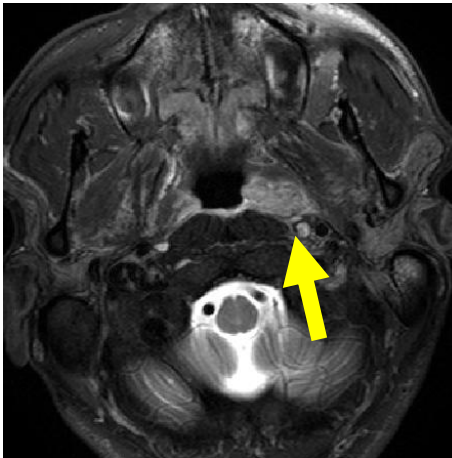
Over and under categorise cases in initial categorisation (7th edition)

Table: LLM and Reader Categorization Accuracy

Model/reader	Type	Model size	Initial categorising		Recategorising	
			T	N	T	N
Gemini 2.5 Flash	General	~200B	74.67%	78.89%	75.00%	74.00%
Gemini 2.5 Pro	General	~20B	79.42%	81.00%	78.00%	79.00%
Claude 3.5 Haiku	General	~20B	62.53%	72.56%	55.00%	60.00%
Claude 3.5 Sonnet	General	~175B	84.70%	84.17%	79.00%	76.00%
DeepSeek V3 0324	General	671B	84.62%	82.49%	84.00%	81.00%
DeepSeek R1	General	685B	85.75%	80.74%	83.00%	82.00%
GPT 4o Mini	General	~20B	53.72%	57.52%	48.00%	60.00%
GPT 4o	General	~200B	83.98%	80.71%	85.00%	67.00%
Gemma 3	General	12B	40.06%	33.65%	39.29%	23.81%
MedGemma	Medical	27B	40.37%	46.70%	36.00%	44.00%
HuatuoGPT-o1	Medical	8B	49.08%	31.40%	31.00%	38.00%
Reader 1	Human	N/A	N/A	N/A	93.00%	88.00%
Reader 2	Human	N/A	N/A	N/A	89.00%	90.00%

Example case 2

#1094, T2w Fat Suppressed, no contrast. **Reference categories: 7th – T1, N1; 8th – T2, N1.**



Original Report (keywords for recategorisation are highlighted):

PRIMARY TUMOUR:

There is a nasopharyngeal carcinoma involving predominantly the left Fossa of Rosenmuller, measuring 23x13mm. The nasal cavity and oropharynx are unremarkable. There is **bulging of the tumour to the left parapharyngeal space, but no definite invasion**. There are no middle ear or mastoid effusions. The skull base is unremarkable. The pterygopalatine fossae and pterygomaxillary fissures are preserved.

The cavernous sinus and sections of the cranium included on the scan are unremarkable. The orbit, infratemporal fossa and hypopharynx are unremarkable.

NODAL METASTASES:

There are two enlarged nodes in the **left** retropharyngeal region (9mm and 8mm) with central necrotic components. There are also metastatic nodes in the **left** upper and mid jugular chain, and left posterior triangle. The largest one measures 21mm in left upper jugular level.

Small nodes in the right upper jugular chain. No enlarged SCF/low jugular nodes.

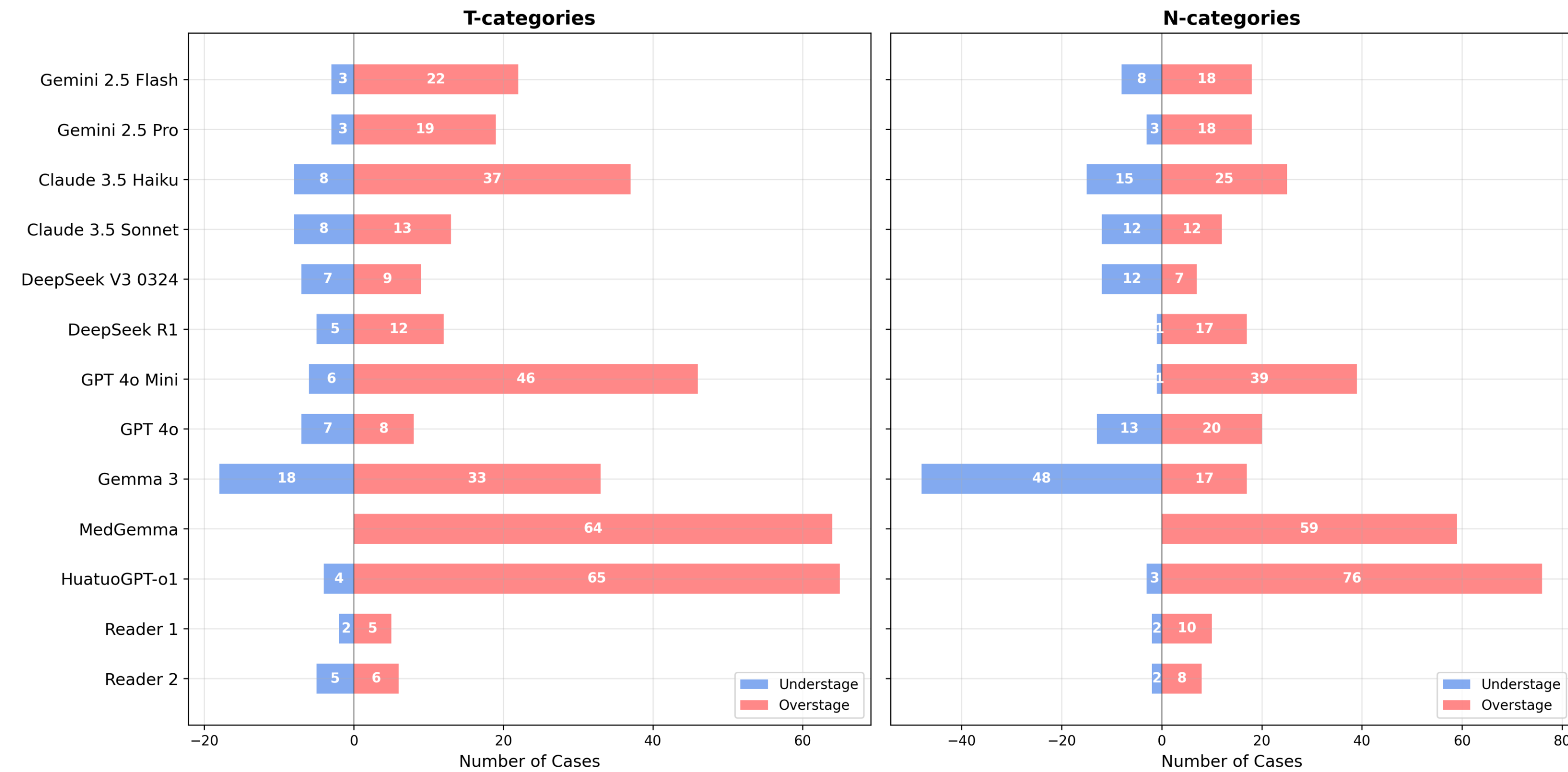
DISTANT METASTASES:

No distant metastases or second primary tumours are identified in the head and neck, upper thorax or bones.

LLM (GPT4o) categories:

Initial categorise (7th) – ☒ T2, ☒ N2; recategorise (8th) – ☒ T2, ☒ N2

- LLM (GPT4o) reasons for recategorise ☒ T2: **Tumor with parapharyngeal extension without definite invasion.**
[☒ **Correct, parapharyngeal space involvement is up-rated to T2**]
- LLM (GPT4o) reasons for recategorise ☒ N2: **Bilateral retropharyngeal nodes and left cervical nodes, all <= 6cm, above caudal border of cricoid cartilage.**
[☒ **Incorrect. Only left-side of the nasopharynx so "bilateral" conclusion is hallucination**]



Over and under categorise cases in recategorisation (8th edition)