LLM in Personalized medicine for Lung Cancer Detection

Sanjuktarani Jena[1]
Department of Computer Science Engineering
Sardar Patel Institute of Technology
Mumbai, India
sanjuktarani.jena@spit.ac.in

Naman Badlani[2]
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
naman.badlani@spit.ac.in

Upendra Kumar[3]
upendra.ietlko@gmail.com

Sai Kiran Oruganti[4]
Lincoln University College
Malaysia
saisharma@lincoln.edu.my

***Abstract*—The problem of lung cancer has been pervasive in the past several years. Cancerous cells are highly difficult to treat due to their unique nature based on the constitutional build-up of each individual patient. Thus, there is an important need to come up with a personalised form of lung cancer treatment that can tackle this issue. In this paper a novel technique has employed combining Vision Transformers (ViT's) and large language models (LLMs). The LLMs have been fine tuned with the Chain-of-Thought prompt engineering technique by training on specialised question answering datasets. Finally, a novel data bank of personalised medical history of real world patients in order to validate our model.**

***Index Terms*—Large language model,Vision Transformer, GPT, Chain of Thoughts, Retrieval Augmented Generation**

I. INTRODUCTION

Lung cancer still ranks among the leading causes of cancer-related deaths globally; therefore, early and accurate detection increases the chance of improving patient outcomes. AI and ML have brought an unprecedented ability to improve medical diagnostics. Furthermore, the recent rise of large language models proposes new avenues for personalized medicine in early cancer detection. Large language models that can handle massive amounts of clinical, imaging, and genomic data actively guide the physician in early diagnosis, risk stratification, and the decision for management. Traditional diagnostic methods utilize imaging techniques like CT scans and biopsies—which are often valuable but can be hampered through subjectivity and accessibility barriers. The incorporation of LLMs in lung cancer detection can aid in bridging such gaps since it gives room for precise, data-driven insights, which are aligned with medical literature, patient records, and real-time clinical data. Additionally, it can allow for patient-specific recommendations with a genetic basis by analyzing genetic markers, lifestyle factors, and treatment responses—thereby advancing personalized medicine. This paper discusses the role of LLMs in lung cancer detection, their incorporation into clinical workflows, advantages, and brought forth challenges. In the discussion, we elaborate on the prospect of enhancing diagnosis accuracy and treatment pathways while contributing to personalized healthcare solutions. Interpretive analysis of the model, ethical issues, and regulatory compliance are also acknowledged in the context of LLMs in clinical settings. Having AI-generated insights, LLMs could very well revolutionize lung cancer detection and treatment approaches, with a good end. This research will be conducted through a systematic review of cancer related papers with different large language models .This paper will be helping researchers to find research gap and the different techniques used in previous research. The review will be helpful in proceeding the new research as well. The organization of the paper will be as follows: Section 2: Methodology used for conducting the review; Section 3: Related work contribution; Section 4: Results & Discussion analysis of answer to the research questions pose as general or specific for the review; Section 5: Conclusion and Limitation of the article.

II. METHODOLOGY

The review research was carried out with selection of data source with strategy for selecting the sources. The author searched for different database on medicine with specific scopus indexed journal and conference publishers like Springer, IEEE, PubMed, Elsevier, Web of Sciences publications. Google scholar and PubMed search engines were used to get specific article in medical or clinic. Different document types journal and conferences had been chosen for selecting the papers for review. Total of 33 articles and conference papers had been retrieved for review.

In this review author answered the following research questions:

A. General Questions

- What are the datasets used for carrying out research in the NLP domain?
- What are the techniques used to extract data in NLP?
- What are the modern technologies used in the healthcare domain for NLP?

B. *Specific Questions*

- Which large language models are used for cancer detection?
- Which modality of dataset is chosen such as text dataset or image dataset?
- Which prompting systems are used for fine tuning the model?

III. Related Works

Krishnan and Krishnan et al [1] explored the use of Vision Transformers (ViTs) for detecting COVID-19 from chest X-ray images. They fine-tuned a pretrained ViT model on a dataset of chest X-rays, achieving an accuracy of 97.61%, a precision of 95.34%, a recall of 93.84%, and an F1-score of 94.58%. The study demonstrated that ViTs could effectively capture global context and spatial relationships in medical images, outperforming traditional convolutional neural networks (CNNs).

Van Veen et al. [2] investigated the performance of adapted Large Language Models (LLMs) in summarizing clinical texts. Domain-specific adaptation methods are applied to LLMs:for the tasks radiology reports, patient questions, progress notes, and doctor-patient dialogues. Ten physicians were involved for evaluation of the summaries based on completeness, correctness, and conciseness. As a result the adapted LLMs' summaries were either equivalent (45%) or superior (36%) to those generated by medical experts.

Angara and Thirunagaru et al [6] explored how well Vision Transformers (ViTs) can detect pneumonia from chest X-ray images. By leveraging transfer learning with pretrained ViT models, they achieved an impressive accuracy ranging from 98.75% to 99.5%. Their findings emphasize the ability of ViTs to recognize intricate patterns in medical images, making them a promising tool for pneumonia diagnosis. When it comes to multi-organ and cardiac segmentation, transformer-based models like TransUNet, MISSFormer, and Segtran have shown superior performance compared to traditional CNNs. TransUNet blends transformers with the U-Net architecture, using a hybrid CNN-Transformer encoder to preserve fine spatial details while capturing a broader context. MISSFormer enhances segmentation accuracy by incorporating an improved transformer block that balances long-range dependencies with local context, proving effective in tasks like multi-organ and cardiac segmentation. Segtran takes a unique approach with its Squeeze-and-Expansion transformer, specifically designed for medical image segmentation. This architecture efficiently captures global context while preserving high spatial resolution, leading to better segmentation results in applications such as optic disc/cup segmentation and polyp detection.

Tozuka et al [8] performed using GPT-4 Omni (GPT-4o), a gold-standard large language model, both with and without the REK. Unlike NotebookLM, which used retrieval-augmented generation (RAG), GPT-4o received the REK directly within the prompt. The results showed that NotebookLM significantly outperformed GPT-4o in the lung cancer staging experiment, achieving an impressive 86% diagnostic accuracy. In comparison, GPT-4o reached only 39% accuracy when given the REK and just 25% without it. Additionally, NotebookLM excelled in locating reference information within the REK, demonstrating 95% accuracy in retrieving relevant details.

Schmidl, B et al [9] proposed ChatGPT 4.0 which was successfully identified leukoplakia cases using image recognition However, it couldnot diagnose squamous cell carcinoma (SCC). The improved accuracy was achieved with clinical history in the prompt.

Ralevski et al [10] evaluated GPT-3.5 and GPT-4'sability , from 25,217 clinical notes on 795 pregnant women. The performance of these LLMs Compared with GPT-3.5 and the NER model, GPT-4 out performed with higher recall (0.924) than human annotators (0.702),However, while GPT-4 had strong recall, its precision (0.850) was lower than that of human annotators (0.971). A chain-of-thought prompt had been designed for improving the accuracy that required the LLMs to provide evidence and justification for each identified case .The results showed that GPT-4 outperformed GPT-3.5 and the NER model, achieving the highest recall (0.924) in detecting housing instability—significantly higher than human annotators (0.702). However, while GPT-4 had strong recall, its precision (0.850) was lower than that of human annotators (0.971), meaning it identified more relevant cases but also produced more false positives.

J Wei et al [11] Compared three CoT models GSM8K, SVAMP, and MAWPS with different LLM model such as LaMDA, GPT, PaLM where PaLM models outperformed with scaling.

D Oniani et al [12] generated synthetic patient descriptions and conducted a study to evaluate the responses of four large language models (LLMs)—GPT-4, GPT-3.5 Turbo, LLaMA, and PaLM 2—through both automatic and human assessments. The results showed that all four LLMs performed better when supplemented with Clinical Practice Guidelines (CPGs) compared to using Zero-Shot Prompting (ZSP) alone. Among the different prompting strategies, Binary Decision Tree (BDT) achieved the highest performance in automatic evaluations, outperforming Chain-of-Thought Few-Shot Prompting (CoT-FSP) and Program-Aided Graph Construction (PAGC). Human evaluations also confirmed strong overall performance across all methods. Ultimately, LLMs enhanced with CPGs provided more accurate recommendations than their plain ZSP counterparts.

Y Hsueh, J et al [13] A study analyzed postoperative complications from 944 renal surgeries performed between August 2005 and March 2022, comparing GPT-4's analysis with a human-curated dataset. The results showed a high match rate of 79.6% between GPT-4 and human assessments. Individually, both methods demonstrated strong accuracy in identifying complications, with GPT-4 achieving 86.7% accuracy and the human-curated dataset reaching 92.9%. However, GPT-4's occasional misclassifications were attributed to the way information was provided in the patient's prompt, highlighting the model's sensitivity to input phrasing and context.

A Kumar et al [14] ViT-based system shows encouraging results with 93% training accuracy and 91% robust testing accuracy in comparison with various CNN architectures such as ResNet50, MobileNet, and EfficientNet,with test accuracy 84%, 85%, 82% respectively These remarkable results highlight the revolutionary potential of Vision Transformers in transforming the detection of breast cancer and related activities.

D Ferber [15] evaluated the model GPT-4 with Vision capabilities (GPT-4V) with few shot sampling (Zero, three, five) on cancer image processing. Cancer histopathology tasks like: Classification of tissue subtypes in colorectal cancer, colon polyp subtyping and breast tumor detection in lymph node had been performed with this in-context learning models. GPT4-V result was outperformed by neural networks like KNN, and Random Forest trained for particular tasks.

Xuxin Chen et al. [16] introduced Mammo-CLIP, evaluated through 5-fold cross-validation. The model was fine-tuned and internally assessed using a dataset of 470 malignant and 479 benign cases. To test its generalizability, a separate dataset comprising 60 malignant and 294 benign cases was utilized. Mammo-CLIP demonstrated superior performance over the state-of-the-art cross-view transformer, achieving higher AUC scores (0.841±0.017 vs. 0.817±0.012 and 0.837±0.034 vs. 0.807±0.036) across both datasets. Additionally, it outperformed other CLIP-based methods, improving AUC by 20.3% and 14.3%.

Aryan Dixit et al [17] utilized LLM models like GPT and Llama along with the pretrained models, ResNET50, and VGG16, DenseNET121, inception with Mobile Net. 81,584 images from TCIA had been taken. A pre-trained LLM model was used to generate doctor prescription and comprehensive medical literature was used for fine tuning. Finally model accuracy was 99.58

CH change et al [18] Large language models (LLMs) Llama-2-70b-chat ,Med42-70B and ClinicalCamel-70B were used to extract pathologic tumor-node-metastasis (pTNM) staging information from real-world pathology reports. Then finally all these LLMs were compared with a pre-trained Clinical-BigBird model. Comparing all prompting approach few shot approach did not improve the performances of LLMs significantly. It was also observed that LLMs are sensitive to prompting which required rigorous prompt testing and engineering.

T Wang et al [19] proposed a PneuNet which is Vision Transformer (VIT). The model used lung X Ray images and backed by channel-based attention. Multi-head attention is applied on channel patches rather than feature patches. It performed well with an accuracy result 94.96%.

K Nasiri et al [20] described different LLM models with their architectures BERT, RoBERTa, GPT-3, GPT-4, Bloom, PaLM, LLaMA, StableLM, T5, BART. Differentiated between classic fine-tuning and prompt tuning. Mentioned evolution of GPT in healthcare, described the application of transformer in healthcare industry, summarized different medical dataset used for LLM models.

X Meng et al [21] summarized LLMs applications in medical research by different organ system with different study area, explained how globally LLM is spreaded over for medical research, and also studied LLMs medical contributions across various nations.

J Haltaufderheide et al [22] Illustrated different aspects of applications like Predictive Analysis and Risk Assessment, Patient Consultation and Communication, Treatment Planning, Professional Support and Research.

Lun-Hsiang Yuan et al [23] used CT scan , MRI scan, bone scan and biopsy pathology of a Forth stage prostate cancer patient. Research was performed with four LLMs (ChatGPT-4-turbo, Claude-3-opus, Gemini-Pro-1.0, ChatGPT-3.5-turbo) on three Risk Assessment (RA) tasks (LATITUDE, CHAARTED, TwNHI) and seven Information retrieval (IR) tasks. Zero-shot chain-of-thought prompting was used via application programming interface in LLM models. ChatGPT-4-turbo outperformed by other LLMs with accuracy the 91.6% in 3 RA tasks.

Mathew Renze et al [25] described the difference between Chain of Thought (CoT) and (CCoT) Concise chain of Thought. Two kinds of CoTs Zero shot and Few shots were mentioned. Zero Shots instructs LLM to think step by step whereas Few shots Cot provides series of example solutions. CCoT gives instruction to think step by step but to be precised. As a result CCoT reduced average response length by 48.70% for both GPT-3.5 and GPT-4.

Sumit Madan et al [26] described the different transformer model in different domain like biomedical graph, biomedical images biological sequencing, biomedical natural language, biomedical multimodal data, structured longitudinal EHR in healthcare. Explained LLMs including Generative Pre-trained Transformer(GPT), Bidirectional Encoder Representations from Transformers (BERT) Large Language Model Meta AI (LLamA) and BigScience Large Open-science Open-access Multilingual Language Model (BLOOM).

Subhash Nerellaa et al [27] used Clinical notes, Electronic health records to describe transformer application in different clinical applications like clinical word embedding, clinical information extraction with Name Entity Recognition(NER), Clinical coreference resolution (CR), Clinical relationship extraction (CRE), Question answering (QA), Biomedical entity normalization (BEN), Semantic text similarity (STS), Automatic international statistical classification of diseases (ICD) coding.

Rubin R et al [28] proposed Deep ViT used Multi head self attention(MHSA) and a fed forward network to perform the task. In comparison with other pretrained model Deep ViT outperformed than RES Net, Mobile inception, Deep ViT with accuracy 97% .

Kaixuan Li et al [29] explain Occlusion Transgate which is a transformer based gate recognition network. The network combined with Vision Transformer (ViT) and Swin transformer base structure. The Deep ViT used occlusion dataset. The Swin Transformer used different back bone type NM, BG, CL and got improved accuracy with 0.4% on NM, 1.0% on BG and 1.4% on CL.

TABLE I
COMPARISON TABLE OF PROPOSED WORK WITH PREVIOUS WORK

Sl No	Ref.No	GPT 2	GPT3.5/ GPT4	ViT	Both GPT and ViT	CoT Prompting
1	[1]	X	X	yes	X	X
2	[6]	X	X	yes	X	X
3	[8]	X	Yes	X	X	X
4	[9]	X	Yes	X	X	X
5	[1]	X	Yes	X	X	Yes
6	[12]	X	Yes	X	X	Yes
7	[13]	X	Yes	X	X	X
8	[14]	X	x	Yes	X	X
9	[15]	X	Yes	X	X	X
10	[17]	X	Yes	X	X	X
11	[19]	X	Yes	X	Yes	X
12	[20]	X	Yes	X	X	X
13	[23]	X	Yes	X	X	X
14	[25]	X	Yes	X	X	X
15	[28]	X	Yes	X	X	X
16	[30]	Yes	X	X	X	X
17	[31]	X	Yes	X	X	X
18	[32]	X	Yes	X	X	X
19	[33]	X	Yes	X	X	Yes
20	Proposed Model	X	Yes	X	Yes	Yes

Ken Cheligeer et al [30] used 351 breast cancer patients underwent with NeoAdjuvant Chemotherapy(NAC) and subsequent surgery. 15 different transformer-based models were used followed by logistic regression to classify on these embeddings to classify pCR. Generative Pre-trained Transformer-2 (GPT-2) Model was fine tuned by attaching a simple feed forward network. result was shown with 95.3% sensitivity, 90.9% of positive predictive value which outperforms by traditional machine learning methods.

Pranab Sahoo et al [31] used chest CT scan images of SARS-CoV-2 datasets for covid detection. Vision Transformer model was proposed for this detection rather than traditional machine learning and other pretrained CNN methods. Accuracy of 98.39% was achieved with this ViT model.

Kobiljon Ikromjanov et al [32] used multiple patches of images extracted from whole slide image (WSI) of prostate biopsy with Gleason grading system. First WSI images were divided into patches. ROI was used to extract the patches then ViT was used for classification.

Musarrat Hussain et al [33] proposed CoT-STS where they focused on four primary factors theme similarity, participating object similarity, similarity of the activity being carried out, and the evaluation of other factors which are played an important role in NLP task. The experiment was performed with BIOSSES dataset.Result was shown with Pearson's correlation of 0.72 outperfoms the standard prompting and prompting used by zero-shot CoT respectively 0.45 and 0.71.

IV. RESULTS & DISCUSSION

Result section describes the overview of the papers and answer to the general and specific Questions.

TABLE II
DATASET REFERENCES

SNo	Ref.No	Dataset
1	[32]	Kaggle PANDA challenge
2	[10]	EHR, SDoH
3	[31]	SARS-CoV-2
4	[30]	ACR, SCM
5	[29]	Occlusion masked dataset
6	[28]	Raabbin Database
7	[27]	PubMedQA, SQuAD, BioASQ, BioCreative CDR, MIMIC-III, NCBI, TAC2017ADR, EHR, clinical notes
8	[3]	Chest XRay
9	[34]	'PlantVillage' public dataset
10	[14]	PatchCamelyon
11	[7]	hematoxylin and eosin (H&E) stained microscopic images
12	[17]	TCIA
13	[18]	TNM, pTNM
14	[19]	Lung XRay image
15	[23]	CT scan, MRI scan
16	[26]	EHR
17	[27]	Clinical notes, EHR

- GQ1: What are the dataset used for carrying out the research in NLP domain?

TABLE III
AI MODELS USED WITH REFERENCES

S.No	Ref.No	AI Model
1	[3]	ViT
2	[10]	GPT-3.5, GPT-4
3	[11]	LaMDA, GPT, PaLM
4	[12]	GPT-4, GPT-3.5 Turbo, LLaMA, and PaLM 2
5	[13]	GPT-4
6	[14]	ViT
7	[15]	GPT4-V
8	[17]	GPT and LLaMA, ResNET50, VGG16, DenseNET121, Mobile Net
9	[18]	Llama-2-70b-chat, Med42-70B and ClinicalCamel-70B
10	[19]	ViT
11	[20]	BERT, RoBERTa, GPT-3, GPT-4, Bloom, PaLM, LLaMA, StableLM, T5, BART
12	[23]	ChatGPT-4-turbo, Claude-3-opus, Gemini-Pro-1.0, ChatGPT-3.5-turbo
13	[25]	Chain of Thought (CoT) and (CCoT) Concise chain of Thought
14	[26]	GPT, BERT, LLaMA, and BLOOM
15	[27]	PaLM, GPT3, GPT4, LLaMA, BERT, PaLM2, ViT, BioBERT
16	[28]	Deep ViT
17	[29]	ViT, Swin Transformer, YOLOX, LISA
18	[30]	GPT-2
19	[31]	ViT
20	[32]	ViT
21	[34]	ViT

Ans. The datasets used to carry out the review are Kaggles Panda challenge, SARS-CoV-2, Alberta Cancer Registry (ACR) database, Sunrise Clinical Manager (SCM), Electronic Health Records (EHRs), Social Determinants of Health (SDoH), Rabbin WBC Database, PubMedQA, SQuAD, BioASQ, BioCreative CDR, MIMIC-III, NCBI, TAC2017ADR, EHR, clinical notes.

- GQ2: What are the techniques used to extract the data in NLP?
  Ans. To extract data in NLP several Techniques are used:
  - Rule based Techniques (regex, template matching, dictionary based approach)
  - Statistical and Machine Learning-Based Techniques (NER, POS, Dependency parsing)
  - Deep Learning-Based Techniques (Transformer based model, NER with deep learning).
- GQ3: What are the Modern technologies used in health-care domain for NLP?
  Ans. Different transformer based models BERT, GPT based-models, BART, NER model, spaCy, scispaCY, med7, medical chatbots are used.

- SQ1: Which large language models are used for cancer detection?
  Ans. LaMDA, GPT, PaLM, GPT-4, GPT-3.5 Turbo, LLaMA, and PaLM 2

- SQ2: Which modality of dataset is chosen?
  Ans. Text dataset or image dataset.

- SQ3. Wchich prompting systems are used for fine tuning the model?
  Ans. Zero shot, One shot , few shot prompting, Chain of thought(CoT), Concise chain of thought (CCoT), RAG

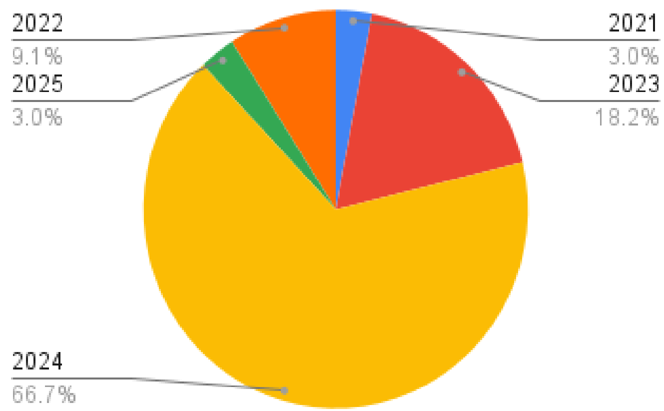

Fig. 1. Pie Chart for Year Wise Publications

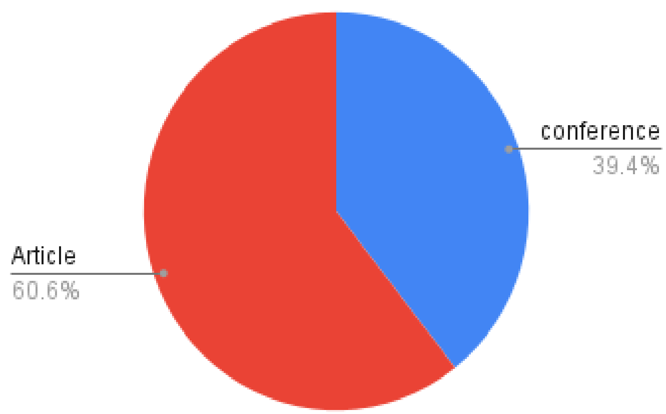

Fig. 2. Pie Chart Document Type Publications

V. CONCLUSION

This review delves into prominent datasets, data extraction methods, advanced technologies, and prompt systems in NLP, especially in healthcare. MIMIC-III, PubMedQA, and electronic health records are some of the datasets supporting clinical NLP research. Data extraction is based on rule-based, statistical, and deep learning methods. Models like BERT, GPT, and BioBERT based on transformers greatly boost NLP applications in healthcare, such as cancer detection. Moreover, sophisticated prompting methods like zero-shot, few-shot, and chain-of-thought (CoT) prompting assist in fine-tuning large language models. These developments together help enhance automated medical analysis and decision-making systems. In future author will try to implement BigScience Large Open-science Open-access Multilingual Language Model(BLOOM) which supports 46 languages and 13 programming languages.

References

[1] Krishnan, K. S., & Krishnan, K. S. (2021, October). Vision transformer based COVID-19 detection using chest X-rays. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 644-648). IEEE.

[2] Zhao, H., Yilahun, H., & Hamdulla, A. (2023, November). Pipeline chain-of-thought: A prompt method for large language model relation extraction. In 2023 International Conference on Asian Language Processing (IALP) (pp. 31-36). IEEE.

[3] Wang, T., Nie, Z., Wang, R., Xu, Q., Huang, H., Xu, H., ... & Liu, X. J. (2023). PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using Vision Transformer. Medical & Biological Engineering & Computing, 61(6), 1395-1408.

[4] Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J. B., Aali, A., Bluethgen, C., ...& Chaudhari, A. S. (2024). Adapted large language models can outperform medical experts in clinical text summarization. Nature medicine, 30(4), 1134-1142

[5] De Brito, B. A. O., Barroca Filho, I. D. M., De Lima, J. M. M., Gurgel, A. M., & Malaquias, R. (2024, June). Natural Language Processing in the Context of Cancer: A Systematic Review. In 2024 IEEE Symposium on Computers and Communications (ISCC) (pp. 1-6). IEEE.

[6] Angara, S., Mannuru, N. R., Mannuru, A., & Thirunagaru, S. (2024). A novel method to enhance pneumonia detection via a model-level ensembling of CNN and vision transformer. arXiv preprint arXiv:2401.02358.

[7] Pu, Q., Xi, Z., Yin, S., Zhao, Z., & Zhao, L. (2024). Advantages of transformer and its application for medical image segmentation: a survey. BioMedical engineering online, 23(1), 14.

[8] Tozuka, R., Johno, H., Amakawa, A., Sato, J., Muto, M., Seki, S., ... & Onishi, H. (2024). Application of NotebookLM, a large language model with retrieval-augmented generation, for lung cancer staging. Japanese Journal of Radiology, 1-7.

[9] Schmidl, B., Hütten, T., Pigorsch, S., Stögbauer, F., Hoch, C. C., Hussain, T., ... & Wirth, M. (2025). Artificial intelligence for image recognition in diagnosing oral and oropharyngeal cancer and leukoplakia. Scientific Reports, 15(1), 3625.

[10] Ralevski, A., Taiyab, N., Nossal, M., Mico, L., Piekos, S. N., & Hadlock, J. (2024). Using Large Language Models to Annotate Complex Cases of Social Determinants of Health in Longitudinal Clinical Records. medRxiv.

[11] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.

[12] Oniani, D., Wu, X., Visweswaran, S., Kapoor, S., Kooragayalu, S., Polanska, K., & Wang, Y. (2024, June). Enhancing large language models for clinical decision support by incorporating clinical practice guidelines. In 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI) (pp. 694-702). IEEE.

[13] Hsueh, J. Y., Nethala, D., Singh, S., Linehan, W. M., & Ball, M. W. (2024, September). Investigating the clinical reasoning abilities of large language model GPT-4: an analysis of postoperative complications from renal surgeries. In Urologic Oncology: Seminars and Original Investigations (Vol. 42, No. 9, pp. 292-e1). Elsevier.

[14] Kumar, A., Ojha, S., Tyagi, S., Yadav, U. K., & Mittal, S. (2024, August). Histopathological Breast Cancer Detection: A Vision Transformer and Attention-Based Approach. In 2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET) (pp. 1-6). IEEE.

[15] Ferber, D., Wölflein, G., Wiest, I. C., Ligero, M., Sainath, S., Ghaffari Laleh, N., ... & Kather, J. N. (2024). In-context learning enables multimodal large language models to classify cancer pathology images. Nature Communications, 15(1), 10104.

[16] Chen, X., Li, Y., Hu, M., Salari, E., Chen, X., Qiu, R. L., ... & Yang, X. (2024). Mammo-clip: Leveraging contrastive language-image pre-training (clip) for enhanced breast cancer diagnosis with multi-view mammography. arXiv preprint arXiv:2404.15946.

[17] Dixit, A., Mishra, S., Goyal, A., & Jain, V. (2024, November). Lung Cancer Detection: Leveraging Transfer Learning & AI Chatbot for Doctor Prescription. In 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 35-41). IEEE.

[18] Change, C. H., Lucas, M. M., Lu-Yao, G., & Yang, C. C. (2024, June). Classifying cancer stage with open-source clinical large language models. In 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI) (pp. 76-82). IEEE.

[19] Wang, T., Nie, Z., Wang, R., Xu, Q., Huang, H., Xu, H., ... & Liu, X. J. (2023). PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using Vision Transformer. Medical & Biological Engineering & Computing, 61(6), 1395-1408.

[20] Nassiri, K., & Akhloufi, M. A. (2024). Recent advances in large language models for healthcare. BioMedInformatics, 4(2), 1097-1143.

[21] Meng, X., Yan, X., Zhang, K., Liu, D., Cui, X., Yang, Y., ... & Tang, Y. D. (2024). The application of large language models in medicine: A scoping review. Iscience, 27(5).

[22] Haltaufderheide, J., & Ranisch, R. (2024). The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ digital medicine, 7(1), 183.

[23] Yuan, L. H., Huang, S. W., Chou, D., & Tsai, C. Y. (2024). The In-depth Comparative Analysis of Four Large Language AI Models for Risk Assessment and Information Retrieval from Multi-Modality Prostate Cancer Work-up Reports. The World Journal of Men's Health, 42.

[24] Preiksaitis, C., Ashenburg, N., Bunney, G., Chu, A., Kabeer, R., Riley, F., ... & Rose, C. (2024). The role of large language models in transforming emergency medicine: scoping review. JMIR medical informatics, 12, e53787.

[25] Renze, M., & Guven, E. (2024, November). The benefits of a concise chain of thought on problem-solving in large language models. In 2024 2nd International Conference on Foundation and Large Language Models (FLLM) (pp. 476-483). IEEE.

[26] Madan, S., Lentzen, M., Brandt, J., Rueckert, D., Hofmann-Apitius, M., & Fröhlich, H. (2024). Transformer models in biomedicine. BMC Medical Informatics and Decision Making, 24(1), 214

[27] Nerella, S., Bandyopadhyay, S., Zhang, J., Contreras, M., Siegel, S., Bumin, A., ... & Rashidi, P. (2024). Transformers and large language models in healthcare: A review. Artificial intelligence in medicine, 102900.

[28] Rubin, R., Anzar, S. M., Panthakkan, A., & Mansoor, W. (2023, November). Transforming healthcare: Raabin white blood cell classification with deep vision transformer. In 2023 6th International Conference on Signal Processing and Information Security (ICSPIS) (pp. 212-217). IEEE.

[29] Li, K., & Meng, S. (2023, November). Transgait: Vision transformer based gait recognition network. In 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 339-343). IEEE.

[30] Cheligeer, K., Wu, G., Laws, A., Quan, M. L., Li, A., Brisson, A. M., ... & Xu, Y. (2024). Validation of large language models for detecting pathologic complete response in breast cancer using population-based pathology reports. BMC Medical Informatics and Decision Making, 24(1), 283.

[31] Sahoo, P., Saha, S., Mondal, S., & Gowda, S. (2022, September). Vision transformer based covid-19 detection using chest ct-scan images. In 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 01-04). IEEE

[32] Ikromjanov, K., Bhattacharjee, S., Hwang, Y. B., Sumon, R. I., Kim, H. C., & Choi, H. K. (2022, February). Whole slide image analysis and detection of prostate cancer using vision transformers. In 2022 international conference on artificial intelligence in information and communication (ICAIIC) (pp. 399-402). IEEE.

[33] Hussain, M., Rehman, U. U., Nguyen, T. D., & Lee, S. (2023, December). CoT-STS: A Zero Shot Chain-of-Thought Prompting for Semantic Textual Similarity. In Proceedings of the 2023 6th Artificial Intelligence and Cloud Computing Conference (pp. 135-139).

[34] Maurya, R., Pandey, N. N., Singh, V. P., & Gopalakrishnan, T. (2023, May). Plant disease classification using interpretable vision transformer network. In 2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON) (pp. 688-692). IEEE.