International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5756
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

Hybrid Multitask Learning with LLM Embeddings for Lung Cancer Subtype Classification: A Narrative Review

Anandhakumar Kaluvakolli[1]
[1]*Harrisburg University of Science and Technology, Pennsylvania, US*
Email ID: *anandusa007@gmail.com*[1]

Abstract

The use of histological subtyping of lung cancer for first-line therapy, targeted agents and prognostic counselling is supported by evidence that exists in a wide variety of data types: unstructured narrative in pathology and clinical records, coded clinical variables, molecular panels and digital slides. Three research axes have emerged, which have been developed in parallel to the problem of fragmentation: domain adaptation of large language models (LLMs) embeddings to clinical narrative, multimodal fusion of heterogeneous features, and multitask learning to overcome the challenge of limited labels. This narrative review combines those threads and proposes that their intersection, which we call hybrid multitask learning with LLM embeddings, is a unified and less explored design space for subtype classification. We present the evolution of clinical text representation from bag-of-words features to domain-adapted transformer encoders and generative foundation models, the different integration strategies in the literature including early concatenation, late fusion, learned gating and cross-modal co-attention, and single-task, hard-sharing and uncertainty-weighted multitask formulations. There are three consistent trends throughout the literature: in-domain pretraining is more important than the number of parameters; learned modality interaction is more effective than simply concatenation; and auxiliary clinical information like stage and survival is an effective regulariser in situations where supervision is scarce in the minority classes. Next, we discuss the challenges that hinder translation — size of cohorts, imbalance in classes, leakage of labels from diagnostic conclusions within the source text, lack of external multi-site validation, expectations of interpretability in pathology, and privacy restrictions of hosted models — and outline reporting practices and research priorities that would make results in this area comparable.

Keywords: *lung cancer; histological subtype; narrative review; large language models; clinical text embeddings; multitask learning; multimodal fusion; computational oncology*

1. Introduction

Lung cancer results in more cancer deaths than any other cancer, about one in five cancer deaths globally and continues to be the leading cause of cancer deaths in high and middle-income countries [1,2]. It is an unusual management for which the histological classification plays a role. Small cell and non-small cell carcinoma have different treatment strategies and within non-small cell, adenocarcinoma vs squamous cell dictates pemetrexed-based therapy and anti-angiogenic therapy and some molecularly targeted agents. The successive World Health Organization revisions have more clearly defined these categories and have added even more immunohistochemical and molecular parameters, which have improved the diagnostic accuracy and the burden of interpretation for the reporting pathologist [3,4]. In this context, computational support has been rapidly developed as part of a general trend towards data-driven clinical decision support [5]. But growth has not been consistent for all data types. The work on image-based artefacts is well-established and well benchmarked, while the textual record - the one artefact that virtually every

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

institution has - has received much less systematic study for subtyping in particular. In the meantime, two methodological literatures with direct implications to the problem have been created, namely: multimodal fusion and multitask learning, both largely outside the oncology realm and sparsely referenced in thoracic applications. This review seeks to mend that rift. It is designed to be intentionally limited to methods that use transformer-based language models to represent clinical narrative, incorporate both clinical and molecular structured variables with clinical narrative, and train multiple clinical objectives simultaneously. We seek to characterise each strand and its contribution; to compare the available options on three dimensions and determine what combinations the evidence suggests; to identify common methodological weaknesses that hinder comparability; and to recommend reporting conventions and research priorities. This is a narrative review, not a systematic review: the evidence base has a variety of definitions of the tasks and reporting, so quantitative pooling would be inappropriate, and the contribution is conceptual organisation, not meta-analysis.

2. Scope and approach

The literature was retrieved from PubMed, Scopus, IEEE Xplore and the ACL Anthology using combinations of terms related to lung cancer and histological subtype, clinical natural language processing, contextual/sentence embeddings, multimodal fusion and multitask/auxiliary-task learning, and by backward and forward citation search from the key papers in each thread. Peer-reviewed studies are given preference; a few highly influential preprints and conference papers are included where they published a method with no subsequent journal publication. Works were excluded if they involved detection or screening without histological typing, or no evaluation of clinical data. Since the strands vary in how they define outcome, cohorts and choice of metric, results are synthesised qualitatively and reported as patterns that emerge across the strands, not pooled effect estimates.

3. The classification problem and its data

In practice, subtype assignment is a reasoning process and is written down as a text. A pathological report usually records the growth pattern, morphology of the cells, characterisation of the mitoses, a panel of various immunohistochemical stains and then comes to a conclusion, the clinical record adds smoking exposure, distribution of the disease on the radiograph and treatment context. These categories are associated with truly different biology, as demonstrated by molecular profiling, which revealed that adenocarcinoma and squamous carcinoma have different mutational and pathway landscapes [8,9] and that panel features have independent discriminative signal, not just a restatement of morphology. There is a characteristic strength and a characteristic gap for each modality that is available. Deep models over slides can discriminate between adenocarcinoma and squamous carcinoma at the level of expert readers and can also infer selected driver mutations from morphology; the limitation is infrastructure, as research-ready digitised slides are still focused in the tertiary centres [6]. Computed tomography has good reported performance for detection and risk stratification, but has poor reliability for resolving histology [7]. Coded clinical fields are quantitative, inexpensive to process, but surveys of electronic health record modelling reveal a consistent upper limit due to the nature of what is coded [10]. Narrative text is almost ubiquitous and includes qualifiers, uncertainty and morphological nuances that are not captured by any coded field, but it is stylistically variable, institution-specific and, importantly, often includes the diagnostic conclusion to be predicted by the model in the supervised learning setting. The data has two properties that give each method its structure. Cohorts with related story, structured variables and outcome labels are typically no larger than a few thousand patients, with the distribution of subtypes often being highly skewed, such as an order of magnitude fewer small cell carcinoma patients than adenocarcinoma in surgical cohorts. Both properties penalize high capacity single objective models and provide an explanation for the convergence to frozen representations and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

auxiliary supervision.

4. Representing clinical narrative

The representation axis has gone through 4 distinct phases, outlined in the top row of Figure 1. While sparse lexical features like TF-IDF are still a legitimate baseline and are interpretable on a term-by-term basis, they fail to distinguish between negated and asserted findings, and cannot resolve the abbreviations which are common in pathology reporting. Static word embeddings provided distributional similarity but no contextual disambiguation. This pivotal shift occurred with the introduction of self-attention architectures [11] and bidirectional pretraining [12] which resulted in token representations that could be conditioned on the surrounding report and transfer to small labelled sets. The most reproducible empirical result in the clinical NLP literature is that domain adaptation of these encoders is more important than scale in the ranges that are typically used. BioBERT, a model pretrained on biomedical literature, has been shown to perform better than its general domain initialisation on biomedical entity and relation extraction tasks [13] and PubMedBERT demonstrated that pretraining from scratch on in-domain text is better than continual pretraining on a biomedical benchmark suite [14]. Encoders specially designed for hospital notes, instead of literature, capture a different register — one that is dense in abbreviations, telegraphic and template-influenced — and better predict from admission records accordingly [15]. This register sensitivity is directly consequential for subtype classification, in which the discriminative evidence is in the staining and morphology vocabulary. Two preprocessing decisions are not discussed to the extent that they deserve. Pathology narrative is highly abbreviated and punctuation is inconsistent, and normalisation of variant forms (expansion of shorthand for tumour types, harmonisation of stain names, removal of institutional headers and de-identification placeholders) measurably alters what an encoder can recover, but is not often described with sufficient detail to reproduce. The second is document length: reports often fall outside of the several hundred token window of these classes of encoders, and studies either truncate or segment with overlap and pool chunk representations. Furthermore, truncation and pooling are not the same, and it is rarely specified with results, making it difficult to make comparisons between different studies. One of the questions that can be asked is fine-tune or freeze? Sentence-level embedding objectives also enable the use of the encoders as fixed feature extractors [16] which is often the safer path to take given the cohort sizes found in thoracic oncology: a few hundred labelled patients, a hundred million trainable parameters, and instability. Generative and foundation models are the current state of the art, and have been shown to encode significant clinical knowledge [17] and are increasingly being applied in the medical field [18,19], but they come with their own costs, governance and reproducibility concerns, and for a well-specified three class problem, their utility over a frozen domain adapted encoder has not been well established. As shown in Figure 1 Proposed hybrid multitask learning framework for lung cancer subtype classification.

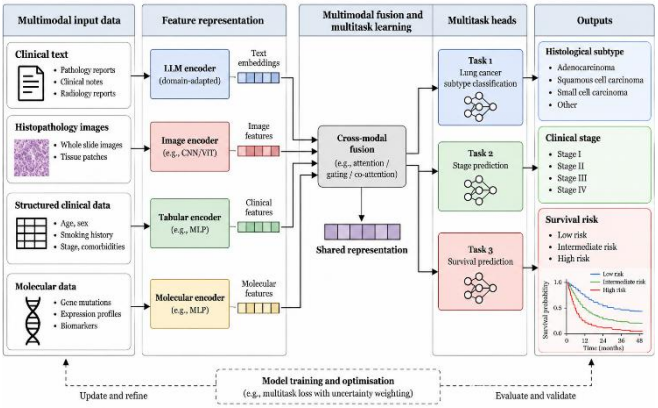


Figure 1 Proposed hybrid multitask learning framework for lung cancer subtype classification

5. Integrating narrative with structured data

The fusion literature has an established taxonomy of when and how modality-specific data is fused [24] and the categories carry over to the text-plus-tabular case; the trade-offs are summarized in Table 2. The most widely-used and most basic method is to concatenate a text embedding with a structured feature vector, but this assigns the same weight to

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

text and tabular modalities for all patients and allows a high-dimensional text embedding to outnumber a short tabular embedding. Late fusion keeps the modularity and allows for missing modalities, but never learns cross-modal interaction. Learned gating is able to compute a weight from both modalities and multiply each element of the representations by the corresponding weight, giving more weight to the narrative evidence in more richly documented cases and more weight to the structured/molecular features in more templated or truncated cases, reflecting the variability in how cases are documented in real health systems. Co-attention mechanisms allow for the most elaborate interaction and have been shown to outperform unimodal inputs in multimodal survival modelling in oncology [25], but require many parameters which are hard to justify for less than a few thousand patients. In this domain, missingness is structural, and should be treated separately. The ordering of a molecular panel is related to care pathway, disease burden and institutional resources, and the absence of a panel is also informative and imputing it would add optimistic bias if imputation statistics were calculated for the entire dataset. Simple precautions such as explicitly representing absence, using an indicator per field, and fitting all imputation parameters within training folds, are precautions taken inconsistently in studies in this area. The different fusion designs have varying degrees of gracefulness in their ability to deal with a missing modality: late fusion is natural, gating can learn to shift weight away from a missing source, and early concatenation and co-attention generally require explicit masking to behave sensibly. The common finding in all the studies is learned interaction between modalities is superior to fixed combination, and the right level of interaction depends on the size of the cohort. Thus, gating is an appealing intermediate for subtype classification, which is more expressive than concatenation and much lower cost than co-attention.

Table 2 Design choices for hybrid multitask models over clinical narrative and structured data, with reported advantages and limitations specific to small, class-imbalanced oncology cohorts

Design choice	Mechanism	Reported advantages	Limitations under small, imbalanced cohorts
Text-only or tabular-only	A single encoder over one modality	Simple, easy to audit, low data requirement	Discards complementary evidence; performance plateaus well below fused models
Early concatenation	Feature vectors joined before a shared classifier	Minimal added parameters; a strong and common baseline	Fixed, patient-independent weighting; a high-dimensional text vector can dominate a short tabular vector
Late fusion / ensembling	Separate models combined at the decision level	Modular; tolerates missing modalities; each component testable	No shared representation, so cross-modal interactions are never learned
Learned gating	A gate computed from both modalities weights them element-wise per patient	Adapts to documentation quality; degrades gracefully when text is sparse or templated	Gate can collapse to a constant if the training set is too small or too homogeneous

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

Cross-modal co-attention	Tokens of one modality attend to the other [25]	Richest interaction; strongest results on large multimodal cohorts	Parameter-hungry and prone to overfitting below a few thousand patients
Single-task supervision	One loss, one output head	Direct optimisation of the target metric	High variance; minority classes such as SCLC are learned poorly
Hard parameter sharing	A shared trunk with task-specific heads	Auxiliary labels act as inductive bias [20,21]	Negative transfer if auxiliary tasks are weakly related to the target
Uncertainty-weighted multitask	Per-task learned noise parameters scale each loss [23]	Removes manual loss-weight search; down-weights noisily labelled tasks automatically	Adds tuning-free but opaque weights; sensitive to task-label completeness

6. Multitask formulations

To improve generalisation, multitask learning was proposed, requiring one representation to serve multiple related tasks; the auxiliary tasks can also be viewed as an implicit regulariser [20]. Later, the distinction between hard and soft parameter sharing was formalized, where hard parameter sharing is where a shared trunk is used for a task with task-specific heads, and soft sharing is where separate models are coupled by constraints or adapters [21,22]. In oncology there are natural auxiliary candidates, which are downstream consequences of tumour biology: stage grouping, grade, treatment response and survival. These rely on the same kind of latent properties of aggressiveness and dissemination that underlie subtype, and drive the shared representation into biological structure, rather than surface lexical cues, as is most commonly proposed to explain gains on rare classes. The choice of auxiliary tasks is far from decided. The most common is stage grouping, which is widely used because annotation is widely available, and there is a clinically intelligible relationship between stage and subtype, although stage is also affected by the timing of detection, which reduces the commonality of the signal. Survival endpoints are more directly biological and can be either a discrete horizon or a continuous risk with a partial likelihood loss, but are censored and require longer follow-up than many retrospective cohorts. Treatment response and grade are occasionally mentioned, and mutation status has proven to be successful as an auxiliary target in image-based work [6], which indicates a similar function for the panel-derived labels in text-plus-structured models. The number of auxiliary heads is limited by the completeness of the labels, not by the capacity, and two or three well-justified tasks are more frequently reported than larger sets. Loss balancing is the constant challenge. Manual weights are hard coded and will need searching which can't be undertaken with small validation sets. Uncertainty weighting addresses this by using each task's weight as a learned homoscedastic noise parameter, which results in the automatic down-weighting of noisily or incompletely labelled auxiliary tasks [23] – this is now the recommended approach when auxiliary label completeness differs across contributing sources, which is usually the case in pooled oncology cohorts. Two cautions recur. Negative transfer is when the auxiliary task is not strongly related to the target task, and auxiliary selection should be made for clinical reasons, not convenience. The gains are focused where supervision is lacking: reported gains are highest for minority subtypes and lowest for the dominant class, which is exactly why a single number of accuracy is a bad headline number for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

this problem.

7. Synthesis

Figure 1 shows the three axes plotted and the area favoured by the evidence. When read together, the literature suggests a fairly specific configuration: a frozen domain-adapted encoder for narrative text, a learned per-patient gate that combines this to an encoded structured and molecular vector, and a small number of clinically justified auxiliary heads, trained with uncertainty weighting.

Table 1 Research strands contributing to hybrid multitask subtype classification, with their data modality, supervision structure and specific relevance to the reviewed problem

Research strand	Representative work	Data modality	Supervision	Relevance to subtype classification
Histopathology deep learning	Coudray et al. [6]	Whole-slide images	Single-task and mutation prediction	Established that LUAD/LUSC separation is learnable at expert level; sets the accuracy reference but requires digitised slides
Radiological screening	Ardila et al. [7]	Low-dose CT volumes	Single-task risk prediction	Demonstrated end-to-end learning on imaging; addresses detection rather than histological typing
Molecular characterisation	TCGA LUAD [8]; TCGA LUSC [9]	Transcriptomic and mutational profiles	Unsupervised and descriptive	Provides the biological grounding for subtype labels and the curated features reused as structured inputs
Structured EHR modelling	Shickel et al. [10]	Coded clinical variables	Varied, mostly single-task	Documents the strengths and ceiling of coded-field models; motivates adding narrative text
General-domain language models	Vaswani et al. [11]; Devlin et al. [12]	Free text	Self-supervised pretraining	Supplies the contextual encoder architecture on which all clinical text representations now rest
Domain-adapted encoders	Lee et al. [13]; Gu et al. [14]; Alsentzer et al. [15]	Biomedical literature and hospital notes	Self-supervised, then fine-tuned	Shows that in-domain pretraining, not parameter count alone, drives biomedical transfer; the usual embedding source
Frozen sentence-level embeddings	Reimers and Gurevych [16]	Free text	Contrastive sentence objectives	Makes encoders usable as fixed feature extractors, the safer option for small oncology cohorts

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

Generative and foundation models	Singhal et al. [17]; Thirunavukarasu et al. [18]; Moor et al. [19]	Free text, multimodal	Instruction tuning, prompting	Encodes substantial clinical knowledge but raises cost, governance and reproducibility concerns
Multitask learning theory	Caruana [20]; Ruder [21]; Zhang and Yang [22]; Kendall et al. [23]	Modality-agnostic	Joint multitask objectives	Frames auxiliary tasks as regularisers and resolves loss balancing through learned uncertainty weights
Multimodal fusion	Baltrusaitis et al. [24]; Chen et al. [25]	Images, text, molecular	Single- and multi-task	Provides the fusion taxonomy and evidence that learned cross-modal interaction beats naive concatenation

Each of these elements is based on findings that are independent of each other, and none of them depend on digitised slides, or the very large cohorts that most institutions do not have: domain adaptation over scale [13,14], learned interaction over concatenation [24,25], auxiliary regularisation with learned loss weights [20,23]. The contributing strands and their respective roles are mapped in table 1. As shown in Table 1 Research strands contributing to hybrid multitask subtype classification, with their data modality, supervision structure and specific relevance to the reviewed problem. LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas; CT, computed tomography; EHR, electronic health record; LLM, large language model. There are two caveats to this synthesis. First, the constituent evidence is mostly from neighbouring problems, such as imaging and molecular survival modelling, and computer vision and general clinical prediction. There is currently limited direct comparative evidence for the entire combination on lung cancer subtyping, and the area highlighted in Figure 1 should therefore be interpreted as a hypothesis ripe for testing by the field. Second, reporting heterogeneity makes cross-study comparison unreliable (discussed below).

8. Open challenges

8.1.Label leakage from source text

It is the most serious and least regularly reported issue in the region. Pathology reports typically include a diagnosis, and a text model can be almost perfect with just a phrase look-up, without having to interpret descriptive evidence. Different studies have different policies on whether or not to filter such spans, and if they do, how aggressively, and whether or not to even report the filter, which can contribute significantly to the apparent difference in performance. The leakage-control procedure should be explicitly described and preferably an ablation should be used to quantify the leakage effect.

8.2.Cohort size, imbalance and validation

Cohorts are small, minority cohorts are rarely studied, and external multi-site validation is uncommon; most studies cross-validate within a single data source, which means validation is limited to data source-specific documentation conventions. If the minority class is from a single contributing site, then the documentation style can be correlated with the label itself. Macro-averaged metrics and per-class reporting is necessary and, at least external evaluation, if not prospective, should precede clinical claims.

8.3.Interpretability, privacy and governance

Pathology is an explanatory discipline and a prediction of a subtype without any evidence to trace back to it is of very limited practical use; a prediction by attribution over fused representations and by inspection of gate behaviour provides only part of the solution, and neither is a diagnostic rationale. There are additional reasons for compact

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

in-domain models [14,19] as hosted generative models introduce additional deployment constraints, such as transporting identifiable clinical narrative, which locally hosted frozen encoders tend to avoid. There are also less developed regulatory processes for text-based diagnostic support than for imaging devices.

8.4.Reporting practice

With a few conventions, however, comparisons could be made much more meaningful: per-class and macro-averaged metrics; an explicit statement about leakage control; whether encoders were frozen or fine-tuned, and how many trainable parameters; ablation of fusion and of each auxiliary task separately, as contributions of the two are often mixed; and description of what source institutions and documentation templates are represented in the cohort.

9. Future directions

The gaps above are the basis for four priorities. Multi-institutional benchmarks with linked narrative, structured and outcome data and with explicit leakage-controlled splits, would do more for the field than further architectural variation. Systematic comparison of frozen domain-adapted encoders and prompted generative models on the same subtyping cohorts would resolve a question answered by assumption. The natural next step is extension to slide-level, where the diagnostic reference standard is used in conjunction with the text-plus-structured fusion described here, but this has not yet been tried to any great extent. Federated or privacy-preserving training should be considered, as it is the one that has the greatest potential to create cohorts of sufficient size without transferring identifiable narrative between institutions. In addition to these, there are two methodological questions that are suitable for direct study. One is how far out the auxiliary-task benefit does carry: if it is indeed a variance-reduction mechanism, then it should diminish with scale, and knowing where the crossover is would give practitioners an indication of when the extra complexity is worthwhile. The second is whether gate behaviour can be transformed into a data-quality signal that can be used. When documentation is templated or truncated, the learned weighting, if it reliably moves away from narrative text, can also be interpreted as an indicator of report informativeness, and may be used for documentation audit, apart from the classification task.

Conclusion

The classification of lung cancer subtypes is at the intersection of three literatures that have developed independently. This review has followed the evolution of clinical text representation towards domain-adapted transformer encoders, the fusion taxonomy that dictates how narrative evidence and structured evidence are combined, and multitask formulations that use related clinical labels as regularisers, and identified the configuration their combined evidence points to for the small, imbalanced cohorts common in thoracic oncology. The methodological ingredients are in place; the missing pieces are direct comparative evidence for the combination of the methodological ingredients and the discipline of reporting and common benchmarks that would make such evidence interpretable. Nowadays, the development of this field is more and more relying on leakage control multi-institutional evaluation, and the routine ablation of fusion and auxiliary-task contributions.

References

[1]. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–263.

[2]. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.

[3]. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10(9):1243–1260.

[4]. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. J

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

Thorac Oncol. 2022;17(3):362–387.
[5]. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
[6]. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24(10):1559–1567.
[7]. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(6):954–961.
[8]. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543–550.
[9]. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489(7417):519–525.
[10]. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform. 2018;22(5):1589–1604.
[11]. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS) 30. 2017. p. 5998–6008.
[12]. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. 2019. p. 4171–4186.
[13]. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36(4):1234–1240.
[14]. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthc. 2021;3(1):1–23.
[15]. Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop, NAACL. 2019. p. 72–78.
[16]. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of EMNLP-IJCNLP. 2019. p. 3982–3992.
[17]. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172–180.
[18]. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930–1940.
[19]. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259–265.
[20]. Caruana R. Multitask learning. Mach Learn. 1997;28(1):41–75.
[21]. Ruder S. An overview of multi-task learning in deep neural networks. arXiv:1706.05098 [cs.LG]. 2017.
[22]. Zhang Y, Yang Q. A survey on multi-task learning. IEEE Trans Knowl Data Eng. 2022;34(12):5586–5609.
[23]. Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. p. 7482–7491.
[24]. Baltrusaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell. 2019;41(2):423–443.
[25]. Chen RJ, Lu MY, Weng WH, Chen TY, Williamson DFK, Manz T, et al. Multimodal co-attention transformer for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 10 October 2026
Page No: 5786 - 5795
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0751

survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021. p. 4015–4025.