Extended Results for: Enhancing Abstract Screening Classification in Evidence-Based Medicine: Incorporating Domain Knowledge into Pre-trained Models*

Regina Ofori-Boateng[1,*], Magaly Aceves-Martins[2], Nirmalie Wirantuga[1] and Carlos Francisco Moreno-García[1]

[1]School of Computing, Robert Gordon University, Aberdeen, Scotland
[2]The Rowett Institute, University of Aberdeen, Scotland

Abstract

Evidence-based medicine (EBM) is a foundational element in medical research, playing a crucial role in shaping healthcare policies and clinical decision-making. However, the rigorous processes required for EBM, particularly during the abstract screening phase, pose substantial challenges to researchers. While many have sought to automate this stage using Pre-trained Language Models (PLMs), these efforts often face obstacles due to the specificity of the domain, especially when dealing with EBM studies related to both human and animal subjects. To address this, our initial research presented a state-of-the-art (SOTA) transfer learning approach that enhanced four abstract screening by embedding domain-specific knowledge into PLMs without modifying their base weights utilising the concepts of adapters. Extending the previous work, in this study, we evaluate the same methodology on four animal and human EBM datasets. Our evaluation, conducted on the further four EBM abstract screening datasets, demonstrates that the proposed method significantly improves the screening process and outperforms strong baseline PLMs.

Keywords

Evidence-Based Medicine, Domain Integration, Large/Pre-trained Language Models, Transfer Learning,

1. Introduction

Evidence-based medicine(EBM) is regarded as one of the most reliable methods for informing healthcare policies and guiding clinical decision-making, providing a structured approach to synthesising research findings [1]. The EBM process typically involves several key stages: (i) formulating a protocol that outlines the review's objectives and methodologies, (ii) defining a precise research question using frameworks such as PICO (Population, Intervention, Comparison, Outcome measures) to establish inclusion and exclusion criteria [2], (iii) conducting comprehensive literature searches, (iv) screening abstracts for relevance, (v) extracting and analysing data from selected studies, and (vi) interpreting and publishing the findings. Despite its well-defined process, EBM remains a labour-intensive and time-consuming endeavour, a challenge further compounded by the exponential increase in the volume of published research [3, 4]. On average, completing and publishing an EBM study takes approximately 15 months [5], which often results in reviews becoming outdated soon after their completion, necessitating frequent updates and revisions [6].

Among the various stages of EBM, the abstract screening phase is frequently identified as the most demanding [6, 7]. Empirical evidence indicates that an experienced researcher may take between 30 to 90 seconds to screen a single abstract [8], with screening approximately 5,000 publications requiring an estimated 80 to 125 hours of effort [9]. To address these challenges, numerous methodologies have been proposed for automating the abstract screening stage, ranging from text classification techniques

SICSA REALLM Workshop 2024
*The original paper can be found in: https://link.springer.com/chapter/10.1007/978-3-031-66538-7_26
*Corresponding author.
r.ofori-boateng@rgu.ac.uk (R. Ofori-Boateng)
0000-0002-0319-773X (R. Ofori-Boateng); 0000-0002-9441-142X (M. Aceves-Martins); 0000-0003-4040-2496 (N. Wirantuga); 0000-0001-7218-9023 (C. F. Moreno-García)

to prioritisation methods, all aiming to achieve a recall rate of 95% or higher [10].

Despite these advancements, existing approaches are often limited by domain specificity, particularly in complex studies that involve both human and animal data [11]. Recently, the development of Large/Pre-trained language models (PLMs) tailored to specific domains, such as SciBERT [12], PubMedBERT [13], ClinicalBERT [14], and BioBERT [15], has significantly advanced the field of abstract screening. These models, when fully fine-tuned (FFT) on abstract datasets, have demonstrated improved performance [16]. However, since PLMs are often pre-trained on unstructured and unrestricted corpora, they may lack the structured domain knowledge necessary for effective performance on biomedical tasks [17, 18]. As a result, these models tend to treat biomedical entities and concepts as generic tokens, which can limit their efficacy [19]. Furthermore, the application of FFT PLMs involves updating a substantial number of parameters whenever a new EBM dataset is introduced, leading to increased computational costs, memory requirements, and resource utilisation.

To address these limitations, in our previous work [20], we presented a domain knowledge integrated PLM using the concept of adapters. In the previous work, we showed that knowledge-based adapters outperform FFT PLMs on the abstract screening tasks. As an extended version of the work done previously, this study further investigates the method's potential on four extra tasks. Thus, the same research questions **(RQs)** are also used here. To summarise, in our previous work, the **RQs** asked were

1. How can the diverse domain knowledge crucial for abstract screening tasks be integrated into a base PLM without adjusting model parameters?
2. What is the effect of different configurations of the knowledge layers (where they are inserted) on the downstream task?
3. Can adapter-based tuning perform better than SOTA FFT PLMs proposed for EBM abstracts?

2. Methodology

Extending our initial research, the same method, implementation, experimental setup, and approach used in the previous work [20] were applied in this research. As such, readers are referred to our preceding work [20] for a contextual understanding. Popular metrics such as recall, precision and F1 scores were used for evaluations. To summarise, the methodology from our previous work, to address **RQ1** we integrated small neural networks (referred to as knowledge layers or adapters) into the layers of a base PLM, SciBERT, and trained them on three domain-specific knowledge (PICO entities, two biomedical Q&A datasets, PubMedQA[1] and BioASQ 7B[2]). SciBERT was chosen due to its broad coverage of the biomedical domain, making it well-suited for this task. To address **RQ2**, we explored the impact of different configurations of these knowledge layers—specifically, where they are inserted—on the downstream task by comparing three configurations. The configurations we investigated were: (a) the Houlsby Configuration (H) [21], where adapter modules are inserted before the multi-head attention mechanism and the FeedForward layer of the SciBERT model [22], (b) the Pfeiffer Configuration (Pf) [15], where adapters are placed only after the FeedForward layer, and (c) the Compacter Configuration (C), which is similar to the Houlsby configuration but replaces the standard linear FFD and FFU with a Parameterised Hypercomplex Multiplication (PHM) layer [14]. The PHM layer computes its weights using the Global Multiplier of the Kronecker Product (GMKP) between two smaller matrices. For a detailed explanation of the GMKP and PHM, refer to [23]. Lastly, to address **RQ3**, we investigated whether adapter-based tuning outperforms SOTA fine-tuned PLMs designed for the EBM abstracts task. As such, we empirically evaluate the performance of the trained knowledge layers against FFT SciBERT and assess the transferability and modularity of the method by inserting the trained networks into other variants like ClinicalBERT, PubMedMBERT, and BioBERT, then comparing them with their FFT-tuned versions.

Adapters are small neural networks integrated into PLMs to allow task-specific fine-tuning with minimal changes. They include a FeedForward Down Projection (FFD) and Up Projection (FFU) to

[1] https://pubmedqa.github.io/
[2] http://participants-area.bioasq.org/datasets

Algorithm 1 Overview of a base PLM with Knowledge Layer/Adapter Mechanism

Require: Input vector x

Ensure: Output vector $A(y)$ after Adapter processing

Initialization:

Weight matrices and biases: $W_{\text{down}}, b_{\text{down}}, W_{\text{up}}, b_{\text{up}}$

Step 1: Process input x through the SubLayer e.g the feed-forward network of the PLM

$x \leftarrow$SubLayer(x)

Step 2: Add and Normalize

$y \leftarrow \text{Norm}(x + \text{SubLayer}(x))$

Step 3: Adapter processing

{Component 1: Down-Projection}

$z \leftarrow W_{\text{down}} \cdot y + b_{\text{down}}$

{Component 2: Apply Activation function}

$a \leftarrow \text{LeakyReLU}(z)$

{Component 3: Up-Projection}

$u \leftarrow W_{\text{up}} \cdot a + b_{\text{up}}$

{Component 4: Residual Connection}

$A(y) \leftarrow u + y$

return $A(y)$ where $A(y)$ is passed as input to the subsequent layers in the PLM

reduce dimensionality, a LeakyReLU activation to handle negative inputs, and a skip residual to retain essential information. These components enable efficient adaptation of PLMs without altering the core model structure. Algorithm 1 provides a mathematical description of how adapters function within the PLM.

2.1. SLR Datasets for Evaluation

The proposed PLM, consisting of the knowledge integrated (PICO entities, PubMedQA and BioASQ dataset) through adapters underwent fine-tuning and evaluation on four extra complex medical SLR abstract datasets encompassing both human and animal studies. In addition to the, four datasets in our previous studies (Aceves-Martins_2022 [24], Leenars_2019 [25], Van_Dis [26], Appenzeller-Herzog dataset [27]), the impact of the proposed method was investigated on four extra datasets; (Aceves-Martins_2022 [28], Muthu_2022 [29], Wassenaar_2017 [30], Menon_2022 [31], Oud_2018 [32], and Nelson_2006 [33]—publicly available on GitHub[3] repository. A summary is provided in Table 1.

Name_of_dataset	Subject	Total	Relevant	Irrelevant	IR
Aceves-Martins_2021(AM_21)	Oral Health and obesity in children	807	18	789	1:44
Menon_2022 (MN)	Toxicology and environmental epidemiology	975	74	901	1:12
Oud_2018 (OU)	Specialised psychotherapies for adults	952	20	932	1:47
Wassenaar_2017(WR)	Obesity-Related Outcomes in Rodents	7668	111	7557	1:68

Table 1
Summary of the datasets ranging from human to animal study, where IR = Imbalance Ratio

3. Results and Discussion

Figure 1 illustrates the outcomes of comparing adapter-based tuning with FFT biomedical variant PLMs and established knowledge-integrated PLM baselines (CODER-BERT and KRISSBERT). Overall, the findings consistently show that tuning the various FPBPA configurations(Compacter(C), Pfeiffer(Pf), and Houlsby(H)) within these PLMs results in significant metric improvements over the baseline models. This observation addresses **RQ3** and underscores the efficacy of FPBPA in the context of SLR abstract

[3]https://github.com/asreview/synergy-dataset/tree/master

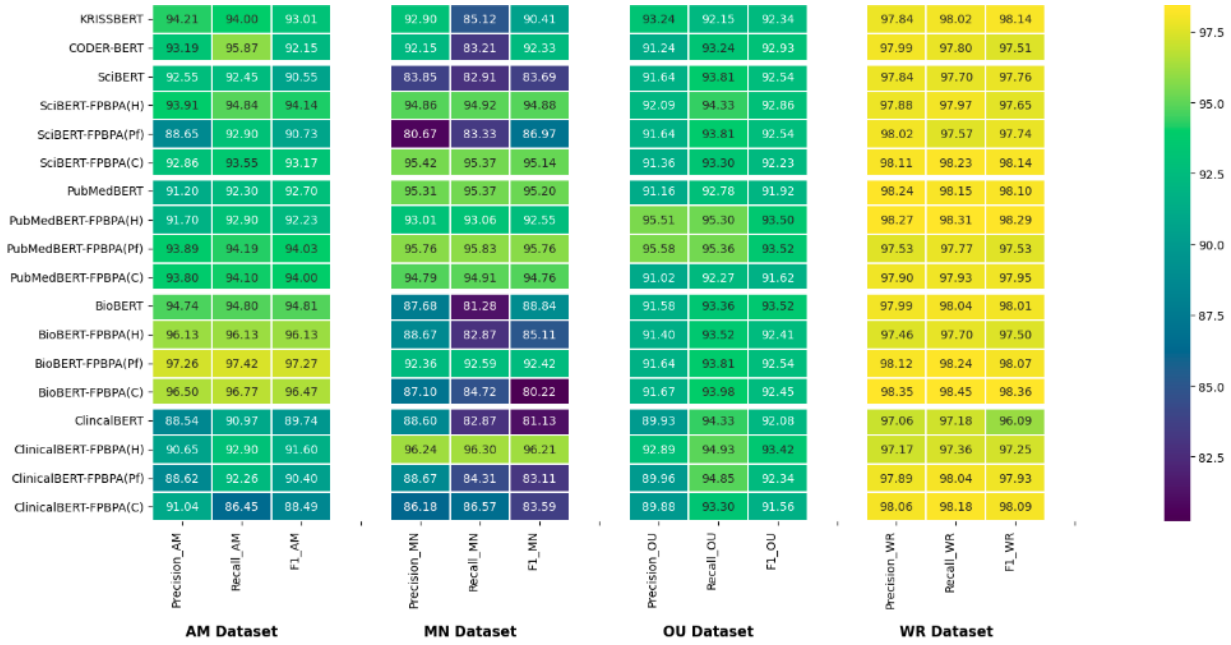

Figure 1: Heatmap of results obtained for the downstream datasets with imbalance ratio (IR) < 90

screening tasks. In this section, we discuss the advantages and relevance of adapter-based tuning in relation to **RQs** when compared to FFT PLMs.

3.1. Computational Efficiency

Table 2 offers a comparative analysis of trainable parameters for different adapter configurations, such as Compacter, Pfeiffer, and Houlsby, relative to the SciBERT PLM. The results indicate that the integrated adapters (PICO, PubMedQA, and BioASQ) significantly reduce the number of trainable parameters compared to the base PLM. For example, the PICO adapter's parameter count ranges from just 57,088 in the Compacter configuration to roughly 1.79 million in the Houlsby configuration, which is considerably fewer than the base SciBERT PLM. Even in the FPBPA setup, where multiple adapters are combined to utilise diverse knowledge, the trainable parameter count increases but still remains below that of the base SciBERT PLM. This highlights that the adapter approach offers a viable solution for improving computational efficiency in machine learning applications by substantially lowering trainable parameters while preserving model performance. This not only saves computational resources but also allows for the seamless integration of varied and evolving knowledge bases into PLMs, addressing the ever-changing demands of biomedical research and its applications.

Type	Compacter	Pfeiffer	Houlsby
SciBERT PLM	109920002	109920002	109920002
PICO Adapter	57088	894528	1789056
PubMedQA Adapter	59968	1189632	2379264
BioASQ Adapter	57088	894528	1789056
All Fused Adapter FPBPA	42678336	24230784	48461568

Table 2
Total Number of Trainable Parameters of the Knowledge Adapters and base PLM

3.2. Can Adapter-Based Tuning Outperform SOTA FFT PLMs?

This section evaluates whether adapter-based tuning can match or exceed the performance of FFT PLMs by analysing the balance between evaluation metrics and the number of task-specific parameters

trained. Across both types of datasets, adapter-based tuning with various configurations (H, Pf, C) consistently outperformed FFT PLMs (such as SciBERT, ClinicalBERT, PubMedBERT, and BioBERT) and knowledge-integrated baselines (like KRISSBERT and CODER-BERT), based on the evaluation metrics. These results also highlight the adaptability of adapters when transferred across different base PLMs. The findings presented in Figure 1 show that adapter-based tuning (specifically, FPBPA configurations) generally enhances the performance of baseline models such as SciBERT, PubMedBERT, and BioBERT, particularly in precision, recall, and F1 scores across diverse datasets. This improvement is especially noticeable in datasets with higher imbalance ratios, indicating that adapter-based methods are more effective at handling imbalanced data scenarios.

For example, in Figure 1, focusing on the **AM_21 (IR:44)** dataset, SciBERT-FPBPA(H) achieved a precision of 93.91%, surpassing the base SciBERT model, which achieved 92.55%. This trend of improved performance using FPBPA variants holds true across other PLMs, showing notable gains in recall and F1 scores. For instance, the F1 scores for BioBERT-FPBPA configurations (Pf, H, and C) were 97.27%, 96.13%, and 96.47%, respectively, compared to 94.84% for the base BioBERT model. Similarly, for the **MN(IR 1:12)** dataset, ClinicalBERT-FPBPA(H) demonstrated a significant F1 score, showcasing the effectiveness of adapter-based tuning in moderately imbalanced scenarios. In the case of the **OH(IR 1:47)** dataset, BioBERT and its FPBPA variants showed strong performance, with BioBERT-FPBPA(Pf) achieving the highest F1 score. Lastly, for the **WR (IR 1:68)** dataset, FPBPA variants of PubMedBERT and BioBERT exhibited high precision and recall.

3.3. Impact of Different Knowledge Adapter Configurations

Different FPBPA configurations (H, Pf, C) had varying impacts on different datasets, as seen in Figure 1. Generally, FPBPA(H) performed better in datasets with moderate to high IRs, while FPBPA(Pf) showed superior performance in highly imbalanced datasets as seen in our previous work.

3.4. Practical Implications

For real-world SLR abstract automation in biomedical research involving human and animal studies, particularly those with highly imbalanced datasets, adapter-based tuning methods offer a more effective alternative to traditional SOTA FFT models. The ability of adapter-based models to handle imbalanced data is noteworthy, suggesting their practical applicability in real-world scenarios where data imbalance is common. The choice of model and adapter configuration should be tailored to the specific dataset characteristics, especially the imbalance ratio. As discussed, FPBPA (Pf) configurations are more suitable for datasets with extreme imbalance, whereas FPBPA (H) is preferable for moderate imbalance scenarios. For SLRs using broad search strings (e.g., topics like prisoners or adolescents), FPBPA (Pf) would provide an advantage due to its capability to manage a large volume of irrelevant data. These findings can also extend to other fields facing similar challenges with imbalanced data, highlighting the broad applicability of adapter-based tuning methods.

4. Conclusion and Future Works

This study, which is amn extended work, proves the potential of integrated domain-specific knowledge into pre-trained language models (PLMs) using adapters. By employing the PICO framework and leveraging resources such as PubMedQA and BioASQ Q&A datasets, our method enhances the capabilities of PLMs for EBM abstract screening. This improvement is vital for advancing clinical decision-making and policy development. Our extensive experimental results demonstrate that the proposed approach achieves strong performance across various metrics, including precision, recall, F1 score.

One of the future woreks stated in our previous work was to extend the proposed method on more datasets on more datasets, which was the aim of this study. In future work, we aim to incorporate additional domain-specific knowledge bases, such as the Unified Medical Language System (UMLS), DisGeNET, and the UNIPROT knowledge database, to further enrich our model's domain expertise.

Although our current research is focused on the BERT model, we plan to extend this work to other cutting-edge PLMs, such as GPT and LLaMA, to explore broader applications. Additionally, we intend to conduct comparative studies between our approach and traditional baseline models, including Support Vector Machines (SVM) and Naive Bayes (NB) with and without UMLS integration. We will also explore the effectiveness of keyword-based search methods, such as those using cTAKES or MetaMap, and investigate models leveraging TF-IDF or n-gram analysis for abstract screening.

Acknowledgments

The authors thank members of the Childhood Obesity in Mexico (COMO)[4] project for supporting this research.

References

[1] P. B. Burns, R. J. Rohrich, K. C. Chung, The levels of evidence and their role in evidence-based medicine, Plastic and Reconstructive Surgery 128 (2011) 305–310. URL: https://doi.org/10.1097%2Fprs.0b013e318219c171. doi:10.1097/prs.0b013e318219c171.

[2] A. M. Methley, S. Campbell, C. Chew-Graham, R. McNally, S. Cheraghi-Sohi, PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews, BMC Health Services Research 14 (2014). URL: https://doi.org/10.1186/s12913-014-0579-0. doi:10.1186/s12913-014-0579-0.

[3] I. J. Marshall, B. C. Wallace, Toward systematic review automation: a practical guide to using machine learning tools in research synthesis, Systematic Reviews 8 (2019). URL: https://doi.org/10.1186%2Fs13643-019-1074-9. doi:10.1186/s13643-019-1074-9.

[4] A. Bannach-Brown, P. Przybyła, J. Thomas, A. S. Rice, S. Ananiadou, J. Liao, M. R. Macleod, Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error, Systematic reviews 8 (2019) 1–12. URL: https://doi.org/10.1186/s13643-019-0942-7.

[5] L. Bornmann, R. Mutz, Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references, Journal of the Association for Information Science and Technology 66 (2015) 2215–2222. URL: https://doi.org/10.48550/arXiv.1402.4578. doi:10.1002/asi.23329. arXiv:1402.4578.

[6] R. van de Schoot, J. de Bruin, R. Schram, P. Zahedi, J. de Boer, F. Weijdema, B. Kramer, M. Huijts, M. Hoogerwerf, G. Ferdinands, A. Harkema, J. Willemsen, Y. Ma, Q. Fang, S. Hindriks, L. Tummers, D. L. Oberski, An open source machine learning framework for efficient and transparent systematic reviews, Nature Machine Intelligence 3 (2021) 125–133. URL: http://dx.doi.org/10.1038/s42256-020-00287-7. doi:10.1038/s42256-020-00287-7.

[7] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, S. Ananiadou, Using text mining for study identification in systematic reviews: a systematic review of current approaches, Systematic reviews 4 (2015) 1–22. URL: https://doi.org/10.1186/2046-4053-4-5.

[8] B. E. Howard, J. Phillips, A. Tandon, A. Maharana, R. Elmore, D. Mav, A. Sedykh, K. Thayer, B. A. Merrick, V. Walker, A. Rooney, R. R. Shah, SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation, Environment International 138 (2020) 105623. URL: https://doi.org/10.1016/j.envint.2020.105623. doi:10.1016/j.envint.2020.105623.

[9] P. Przybyła, A. J. Brockmeier, G. Kontonatsios, M. A. Le Pogam, J. McNaught, E. von Elm, K. Nolan, S. Ananiadou, Prioritising references for systematic reviews with RobotAnalyst: A user study, 2018. URL: https://doi.org/10.1002/jrsm.1311. doi:10.1002/jrsm.1311.

[10] R. van Dinter, B. Tekinerdogan, C. Catal, Automation of systematic literature reviews: A systematic literature review, Information and Software Technology 136 (2021) 106589. URL: https://doi.org/10.1016/j.infsof.2021.106589.

[4] https://www.comoprojectmx.com/collaborators

[11] A. Natukunda, L. K. Muchene, Unsupervised title and abstract screening for systematic review: a retrospective case-study using topic modelling methodology, Systematic Reviews 12 (2023). URL: http://dx.doi.org/10.1186/s13643-022-02163-4. doi:10.1186/s13643-022-02163-4.

[12] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, M. McDermott, Publicly available clinical, in: Proceedings of the 2nd Clinical Natural Language Processing Workshop, Association for Computational Linguistics, 2019. URL: https://doi.org/10.18653%2Fv1%2Fw19-1909. doi:10.18653/v1/w19-1909.

[13] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, Domain-specific language model pretraining for biomedical natural language processing, ACM Transactions on Computing for Healthcare 3 (2021) 1–23. URL: https://doi.org/10.1145%2F3458754. doi:10.1145/3458754.

[14] K. Huang, J. Altosaar, R. Ranganath, Clinicalbert: Modeling clinical notes and predicting hospital readmission, 2020. URL: https://doi.org/10.1093%2Fbioinformatics%2Fbtz682. doi:10.1093/bioinformatics/btz682. arXiv:1904.05342.

[15] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics 36 (2019) 1234–1240. URL: https://doi.org/10.1093%2Fbioinformatics%2Fbtz682. doi:10.1093/bioinformatics/btz682.

[16] C. F. Moreno-Garcia, C. Jayne, E. Elyan, M. Aceves-Martins, A novel application of machine learning and zero-shot classification methods for automated abstract screening in systematic reviews, Decision Analytics Journal 6 (2023) 100162. URL: http://dx.doi.org/10.1016/j.dajour.2023.100162. doi:10.1016/j.dajour.2023.100162.

[17] A. Rogers, O. Kovaleva, A. Rumshisky, A primer in BERTology: What we know about how BERT works, Transactions of the Association for Computational Linguistics 8 (2020) 842–866. URL: https://doi.org/10.1162%2Ftacl_a_00349. doi:10.1162/tacl_a_00349.

[18] R. Wang, D. Tang, N. Duan, Z. Wei, X. Huang, J. Ji, G. Cao, D. Jiang, M. Zhou, K-adapter: Infusing knowledge into pre-trained models with adapters, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, 2021. URL: https://doi.org/10.18653%2Fv1%2F2021.findings-acl.121. doi:10.18653/v1/2021.findings-acl.121.

[19] Q. Xie, J. A. Bishop, P. Tiwari, S. Ananiadou, Pre-trained language models with domain knowledge for biomedical extractive summarization, Knowledge-Based Systems 252 (2022) 109460. URL: http://dx.doi.org/10.1016/j.knosys.2022.109460. doi:10.1016/j.knosys.2022.109460.

[20] R. Ofori-Boateng, M. Aceves-Martins, N. Wirantuga, C. F. Moreno-García, Enhancing Abstract Screening Classification in Evidence-Based Medicine: Incorporating Domain Knowledge into Pre-trained Models, Springer Nature Switzerland, 2024, p. 261–272. URL: http://dx.doi.org/10.1007/978-3-031-66538-7_26. doi:10.1007/978-3-031-66538-7_26.

[21] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. de Laroussilhe, A. Gesmundo, M. Attariyan, S. Gelly, Parameter-efficient transfer learning for nlp, 2019. URL: https://doi.org/10.48550/arXiv.1902.00751. arXiv:1902.00751.

[22] J. Pfeiffer, A. Kamath, A. Rücklé, K. Cho, I. Gurevych, AdapterFusion: Non-destructive task composition for transfer learning, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, 2021. URL: https://doi.org/10.18653%2Fv1%2F2021.eacl-main.39. doi:10.18653/v1/2021.eacl-main.39.

[23] R. K. Mahabadi, J. Henderson, S. Ruder, Compacter: Efficient low-rank hypercomplex adapter layers, 2021. arXiv:2106.04647.

[24] M. Aceves-Martins, N. L. Godina-Flores, Y. Y. Gutierrez-Gómez, D. Richards, L. López-Cruz, M. García-Botello, C. F. Moreno-García, Obesity and oral health in mexican children and adolescents: systematic review and meta-analysis, Nutrition Reviews 80 (2022) 1694–1710. URL: http://dx.doi.org/10.1093/nutrit/nuab088. doi:10.1093/nutrit/nuab088.

[25] C. H. C. Leenaars, C. Kouwenaar, F. R. Stafleu, A. Bleich, M. Ritskes-Hoitinga, R. B. M. D. Vries, F. L. B. Meijboom, Animal to human translation: a systematic scoping review of reported concordance rates, Journal of Translational Medicine 17 (????). URL: https://doi.org/10.1186/s12967-019-1976-2.

[26] E. A. M. van Dis, S. C. van Veen, M. A. Hagenaars, N. M. Batelaan, C. L. H. Bockting, R. M. van den Heuvel, P. Cuijpers, I. M. Engelhard, Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: A systematic review and meta-analysis, JAMA Psychiatry 77 (2020) 265. URL: http://dx.doi.org/10.1001/jamapsychiatry.2019.3986. doi:10.1001/jamapsychiatry.2019.3986.

[27] C. Appenzeller-Herzog, T. Mathes, M. L. Heeres, K. H. Weiss, R. H. Houwen, H. Ewald, Comparative effectiveness of common therapies for wilson disease: A systematic review and meta-analysis of controlled studies, Liver International 39 (2019) 2136–2152. URL: https://doi.org/10.1111%2Fliv.14179. doi:10.1111/liv.14179.

[28] M. Aceves-Martins, L. López-Cruz, M. García-Botello, Y. Y. Gutierrez-Gómez, C. F. Moreno-García, Interventions to prevent obesity in mexican children and adolescents: Systematic review, Prevention Science 23 (2021) 563–586. URL: http://dx.doi.org/10.1007/s11121-021-01316-6. doi:10.1007/s11121-021-01316-6.

[29] V. Muthu, R. R. Gandra, S. Dhooria, I. S. Sehgal, K. T. Prasad, H. Kaur, N. Gupta, A. Bal, B. Ram, A. N. Aggarwal, A. Chakrabarti, R. Agarwal, Role of flexible bronchoscopy in the diagnosis of invasive fungal infections, Mycoses 64 (2021) 668–677. URL: http://dx.doi.org/10.1111/myc.13263. doi:10.1111/myc.13263.

[30] P. N. H. Wassenaar, L. Trasande, J. Legler, Systematic review and meta-analysis of early-life exposure to bisphenol a and obesity-related outcomes in rodents, Environmental Health Perspectives 125 (????). URL: https://doi.org/10.1289/ehp1233.

[31] J. M. L. Menon, F. Struijs, P. Whaley, The methodological rigour of systematic reviews in environmental health, Critical Reviews in Toxicology 52 (????) 167–187. URL: https://doi.org/10.1080/10408444.2022.2082917.

[32] M. Oud, A. Arntz, M. L. M. Hermens, R. Verhoef, T. Kendall, Specialized psychotherapies for adults with borderline personality disorder: A systematic review and meta-analysis, Australian & New Zealand Journal of Psychiatry 52 (????) 949–961. URL: https://doi.org/10.1177/0004867418791257.

[33] E. Nelson, S. O'Meara, D. Craig, C. Iglesias, S. Golder, J. Dalton, K. Claxton, S. Bell-Syer, E. Jude, C. Dowson, R. Gadsby, P. O'Hare, J. Powell, A series of systematic reviews to inform a decision analysis for sampling and treating infected diabetic foot ulcers, Health Technology Assessment 10 (2006). URL: http://dx.doi.org/10.3310/hta10120. doi:10.3310/hta10120.