Objectives

The exponential growth of pharmacological information in unstructured sources such as scientific publications and clinical trial reports presents a barrier to large-scale pharmacometric data integration. Current biomedical text mining systems remain sensitive to annotation sparsity and limited training data, limiting reliable extraction of quantitative pharmacokinetic (PK) parameters and associated covariates required for model-informed drug development. Improving robustness under data-constrained conditions is therefore critical.

We present an enhanced BioBERT-based multi-task, multi-granularity named entity recognition (NER) framework designed to improve learning stability through homoscedastic uncertainty-based task weighting. The model jointly learns token-level PK entity recognition alongside auxiliary token- and sentence-level objectives, with adaptive uncertainty weighting, regulating task contributions during optimisation. This formulation acts as an uncertainty-aware regulariser, stabilising gradient updates and reducing overfitting in low-resource regimes. Applied to a gold-standard corpus of PubMed-indexed PK articles, the approach improves robustness in extracting PK entities and associated numerical values and units, particularly under simulated data scarcity.

Our objectives were: (i) to evaluate multi-task, multi-granularity learning for PK NER; (ii) to assess the application of joint multi-task homoscedastic uncertainty-based loss weighting to a hierarchical primary-auxiliary setting and (iii) to determine robustness and data efficiency under data scarcity.

Methods

We adapted the multi-task, multi-granularity BioBERT framework of Tong et al. [1] to PK parameter extraction, using hard parameter sharing across four classification heads built on a base BioBERT model: (i) primary token-level NER; (ii) token-level multi-token classification (mtCLS) distinguishing single- vs multi-token entities; (iii) sentence-level binary classification (bCLS) for PK presence; and (iv) sentence-level multi-class classification (mCLS) estimating mentions per sentence.

We compared: (1) a fine-tuned single-task BioBERT baseline [2]; (2) multi-task learning with heuristic loss weighting; (3) homoscedastic uncertainty-weighted multi-task learning in a symmetric, asymmetric and higher-granularity uncertainty schema [3].

Robustness was evaluated via progressive downsampling (75%, 50%, 25%, 10%, 7.5%, 5%, 2.5%) with class balance preserved. We also conducted auxiliary ablation and pairwise task interaction analysis using cosine similarity of shared encoder representations, as well as estimation of data efficiency. Further, we explore homoscedastic uncertainty trajectories and their relation to minority class entity recovery.

Results

Under full-data conditions, differences between the multi-task variants and the single-task baseline were minimal, consistent with a ceiling effect imposed by the BioBERT encoder. The advantages of multi-tasking emerged only as data became scarce. In reduced-data settings, training size significantly influenced performance, and multi-task models consistently outperformed the single-task baseline at dataset sizes below 25% of the original training set. In extreme low-data regimes this gap widened: the single-task baseline exhibited greater variance across seeds, whereas the multi-task models, particularly the uncertainty-weighted variant, scaled more stably and delivered clear gains over the single-task state of the art. Overall, multi-tasking was up to twice as data efficient.

Auxiliary task ablation produced only minor effects, indicating that robustness was not driven by any single objective. This picture is reinforced by several related observations: different architectures struggled on different tasks, and the tasks that mattered most varied with the weighting scheme. Crucially, performance was governed not by how frequently a class was seen but by the difficulty of that class, and finer-grained weighting within the NER task proved beneficial.

Conclusions

Multi-task, multi-granularity learning did not significantly outperform a strong single-task BioBERT baseline when data were abundant. Its primary benefit emerged under data-scarce conditions, where auxiliary supervision acted as a regulariser, reducing variance and slowing degradation in low-resource regimes. Overall, for PK NER, multi-task learning appears most valuable for improving generalisation and stability under annotation scarcity rather than maximising peak F1 in high-resource settings. Robust low-resource extraction of PK entities and quantitative attributes strengthens the evidentiary foundation for integrating AI-driven methods into large-scale parameter aggregation, thereby supporting more reliable pharmacometric modelling.