EVALUATION OF NAMED ENTITY RECOGNITION FOR AUTOMATED EXTRACTION OF PRESENT TUMOUR SIZE AND PERSONAL NAMES FROM RADIOLOGY REPORTS USING SPACY

Lorena Garcia-Foncillas Macias*, Theodore Barfoot*, Tom Vercauteren*, Jonathan Shapey*‡

* School of Biomedical Engineering & Imaging Sciences, King's College London, ‡ Department of Neurosurgery, King's College Hospital NHS Foundation Trust

INTRODUCTION

Meningiomas are the most common primary CNS tumours. Patients typically undergo multiple MRI scans to monitor growth, with results recorded in radiology reports, mostly in unstructured text.

De-identification of clinical records is essential for research collaboration.

- **HIPS** (Hiding in Plain Sight) [1] replaces identifiable data, like names, with realistic but fictitious alternatives.

NLP techniques, such as Named Entity Recognition (NER), extract structured data from unstructured text.

Project goals:

- Develop an NLP pipeline to **extract personal names** and **tumour sizes** from meningioma **radiology reports**
- **Implement HIPS** to de-identify personal names

METHODS

Dataset:
- 9,175 MRI radiology reports (2011–2020) from meningioma patients at King's College Hospital.
- **400 manually annotated** for personal names and tumour sizes using Label Studio [2].

Benchmark:
- A rule-based pipeline was developed to extract personal names only.

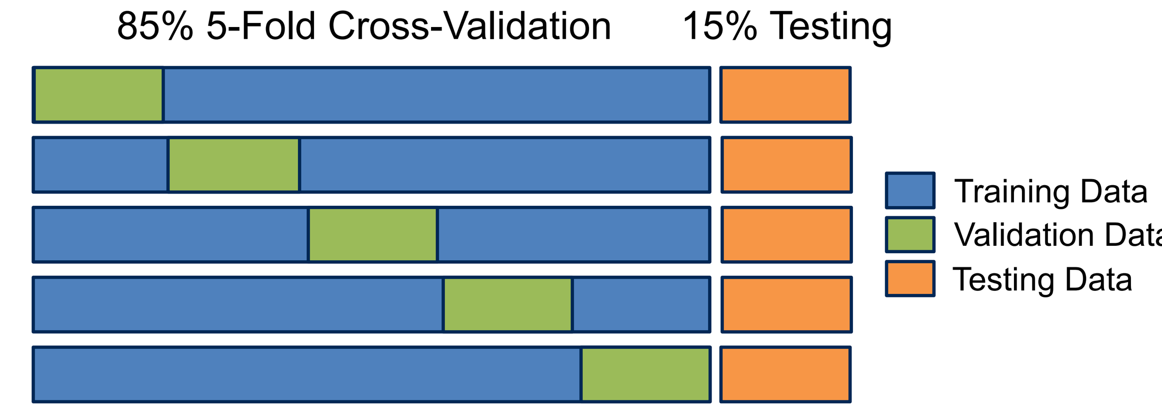
NER Models:

3 models from the **spaCy** toolkit [3] were trained to extract:
- Personal names
- Tumour sizes
- Both entities

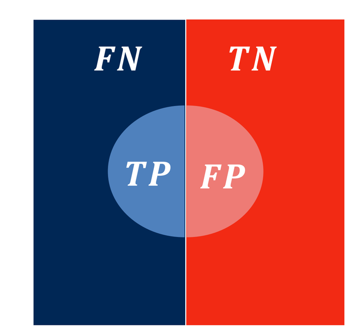
Evaluation:
- **5-fold cross-validation** on 85% of annotated data; 15% reserved for testing.
- Performance measured:
  - **Word level**: Macro-average precision, recall, F1-score
  - **Report level**: Exact matches between predictions and ground truth

De-identification:
- The best-performing model was used for HIPS implemented with Python library Faker [4].

spaCy Model Training



Word Level Evaluation Metrics



$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - Score = \frac{2TP}{2TP + FN + FP}$$

$$Macro - average = \frac{1}{N}\sum_{i=1}^{N} M_i$$

M_i = Metric value for an entity class
N = Total number of entity classes ($N = 3$)

RESULTS

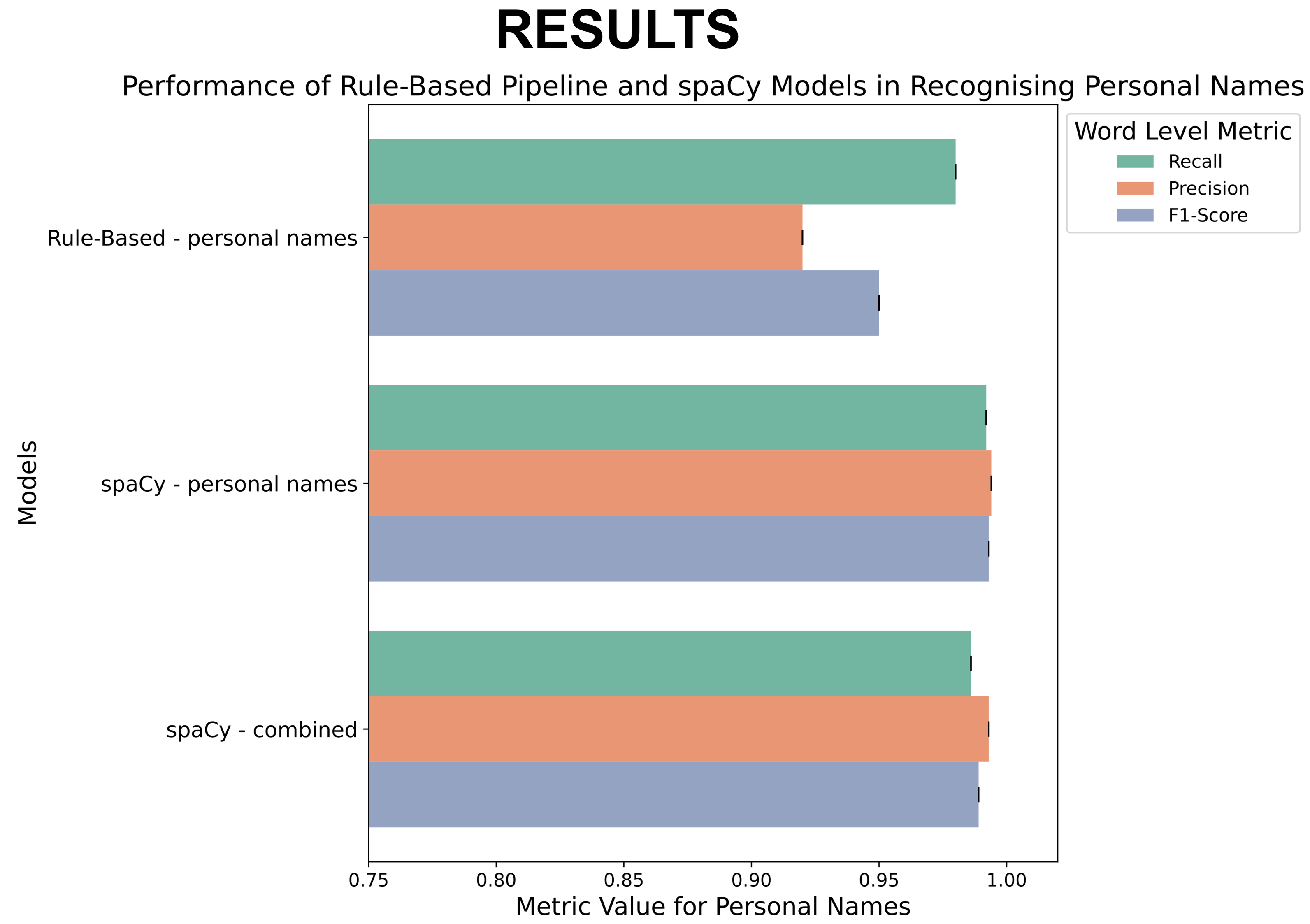


Figure 1: Word level performance comparison of the rule-based and spaCy models for extracting personal names on the testing dataset. F1- score, recall, and precision with error bars representing the standard deviation of the metrics from the cross-validation.

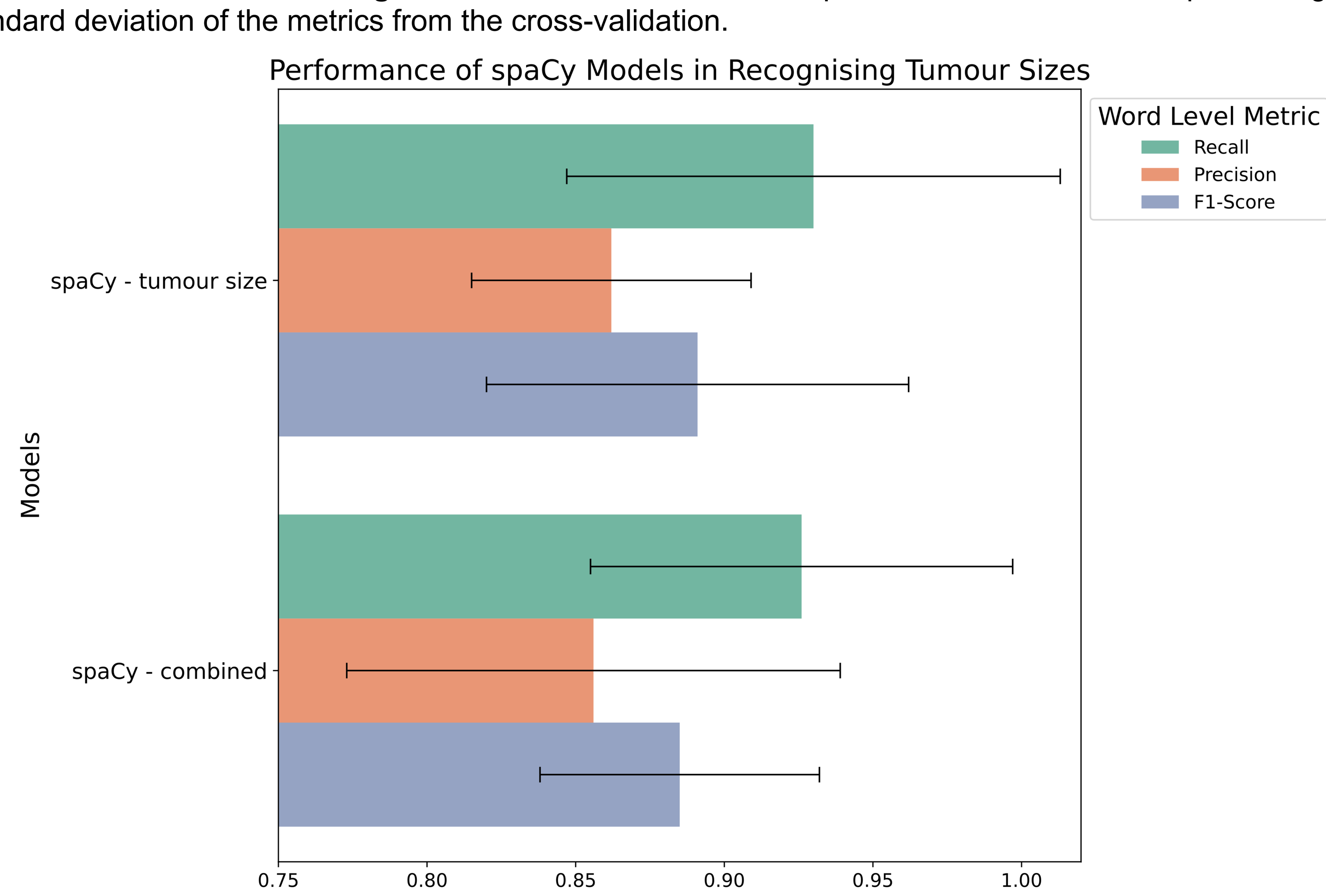
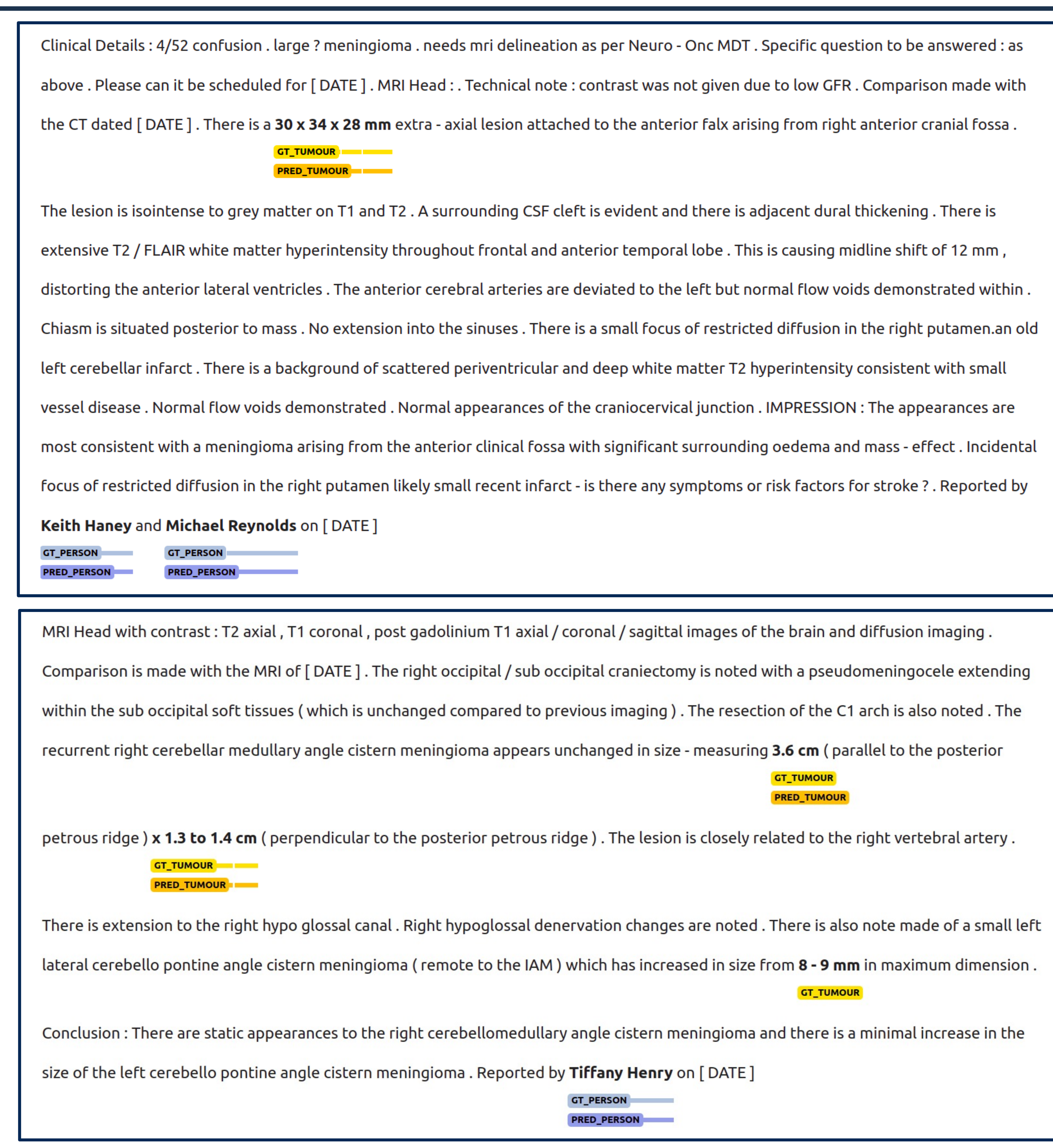


Figure 2: Word level performance comparison of the spaCy models for extracting present tumour sizes on the testing dataset. F1- score, recall, and precision with error bars representing the standard deviation of the metrics from the cross-validation.



Clinical Details : 4/52 confusion . large ? meningioma . needs mri delineation as per Neuro - Onc MDT . Specific question to be answered : as above . Please can it be scheduled for [DATE] . MRI Head : . Technical note : contrast was not given due to low GFR . Comparison made with the CT dated [DATE] . There is a **30 x 34 x 28 mm** extra - axial lesion attached to the anterior falx arising from right anterior cranial fossa .

GT_TUMOUR
PRED_TUMOUR

The lesion is isointense to grey matter on T1 and T2 . A surrounding CSF cleft is evident and there is adjacent dural thickening . There is extensive T2 / FLAIR white matter hyperintensity throughout frontal and anterior temporal lobe . This is causing midline shift of 12 mm , distorting the anterior lateral ventricles . The anterior cerebral arteries are deviated to the left but normal flow voids demonstrated within . Chiasm is situated posterior to mass . No extension into the sinuses . There is a small focus of restricted diffusion in the right putamen.an old left cerebellar infarct . There is a background of scattered periventricular and deep white matter T2 hyperintensity consistent with small vessel disease . Normal flow voids demonstrated . Normal appearances of the craniocervical junction . IMPRESSION : The appearances are most consistent with a meningioma arising from the anterior clinical fossa with significant surrounding oedema and mass - effect . Incidental focus of restricted diffusion in the right putamen likely small recent infarct - is there any symptoms or risk factors for stroke ? . Reported by **Keith Haney** and **Michael Reynolds** on [DATE]

GT_PERSON GT_PERSON
PRED_PERSON PRED_PERSON

MRI Head with contrast : T2 axial , T1 coronal , post gadolinium T1 axial / coronal / sagittal images of the brain and diffusion imaging . Comparison is made with the MRI of [DATE] . The right occipital / sub occipital craniectomy is noted with a pseudomeningocele extending within the sub occipital soft tissues (which is unchanged compared to previous imaging) . The resection of the C1 arch is also noted . The recurrent right cerebellar medullary angle cistern meningioma appears unchanged in size - measuring **3.6 cm** (parallel to the posterior

GT_TUMOUR
PRED_TUMOUR

petrous ridge) **x 1.3 to 1.4 cm** (perpendicular to the posterior petrous ridge) . The lesion is closely related to the right vertebral artery .

GT_TUMOUR
PRED_TUMOUR

There is extension to the right hypo glossal canal . Right hypoglossal denervation changes are noted . There is also note made of a small left lateral cerebello pontine angle cistern meningioma (remote to the IAM) which has increased in size from **8 - 9 mm** in maximum dimension .

GT_TUMOUR

Conclusion : There are static appearances to the right cerebellomedullary angle cistern meningioma and there is a minimal increase in the size of the left cerebello pontine angle cistern meningioma . Reported by **Tiffany Henry** on [DATE]

GT_PERSON
PRED_PERSON

Figure 3: Examples of ground truth and spaCy predictions on radiology reports. Ground truth labels are marked with "GT," and predictions with "PRED." Personal names ("PER") replaced with HIPS are highlighted in blue, and tumour sizes ("TUMOUR") in yellow. "[DATE]" is used as a placeholder for visualisation.

Method	Perfect Report Match (%)
Rule-Based - personal names	83.6
spaCy - personal names	97.06 ± 3.16
spaCy - tumour size	87.12 ± 5.66
spaCy - combined	81.72 ± 2.89

Table 1: Report level performance evaluation on the testing dataset. The average and standard deviation of perfectly predicted reports from models trained with 5-fold cross-validation.

CONCLUSION

The **spaCy model** trained to extract **both entities** demonstrated strong performance, achieving a **macro-averaged recall of 0.97 ± 0.02, precision of 0.95 ± 0.03,** and **an F1-score of 0.96 ± 0.01**. The model perfectly matched **81.72 ± 2.89%** of cases at the **report level.**

Challenges in NER tasks:
- **Easier task**: Identifying personal names.
- **Harder task**: Extracting tumour sizes.
- **Dataset imbalance**: Reports average 170 words, with 1.68 personal names vs. 0.68 tumour sizes mentions per report.
- **Dataset ambiguity**: References to other lesions (e.g., cysts) and reporting variability.

BERT models:
- Performance was **limited by the small dataset**.
- Expanding the dataset could improve results and allow for the use of Large Language Models (LLMs).

Future work:
- **Incorporate other identifiable data** (e.g., addresses, dates of birth) to enhance de-identification.

Reference

[1] D. Carrell et al., "Hiding in plain sight: use of realistic surrogates to reduce exposure of protected health information in clinical text," J Am Med Inform Assoc, vol. 20, no. 2, p. 342, 2013, doi: 10.1136/AMIAJNL-2012-001034. [2] "Open Source Data Labeling | Label Studio." Accessed: Aug. 09, 2024. [Online]. Available: https://labelstud.io/ [3] "spaCy · Industrial-strength Natural Language Processing in Python." Accessed: Aug. 09, 2024. [Online]. Available: https://spacy.io/ [4] Daniele Faraglia. Faker · pypi. Accessed: 2024-08-2

Acknowledgements & Funding

EPSRC CDT [EP/S022104/1]