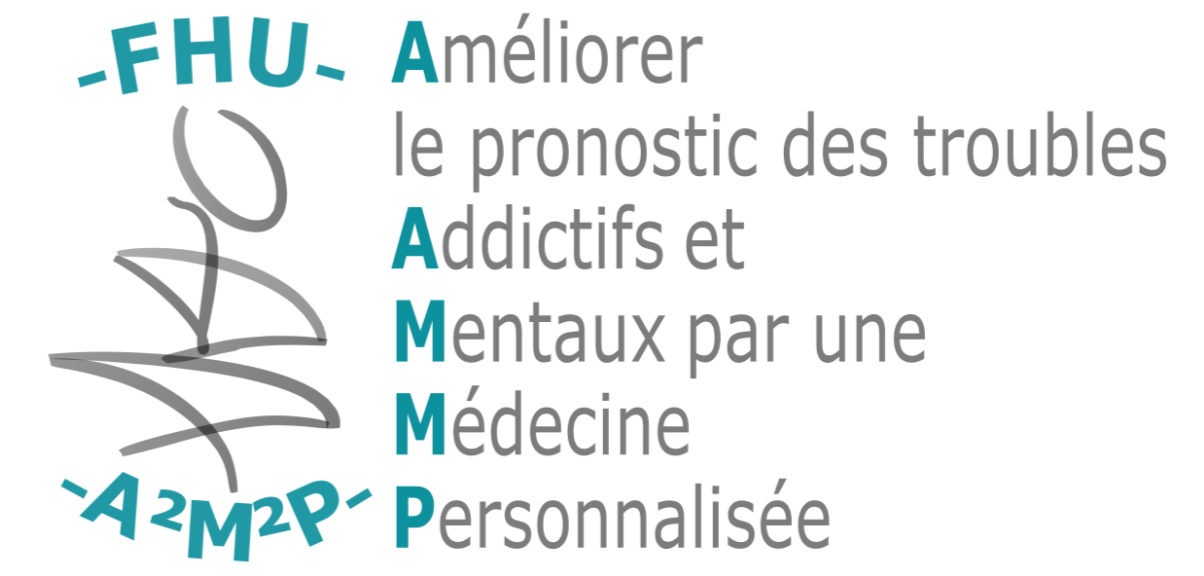


Analyse de l'estimation du niveau de dépression basée sur les symptômes à travers le prisme de l'expertise des psychiatres

Navneet Agarwal, Kirill Milintsevich, Lucie Metvier, Maud Rothärmel, Gaël Dias, Sonia Dollfus



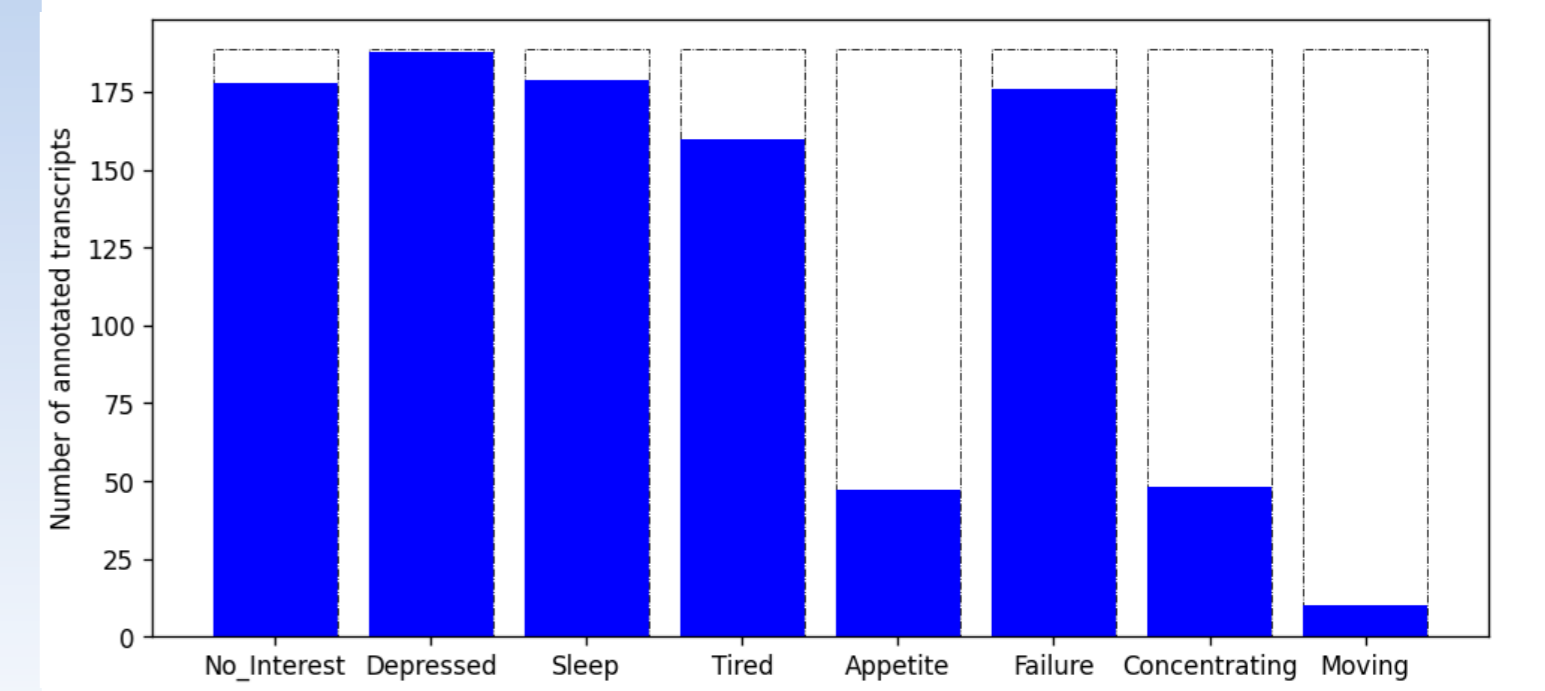
Abstract

Le nombre sans cesse croissant de patients souffrant de troubles mentaux a motivé d'importantes initiatives de recherche visant à automatiser l'estimation de la dépression. Malgré la nature multidisciplinaire de la tâche, très peu de ces approches ont inclus des professionnels de la santé dans le processus de recherche. Dans cet article, nous proposons de ramener l'expertise des psychiatres dans la boucle en incorporant leurs connaissances sous forme d'annotations dans l'ensemble de données DAIC-WOZ. Les résultats globaux démontrent une forte corrélation entre les annotations des psychiatres et les architectures à base de transformateurs, ce qui indique leur fiabilité en tant que modèles prédictifs.

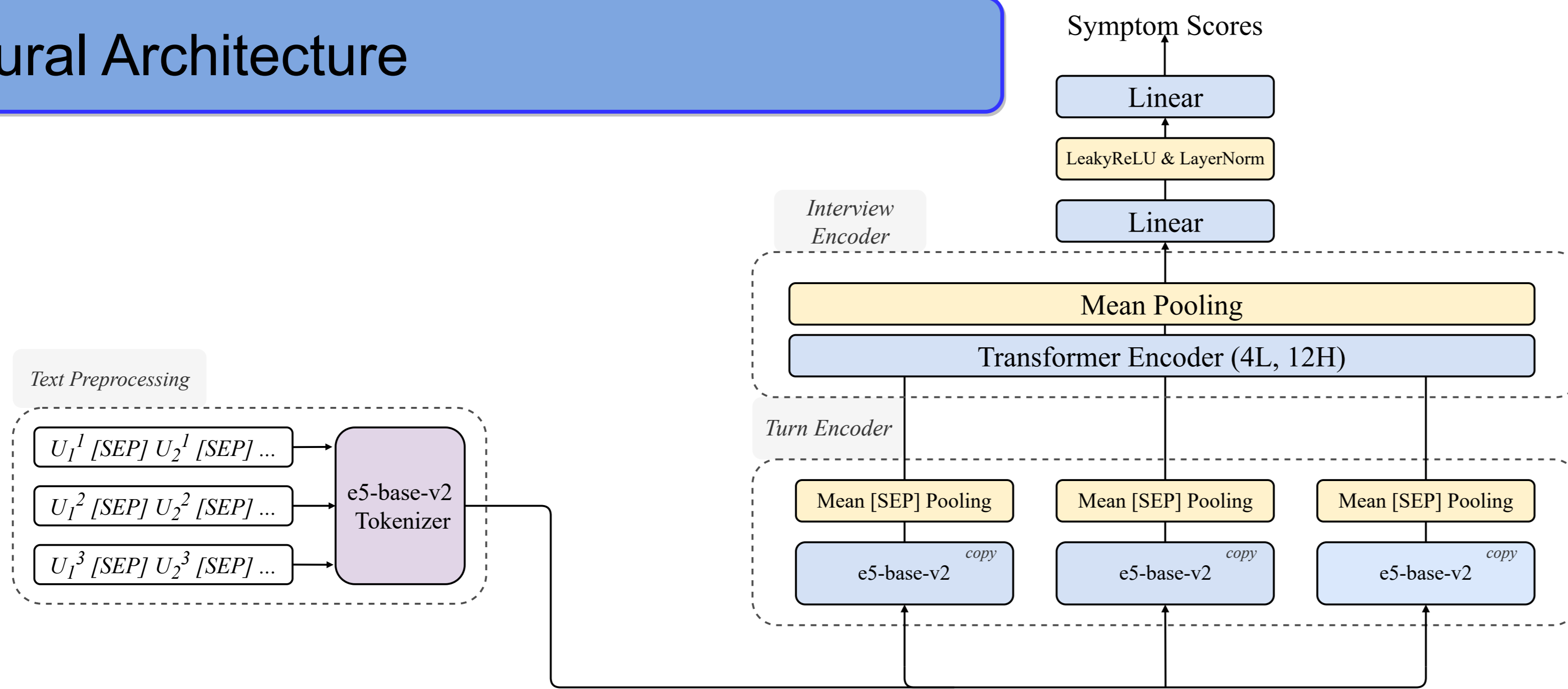
Annotations

- Réalisées par 3 psychiatres : un étudiant en doctorat, un médecin débutant et un professeur expérimenté.
- Nous sommes limités à une annotation de psychiatre par transcription en raison de contraintes de temps.
- Les annotateurs ont été totalement libres de souligner tout ce qu'ils jugeaient important, sans aucune contrainte prédéfinie, afin d'éviter toute partialité.
- Ils ont été encouragés à annoter non seulement le texte représentant la présence de la dépression, mais aussi celui représentant son absence. Aucune distinction n'a été faite entre les deux catégories.
- Les psychiatres ont également rempli le formulaire d'auto-évaluation PHQ-8 au nom des patients.

ELLIE: *how close are you to your family*
 PARTICIPANT: @@ very close @@ even though i don't live with them @@ i try to see them as much as possible @@
 ELLIE: *mhm*
 ELLIE: *how do you like your living situation*
 PARTICIPANT: *uh it's ok*



Neural Architecture



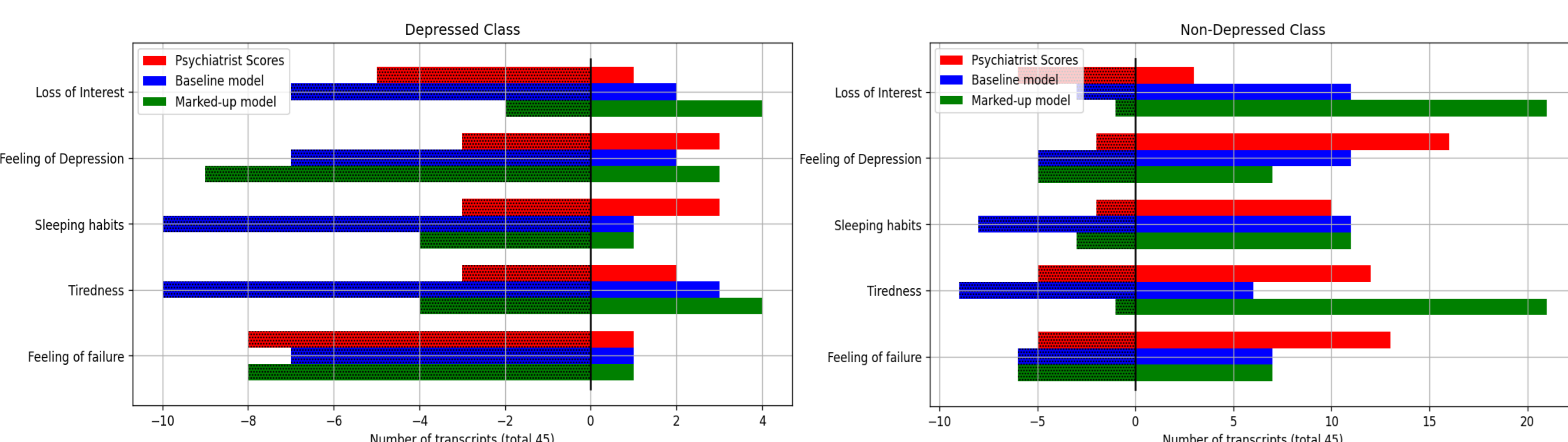
Over-all Performance

- Nous calculons les MAE entre les scores PHQ-8 des psychiatres et les auto-évaluations des patients afin d'avoir une perspective.
- Le modèle de base surpasse le modèle de marquage pour la plupart des symptômes, ce qui est attendu puisque les modèles automatisés s'occupent déjà des phrases annotées.

Symptoms	Psychiatrist Prediction		Baseline Model		Marked-up Model	
	Depr.	Non-Depr.	Depr.	Non-Depr.	Depr.	Non-Depr.
Loss of interest	0.615	0.366	0.611	0.431	0.699	0.485
Feeling of depression	0.571	0.696	0.884	0.443	0.939	0.465
Sleeping habits	0.615	0.533	0.761	0.691	0.651	0.808
Tiredness	0.727	0.689	0.797	0.711	0.812	0.666
Feeling of failure	1.083	0.800	0.820	0.543	0.786	0.573
Lack of concentration	-	-	1.332	0.521	1.316	0.475
Loss of appetite	-	-	0.932	0.745	1.037	0.628
Lack of movement	-	-	1.008	0.105	0.964	0.125

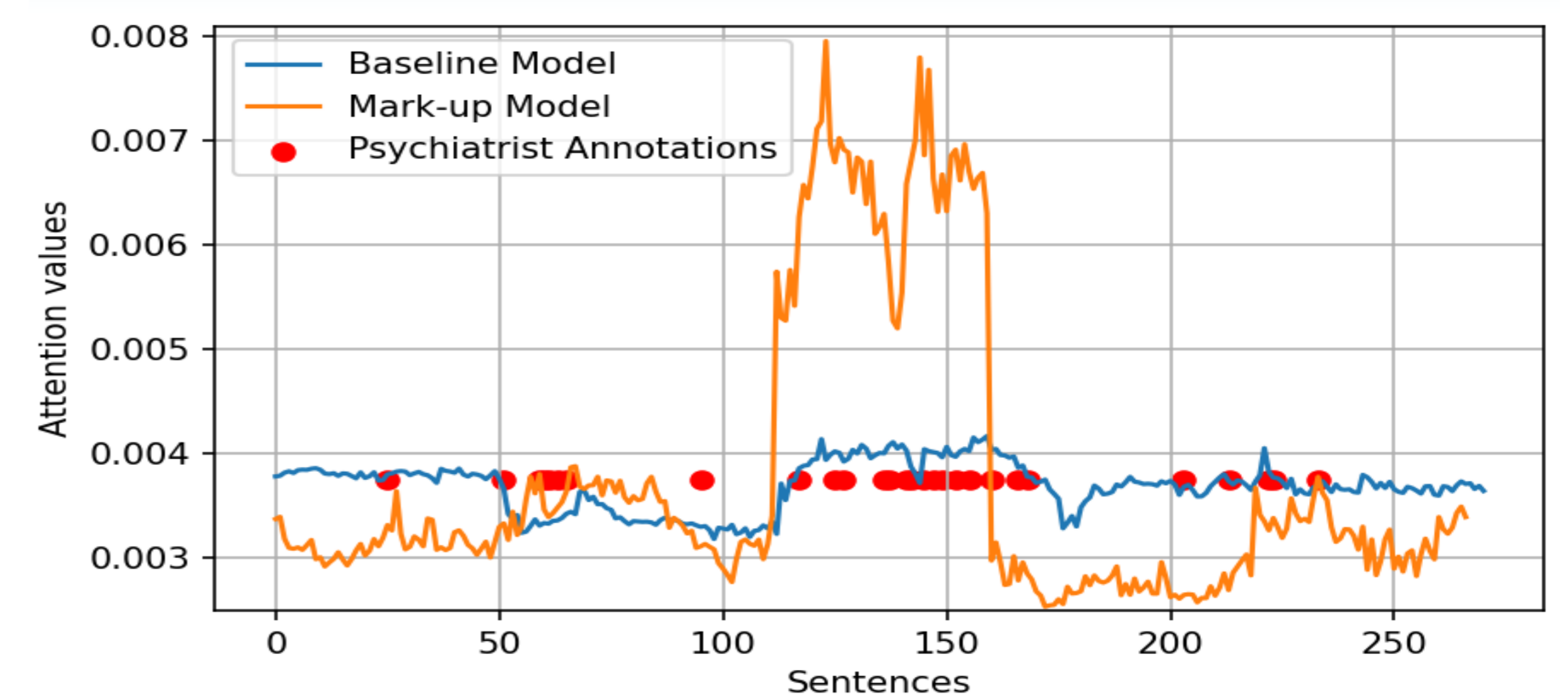
Psychiatrist scoring and self assessment

- La recherche médicale montre que les psychiatres ont tendance à surévaluer les scores des symptômes des patients non dépressifs et à sous-évaluer ceux des patients dépressifs.
- Nous observons le même comportement pour le modèle de base et le modèle majoré.
- Cela renforce l'affirmation selon laquelle les modèles automatisés et les psychiatres partagent les mêmes tendances psychologiques en ce qui concerne l'estimation de la dépression.



Model Attention and Psychiatrist Annotations

- Nous proposons de vérifier si les modèles automatisés s'intéressent aux mêmes informations que le psychiatre ou s'ils se concentrent sur des informations complémentaires.
- Les psychiatres ne s'intéressent qu'aux propos du patient alors que les modèles automatisés s'intéressent à la fois aux propos du patient et à ceux du thérapeute au cours de la conversation. Cela correspond au comportement attendu.
- Notre analyse montre une forte corrélation entre les deux agents, les réponses annotées recevant des scores d'attention relativement élevés dans les modèles automatisés



Class	Metric	Ques.	Non-ann.	Ann.
		12.84	12.93	13.60
Non-depressed	min.	12.84	12.93	13.60
	max.	137.50	136.76	135.35
	med.	42.03	42.10	42.25
	avg.	30.85	31.01	31.25
Depressed	min.	15.29	15.02	15.37
	max.	103.88	102.83	110.89
	med.	37.96	38.50	38.82
	avg.	12.18	12.18	12.29

References

- Milintsevich, Kirill, Kairit Sirts, and Gaël Dias. "Towards automatic text-based estimation of depression through symptom prediction." *Brain Informatics* 10.1 (2023): 1-14.
- Domken, Marc, Jan Scott, and Peter Kelly. "What factors predict discrepancies between self and observer ratings of depression?." *Journal of affective disorders* 31.4 (1994): 253-259.
- Gratch J, Artstein R, Lucas GM, Stratou G, Scherer S, Nazarian A, Wood R, Boberg J, DeVault D, Marsella S, Traum DR. The Distress Analysis Interview Corpus of Human and Computer Interviews. In *Proceedings of LREC 2014 May* (pp. 3123-3128).

Acknowledgements

This research is supported by the FHU A²M²P project funded by the G4 University Hospitals of Amiens, Caen, Lille and Rouen (France). We would like to thank the High Performance Computing Center of the University of Tartu for providing the GPUs for training.