



Soutenance de Mémoire de Fin d'Etude

en vue de l'obtention du diplôme de Master

SCIENCE DE DONNEES ET BIG DATA

LA DETECTION DES FAKE NEWS EN UTILISANT DEEP LEARNING

- ❑ Figurer une solution pour détecter fake news en utilisant deep Learning
- ❑ Discuter les différents modèles des contributions précédentes
- ❑ Proposer une contribution pour lutter contre Fake News
- ❑ Evaluer la contribution avec différentes métriques

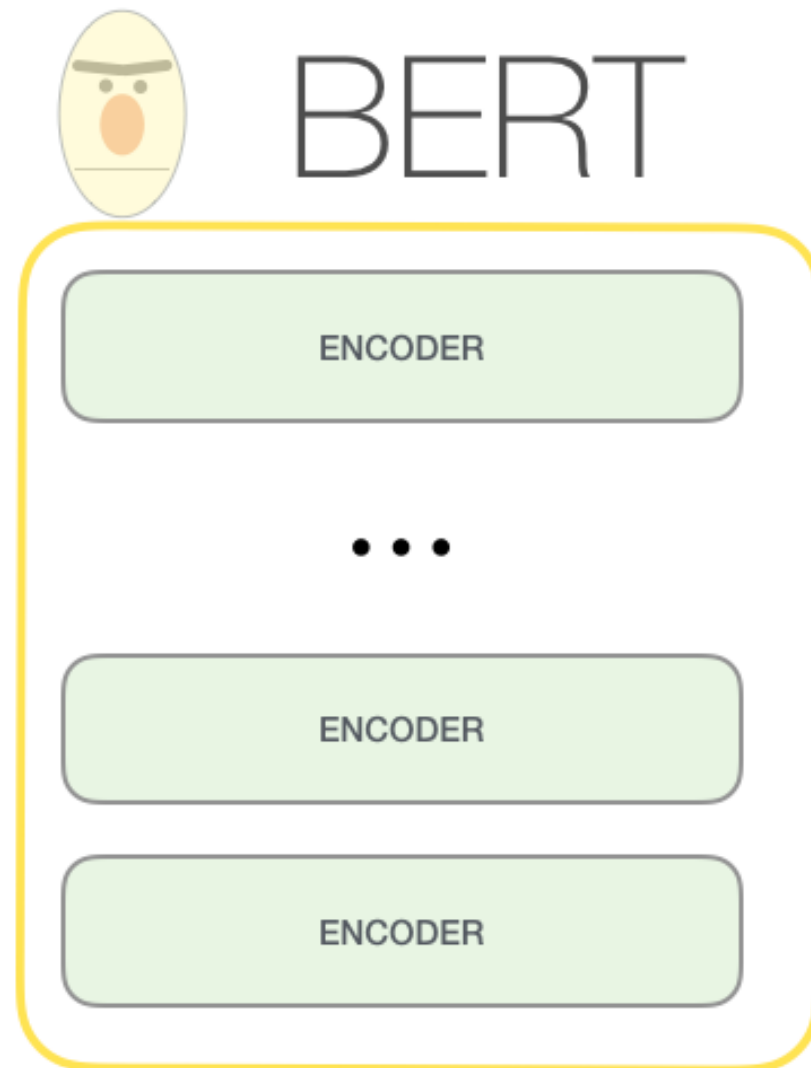
3. Objectifs

BERT Embeddings

L'architecture des transformateurs

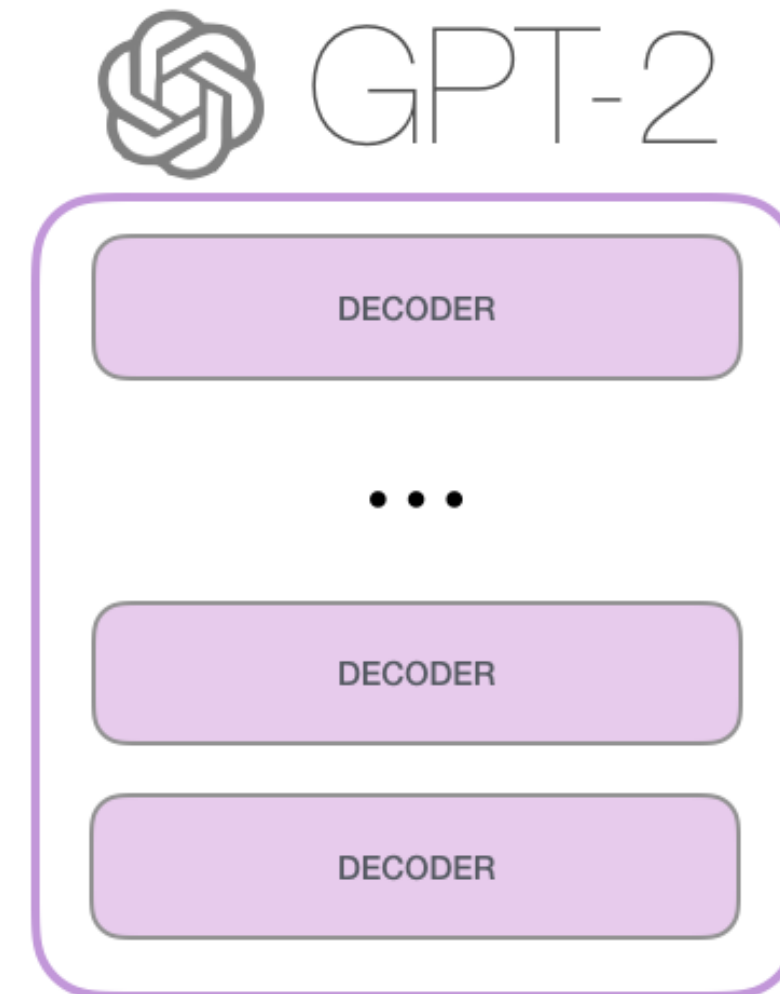
Encoders:

Génère des embeddings pour chaque mot
(langue, la grammaire et le contexte)



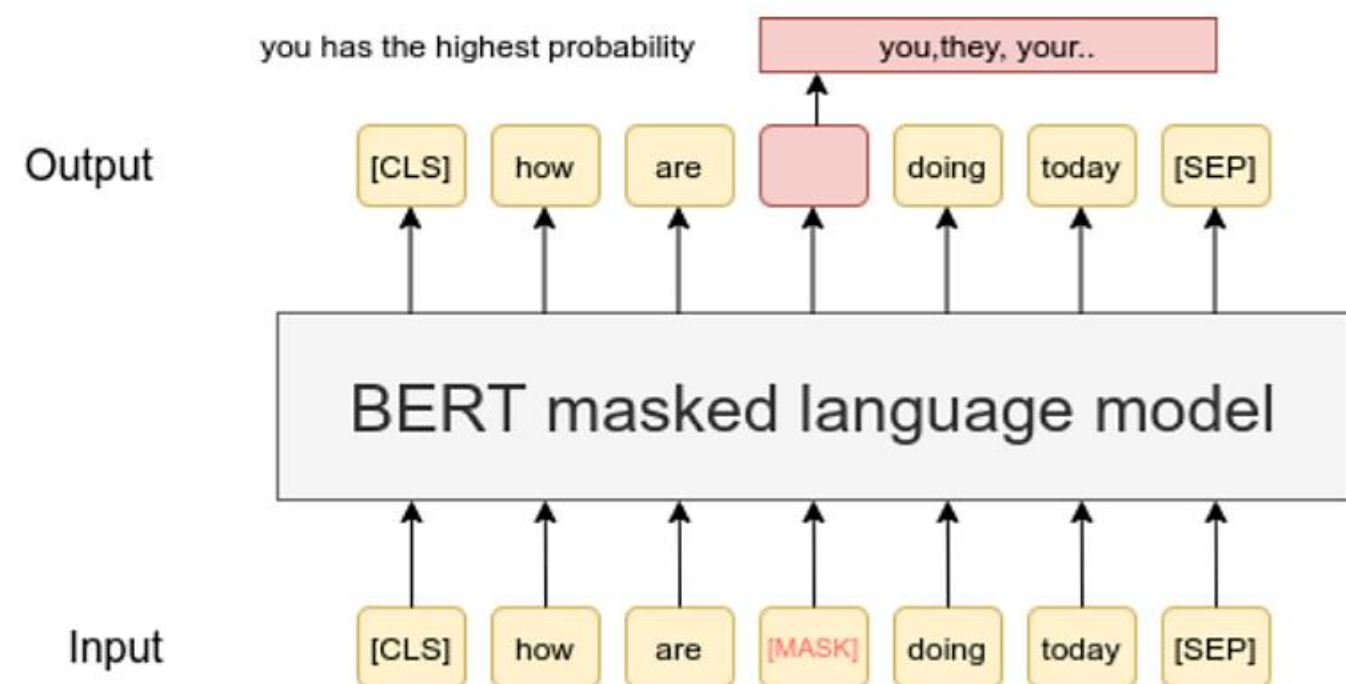
Decoders:

Utilise les embeddings de l'encodeur pour
générer le prochain mot (relation entre les mots)

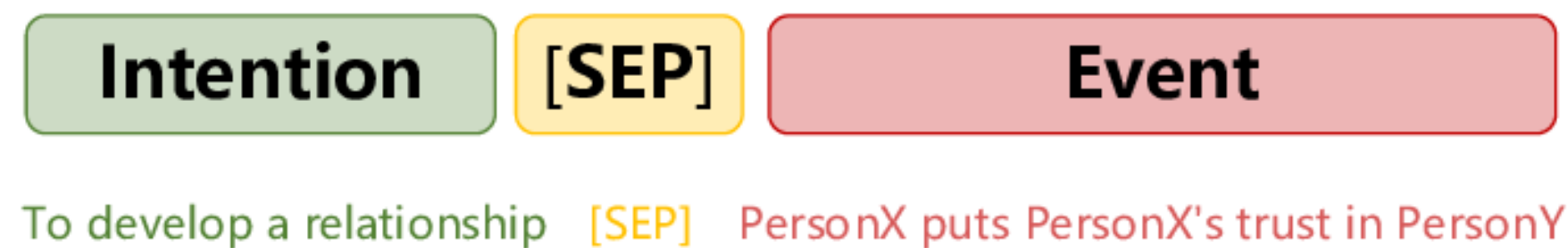


PRETRAINING BERT

- Mask language Modeling



- Next sentence prediction

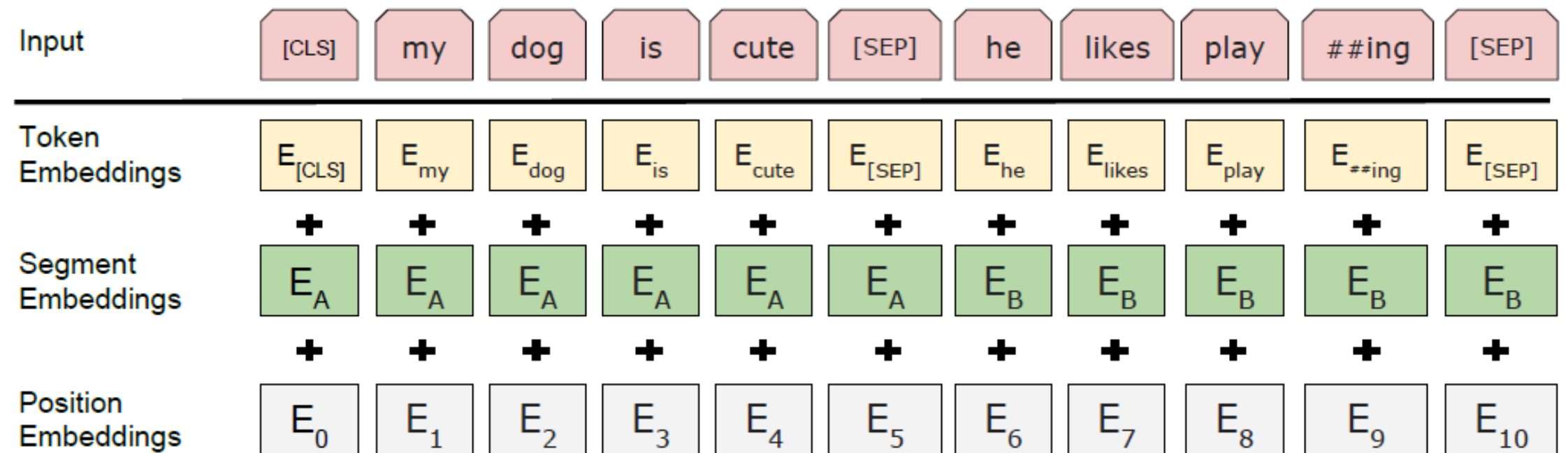


BERT Embeddings

- **Embeddings de Tokens** : sont les embeddings pré-entraînées pour différents mots

- **Les embeddings de segments**: sont essentiellement le numéro de la phrase qui est codé dans un vecteur

- **Les embeddings de position**: sont la position du mot dans la phrase qui est encodée dans un vecteur



Mise en œuvre

News	Size (Number of articles)	Subjects	
Real-News	21417	Type	Articles size
		<i>World-News</i>	<i>10145</i>
		<i>Politics-News</i>	<i>11272</i>
Fake-News	23481	Type	Articles size
		<i>Government-News</i>	<i>1570</i>
		<i>Middle-east</i>	<i>778</i>
		<i>US News</i>	<i>783</i>
		<i>left-news</i>	<i>4459</i>
		<i>politics</i>	<i>6841</i>
		<i>News</i>	<i>9050</i>

DATASET ISOT



- Les vrais articles ont été obtenus de Reuters
- Les articles de fausses nouvelles sont obtenus à partir de différentes sources telles que des sites Web qui ont été signalés par Politifact
- Titre ,sujet,text,label



Hyperparametres

Modèle de BERT	Bert uncased L-12 H-768 A-12 (12 layers)
Fonction de perte	Binary Cross Entropy
Fonction d'activation (couche cachées)	RELU
Fonction d'activation (couche de sortie)	Sigmoïde
Batch size	(1,32,64,128,512)
Optmisateur	Adam,Adagrad
Dropout	0,5

Outils utilisés

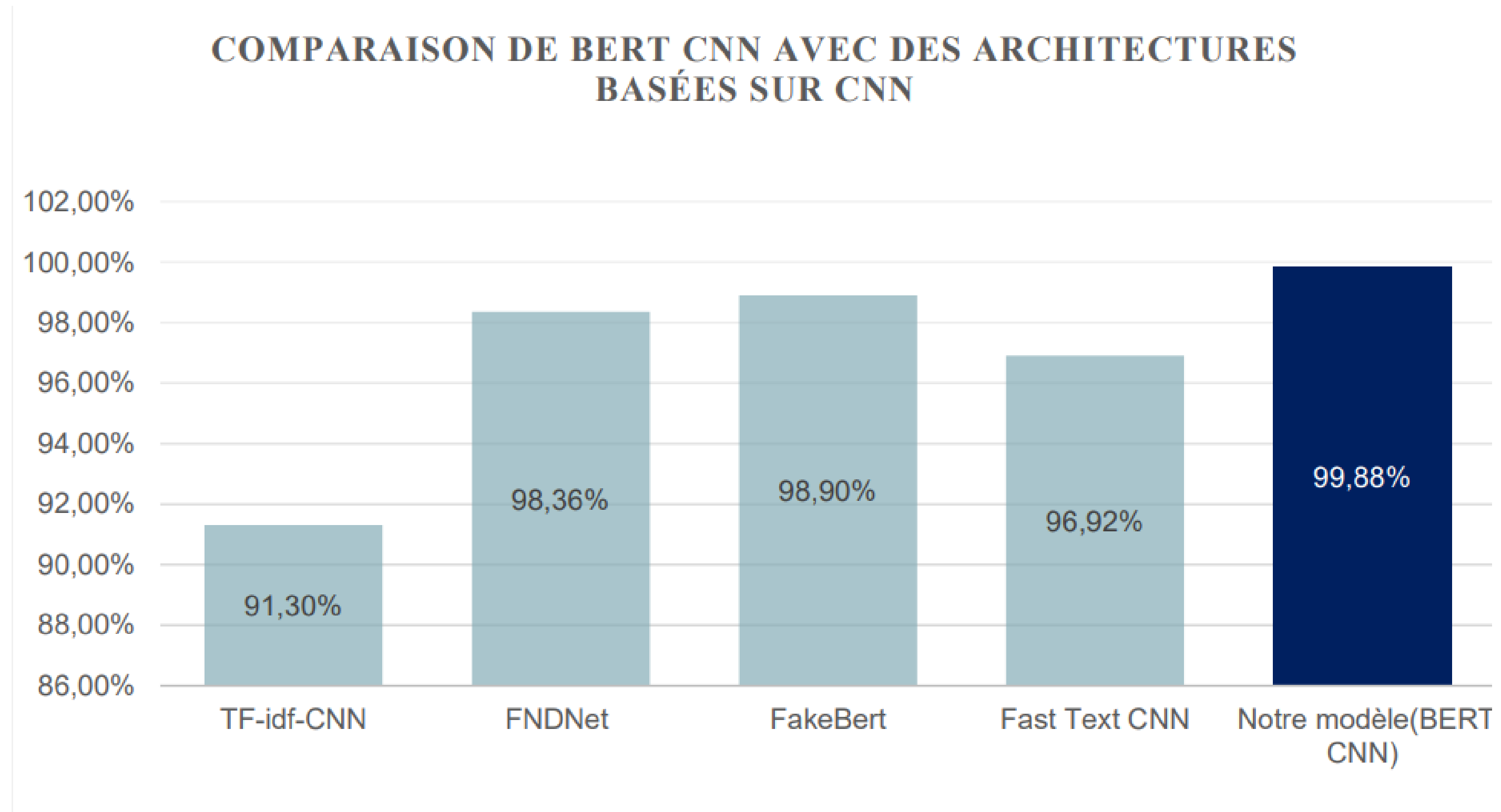




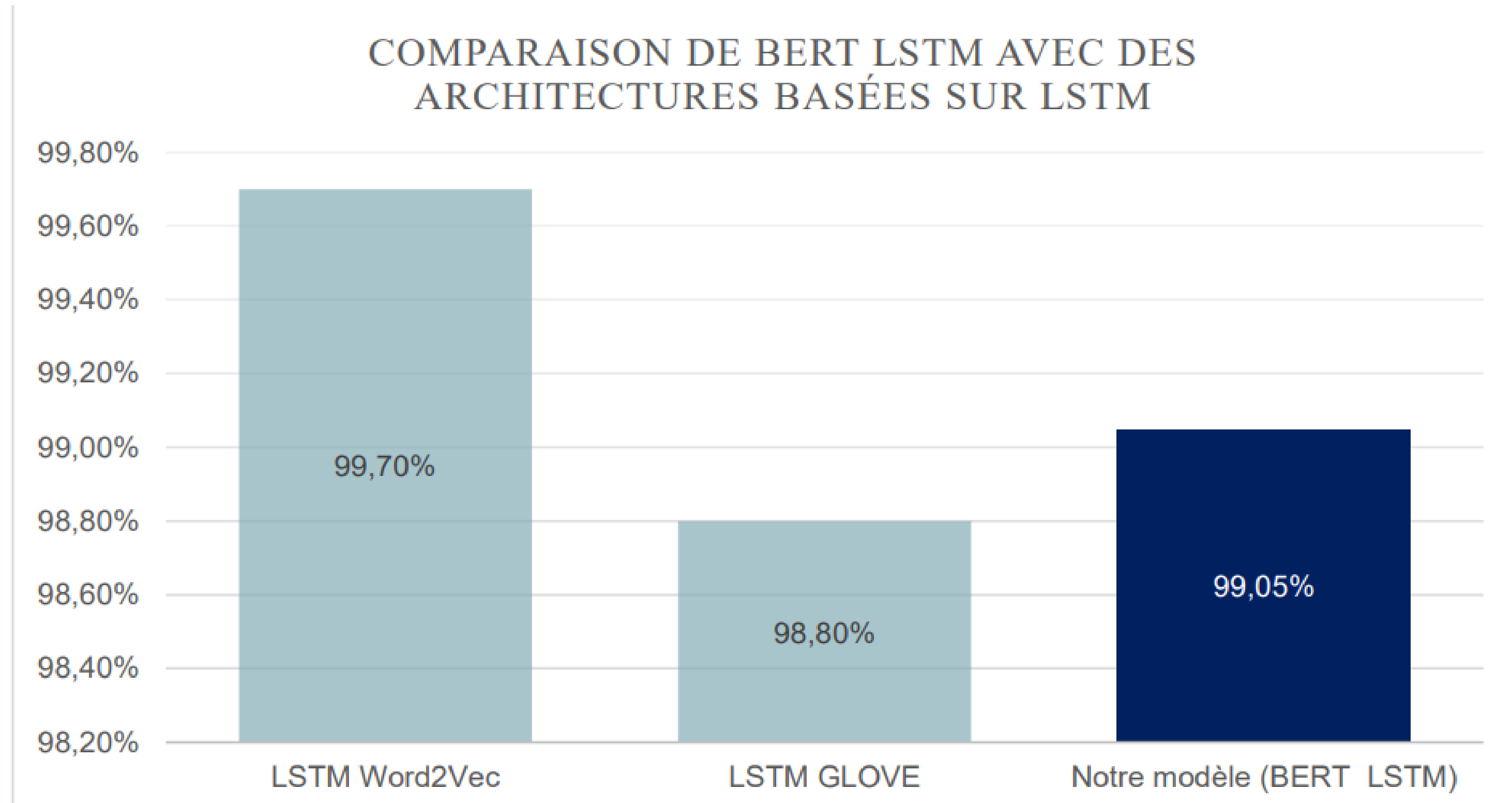
Résultats et discussion

LORS DE MON PROJET DE RECHERCHE, J'AI DISCUTÉ EN DÉTAIL DES RÉSULTATS QUE J'AI OBTENUS À PARTIR DE MES MODÈLES. EN PLUS DE PRÉSENTER MES RÉSULTATS, JE LES AI COMPARÉS À CEUX DE LA REVUE DE LA LITTÉRATURE QUI UTILISAIT LE MÊME ENSEMBLE DE DONNÉES. CE FAISANT, J'AI PU VALIDER LA PRÉCISION DE MES RÉSULTATS ET OBTENIR DES INFORMATIONS SUR LES RECHERCHES EXISTANTES. CELA M'A PERMIS DE FOURNIR UNE ANALYSE PLUS SOLIDE DE MES RÉSULTATS ET D'ÉVALUER LA PERTINENCE DE MA RECHERCHE DANS LE CONTEXTE PLUS LARGE DU DOMAINE. DANS L'ENSEMBLE, LA COMPARAISON DE MES RÉSULTATS AVEC CEUX DE L'ANALYSE DOCUMENTAIRE M'A PERMIS D'ENRICHIR LES CONNAISSANCES EXISTANTES ET DE CONTRIBUER AU DÉVELOPPEMENT DE LA RECHERCHE.

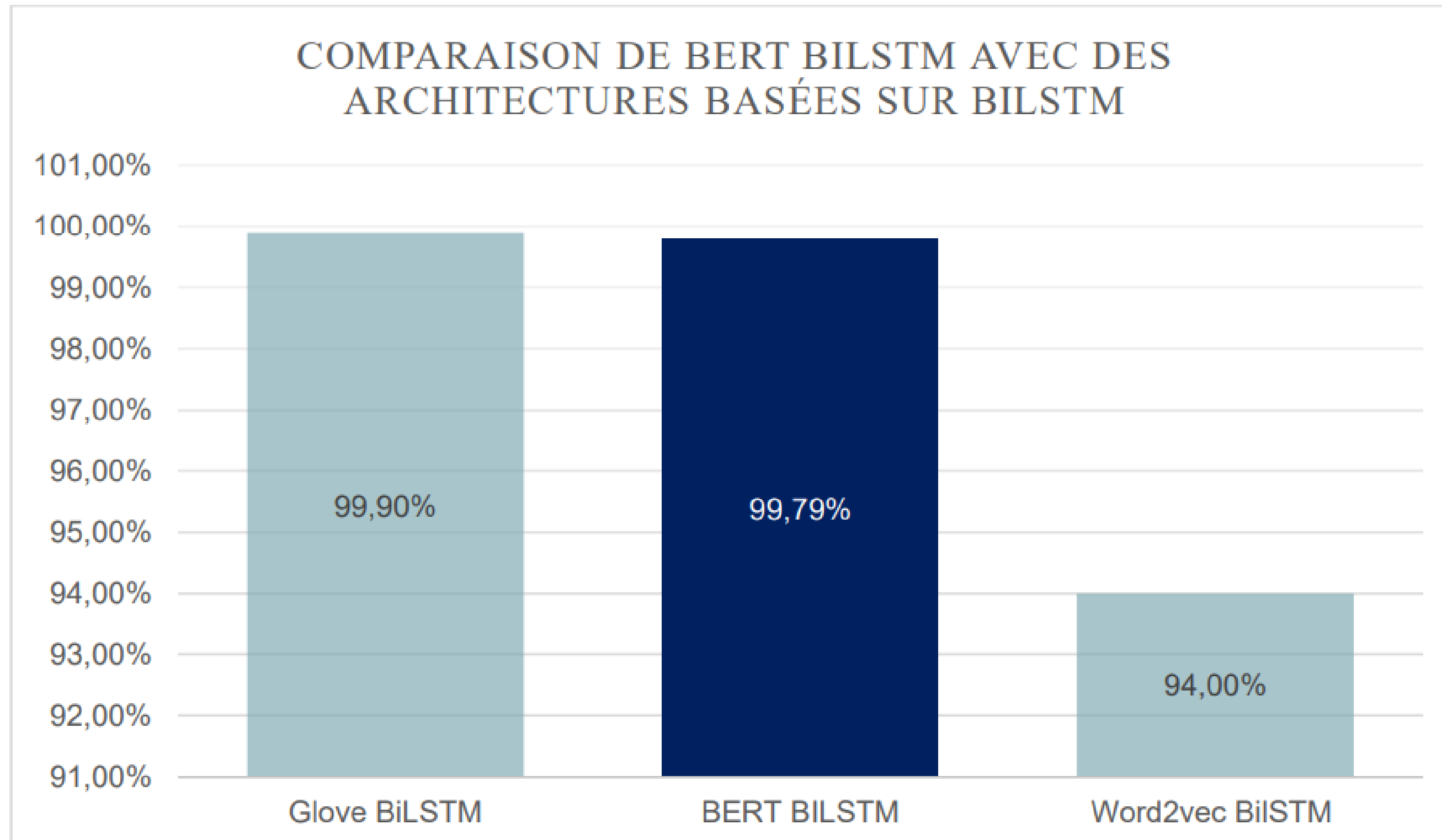
1 Evaluation de BERT CNN



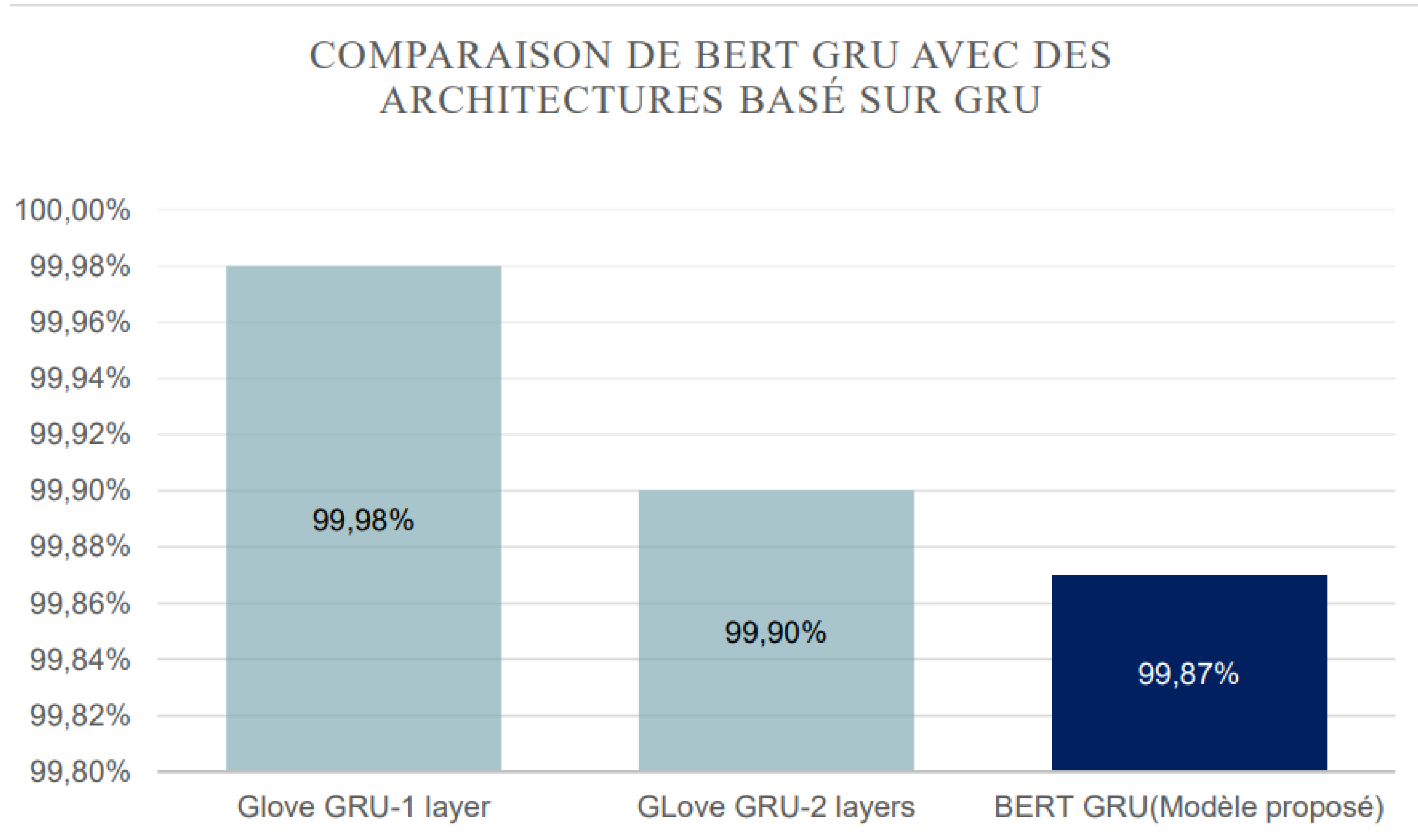
2 Evaluation de BERT LSTM



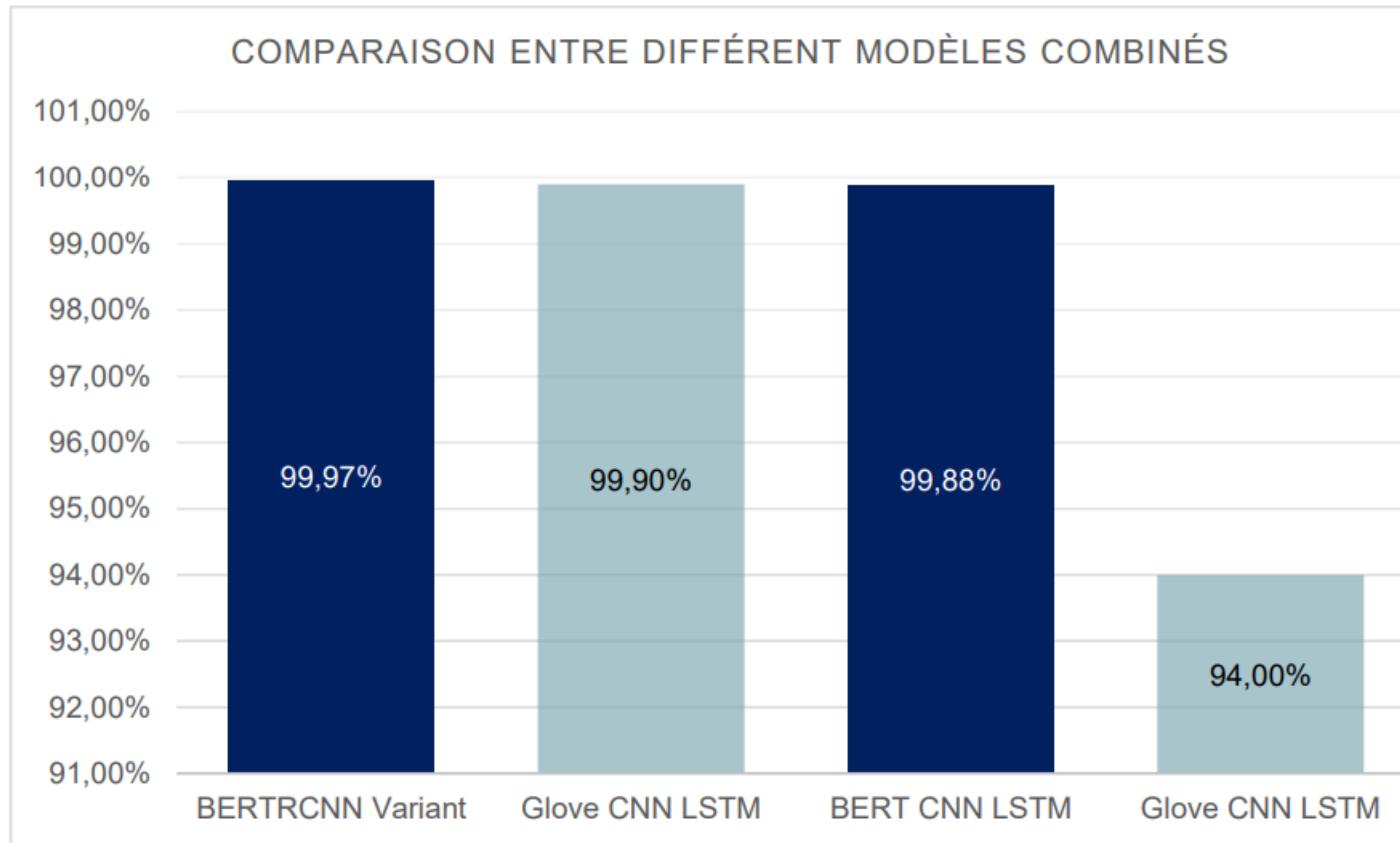
3 Evaluation de BERT BiLSTM



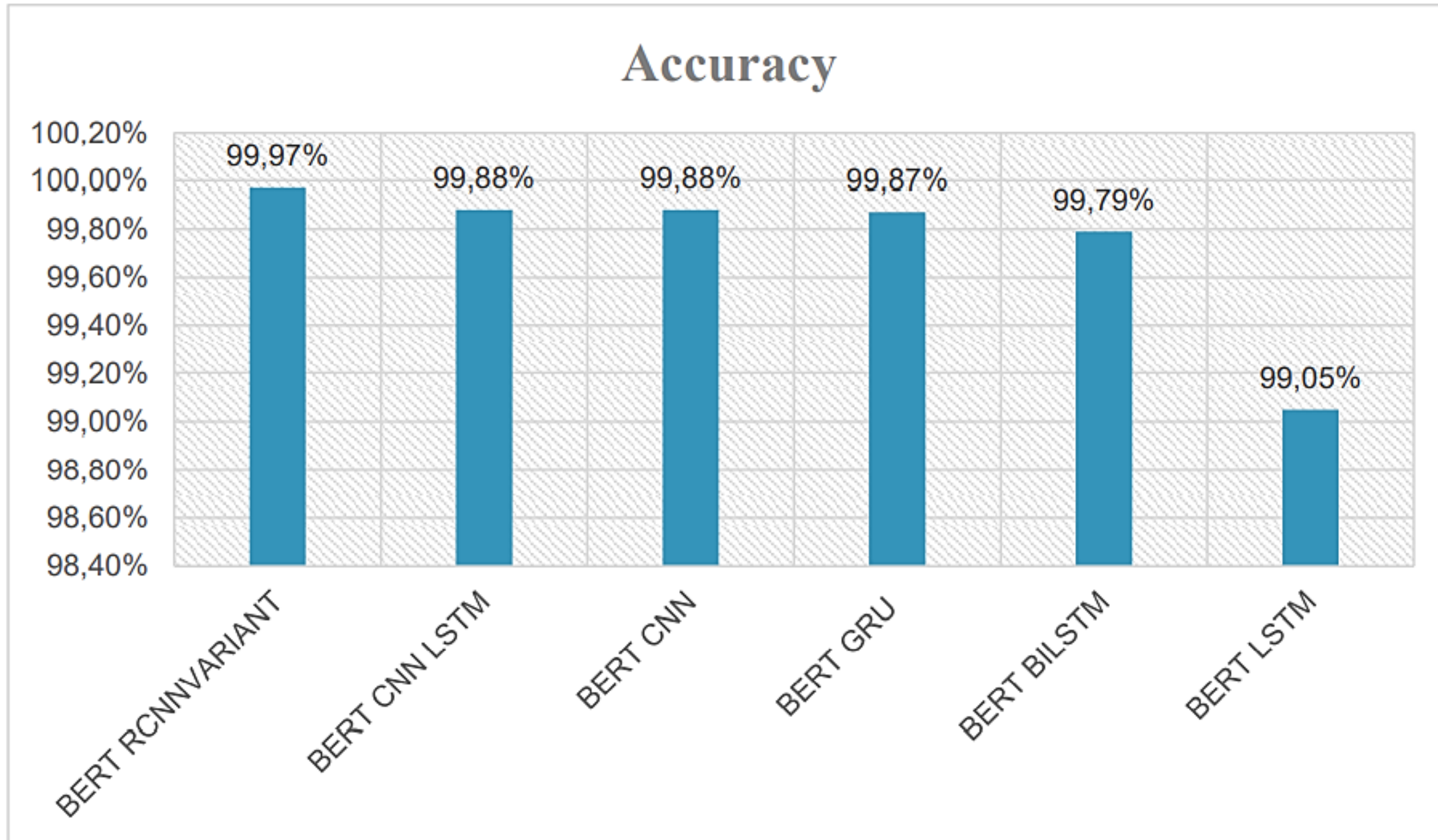
4 Evaluation de BERT GRU



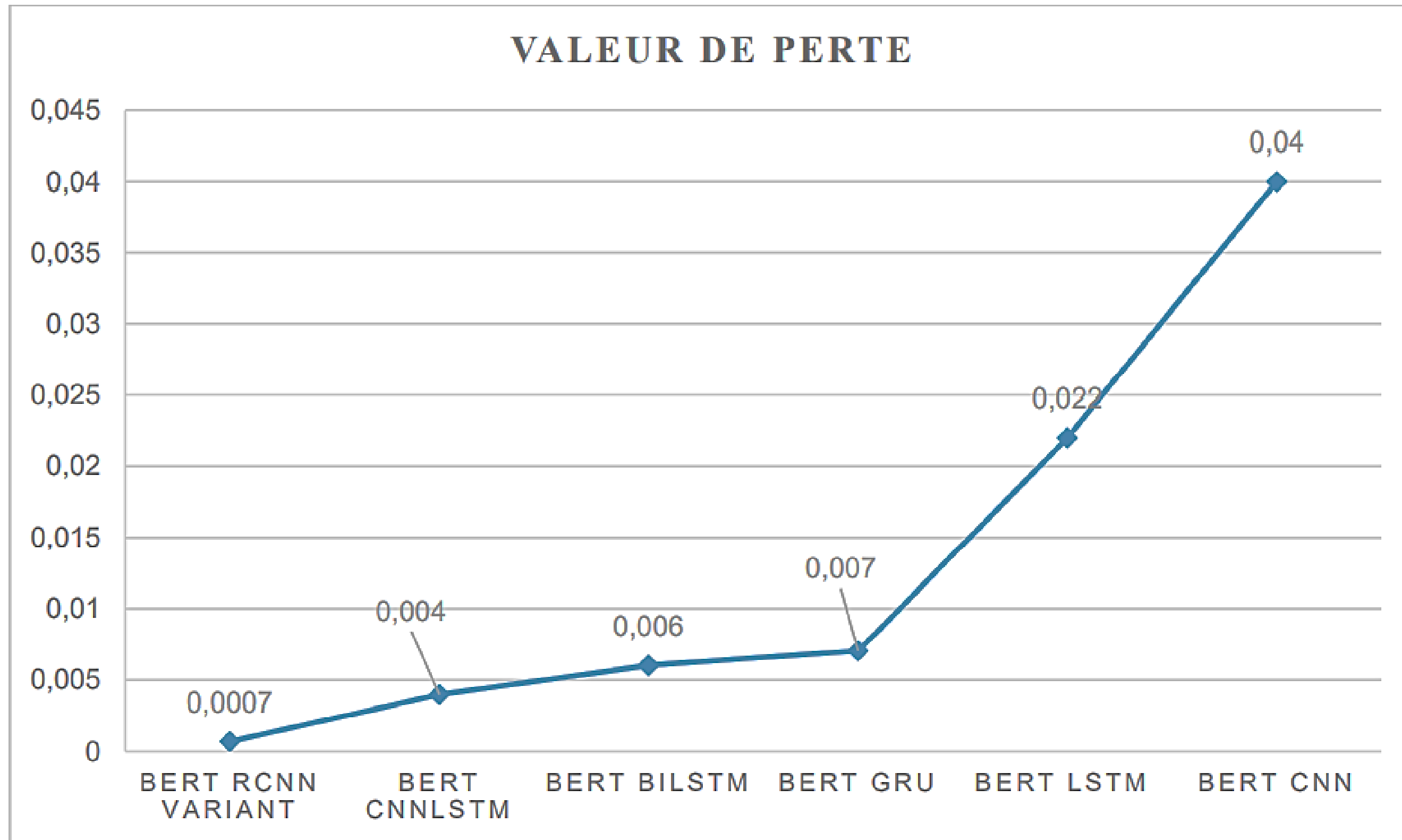
5 Evaluation de BERT CNN-LSTM et BERT RCNN VARIANT



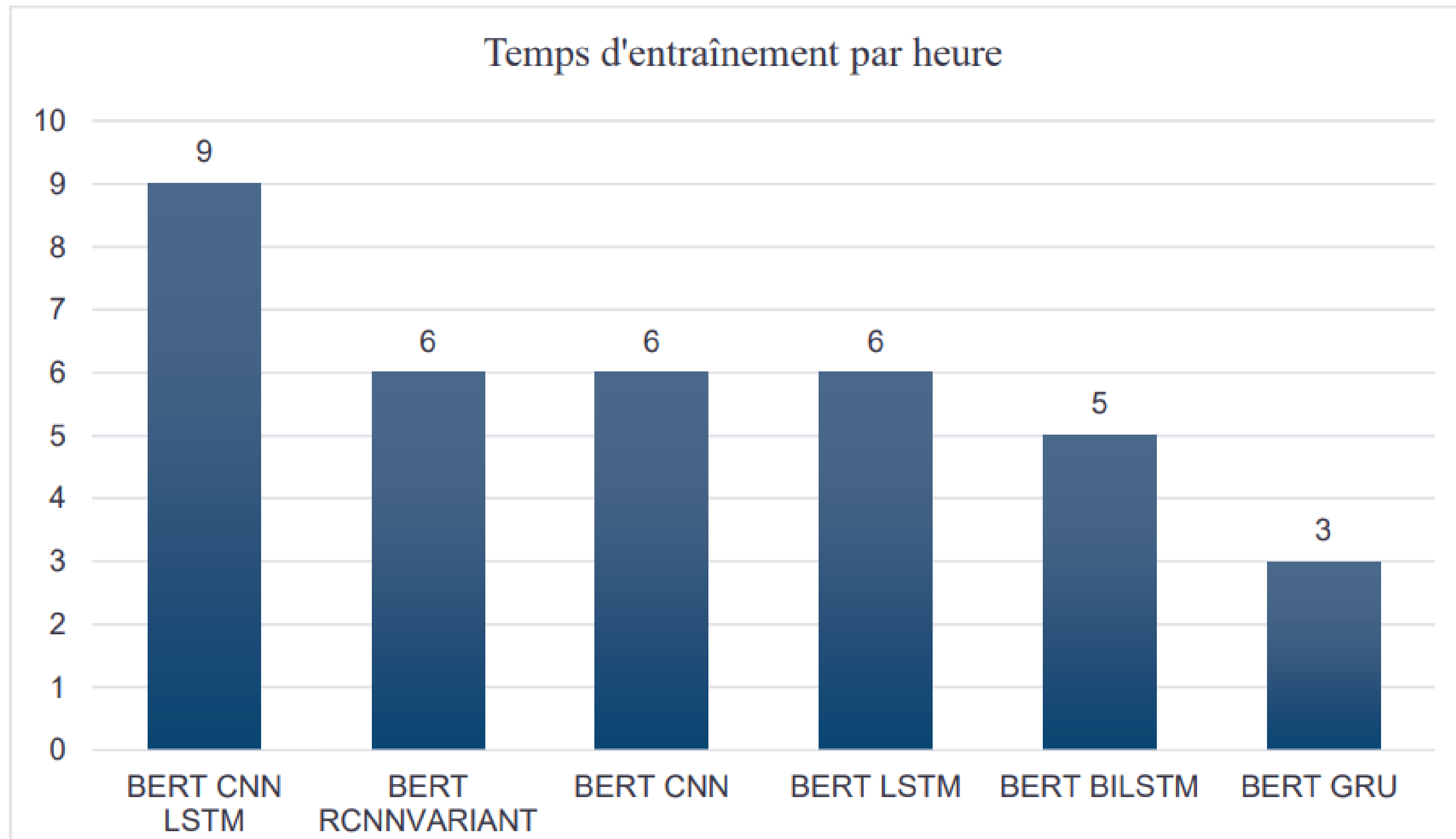
6 Comparaison des entre les contributions (Accuray)



7 Comparaison des entre les contributions (Valeur de perte)



8 Comparaison des contributions (Temps d'entraînement)





Conclusion

- ❑ Le choix de la technique d'embeddings joue un grand rôle pour la performance de modèles presque ils capture la signification des mot.
- ❑ BERT pour fournir les meilleures représentations contextuelles des mots.
- ❑ La contribution de RCNNVariant a montrer une supériorité pour la détection de fake News

Conclusion