

# Classification automatique de rapports médicaux par spécialité

## Introduction

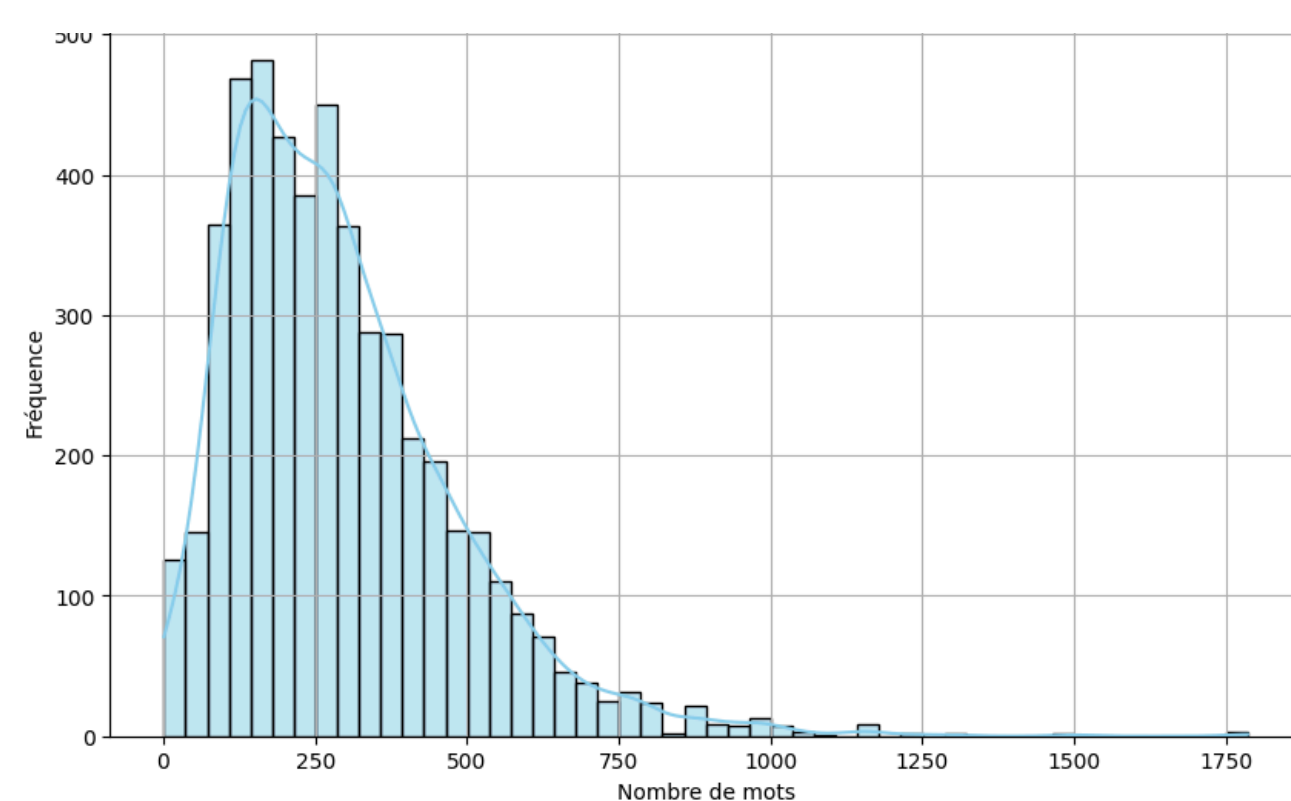
Dans les établissements de santé, les professionnels rédigent quotidiennement de nombreux rapports médicaux sous forme de textes libres. Ces documents, bien que riches en informations, sont souvent non structurés, ce qui complique leur exploitation. Classer ces rapports selon leur spécialité médicale est une tâche essentielle mais fastidieuse si elle est effectuée manuellement. Ce projet vise à répondre à ce besoin en développant une solution automatique, rapide et fiable pour faciliter l'organisation et l'analyse de ces données textuelles.

## Methodologie

Le projet repose sur le jeu de données Medical Transcriptions de Kaggle, composé de rapports médicaux courts et anonymisés, chacun étiqueté par spécialité (cardiologie, radiologie, etc.). Après un nettoyage des textes (suppression de la ponctuation, des mots vides, etc.), ceux-ci sont vectorisés à l'aide de la méthode TF-IDF. Un modèle de classification supervisée de type Fully Connected Neural Network (FCNN) a ensuite été entraîné. Les performances du modèle ont été évaluées à l'aide de métriques standard telles que l'accuracy, ainsi que par une matrice de confusion.

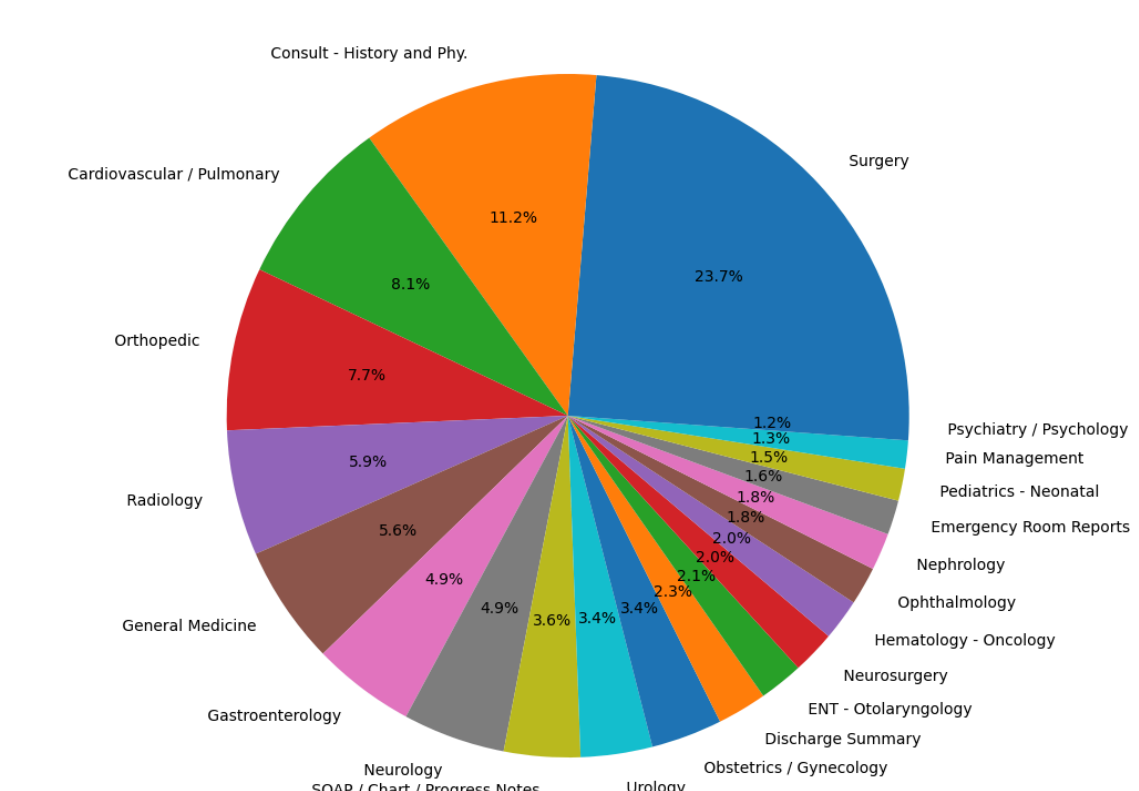
## Résultats

### Distribution de la longueur des transcriptions



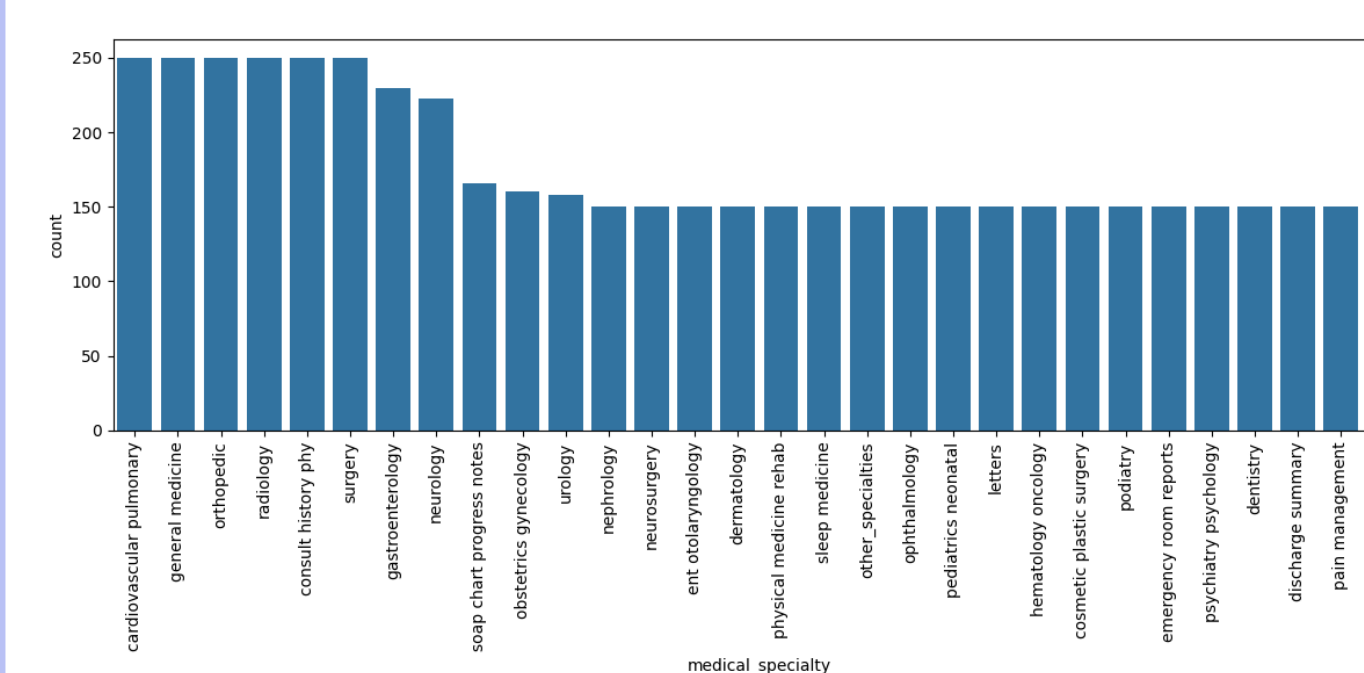
La majorité des transcriptions contiennent entre 100 et 400 mots, avec quelques cas plus longs.

### Distribution des rapports médicaux selon la spécialité



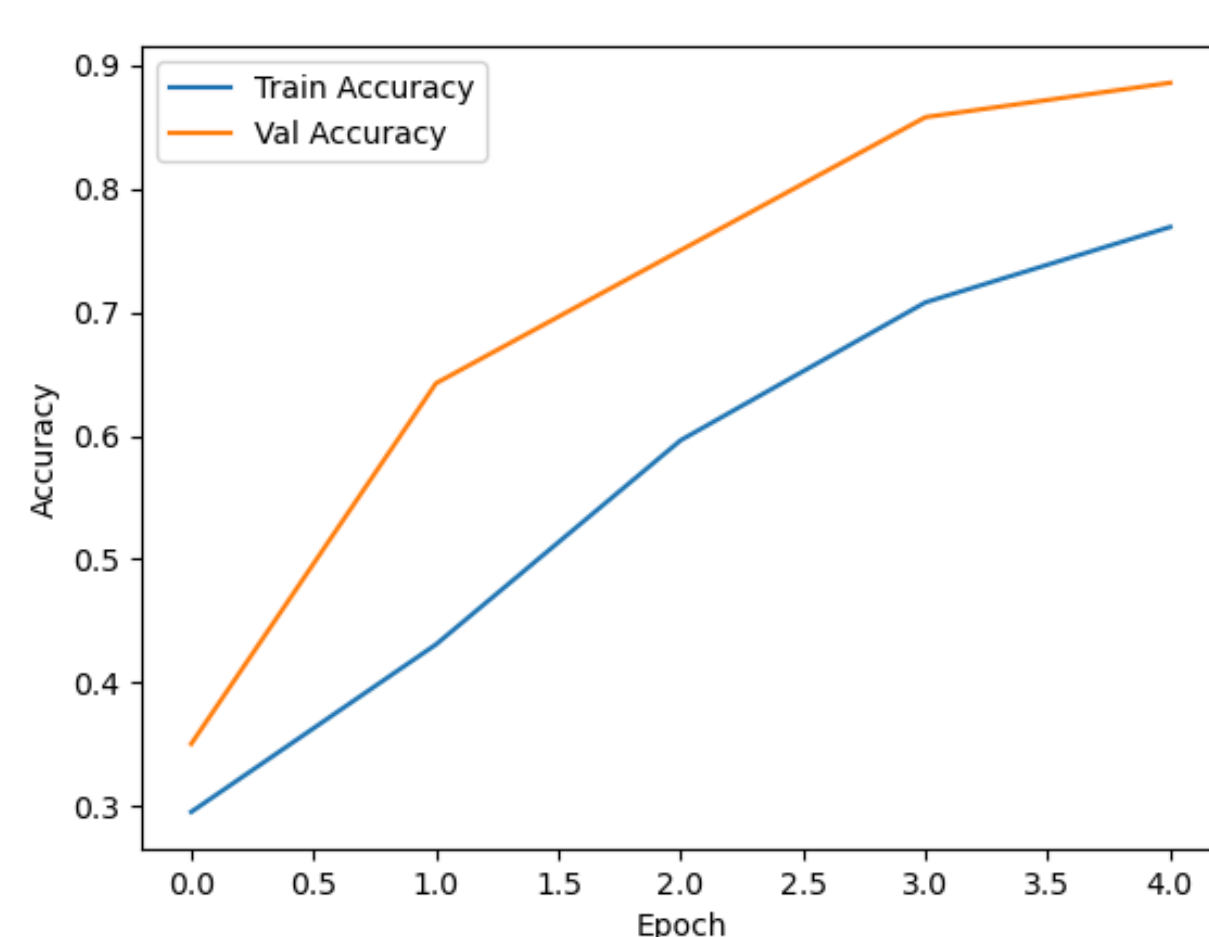
La chirurgie domine, suivie de l'anamnèse et des pathologies cardiovasculaires, reflétant la priorité des actes cliniques et chirurgicaux.

### Équilibrage des données



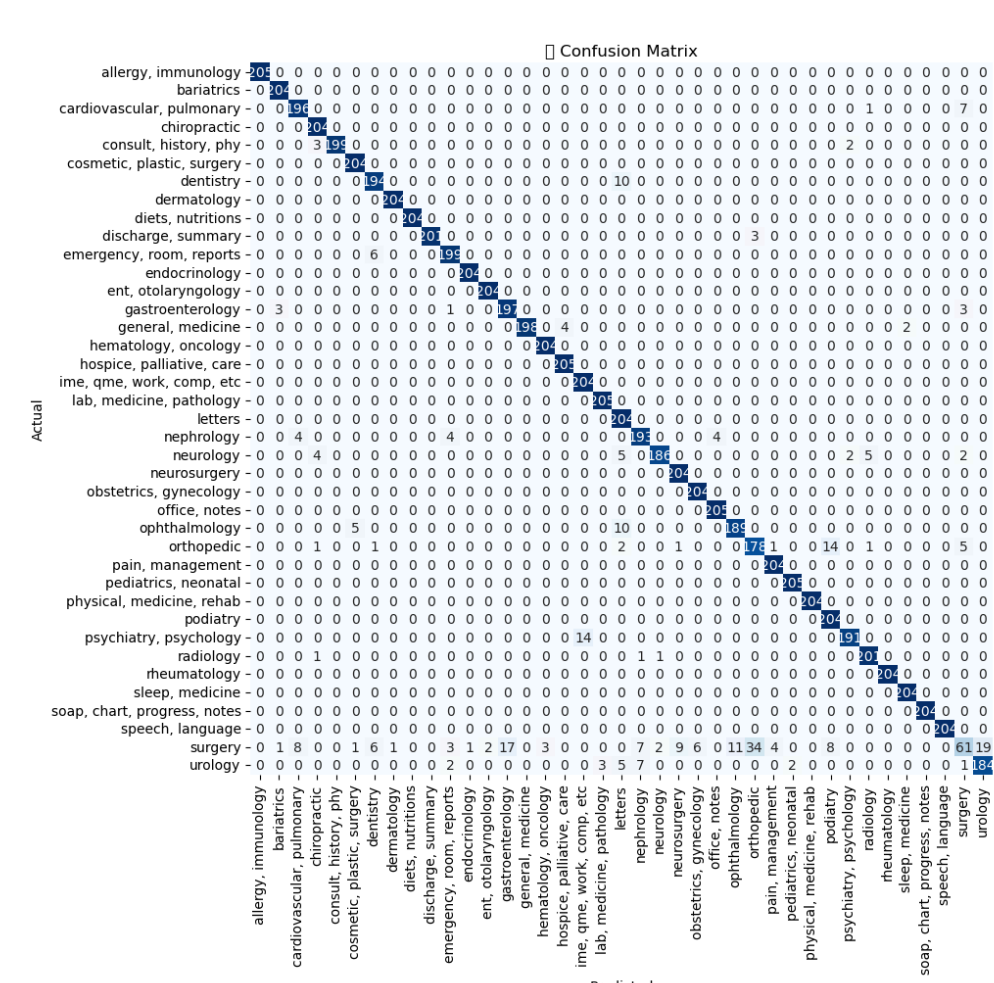
La distribution des spécialités médicales a été équilibrée, ce qui favorise un apprentissage sans biais significatif.

### Progression de la précision du modèle



Le graphique montre une amélioration continue de la précision du modèle FCNN, atteignant 87 % en validation sans surapprentissage.

### Matrice de Confusion



La matrice de confusion montre une bonne performance globale du modèle avec des prédictions majoritairement correctes.

## Interface

**Medical Specialty Classifier**  
Advanced AI-powered classification of medical transcriptions

**Quick Start**  
How to use:  
1. Enter or paste a medical transcription  
2. Click 'Analyze Transcription'  
3. View predicted specialty and confidence score

**Model Info**  
Total Samples: 39

**Sample Transcripts**  
Try a sample: Cardiovascular, Pulmonary

**Session Stats**  
Predictions Made: 0

**Clinical Transcription Input**  
Enter medical transcription:  
Patient presents with chest pain radiating to left arm. ECG shows ST elevation in leads I, II, aVF. Troponin levels elevated at 2.5 ng/mL. Patient has history of hypertension and diabetes. Recommend cardiac catheterization and initiate dual antiplatelet therapy.

**About This Tool**  
Purpose: This AI model classifies medical transcriptions into their appropriate medical specialties.  
Accuracy: Trained on thousands of medical documents for high precision.  
Use Cases:  
• Clinical documentation routing  
• Medical record organization  
• Specialty referral assistance

**Prediction Results**  
Primary Specialty: cardiovascular, pulmonary  
Confidence Score: 36.8%

Une interface intuitive permettant à l'utilisateur de saisir un rapport médical et d'obtenir instantanément la spécialité correspondante grâce au modèle de classification.

## Conclusion

Ce projet a pour objectif de classer automatiquement des rapports médicaux courts et anonymisés selon leur spécialité (telles que la cardiologie, la radiologie ou l'oncologie). En s'appuyant sur des techniques de traitement automatique du langage naturel (NLP), un pipeline a été mis en place incluant le prétraitement des textes, leur vectorisation via TF-IDF, et l'entraînement de modèles de classification supervisée. Parmi les approches testées, un réseau de neurones entièrement connecté (FCNN) a été utilisé pour améliorer la performance sur des classes parfois proches ou déséquilibrées. Les résultats obtenus montrent que ce type de modèle peut efficacement structurer et organiser des données médicales non structurées, ouvrant la voie à une intégration dans des systèmes d'information hospitaliers.

