

Title: The Use of EffectiveCAN with Confidence Levels for Automated ICD-10-CM/PCS Coding of French Hospital Stay Records.

Introduction Healthcare budgets, particularly in hospitals, are increasingly under pressure. Coupled with the challenge of recruiting, training, and retaining qualified coding staff, there is a growing need for AI-driven solutions to support coding professionals. Currently, no reliable AI-based coding system exists for French medical documents. This study evaluates the application of the Effective Convolutional Attention Network (EffectiveCAN) for the automated coding of French hospital stay records.

Methods EffectiveCAN is a deep learning model designed for multi-label document classification (MLDC), with a focus on medical code prediction. It employs a deep convolution-based encoder integrating squeeze-and-excitation (SE) networks and residual connections to enhance text representations. Multi-layer attention is applied to extract informative features from different encoding layers, and sum-pooling attention is used for datasets with limited training samples. To improve performance on infrequent labels, the model combines binary cross-entropy loss with focal loss. EffectiveCAN has previously demonstrated strong performance, achieving F1 scores of 67.6% for diagnoses and 65.5% for procedures on Dutch ambulatory hospital stay records.

For this study, the model was trained on a multi-year corpus of de-identified French hospital documents coded by experienced ICD-10-CM/PCS professionals. Patient names were anonymized while preserving eponyms (e.g., "Huntington's disease" remained unchanged, whereas "the disease of our patient Mr. Huntington" was altered to "the disease of our patient Mr. Smith"). The model was evaluated on its ability to predict the principal diagnosis (ICD-10-CM), main procedure (ICD-10-PCS), and Diagnosis-Related Group (APR-DRG v40, 3M™).

Results Performance metrics were calculated based on confidence levels assigned by the model. The distribution of hospital stays according to confidence levels was:

- High confidence: 18%
- Medium confidence: 16%
- Low confidence: 65%

The results per Confidence Level are as follows:

Confidence Level	Precision	Recall	F1 Score
High	92%	93%	92%
Medium*	54%	59%	56%
Low	75%	54%	63%

*The Medium Confidence category was affected by several records with missing documentation. If these 'empty' records were excluded, the F1 score for Medium Confidence would improve from 56% to 74%.

When tested on historical data, the overall model performance was:

- **Precision:** 74.8%
- **Recall:** 65.9%
- **F1 Score:** 70.1%

Discussion & Conclusion

While the results are promising, fully autonomous AI-driven coding for all French medical texts remains unattainable at this stage. However, leveraging the model's Confidence Levels allows for selective automation of hospital stays with high certainty, thereby reducing the workload of human coders. Continuous improvements in AI methodologies indicate a positive trajectory, bringing the system closer to reliable automated coding. This study demonstrates that NLP-based AI can successfully assist in coding ambulatory hospital stays, although human oversight remains necessary. Future iterations aim to enhance model accuracy, particularly for cases currently classified under Medium Confidence, further bridging the gap toward trustworthy unsupervised medical coding.