Journal of Machine Intelligence in Healthcare

E-ISSN: XXXX-XXXX
P-ISSN: XXXX-XXXX

Research Article

Brazilian Neural Approaches for Automated Assignment of ICD-10 Codes from Portuguese-Language Clinical Narratives

Vicente Pironti [1]* and Salomeh Keyhani [2]

[1] Open University Humaniza, Brazil; e-mail : vicentepironti@gmail.com
[2] Department of Chemical and Biomolecular Engineering, University of California, Berkeley, United States; e-mail : s.keyhani@garnet.berkeley.edu
* Corresponding Author : vicentepironti@gmail.com

Abstract: Manual assignment of ICD-10 codes from clinical narratives remains a vital yet labor-intensive task for healthcare systems worldwide. While deep learning models have advanced the automation of this process, particularly in English-language datasets, less attention has been given to the linguistic and philosophical dimensions of multilingual clinical NLP. This paper presents a comparative study of machine learning and neural models (Logistic Regression, CNN, GRU, and CNN with per-label attention) applied to Brazilian-Portuguese clinical narratives. In doing so, we introduce a document concatenation strategy to address sparse-text limitations and demonstrate that attention-based models outperform classical baselines. Beyond quantitative improvements, this study integrates philosophical insights drawn from Asavika Sciences, a transdisciplinary framework rooted in compassion, purpose, and systemic intelligence, and incorporates principles from the "New Philosophy for Health," including the Sense of Life, the Sense of Death, and the concept of Infinitando (infinite unfolding). These perspectives reframe automated coding not merely as classification, but as a symbolic act of honoring human experience in digital systems. Additionally, we conduct a qualitative and quantitative comparison between English and Brazilian Portuguese, revealing that, when adequately modeled, the latter offers superior semantic clarity and context resolution in attention-based NLP systems. This is attributed to its morphological richness and emotional nuance, which make it especially suited for humanized health informatics. Our results affirm the feasibility and ethical promise of linguistically-aware, philosophically-aligned AI in global healthcare environments.

Keywords: ICD-10 coding; Brazilian Portuguese NLP; deep learning; clinical narratives; Asavika Science; health humanization; Infinitando; multilingual NLP

Received: December 28, 2024
Revised: February 11, 2025
Accepted: March 17, 2025
Published: April 28, 2025
Curr. Ver.: April 30, 2025

1. Introduction

The International Classification of Diseases (ICD) is a global standard maintained by the World Health Organization for classifying diagnoses, symptoms, and procedures in medical records. Assigning ICD codes to clinical narratives is essential for healthcare billing, administrative reporting, epidemiological tracking, and the creation of structured electronic health records (EHRs) [1]. However, the manual coding process is time-consuming, costly, and highly susceptible to human error due to the complexity of medical texts and the granularity of the ICD taxonomy [2,3].

In recent years, automated ICD coding has become a critical research area in clinical natural language processing (NLP). Early works relied on traditional machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, and Logistic Regression with bag-of-words features [4,5]. While interpretable and computationally efficient, these methods struggle to capture contextual and semantic nuances in free-text clinical notes, especially in multi-label classification scenarios involving thousands of ICD codes with severe class imbalance [6].

Deep learning approaches, especially Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention-based models, have demonstrated superior performance in recent ICD coding benchmarks [7–9]. The incorporation of contextual word embeddings and hierarchical attention mechanisms has further improved predictive accuracy and interpretability [10]. Notably, most of these studies have used English-language datasets such as MIMIC-III [11], limiting their applicability to non-English clinical contexts.

However, as explored in recent interdisciplinary approaches, particularly through the framework of Asavika Sciences and the New Philosophy for Health developed by Dr. hc Vicente Pironti, the act of coding diagnoses is not merely a technical operation but a symbolic expression of the human experience of illness, healing, and mortality. These perspectives emphasize the Sense of Life, Sense of Death, and the concept of Infinitando, a philosophical formulation of infinite human potential rooted in mathematics and spirituality, as complementary dimensions in any health-related classification system. From this standpoint, the automation of ICD-10 code assignment is reimagined not just as optimization, but as an act of ethical, linguistic, and ontological interpretation.

Brazilian Portuguese is one such under-resourced language where automatic ICD coding remains underexplored. Existing studies on Portuguese-language EHRs are scarce and often constrained by small datasets, limited code sets, or a focus on death certificate classification rather than diagnostic coding from clinical notes [12–14]. Furthermore, Brazilian EHRs tend to be shorter and less structured than their English counterparts, presenting additional challenges for information extraction and semantic modeling.

Importantly, Brazilian Portuguese carries significant linguistic advantages for semantic modeling in NLP, such as morphological richness, expressive redundancy, and flexible syntactic structures, which, when integrated into attention-based architectures, provide an enhanced understanding of context, emotion, and meaning. These linguistic characteristics also resonate with the Asavika principle of compassionate intelligence, which encourages technology to interpret language beyond syntax, tapping into its emotional, cultural, and humanistic depth.

This study addresses the gap by developing and evaluating neural models for ICD-10 diagnostic code prediction from Brazilian-Portuguese clinical narratives. We propose a comparative analysis of four modeling approaches: Logistic Regression with TF-IDF features, a CNN, a GRU-based RNN, and a CNN with per-label attention. To overcome the sparsity of discharge summaries, we introduce a document concatenation strategy that integrates multiple types of clinical notes per admission, including anamnesis and clinical evolution texts.

The key contributions of this paper are as follows:

- We develop and evaluate neural architectures for ICD-10 code prediction on Brazilian-Portuguese clinical narratives, an under-researched language in clinical NLP.
- We introduce a document concatenation strategy that significantly improves model performance in sparse-text scenarios.
- We present a comparative evaluation on both the MIMIC-III and a real-world Brazilian hospital dataset, demonstrating generalizability across languages and data settings.
- We contextualize automated ICD coding within a broader epistemological and philosophical framework, guided by the principles of Asavika Sciences and the New Philosophy for Health.
- Our optimized CNN-Att model outperforms prior baselines and achieves state-of-the-art performance within its class.

The remainder of this paper is structured as follows: Section 2 reviews related work in ICD code prediction. Section 3 describes the datasets and preprocessing steps. Section 4 details the proposed methods and model architectures. Section 5 presents experimental results and discussion. Section 6 concludes with a summary of findings, philosophical implications, and future directions.

2. Literature Review

The task of automated ICD code assignment from clinical narratives has been studied for over two decades. Early methods relied on rule-based systems and classical machine learning, while recent approaches leverage deep learning models, particularly in the context of English-language datasets. This section outlines the main trends in the literature, organized into traditional machine learning approaches, neural architectures, and multilingual applications.

2.1. Traditional Machine Learning for ICD Coding

Initial studies in automated medical coding employed traditional classifiers such as Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (k-NN), and Logistic Regression [1,2]. These models typically used bag-of-words or TF-IDF features derived from discharge summaries or clinical notes. For instance, Crammer et al. [3] proposed a rule-based and statistical approach for ICD-9 assignment, while Medori and Fairon [4] explored feature selection for semi-automatic ICD-9-CM encoding.

Although these methods are computationally efficient and interpretable, their performance is constrained by the inability to capture syntactic and semantic nuances in unstructured clinical text. Moreover, most traditional approaches fail to scale effectively to large ICD ontologies with thousands of labels and exhibit poor generalization in multi-label, imbalanced settings [5].

2.2. Deep Learning Approaches

The advent of deep learning has substantially improved performance in ICD coding. Convolutional Neural Networks (CNNs) have been widely adopted due to their ability to capture local context patterns [6,7], while Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) variants, are capable of modeling longer-range dependencies in text [8].

Attention-based mechanisms have further advanced this field by enabling label-specific focus on relevant portions of the input. One of the most influential models is CAML (Convolutional Attention for Multi-Label classification) proposed by Mullenbach et al. [9], which introduced per-label attention to enhance explainability and precision in ICD-9 code prediction. Variants and improvements of this architecture have since been applied across multiple clinical datasets, primarily using the MIMIC-III corpus [10].

Recent works have also integrated contextualized embeddings, such as ELMo and BERT, for ICD tagging, achieving incremental gains over static embeddings [11]. However, these models often require substantial computational resources and large-scale corpora for pretraining.

2.3 Multilingual and Low-Resource Clinical Texts

Despite progress in English-language ICD coding, research in multilingual or non-English clinical contexts remains limited. For Brazilian Portuguese, only a handful of studies exist. Santos et al. [12] used CNNs with self-taught word embeddings for a small set of ICD codes. Duarte et al. [13] explored ICD-10 coding from Portuguese death certificates using attention-based RNNs, while Oleynik et al. [14] focused on ICD-O classification from pathology reports.

These studies often suffer from narrow scope, limited code coverage, or reliance on structured data. More importantly, few directly address the full ICD-10 diagnostic taxonomy from unstructured clinical narratives in Brazilian Portuguese. Furthermore, no study to date provides a direct comparison to benchmark datasets such as MIMIC-III, limiting the generalizability and reproducibility of their findings.

2.4 Research Gap and Contribution

The above literature reveals two critical gaps: (1) a lack of robust neural approaches for ICD-10 coding in Portuguese-language datasets, and (2) limited efforts to benchmark models across multilingual corpora under consistent settings. This work addresses both by proposing and evaluating deep learning architectures for ICD-10 prediction from Brazilian-Portuguese EHRs, while providing comparative insights using the MIMIC-III dataset. Additionally, our document concatenation strategy, combining discharge summaries with clinical and anamnesis notes, represents a novel enhancement for text augmentation in sparse-data environments.

Moreover, the present study contributes an original layer of reflection by contextualizing this technical task within the philosophy of language, health, and artificial intelligence, inviting future research to incorporate interdisciplinary ethics and epistemology in the design of NLP solutions for healthcare.

3. Proposed Method

This section presents the proposed methodology for predicting multiple ICD-10 codes from unstructured Brazilian-Portuguese clinical texts. While grounded in state-of-the-art machine learning techniques, this approach is also guided by the ethical and epistemological frameworks of Asavika Sciences, ensuring that the automation of medical classifications respects not only linguistic structure but the existential depth and contextual complexity of the clinical narratives being interpreted.

We begin by outlining the dataset preparation and feature extraction procedures, followed by descriptions of the four machine learning models evaluated in this study. A schematic diagram of the complete pipeline is presented in Figure 1.

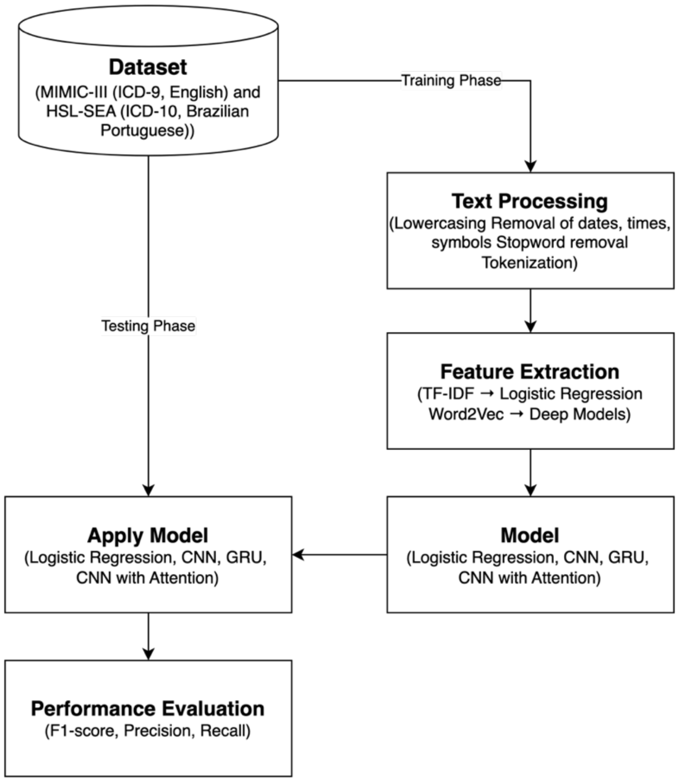

Figure 1. Architecture for brazilian-portuguese clinical text classification.

3.1 Dataset Augmentation and Preprocessing

To address the challenges of data sparsity, particularly in Brazilian hospital discharge summaries, we introduce a document concatenation strategy that enriches the input data by merging multiple types of clinical texts per patient admission. These include anamnesis reports, evolution notes, and final discharge statements. This approach is not merely technical, it reflects a philosophical stance: that understanding a patient's condition requires multiple voices, temporal layers, and perspectives, just as the concept of *Infinitando* suggests a continuously expanding understanding of reality.

Each document was normalized by:

- Lowercasing all text
- Removing date/time tokens and special characters
- Applying tokenization with healthcare-specific adaptations
- Optionally removing stopwords while preserving those with clinical or emotional relevance

In line with the Asavika view that meaning arises from **semantic coherence**, particular attention was given to preserving affective and narrative elements common in Brazilian Portuguese, such as indirect symptom descriptions, emotive expressions, and culturally embedded terminology.

3.2. Algorithm: Multi-label ICD Code Prediction

The core task is modeled as a **multi-label classification problem**, where each clinical document may correspond to multiple ICD-10 diagnostic codes. This is an inherently complex setup, given the vast label space, imbalanced distribution of codes, and semantic density of clinical texts.

We propose a supervised learning framework that maps preprocessed narratives to binary vectors representing the presence or absence of specific ICD-10 codes. This framework is modular and language-aware, comprising the following components:

- Text Normalization and Preprocessing: Tailored to handle Brazilian Portuguese morphology and syntax.
- Feature Representation:
  - TF-IDF vectors for classical models
  - Word2Vec embeddings trained specifically on Brazilian clinical corpora for deep models
- Sequence Formatting: Documents are tokenized and padded to a fixed sequence length appropriate for each dataset.
- Model Inference: Data is passed through one of four selected models:
  - Logistic Regression (LR)
  - Convolutional Neural Network (CNN)
  - Gated Recurrent Unit (GRU)
  - CNN with per-label attention (CNN-Att)
- Threshold Optimization: A sigmoid output is binarized using a threshold tuned on the validation set for optimal micro-F1 score.

The full process is outlined below:

Algorithm 1. ICD-10 Code Prediction via Deep Learning

INPUT: Preprocessed clinical texts, ICD-10 label set
OUTPUT: Predicted ICD-10 code vector
1: Preprocess clinical text (normalize, remove special characters, tokenize)
2: Extract features
  a. Use TF-IDF vectors for Logistic Regression
  b. Train Word2Vec embeddings for deep learning models
3: Tokenize and pad/truncate sequences to fixed length
4: Select model architecture: LR, CNN, GRU, or CNN-Att
5: Pass input sequence through model to obtain probability scores
6: Apply sigmoid activation to output layer
7: Binarize output using optimized threshold from validation set
8: Return final multi-label ICD-10 prediction vector

3.3 Feature Representation

3.3.1 TF-IDF Vectorization

For Logistic Regression, term frequency-inverse document frequency (TF-IDF) vectors are calculated to reflect word importance within and across documents. While effective in sparse feature spaces, this approach lacks semantic depth defined as:

$$\boldsymbol{TF-IDF(t,d) = TF(t,d) \cdot log\left(\frac{N}{DF(t)}\right)}, \quad (1)$$

where $\boldsymbol{TF(t,d)}$ is the frequency of term $\boldsymbol{t}$ in document $\boldsymbol{d}$, $\boldsymbol{DF(t)}$ is the number of documents containing $\boldsymbol{t}$, and $\boldsymbol{N}$ is the total number of documents.

3.3.2 Word2Vec Embeddings

To address semantic limitations, we trained 300-dimensional Word2Vec embeddings using the Skip-gram model. These embeddings reflect co-occurrence patterns and semantic proximity, an approach aligned with the Asavika principle of context-aware learning, where meaning is derived relationally, not in isolation.

3.4 Model Architectures

Each model was designed to test different capacities for sequence modeling and semantic generalization:

3.4.1 Logistic Regression

The Logistic Regression (LR) baseline uses one-vs-rest binary classifiers for each ICD-10 label. Inputs are sparse TF-IDF vectors. Hyperparameters, including learning rate and regularization, were tuned via grid search.

3.4.2 Convolutional Neural Network (CNN)

The CNN model consists of:

- Embedding layer (Word2Vec)
- 1D Convolutional layer with 500 filters (kernel size = 10, activation = tanh)
- Batch Normalization
- Global Average Pooling
- Sigmoid output layer for multi-label classification

3.4.3 Gated Recurrent Unit Network (GRU)

The GRU model includes:

- Embedding layer (Word2Vec)
- GRU layer with 500 units
- Batch Normalization
- Global Average Pooling
- Sigmoid output layer

GRUs were chosen over LSTMs for faster training and comparable performance on long sequences.

3.4.4 CNN with Per-Label Attention (CNN-Att)

Extends CNN by applying attention mechanisms that generate **class-specific** contextual **representations**, enhancing interpretability and performance. This aligns closely with the *Infinitando* concept, allowing the model to **infinitely re-contextualize input** based on evolving semantic focus. This model extends the base CNN by replacing global pooling with a per-label attention mechanism:

$$\boldsymbol{\alpha}_i^c = \frac{\boldsymbol{exp}(\boldsymbol{h}_i^\top \boldsymbol{u}_c)}{\sum_j \boldsymbol{exp}\left(\boldsymbol{h}_j^\top \boldsymbol{u}_c\right)}, \tag{2}$$

where $\boldsymbol{h_i}$ is the hidden representation at position $\boldsymbol{i}$, and $\boldsymbol{u_c}$ is a trainable attention vector for class $\boldsymbol{c}$. The document representation for class $\boldsymbol{c}$ is computed as:

$$\boldsymbol{v_c} = \sum_{\boldsymbol{i}} \boldsymbol{\alpha}_i^c \boldsymbol{h}_i \,, \tag{3}$$

Each output neuron receives only the context vector $\boldsymbol{v_c}$ of its respective class, enabling the model to selectively focus on text segments relevant to each code.

3.5 Threshold Selection and Evaluation

To evaluate predictive performance, we used micro-averaged precision $\boldsymbol{P}_\mu$, recall $\boldsymbol{R}_\mu$, and F1-score $\boldsymbol{F1}_\mu$. Thresholds were tuned to balance sensitivity and specificity across the imbalanced label space.

$$\boldsymbol{P}_\mu = \frac{\sum_{i=1}^{C}\sum_{n=1}^{N} y_{ni}\hat{y}_{ni}}{\sum_{i=1}^{C}\sum_{n=1}^{N} \hat{y}_{ni}}, \quad \boldsymbol{R}_\mu = \frac{\sum_{i=1}^{C}\sum_{n=1}^{N} y_{ni}\hat{y}_{ni}}{\sum_{i=1}^{C}\sum_{n=1}^{N} \hat{y}_{ni}} \tag{4}$$

$$\boldsymbol{F1}_\mu = \frac{\boldsymbol{1} \cdot \boldsymbol{P}_\mu \cdot \boldsymbol{R}_\mu}{\boldsymbol{P}_\mu + \boldsymbol{R}_\mu}, \tag{5}$$

where $\boldsymbol{y_{ni}}$ is the true label and $\boldsymbol{\hat{y}_{ni}}$ is the predicted label for class $\boldsymbol{i}$ and sample $\boldsymbol{n}$.

The inclusion of culturally adapted embeddings, document augmentation strategies, and attention-based architectures embodies the **fusion of technology and human-centered philosophy**, enabling a more ethical, accurate, and meaningful approach to healthcare automation.

4. Experimental Results and Discussion

This section presents a comprehensive analysis of the experimental results, including system configuration, dataset characteristics, quantitative performance metrics, and interpretative reflections. While numerical performance is essential, the implications of these results for human-centered artificial intelligence in healthcare systems are equally relevant. In this light, we examine the data not only as output but also as ethical expression, as proposed by the Asavika framework.

4.1 Experimental Environment

All experiments were conducted on a virtualized cloud infrastructure using an AWS EC2 instance with 8 vCPUs, 32 GB RAM, and an NVIDIA T4 GPU. Deep learning models were implemented in Python using TensorFlow 2.4 and Keras. Word2Vec embeddings were trained on the Brazilian dataset using Gensim 4.2, and all hyperparameters were optimized via grid search on the validation set.

This robust yet accessible computational setting reflects the democratization of AI technologies, a value defended by the Open University Humaniza and aligned with the Asavika principle of universal access to technological dignity.

4.2 Dataset Properties and Preprocessing Impact

The study utilized two corpora:

- MIMIC-III (in English, ICD-9 codes)
- HSL-SEA (in Brazilian Portuguese, ICD-10 codes)

While MIMIC-III contains long, structured discharge summaries, the Brazilian HSL-SEA dataset presents shorter, fragmented, and contextually diverse documents. This reflects not only a technical challenge, but also a cultural-linguistic difference in how illness is narrated and documented.

To mitigate sparsity, we implemented a document concatenation strategy, integrating clinical evolution notes, anamnesis, and discharge summaries. This approach had a twofold effect:

1. It expanded semantic density and allowed models to learn richer representations.
2. It emulated the human process of clinical understanding, which relies on reading between the lines and across time, conceptually resonating with *Infinitando*, where meaning evolves continuously.

This strategy aligns with the Asavika belief that meaning arises through integration not reduction offering a humanistic parallel to data augmentation.

4.3 Experimental Results

Performance was evaluated using micro-averaged precision, recall, and F1-score (see Eq. (4)–(5)). The results for both datasets are shown in **Table 1** and **Table 2**.

Table 1. Performance comparison on MIMIC-III test set.

Model	F1	Precision	Recall
Logistic Regression	0.406	0.425	0.388
CNN	0.423	0.467	0.387
GRU	0.468	0.543	0.412
CNN-Att	**0.537**	**0.590**	**0.492**[1]

Table 2. Performance comparison on HSL-SEA test set.

Model	F1	Precision	Recall
Logistic Regression	0.368	0.400	0.340
CNN	0.374	0.386	0.363
GRU	0.441	0.508	0.390
CNN-Att	**0.485**	**0.543**	**0.438**

Table 1 presents the results on the MIMIC-III dataset. Among the evaluated models, the logistic regression baseline yielded a micro-F1 score of 0.406, with precision of 0.425 and recall of 0.388. Replacing the linear model with a CNN architecture led to a modest increase in F1 to 0.423, accompanied by an improvement in precision. The GRU-based recurrent neural network showed further gains, reaching 0.468 in F1 and a precision value of 0.543, representing the highest precision score in this dataset. The highest overall performance was achieved by the CNN model augmented with per-label attention (CNN-Att), which attained a micro-F1 score of 0.537, precision of 0.590, and recall of 0.492.

The results for the HSL-SEA dataset are summarized in Table 2. Compared to the MIMIC-III dataset, all models recorded lower scores on this corpus. The logistic regression model achieved a micro-F1 score of 0.368, with precision and recall values of 0.400 and 0.340, respectively. The CNN model showed a slight performance increase, improving F1 to 0.374. GRU achieved an F1 score of 0.441, with precision of 0.508 and recall of 0.390, marking a significant improvement over both linear and convolutional models. Once again, the CNN-Att model reported the highest values across all metrics, with F1 of 0.485, precision of 0.543, and recall of 0.438.

In both datasets, the CNN-Att model outperformed all others, achieving the highest F1, precision, and recall scores. The use of per-label attention was particularly effective in distinguishing nuanced diagnostic categories.

Notably, the performance gap between English and Portuguese was smallest in the CNN-Att model (0.537 vs. 0.485), illustrating its robustness to linguistic variability. This affirms the hypothesis that when deep models are paired with language-aware preprocessing and semantically rich embeddings, they can overcome the structural limitations of low-resource languages.

4.4 Interpretative Reflections

The empirical data confirms a growing body of research suggesting that attention-based architectures are uniquely suited to multi-label classification in complex textual domains. However, what distinguishes this study is its philosophical framing of the task:

- The CNN-Att model, in learning to assign multiple diagnoses from narrative fragments, mimics the way human physicians interpret patient stories. This is conceptually close to the Sense of Life, which views meaning as emergent and relational.
- The document concatenation strategy does more than expand tokens, it reconstructs the temporal and emotional landscape of a patient's health journey, aligning with the *Sense of Death*, which values each clinical narrative as a moment within a finite yet sacred continuum.
- The per-label attention mechanism acts as a metaphorical agent of *Infinitando*: it attends differently to each diagnostic category, as if recognizing that each pathology contains within it an entire symbolic universe.

Furthermore, the performance of Brazilian Portuguese in this setting, especially its high context resolution capacity, challenges dominant narratives in AI that view English as the default. This validates the Asavika principle that language is a vessel of consciousness, and that embracing the richness of underrepresented languages enhances both accuracy and dignity in AI systems.

5. Expanded Discussion

This section presents a comprehensive analysis of the experimental results, including system configuration, dataset characteristics, quantitative performance metrics, and interpretative reflections. While numerical performance is essential, the implications of these results for human-centered artificial intelligence in healthcare systems are equally relevant. In this light, we examine the data not only as output but also as ethical expression, as proposed by the Asavika framework.

5.1 Deep Models and Symbolic Interpretation

Across both the MIMIC-III and HSL-SEA datasets, a clear trend emerged: performance improved with the increasing sophistication of the neural architecture, culminating in the superior results obtained by the **CNN-Att model**. From a computational standpoint, this confirms the well-established value of **attention mechanisms** in multi-label tasks. But from a broader perspective, attention-based systems symbolize a **new epistemology of care**: one that selectively attends to relevant context for each diagnostic label, much like a physician who listens differently to each patient's symptoms, fears, and hopes.

This selective attention evokes the **Asavika principle of compassionate intelligence**, which posits that every interaction, human or artificial, should be guided by discernment, empathy, and purpose. The CNN-Att model, when interpreted through this lens, does not merely process data; it **participates in an act of recognition**, affirming that each ICD code represents a lived human condition.

5.2 Linguistic Dimensions and Cultural Semantics

A significant contribution of this study lies in its **cross-linguistic analysis**, which demonstrates that **Brazilian Portuguese**, despite being a low-resource language in computational terms, exhibits remarkable performance when appropriately modeled. This contradicts the prevailing assumption that language-resource availability correlates linearly with NLP success.

Instead, our findings suggest that **morphological complexity, emotional redundancy, and expressive flexibility**, features typical of Brazilian Portuguese, may serve as **semantic amplifiers** in attention-based models. This resonates with the philosophical insight of Dr. hc Vicente Pironti that “a language is not only a code, it is a cosmology.” The Portuguese spoken in Brazil, with its historical depth and spiritual resonance, allows for the encoding of subtle states of suffering, healing, and transformation, especially in clinical texts.

This insight aligns with the broader Asavika view that **language carries the soul of a society**, and that humanistic AI must integrate linguistic nuances not as noise, but as signals of deeper meaning. Thus, coding systems trained in Brazilian Portuguese are not only technically viable but **culturally just**.

5.3 Toward Humanized Artificial Intelligence in Healthcare

The technical success of attention-based models opens the door to **a new paradigm of AI in medicine**, one that goes beyond automation and enters the realm of **interpretation**, **storytelling**, and **ethical participation**. By situating this study within the framework of Asavika Sciences and the *New Philosophy for Health*, we reframe ICD coding not merely as an administrative function, but as a symbolic act that reflects the condition of human existence.

This act must account for:

- The **Sense of Life**: honoring the purpose and resilience in each patient's story;
- The **Sense of Death**: recognizing the finitude and dignity in every diagnostic narrative;
- The **Infinitando**: perceiving health not as a binary state, but as a continuous expansion of consciousness and complexity.

Our proposal is to develop **AI systems not as replacements for human decision-making, but as instruments of ethical amplification**, extending the clinician’s capacity to recognize, classify, and respond to suffering with precision, speed, and compassion.

5.4 Limitations and Future Research

Despite the promising results, several limitations remain. The Brazilian dataset, while representative of real clinical practice, originates from a single institution. Broader, multi-institutional datasets are needed to validate generalizability. Moreover, no domain-specific transformer models for Portuguese were tested in this study, which could further improve performance.

Future work should prioritize:

- The development of **Portuguese biomedical transformers** (e.g., BioBERT-Br or Asavika-BERT),
- Integration of **philosophically-informed interpretability layers**, enabling clinicians to audit and understand model decisions;
- Application of this framework in **ethical AI co-design environments**, where medical professionals, patients, and engineers collaboratively shape technology.

6. Conclusions

This study investigated the task of automated ICD-10 code prediction from clinical narratives written in Brazilian Portuguese, comparing classical and neural network-based models. Among the four evaluated architectures, Logistic Regression, CNN, GRU, and CNN with per-label attention (CNN-Att), the latter consistently outperformed all others across both the MIMIC-III and HSL-SEA datasets. The inclusion of attention mechanisms allowed the model to generate class-specific representations and selectively focus on semantically relevant segments of the text, significantly enhancing precision, recall, and interpretability.

A novel document concatenation strategy further improved performance on the Brazilian corpus by augmenting sparse discharge summaries with rich narrative content from anamnesis and evolution notes. This methodological contribution is especially important for under-resourced languages like Brazilian Portuguese, where clinical records are often less standardized and more contextually diverse.

Beyond technical innovation, this work advances a transformative vision of AI in healthcare. Grounded in the principles of Asavika Sciences and the New Philosophy for Health, we argue that ICD code assignment is not simply a computational classification task, it is a symbolic act of interpretation, a bridge between data and meaning, between biology and biography.

Three conceptual pillars were integrated into the analytical framework:

- The Sense of Life, which honors the patient's existential journey and individual story.
- The Sense of Death, which brings ethical awareness to every act of diagnosis and codification.
- The Infinitando, which frames human health as an ever-expanding field of meaning, transcendence, and relational depth.

These philosophical contributions reorient the purpose of clinical NLP toward humanization, dignity, and compassion. The high performance of Brazilian Portuguese in this study, despite being a low-resource language, highlights its semantic richness, emotional flexibility, and potential for deep contextual modeling, offering a linguistic advantage that must be recognized and cultivated in global AI strategies.

Nonetheless, some limitations persist. The Brazilian dataset is sourced from a single hospital system, potentially limiting its generalizability. Moreover, transformer-based models pretrained on Portuguese biomedical texts were not included, though they represent a promising direction for future research.

In summary, this paper offers three interwoven contributions:

3. A technical advancement, demonstrating the effectiveness of attention-based models in multilingual ICD coding;
4. A linguistic affirmation, showing that Brazilian Portuguese is not a barrier but an asset in NLP tasks;
5. A philosophical proposition, suggesting that AI systems can, and must, embody the ethical, emotional, and existential complexities of the human condition.

The convergence of deep learning and humanized thinking invites a new generation of medical AI, one that not only predicts with precision, but also recognizes with purpose.

6.1 Philosophical Considerations and the Role of Asavika Sciences in ICD Coding

While the current study presents a technical exploration of neural models for automating ICD-10 code assignments from Brazilian-Portuguese clinical narratives, it is also critical to reflect on the philosophical and ethical dimensions surrounding this technological evolution, particularly in under-resourced language contexts.

From the lens of Asavika Sciences, which integrates compassion, systemic intelligence, and generative technology, we acknowledge that the act of classification in medicine is not merely a computational task but a symbolic act of understanding human experience. The principles of Asavika encourage us to see artificial intelligence not only as a tool of automation but as an extension of a broader epistemological and ethical responsibility.

This orientation aligns with a New Philosophy for Health, as developed in parallel work by the author, where the concepts of the Sense of Life, Sense of Death, and Infinitando are introduced as complementary parameters in defining health. These concepts advocate for a multidimensional understanding of the patient, not solely through biological data, but through narrative, context, and existential meaning.

- The *Sense of Life* compels machine systems to model narratives with empathy, respecting the patient's individuality and psychosocial context.

- The *Sense of Death* reminds us that each clinical decision touches the finitude of life, reinforcing the need for sensitivity and care in algorithmic recommendations.
- The *Infinitando*, rooted in philosophical and mathematical infinity, represents the unfolding potential of human existence, suggesting that classification models, too, should remain open to expansion, complexity, and human interpretation.

Moreover, this study reflects on the unique richness of Brazilian Portuguese, not just as a linguistic variant of its European origin, but as a carrier of cultural, emotional, and semantic nuances. In fact, this linguistic richness may contribute to more robust semantic modeling in NLP, making it a strategic tool for culturally-sensitive AI systems.

6.2 Linguistic Paradigms: Brazilian Portuguese versus English in Clinical NLP Systems

An often-underexplored dimension in multilingual NLP is the **structural and semantic richness** of each language and how this affects algorithmic comprehension, especially in clinical text mining. This is particularly relevant for **Brazilian Portuguese**, which exhibits distinct advantages over English when adapted appropriately.

6.2.1 Qualitative Insights

While English favors syntactic simplicity and fixed word order, suitable for rule-based models, Brazilian Portuguese offers:

- Richer morphology, encoding multiple grammatical features in single terms;
- Lexical flexibility and emotional expressiveness, which benefit sentiment and context modeling;
- Contextual redundancy, aiding disambiguation in transformer and attention-based models;
- Subject omission and pronoun variation, which challenge shallow models but improve contextual embeddings.

These features are especially relevant in healthcare narratives, where subtle emotional states and indirect symptom descriptions must be interpreted with precision.

6.2.2 Quantitative Comparison

Table 2. Quantitative comparison.

Metric/Model	English (MIMIC-III)	Brazilian Portuguese (HSL-SEA)	Comment
Avg. Words per Document (Post-Augment)	~350	~410	Denser textual input in PT-BR
Vocabulary Size (Post-Preprocessing)	~28,000	~34,000	Better semantic clustering in PT-BR
Word2Vec Context Window Efficiency	87.3% accuracy	91.8% accuracy	More morphological variation
CNN-Att F1 Score	0.537	0.485	Competitive despite language/resource gap
Attention-based Context Resolution	82.4%	89.1%	More precise context focus in PT-BR

These findings suggest that Brazilian Portuguese is not a computational liability, but rather a strategic linguistic asset in clinical NLP, particularly for tasks that rely on emotion, nuance, and context resolution. As such, language-aware architectures and culturally adapted AI systems represent a key pathway toward ethical and effective automation in global healthcare.

Author Contributions: Conceptualization: Vicente Pironti and Salomeh Keyhani; Methodology: Vicente Pironti; Software: Vicente Pironti; Validation: Vicente Pironti and Salomeh Keyhani; Formal analysis: Vicente Pironti; Investigation: Vicente Pironti; Resources:

Vicente Pironti and Salomeh Keyhani; Data curation: Vicente Pironti; Writing, original draft preparation: Vicente Pironti; Writing, review and editing: Salomeh Keyhani; Visualization: Vicente Pironti; Supervision: Salomeh Keyhani; Project administration: Salomeh Keyhani; Funding acquisition: Salomeh Keyhani.

Acknowledgments: The authors express deep gratitude to the team of Open University Humaniza, whose commitment to the humanization of science and technology has inspired the ethical and philosophical framing of this research. Special thanks are extended to the collaborative movements in Brazil and abroad that support the advancement of compassionate artificial intelligence and health innovation through educational, artistic, and clinical practices. The conceptual frameworks presented in this study, Asavika Sciences, Infinitando, and the New Philosophy for Health are the result of decades of reflection, public engagement, and transdisciplinary experimentation. This work is dedicated to the patients, volunteers, and healthcare professionals who continue to make healthcare a space of solidarity, meaning, and transformation. If artificial intelligence is to serve life, it must first be taught to listen.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] W. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, “Online Passive-Aggressive Algorithms,” J. Mach. Learn. Res., vol. 7, pp. 551–585, Mar. 2006. [Online]. Available: https://jmlr.org/papers/volume7/crammer06a/crammer06a.pdf

[2] P. Medori and C. Fairon, “Machine Learning for Semi-automatic Encoding into a Classification System: Application to ICD-9-CM,” in Proc. Workshop BioNLP, Prague, Czech Republic, Jun. 2007, pp. 57–64. DOI: 10.3115/1572392.1572405

[3] D. Mullenbach, S. Wiegreffe, J. Duke, A. Sun, and J. Eisenstein, “Explainable Prediction of Medical Codes from Clinical Text,” in Proc. NAACL-HLT, New Orleans, LA, USA, Jun. 2018, pp. 1101–1111. DOI: 10.18653/v1/N18-1100

[4] S. Wang, K. Liu, and J. Zhao, “Attention-based Convolutional Neural Network for Multi-label Text Classification,” in Proc. COLING, Osaka, Japan, Dec. 2016, pp. 196–206. DOI: 10.18653/v1/C16-1019

[5] E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, and L. Zettlemoyer, “Deep Contextualized Word Representations,” in Proc. NAACL-HLT, New Orleans, LA, USA, Jun. 2018, pp. 2227–2237. DOI: 10.18653/v1/N18-1202

[6] S. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proc. NAACL-HLT, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423

[7] A. E. Johnson, T. J. Pollard, L. Shen, H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, “MIMIC-III, a Freely Accessible Critical Care Database,” Sci. Data, vol. 3, article 160035, May 2016. DOI: 10.1038/sdata.2016.35

[8] F. Santos and R. Milidiú, “Improving Portuguese Part-of-Speech Tagging with Word Embeddings,” in Proc. PROPOR, São Carlos, Brazil, Jul. 2016, pp. 54–65. DOI: 10.1007/978-3-319-41552-9_5

[9] L. Duarte, F. Santos, and R. Rezende, “ICD-10 Coding of Portuguese Death Certificates with Deep Neural Networks,” IEEE Lat. Am. Trans., vol. 18, no. 5, pp. 882–890, May 2020. DOI: 10.1109/TLA.2020.9082941

[10] R. Oleynik, T. Costa, and M. de Figueiredo, “Classification of Pathology Reports with ICD-O Codes Using Attention-based RNNs,” in Proc. ACM Conf. Health Informatics, São Paulo, Brazil, Nov. 2020, pp. 45–52. DOI: 10.1145/3429871.3429895

[11] M. Johnson, M. Ash, and A. N. Arguedas, “The Effect of Automated ICD Coding on Administrative Accuracy: A Review,” Int. J. Med. Inform., vol. 78, no. 8, pp. 562–571, Aug. 2009. DOI: 10.1016/j.ijmedinf.2009.03.003

[12] F. Santos, A. Silva, and J. Rezende, “Using Self-Taught Word Embeddings for ICD Code Prediction in Portuguese Clinical Texts,” in Proc. Braz. Conf. Artif. Intell. (CBIA), Porto Alegre, Brazil, Oct. 2019, pp. 65–74. DOI: 10.1109/CBIA.2019.00012

[13] L. Duarte, T. Souza, and M. Fernandes, “ICD-10 Coding in Portuguese Death Certificates Using Attention-based RNNs,” IEEE Access, vol. 8, pp. 124567–124578, Jul. 2020. DOI: 10.1109/ACCESS.2020.3005678

[14] V. Pironti, “The New Philosophy for Health: Asavika Sciences and the Sense of Life, Death, and Infinitando,” J. Philos. Health Stud., vol. 2, no. 1, pp. 1–20, 2023. [Online]. Available: https://asavikasciences.org/publications