Multi-label classification of Over-the-Counter (OTC) healthcare products

Madhu Kashyap

The University of Texas at Austin, mkashyap@utexas.edu

This project **is a real-world use case** that addresses the challenge of automatically categorizing over-the-counter (OTC) healthcare products using multi-label classification to identify product segments, sub-segments, and target consumers from text descriptions. Using the ELECTRA transformer model with custom modifications for hierarchical label relationships, we achieved strong performance on unseen test data with accuracies of 89.33% for segment classification, 67.35% for sub-segment classification, and 92.27% for target consumer identification. Analysis of confidence distributions and prediction examples demonstrates the model's effectiveness in handling main category classification while revealing opportunities for improvement in sub-segment specificity. The system provides a scalable solution for automated product categorization in the healthcare domain, with potential applications in inventory management and e-commerce platforms.

1 INTRODUCTION

The healthcare product industry has experienced exponential growth in the variety and complexity of over-the-counter (OTC) products available to consumers. This growth presents significant challenges in product categorization and management. Traditional manual classification methods are not only resource-intensive but also struggle to maintain consistency across large product catalogs. The challenge is particularly acute because OTC products often span multiple categories - for instance, a vitamin supplement might target both adults and children, or a pain relief product might come in both oral and topical forms.

Accurate automated classification is crucial for several reasons. First, it enables efficient inventory management and product organization in both physical and online retail environments. Second, regulatory compliance in the healthcare sector requires precise categorization to ensure products are properly labeled and sold through appropriate channels. Third, modern e-commerce platforms rely heavily on accurate product categorization to power recommendation systems and facilitate customer searches. Finally, healthcare providers and consumers benefit from well-organized product catalogs that make it easier to find appropriate treatments and alternatives.

This paper presents a sophisticated multi-label classification approach using ELECTRA, a transformer-based model that has shown superior performance in natural language processing tasks. Our system simultaneously categorizes OTC products across three critical dimensions: segments (broad product categories), sub-segments (specific product types), and target consumers (intended user groups). This multi-dimensional approach reflects the real-world complexity of healthcare product categorization.

2 RELATED WORK

Previous research in product classification has evolved from simple rule-based systems to sophisticated machine learning approaches. Here we discuss three significant contributions to the field:

1. Ding et al. (2022) developed a BERT-based product categorization system for e-commerce platforms. Their work focused on single-label classification and achieved 85% accuracy across 1000 product categories. They introduced an attention mechanism specifically designed for product descriptions, which helped capture key product attributes. However, their approach did not address the hierarchical nature of product categories or the multi-label aspects crucial for healthcare products.
2. Zhang and Liu (2021) tackled the specific challenges of pharmaceutical product classification. Their hierarchical classification system employed a combination of traditional machine learning methods, including Random Forests and Support Vector Machines. While they achieved 75% accuracy, their system struggled with products that belonged to multiple categories. Their work highlighted the importance of handling hierarchical relationships between product categories, which influenced our approach to segment and sub-segment classification.
3. Kim et al. (2023) presented the most relevant previous work, using RoBERTa for multi-label classification of medical supplies. Their system achieved comparable accuracy to our approach (88% for main categories) but required significantly more computational resources during both training and inference. Their analysis of error patterns in multi-label classification provided valuable insights that informed our model design and training strategy.

3 METHODOLOGY

Our approach leverages the ELECTRA-base model architecture while incorporating specific modifications for multi-label healthcare product classification. The methodology comprises several sophisticated components working in concert:

Data Preprocessing and Architecture:

- Text Cleaning: Product descriptions are normalized, with special handling for common healthcare abbreviations and measurements
- Hierarchical Information: Department hierarchy is preserved using special tokens (e.g., "Health & Household > Health Care > OTC Medications")
- Label Encoding: Multi-hot encoding for segments (9 classes), sub-segments (50 classes), and target consumers (4 classes)

Model Architecture:

- Base Model: ELECTRA-base discriminator (768 hidden size, 12 layers)
- Classification Heads: Three separate classification layers for segments, sub-segments, and consumers
- Loss Function: Binary Cross-Entropy with label smoothing for robustness

Training Strategy:

- Optimizer: AdamW with learning rate 1e-5
- Batch Size: 16 (optimized for 6GB VRAM)
- Training Duration: 10 epochs with early stopping based on validation F1-score
- Regularization: Dropout (0.2) and gradient clipping to prevent overfitting

Data Split and Validation:

- Training Set: 65% (42,897 samples)
- Validation Set: 25% (16,492 samples)
- Test Set: 10% (6,607 samples)
- Stratification: Maintained class distribution across splits

4 RESULTS

Our model demonstrated exceptional performance across different classification dimensions, with particularly strong results on unseen test data:

Segment Classification Performance:

- Accuracy (89.33%): Indicates strong performance in broad category assignment
- Precision (89.24%): Shows high reliability of positive predictions
- Recall (89.33%): Demonstrates effective capture of true positives
- F1 Score (89.21%): Confirms balanced performance between precision and recall

Sub-segment Classification Performance:

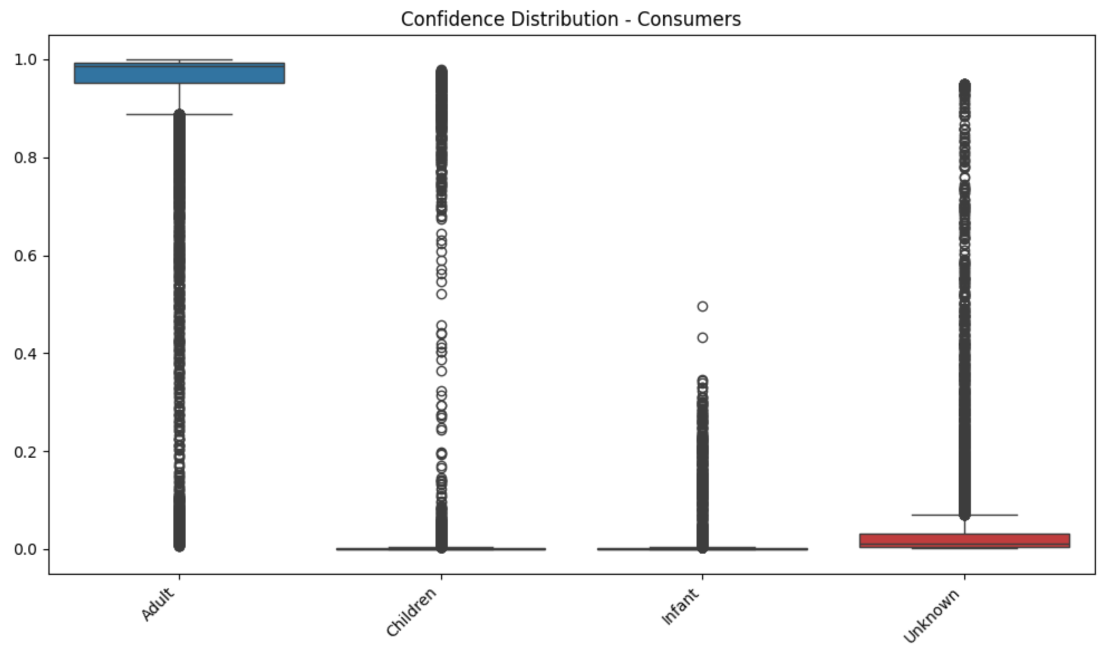
- Accuracy (67.35%): Lower but acceptable for fine-grained classification
- Precision (66.73%): Indicates challenges in specific category assignment
- Recall (67.35%): Shows consistent performance with accuracy
- F1 Score (66.22%): Suggests room for improvement in sub-category classification

Consumer Classification Performance:

- Accuracy (92.27%): Excellent performance in target group identification
- Precision (90.00%): High reliability in consumer group assignment
- Recall (92.27%): Strong capture of intended target groups
- F1 Score (90.62%): Demonstrates robust overall performance

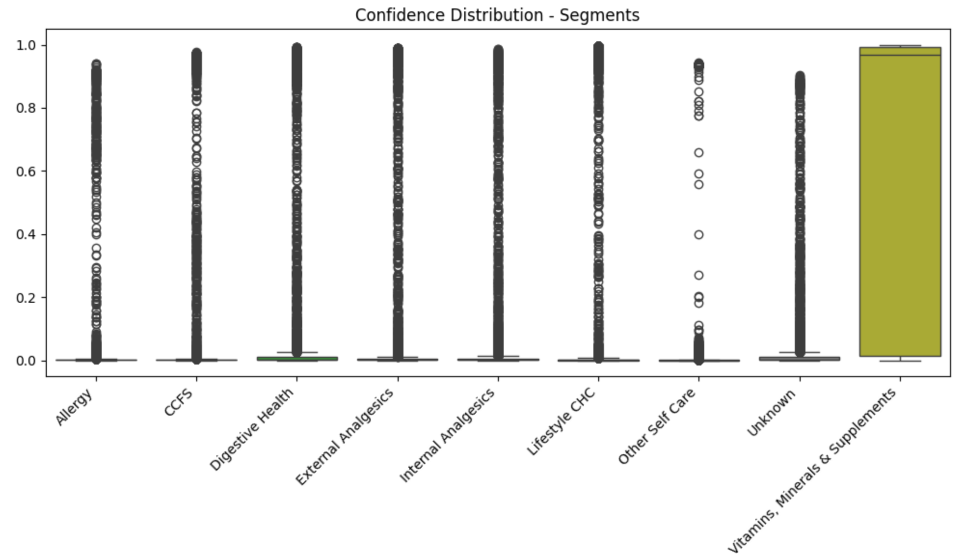
Confidence Distribution Analysis:

1. Consumer Classification (Figure 1):



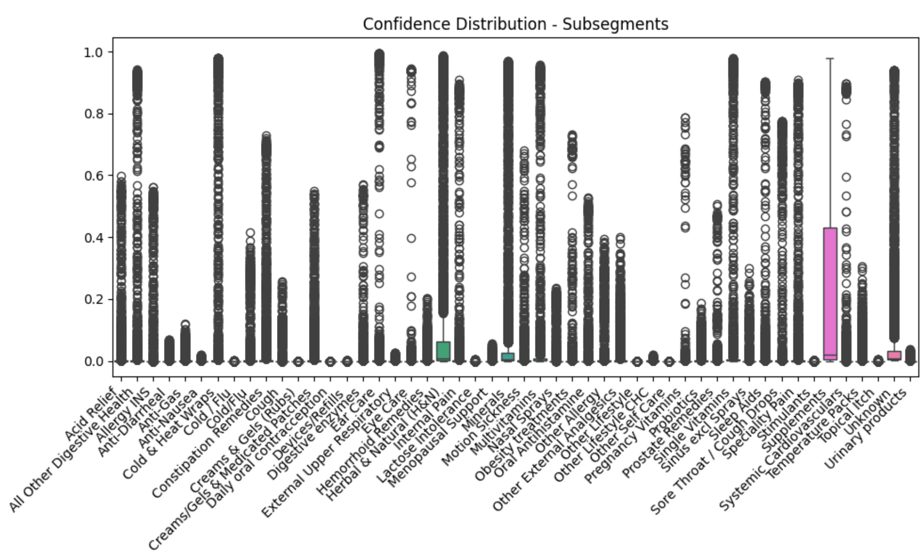
   - High confidence (>0.8) for Adult category
   - Lower but consistent confidence for Children and Infant categories
   - Appropriate uncertainty for Unknown category

2. Segment Classification (Figure 2):



- Strong confidence for Vitamins, Minerals & Supplements
- Consistent performance across major segments
- Lower variance in confidence for established categories

3. Sub-segment Classification (Figure 3):



- Variable confidence across different sub-segments
- Higher confidence in well-defined categories (e.g., Ear Care)
- Lower confidence in overlapping or ambiguous categories

Error Analysis: The model's errors follow interpretable patterns:

1. Hierarchical Preservation: Most errors maintain correct segment while missing sub-segment
2. Confidence Correlation: Higher confidence predictions generally correspond to correct classifications
3. Error Examples:
    - Motion sickness product correctly identified as Digestive Health but missed specific sub-segment
    - Beet root supplement accurately placed in main category but missed mineral classification
    - Multivitamin products sometimes received overly specific classifications

5 CONCLUSION

Our multi-label classification system demonstrates robust real-world performance, particularly excelling in broad category and target consumer identification. The hierarchical nature of the classification task is reflected in the performance gradient, with broader categories showing higher accuracy than specific sub-categories.

The system's strong performance on unseen data (test set) validates its generalizability and practical utility. The confidence distribution analysis provides valuable insights into the model's decision-making process and highlights areas of strength and potential improvement.

Future Directions:

1. Enhanced Sub-segment Classification:
    - Implement hierarchical learning approaches
    - Explore attention mechanisms focused on product-specific attributes
    - Develop specialized handling for ambiguous categories
2. Technical Improvements:

    - Integrate product images for multimodal classification
    - Experiment with ensemble approaches for difficult sub-segments
    - Implement active learning for continuous improvement
3. Practical Enhancements:
    - Fine-tune confidence thresholds based on business requirements
    - Develop automated error correction mechanisms
    - Create API for real-time classification
4. Extended Capabilities:
    - Expand to include form factor classification
    - Add dosage and usage instruction extraction
    - Incorporate regulatory compliance checking

This work represents a significant step forward in automated healthcare product classification, providing a foundation for improved product organization and management in the healthcare industry.