IJIRIS: International Journal of Innovative Research in Information Security
Volume 11, Issue 02, April 2025
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Automated Medical Device Classification

Roopalakshmi S, N S Paranjay Reddy, Deelaksha K
Mythra K, Mrudula R
Computer Science and Engineering
Vemana Institute of technology, Bengaluru, India
roopalakshmi.s@vemanait.edu.in, paranjaypandu133@gmail.com
deelakshadeekshith@gmail.com, Mythrak21@gmail.com
Mrudula2155@gmail.com

Publication History:
Manuscript Reference No: IJIRIS/RS/Vol.11/Issue02/APIS10086
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.11/Issue02/APIS10086
Received: 02,April 2025 Revised: 14, April 2025 Accepted: 25, April 2025 Published Online: 05, May 2025, Volume 2025 Article ID APIS10086 https://www.ijiris.com/volumes/Vol11/iss-02/07.APIS10086.pdf
Article Citation: Roopalakshmi,Paranjay,Deelaksha,Mythra,Mrudula (2025). Automated Medical Device Classification. International Journal of Innovative Research in Information Security, Volume 11, Issue 02, Pages 106-109
doi:> https://doi.org/10.26562/ijiris.2025.v1102.07
BibTex key: Roopalakshmi@2025Automated

Abstract: Manual categorization of medical devices according to regulatory guidelines such as CDSCO (Central Drugs Standard Control Organization) is generally time-consuming, variable, and susceptible to human error. This article describes an automated system of classification based on Microsoft's MiniLM-L12-H384-uncased model, fine-tuned using a dataset of device descriptions, classes, and respective departments. The model is enhanced further by using TF-IDF vectorization and cosine similarity for the determination of the nearest matching device. Experimental evidence shows better classification accuracy and robustness, enabling this technique for large-scale deployment in a regulatory context.
Keywords: LLM, medical device classification, MiniLM, CDSCO, TF-IDF, cosine similarity, NLP, model training.

I. INTRODUCTION

Over the past few years, the health sector has seen an upsurge in innovation in medical devices, and this has created an increased demand for organized and precise regulatory supervision. The regulation of these products depends mainly on their classification to determine the level of strictness with which they should be controlled in order to guarantee safety to patients and clinical performance. As conventional manual classification techniques fall behind the speed of innovation, interest is building in using new AI methods specifically Natural Language Processing (NLP) to automate and streamline this process.

A. Background and Significance of Medical Device Classification

The classification of medical devices is a central element of healthcare regulation and patient safety because it determines the level of regulatory control necessary to guarantee device safety and effectiveness. Regulatory bodies like the FDA and CDSCO categorize devices into risk-based classes – Class I, II, or III in the United States, and Class A, B, C, or D in india-based on the extent of possible harm to patients and users. Class A or I is the one with the least risk, while Class D or III is the riskiest [2].

B. Limitations of Conventional Classification Methods

1. Inefficiency and inconsistency: Manual classification processes, which have long depended on expert judgement, are becoming less effective with the growth in diversity and complexity of medical devices. With new technologies appearing at a rapid pace, time and expertise needed to examine and classify each device strain regulatory process [3].

2. Need for Automation: This increasing problem has shown the way to the use of automated solutions fueled by Natural Language Processing (NLP) and Machine Learning (ML) advances in different applications, especially those involving unstructured regulatory and medical documents.

1. Regulatory Hierarchy of Medical Device Classification

C. Project Motivation and Objectives

This project investigates the application of transformer-based language models, specifically the MiniLM architecture, to automate medical device classification according to regulatory standards such as those released by CDSCO [1]. The objective is to maximize accuracy and consistency in classification and reduce dependence on manual expert labor.

1. Interpretability and trust: Aside from making class predictions, the system also provides interpretable outcomes by way of TF-IDF similarity-based comparisons to enable regulatory experts to appreciate the justification of classifications and aid compliance and device risk stratification [3].

II. RELATED WORK

A. NLP Applications in Regulatory Healthcare

In the past few years, Natural Language Processing (NLP) has been in growing use in the health care industry for applications including clinical summarization, diagnostics, and extracting medical data. Large Language Models (LLMs) such as BERT,GPT, and smaller architectures such as MiniLM are made it possible to automate manually intensive processes [1]. In regulation terms, there's increasing interest in using these models to tokenize and analyze health-related documents such as Electronic Health Records (EHRs), approval guidelines, and drug/device descriptions to facilitate classification acceleration, compliance analysis, and risk stratification [3].

B. Transformer Models and Performance Improvements

Recent work points to the utility of domain-specific fine-tuning for LLMs. Gupta et al. [3] showed that transformer-based models that have been trained on structured and unstructured medical text outperform conventional machine learning models in classification, thanks to their context-sensitive architecture.

2. Regulatory Hierarchy of Medical Device Classification

Likewise, in regulatory environments, Das et al. [4] investigated the application of models such as ChatGPT towards automating FDA document processing and policy deciphering to enable quicker compliance verification. These developments have created room for automation in areas such as device classification, which previously relied on expert opinion and strict rule-based processes.

C. MiniLM and Lightweight LLM Alternatives

While extensive models such as GPT-3 and Falcon-40B [5] provide better understanding, their extent and latency could prove to be a limiting factor for use in real-world systems. MiniLM, designed by Microsoft, is a compromise between performance and efficiency as it extracts knowledge from more extensive models in a lean structure [1]. It has proven to be an effective tool for real-time processes like question answering and semantic search. This work utilizes MiniLM to construct an efficient, accurate low-latency system for CDSCO-compliant medical device classification with a balance of efficiency and interpretability.

III. METHODOLOGY

The proposed framework presents a dual-layered approach to medical device classification under CDSCO regulation: a transformer-based classification pipeline coupled with an interpretable retrieval-based layer based on cosine similarity over TF-IDF vectors. Together, they provide both the predictive accuracy and the transparency required by regulators, allowing stakeholders to check the justification for each automated classification [2].

A. Dataset preparation

1. **Description and Labelling:** To build a sound basis for model training, a carefully curated dataset of more than 1000 device descriptions was collected. Every record was annotated manually with:

- A corresponding risk classification (Class A–D) according to CDSCO standards.
- A specialty-specific department (e.g., Cardiology, Orthopedics, Radiology), indicative of domain responsibility

These annotations were strictly aligned with the Medical Device Rules, 2017, to maintain consistency with the official regulatory taxonomy [2].

2. Preprocessing

Textual data passed through a large preprocessing pipeline. First, it was normalized punctuation, special characters, and unnecessary tokens were stripped away. Descriptions were tokenized with HuggingFace's MiniLM-specific AutoTokenizer. Training phase 2 incorporated department tags as contextual features, enriching the semantic space of embeddings and improving inter-class separability.

B. Model Architecture and Pipeline

The backbone of the system is a light but semantically strong transformer model MiniLM-L12-H384 uncased that hits a good balance between representational depth and computational efficiency.

1. Transformer-Drives Risk Classification

Every input description is MiniLM vectorized into dense contextual embeddings and then fed into a multi-layer feed forward classification head, which projects them onto a four-class output layer for CDSCO's risk levels. The model is trained end-to-end with categorical cross-entropy loss.

2. TF-IDF-Enhanced Semantic Retrieval

To support explainability, the system completes transformer predictions with a vector similarity engine:

- All training instances are indexed by TF-IDF (Term Frequency–Inverse Document Frequency)
- At inference time, the input vector is matched against this index with cosine similarity
- The k-most similar descriptions are retrieved, and the top-1 match, along with its label, is rendered for user reference

[4] This two-way feedback loop not only assists in justification but also establishes trust for regulatory stakeholders who require open model behavior.

C. Training Phases and Experimental Setup

1. Phase 1 – Baseline Architecture

During the first training phase, raw device descriptions were fed as input without contextual augmentation. This configuration was a baseline, used to assess how well the transformer would generalize from sparse, unstructured data.

Input: Raw, unprocessed text.

Goal: Set up baseline performance.
Limitation: Moderate accuracy and low interpretability due to absence of domain cues.

2. Phase 2 – Context-Aware Pipeline

During Phase 2, the architecture was augmented with departmental tags and TF-IDF explanations. The classifier gained considerable boosts in precision, recall, and explain ability.

I. COMPARATIVE ANALYSIS OF TRAINING PHASES WITH QUANTITATIVE METRICS

Input: Cleaned department-tagged descriptions.
Improvements: TF-IDF integration to provide justification and semantic agreement.
Result: Enhanced classifier confidence and stakeholder trust [3].

Phase	Input Features	Accuracy (%)	Precision(%)	Recall(%)	F1-Score(%)	Interpretability (Top-1 Match Confidence)
Phase 1	Raw Descriptions	88.2	87.0	85.5	86.2	Not Available
Phase 2	Cleaned + Department Context Tags	94.7	93.9	94.2	94.0.	Top-1: 92.6% Similarity Score

IV. RESULTS AND EVALUATION

The performance of the model was evaluated on a held-out test set that made up 10% of the dataset. Two experimental phases were performed to test the effect of contextual preprocessing and interpretability improvements on classification accuracy.

A. Quantitative Metrics

These findings in table 1 suggest that Phase 2—having utilized cleaned text and department-sensitive tagging—resulted in dramatic improvements in all the evaluation metrics.

B. ROC Curve Analysis

The ROC curve in Phase 2 showed a very close-to-perfect AUC of 0.95, which verifies the strong discriminatory ability of the model between the four risk classes. This supports the classifier's value for high-stakes regulatory applications [3], [4].

3. ROC – Curve updated after fine-tuning

C. Interpretability Evaluation

Along with better metrics, the TF-IDF + Cosine Similarity mechanism obtained an average top-1 similarity confidence score of 92.6%, supporting traceable predictions. This interpretability aspect enables regulatory review—a requirement highlighted in CDSCO's transparency requirements [2].

V. SYSTEM TESTING

The system was tested stringently not only for accuracy, but also for robustness, latency, class confusion, and interpretability. The objective was to confirm the viability of implementing this solution in real-time regulatory settings.

A. Robustness Across Random Splits

For guaranteeing model generalizability, the dataset was shuffled randomly and divided into test and training partitions in five repeated runs. Standard deviation of F1-score was ±0.8%, representing very little variance and implying that the model is not overfitted to some patterns [1], [3].

B. Error Analysis Using Confusion Matrix

A 4-class confusion matrix was created in order to compute misclassifications. The bulk of the mistakes fell between Class B and Class C, not unexpected due to functional and text-based similarity across medium- and moderate-risk gadgets. In order to gain insight into inter-class misclassification patterns, a 4-class confusion matrix for predictions in Phase 2 was produced. The model is very accurate for all classes, and slight misclassification is seen between Class B and C due to semantic overlap within mid-risk categories. Phase 2 Predictions – Confusion Matrix

C. Latency and Scalability Measurement

- Average Inference Time (per request): 1.2 seconds (inclusive of similarity matching).
- Peak Load Test: 100 reqs/minute without GPU throttling or delay at batch-level.
- Memory Footprint: ~512MB VRAM for inference with MiniLM.

These figures indicate the architecture is light enough to be deployed in hospitals or regulatory portals with minimal cloud reliance [5], [9].

D. Regulatory Readiness Checklist

The solution was also tested against global and domestic AI governance frameworks like:

- CDSCO guidelines [2].
- WHO model framework for device regulation [8].
- FDA device classification protocol (mapped for future use) [10].

These verifications suggest that the system is appropriate for real-world Implementation in compliance processes, particularly in India and possibly in low-to-middle income countries (LMICs) with comparable classification issues.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

VI. CONCLUSION

This paper introduces a hybrid NLP-based architecture that utilizes both transformer-based modeling (MiniLM) and TF-IDF based semantic similarity to automate medical device classification according to CDSCO regulations. The objective was not only to enhance classification accuracy but also to include interpretability an essential requirement for real-world deployment in regulatory scenarios. The last system realized a classification rate of 94.7%, well above the baseline architecture. By leveraging the use of domain-specific tags and semantic similarity scoring, the solution provides explainable predictions, which is an essential aspect when dealing with risk-sensitive decisions such as medical device classification.

REFERENCES

1. Z. Huang, A. Mahajan, D. J. Weller, and Y. Zhang, “A Review of Applying Large Language Models in Healthcare,” Proc. IEEE IC AIM, pp. 101–115, 2023.
2. Central Drugs Standard Control Organization (CDSCO), Medical Device Rules, 2017, Government of India, Ministry of Health and Family Welfare.
3. M. Gupta, S. Dey, and N. Khanna, “Large Language Models in Healthcare: Bridging Gaps in Compliance, Documentation and Clinical Decision Support,” IEEE Transactions on Biomedical Engineering, vol. 70, no. 6, pp. 1352–1364, Jun. 2023.
4. A. Das, R. Bhattacharya, and V. Kulkarni, “ChatGPT and Beyond: Applying LLMs for Regulatory Intelligence in Healthcare,” Proc. IEEE GCAIM, pp. 88–96, 2023.
5. Technology Innovation Institute, “Falcon-40B: Open-Source Large Language Model for Multilingual Applications,” Abu Dhabi, UAE, Whitepaper, 2023.
6. B. Rieder, "Engines of Order: A Mechanology of Algorithmic Techniques," Amsterdam Univ. Press, 2020.
7. I. Boglaev, "Solving Nonlinear Integro-Differential Equations," J. Comput. Math., 2016.
8. WHO Global Model Regulatory Framework for Medical Devices Including IVDs, World Health Organization, 2020.
9. S. Liu et al., "Deployment of Low-latency AI Models in Healthcare Systems," ACM HealthTech, 2022.
10. FDA, "510(k) Device Classification Database", Center for Devices and Radiological Health, 2021.