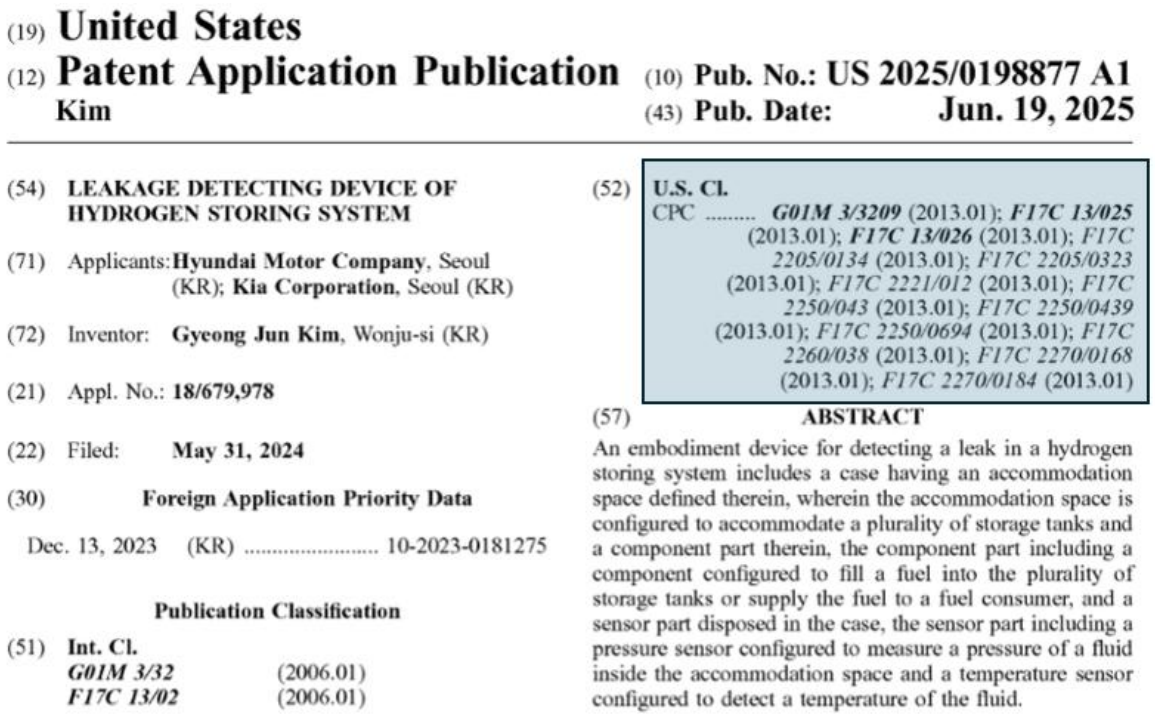
Pairwise Contrastive Fine-Tuning for Patent Classification

By Mridul Jain and Lynne Wang

July 2025 - NLP Course (MIDS 266), UC Berkeley

https://github.com/lwang9/mids-w266-final-project

CPC Hierarchy: Structure, Size, and Complexity



(19) **United States**

(12) **Patent Application Publication**
Kim

(10) **Pub. No.: US 2025/0198877 A1**

(43) **Pub. Date: Jun. 19, 2025**

(54) **LEAKAGE DETECTING DEVICE OF HYDROGEN STORING SYSTEM**

(71) Applicants: **Hyundai Motor Company**, Seoul (KR); **Kia Corporation**, Seoul (KR)

(72) Inventor: **Gyeong Jun Kim**, Wonju-si (KR)

(21) Appl. No.: **18/679,978**

(22) Filed: **May 31, 2024**

(30) **Foreign Application Priority Data**

Dec. 13, 2023 (KR) 10-2023-0181275

Publication Classification

(51) **Int. Cl.**
G01M 3/32 (2006.01)
F17C 13/02 (2006.01)

(52) **U.S. Cl.**
CPC ***G01M 3/3209*** (2013.01); ***F17C 13/025*** (2013.01); ***F17C 13/026*** (2013.01); *F17C 2205/0134* (2013.01); *F17C 2205/0323* (2013.01); *F17C 2221/012* (2013.01); *F17C 2250/043* (2013.01); *F17C 2250/0439* (2013.01); *F17C 2250/0694* (2013.01); *F17C 2260/038* (2013.01); *F17C 2270/0168* (2013.01); *F17C 2270/0184* (2013.01)

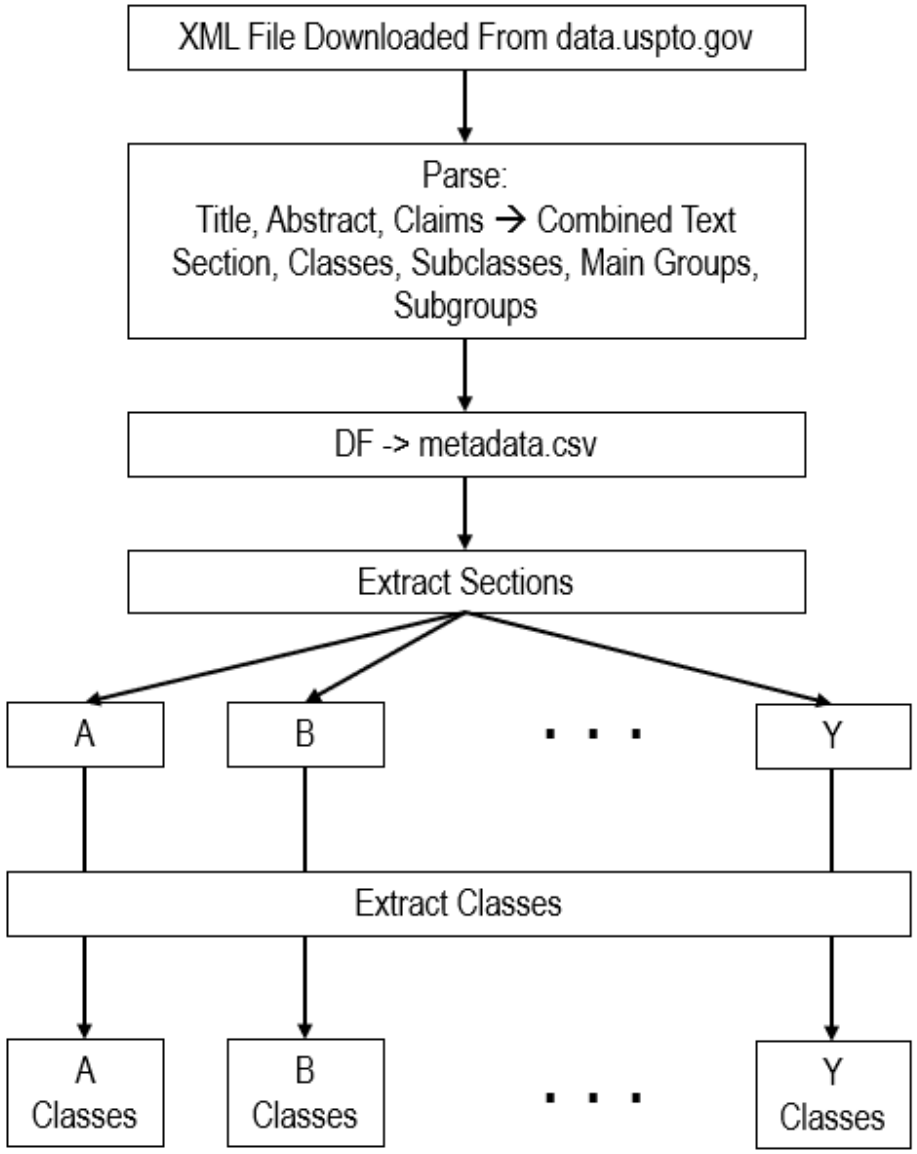
(57) **ABSTRACT**

An embodiment device for detecting a leak in a hydrogen storing system includes a case having an accommodation space defined therein, wherein the accommodation space is configured to accommodate a plurality of storage tanks and a component part therein, the component part including a component configured to fill a fuel into the plurality of storage tanks or supply the fuel to a fuel consumer, and a sensor part disposed in the case, the sensor part including a pressure sensor configured to measure a pressure of a fluid inside the accommodation space and a temperature sensor configured to detect a temperature of the fluid.

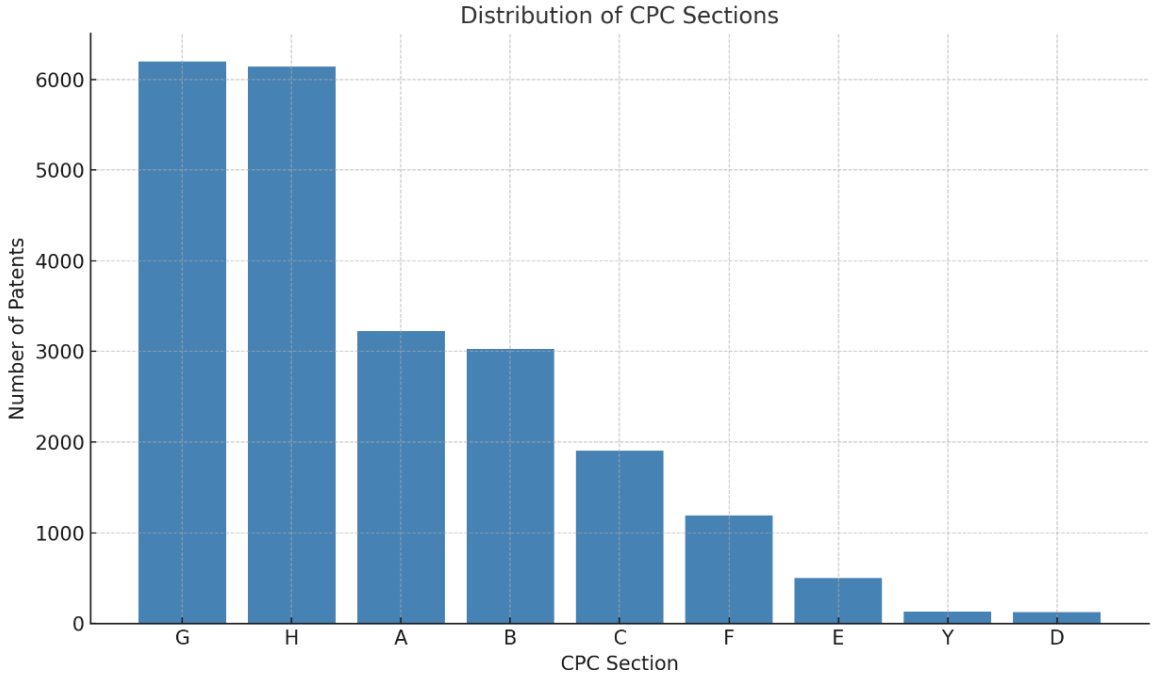
G - Physics
└── G01 - Measuring; Testing
  └── G01M - Testing static or dynamic balance of machines or structures; Testing of structures or apparatus, not otherwise provided for
    └── G01M 3/3209 - Leak testing using fluid detection, etc.

F - Mechanical Engineering; Lighting; Heating; Weapons; Blasting
└── F17 - Storing or distributing gases or liquids
  └── F17C - Vessels for storing or distributing compressed, liquefied or solidified gases
    ├── F17C 13/025 - Arrangements for detecting or preventing leakage
    ├── F17C 13/026 - Arrangements for preventing corrosion
    ├── F17C 2205/0134 - Type of vessel: Rigid vessel with outer jacket
    ├── F17C 2205/0323 - Material: Metal only (e.g., aluminum, steel)
    ├── F17C 2221/012 - Insulating means: Vacuum insulation
    ├── F17C 2250/043 - Leak detection using pressure or vacuum change
    ├── F17C 2250/0439 - Leak detection by means of acoustic sensing
    ├── F17C 2250/0694 - Protective devices or arrangements (e.g., relief valves)
    ├── F17C 2260/038 - Use or application: Cryogenic liquefied gases (e.g., LNG, liquid nitrogen)
    ├── F17C 2270/0168 - Features related to maintenance: Monitoring of physical parameters
    └── F17C 2270/0184 - Features related to maintenance: Data processing or control arrangements

Data Preprocessing



Highly Imbalanced Distribution



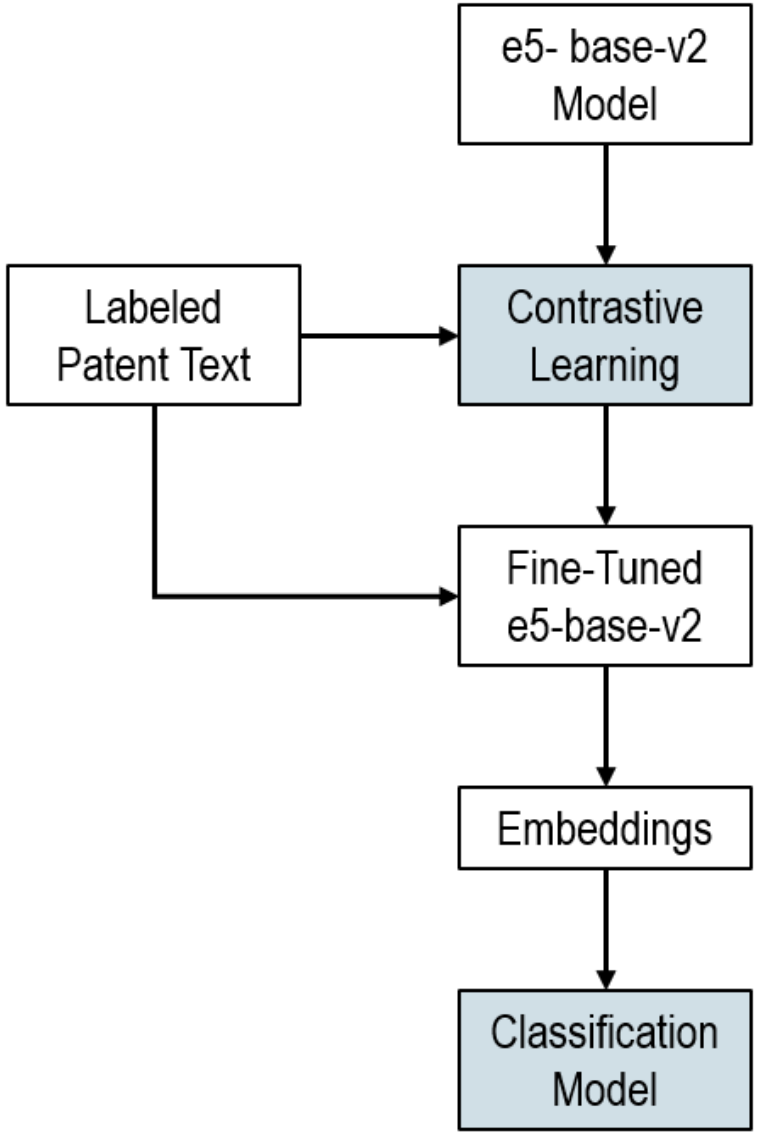
Two-Stage Process: Contrastive Fine-Tuning and Mixture of Experts Classification

Contrastive Fine-Tuning

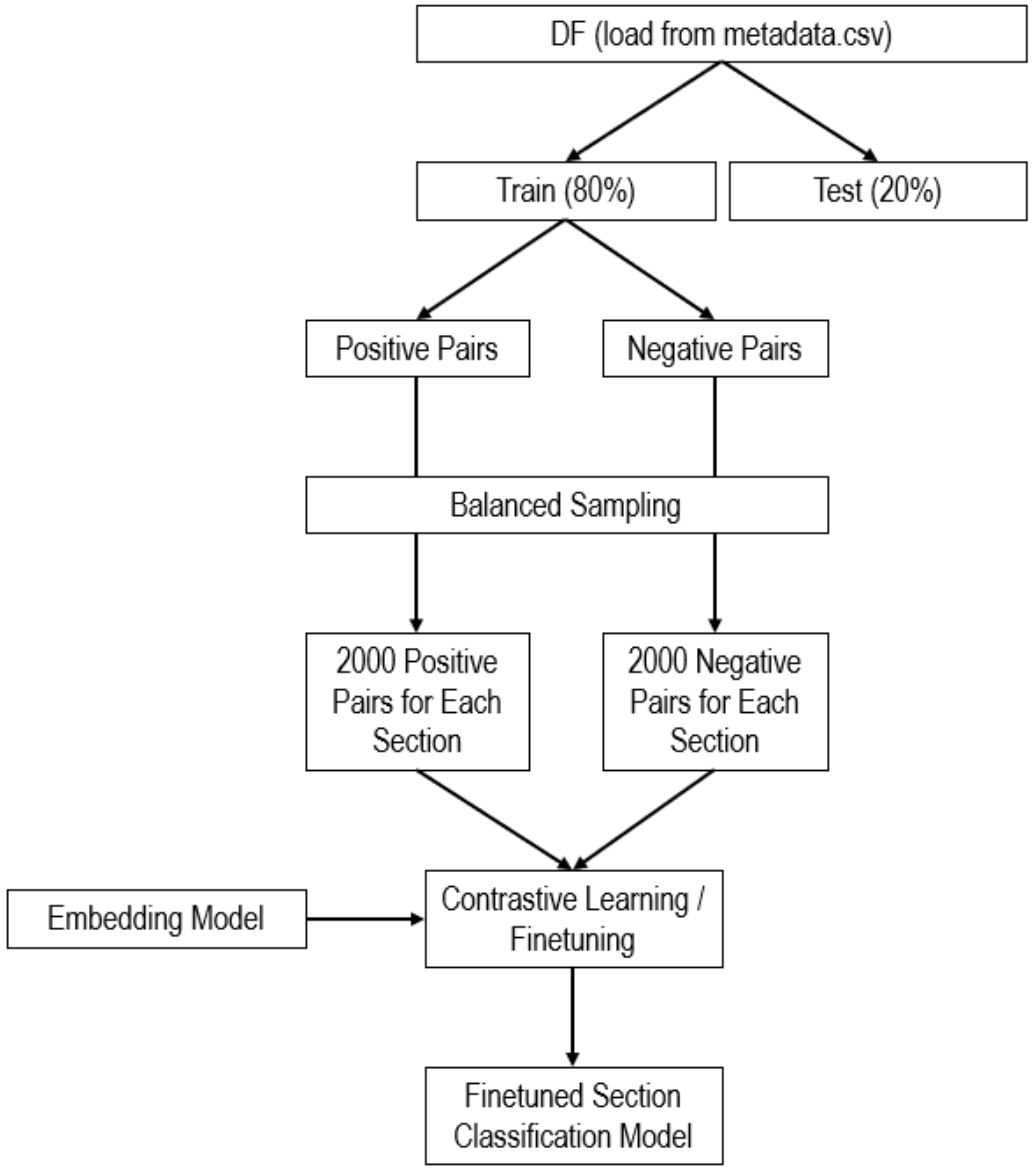
Fine-tuning E5-Base-V2 embeddings enhances semantic separation through contrastive learning techniques.

Mixture of Experts Classifier

Expert classifier integrates embeddings and taxonomy features to accurately predict CPC sections.



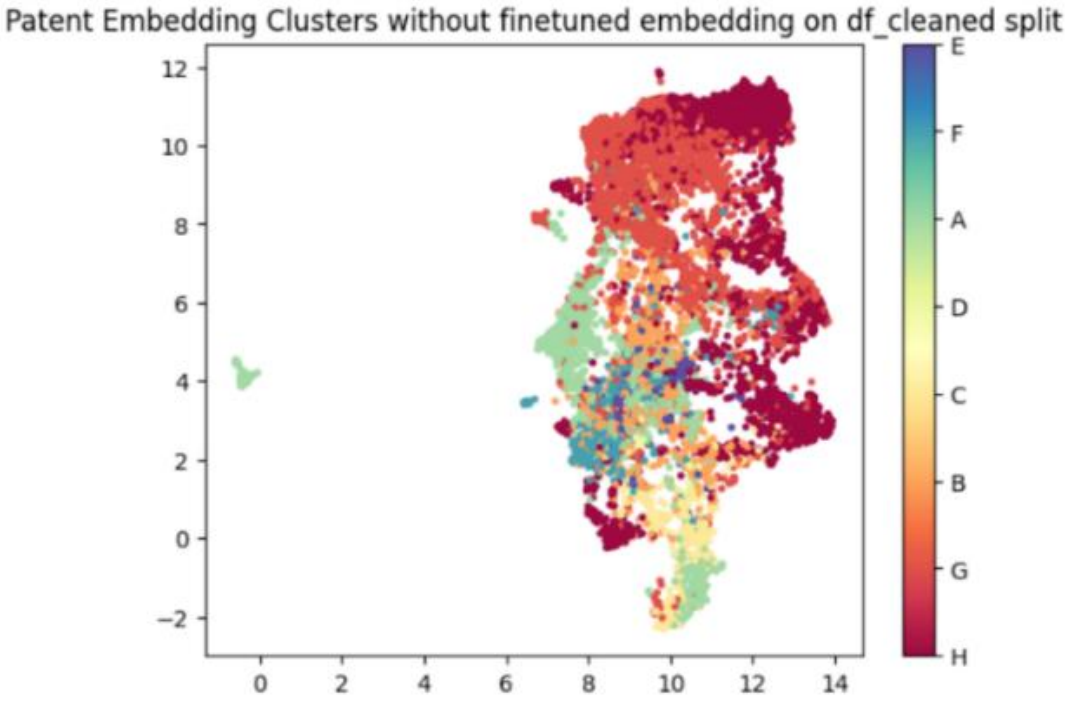
Stage 1: Contrastive Fine-Tuning of E5-Base-V2 Embeddings

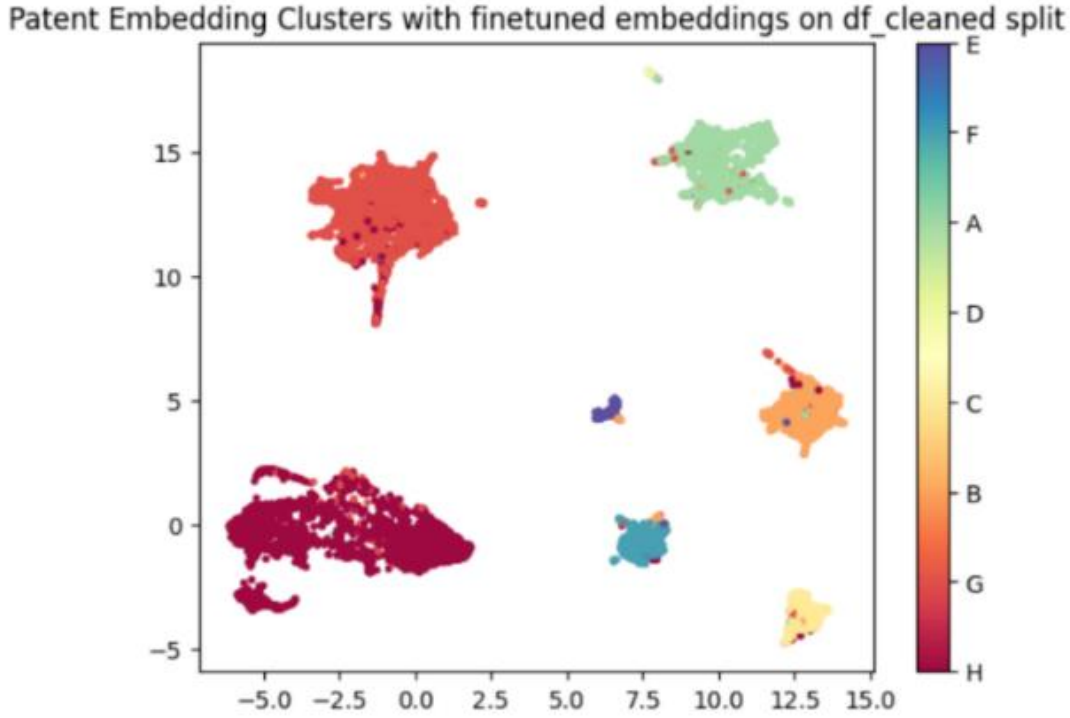


Incorporation of Taxonomy-Aware Features Using CPC Descriptions

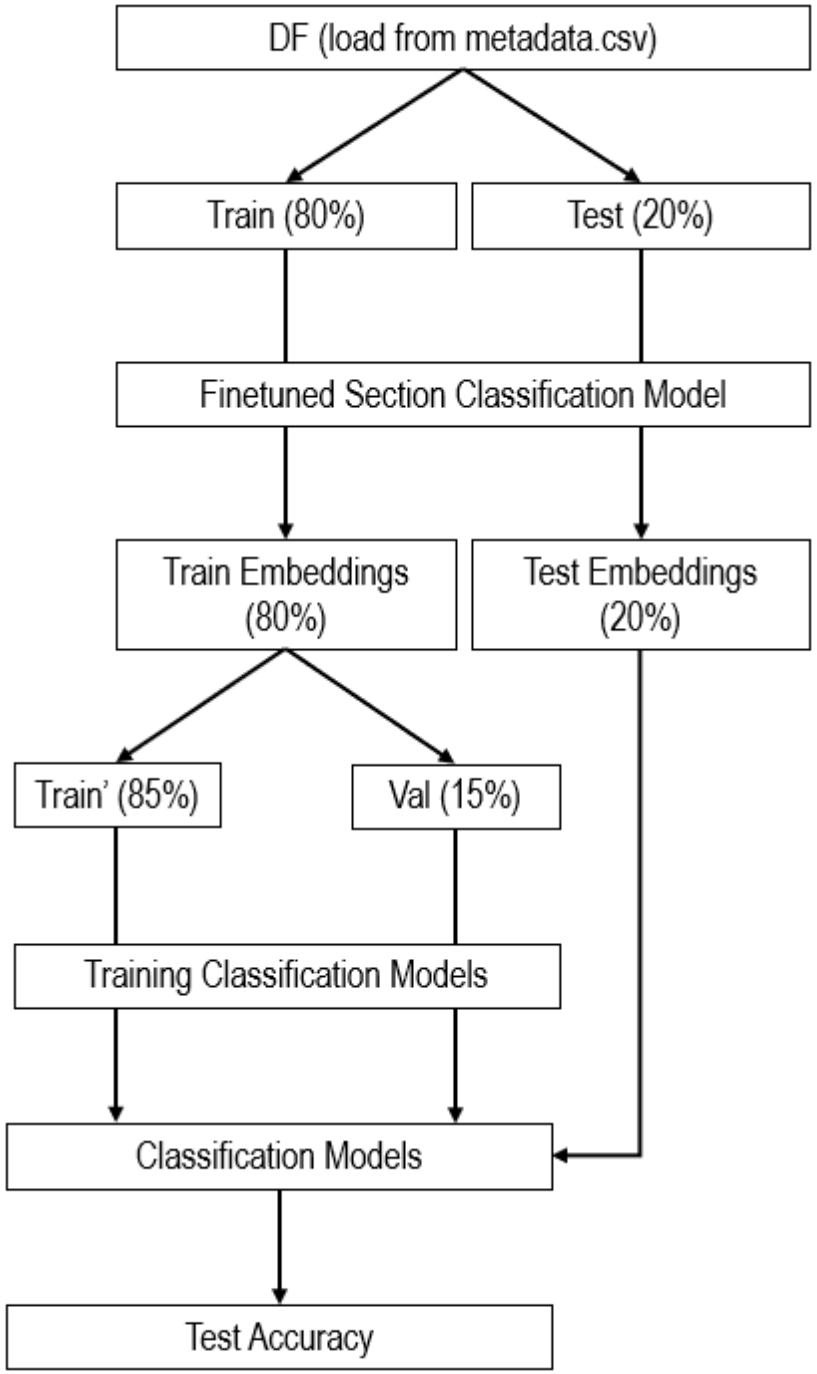
```
cpc_class_list = [
   ["A01", "Agriculture; Forestry; Animal Husbandry; Hunting; Trapping; Fishing"],
   ["A21", "Baking; Equipment for making or processing doughs; Doughs for baking"],
   ["A22", "Butchering; Meat treatment; Processing poultry or fish"],
   ["A23", "Foods or foodstuffs; Their treatment, not covered by other classes"],
   ["A24", "Tobacco; Cigars; Cigarettes; Smokers' requisites"],
   ["A41", "Wearing apparel"],
   ["A42", "Headwear"],
   ["A43", "Footwear"],
   ["A44", "Haberdashery; Jewelry"],
   ["A45", "Hand or travelling articles"],
   ["A46", "Brushware"],
   ["A47", "Furniture; Domestic articles or appliances"],
   ["A61", "Medical or veterinary science; Hygiene"],
   ["A62", "Life-saving; Fire-fighting"],
   ["A63", "Sports; Games; Amusements"],
. . .
```

Embedding Space Before vs. After Fine-Tuning





Stage 2: Training Classification Models



Results & Discussion

Section	Our Model (MoE)	Our Model (Taxonomy Aware)	PatentSBERTa (2024)	Shajalal et al. (2023)
A	0.87	0.85	N/A	0.85
B	0.75	0.72	0.76	0.70
C	0.80	0.76	0.86	0.81
D	0.37	0.16	0.64	0.73
E	0.64	0.64	0.74	0.67
F	0.72	0.68	0.78	0.70
G	0.83	0.83	0.85	0.82
H	0.84	0.84	0.86	0.82
Y	0.07	0.00	0.56	0.41
Micro Avg.	0.81	0.80	0.80	0.78
Macro Avg.	0.65	0.61	0.80	N/A
Instance Avg.	0.84	0.83	0.82	N/A

Next Steps

- Expand training data for rare sections.
- Extend to subclass, main group, subgroup.
- Explore zero-shot classification using taxonomy embeddings.