Automatic Medical Concept Detection on Images: Dividing the Task into Smaller Ones

Notebook for the ImageCLEFmedical Caption 2024. Contributions of the UACH-VisionLab Team.

Axel Moncloa-Muro[1,†], Graciela Ramirez-Alonso[1,*,†] and Fernando Martinez-Reyes[1,†]

[1]*Facultad de Ingeniería, Universidad Autónoma de Chihuahua, Circuito Universitario Campus II, 31125 Chihuahua, Mexico*

Abstract

This paper describes the approach proposed by the UACH-VisionLab team for the ImageCLEFmedical Concept Detection subtask 2024. The objective of this subtask is to assign medical concepts to images automatically. In particular, 1,945 distinct Clinical Concepts of Unique Identifiers (CUIs) must be associated with medical images representing a multi-label classification (MLC) problem. In this context, the ImageCLEFmedical Concept Detection subtask provides a multi-label dataset in which a medical image may contain multiple descriptive labels. The class imbalance problem in MLC poses a challenge where the samples and their corresponding labels are not uniformly distributed over the dataset. To address this challenge, our approach employs an ensemble of five EfficientNet B0 (ENB0) neural architectures. An initial neural network, ENB0, classifies each image into all possible labels. Based on the classification results, we create subgroups of multi-label datasets considering specific CUIs, such as *ultrasonography, bone structure of the cranium, angiogram,* and *lower extremity*. A separate ENB0 architecture is trained for each of these subgroups. Finally, the outputs of these five neural architectures are combined to generate the final prediction results. Our proposal ranks 5th place in the ImageCLEFmedical Concept Detection subtask, achieving an F1-score of 0.59. The code to implement our proposal can be found in https://github.com/axelm11/CLEF-ImageCLEF-2024.

Keywords

Multi-label classification, imbalanced data, EfficientNet, ImageCLEFmedical, ensemble

1. Introduction

ImageCLEF is an ongoing evaluation event launched in 2003 as part of the Cross Language Evaluation Forum (CLEF) [1]. In 2024, the ImageCLEFmedical Lab presents the 8th edition of the automatic image captioning task, which consists of two subtasks: concept detection and caption prediction [2]. The objective of the concept detection subtask is to identify the Unified Medical Language System (UMLS) concepts of each image. These concepts are unique identifiers assigned to different medical-related terms. The training, validation, and test datasets for this subtask comprise 70,108, 9,972, and 17,237 images, respectively. This subtask is considered a multi-label classification problem, where 1,945 different concepts must be detected and a single medical image can be associated with multiple labels. The dataset is highly imbalanced, with four of the most prevalent concepts having a frequency of occurrence in the training set of 24,227, 19,363, 11,296, and 9,870, in contrast to 306 classes that have ten or fewer images. For these reasons, this dataset is particularly challenging and complex, providing an ideal setting for the development of new deep learning (DL) approaches where robust solutions must be capable of identifying the different concepts for each medical image.

In this work, we present our approach, which we submit as part of the UACH-Vision Lab group for the ImageCLEFmedical Concept Detection subtask. This proposal consists of an ensemble of five deep learning models based on the EfficientNet B0 (ENB0) architecture [3]. An initial ENB0 associates each

---

image with 1,945 possible medical concepts. Given the high imbalance of the dataset, an additional four ENB0 models were trained to identify specific concepts and improve the performance of our proposal.

The rest of this paper is organized as follows: Section 2 presents a general description of the ImageCLEFmedical dataset, Section 3 introduces our approach, and Sections 4 and 5 provide results and conclusions.

2. Dataset

The multimodal data utilized in the ImageCLEFmedical Lab is derived from the Radiology Object in Context version 2 (ROCOv2) dataset [4]. This dataset consists of radiological images accompanied by their respective medical concepts and captions. It is comprised of three distinct subsets: the training set, the validation set, and the test set. The training and validation datasets are accompanied by comma-separated value (CSV) files, which contain the medical image identifiers and the corresponding Concept Unique Identifiers (CUIs). The objective of the concept detection task is to automatically assign the corresponding CUIs to the different images of the dataset. Figure 1 shows a visual representation of the medical concepts associated with the different CUIs. In this case, the size of each word is related to its frequency. Among the most frequently occurring concepts are *X-Ray Computed Tomography, Plain x-ray, Ultrasonography, Magnetic Resonance Imaging,* and *Chest*, to mention some.

The task of assigning the 1,945 possible medical concepts to each image in the ImageCLEFmedical dataset is highly challenging, given the high level of complexity involved. For instance, images obtained from the same image modality may describe different conditions affecting different parts of the body. This is exemplified in Figure 2, where images corresponding to the same modality, *X-Ray Computed Tomography*, show different parts of the body emphasizing different medical concepts.

Another case is presented in Figure 3, where different image modalities present the same medical CUI. In this case, an *angiogram, plain x-ray,* and *magnetic resonance imaging* are associated with the CUI *heart*. Therefore, it is possible that one CUI can be present in different image modalities.

Figure 4 shows an additional challenging scenario where images that appear to be highly similar may, in fact, have different CUIs.

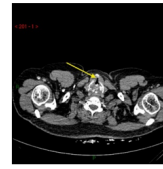**Figure 1:** Word cloud of the medical concepts present on the ImageCLEFmedical dataset.

CC BY-NC [Nghiem et al. (2014)]
C0040405 X-Ray Computed Tomography
C0340161 Hematomediastinum

CC BY-NC [Unterstell et al. (2013)]
C0040405 X-Ray Computed Tomography
C0817096 Chest
C0497156 Lymphadenopathy

CC BY [Muacevic et al. (2021)]
**C0040405 X-Ray Computed Tomography**

Figure 2: Images obtained with the same imaging modality yet showing different anatomical regions of the body emphasizing different medical concepts. CC BY-NC [Nghiem et al. (2014)], CC BY-NC [Unterstell et al. (2013)], CC BY [Muacevic et al. (2021)].
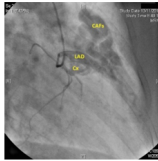
CC BY [Lacalzada-Almeida et al. (2018)]
C0002978 Angiogram
C0018787 Heart
C0003855 Arteriovenous fistula
C0226032 Anterior descending branch
       of left coronary artery
C0205042 Coronary artery
C0034052 Pulmonary artery structure

CC BY-NC [Biharas Monfared et al. (2015)]
C1306645; Plain x-ray
C0817096 Chest
C0018787 Heart
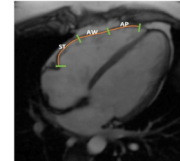
CC BY [Bourfiss et al. (2017)]
C0024485 Magnetic Resonance Imaging
C0018827 Heart Ventricle
C0018787 Heart

Figure 3: The CUI associated with the *heart* concept is present in different image modalities. CC BY [Lacalzada-Almeida et al. (2018)], CC BY-NC [Biharas Monfared et al. (2015)], CC BY [Bourfiss et al. (2017)].

CC BY [Yuasa et al. (2015)]
C1306645 Plain x-ray
C0023216 Lower Extremity
C0588193 Bone structure of shaft of femur

CC BY [Yuasa et al. (2015)]
C1306645 Plain x-ray
C0023216 Lower Extremity
C0029408 Degenerative polyarthritis
C0524470 Right hip region structure

CC BY [Yuasa et al. (2015)]
C1306645 Plain x-ray
C0030797 Pelvis
C0005971 Bone plates

Figure 4: Similar images with different CUIs. CC BY [Yuasa et al. (2015)].

3. Methods

Our proposal is based on the baseline model provided by the ImageCLEFmedical 2024 organizers, an EfficientNet B0 (ENB0) neural architecture. Our team evaluates different neural architecture models, such as ResNet [5], DenseNet [6], the Vision Transformer (ViT) [7], and Convolutional vision Transformer (CvT) [8]. However, the one proposed by the organizers yielded the best F1-scores with the validation set. The results of the ENB0 model indicate that certain CUIs exhibit highly accurate F1 performance while others exhibit zero performance. This discrepancy is primarily attributed to the multi-label class imbalance issue inherent in real-world application datasets [9, 10, 11, 12, 13]. Table 1 presents the top eight best F1-score performances. Based on these results, we select specific CUIs to create four multi-label subgroups to train and validate separate ENB0 models. The number of support samples and visual similarities in the images were considered when selecting these CUIs. For example, the categories

Table 1
Canonical name, CUI, F1-score, and support set of the top eight best classification results obtained with an ENB0 neural architecture.

Canonical name	CUI	F1-score	Support
Ultrasonography	C0041618	0.9943	1,606
X-Ray Computed Tomography	C0040405	0.9737	3,625
Plain x-ray	C1306645	0.9551	2,741
Magnetic Resonance Imaging	C0024485	0.9535	1,437
Bone structure of cranium	C0037303	0.9296	393
Lower Extremity	C0023216	0.8411	463
Angiogram	C0002978	0.8366	421
Upper Extremity	C1140618	0.8060	178

bone structure of cranium, lower extremity and *angiogram* exhibit a comparable number of samples. In contrast, *ultrasonography* is a particularly interesting image modality, given the homogeneity of the images within this subgroup.

Figure 5 shows a block diagram of the proposed approach. First, an initial ENB0 model is trained to classify all the images of the training dataset on all the possible CUIs of the challenge. The output of this model is a vector of dimensionality 1,945. Then, four subgroups are defined based on the classification results of the *ultrasonography, bone structure of cranium, lower extremity* and *angiogram* CUIs. If an image is classified within any of the four aforementioned concepts, it is considered to be part of a specific subgroup. Once the subgroups have been defined, they are trained with a separate ENB0 model to identify the possible medical concepts they contain. During training, we consider it appropriate to eliminate those CUIs with a very high or low-frequency appearance to avoid severe class imbalance issues.
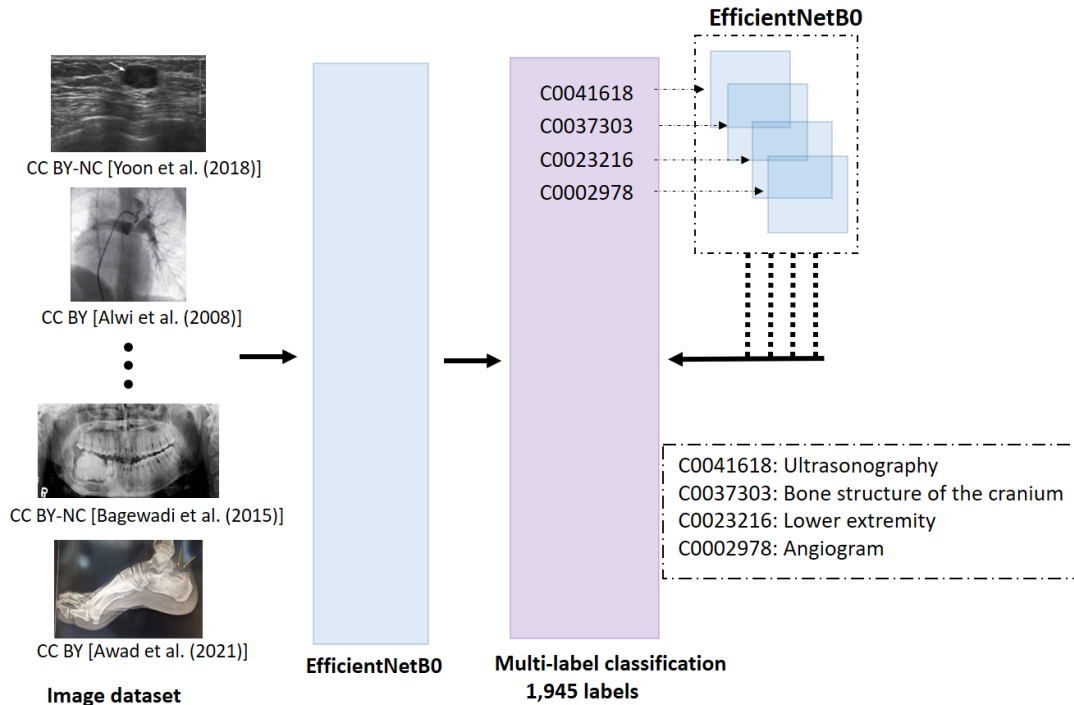
Figure 5: Block diagram of our proposal. An initial ENB0 detects all possible labels. If one of these labels corresponds to the concepts *ultrasonography, bone structure of cranium, lower extremity* or *angiogram*, the initial prediction will be improved with the output of the corresponding ENB0 model. CC BY-NC [Yoon et al. (2018)], CC BY [Alwi et al. (2008)], CC BY-NC [Bagewadi et al. (2015)], CC BY [Awad et al. (2021)].

For example, the concept *plain x-ray* is a very common concept. Therefore, it is eliminated from all the subgroups. For low-frequency concepts, we consider those CUIs with a support set of at least 50 samples and a maximum of 20 concepts to predict for each model.

Then, the proposed methodology is as follows. If the initial ENB0 identifies that the input medical image contains a CUI associated with the concepts of *ultrasonography, bone structure of cranium, lower extremity* or *angiogram*, then the ENB0 model trained with the specific subgroup will also analyze this input image and will produce an output prediction. All possible predictions identified by the second ENB0 will be included in the initial prediction. In other words, four ENB0 neural architectures are employed to enhance the outcome of the initial model. To ensure a precise final prediction, it is essential to exercise caution in determining the location of the CUI, as the output dimensionality of these models differs. Figure 6 illustrates this procedure. In this example, the *angiogram* concept is identified, and the prediction of the model trained with this specific subgroup is utilized to generate the final prediction result. In this case, the second ENB0 model detects four new concepts included in the final prediction.

Once we define the four subgroups, we proceed to analyze the relationship between the different CUIs they contain. Figure 7 shows the chord diagram of the *angiogram* concept. This figure illustrates the relationship between the CUIs within this subgroup. The nodes represent the different concepts, and the width of the edges is proportional to the relationship between the two nodes. Table 2 provides a more detailed overview of the different concepts within this subgroup and the support set of each of them. The most frequent concepts are the *anterior descending branch of left coronary artery, stent device, right coronary artery structure* and *stenosis*. As can be observed in Figure 7, the *anterior descending branch of*

CUI	Canonical Name	1*	2*
C0040405	X-Ray Computed Tomography	1	1
C1306645	Plain x-ray	0	0
C0024485	Magnetic Resonance Imaging	0	0
C0041618	Ultrasonography	0	0
C0817096	Chest	1	1
C0002978	Angiogram	1	1
C0037303	Bone structure of cranium	0	0
C0000726	Abdomen	0	0
C0030797	Pelvis	0	0
C0023216	Lower Extremity	0	0
C0018787	Heart	0	0
C0034052	Pulmonary artery structure	0	1
C0003483	Aorta	0	0
C0006663	Calcinosis	0	0
C0013604	Edema	0	0
C0225883	Right ventricular structure	0	0
C1947917	Occluded	0	1
C0005682	Urinary Bladder	0	0
C1510420	Cavitation	0	0
⋮	⋮		
C0087086	Thrombus	0	0
C0018827	Heart Ventricle	0	0
C0205271	Irregular	0	0
C0021852	Intestines, Small	0	0
C0038257	Stent, device	0	0
C0442800	Enlarged	0	0
C0024687	Mandible	0	0
C0016169	pathologic fistula	0	0
C0085590	catheter device	0	1
C0205054	Hepatic	0	0
C0015811	Femur	0	0
C3714551	Structure of omental adipose tissue (body structu	0	0
C0032743	Positron-Emission Tomography	0	0
C1261287	Stenosis	0	0
C0022646	Kidney	0	0
C0226032	Anterior descending branch of left coronary arter	0	1
C0007776	Cerebral cortex	0	0
C0021102	Implants	0	0

1* Prediction of the initial ENB0 model
2* Final prediction resulting from the incorporation of the Angiogram subgroup model

Angiogram	
Anterior descending branch of left coronary ar	1
Stent, device	0
Right coronary artery structure	0
Stenosis	0
Pulmonary artery structure	1
catheter device	1
Occluded	1
Complete obstruction	0
Aneurysm	0
Pseudoaneurysm	0
Structure of circumflex branch of left coronary	0
Heart	0
Vessel Positions	0
Left coronary artery structure	0
pathologic fistula	0
Caudal	0
Collateral branch of vessel	0

Lower Extremity	
Femur	0
Screw	0
Pelvis	0
Joint Capsule	0
Head of femur	0
Structure of right knee region	0
Right hip region structure	0
Cerebral cortex	0
Ankle region	0
Bone plates	0
Structure of left hip	0
Axis vertebra	0
Implants	0
Structure of left knee region	0
Metatarsal bone structure	0

Figure 6: Example prediction of our proposal. The initial ENB0 model generates an output vector with all possible predictions. In this example, the *angiogram* concept is detected, then the output of a second ENB0 model is incorporated into the initial prediction. Special care must be taken with regard to the dimensions of the output vector of each model.
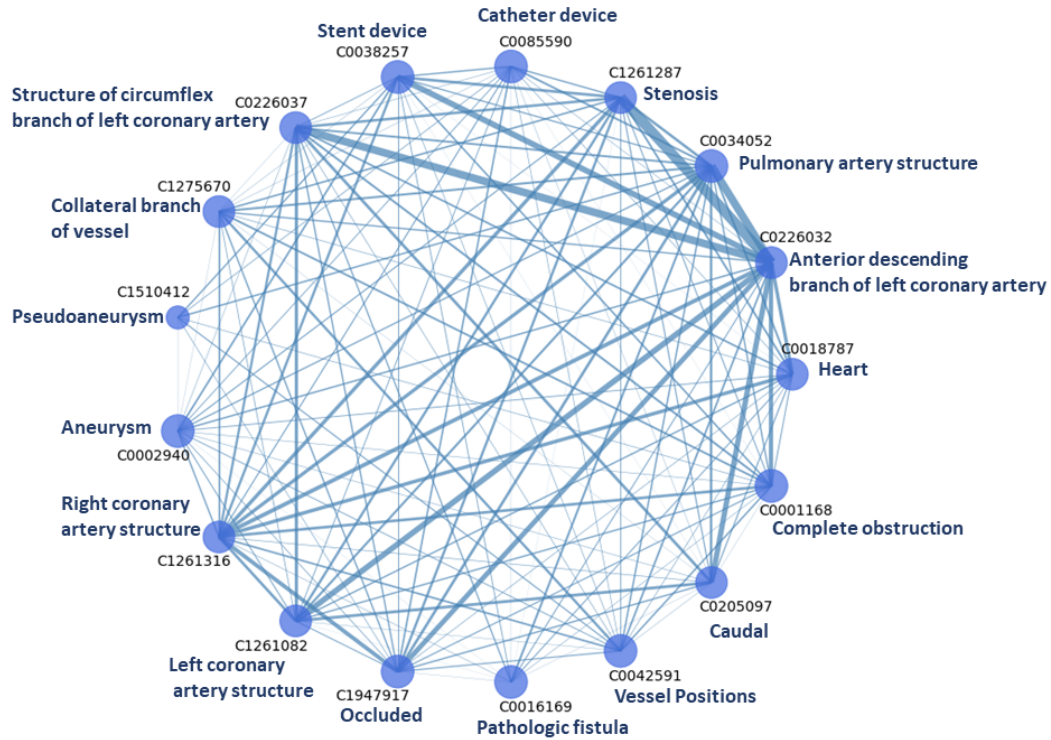
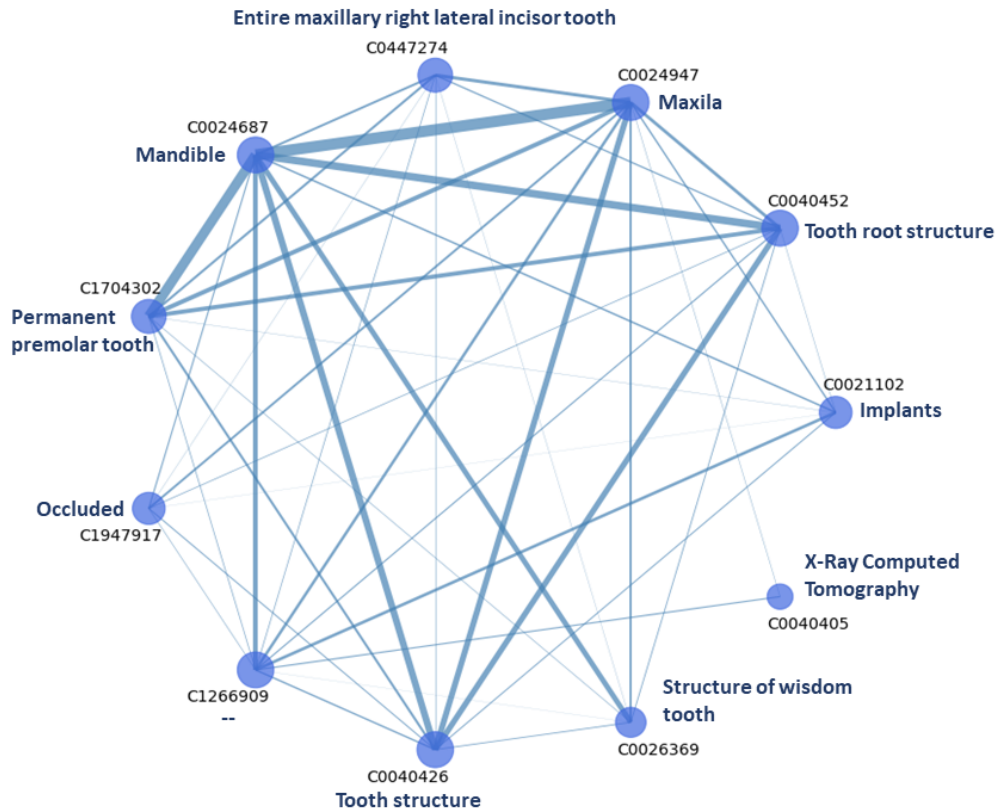**Figure 7:** CUIs and canonical names relationship in the Angiogram subgroup.

Table 2
CUIs, canonical names, and support set of the Angiogram subgroup.

CUI	Canonical Name	Support
C0226032	Anterior descending branch of left coronary artery	448
C0038257	Stent, device	355
C1261316	Right coronary artery structure	302
C1261287	Stenosis	300
C0034052	Pulmonary artery structure	258
C0085590	Catheter device	231
C1947917	Occluded	229
C0001168	Complete obstruction	200
C0002940	Aneurysm	194
C1510412	Pseudoaneurysm	185
C0226037	Structure of circumflex branch of left coronary artery	156
C0018787	Heart	145
C0042591	Vessel Positions	134
C1261082	Left coronary artery structure	129
C0016169	Pathologic fistula	126
C0205097	Caudal	111
C1275670	Collateral branch of vessel	104

left coronary artery has a strong relationship with *stenosis, pulmonary artery structure*, and *structure of circumflex branch of left coronary artery*. Furthermore, it is noteworthy that the *right coronary artery structure* is a frequent medical concept in this subgroup that exhibits a constant relationship with the majority of other concepts, with the exception of *pseudoaneurysm*.

Figure 8 shows the chord diagram of the medical concept *bone structure of cranium*. Table 3 shows the specific canonical names of this subgroup and their support set. As can be observed, *mandible* is the more common medical concept. It has a strong relationship with *permanent premolar tooth,* and *maxilla*

Figure 8: CUIs and canonical names relationship in the Bone Structure of Cranium subgroup.

Table 3
CUIs, canonical names, and support set of the Bone Structure of Cranium subgroup.

CUI	Canonical Name	Support
C0024687	Mandible	472
C0040426	Tooth structure	273
C0024947	Maxilla	265
C1266909	–	174
C0040452	Tooth root structure	172
C0021102	Implants	171
C1704302	Permanent premolar tooth	140
C0026369	Structure of wisdom tooth	81
C1947917	Occluded	67
C0447274	Entire maxillary right lateral incisor tooth	61
C0040405	X-Ray Computed Tomography	61

but also, the concepts *tooth structure, tooth root structure* and *structure of wisdom tooth* are related to it. On the contrary, *X-Ray Computed Tomography* is only slightly related to *maxilla* and the CUI C1266909 (this CUI does not present a canonical name associated with it).

Figure 9 and Table 4 show the chord diagram and CUIs, canonical names, and support set of the *lower extremity* subgroup. *Femur* is the most frequent concept with a strong relationship with *cerebral cortex, axis vertebra,* and *head of femur.* We would like to point out that we are not sure if the *cerebral cortex* should be the correct canonical name of *C0007776.* Furthermore, it can be observed that the medical concepts of *bone plates* and *screw* are closely related.

Ultrasonography is our last subgroup. Figure 10 shows its relationship chord diagram, and Table 5 presents the canonical names and support set of this subgroup. *Left ventricular structure* and *right*
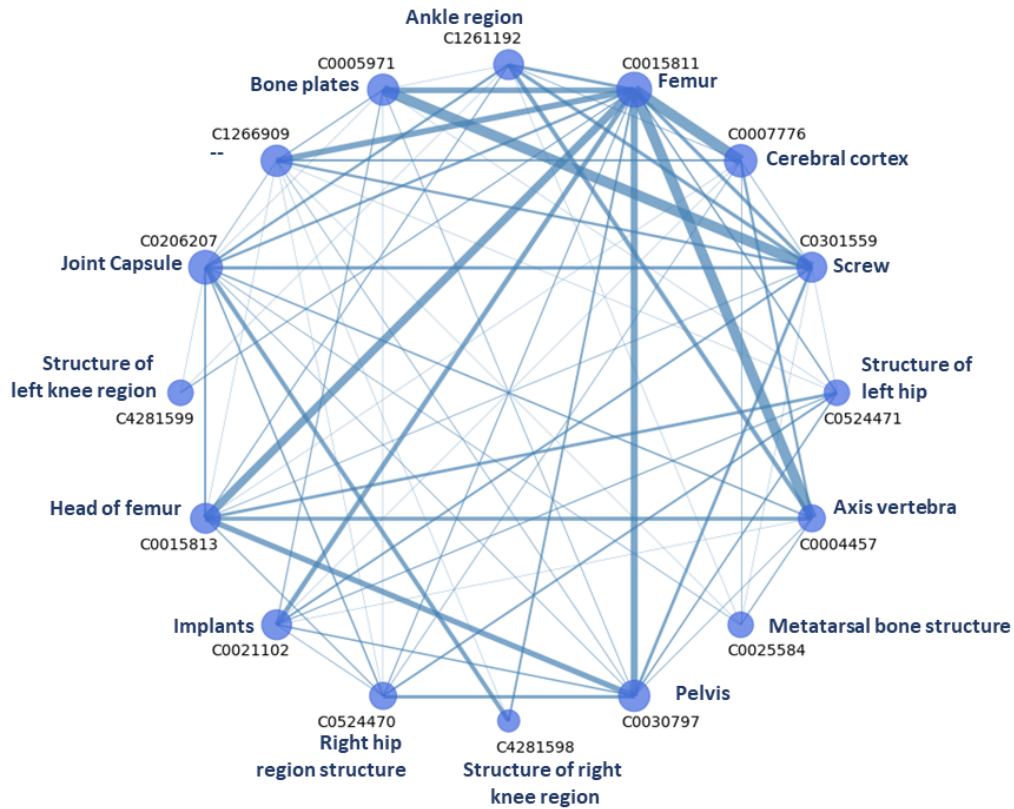
Figure 9: CUIs and canonical names relationship in the Lower Extremity subgroup.

Table 4
CUIs, canonical names, and support set of the Lower Extremity subgroup.

CUI	Canonical Name	Support
C0015811	Femur	318
C0301559	Screw	119
C0030797	Pelvis	116
C0206207	Joint Capsule	103
C1266909	–	102
C0015813	Head of femur	93
C4281598	Structure of right knee region	91
C0524470	Right hip region structure	83
C0007776	Cerebral cortex	78
C1261192	Ankle region	77
C0005971	Bone plates	75
C0524471	Structure of left hip	74
C0004457	Axis vertebra	72
C0021102	Implants	69
C4281599	Structure of left knee region	64
C0025584	Metatarsal bone structure	50

ventricular structure are the more common concepts and present a high relationship between them. *Right atrial structure* is another common concept, and it can be observed that it is associated with the concepts *left ventricular structure, right ventricular structure* and *left atrial structure*.
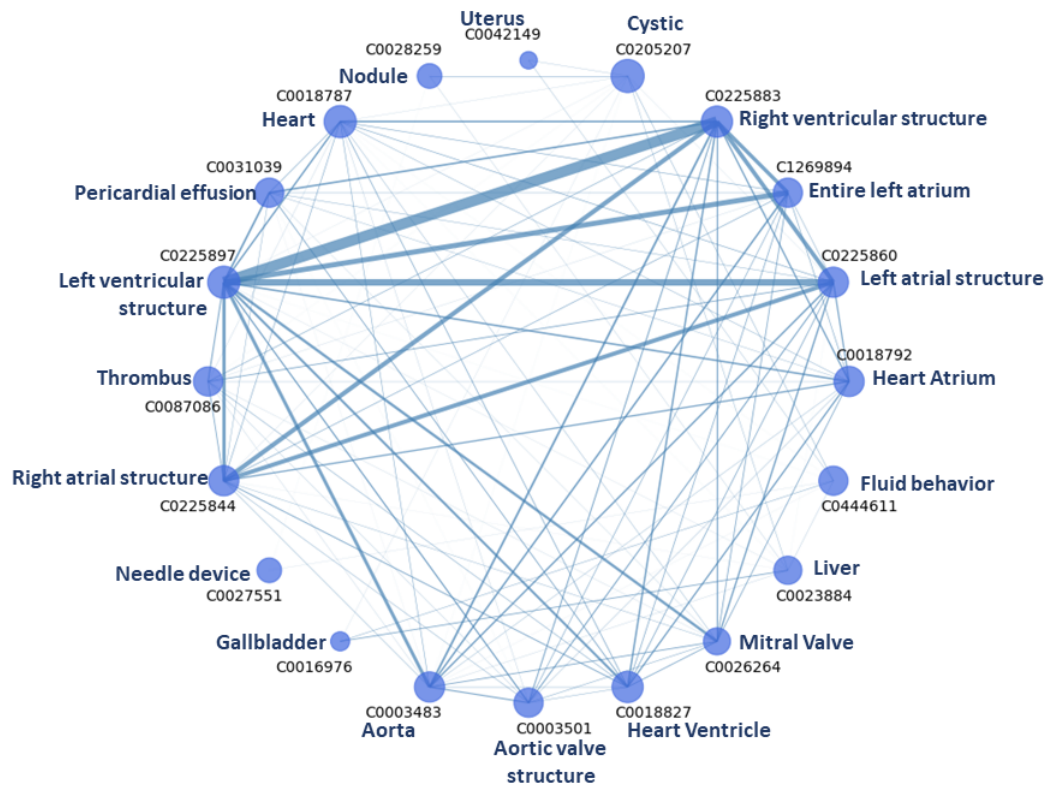
Figure 10: CUIs and canonical names relationship in the Ultrasonography subgroup.

Table 5
CUIs, canonical names, and support set of the Ultrasonography subgroup.

CUI	Canonical Name	Support
C0225897	Left ventricular structure	671
C0225883	Right ventricular structure	538
C0225860	Left atrial structure	380
C0205207	Cystic	340
C0018827	Heart Ventricle	332
C0225844	Right atrial structure	319
C0003483	Aorta	294
C0018792	Heart Atrium	278
C0031039	Pericardial effusion	253
C0026264	Mitral Valve	247
C0444611	Fluid behavior	241
C0023884	Liver	237
C0087086	Thrombus	235
C1269894	Entire left atrium	233
C0018787	Heart	214
C0003501	Aortic valve structure	207
C0016976	Gallbladder	206
C0027551	Needle device	193
C0042149	Uterus	190
C0028259	Nodule	190

4. Results

All the neural models were trained on an NVIDIA GeForce RTX 3080 Ti 12GB GPU using the PyTorch framework and the Adam Optimizer, with an initial learning rate of 1e-3 using a batch size of 64.

Table 6 shows the results of our team, UACH-VisionLab, with the test partition dataset. These results were provided by the ImageCLEFmedical Lab 2024 organizers. The F1-score is a measure of the harmonic mean of precision and recall. A secondary F1-score was calculated using a subset of concepts that was manually curated. Two runs were submitted by our team. The first run use a *drop path rate* of 0.2 while the second a *drop path rate* of 0.3, with a *weight decay factor* of 1e-5.

The results presented in Table 6 demonstrate that the first run achieves a superior performance. The increase in the *drop path rate* and the use of the L2 regularization method affect the performance of the model, reducing its generalization ability with test data.

In order to gain a deeper understanding of the manner in which the incorporation of the four ENB0 models enhances the performance of our approach, Table 7 presents the results of the precision, recall, and F1-score metrics on randomly selected CUIs. The first three columns show the results obtained when only one ENB0 model is employed, defined as the "Base" model. Subsequently, the approach was further enhanced by incorporating the training of the *lower extremity* (LE) subgroup defining the "Base+LE" approach. The "Base+LE+Angio" approach was created by additionally including the *angiogram* subgroup. The "Base+LE+Angio+Ultrasono" approach was constructed by combining the LE and *angiogram* subgroups with *ultrasonography*. Finally, the "Base+LE+Angio+Ultrasono+Cranium" approach integrates the *bone structure of cranium* subgroup.

A green highlight in Table 7 indicates a metric improvement, whereas a yellow highlight indicates a metric decrease. It is important to note that the improvements in the F1-score are mainly related to an increase in the recall score. The recall metric measures how often a true positive image is identified, whereas the precision metric considers how many positive predictions are true positive samples. Consequently, if the model detects only one true positive sample with a specific CUI, the precision metric will be high. In contrast, the recall metric will exhibit a low performance (as observed, for example, in the third row of Table 7 where many false negative samples are detected). Consequently, with fewer false negative detections but more false positives, the precision metric will decrease (highlighted in yellow), while the recall metric will increase, resulting in an improved F1-score metric (highlighted in green).

The improvements in the F1-score metric resulting from the incorporation of the *lower extremity* subgroup (Base+LE apporach) are *structure of left hip, femur, joint capsule, screw*, and *head of femur*. All of these medical concepts are considered in the training of this subgroup.

The improvement in the concepts detection resulting from the incorporation of the *angiogram* subgroup (Base+LE+Angio approach) includes the *stent device, caudal, structure of circumflex branch of left coronary artery, collateral branch of vessel, pseudoaneurysm* and *vessel positions*. It should be noted that all the aforementioned improvements, which had been reported in the previous approach (Base+LE), are maintained in this one, but only those that are new are highlighted in these three columns. This same reporting strategy is used in the remaining approaches.

The training and incorporation of the *ultrasonography* subgroup results in the Base+LE+Angio+Ultrasono approach. The concepts that demonstrate an improvement in the F1-score metric are *liver, heart atrium, right atrial structure, aorta, mitral valve, right ventricular structure, uterus, heart ventricle, thrombus,* and *pericardial effusion*. The medical concept *heart atrium* was also slightly modified with the training and incorporation of the *bone structure of cranium* subgroup. However, this is the only concept that was modified. No additional improvements could be identified with the Base+LE+Angio+Ultrasono+Cranium approach.

Table 6

Test results of the Concept Detection subtask on the ImageCLEFmedical Concept Lab 2024. Two runs were submitted by our team. The first run use a *drop path rate* of 0.2 while the second a *drop path rate* of 0.3, with a *weight decay factor* of $1e\text{-}5$.

Team	F1-score	Secondary F1-score
1st run - UACH-VisionLab	0.59876	0.93631
2nd run - UACH-VisionLab	0.52921	0.84224

Table 7

Comparison of precision, recall, and F1-score across the different approaches with the validation dataset. The Base model corresponds to employing only one ENB0 model, Base+LE incorporates the training of the *lower extremity* (LE) subgroup, Base+LE+Angio includes the training of the *angiogram* subgroup, Base+LE+Angio+Ultrasono combines the LE and *angiogram* subgroups with the *ultrasonography*, and Base+LE+Angio+Ultrasono+Cranium integrates the *bone structure of cranium* subgroup. A green highlighting is related to a metric improvement, whereas a yellow highlight indicates a metric decrease.

Base			Base+LE			Base+LE+Angio			Base+LE+Angio+Ultrasono			Base+LE+Angio+Ultrasono+Cranium			CUI	Canonical Name
precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score		
0.4583	0.0873	0.1467	0.4583	0.0873	0.1467	0.4583	0.0873	0.1467	0.2268	0.1746	0.1973	0.2268	0.1746	0.1973	C0023884	Liver
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1591	0.1148	0.1333	0.1167	0.1148	0.1157	C0018792	Heart Atrium
1.0000	0.0130	0.0256	1.0000	0.0130	0.0256	1.0000	0.0130	0.0256	0.2857	0.1299	0.1786	0.2857	0.1299	0.1786	C0225844	Right atrial structure
0.0000	0.0000	0.0000	0.0909	0.0435	0.0588	0.0909	0.0435	0.0588	0.0909	0.0435	0.0588	0.0909	0.0435	0.0588	C0524471	Structure of left hip
0.0000	0.0000	0.0000	0.1264	0.1392	0.1325	0.1264	0.1392	0.1325	0.1264	0.1392	0.1325	0.1264	0.1392	0.1325	C0015811	Femur
0.2174	0.0500	0.0813	0.2174	0.0500	0.0813	0.2174	0.0500	0.0813	0.1739	0.1200	0.1420	0.1739	0.1200	0.1420	C0003483	Aorta
0.4286	0.0361	0.0667	0.4286	0.0361	0.0667	0.2400	0.1446	0.1805	0.2400	0.1446	0.1805	0.2400	0.1446	0.1805	C0038257	Stent, device
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3571	0.1087	0.1667	0.3571	0.1087	0.1667	0.3571	0.1087	0.1667	C0205097	Caudal
0.0000	0.0000	0.0000	0.0455	0.0179	0.0256	0.0455	0.0179	0.0256	0.0455	0.0179	0.0256	0.0455	0.0179	0.0256	C0206207	Joint Capsule
0.5000	0.0154	0.0299	0.1667	0.0462	0.0723	0.1667	0.0462	0.0723	0.1667	0.0462	0.0723	0.1667	0.0462	0.0723	C0301559	Screw
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2500	0.1351	0.1754	0.2500	0.1351	0.1754	C0026264	Mitral Valve
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6000	0.1111	0.1875	0.6000	0.1111	0.1875	0.6000	0.1111	0.1875	C0226037	Structure of circumflex branch of left coronary artery
1.0000	0.0500	0.0952	1.0000	0.0500	0.0952	0.3333	0.1000	0.1538	0.3333	0.1000	0.1538	0.3333	0.1000	0.1538	C1275670	Collateral branch of vessel
0.5769	0.1282	0.2098	0.5769	0.1282	0.2098	0.5769	0.1282	0.2098	0.3095	0.3333	0.3210	0.3047	0.3333	0.3184	C0225883	Right ventricular structure
0.2500	0.0156	0.0294	0.2500	0.0156	0.0294	0.2500	0.0156	0.0294	0.1648	0.2344	0.1935	0.1042	0.2344	0.1442	C0042149	Uterus
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1609	0.1359	0.1474	0.1609	0.1359	0.1474	C0018827	Heart Ventricle
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1364	0.1034	0.1176	0.1364	0.1034	0.1176	0.1364	0.1034	0.1176	C1510412	Pseudoaneurysm
0.2308	0.1429	0.1765	0.2000	0.1667	0.1818	0.2000	0.1667	0.1818	0.2000	0.1667	0.1818	0.2000	0.1667	0.1818	C0015813	Head of femur
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0667	0.0288	0.0403	0.0694	0.0481	0.0568	C0087086	Thrombus
0.7143	0.2353	0.3540	0.7143	0.2353	0.3540	0.7143	0.2353	0.3540	0.4359	0.4000	0.4172	0.1545	0.4235	0.2264	C0031039	Pericardial effusion
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1000	0.0208	0.0345	0.1000	0.0208	0.0345	0.1000	0.0208	0.0345	C0042591	Vessel Positions

5. Conclusion

This working note paper presents the approach and results of the UACH-VisionLab team on the ImageCLEFmedical 2024 Concept Detection subtask. An analysis of the results yielded by the baseline code provided by the organizers reveals a significant imbalance issue in the context of multi-label classification. Therefore, we consider it appropriate to define subgroups with the aim of reducing this class imbalance problem. The medical concepts of *ultrasonography, bone structure of the cranium, lower extremity* and *angiogram* are identified as appropriate for use in the construction of these subgroups. Each subgroup is trained separately, and their results are merged with those produced by an initial ENB0 neural model.

Upon examination of the validation results obtained in the various iterations of our experiments, we observe an increase in the recall metric. This indicates that our approach has reduced the number of false negative detections, which is the behavior we are looking for in class imbalance datasets. However, it has also resulted in an increase in the number of false positives, decreasing the precision metric. The only subgroup that does not produce an improvement in the metric results is the *bone structure of cranium*. Further investigation is required in order to gain an understanding of this behavior.

A chord diagram of the formed subgroups provides a more comprehensive understanding of the diverse concepts within them and their interconnections. Unfortunately, due to time constraints, we were unable to incorporate this crucial knowledge into the training of the models. However, we consider it to be of paramount importance, and we intend to incorporate this information into future approaches.

References

[1] B. Ionescu, H. Müller, A. Drăgulinescu, J. Rückert, A. Ben Abacha, A. García Seco de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, T. M. G. Pakull, H. Damm, B. Bracke, C. M. Friedrich, A. Andrei, Y. Prokopchuk, D. Karpenka, A. Radzhabov, V. Kovalev, C. Macaire, D. Schwab, B. Lecouteux, E. Esperança-Rodier, W. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, S. A. Hicks, M. A. Riegler, V. Thambawita, A. Storås, P. Halvorsen, M. Heinrich, J. Kiesel, M. Potthast, B. Stein, Overview of ImageCLEF 2024: Multimedia retrieval in medical applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 15th International Conference of the CLEF Association (CLEF 2024), Springer Lecture Notes in Computer Science LNCS, Grenoble, France, 2024.

[2] J. Rückert, A. Ben Abacha, A. G. Seco de Herrera, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, B. Bracke, H. Damm, T. M. G. Pakull, C. S. Schmidt, H. Müller, C. M. Friedrich, Overview of ImageCLEFmedical 2024 – Caption Prediction and Concept Detection, in: CLEF2024 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org, Grenoble, France, 2024.

[3] M. Tan, Q. V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, in: Proceedings of the 36th International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

[4] J. Rückert, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, C. S. Schmidt, S. Koitka, O. Pelka, A. B. Abacha, A. G. S. de Herrera, H. Müller, P. A. Horn, F. Nensa, C. M. Friedrich, ROCOv2: Radiology Objects in COntext version 2, an updated multimodal image dataset, Scientific Data (2024). URL: https://arxiv.org/abs/2405.10004v1. doi:10.1038/s41597-024-03496-6.

[5] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[6] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269. doi:10.1109/CVPR.2017.243.

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words:

Transformers for Image Recognition at Scale, in: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, OpenReview.net, 2021.

[8] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, L. Zhang, CvT: Introducing Convolutions to Vision Transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 22–31. doi:`10.1109/ICCV48922.2021.00009`.

[9] J. Ye, L. Jiang, S. Xiao, Y. Zong, A. Jiang, Multi-Label Image Classification Model Based on Multiscale Fusion and Adaptive Label Correlation, Journal of Shanghai Jiaotong University (Science) (2024) 1–10. doi:`10.1007/s12204-023-2688-6`.

[10] H. Liz, J. Huertas-Tato, M. Sánchez-Montañés, J. Del Ser, D. Camacho, Deep learning for understanding multilabel imbalanced Chest X-ray datasets, Future Generation Computer Systems 144 (2023) 291–306. doi:`10.1016/j.future.2023.03.005`.

[11] L. Chen, Y. Wang, H. Li, Enhancement of DNN-based multilabel classification by grouping labels based on data imbalance and label correlation, Pattern Recognition 132 (2022) 108964. doi:`10.1016/j.patcog.2022.108964`.

[12] J. Duan, X. Yang, S. Gao, H. Yu, A partition-based problem transformation algorithm for classifying imbalanced multi-label data, Engineering Applications of Artificial Intelligence 128 (2024) 107506. doi:`10.1016/j.engappai.2023.107506`.

[13] K. Zhang, Z. Mao, P. Cao, W. Liang, J. Yang, W. Li, O. R. Zaiane, Label correlation guided borderline oversampling for imbalanced multi-label data learning, Knowledge-Based Systems 279 (2023) 110938. doi:`10.1016/j.knosys.2023.110938`.