IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 January 2024, accepted 21 March 2024, date of publication 8 April 2024, date of current version 26 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3384701

RESEARCH ARTICLE

Predicting Semantic Concepts in Heterogeneous Radiographic Images

DIANA MIRANDA[1], VEENA THENKANIDIYOOR[1], (Member, IEEE), AND DILEEP AROOR DINESH[2], (Member, IEEE)

[1]Department of Computer Science and Engineering, National Institute of Technology Goa, Ponda, Goa 403401, India
[2]Department of Computer Science and Engineering, Indian Institute of Technology Dharwad, Dharwad, Karnataka 580011, India

Corresponding author: Dileep Aroor Dinesh (addileep@iitdh.ac.in)

ABSTRACT Concept detection in a radiological image involves the identification of biomedical semantic entities within the given image. However, different modalities of radiological images make it difficult to design a single suitable approach that can handle this heterogeneity. Such imaging data also suffers from the problem of underrepresented sparse concepts where several concepts are present in very few training images making it difficult to train machine learning models to correctly predict their occurrence. This paper proposes a hierarchical approach for concept detection in radiological images using deep features extracted from the layers of convolutional neural networks. At the first level, the modality of the radiological image is identified. The second level of classification detects concepts present in the input image using multi-label classification by considering only those concepts that are relevant to images belonging to the same modality as the input image. This multi-label classification is performed using suitable classifiers that efficiently handle underrepresented sparse concepts. The proposed hierarchical approach for concept detection in radiological images outperforms state-of-the-art methods for different datasets.

INDEX TERMS Medical image modalities, radiological images, semantic concept detection, underrepresented sparse concepts.

I. INTRODUCTION

Radiological images involve the use of different radiographic techniques to visualize the internal structure of a human's body. A low dosage of radiation is given to the patient to obtain such images that aid in the diagnosis of his/her condition. However, a qualified radiologist is required to correctly interpret the findings in the radiological images. It will be helpful to radiologists if routine normal images are automatically identified so that they can focus only on the critical ones. This corresponds to automating diagnosis using artificial intelligence techniques. Towards this, it is imperative to identify the biomedical semantic concepts in such radiographic images. This task is called concept detection. Once such biomedical semantic concepts or semantic entities are determined, they could also be used to facilitate indexing for content-based image retrieval and search engines that attempt to retrieve medical images similar to a given image [1]. The proposed work focuses on the automatic detection of biomedical semantic concepts in radiographic images.

Automating the task of concept detection in radiological images is challenging due to the heterogeneous nature of such images. Radiological images can belong to one of several different imaging modalities that use different techniques to obtain images [2], [3]. For example, X-ray scans use X-rays while ultrasound scans use sound waves to produce images. Computed tomographic (CT) scans use X-rays in addition to the use of computing algorithms to produce enhanced reconstructions of body parts. Magnetic fields and radio frequency signals are used to produce magnetic resonance imaging (MRI) scans. Even the intensity levels of radiation vary for images belonging to the same modality. Variations in the specifications of biomedical devices may produce images that appear significantly different in terms of their visual characteristics. These factors could confuse

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Zuo.

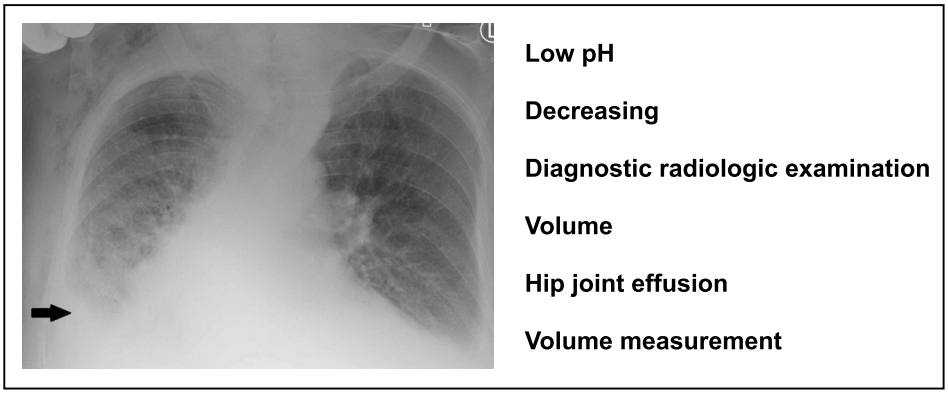

FIGURE 1. A chest radiographic image along with its associated biomedical semantic concepts.

a machine-learning model built to detect concepts in such images. Another challenge with radiological image concept detection is the presence of concepts that are not obvious in the image. In Fig. 1, a chest radiograph is shown that contains six concepts associated with the given image. One of the associated concepts is "decreasing". Although the picture indicates fluid in the lungs, it is difficult to ascertain if the fluid has decreased unless the history of the patient is provided. It is also important to note that concept detection is different from object detection since concepts inherent to the image may not be directly visible in the image. Such concepts are called biomedical semantic concepts. Therefore, segmentation-based approaches may not be of much help in detecting these concepts.

Identification of biomedical semantic concepts in radiological images suffers from two main issues, namely, (1) heterogeneity of images, and (2) underrepresented sparse concepts. This heterogeneous nature of the images results in significant intra-concept variability and inter-concept similarity. Intra-concept variability means that images containing the same concept may appear visually very different. Inter-concept similarity indicates that images containing two different concepts could have the same visual appearance. Automating the task of predicting concepts in such heterogeneous data can be difficult. It is also observed that images belonging to the same radiological modality usually have several common concepts. In situations of high inter-class similarity, hierarchical approaches are found to be effective [4]. This approach involves considering the classes exhibiting high inter-class similarity collectively as a single class in the first level of hierarchical classification, and in the second level, to consider individual classes within the group formed at the first level. Inspired by this, we propose to consider a hierarchical approach for biomedical semantic concept detection. It is also observed that certain unique biomedical concepts are associated with every image captured using a particular modality.

In the proposed hierarchical approach to biomedical semantic concept detection, we propose to identify the modality of a radiological image at the first level of hierarchy. Once the modality is identified, the semantic concept detection concerning a specific modality is proposed at the second level of the hierarchy. This reduces the search space for semantic concepts belonging to a particular modality, which in turn reduces the ambiguity in concept detection. We propose to consider support vector machine (SVM) based classification of deep features extracted from the penultimate layers of a convolutional neural network (CNN), to identify the modality of the radiological image at the first level of the proposed hierarchy. At the second level of the proposed hierarchical approach, we need to have biomedical concept detectors. We propose to have modality-specific concept detectors. The concept detection is a multi-label classification task since multiple biomedical semantic concepts are associated with a single medical image. There are different methods to identify such biomedical semantic concepts in medical images. Biomedical semantic concepts occurring in similar training images can be associated with a test image. In some approaches, biomedical semantic concepts are predicted as words using a caption prediction approach [5]. However, we propose to use multi-label classification since it is effective in identifying common as well as rare biomedical semantic concepts.

An important issue in biomedical semantic concept detection is based on the observation that some biomedical semantic concepts are found in very few training examples. These are called underrepresented sparse concepts. It becomes challenging for machine learning-based approaches to predict the occurrence of such concepts due to significant bias generated in the absence of sufficient training data. These underrepresented sparse concepts still hamper the performance of the multi-label classification task. Such concepts occur in very few images but may be highly clinically significant. Therefore, it is important for the concept detection method to successfully annotate images with these concepts if present. In this work, we propose to use the Rakel algorithm to handle such sparse underrepresented concepts [6]. This algorithm is found to work well for large

datasets that have rare concepts. The contributions of this work are as follows:

- A hierarchical approach to biomedical semantic concept detection is proposed. At the first level of the hierarchy, the deep features of the radiological image are used to identify the radiological image modality using an SVM-based classification approach.
- At the second level of the hierarchy, an approach to biomedical semantic concept detection that involves modality-specific biomedical concept detectors is used to effectively handle underrepresented sparse concepts.

This paper is organized as follows: Section II surveys the recent work done in detecting concepts in medical images. Section III outlines the proposed work. Section IV describes the experimental results and provides an analysis of the performance of the proposed approach. Section V concludes the paper.

II. RELATED WORK

Conventionally, handcrafted features like scale-invariant feature transform (SIFT), local binary pattern (LBP), color and edge directivity descriptor (CEDD), gray level co-occurrence matrix (GLCM), and quad bag-of-colors (QBoC) were popular for such machine learning applications that involve medical images [7], [8], [9], [10], [11], [12]. However, there are two main problems with the use of such features for radiological images containing different imaging modalities. Firstly, a given handcrafted feature may not be suitable for capturing the visual characteristics of all the radiological image modalities. Secondly creating different handcrafted features for each modality is a cumbersome task. In contrast to this, in recent times, most of the research work on medical images suggests that deep features extracted from CNNs prove to be more effective if the dataset has heterogeneous visual features [13].

The main advantage of using CNNs is that they can serve dual functions of feature extraction as well as classification [14]. To maximize its performance, a CNN must be trained on huge volumes of data. In the absence of sufficient training data, a pre-trained CNN can be used [13]. A CNN that is already trained using huge amounts of data for a specific classification task is called a pre-trained CNN. The same CNN can then be used to perform classification for a similar task by fine-tuning its weights to classify data for another classification task. Obtaining a CNN that is pre-trained on a similar dataset is crucial. Most deep learning applications use CNNs that are pre-trained on the ImageNet dataset [15] that contains natural images. In [17] and [16], the authors used a DenseNet CNN that is pre-trained on chest X-ray images to perform concept detection in radiological images. This has shown significant improvement in performance as compared to CNNs pre-trained on natural images [18], [19].

The existing work in concept detection for radiological images can be broadly classified into four major categories based on the approaches followed: (1) multi-label classification-based approaches, (2) similarity search-based-approaches, (3) encoder-decoder based approaches, and (4) caption prediction based approaches [5]. In multi-label classification, the test radiological image is associated with multiple labels or concepts [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30]. Most multi-label classification approaches, fine-tune a pre-trained CNN to predict if the images contain the given concepts present in the dataset [31], [32], [33], [34], [35], [36], [37]. Other approaches perform multi-label classification using deep features extracted from transformers [38], [39], [40], [41]. Similarity-based retrieval techniques identify similar training images and determine the concepts associated with these training images [19], [43], [44], [45], [46], [47]. The concepts that are most relevant to the test image are then selected and assigned to the test image. The encoder-decoder-based approach encodes images using a CNN and detects entities using a recurrent neural network (RNN) or long short-term memory (LSTM) decoder [48], [49], [50], [51], [52]. Some methods predict the caption associated with the image and then extract relevant biomedical entities as concepts from these captions [48], [51]. Similarity-based approaches are effective in identifying rare concepts. However, they perform poorly in detecting frequently occurring concepts as compared to multi-label classification [7], [16], [17]. Methods using an encoder-decoder-based approach are effective in identifying correlated biomedical semantic concepts. Approaches that extract biomedical semantic concepts from predicted captions suffer from the problem of caption prediction task errors propagating to the concept detection task. Therefore, the multi-label classification-based method outperforms the other approaches if common as well as underrepresented concepts need to be correctly identified.

Most of the research done in radiological image concept detection is part of the ImageCLEF challenges that began in 2017 [53]. Following the initial challenges, the organizers concluded that it was better to restrict the number of medical image modalities used in the challenge to maximize the performance of the proposed solutions. It was also reported that general concepts such as ''relationship conjunction - and'' and ''medical image'' do not have any clinical relevance to the task [54]. During subsequent challenges, such concepts were dropped as they did not offer any medical utility. Several concepts have very few training images annotated with them as their ground truth. This introduces significant bias in the learning model leading to incorrect predictions, especially if CNNs are used. Due to this uneven distribution of training examples for the concepts, several authors propose different strategies to combat this bias. Some prune the dataset to include only those concepts whose frequency exceeds a minimum threshold value [54], [55]. The authors in [56] propose to identify frequently occurring concepts as major concepts and proceed to first detect them. Using these predictions, they predict infrequently occurring minor

TABLE 1. Details of the major concepts associated with some radiological medical image modalities.

Modality	Associated Concept
Angiogram (DRAN)	Angiogram
Combined modalities in one image (DRCO)	-
CT scan (DRCT)	Tomography, Emission-Computed and X-Ray Computed Tomography
MRI scan (DRMR)	Magnetic Resonance Imaging
PET scan (DRPE)	Positron-Emission Tomography
Ultrasound scan (DRUS)	Ultrasonography
XRay scan (DRXR)	Diagnostic radiologic examination

concepts. Other approaches reduce the number of concepts to a manageable level by identifying the modality of the radiological image and then training the classifiers to predict only those concepts that are relevant to that particular radiological image modality [17]. However, it is observed that although this approach is successful in identifying biomedical semantic concepts that are common to images belonging to a single radiological image modality, the concept detection model still finds it difficult to predict underrepresented sparse concepts. Therefore, we propose to consider a hierarchical approach in which the modality of a radiological image is determined at the first level using the deep features of the image. At the second level, multi-label classification is performed using suitable classifiers that handle the problem of underrepresented sparse concepts.

III. PROPOSED APPROACH

The proposed approach involves a hierarchical method comprising modality identification followed by concept detection. This is based on our observation on the ImageCLEF 2020 concept detection dataset that images belonging to different modalities have different sets of major concepts associated with them. Upon observing the data, it is seen that most of the images belonging to a single modality have certain major concepts associated with them. Table 1 shows these major concepts present in the images belonging to radiological image modalities. For example, two major concepts occur in all the CT scan images which are "Tomography, Emission-Computed" and "X-Ray Computed Tomography". This also indicates that the name of the modality to which the image belongs is one of the semantic concepts associated with it. Hence, we propose to use a hierarchical approach. At the first level of the hierarchy, we propose to detect the modality of a radiological image. At the next level of the hierarchy, we propose to detect concepts in a radiological image based on the modality to which the image belongs. If an image I belongs to a modality m, the set of all q biomedical semantic concepts associated with modality m given as $C^m = \{c_1^m, c_2^m, \ldots, c_q^m\}$ are considered as potential concepts for the concept detection task. Here, $C^m \subseteq C$ and $q \leq j$ where C is the set of all biomedical semantic concepts $C = \{c_1, c_2, \ldots, c_j\}$ in the dataset.

In this section, we first present the proposed approach for modality identification. Then, we present the proposed approach for concept detection. Lastly, an approach to handle underrepresented sparse concepts is elucidated.

A. RADIOLOGICAL IMAGE MODALITY IDENTIFICATION

Radiological image modality identification involves the identification of the modality of imaging used for a given radiological image [13]. This proposed approach is illustrated in Fig. 2. Radiological image modality identification involves first extracting suitable features from radiological images and then using a suitable classifier to identify the modality. A radiological image I is classified to belong to one of the s radiological image modalities m, where $m \in M = \{M_1, \ldots, M_s\}$. As illustrated in Fig. 2, we propose to represent a radiological image using deep features extracted using convolutional neural networks (CNNs) as they were found to be effective in [13]. When an image is passed through a CNN, it undergoes a series of transformations at each layer of the CNN. Therefore, an output of any layer of the CNN can be used to represent the image [14].

We propose to consider support vector machine (SVM) based classifiers for modality identification as shown in Fig. 2. An SVM is a binary classifier that involves transforming the data to a higher dimension space where two classes can be easily separated by a linear boundary and builds a classifier [57]. It determines a hyperplane having the maximum margin that separates data of a pair of classes. A binary classifier can be used for multi-class classification using either the "one-versus-rest" approach or the "one-versus-one" approach. In this work, we consider the one-versus-rest strategy. It is important to use suitable kernels for the effectiveness of SVM-based classifiers. In this work, we consider two kernels, namely a linear kernel and a Tanimoto kernel [58], [59].

Consider two images A and B where $\mathbf{a} = [a_1\ a_2\ \ldots\ a_n]^T$ and $\mathbf{b} = [b_1\ b_2\ \ldots\ b_n]^T$ are the d-dimensional deep features extracted after passing images A and B through the CNN. A linear kernel between images A and B is computed as:

$$\mathrm{K}_{\mathrm{LIN}}(\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b} \tag{1}$$

The Tanimoto kernel [59] is found to be effective while performing medical image modality classification in [13]. It considers the visual features as molecules and compares them. The Tanimoto kernel between images A and B is computed as:

$$\mathrm{K}_{\mathrm{TAN}}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - \mathbf{a} \cdot \mathbf{b}} \tag{2}$$

Once the modality of a radiological image is identified, we propose to find the concepts associated with it. In the next section, we present the proposed approach for concept detection corresponding to a radiological image belonging to a modality.

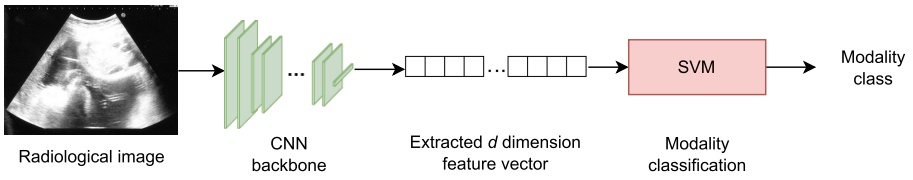

FIGURE 2. Proposed approach to identify the modality of a radiological image.

B. CONCEPT DETECTION FOR A RADIOLOGICAL IMAGE BELONGING TO A MODALITY

For a radiological image belonging to a modality m, we propose to consider the q biomedical semantic concepts associated with modality m, $C^m = \{c_1^m, c_2^m, \ldots, c_q^m\}$. However, it is also necessary to decide a subset of t biomedical semantic concepts $C_I^m = \{c_1, c_2, \ldots, c_t\}$ where $C_I^m \subseteq C^m$ that are prominent and most suitable to the image. To carry out this we need to use a suitable method. The detailed algorithm for this is given in Algorithm 1. This proposed approach to concept detection is illustrated in Fig. 3. This involves first representing a radiological image extracting suitable features and then building classifiers to detect concepts. Here, we propose to extract deep features for a radiological image from a chosen layer of a convolutional neural network (CNN). A radiological image I belonging to modality m is passed through a CNN. The d-dimensional output of one of the penultimate layers of the CNN is considered as the feature vector representation $\mathbf{i}$ of the input image I.

Algorithm 1 Concept Detection Using Proposed Hierarchical Approach

Input: Image I with modality m
Output: r concepts $\tilde{C}_I^m = \{\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_r\}$ associated with image I

$\mathbf{i} \leftarrow d$-dimensional output of an intermediate layer of the CNN, given I as input
$\tilde{C}_I^m = \phi$
Identify the q concepts associated with images belonging to m as $C^m = \{c_1^m, c_2^m, \ldots, c_q^m\}$
Generate $\lceil \frac{q}{k} \rceil$ random concept sets from C^m as R_l where $l = 1, \ldots, \lceil \frac{q}{k} \rceil$ of size q, such that $\bigcap_{l=1}^{\lceil \frac{q}{k} \rceil} R_l = \phi$
for each R_l **do**
    Generate 2^k possible combinations of the k concepts in R_l
    Perform classification with a base classifier using 2^k classes generated in the previous step, to find the most probable combination of concepts as $\hat{C}_l$
    for each concept c in the combination $\hat{C}_l$ **do**
        $\tilde{C}_I^m = \tilde{C}_I^m \cup \{c\}$
    end for
end for

An important issue for the concept detection method proposed in this section is underrepresented sparse concepts.

TABLE 2. Frequency of images containing concepts in the ImageCLEF 2020 concept detection training dataset. Here, 1624 biomedical semantic concepts belonging to the DRAN modality have 1-10 images associated with them.

Frequency	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR	Common
1-10	1624	1561	573	943	1325	1300	666	0
11-20	450	80	535	614	79	507	505	0
21-30	149	19	362	281	17	301	369	0
31-50	144	8	420	432	60	298	369	296
51-100	117	4	550	459	6	242	580	1007
101-200	58	3	351	199	3	148	306	691
201-500	35	0	210	51	0	55	176	734
501-1000	0	0	10	0	1	25	14	297
1001-20000	1	0	0	1	0	1	1	20
>20000	0	0	2	0	0	0	0	2
Total concepts	2578	1675	3013	2980	1491	2877	2986	3047
Total images	4713	487	20031	11447	502	8629	18944	64753

A biomedical semantic concept is said to be an underrepresented sparse concept if only a few images in the training data contain that concept [60]. Table 2 shows the frequency of images containing the biomedical semantic concepts for the ImageCLEF 2020 concept detection training dataset. The frequency column indicates the number of images in which a particular biomedical semantic concept occurs. The next seven columns indicate the number of concepts having the frequency mentioned in the first column for each modality. For example, in the first row, 1624 concepts occur in less than or equal to 10 angiographic images (DRAN) while 1561 out of 1675 concepts occur in less than 11 combined modality (DRCO) images. The last column shows the number of concepts having the frequency mentioned in the first column when the modality of the images is not considered. When the modality of the radiological images is not considered, i.e. if we do not classify the images as per their modalities and consider the entire training dataset as a whole (i.e. 64753 images), it is seen that 1303 out of 3047 biomedical semantic concepts, or rather 43% of the total concepts occur in less than 100 out of 64753 training images. Such few training samples are not sufficient to train a deep neural network to make accurate predictions for underrepresented sparse concepts. Even for the training images belonging to a single modality, it is seen that most of the modalities have less than 10 training images per concept. An approach that specifically handles underrepresented sparse concepts is needed.

The most common approach to handling unbalanced datasets with underrepresented sparse concepts is the

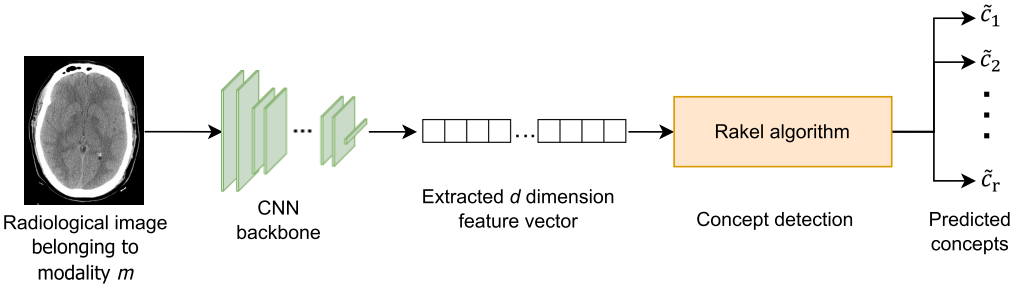

FIGURE 3. Proposed approach to determine the concepts associated with a radiological image belonging to modality m.

augmentation of the training data [49]. However, this may not work when there is a paucity of time and computational resources. For multi-label classification, methods such as binary relevance (BR) [61], and label powerset (LP) [62] are popular. Given q concepts, the BR method creates q binary classifiers where a classifier for concept c considers all training examples annotated with concept c as positive examples, and all other training examples that are not annotated with concept c as negative examples. Using this, binary classification is performed and only those test examples that are predicted as positive are annotated with the concept c. This method, however, does not consider the correlation of concepts in the images [63]. Also, when the dataset contains underrepresented sparse concepts, the number of negative training examples is much more than the number of positive training examples. This introduces significant bias in the model. The LP method transforms a multi-label classification problem into a multi-class classification problem. In a multi-label classification problem, a given example can be annotated with one or more concepts from a set of q concepts where $q \geq 2$. In a multi-class classification problem, a given example is annotated with only one concept from a set of q concepts where $q > 2$. The LP method considers every possible combination of concepts in the training data as a single class and performs multi-class classification. This means that if the training data contains q concepts, the multi-class classification problem will have 2^q classes and the classifier will predict only one out of the 2^q classes. The combination of concepts associated with that class will indicate the predictions for the example. This is highly computationally intensive [63]. In our work, we propose to use the disjoint Rakel algorithm to perform multi-label classification [6]. The Rakel algorithm considers k out of q concepts at a time to form $\lceil \frac{q}{k} \rceil$ concept sets R_l, where $l = 1, \ldots, \lceil \frac{q}{k} \rceil$. If $\lceil \frac{q}{k} \rceil$ is not an integer, then the $R^{th}_{\lceil \frac{q}{k} \rceil}$ concept set will contain less than k concepts. As shown in Algorithm 1, a separate classifier is trained for each of the R_l concept sets using the LP method. For each classifier, 2^k classes are generated using all the possible combinations of the k concepts present in the concept set. Then for each concept set R_l, a base classifier is trained to predict one out of the 2^k outputs, that contains the most probable combination of concepts $\hat{C}_l$ relevant to the input image. In this way, $\lceil \frac{q}{k} \rceil$ random disjoint sets of concepts are generated, each containing k concepts, and the LP method of multi-label classification is performed on each concept set. The results of the individual classification tasks are merged to predict the most probable concepts $\tilde{C}^m_I$ relevant to image I across all the concepts in C_m. Although the grouping of the concepts is done using the disjoint Rakel algorithm, the method of classification depends on the base classifier. In [6], Tsoumakas et. al have suggested that the ideal value of k is 3. Therefore, we have initialized $k = 3$ for our experiments. In this work, we have performed classification using the Rakel algorithm with the following base classifiers:

1) Support vector machines (SVM): The SVM classifier uses a maximum margin hyperplane to discriminate the training examples belonging to different classes [57]. For each of the $\lceil \frac{q}{k} \rceil$ concept sets, a separate Tanimoto kernel-based SVM classifier is trained to predict one out of the 2^k combination generated from the k concepts using a one-versus-rest strategy as $\hat{C}_l$. The results obtained across all concept sets R_l are combined to produce $\tilde{C}^m_I$.
2) Gaussian naive Bayes classifier: A naive Bayes classifier is trained for each concept set R_l. It uses the Bayes theorem to calculate the probability of an example containing each of the 2^k combinations of the k concepts in R_l. The combination of concepts with the highest probability is determined as $\hat{C}_l$. The concepts occurring in this most probable combination are assigned to the example. In a Gaussian naive Bayes classifier, the features are assumed to be independent and the data is assumed to follow a Gaussian distribution [64].
3) Logistic regression classifier: The logistic regression classifier uses an "S" shaped sigmoid function fitted to the input data and predicts the probability of a test example belonging to a particular class [65]. The class with the highest probability is assigned to the test example. When used as a base classifier for the disjoint Rakel algorithm, there are 2^k classes or combinations formed out of k concepts. The logistic regression classifier performs this classification $\lceil \frac{q}{k} \rceil$ times for each of the R_l concept sets. For each classification,

TABLE 3. Details of the dataset used for radiological image concept detection in the ImageCLEF concept detection challenges.

Year of challenge	No. of training images	No. of validation images	No. of test images	Total no. of images
2020 [19]	64753	15970	3534	84257
2021 [67]	2756	500	444	3700
2022 [68]	83275	7645	7645	98565

the most probable combination $\hat{C}_l$ is determined and assigned to the test example. The results of each of the individual classification tasks are combined to predict the association of all the q concepts with the test example.

4) Random forest classifier: The q concepts are divided into $\lceil \frac{q}{k} \rceil$ concept sets. In a concept set R_l, 2^k combinations of the q concepts present in R_l are created. These are the classes used to classify the example using an ensemble of decision trees. A majority vote is taken to predict the class to be assigned to the test example [66]. The results obtained for each of the $\lceil \frac{q}{k} \rceil$ classification tasks are combined to give the final result $\tilde{C}_I^m$.

IV. EXPERIMENTAL STUDIES

In this section, the studies performed for the biomedical semantic concept detection task are presented. The details of the datasets used for the experiments are provided first. Thereafter, the studies conducted are discussed.

A. DATASETS

The medical images used for studies in this work belong to the ImageCLEF 2020, 2021, and 2022 concept detection challenges [19], [67], [68]. ImageCLEF organizes a concept detection challenge every year since 2017 [18], [18], [19], [53], [67], [68]. After the initial challenges, it was decided that the number of modalities in the dataset should be reduced to improve the performance of the proposed approaches [53]. The later challenges focus on the concept detection task for radiological images only. The images are taken from the Radiology Objects in COntext (ROCO) dataset which are obtained from various biomedical articles and literature in the PubMed repository [70]. The details of the ImageCLEF concept detection challenge datasets for the years 2020, 2021, and 2022 are provided in Table 3. Since the concepts associated with the test images of the ImageCLEF 2022 concept detection are not available, the training and validation datasets were combined and 20% of this combined dataset was randomly chosen for testing while the remaining 80% of the images were kept for training purposes. Therefore, we used 72733 training images and 18187 test images while performing experiments on the ImageCLEF 2022 concept detection dataset as mentioned in Table 8.

Each image in these datasets has a set of biomedical semantic concepts associated with it. The ImageCLEF challenge dataset for the year 2020 also has the modality label for every radiological image. The details of these modalities

TABLE 4. Details of the dataset used for the ImageCLEF 2020 medical image concept detection challenge [19].

Modality	No. of training images	No. of validation images	No. of test images
DRAN	4713	1132	325
DRCO	487	73	49
DRCT	20031	4992	1140
DRMR	11447	2848	562
DRPE	502	74	38
DRUS	8629	2134	502
DRXR	18944	4717	918

for the ImageCLEF 2020 challenge dataset are given in Table 4. As given in Table 4, the ImageCLEF 2020 dataset consists of 84,257 radiographic images belonging to 7 classes with 6 modalities and 1 class of multiple modalities in a single image. Here, images having multiple modalities on a single image correspond to compound images that comprise multiple sub-images coming from different modalities. As seen in Table 3, the ImageCLEF 2021 dataset was much smaller in size with only 3,700 radiographic images while the dataset used for the 2022 ImageCLEF concept detection challenge had 98,565 radiographic images. One thing to note is that both the ImageCLEF 2021 and 2022 concept detection datasets did not have images annotated with their modalities. We propose to build modality classifiers using the ImageCLEF 2020 dataset and use those classifiers to identify the modalities of images in the ImageCLEF 2021 and 2022 datasets. In the next section, we present the studies on modality identification.

B. STUDIES ON MODALITY IDENTIFICATION FOR RADIOLOGICAL IMAGES

We propose to build a modality identifier using the ImageCLEF 2020 dataset that is comprised of radiological images from seven modalities as given in Table 4. Every image is represented using deep features that are extracted from the penultimate layers of a pre-trained convolutional neural network (CNN). In our proposed work, we use five pre-trained CNNs. The first four CNNs are ResNet50 [71], NASNet-Mobile [72], VGG16 [73], and EfficientNetV2L [74]. These are pre-trained on natural images belonging to the ImageNet dataset [15]. To observe the performance of features extracted from a CNN that is pre-trained on radiographic images, we have also used CheXNet [75], a 121-layer DenseNet CNN that is pre-trained on a chest X-ray dataset [76]. The outputs of the penultimate layers of these five CNNs are extracted, and the details of the dimensions are presented in Table 5. Let every radiological image be represented by a d-dimension feature vector that is extracted from a chosen layer of a CNN. To identify the modality of a radiological image, it is necessary to build a suitable classifier. In this work, we propose to build support vector machine (SVM)-based classifiers [57]. For SVM-based classifiers to be effective, it is important to use suitable kernel functions. In this work, we propose to explore the linear kernel and Tanimoto

TABLE 5. Details of the feature vectors obtained after feature extraction from the layers of the CNNs.

CNN	Name of Layer	Dimensions
ResNet50	average pooling after last convolution layer	2048
NASNetMobile	average pooling after last activation layer	1056
VGG16	second fully connected layer	4096
EfficientNetV2L	average pooling after last activation layer	1280
CheXNet	average pooling after last activation layer	1024

TABLE 6. Accuracy (in %) of radiological image modality detection for the ImageCLEF 2020 concept detection dataset.

CNN used for feature extraction	Linear kernel-based SVM	Tanimoto kernel-based SVM
ResNet50	93.04	**94.59**
NASNetMobile	91.31	91.91
VGG16	90.78	92.87
EfficientNetV2L	91.00	92.67
CheXNet	84.41	81.78

kernel for building SVM-based classifiers. The binary classification strategy of SVM is extended to multi-class modality classification using the one-versus-rest strategy. The performance of the modality classifier is measured using accuracy as a metric. The accuracy of a classifier is measured as given in Equation 3 [77]. The performance of the proposed SVM-based modality identification for the ImageCLEF 2020 concept detection dataset is given in Table 6. In Fig. 4 and Fig. 5, the confusion matrices for each of the proposed approaches to radiological image modality identification are given.

$$accuracy = \frac{no.\ of\ images\ whose\ modality\ is\ correctly\ predicted}{total\ no.\ of\ test\ images\ in\ the\ dataset} \tag{3}$$

It is seen from Table 6, Fig. 4, and Fig. 5 that SVM-based classifiers using image representations extracted from the ResNet50 CNN perform better than those using the representation extracted from the DenseNet-121 CNN. It is also observed that the Tanimoto kernel-based SVM classifier performs better than the linear kernel-based SVM. Overall, it is seen from Table 6 that the proposed SVM-based modality classifier that uses the Tanimoto kernel on features extracted from the ResNet50 CNN gives the best performance. The proposed approach to radiological image modality identification is compared with existing approaches that use the same test data in Table 7. The approach that obtained an accuracy of 62.08% proposed to perform k-nearest neighbour classification of features extracted from a ResNet18 CNN [19]. From Tables 6 and 7, it can be seen that the proposed approach to perform radiological image modality classification gives better performance than other existing approaches.

TABLE 7. Classification accuracy (in %) of radiological modality detection for the ImageCLEF 2020 concept detection dataset.

Approach Used	Accuracy
k-NN on ResNet18 embeddings [42]	62.08
Feedforward neural network over variation of DenseNet-121 and CheXNet [17]	59.73
Fine-tuning of DenseNet-121 [45]	56.34
Fine-tuning of Xception CNN [78]	50.08
Multi-label classification using ResNet50 [79]	47.06
Fine-tuning DenseNet-169 using concept groups [80]	2.06
Multi-label classification using a CNN [81]	1.39
Proposed approach using a Tanimoto kernel-based SVM classification on features extracted from a ResNet50 CNN	**94.59**

TABLE 8. Distribution of images across radiological modalities for the ImageCLEF 2021 and 2022 concept detection datasets.

Modality	No. of training images in the 2021 ImageCLEF dataset	No. of test images in the 2021 ImageCLEF dataset	No. of training images in the 2022 ImageCLEF dataset	No. of test images in the 2022 ImageCLEF dataset
DRAN	88	17	5264	1317
DRCO	4	1	606	152
DRCT	1087	157	23079	5770
DRMR	660	89	12484	3122
DRPE	8	1	466	117
DRUS	271	39	9818	2455
DRXR	638	140	21016	5254
Total	2756	444	72733	18187

It is also seen that SVM-based modality classifiers built using features extracted from ResNet50 outperform the SVM-based modality classifiers that used features extracted from other CNNs. Features extracted from CheXNet, which is the DenseNet-121 CNN pre-trained on radiological images, obtain the lowest accuracy for the modality classification task. This leads to the observation that pre-training the CNN model on radiographic images does not provide any added advantage in terms of classification accuracy as compared to a CNN model pre-trained on natural images. We propose to use the Tanimoto kernel-based SVM classifier on deep features extracted from the ResNet50 CNN for the identification of the modality of a radiological image.

The images in the 2021 and 2022 ImageCLEF concept detection datasets are not annotated with their respective modalities. Therefore, the Tanimoto kernel-based SVM classifier trained on the features extracted from the average pooling layer after the last convolution layer of the pre-trained ResNet50 CNN is used to classify the training and test images of the 2021 and 2022 ImageCLEF concept detection datasets. The distribution of images across the radiological modalities after performing modality identification using the proposed approach for both datasets is given in Table 8. It is observed in Tables 4 and 8 that the distribution of radiological modalities for both these datasets is similar to the distribution of radiological modalities in the ImageCLEF 2020 concept detection dataset with CT scans being the most common modality across all three datasets. In the next section, we present the studies on modality-specific concept detection.

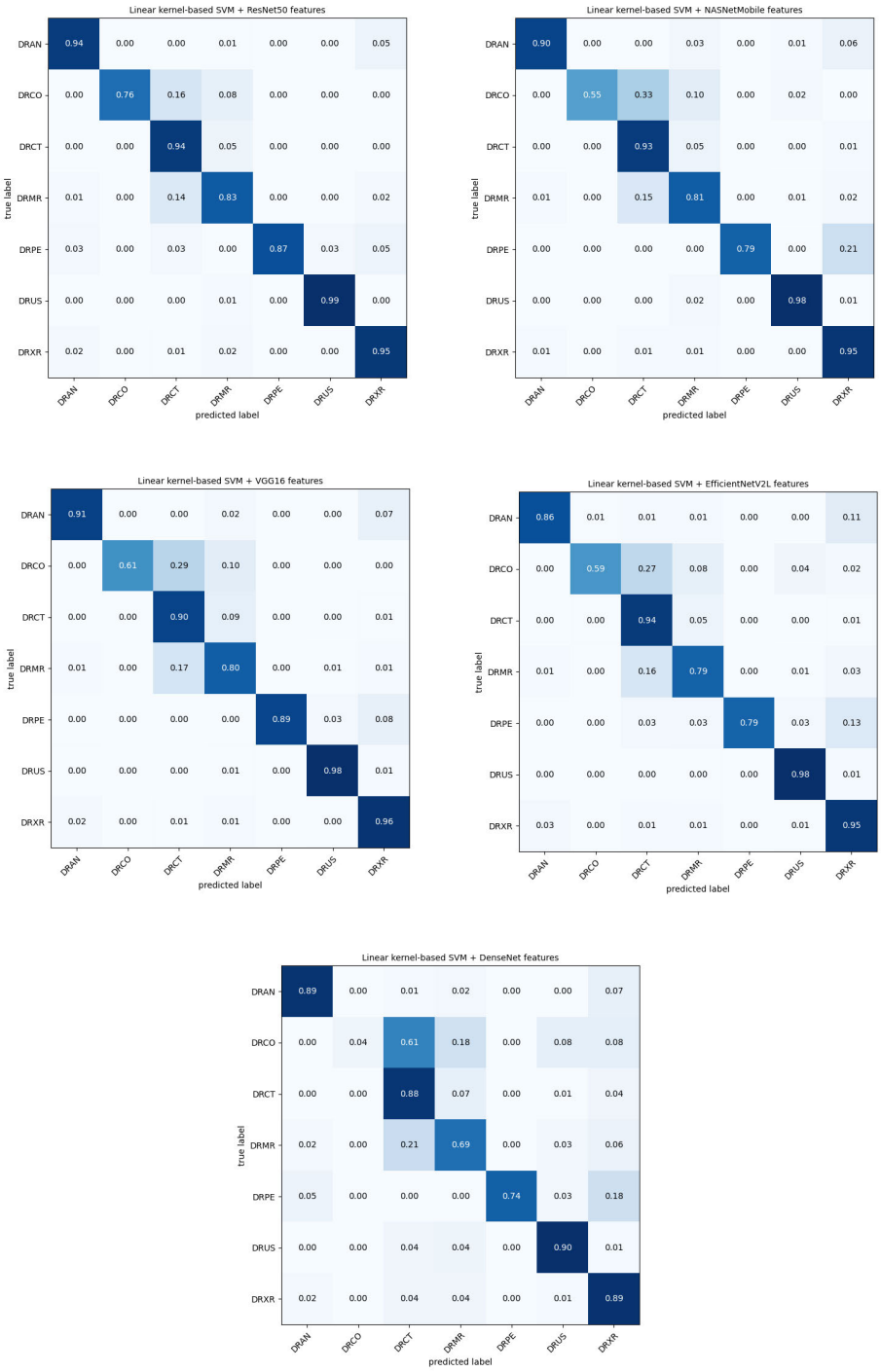

FIGURE 4. Confusion matrices for the approaches proposed for radiological modality identification using linear kernel-based SVM classification performed on the radiological images of the ImageCLEF 2020 concept detection dataset.

C. STUDIES ON MODALITY SPECIFIC CONCEPT DETECTION

In this section, we present the studies on the proposed approach to concept detection specific to a modality. Once the modality of the radiological image is identified, we build multi-label classifiers to detect modality-specific concepts. For this classification task, we have considered features extracted from four different CNNs for semantic concept detection, namely, ResNet50 [71], NASNetMobile [72], VGG16 [73], and EfficientNetV2L [74]. Since the features extracted from the CheXNet CNN did not obtain a high accuracy for the modality identification task as compared

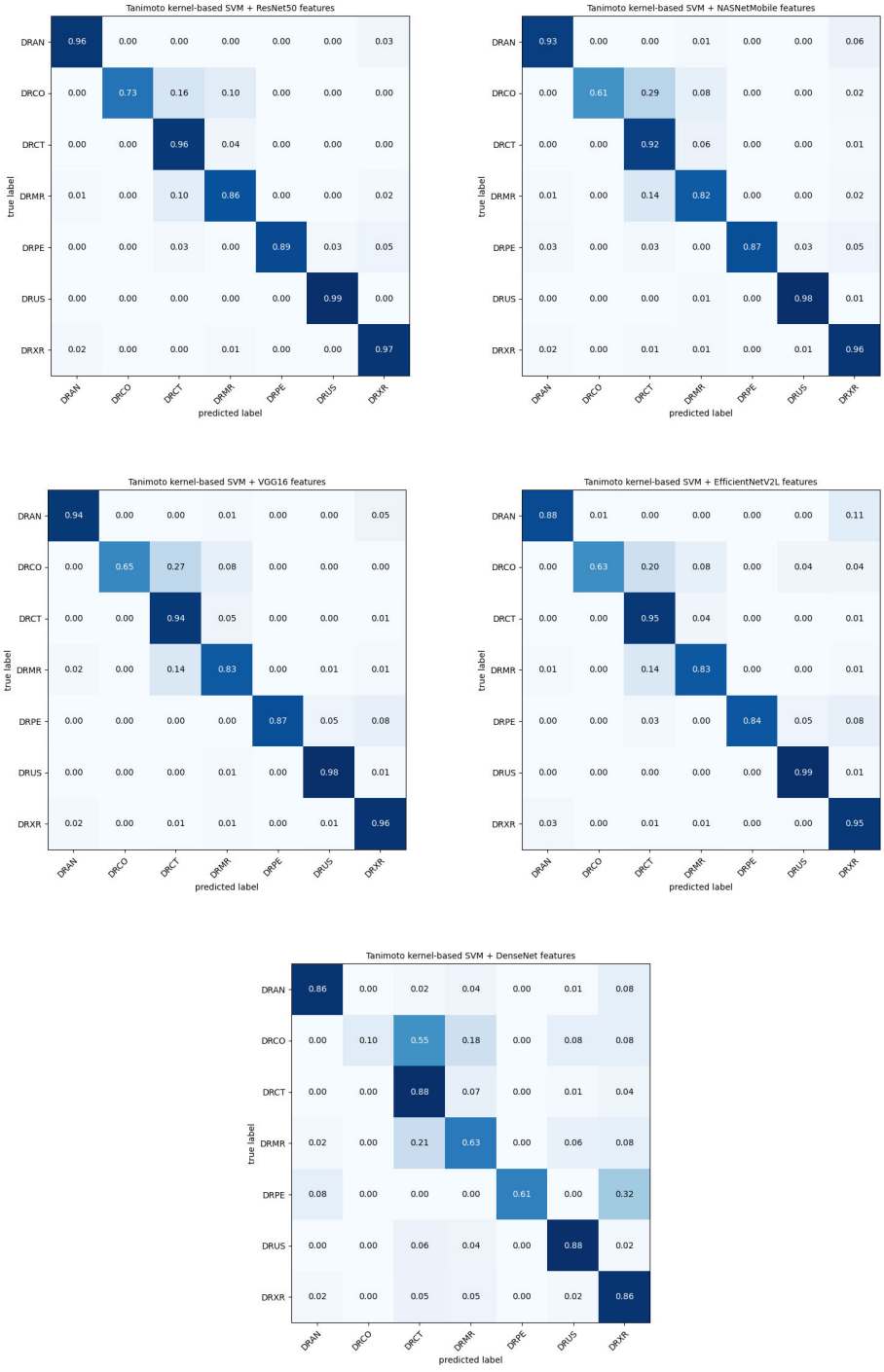

FIGURE 5. Confusion matrices for the approaches proposed for radiological modality identification using Tanimoto kernel-based SVM classification performed on the radiological images of the ImageCLEF 2020 concept detection dataset.

to features extracted from the other four CNNs, we did not consider them to represent features for modality-specific concept detection. To evaluate the performance of the proposed approach for modality-specific concept detection, we consider precision, recall, and F_1-score as the metrics.

For a given concept $c \in C$, precision as shown in Equation 4, indicates how many images that are predicted to contain concept c have c in their ground truth. Recall as given in Equation 5, for a concept c determines the number of images that are correctly predicted to contain concept c in their ground truth. The F_1-score is computed for each unique

TABLE 9. Mean F_1-scores for the modality-specific concept detection task performed on the ImageCLEF 2020 concept detection dataset.

Tanimoto kernel based SVM							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.2921	0.0096	0.4171	0.3031	0.2242	0.2557	0.3089
NASNetMobile	0.2878	0.0041	0.4154	0.3017	0.2192	0.2514	0.3019
VGG16	0.2879	0.0177	0.4189	0.3017	0.2268	0.2622	0.3079
EfficientNetV2L	0.2863	0.0119	0.4202	0.2991	0.2183	0.2582	0.3119
Gaussian Naïve Bayes Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.1101	0.0897	0.0591	0.0741	0.2704	0.0899	0.0562
NASNetMobile	0.1001	0.0461	0.0467	0.0562	0.2423	0.0767	0.0458
VGG16	0.1177	0.0503	0.0717	0.0816	0.2533	0.0982	0.0579
EfficientNetV2L	0.0953	0.0513	0.0454	0.0566	0.2688	0.0734	0.0514
Logistic Regression Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.2419	0.0658	0.3262	0.2613	0.2585	0.2489	0.2386
NASNetMobile	0.2331	0.0439	0.3501	0.2346	0.2515	0.2160	0.2451
VGG16	0.2513	0.0437	0.3543	0.2793	0.2383	0.2479	0.2640
EfficientNetV2L	0.2771	0.0599	0.3975	0.2805	0.2754	0.2555	0.2904
Random Forest Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.2809	0	0.4098	0.2938	0.2171	0.2478	0.2940
NASNetMobile	0.2809	0.0021	0.4098	0.2938	0.2078	0.2481	0.2934
VGG16	0.2809	0.0181	0.4098	0.2938	0.2192	0.2481	0.2938
EfficientNetV2L	0.2822	0.0008	0.4098	0.2938	0.2088	0.2483	0.2934

biomedical semantic concept $c \in C$ found in the training data as provided in Equation 6 [82]. It is the harmonic mean of the precision and recall values for a concept c. The mean F_1-score is the average F_1-score across all $c \in C$. To evaluate the performance of the proposed approach, the mean F_1-score performance metric is used as shown in Equation 7.

$$precision_c = \frac{no.\ of\ images\ correctly\ associated\ with\ concept\ c}{no.\ of\ images\ predicted\ to\ contain\ concept\ c} \quad (4)$$

$$recall_c = \frac{no.\ of\ images\ correctly\ associated\ with\ concept\ c}{no.\ of\ images\ that\ actually\ contain\ the\ concept\ c} \quad (5)$$

$$F_1 - score_c = \frac{2 \times precision_c \times recall_c}{precision_c + recall_c} \quad (6)$$

$$mean\ F_1 - score = \frac{\sum_{i=1}^{j} F_1 - score_{c_i}}{j} \quad (7)$$

We now present the studies conducted for the concept detection task on the ImageCLEF 2020, 2021, and 2022 datasets respectively.

1) STUDY ON MODALITY-SPECIFIC CONCEPT DETECTION TASK FOR RADIOLOGICAL IMAGES OF THE IMAGECLEF 2020 DATASET

The mean F_1-scores obtained for the modality-specific concept detection task performed on the ImageCLEF 2020 concept detection dataset are presented in Table 9. The corresponding box plot is given in Fig. 6. From Table 9 and the box plot in Fig. 6, it is evident that the Tanimoto kernel-based SVM classifier outperforms the other three approaches. The Gaussian naive Bayes classifier performs the worst among the four classification methods. It is seen that the mean F_1-score for the DRCT modality (computed tomography) is the highest among all the individual modalities. The second highest mean F_1-score is obtained for the DRXR (X-ray scan) modality. In the training dataset of the ImageCLEF 2020 concept detection challenge, 31% of the images belong to the DRCT modality while 21% of the images belong to the DRXR modality. The mean F_1-scores for the remaining modalities follow the pattern of being proportional to the number of training images in the dataset. The only exceptions to this are the DRAN (angiogram) and DRUS (ultrasound) modalities. The number of training examples from the ultrasound modality is almost twice that of the angiography samples. Despite this, the concept detection task has a higher mean F_1-score for DRAN as compared to DRUS. If the mean F_1-score obtained for DRCO (combined modality) is compared with that obtained for DRPE (PET scan), it is observed that the number of training image to test image ratio is higher for the DRCO modality as compared to the DRPE modality. The DRCO modality has 1673 unique concepts while the DRPE modality has 1491 unique biomedical concepts. The highest mean F_1-score for the DRPE modality is 0.2268, while the highest mean F_1-score for the DRCO modality is only 0.0119. This indicates that detecting semantic concepts in images that contain multiple radiological modalities is more difficult compared to images containing a single modality. For this dataset, concept detection performed on features extracted from the VGG16 CNN gave the best mean F_1-scores.

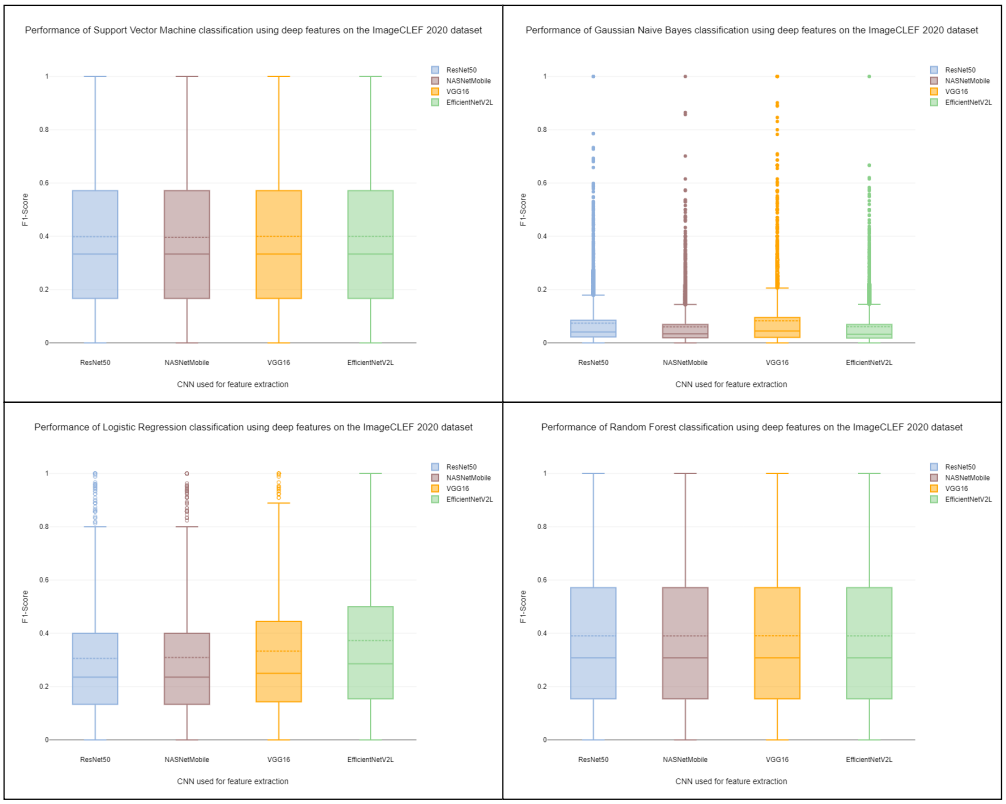

FIGURE 6. Box plots corresponding to the concept detection task performed on the ImageCLEF 2020 concept detection dataset.

2) STUDY ON MODALITY-SPECIFIC CONCEPT DETECTION TASK FOR RADIOLOGICAL IMAGES OF THE IMAGECLEF 2021 DATASET

The mean F_1-scores obtained for the modality-specific concept detection task performed on the ImageCLEF 2021 concept detection dataset are presented in Table 10. The corresponding box plot is given in Fig. 7. The ImageCLEF 2021 dataset for the concept detection challenge is the smallest dataset with only 3700 images in total as compared to the 2020 and 2022 datasets. The mean F_1-scores of the individual modalities as seen in Table 10 and Fig. 7 are much higher when compared to those obtained for the other two datasets, with DRAN (angiogram) images obtaining the highest mean F_1-score of 0.6634 for the concept detection task, despite having only 88 training images and 17 test images. The second highest mean F_1-score is observed for the DRUS (ultrasound) modality which contains only 271 training images and 39 test images. Although the DRCT (computed tomography) modality had the most training and test images, it obtained the fifth highest mean F_1-score of only 0.5720 when compared to the mean F_1-scores obtained by the other modalities. There are only 8 DRPE (PET scan) training images and 1 test image. However, the DRPE modality achieved a maximum mean F_1-score of 1. On the other hand, for the DRCO category (combined modalities), the proposed approach is unable to predict the occurrence of even a single semantic concept correctly. It is also seen in Fig. 7 that the performance of the Gaussian naive Bayes classifier is the worst as compared to the other classification techniques. In Table 10, it is seen that the logistic regression classifier performs significantly better than the random forest classifier for modalities like DRAN and DRCT. The performance of the logistic regression classifier is marginally better than the Tanimoto-based kernel SVM for the DRCT and DRUS modalities. However, when we consider the overall performance as shown in Table 12, the SVM classifier outperforms all the other methods.

3) STUDY ON MODALITY-SPECIFIC CONCEPT DETECTION TASK FOR RADIOLOGICAL IMAGES OF THE IMAGECLEF 2022 DATASET

The mean F_1-scores obtained for modality-specific concept detection task performed on the ImageClEF 2022 concept detection dataset are presented in Table 11. The corresponding box plot is given in Fig. 8. As it is seen in Table 11 and Fig. 8, except for the DRCO (combined modality) category, all other modalities obtained mean F_1-scores greater than 0.4. DRXR (X-ray scan) images achieved the highest mean F_1-score of 0.5621 which is much higher than the second highest mean F_1-score obtained by the ultrasound modality. The performance for the DRCT (computed tomography) modality is ranked fifth with a mean

TABLE 10. Mean F_1-scores for the modality-specific concept detection task performed on the ImageCLEF 2021 concept detection dataset.

Tanimoto kernel based SVM							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.6634	0	0.5595	0.5633	0.4	0.6321	0.5491
NASNetMobile	0.6271	0	0.5242	0.5243	0.4	0.5728	0.5338
VGG16	0.5699	0	0.5588	0.5308	0.4	0.6066	0.5496
EfficientNetV2L	0.5653	0	0.5709	0.5581	0.4	0.6222	0.5377
Gaussian Naïve Bayes Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.5193	0	0.4072	0.4330	0.4	0.5383	0.4414
NASNetMobile	0.5275	0	0.3699	0.4013	0.4	0.5579	0.4272
VGG16	0.4490	0	0.4242	0.3722	0.4	0.5830	0.4421
EfficientNetV2L	0.4654	0	0.3437	0.3757	0.4	0.5717	0.4176
Logistic Regression Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.6222	0	0.5720	0.5375	0.4	0.6106	0.5531
NASNetMobile	0.6049	0	0.5138	0.4707	0.4	0.5945	0.5329
VGG16	0.5978	0	0.5564	0.5170	1	0.6605	0.5404
EfficientNetV2L	0.5155	0	0.5305	0.5219	0.4	0.6427	0.5328
Random Forest Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.5078	0	0.4621	0.4769	0.4	0.5812	0.5334
NASNetMobile	0.4490	0	0.4581	0.4680	0.4	0.5812	0.5329
VGG16	0.4490	0	0.4629	0.4708	0.4	0.5897	0.5245
EfficientNetV2L	0.5471	0	0.4568	0.4708	0.4	0.5897	0.5194

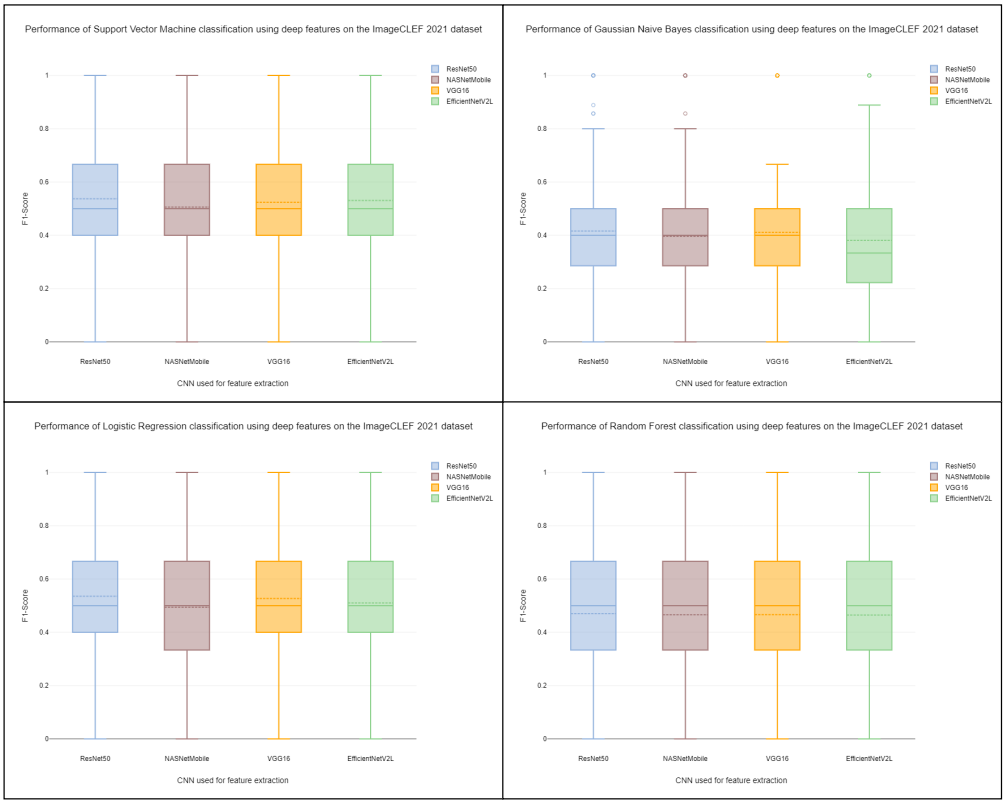

FIGURE 7. Box plots corresponding to the concept detection task performed on the ImageCLEF 2021 concept detection dataset.

F_1-score of 0.4106. However, there is not much difference between the performance for the DRCT modality when compared with the average performance obtained for the other individual modalities across the dataset. The mean F_1-score for the DRCO modality is 0.1952 which is the highest compared to the mean F_1-scores obtained for the DRCO modality in the other two datasets. However, there is a huge difference between the mean F_1-scores for the DRCO

TABLE 11. Mean F_1-scores for the modality-specific concept detection task performed on the ImageCLEF 2022 concept detection dataset.

Tanimoto kernel based SVM							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.4315	0	0.4081	0.4039	0.3923	0.4252	0.5513
NASNetMobile	0.4327	0.0795	0.4089	0.4050	0.3956	0.4297	0.5362
VGG16	0.4340	0.0994	0.4102	0.4064	0.3911	0.4346	0.5447
EfficientNetV2L	0.4352	0.0881	0.4106	0.4070	0.3920	0.4365	0.5497
Gaussian Naïve Bayes Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.1127	0.1295	0.0627	0.3536	0.0775	0.0707	0.0369
NASNetMobile	0.1003	0.1949	0.0392	0.0632	0.3581	0.0675	0.0590
VGG16	0.1367	0.1753	0.0562	0.0831	0.4104	0.1232	0.0956
EfficientNetV2L	0.0793	0.1736	0.0338	0.0508	0.3112	0.0632	0.0559
Logistic Regression Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.3866	0.1269	0.4079	0.4006	0.3606	0.4129	0.5489
NASNetMobile	0.4313	0.2086	0.4057	0.4012	0.4133	0.4236	0.5159
VGG16	0.4261	0.1668	0.4044	0.4021	0.3737	0.4227	0.5447
EfficientNetV2L	0.4322	0.1869	0.4063	0.4023	0.4201	0.4227	0.5287
Random Forest Classifier							
CNN	DRAN	DRCO	DRCT	DRMR	DRPE	DRUS	DRXR
ResNet50	0.4307	0.1857	0.4074	0.4033	0.3994	0.4244	0.5448
NASNetMobile	0.4319	0.2251	0.4069	0.4034	0.4173	0.4255	0.5293
VGG16	0.4319	0.2227	0.4067	0.4018	0.4098	0.4278	0.5387
EfficientNetV2L	0.4334	0.2089	0.4059	0.4010	0.4235	0.4255	0.5371

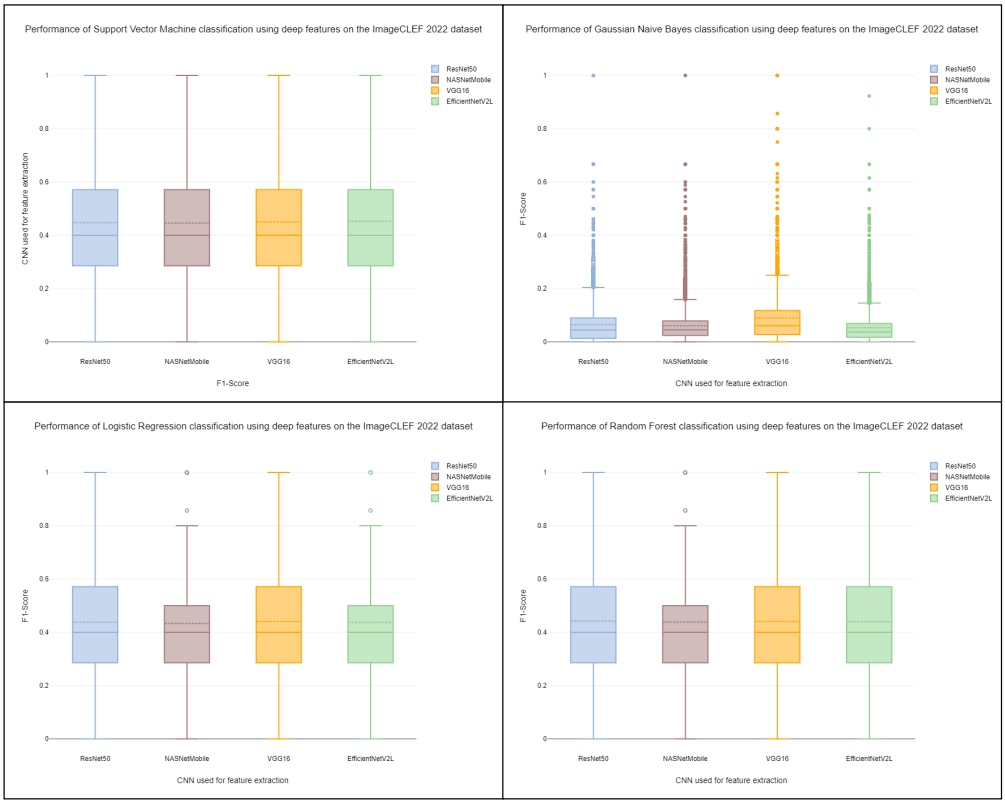

FIGURE 8. Box plots corresponding to the concept detection task performed on the ImageCLEF 2022 concept detection dataset.

modality and the other six modalities. It is seen in Table 11 that the Gaussian naive Bayes classifier exhibits the worst performance. The performance of the logistic regression classifier and the random forest classifier are similar. For this dataset too, the Tanimoto kernel-based SVM classification method outperforms the other methods.

The overall mean F_1-scores for the modalities across all three datasets show a similar trend. The results indicate the

TABLE 12. Comparison of the mean F_1-scores for the concept detection task using deep features extracted from different CNNs and multi-label classification performed using different base classifiers for the Rakel algorithm with state-of-the-art methods.

Approach	ImageCLEF 2020 dataset	ImageCLEF 2021 dataset	ImageCLEF 2022 dataset
ResNet50+SVM	0.3987	**0.5371**	0.4476
ResNet50+Gaussian Naïve Bayes	0.0739	0.4160	0.0643
ResNet50+Logistic Regression	0.3056	0.5355	0.4404
ResNet50+Random Forest	0.3904	0.4701	0.4423
NASNetMobile+SVM	0.3959	0.5059	0.4459
NASNetMobile+Gaussian Naïve Bayes	0.0601	0.3960	0.0602
NASNetMobile+Logistic Regression	0.3092	0.4940	0.4383
NASNetMobile+Random Forest	0.3903	0.4659	0.4333
VGG16+SVM	**0.3997**	0.5239	0.4503
VGG16+Gaussian Naïve Bayes	0.0828	0.4111	0.0895
VGG16+Logistic Regression	0.3332	0.5268	0.4404
VGG16+Random Forest	0.3908	0.4663	0.4409
EfficientNetV2L+SVM	0.3996	0.5307	**0.4528**
EfficientNetV2L+Gaussian Naïve Bayes	0.0606	0.3809	0.0529
EfficientNetV2L+Logistic Regression	0.3728	0.5102	0.4371
EfficientNetV2L+Random Forest	0.3903	0.4648	0.4399
state-of-the-art Method	0.3940 [17]	0.505 [31]	0.4511 [30]

relation between the number of training examples and the performance of the multi-label classifiers. Images belonging to modalities such as DRCT and DRXR that have more training examples have higher mean F_1-scores as compared to those obtained for modalities like DRCO and DRPE which have fewer training examples per concept. The mean F_1-scores for the DRCO modality are the lowest among all modalities across the three datasets. This indicates that the occurrence of multiple modalities within a single image makes it difficult to predict biomedical concepts for such images. A possible solution to this problem could be the splitting of every single compound image belonging to the DRCO modality into multiple sub-images where each sub-image belongs to a different modality [5]. The proposed hierarchical approach for concept detection could then be applied separately for each sub-image. When considering the feature extraction methods used in the proposed approach, deep features extracted from the EfficientNetV2L CNN give the best results for the concept detection task for the 2022 dataset which is the largest dataset. Deep features extracted from the ResNet50 CNN give better results for the 2021 dataset which is much smaller and consists only of 3700 images. When the overall performance of the proposed approach on the individual datasets is compared, it is seen that deep features extracted from the EfficientNetV2L CNN give the best results as compared to deep features extracted from the other CNNs. Among the classification techniques considered, the Tanimoto kernel-based SVM classifier outperforms all the other techniques for all three datasets. The Gaussian naive Bayes classifier obtains the lowest mean F_1-scores for all three datasets. The box plots corresponding to the Gaussian naive Bayes classifier given in Fig. 6, Fig. 7, and Fig. 8 show inconsistency in performance. This could probably be because the data does not follow a Gaussian distribution and the features are not independent of each other. The random forest classifier achieves a slightly better performance as compared to the logistic regression classifier for the ImageCLEF 2020 and 2022 concept detection datasets. This could be due to the ensemble nature of the random forest which takes a majority vote while predicting the class labels.

Since concept detection in radiological images involves underrepresented sparse concepts, we proposed to use the Rakel algorithm over the base classifiers to effectively detect the underrepresented sparse concepts. In Table 12, the proposed approach using deep features and Rakel multi-label classification is compared with the state-of-the-art methods. The proposed approach that uses a Tanimoto kernel-based classifier outperforms the state-of-the-art methods for all three datasets. All three state-of-the-art methods are proposed by researchers at the Athens University of Economics and Business. The approach in [17] uses a feedforward neural network that acts as a classification layer over CheXNet which is a variation of the DenseNet-121 CNN. The CNN is fine-tuned to the concept detection dataset and a separate model is trained for each modality. This approach produces a mean F_1-score of 0.3940 while our proposed approach obtains a mean F_1-score of 0.3997. Although the approach proposed by Karatzas et al. in [17] uses the modality of the image to narrow down the search space for concept detection, it does not consider any mechanism to handle the problem of sparse semantic concepts. All proposed approaches that used deep features with the Tanimoto kernel-based SVM classifier performed better than the approach proposed by [17]. This shows the effectiveness of using a suitable kernel-based SVM for classifying the data. The approach that gives the best results for the ImageCLEF 2021 concept detection dataset uses 4 different encoders. The first encoder is the state-of-the-art method for the ImageCLEF 2020 concept detection challenge. The other 3 encoders are ResNet-50, DenseNet-201, and EfficientNet-B0 which are pre-trained on ImageNet and fine-tuned on the 2021 concept detection training data. 5 models are trained for each encoder resulting in 20 models. Image embeddings are extracted from the last average pooling layer of each of these models and the nearest 20 training images are retrieved using the cosine

similarity between the image embeddings of the training images and the test image. Each concept, if present in the majority of the training images, is assigned to the test image [31]. This method obtains a mean F_1-score of 0.505 while our proposed approach produces a mean F_1-score of 0.537. The proposed approaches to use deep features with the SVM classifier and the logistic regression classifier perform better than the approach proposed by [31]. However, unlike the other two datasets, the random forest classifier performed worse than the logistic regression classifier for the ImageCLEF 2021 dataset. The state-of-the-art method proposed by Charalampakos et al. in [30] for the ImageCLEF 2022 concept detection dataset uses 2 instances of an EfficientNetV2-B0 CNN that are pre-trained on ImageNet. The image embeddings are extracted from the last convolutional layer followed by a generalized mean global pooling layer and a sigmoid activation layer that performs multi-label classification of the concepts in the training dataset. The union of the concepts obtained by both instances of the CNN is then assigned to the image. This approach achieves an F_1-score of 0.4511 while the proposed hierarchical multi-label classification approach achieves a mean F_1-score of 0.4528. Although the approach proposed in [17] considers the modality of the radiological image before performing concept detection, the other state-of-the-art methods do not consider the radiological image modality. From Table 12, it is evident that the proposed approach that performs hierarchical classification by identifying the modality of an image and then performing multi-label classification using the Rakel algorithm outperforms all the state-of-the-art methods.

V. CONCLUSION

In this paper, the heterogeneity of radiological images is exploited to identify biomedical semantic concepts in radiological images. In the proposed approach, the radiological modality of the input images is first determined by performing SVM-based classification using deep features extracted from the penultimate layers of CNNs. Deep features extracted from CNNs pre-trained on natural images perform better at the modality identification task as compared to deep features extracted from CNNs pre-trained on radiographic images as well as other state-of-the-art methods. For the concept detection task, we propose to classify deep features extracted from the penultimate layers of CNNs using the Rakel algorithm. The Rakel algorithm is used with base classifiers such as logistic regression classifier, random forest classifier, Gaussian naive Bayes classifier, and Tanimoto-kernel based SVM classifier. The mean F_1-scores obtained indicate that deep features extracted from the EfficientNetV2L CNN perform better at the multi-label classification task. The proposed approach that uses Taniomoto-kernel based SVM classifiers outperforms the state-of-the-art methods for the ImageCLEF 2020, 2021, and 2022 concept detection datasets.

This indicates that determining the modality of a radiological image before performing concept detection improves the performance of concept detection approaches. As part of our future work, we would like to explore other possible applications in radiological image processing that could be performed after detecting such semantic concepts. These could include caption prediction, diagnostic report generation, and image retrieval methods.

REFERENCES

[1] P. Shamna, V. K. Govindan, and K. A. A. Nazeer, "Content-based medical image retrieval by spatial matching of visual words," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 2, pp. 58–71, Feb. 2022, doi: 10.1016/j.jksuci.2018.10.002.
[2] Q. B. Carroll, *Radiography in the Digital Age: Physics-Exposure-Radiation Biology*. Springfield, IL, USA: Charles C. Thomas, 2018.
[3] W. E. Brant and C. A. Helms, *Fundamentals of Diagnostic Radiology*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.
[4] Y. Seo and K.-S. Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert Syst. Appl.*, vol. 116, pp. 328–339, Feb. 2019.
[5] D. Miranda, V. Thenkanidiyoor, and D. A. Dinesh, "Review on approaches to concept detection in medical images," *Biocybernetics Biomed. Eng.*, vol. 42, no. 2, pp. 453–462, Apr. 2022.
[6] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 7, pp. 1079–1089, Jul. 2011, doi: 10.1109/TKDE.2010.164. https://doi.org/10.1109/TKDE.2010.164
[7] E. Pinho and C. Costa, "Feature learning with adversarial networks for concept detection in medical images: UA.PT bioinformatics at ImageCLEF 2018," in *Proc. CLEF Work. Notes*, 2018.
[8] A. J. Gonçalves, E. Pinho, and C. Costa, "Informative and intriguing visual features: UA.PT bioinformatics in ImageCLEF caption 2019," in *Proc. CLEF Work. Notes*, 2019.
[9] L. Valavanis and S. Stathopoulos, "IPL at ImageCLEF 2017 concept detection task," in *Proc. CLEF Work. Notes*, 2017.
[10] L. Valavanis and T. Kalamboukis, "IPL at ImageCLEF 2018: A kNN-based concept detection approach," in *Proc. CLEF Work. Notes*, 2018.
[11] E. Pinho, J. F. Silva, J. M. Silva, and C. Costa, "Towards representation learning for biomedical concept detection in medical images: UA.PT bioinformatics in ImageCLEF 2017," in *Proc. CLEF Work. Notes*, 2017.
[12] S.-K. Lam et al., "Multi-organ omics-based prediction for adaptive radiation therapy eligibility in nasopharyngeal carcinoma patients undergoing concurrent chemoradiotherapy," *Frontiers Oncol.*, vol. 11, Jan. 2022, Art. no. 792024.
[13] D. Miranda, V. Thenkanidiyoor, and D. A. Dinesh, "Detecting the modality of a medical image using visual and textual features," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104035.
[14] M. I. Daoud, S. Abdel-Rahman, T. M. Bdair, M. S. Al-Najar, F. H. Al-Hawari, and R. Alazrai, "Breast tumor classification in ultrasound images using combined deep and handcrafted features," *Sensors*, vol. 20, no. 23, p. 6838, Nov. 2020.
[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[16] V. Kougia, J. Pavlopoulos, and I. Androutsopoulos, "AUEB NLP group at ImageCLEFmed caption 2019," in *Proc. CLEF Work. Notes*, 2019.
[17] B. Karatzas, J. Pavlopoulos, V. Kougia, and I. Androutsopoulos, "AUEB NLP group at ImageCLEFmed caption 2020," in *Proc. CLEF Work. Notes*, Thessaloniki, Greece, Sep. 2020.
[18] O. Pelka, C. M. Friedrich, A. G. S. de Herrera, and H. Müller, "Overview of the ImageCLEFmed 2019 concept detection task," in *Proc. CEUR Workshop*, 2019.
[19] O. Pelka, C. M. Friedrich, A. G. S. de Herrera, and H. Müller, "Overview of the ImageCLEFmed 2020 concept prediction task: Medical image understanding," in *Proc. CLEF Work. Notes, CEUR Workshop*. Thessaloniki, Greece: CEUR-WS.org, 2020.
[20] M. Tatusch, "Approaches for the improvement of the multilabel multiclass classification with a huge number of classes," in *Proc. Grundlagen von Datenbanken*, 2018, pp. 65–70.
[21] S. Singh, S. Karimi, K. Ho-Shon, and L. Hamey, "Biomedical concept detection in medical images: MQ-CSIRO at 2019 ImageCLEFmed caption task," in *Proc. CLEF Work. Notes*, 2019.

[22] K. Dimitris and K. Ergina, "Concept detection on medical images using deep residual learning network," in *Proc. CLEF Work. Notes*, 2017.

[23] P. Sinha, S. Purkayastha, and J. Gichoya, "Full training versus fine tuning for radiology images concept detection task for the ImageCLEF 2019 challenge," in *Proc. CLEF Work. Notes*, 2019.

[24] L.-D. Stefan, B. Ionescu, and H. Müller, "Generating captions for medical images with a deep learning multi-hypothesis approach: MedGIFT-UPB participation in the ImageCLEF 2017 caption task," in *Proc. CEUR Workshop*, vol. 1856, 2017.

[25] X. Wang, Y. Zhang, Z. Guo, and J. Li, "Identifying concepts from medical images via transfer learning and image retrieval," *Math. Biosciences Eng.*, vol. 16, no. 4, pp. 1978–1991, 2019.

[26] Z. Guo, X. Wang, Y. Zhang, and J. Li, "ImageSem at ImageCLEFmed caption 2019 task: A two-stage medical concept detection strategy," in *Proc. CLEF Work. Notes*, 2019.

[27] N. N. Hoavy, J. Mothe, and M. I. Randrianarivony, "IRIT & MISA at image CLEF 2017-Multi label classification," in *Proc. CEUR Workshop*, 2017.

[28] Abacha, Asma Ben, Seco, S. Gayen, D. Demner-Fushman, and S. Antani, "NLM at ImageCLEF 2017 caption task," in *Proc. CEUR Workshop*, 2017.

[29] E. Pinho and C. Costa, "Unsupervised learning for concept detection in medical images: A comparative analysis," *Appl. Sci.*, vol. 8, no. 8, p. 1213, Jul. 2018.

[30] F. Charalampakos, G. Zachariadis, J. Pavlopoulos, V. Karatzas, C. Trakas, and I. Androutsopoulos, "AUEB NLP group at ImageCLEFmedical caption 2022," in *Proc. CEUR Workshop*, Bucharest, Romania, 2022, pp. 1184–1200.

[31] F. Charalampakos, V. Karatzas, V. Kougia, J. Pavlopoulos, and I. Androutsopoulos, "AUEB NLP group at ImageCLEFmed caption tasks 2021," in *Proc. CEUR Workshop*, Bucharest, Romania, 2021.

[32] J. Jacutprakart, F. P. Andrade, R. Cuan, A. A. Compean, G. Papanastasiou, and A. Garcia, "NLIP-Essex-ITESM at ImageCLEFcaption 2021 task: Deep learning-based information retrieval and multi-label classification towards improving medical image understanding," in *Proc. CEUR Workshop*, Bucharest, Romania, 2021.

[33] G. Schuit, V. Castro, P. Pino, D. Parra, and H. Lobel, "PUC Chile team at concept detection: K nearest neighbors with perceptual similarity," in *Proc. CEUR Workshop*, Bucharest, Romania, 2021.

[34] B. Djamila-Romaissa, O. Mourad, and S. Tapio, "Attention-based CNN-GRU model for automatic medical images captioning: ImageCLEF 2021," in *Proc. CEUR Workshop*, Bucharest, Romania, 2021.

[35] X. Wang, Z. Guo, C. Xu, L. Sun, and J. Li, "ImageSem group at ImageCLEFmed caption 2021 task: Exploring the clinical significance of the textual descriptions derived from medical images," in *Proc. CEUR Workshop*, Bucharest, Romania, 2021.

[36] D. Serra, F. Deligianni, J. Dalton, and A. Q. O'Neil, "CMRE-UoG team at ImageCLEFmedical caption 2022: Concept detection and image captioning," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[37] I. Rio-Torto, C. Patrício, H. Montenegro, and T. Gonçalves, "Detecting concepts and generating captions from medical images: Contributions of the VCMI team to ImageCLEFmedical 2022 caption," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[38] M. M. Rahmana and O. Layodea, "CS_Morgan at ImageCLEFmedical 2022 caption task: Deep learning based multi-label classification and transformers for concept detection & caption prediction," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[39] L. Lebrat, A. Nicolson, S. Cruz, G. Belous, B. Koopman, and J. Dowling, "CSIRO at ImageCLEFmedical caption 2022," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[40] G. Moschovis and E. Fransén, "Neuraldynamicslab at ImageCLEF medical 2022," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[41] A. Mir, K. de Gast, and M. J. Carman, "Polimi-imageclef group at ImageCLEFmedical caption task 2022," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[42] R. Sonker, A. Mishra, P. Bansal, and A. Pattnaik, "Techniques for medical concept detection from multi-modal images," in *Proc. CLEF Work. Notes, CEUR Workshop*. Thessaloniki, Greece: CEUR-WS.org, Sep. 2020.

[43] M. Rahman, T. Lagree, and M. Taylor, "A cross-modal concept detection and caption prediction approach in ImageCLEFcaption track of ImageCLEF 2017," in *Proc. CLEF Work. Notes*, 2017.

[44] M. Hajihosseini, Y. Lotfollahi, M. Nobakhtian, M. M. Javid, F. Omidi, and S. Eetemadi, "IUST_NLPLAB at ImageCLEFmedical caption tasks," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[45] A. G. Seco, F. P. Andrade, L. Bentley, and A. A. Compean, "Essex at image-CLEFcaption 2020 task," in *Proc. CEUR Workshop*, Thessaloniki, Greece, Sep. 2020.

[46] A. Gentili, "ImageCLEFmedical caption task, concept detection, finding duplicates, SDVA-UCSD approach," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[47] R. Tsuneda, T. Asakawa, K. Shimizu, T. Komoda, and M. Aono, "Kdelab at ImageCLEFmedical 2022: Medical concept detection with image retrieval and code ensemble," in *Proc. CEUR Workshop*, Bologna, Italy, 2022.

[48] S. A. Hasan, Y. Ling, J. Liu, R. Sreenivasan, S. Anand, T. R. Arora, V. Datla, K. Lee, A. Qadir, C. Swisher, and O. Farri, "Attention-based medical caption generation with image modality classification and clinical concept mapping," in *Proc. 9th Int. Conf. Labs Eval. Forum*, Avignon, France, Sep. 2018, pp. 224–230.

[49] J. Xu et al., "Concept detection based on multi-label classification and image captioning approach-DAMO at ImageCLEF 2019," in *Proc. CLEF Work. Notes*, 2019.

[50] D. Lyndon, A. Kumar, and J. Kim, "Neural captioning for the ImageCLEF 2017 medical image challenges," in *Proc. CLEF Work. Notes*, 2017.

[51] D. Lyndon, A. Kumar, and J. Kim, "Neural captioning for the ImageCLEF 2017 medical image challenges," in *Proc. CLEF Work. Notes*, 2017.

[52] M. Rahman, "A cross modal deep learning based approach for caption prediction and concept detection by CS Morgan State," in *Proc. CLEF Work. Notes*, 2018.

[53] C. Eickhoff, I. Schwall, A. G. S. De Herrera, and H. Müller, "Overview of ImageCLEFcaption 2017—Image caption prediction and concept detection for biomedical images," in *Proc. CEUR Workshop*, 2017.

[54] Y. Zhang, X. Wang, Z. Guo, and J. Li, "ImageSem at ImageCLEF 2018 caption task: Image retrieval and transfer learning," in *Proc. CLEF CEUR Workshop*, Avignon, France, 2018.

[55] X. Wang, Z. Guo, Y. Zhang, and J. Li, "Medical image labelling and semantic understanding for clinical applications," in *Proc. Int. Conf. Cross-Language Eval. Forum Eur. Lang.*, 2019, pp. 260–270.

[56] X. Wang and N. Liu, "AI600 lab at ImageCLEF 2019 concept detection task," in *Proc. CLEF Work. Notes*, 2019.

[57] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, Sep. 1995, Art. no. 3.

[58] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Aug. 2008.

[59] L. Ralaivola, S. J. Swamidass, H. Saigo, and P. Baldi, "Graph kernels for chemical informatics," *Neural Netw.*, vol. 18, no. 8, pp. 1093–1110, Oct. 2005.

[60] Y. Guo and L. Zhang, "One-shot face recognition by promoting underrepresented classes," 2017, *arXiv:1707.05574*.

[61] K. Brinker, J. Fürnkranz, and E. Hüllermeier, "A unified model for multilabel classification and ranking," in *Proc. 17th Eur. Conf. Artif. Intell.*, Riva del Garda, Italy, Aug. 2006, pp. 489–493.

[62] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, Sep. 2004.

[63] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

[64] C. Bishop and N. Nasrabadi, "Probability distributions," *Pattern Recognit. Mach. Learn.*, vol. 4, no. 4, 2006.

[65] J. Tolles and W. Meurer, "Logistic regression: Relating patient characteristics to outcomes," *J. Amer. Med. Assoc.*, vol. 316, no. 5, pp. 533–534, 2016.

[66] T. Kam Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, Aug. 1995, pp. 278–282.

[67] O. Pelka, A. Ben Abacha, A. G. S. de Herrera, J. Jacutprakart, C. M. Friedrich, and H. Müller, "Overview of the ImageCLEFmed 2021 concept & caption prediction task," in *Proc. CLEF Conf. Labs Eval. Forum*, Bucharest, Romania, Sep. 2021.

[68] O. Pelka, A. B. Abacha, A. G. S. de Herrera, J. Jacutprakart, C. M. Friedrich, and H. Müller, "Overview of the ImageCLEFmed 2021 concept & caption prediction task," in *Proc. CLEF Conf. Labs Eval. Forum—Work. Notes*, Bucharest, Romania, Sep. 2021.

[69] A. G. S. de Herrera, C. Eickhof, V. Andrearczyk, and H. Müller, "Overview of the ImageCLEF 2018 caption prediction tasks," in *Proc. Work. Notes CLEF Conf. Labs Eval. Forum (CLEF)*, Avignon, France, Sep. 2018.

[70] O. Pelka, S. Koitka, J. Rückert, F. Nensa, and C. M. Friedrich, "Radiology objects in COntext (ROCO): A multimodal image dataset," in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis: 7th Joint International Workshop, CVII-STENT 2018 and Third International Workshop, LABELS 2018, Held in Conjunction With MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3*. Springer, 2018, pp. 180–189.
[71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[72] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[73] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[74] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
[75] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*.
[76] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.
[77] C. E. Metz, "Basic principles of ROC analysis," *Seminars Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978.
[78] N. Udas, F. Beuth, and D. Kowerko, "Concept detection in medical images using Xception models—TUC_MC at ImageCLEFmed 2020," in *Proc. CEUR Workshop*, Thessaloniki, Greece, Sep. 2020.
[79] M. Kalimuthu, F. Nunnari, and D. Sonntag, "A competitive deep neural network approach for the ImageCLEFmed caption 2020 task," in *Proc. CEUR Workshop*, Thessaloniki, Greece, Sep. 2020.
[80] R. Sonker, A. Mishra, P. Bansal, and A. Pattnaik, "Techniques for medical concept detection from multi-modal images," in *Proc. CEUR Workshop*, Thessaloniki, Greece, Sep. 2020.
[81] S. D. Arul and K. Srinivasan, "ImageCLEF 2020: Image caption prediction using multilabel convolutional neural network," in *Proc. CEUR Workshop*, Thessaloniki, Greece, Sep. 2020.
[82] A. Tharwat, *Classification Assessment Methods*. Emerald Publishing, 2020.

DIANA MIRANDA received the B.E. degree in computer engineering and the M.E. degree in computer science and engineering from Goa College of Engineering, Goa, India, in 2011 and 2016, respectively. She is currently pursuing the Ph.D. degree in computer science and engineering with the National Institute of Technology Goa. She has worked for three years as a Developer with SAP India Pvt Ltd., Bengaluru, Karnataka, India. She is also an Assistant Professor with Goa College of Engineering. Her research interests include natural language processing, machine learning, pattern recognition, biomedical imaging, and deep learning.

VEENA THENKANIDIYOOR (Member, IEEE) received the B.E. degree from Manipal Institute of Technology, Manipal, and Mangalore University, India, in 1998, the M.S. degree from Manipal Academy of Higher Education, Manipal, in 2000, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Madras, Chennai, India, in 2014. From 2013 to 2018, she was an Assistant Professor with the Department of Computer Science and Engineering, NIT Goa, India, where she has been an Associate Professor, since 2018. Her research interests include deep learning, kernel methods, computer vision, speech processing, content-based information retrieval, and pattern recognition.

DILEEP AROOR DINESH (Member, IEEE) received the B.E. degree in computer science and engineering from Gulbarga University, Karnataka, India, in 2000, and the M.Tech. and Ph.D. degrees in computer science and engineering from the Indian Institute of Technology (IIT) Madras, in 2006 and 2013, respectively. He was with the School of Computing and Electrical Engineering, IIT Mandi, Himachal Pradesh, from 2013 to 2023. Since December 2023, he has been an Associate Professor with the Department of Computer Science and Engineering, IIT Dharwad, Karnataka. His current research interests include machine learning, kernel methods, and deep learning, with applications in speech technology, computer vision, and cloud network resource utilization.