IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 February 2023, accepted 13 March 2023, date of publication 20 March 2023, date of current version 23 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3259062

RESEARCH ARTICLE

Common Thorax Diseases Recognition Using Zero-Shot Learning With Ontology in the Multi-Labeled ChestX-ray14 Data Set

OMURHAN AVNI SOYSAL[1], **MEHMET SERDAR GUZEL**[1], **MEHMET DIKMEN**[2], **AND GAZI ERKAN BOSTANCI**[1]

[1]Department of Computer Engineering, Ankara University, 06830 Ankara, Turkey
[2]Department of Computer Engineering, Başkent University, 06790 Ankara, Turkey

Corresponding author: Omurhan Avni Soysal (omurhan.soysal@ankara.edu.tr)

ABSTRACT Disease detection/recognition with limited data sets and labels in the medical image domain is a very costly and greatest challenge. Although open image data sets have increased recently, researches on this problem still need to be developed. Researches to diversify data sets are both costly and face the problem of subjectivity. Unseen classes can be trained with the Zero-Shot Learning (ZSL) in order to overcome this problem. In this paper, we aimed to strengthen ZSL by using ontology as an auxiliary information for class embeddings. In our approach, ZSL is supported by the image embeddings and class embeddings of the multi-labelled ChestX-ray14 data set, as well as the semantic data from DBpedia. In this paper, which we believe will be pioneering in the medical image domain, the Cosine, Hamming and Euclidean distances were taken into account in order to maximize the similarities. We trained ResNet50 neural network with different parameters on the multi-labelled ChestX-ray14 data set. 23.25% precision value in one-to-one matching and 29.59% precision value in at least one matching were obtained. We think that this paper will make a significant contribution to the medical image domain by detecting/recognizing unseen disease images.

INDEX TERMS Zero-shot learning, ResNet50, ontology, ChestX-ray14.

I. INTRODUCTION

Deep learning methods are so widespread due to developments in the field of hardware as well as improved methods. Graphics Processing Units (GPU) have made it easier to utilize methods with long run times that require big data, such as deep learning methods. The fact that the training phase, which could last for days, could be reduced so that it takes only hours led to both the development of deep learning methods and the ability to use these methods in daily life.

In deep learning, as in traditional machine learning methods, the training data are very important. The variety and definition of the data (e.g. labels) to be used for training affect the performance of the network to be trained. Therefore, the preparation of training data is a critical step. Although different data sets have been prepared to solve many problems in recent years and have been made available to researchers, they are still not sufficient in terms of both content and definition. The manual labeling of data brings with it the problem of subjectivity, and it is costly as well. The problem of diversity comes to the forefront in data sets that claim to have solved the problem of subjectivity. A level of diversity that is high enough to represent the real world is still not included in many data sets. Although these data sets include content related to specific problems, they cannot contain all the content specific to a given problem in the full sense. Yet another problem is the imbalance of classes in data sets: While some classes contain a large amount of data, other classes do not have enough data, resulting in problems during the training of the network. This situation causes the network to learn some classes well while not learning others well, resulting in false positive detections.

The diversity of the data set, the number of samples is the greatest challenge in the medical image domain. Although public data sets have been trying to expand recently, these

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

data sets are still limited and deal with the specific problem space. Enlarging these data sets (both increasing their number and diversifying them to best represent the problem) is a very costly task. Moreover, the procedure to label these data sets is a process that should be worked under the supervision of radiologists and researchers together.

In order to overcome these challenges, there has been considerable progress in the development of methods for the recognition of unseen classes using their auxiliary information. One of these methods is Zero-Shot Learning (ZSL). ZSL, which was developed for the detection/recognition of unseen classes with semantic transfer and still has many challenges that researchers need to overcome, is an important solution. In addition to using ZSL in this paper, we also integrated ontology into our method to strengthen semantic relationships. During the training, we developed a method with high detection power by combining both the capabilities of ZSL, which uses deep learning methods, and the power of ontology to detect/recognize unseen classes.

The ChestX-ray14 data set used in this paper. Fourteen common thoracic pathologies include Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural Thickening, Cardiomegaly, Nodule, Mass and Hernia.

Our main contributions:

1) Detect/recognize unseen classes using Zero-Shot Learning
2) Using the semantic equivalents of these labels instead of labels in the data set
3) Leveraging the capabilities of ontology in class embeddings
4) Overcoming the overfitting problem even if the data set is imbalanced
5) ZSL and ontology have not used on ChestX-ray 14 data set (together or separately) before

In the second part of this article, previous studies on the ZSL are discussed. The third section includes the fundamentals of the ZSL, and the fourth section mentions the fundamentals of ontology. In the fifth section, the definition of the data sets used in this paper and a description of how this data sets were prepared are given. In the sixth section, the formulation of the problem, notation and the proposed method are explained. In the seventh section, the results of the applied method and detailed information about the development environment in which this method was implemented is given. Conclusion containing future work is discussed in Section VIII.

II. RELATED WORK

Previous studies about ZSL were discussed in our previous paper in detail [1].

Studies suggesting the ZSL method frequently appear in the literature each year. Although each new method has been shown to improve upon previous methods, this progress is difficult to measure without an established evaluation protocol [2]. Based on this reality, Xian et al. have defined a new benchmark by combining both evaluation protocols and data splits. They evaluated 10 ZSL methods on five data sets for both ZSL and generalized ZSL settings, provided tests of statistical significance and robustness, and presented other valuable insights from this benchmarking. In this paper, Direct Attribute Prediction (DAP) achieved nearly 22% of Top-1 accuracy.

In this sense, Keshari et al. [3] performed a more comprehensive evaluation compared to previous studies. Based on the finding that the performance of most existing supervised deep neural network (DNN) algorithms degrades for unseen classes in the training set, they suggested using the over-complete distribution (OCD), with a conditional variation autoencoder (CVAE), of both seen and unseen classes to learn a discriminative classifier that performs well in ZSL settings, and they evaluated their method using both ZSL and generalized ZSL protocols on the SUN, CUB and AWA2 data sets. But they didn't evalute their data set in seen and unseen classes separately.

Yu et al. [4] presented an effective episode-based training framework for ZSL. The model is trained in a series of sections, each designed to simulate the ZSL classification task. By training on all the episodes, the model progressively acquires community experience in predicting mimetic unseen classes that will generalize well over real-seeming classes. Extensive experiments on four data sets (AWA1, AWA2, CUB, FLO) have shown that the proposed model outperforms state-of-the-art approaches with wide margins. But they didn't evaluate their proposed method on medical data set.

Wang et al. [5] aimed to recognize human interactions with new objects via ZSL. Unlike previous researchers, they allowed for unseen object categories using semantic word embedding. To do this, they designed a human-object region recommendation network specifically for the human-object interaction detection task. The main concept is to leverage human visual cues to localize objects interacting with humans. They test three semantic embeddings (GloVe, word2vec, and FastText). Word2vec on GoogleNews achieves the overall best performance (11.5%). They evaluated their method on the V-COCO and HICO-DET data sets.

Han et al. [6] suggested learning redundancy-free features for generalized ZSL to reduce redundant information in fine-grained objects. They projected the image embeddings into a new feature space and then restricted the statistical dependence. They removed the redundancy-free features without losing the distinguishing data from the visual features, preserving and even strengthening the relationships in the redundancy-free feature space. They evaluated the proposed method on four different data sets (AWA1, CUB, FLO, SUN).

Huynh and Elhamifar [7] proposed a dense feature-based interest approach that focuses its attention on the image regions that are most related to each attribute and obtains attribute-based features. Instead of aligning the global

TABLE 1. Literature Review. 'v' denotes is used in paper and 'x' denotes is not used.

	ZSL	Ontology	ChestX-ray14
Xian et al. [2]	v	x	x
Keshari et al. [3]	v	x	x
Yu et al. [4]	v	x	x
Wang et al. [5]	v	v	x
Han et al. [6]	v	x	x
Huynh et al. [7]	v	x	x
Huynh et al. [8]	v	x	x
Liu et al. [9]	v	x	x
Prasanna et al. [10]	x	v	x

attribute vector of an image with the semantic vector of the associated class, they studied an attribute embedding technique. Therefore, they computed a vector of attribute scores for each attribute entity in an image that maximized similarity to the real class. They also adjusted each attribute score by using a focusing mechanism on attributes to better capture different attributes. In order to overcome the problem of bias towards classes seen during the testing, they proposed a new self-calibration loss that adjusts the probability of unseen classes to account for training bias. They evaluated their approach on the large-scale DeepFashion data set as well as on CUB, SUN and AWA2.

Huynh and Elhamifar [8] developed a shared multi-focus model for multi-label ZSL. They argued that it is not a trivial task to design an attention mechanism to recognize multiple seen and unseen labels in an image, because there is no training signal to locate the unseen labels, and an image contains only a few available labels out of many possible labels. Thus, instead of drawing attention to unseen labels with unknown behaviors that might draw attention to unrelated regions due to the absence of any training samples, they allowed unseen labels to choose from a set of shared attention trained independently of the label. They evaluated the method on the NUS-WIDE and large-scale Open Images data sets.

Liu et al. [9] have proposed a hyperbolic visual embedding learning network for zero-shot recognition. The network is trained with image embeddings in hyperbolic space, which can preserve the hierarchical structure of semantic classes at low dimensions. They argued that compared to other zero-point learning methods, the network is more robust because embedding in hyperbolic space better represents the class hierarchy, thus avoiding the misconceptions of unrelated siblings.

Prasanna et al. [10] worked on the multi-label classification problem using a Convolutional Neural Network (CNN) and applied the method to a data set consisting of images they collected.

All the papers detailed in this section are listed in Table 1.

Studies in the medical domain with ZSL are quite limited. Although there are studies using deep learning methods on lung images, we could not find any study using ZSL methods on the ChestX-ray14 data set. Mostly, single label data sets were used in the studies. Studies using ontology in ZSL's class embeddings are very limited. Although the precision values we have obtained are low and higher than the studies in the literature, we are aware that it still needs improvement. It will help radiologists and specialists in order to detect/recognize unseen disease classes if ZSL methods in the medical domain increase.

III. FUNDAMENTALS OF ZERO-SHOT LEARNING

Studies using traditional machine learning methods had a limited success rate until 2012. Due to the work of Krizhevsky et al. [11] in the ImageNet competition, the success rate, which was 74% until 2012, increased to 83.6%. In studies carried out in the following years, this success rate has exceeded 90%. At this point, methods that have a margin of error less than the human margin of error have been developed. With the introduction of deep learning into our lives, the data set problem continues, just as it has for traditional machine learning methods. Among the most important factors that allow deep learning to achieve such high success rates, the data used in the training phase plays a critical role, as well as the depth of the trained network.

The main aim of ZSL is to detect/recognize unseen classes through learning matrix mapping that bridges the gap between visual information and semantic attributes [12]. Unlike traditional machine learning methods, in ZSL, the data in the test phase are not included in the training phase. This is one of the most powerful features that distinguishes learning without examples from other methods. In this way, ZSL draws attention both because it is similar to human learning and because it is a more practical solution to potential problems that may be encountered in real life.

The human mind tries to make an object or shape (in short, a "thing") look like them (that object or shape), based on what it has learned before, and carries out an idea. In this way, learning continues while a human being is trying to grasp the new "thing." For some examples in real life, it is not possible to find labeled data to give to the network during the training phase. Instead, performing the training with the help of seen classes and making inferences using the label data of unseen classes is the preferred method in human learning.

Another prominent feature of ZSL is the evaluation of visual data as well as auxiliary data during training and testing. This ensures that the neural network does a better job of representing the data. When current ZSL studies are examined, it is clear that many of them are conducted to find more effective ways to describe visual data. ZSL has a feature that can handle problems in many fields, especially health: It can transfer visual and/or semantic data from unseen classes to seen classes using various methods.

In ZSL, auxiliary information is used in addition to the data sets used in classical machine learning methods. These auxiliary data can be text representations of class names, manually labeled attributes or hierarchical representations, as well as the semantic data of the attributes, the details of which are mentioned in the following sections in this paper. This auxiliary information and the class representations support each other.

IV. FUNDAMENTALS OF ONTOLOGY

The word ‘‘ontology’’ is formed by combining the words on or ontos, which mean ‘‘being’’ or ‘‘existence’’ in Ancient Greek, and logos, which means ‘‘word,’’ ‘‘mind,’’ ‘‘speaking’’ or ‘‘knowledge’’. It is a branch of philosophy [13]. In philosophy, ontology refers to theories about the nature of being and what kinds of things exist. Artificial intelligence and web researchers have chosen to use this term in a slightly different way, and for them an ontology is a document or file that formally defines the relationships between terms [14].

Ontologies are knowledge models that describe knowledge in a specific field using a formal infrastructure and through a set of concepts and the relationships between these concepts. Ontology engineering, on the other hand, is an engineering field that aims to bring together the methods, activities, technologies and toolkits required for the development of ontologies in the most effective way. An ontology can be classified according to the scope of the objects that the ontology defines. For example, the scope of a local ontology is narrower than that of a domain ontology; domain ontologies have more specific concepts than basic reference ontologies, which contain the basic concept of a domain. Core ontologies can be viewed as meta-ontologies that describe high-level concepts or principles used to describe other ontologies. Generic ontologies do not belong to a specific field, so their concepts can be as general as those of basic reference ontologies [15].

Sir et al. [16], in their paper examining the difference between an ontology and a database, state that the closed-world assumption (CWA) and open-world assumption (OWA) are both concepts used in the presentation of information. They state that systems with full information benefit from the CWA, while the OWA is applied in systems with incomplete information. They say that the main difference between ontologies and databases is based on the difference between the OWA and CWA: Ontologies use the OWA knowledge representation system, while the CWA is used by databases. Any missing data in a database has a value of ‘‘0’’. Any missing data in an ontology system is considered unknown.

The Semantic Web appears to be a solution to the problems presented by the traditional web, such as the unstructured and unrelated content that is defined on most web pages and does not enable semantically related content to be retrieved. Current database technologies present several problems in the context of the Semantic Web because information cannot be explained semantically. The content of a database is only shown when the database is queried, and on the other hand, the semantic description of the database is represented using its schema, which is often non-existent or even useless because it often cannot be exploited depending on the format chosen to represent it [17].

Since the zero-shot learning method also allows transfer learning, it has some common ground with the methods in the field of ontology engineering. Within this paper, concepts in the field of health and the ontological relationships of these concepts will be used; the details of these concepts are given in the following sections.

TABLE 2. Distributions of 14 disease categories in ChestX-ray 14 data set (Part-1).

	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass
Atelectasis	4212	369	3269	3259	727
Cardiomegaly	369	1094	1060	583	99
Effusion	3269	1060	3959	3990	1244
Infiltration	3259	583	3990	9552	1151
Mass	727	99	1244	1151	2138
Nodule	585	108	909	1544	894
Pneumonia	243	36	253	571	62
Pneumothorax	772	48	995	943	424
Consolidation	1222	169	1287	1220	602
Edema	221	127	592	979	128
Emphysema	423	44	359	447	212
Fibrosis	220	51	188	345	115
Pleural Thickening	495	111	848	749	448
Hernia	40	7	21	33	25
No Nodule	11535	2772	13307	19871	5476

V. DATA SETS

A. CHestX-ray14

The ChestX-ray14 data set used in this paper was retrieved from the Picture Archiving Communication Systems (PACS) at the National Institute of Health Clinical Health Center. Chest X-ray exam is one of the most frequent and cost-effective medical imaging examination. However, clinical diagnosis of Chest X-ray can be challenging. Although the local pathological image regions have large dimensions, they are quite small when the full image is taken into account. There is no explanation about the interrelationships of these common thoracic diseases or their clinical symptoms. There are 112,120 frontier lung surface images taken from 30,805 individual patients (collected from the year of 1992 to 2015). The data set consists of lung images marked with multiple labels. Wang et al. [18], expanded the ChestX-ray8 data set, which contained eight different thoracic diseases, in 2017, and later, it was expanded to contain 14 different thoracic diseases (1-Atelectasis; 2-Cardiomegaly; 3-Effusion; 4-Infiltration; 5-Mass; 6-Nodule; 7-Pneumonia; 8-Pneumothorax; 9-Consolidation; 10-Edema; 11-Emphysema; 12-Fibrosis; 13-Pleural Thickening; 14-Hernia). The distributions of the 14 disease categories with co-occurrence statistics are shown in figure 1. Sample thoracic pathologies with single label from the each classes from ChestX-ray14 data set are shown in figure 2. The distributions of the 14 disease categories are also shown in table 2 and 3.

Although labeling has begun on some of the images, only 1,000 images have been labeled so far; the entire data set has not been labelled. Therefore, these limited data were not evaluated in this paper.

Although suggestions were made for training and test data in the prepared data set, random data were selected from the

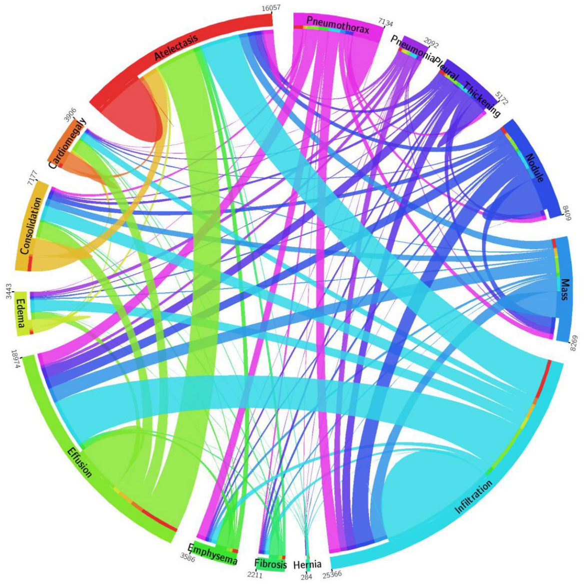
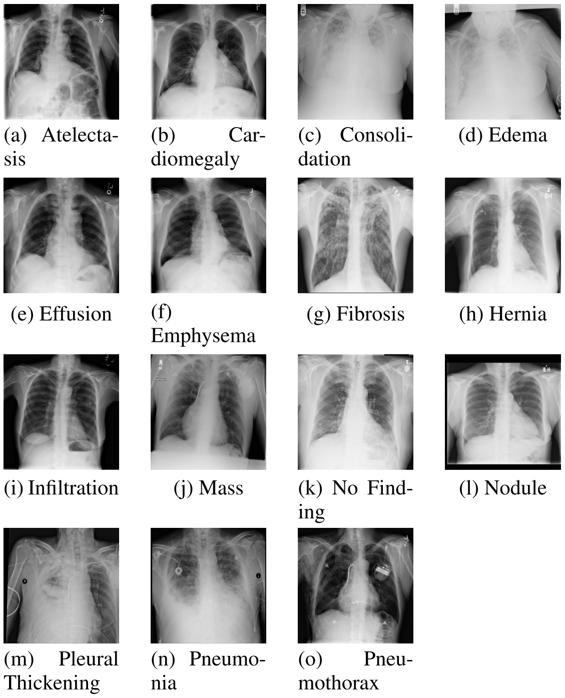


FIGURE 1. The circular diagram shows the proportions of images with multi-labels in each of 14 pathology classes and the labels' co-occurrence statistics [18]: Atelectasis is red, Cardiomegaly is orange, Effusion is chartreuse-green, Infiltration is turquoise, Mass is azure, Nodule is blue, Pneumonia is violet, Pneumothorax is purple, Consolidation is gold, Edema is bright-green, Emphysema is green, Fibrosis is spring-green, Pleural Thickening is blue-violet, Hernia is aquamarine.



(a) Atelectasis (b) Cardiomegaly (c) Consolidation (d) Edema

(e) Effusion (f) Emphysema (g) Fibrosis (h) Hernia

(i) Infiltration (j) Mass (k) No Finding (l) Nodule

(m) Pleural Thickening (n) Pneumonia (o) Pneumothorax

FIGURE 2. Fourteen common thoracic pathologies.

total data set instead: 80% of the total data set was reserved for training and 20% for testing. Unseen classes were included in the test data set, and the matching information in the test data set was taken into account for validation.

Another disadvantage of the data set within the scope of this paper is the absence of feature data. Although the patients' age and gender information were given, they were not evaluated in this paper due to the weak discriminating features of these data. Since there was no physician support, the missing feature data were obtained from DBpedia.

B. DBpedia

The DBpedia Association was established in 2014 to support DBpedia and the DBpedia community. Since then, professionalization efforts for all users have continued due to the contributions of the DBpedia community. DBpedia is a community that aims to gather and construct structured information from Wikipedia and make that information available on the web [19].

The DBpedia community is working hard to localize and internationalize DBpedia and support the release of non-English versions of Wikipedia, as well as to build data communities around specific languages, regions or areas of special interest. Currently, DBpedia has about 20 language sections that deal with improving data extraction from language-specific versions of Wikipedia. These sections are part of the DBpedia administrators and are responsible for contributing to DBpedia's infrastructure. Dutch DBpedia is the first such community to formally become an official DBpedia chapter consortium.

TABLE 3. Distributions of 14 disease categories in ChestX-ray 14 data set (Part-2).

	Nodule	Pneumonia	Pneumothorax	Consolidation	Edema
Atelectasis	585	243	772	1222	221
Cardiomegaly	108	36	48	169	127
Effusion	909	253	995	1287	592
Infiltration	1544	571	943	1220	979
Mass	894	62	424	602	128
Nodule	2706	63	340	428	131
Pneumonia	63	307	34	114	330
Pneumothorax	340	34	2199	222	33
Consolidation	428	114	222	1314	162
Edema	131	330	33	162	634
Emphysema	115	21	746	103	30
Fibrosis	166	11	80	79	9
Pleural Thickening	410	45	289	251	64
Hernia	10	2	9	4	3
No Nodule	6323	1353	5298	4667	2303

The attributes of the 14 class labels in the ChestX-ray14 data set were retrieved from DBpedia. The synonyms of some labels are preferred. A total of 1,258 unique attributes were identified for 836 classes. Then, the label-attribute matrix was prepared.

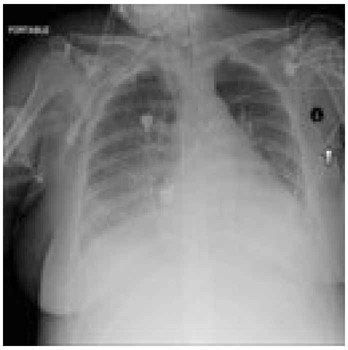


FIGURE 3. This is a sample image from ChestX-ray 14 data set with labelled Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Infiltration, Mass, Nodule.

TABLE 4. Distributions of 14 disease categories in ChestX-ray 14 data set (Part-3).

	Emphysema	Fibrosis	Pleural Thickening	Hernia
Atelectasis	423	220	495	40
Cardiomegaly	44	51	111	7
Effusion	359	188	848	21
Infiltration	447	345	749	33
Mass	212	115	448	25
Nodule	115	166	410	10
Pneumonia	21	11	45	2
Pneumothorax	746	80	289	9
Consolidation	103	79	251	4
Edema	30	9	64	3
Emphysema	895	36	151	4
Fibrosis	36	727	176	8
Pleural Thickening	151	176	1127	8
Hernia	4	8	8	110
No Nodule	2516	1686	3385	227

VI. PROPOSED METHOD

In this section, we define the problem formulation and notation in the first subsection. This subsection also includes details about the data set's classes. Then, the details of the proposed method and the method's steps are given in the second subsection. The architecture of the proposed method (in figure 4) is also included in this same section.

A. PROBLEM FORMULATION AND NOTATIONS

Our data set consists of two parts:

$$Z \cup T \tag{1}$$

Z denotes seen classes, and T denotes unseen classes. We can express the seen classes as follows:

$$Z = (x_i^a, y_i^a)_{i=1}^{M_a} \tag{2}$$

Here, M_a denotes the total number of images in the seen class to be used in the training of the neural network, x_i^a denotes the i. image in the seen class, and y_i^a denotes the label corresponding to this image.

$$T = (x_j^b, y_j^b)_{j=1}^{P_b} \tag{3}$$

TABLE 5. Test scenarios (Part-1).

Scenario	Distance	Batch Size	Learning Rate	Dropout	Epoch Size
1	Euclidean	4	0,25	0,00001	50
2	Euclidean	4	0,5	0,00001	50
3	Cosine	4	0,25	0,00001	50
4	Cosine	4	0,5	0,00001	50
5	Hamming	4	0,25	0,00001	50
6	Hamming	4	0,5	0,00001	50
7	Euclidean	24	0,5	0,00001	50
8	Cosine	24	0,5	0,00001	50
9	Hamming	24	0,5	0,00001	50
10	Euclidean	24	0,5	0,0001	50
11	Cosine	24	0,5	0,0001	50
12	Hamming	24	0,5	0,0001	50
13	Euclidean	24	0,25	0,0001	50
14	Cosine	24	0,25	0,0001	50
15	Hamming	24	0,25	0,0001	50
16	Euclidean	4	0,25	0,0001	50
17	Cosine	4	0,25	0,0001	50
18	Hamming	4	0,25	0,0001	50
19	Euclidean	4	0,75	0,00001	50
20	Euclidean	4	0,75	0,00001	50
21	Euclidean	4	0,75	0,00001	50
22	Euclidean	4	0,75	0,00001	60
23	Euclidean	4	0,75	0,00001	100
24	Euclidean	4	0,75	0,00001	100

Here P_b denotes all the images in the unseen class of the neural network, x_j^b denotes j. image in the unseen class and y_j^b denotes the label corresponding to this image. Images in seen and unseen classes are disjoint. In other words,

$$x^a \cap x^b = \emptyset \tag{4}$$

Attribute vectors corresponding to each $y \in Y$ label are also defined. These vectors were retrieved from DBpedia. Mapping was done with the DBpedia data corresponding to the labels.

Since the ChestX-ray14 data set has multi-label data, the training and test data sets are tailored accordingly. For example, figure 3 has a total of eight labels (Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Infiltration, Mass, Nodule). That is why this image is marked with these eight labels.

In total, 669 classes were created in the training data set and 167 classes were created in the test data set. The size of the attribute vector retrieved from DBpedia was 1,258 and the size of the binary mapping vector was 836 × 1,258.

B. IMPLEMENTATION

After the training and test data sets were prepared, we trained the neural network. The ResNet50 structure was used for the training of the neural network. ResNet50 [20] is a convolutional neural network (CNN) that is 50 layers deep. The layers are explicitly reformulated so that they are residual functions that learn by referencing layer inputs instead of learning unreferenced functions. Shortcut links only perform ID mapping and their output is appended to the stacked layers' outputs (figure 5). The default output of ResNet50's fully connected layer is a 2048-dimensional fully connected vector. In this study, a fully connected layer is added to the output value for each attribute. The output of size 1,258

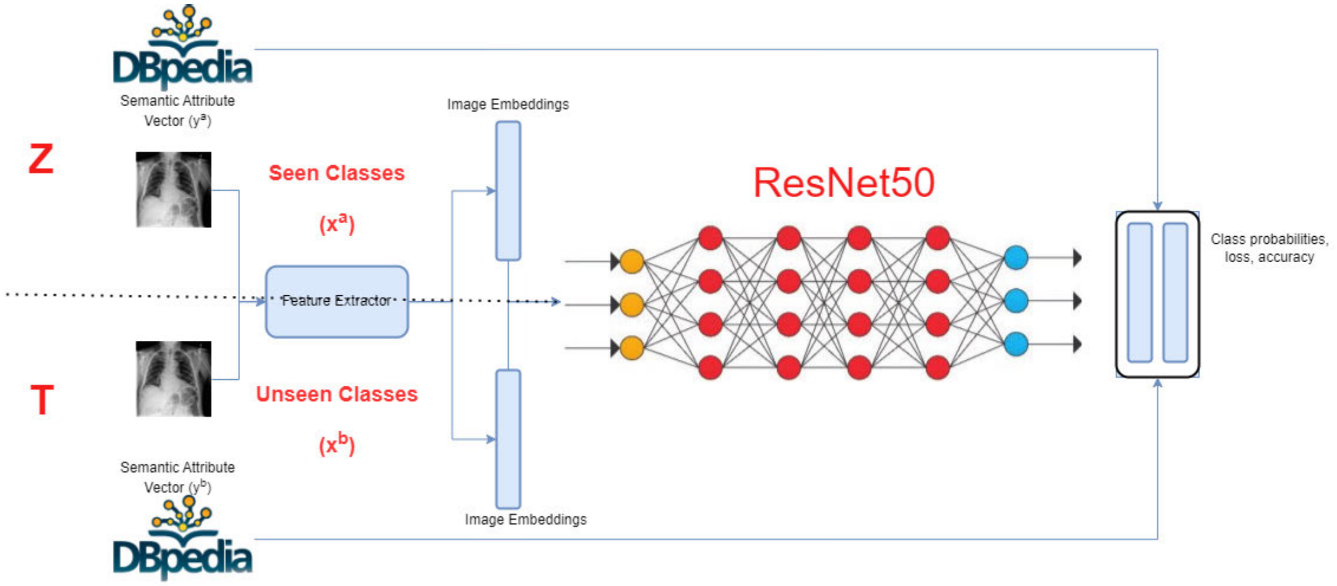


FIGURE 4. Architecture of proposed method. Seen classes (x^a), unseen classes (x^b), semantic attributes of seen classes (y^a), semantic attributes of unseen classes (y^b).

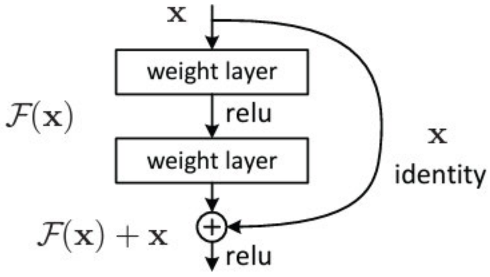


FIGURE 5. Residual learning: a building block.

was passed through a sigmoid activation function to obtain probabilities for each attribute. Since we performed multi-label classification, the binary cross entropy loss was then used to train the network.

When the deep learning literature is examined, it can be seen that the most important criteria in the training of a neural network are the parameters of the training. We trained the neural network with different parameters, and we report the results of this training in Evaluation section: batch size = 4 and 24; dropout = 0.25 and 0.5; epoch size = from 5 to 100, rotation=5 and 15. In order to preprocess: resize=224 and 448, brightness= 0,3 and 0,5 contrast= 0,3 and 0,5 and histogram equalization is applied and not. Adam optimizer were used, and the drop-out was 0,0001 and 0.000001.

VII. EVALUATION

A. PERFORMANCE ANALYSIS

After the training of the neural network, the distance calculation was made using the confusion vector for each image in the test data set. Three different methods were used for distance calculation: the Hamming, Cosine and Euclidean distances. In descriptor-based traditional machine learning techniques, these 3 distance metrics are used frequently. We have chosen these metrics because of their success there. The class in the training data set with the lowest distance in the confusion vector was matched with the corresponding class in the test data set. Tests were carried out for

TABLE 6. Test scenarios (Part-2).

Scenario	Rotation	Resize	Brigthness	Contrast	HistEq
1	15	224	0.3	0.3	X
2	15	224	0.3	0.3	X
3	15	224	0.3	0.3	X
4	15	224	0.3	0.3	X
5	15	224	0.3	0.3	X
6	15	224	0.3	0.3	X
7	15	224	0.3	0.3	X
8	15	224	0.3	0.3	X
9	15	224	0.3	0.3	X
10	15	224	0.3	0.3	X
11	15	224	0.3	0.3	X
12	15	224	0.3	0.3	X
13	15	224	0.3	0.3	X
14	15	224	0.3	0.3	X
15	15	224	0.3	0.3	X
16	15	224	0.3	0.3	X
17	15	224	0.3	0.3	X
18	15	224	0.3	0.3	X
19	15	224	0.3	0.3	X
20	5	224	0.3	0.3	X
21	15	448	0.3	0.3	X
22	15	448	0.3	0.3	X
23	15	448	0.5	0.5	V
24	15	448	0.3	0.3	V

TABLE 7. Precision Results of Test Scenarios (%) from 5th Epoch to 25th Epoch.

Scenario	5. Epoch	10. Epoch	15. Epoch	20. Epoch	25. Epoch
1	13,68	13,67	14,62	15,44	17,12
2	13,66	14,13	13,82	14,26	16,67
3	9,11	8,73	7,02	7,14	5,99
4	8,06	10	6,82	5,29	6,57
5	16,72	16,72	16,72	16,72	16,72
6	16,72	16,72	16,72	16,72	16,72
7	13,66	13,66	13,69	13,67	13,74
8	13,06	7,7	6,77	8,11	8,09
9	16,72	16,72	16,72	16,72	16,72
10	14,69	15,12	15,28	15,15	16,02
11	9,59	6,6	9,26	7,92	11,49
12	16,72	16,72	16,72	16,72	16,72
13	13,72	14,79	16,19	15,55	15,23
14	10,08	10,94	6,19	6,42	8,14
15	16,72	16,72	16,72	16,72	16,72
16	13,66	14,85	15,02	14,73	15,99
17	11,98	10,73	11,42	6,77	9,74
18	16,72	16,72	16,72	16,72	16,72
19	13,66	13,66	14,06	13,87	15,08
20	13,66	13,76	13,65	13,65	14,3
21	13,66	13,54	14,75	17,89	18,89
22	13,66	13,65	13,54	16,36	18,65
23	13,66	13,69	13,49	15,81	17,88
24	13,66	13,66	13,66	18,12	18,54

TABLE 8. Precision Results of Test Scenarios (%) from 30th Epoch to 50th Epoch.

Scenario	30. Epoch	35. Epoch	40. Epoch	45. Epoch	50. Epoch
1	17,97	18,95	17,51	17,7	17,32
2	18,11	17,78	17,54	19,03	18,99
3	5,41	7,34	7,45	10,06	7,34
4	7,35	8,04	7,14	6,49	8,07
5	16,72	16,72	16,72	16,73	16,73
6	16,72	16,72	16,73	16,72	16,72
7	13,73	14,71	14,88	15,16	15,87
8	6,97	8,28	7,09	7,68	7,92
9	16,72	16,72	16,72	16,72	16,72
10	15,88	15,81	16,61	16,41	15,97
11	7,96	7,19	7,81	7,62	9,67
12	16,72	16,72	16,72	16,72	16,72
13	15,8	16,99	16,66	15,92	16,94
14	7,68	9,18	7,35	9,75	11,68
15	16,72	16,72	16,72	16,72	16,72
16	16,81	16,57	17,05	15,43	14,02
17	10,05	9,72	9,49	10,98	11,05
18	16,72	16,72	16,72	16,72	16,74
19	16,47	16,74	19,64	18,17	19,44
20	17,63	17,05	18,61	16,98	17,7
21	19,59	19,06	20,29	20,98	19,99
22	18,73	20,46	19,46	18,4	20,12
23	20,48	21,2	18,46	19,26	20,06
24	19,03	20,83	19,68	19,23	20,66

24 scenarios with different parameters. The different parameters used in these scenarios are shown in the table 5 and 6.

The precision results obtained from 5th and 100th epochs are shown in tables 7, 8, 9 and 10.

The precision was calculated as follows:

$$\frac{T_p}{T_p + F_p} \cdot 100 \tag{5}$$

TABLE 9. Precision Results of Test Scenarios (%) from 55th Epoch to 75th Epoch.

Scenario	55. Epoch	60. Epoch	65. Epoch	70. Epoch	75. Epoch
22	21,54	20,57			
23	19,64	19,78	18,74	21,26	20,22
24	20,44	21,54	20,62	20,07	19,08

TABLE 10. Precision Results of Test Scenarios (%) from 80th Epoch to 100th Epoch.

Scenario	80. Epoch	85. Epoch	90. Epoch	95. Epoch	100. Epoch
22					
23	21,04	20,67	21,61	22,97	18,67
24	**23,25**	21,66	20,04	18,93	18,98

TABLE 11. Precision Results of Test Scenarios (%) at least one matching.

Scenario	Precision (%)
1	22,98
2	24,75
3	9,39
4	10,1
5	20,65
6	20,64
7	21,79
8	9,64
9	20,64
10	21,85
11	12,17
12	20,64
13	22,76
14	14,27
15	20,64
16	18,91
17	14,26
18	20,66
19	25,28
20	23,13
21	25,49
22	28,18
23	**29,59**
24	29,35

True Positive means that an outcome where the model correctly predicts the ones in test classes. False Positive means that an outcome where the model incorrectly predicts the ones in test classes.

The highest precision of 23.25% was obtained in the 80th epoch of scenario 24. So we retrieved 23.25% of the ground-truth images in the test data set.

On the other hand, it was observed that often at least one label was found to match the ground truth, even if some results did not match exactly, since the training was performed with a multi-label data set. For example, an image with the Atelectasis, Consolidation and Effusion labels in the test data set was matched with an image with the Effusion and Infiltration labels in the training data set. As a result, an additional evaluation was made. For this evaluation, precision values were recomputed to take into account at least one label matches from the test results in the above 24 different scenarios, as seen in table 11.

According to these results, a precision of 29.59% was obtained for the "at least one label match" evaluation for scenario 23. Although this evaluation result is specific to the multi-label data set, it is noteworthy in terms of demonstrating the power of ZSL.

B. DEVELOPMENT ENVIRONMENT

The proposed method is implemented in the Python 3.7 environment. Torch 1.10, torchvision 0.10.0 and CUDA Toolkit 11.5 were used together with the open-source conda package and environment management system. GPU card is NVIDIA GeForce RTX 3090/PCIe/SSE2 / NVIDIA GeForce RTX 3090/PCIe/SSE2, Processor is IntelB. Core i9-10900X CPU @ 3.70GHz C- 20 on Ubuntu 20.04.02 LTS operating system.

C. CONSTRAINTS

Train and test image sizes (224 and 448) and batch sizes (4 and 24) are compatible with ImageNet trained models. Batch size couldn't be increased because of the hardware constraints.

VIII. CONCLUSION

In this paper, we have suggested that Zero-Shot Learning (ZSL) should be evaluated as a pioneering method in many fields, especially in the medical image domain. Because disease detection/recognition with limited data sets and labels in the medical image domain is a very costly and greatest challenge. Although open image data sets have increased recently, researches on this problem still need to be developed. In this context, unseen classes have been trained with the ZSL in order to overcome this problem. The ResNet50 structure was used for the training of the neural network with different parameters on multi-labelled ChestX-ray14 data set which contain 14 Common Thorax Diseases. There are 112,120 frontier lung surface images taken from 30,805 individual patients in ChestX-ray14 data set. We aimed to strengthen ZSL by using ontology as auxiliary information for class embeddings. So we have used semantic data related to ChestX-ray14 data set labels from DBpedia. We have tried to maximize the similarities using Cosine, Hamming and Euclidean distances. We have shown that Euclidean distance and related parameters metric have better performance than others. But we are aware of the multi-label data set problem. We think that this is the reason why the precision values are lower than the values in the deep learning literature. In further studies, the ResNet50 architecture will be optimized and re-evaluated with different parameters. In addition, other neural networks from the deep learning literature will also be evaluated within the scope of the method proposed in this work.

ACKNOWLEDGMENT

The ChestX-ray 14 data used to support the findings of this study are available from the author Omurhan Avni Soysal upon request.

REFERENCES

[1] O. A. Soysal and M. S. Guzel, "An introduction to zero-shot learning: An essential review," in *Proc. Int. Congr. Hum.-Comput. Interact., Optim. Robotic Appl. (HORA)*, Jun. 2020, pp. 1–4, doi: 10.1109/HORA49412.2020.9152859.

[2] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning—The good, the bad and the ugly," 2017, *arXiv:1703.04394*.

[3] R. Keshari, R. Singh, and M. Vatsa, "Generalized zero-shot learning via over-complete distribution," 2020, *arXiv:2004.00666*.

[4] Y. Yu, Z. Ji, Z. Zhang, and J. Han, "Episode-based prototype generating network for zero-shot learning," 2019, *arXiv:1909.03360*.

[5] S. Wang, K.-H. Yap, J. Yuan, and Y.-P. Tan, "Discovering human interactions with novel objects via zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11652–11661.

[6] Z. Han, Z. Fu, and J. Yang, "Learning the redundancy-free features for generalized zero-shot object recognition," 2020, *arXiv:2006.08939*.

[7] D. Huynh and E. Elhamifar, "Fine-grained generalized zero-shot learning via dense attribute-based attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4483–4493.

[8] D. Huynh and E. Elhamifar, "A shared multi-attention framework for multi-label zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8776–8786.

[9] S. Liu, J. Chen, L. Pan, C.-W. Ngo, T.-S. Chua, and Y.-G. Jiang, "Hyperbolic visual embedding learning for zero-shot recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9273–9281.

[10] N. L. Prasanna, R. Vaishnavi, V. P. Lakshmi, V. Dakshayani, and T. Keerthana, "Multi label classification for an image using convolutional neural networks," *Int. J. Comput. Sci. Mobile Comput.*, vol. 10, pp. 1–9, Aug. 2021, doi: 10.47760/ijcsmc.2021.v10i07.001.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25. Red Hook, NY, USA: Curran Associates, 2012, pp. 84–90.

[12] Y. Liu, J. Guo, D. Cai, and X. He, "Attribute attention for semantic disambiguation in zero-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6698–6707.

[13] S. Ulas, *Felsefe Sozlugu*, 13th ed. Ankara, Turkey: Bilim ve Sanat Yayinlari, 2002.

[14] T. Berners-Lee, J. Hendler, and O. Lassila, *The Semantic Web*, vol. 284. New York, NY, USA: Scientific American, 2001, pp. 34–43.

[15] G. Falquet and C. Métral, "Ontologies in urban development projects," in *Proc. Adv. Inf. Knowl. Process.*, 2011, pp. 189–196.

[16] M. Sir, Z. Bradac, and P. Fiedler, "Ontology versus database," in *Proc. 13th IFAC IEEE Conf. Program. Devices Embedded Syst.*, vol. 48, Jan. 2015, pp. 220–225.

[17] C. Martinez-Cruz, I. J. Blanco, and M. A. Vila, "Ontologies versus relational databases: Are they so different? A comparison," *Artif. Intell. Rev.*, vol. 38, no. 4, pp. 271–290, Dec. 2012.

[18] 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. 2097–2106, doi: 10.1109/CVPR.2017.369.

[19] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," in *Proc. 6th Int. Semantic Web Conf. (ISWC)*, in Lecture Notes in Computer Science, vol. 4825. Cham, Switzerland: Springer, 2008, pp. 722–735.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

OMURHAN AVNI SOYSAL received the B.Sc. degree from the Department of Electrical and Electronics Engineering, Gazi University, Turkey, in 2004, and the M.Sc. degree from the Department of Computer Engineering, Başkent University, Turkey, in 2016. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, Ankara University, Turkey. He has been a Software Architect.

MEHMET SERDAR GUZEL received the B.S. and M.S. degrees from the Department of Computer Engineering, Ankara University, Ankara, Turkey, and the Ph.D. degree from the Department of Mechanical and Systems Engineering, Newcastle University, U.K., in 2012. From 2006 to 2012, he was a Research Assistant with Ankara University, where he was an Assistant Professor, from 2014 to 2019. His research interests include image processing, software development, control theory, and robotics.

MEHMET DIKMEN received the B.Sc. degree from the Department of Computer Engineering, Başkent University, Turkey, in 2003, the M.Sc. degree from the Department of Electrical and Electronics Engineering, Middle East Technical University, Turkey, in 2006, and the Ph.D. degree from the Department of Geodetic and Geographic Information Technologies, Middle East Technical University, Turkey, in 2014. He has been an Assistant Professor with the Department of Computer Engineering, Başkent University. His research interests include remote sensing, machine learning, computer software, and computer-informatics science and engineering.

GAZI ERKAN BOSTANCI received the B.Sc. and M.Sc. degrees in real-time battlefield simulation from the Department of Computer Engineering, Ankara University, Turkey, in 2007 and 2009, respectively, and the Ph.D. degree from the School of Computer Science and Electronic Engineering, University of Essex, U.K., in 2014. He joined the Department of Computer Engineering, Ankara University, as a Research Assistant. His research interests include different yet closely related aspects of computer science from image processing, computer vision, graphics to artificial intelligence and fuzzy logic, mathematical modeling, and statistical analysis. He has been involved in technical committees of several conferences and organizing an international conference for several years as well as acting as a reviewer of various journals.