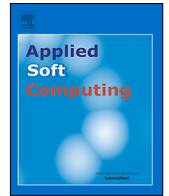




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Towards an effective model for lung disease classification Using Dense Capsule Nets for early classification of lung diseases

Faizan Karim^a, Munam Ali Shah^a, Hasan Ali Khattak^{b,*}, Zoobia Ameer^c, Umar Shoaib^d,
Hafiz Tayyab Rauf^e, Fadi Al-Turjman^f

^a COMSATS University Islamabad, Pakistan

^b National University of Sciences and Technology (NUST), Islamabad, Pakistan

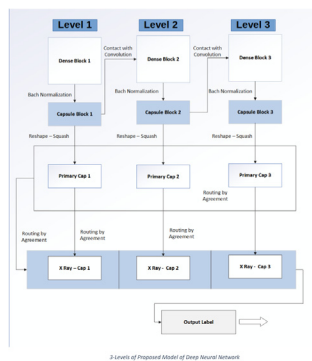
^c Shaheed Benazir Bhutto Women University Peshawar, Pakistan

^d Department of Computer Science, University of Gujrat, Gujrat, Pakistan

^e Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent, United Kingdom

^f Research centre for AI and IoT, Near East University, Nicosia, Mersin 10, Turkey

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 26 March 2021

Received in revised form 27 April 2022

Accepted 24 May 2022

Available online 31 May 2022

Keywords:

Capsule Networks

Deep learning

convolutional Neural Network

Chest X-ray

Lungs X-ray

Structural imaging

ABSTRACT

Machine Learning and computer vision have been the frontiers of the war against the COVID-19 Pandemic. Radiology has vastly improved the diagnosis of diseases, especially lung diseases, through the early assessment of key disease factors. Chest X-rays have thus become among the commonly used radiological tests to detect and diagnose many lung diseases. However, the discovery of lung disease through X-rays is a significantly challenging task depending on the availability of skilled radiologists. There has been a recent increase in attention to the design of Convolution Neural Networks (CNN) models for lung disease classification. A considerable amount of training dataset is required for CNN to work, but the problem is that it cannot handle translation and rotation correctly as input. The recently proposed Capsule Networks (referred to as CapsNets) are new automated learning architecture that aims to overcome the shortcomings in CNN. CapsNets are vital for rotation and complex translation. They require much less training information, which applies to the processing of data sets from medical images, including radiological images of the chest X-rays. In this research, the adoption and integration of CapsNets into the problem of chest X-ray classification have been explored. The aim is to design a deep model using CapsNet that increases the accuracy of the classification problem involved. We have used convolution blocks that take input images and generate convolution layers used as input to capsule block. There are 12 capsule layers operated, and the output of each capsule is used as an input

* Corresponding author.

E-mail addresses: faizankarim2@gmail.com (F. Karim), mshah@comsats.edu.pk (M.A. Shah), hasan.alikhattak@seecs.edu.pk (H.A. Khattak), zoobia.ameer@sbbwu.edu.pk (Z. Ameer), umar.shoaib@uog.edu.pk (U. Shoaib), hafiztayyabrauf093@gmail.com (H.T. Rauf), fadi.alturjman@neu.edu.tr (F. Al-Turjman).

to the next convolution block. The process is repeated for all blocks. The experimental results show that the proposed architecture yields better results when compared with the existing CNN techniques by achieving a better area under the curve (AUC) average. Furthermore, DNet checks the best performance in the ChestXray-14 data set on traditional CNN, and it is validated that DNet performs better with a higher level of total depth.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Pandemic, a word not to be taken or used lightly is a critical situation where an epidemic of a contagious disease eventually spreads across a large area of the world. History has witnessed various pandemics such as smallpox and tuberculosis. If misused or misunderstood, a Pandemic can cause severe issues to the infrastructures and systems that stop functioning and collapse in due course towards human suffering, even deaths.

Corona Virus (COVID-19) is a highly contagious disease that started spreading through Wuhan (China) in late December [1]. As of September 2021, the number of coronavirus cases has crossed 226 million with over 4.6 million reported deaths.¹ Since then, this infectious disease has become the focus of attention, and researchers are looking for different techniques and technology to prevent and control the spread of COVID-19 [2].

In Late December 2019, several patients diagnosed with a severe stage of pneumonia were reported in the province of Wuhan, China. In early January of this year, the organism causing the virus was identified as novel coronavirus, thus named 2019-nCoV and now generally known as COVID-19 [1]. As of September 2021, the number of coronavirus cases has crossed 226 million with over 4.6 million reported deaths. With both symptomatic as well as asymptomatic, patients have been actively playing a significant role in airborne as well as the contact-based spread of the COVID19 disease [3]. Given the spread mechanism so common and unavoidable, the nature of spread has caused the research community to find systems for healthcare to deal with the deadly virus in a safe manner [2].

The weaknesses of the healthcare systems which have come to light could be easily mitigated using the state of the art tools to collect and analyze data coming from various components of the overall government and local healthcare systems [4]. These efforts are directed towards effectively predicting the spread, thus identifying patients as soon as possible to provide better care for the patients with comorbidity [5]. But also these can help governments and infrastructures put in place systems that are better able to cope with giving insights on improvements for future such events.

Different approaches have been witnessed as a reaction from other nations resulting in different outcomes. We have the case of mass testing in South Korea [6], where they were able to isolate the positive patients through proactive testing and tracking contact of each citizen. This continuous testing not only enabled them to have a bigger, relatively complete picture of the contagion and its spread but also possible hot spots [3].

This way, South Korea was able to fully control the spread of the COVID19 virus and return to a better situation than the rest of the world [7]. An almost similar response has been witnessed by other nations such as Taiwan, which also adopted a proactive approach towards testing as well as using extensive data analysis and using machine learning [8].

Similar has been the case for countries that are much more advanced and have better systems. On the same side, we have

countries such as Italy where the system has not been able to cope with the COVID-19 outbreak. One of the reasons was the delay in enforcing the social distancing policy to minimize the spread of disease. The lack of knowledge about the comorbidity has played a significant role in the increase in the number of overall cases as well as massive loss of human life [9]. Taking the example of Taiwan and South Korea, the Italian health care system was required to make decisions not only at the government level but also to take adequate measures in their healthcare system to tackle the massive influx of patients with already known factors in comorbidity were at risk. The healthcare providers were required to make uninformed and on-the-spot decisions about choosing the proper medications as well as choosing the candidate patients who should be on a priority list for taking benefit from the overburdened healthcare system [10,11].

The likelihood of catching the COVID-19 virus can be assessed using health care data so that patients could be flagged for the healthcare operators to evaluate the comorbidity. It is also essential to identify the main hotspots of infected people o control the spread of COVID-19. Identifying the hotspots and infection rates can lead to improved health care responses. This has proven to be very helpful, especially in Wuhan, China, where a patient's mobile phone's location data was used to limit the spread using AI-based methods [12,13].

COVID-19 is a very systematic disease where lethal and severe lung infection complications lead to the overall failure of critical organs. Though there are no effective vaccines for COVID-19 yet, assistive therapeutic procedures improve a patient's condition over time. Machine Learning is among valuable tools for diagnosis. Similarly, the proposed system will be able to assess and give information about the patient survival and their conditions [7,14].

During the last few years, complex computer vision tasks have evolved in different stages, from categorizing unique images to classifying/discovering/dividing multiple situations and multiples categories into more complex cognitive functions that include understanding and describing situations in pictures or videos the rapid and significant improvement of performance is partly driven by public access to significantly large data sets with annotations for the high-quality images. In contrast, there are enough Data sets already publicly available in medical imaging [15].

Traditional means of commenting on natural images, for example, crowd sources, cannot be applied to medical images because such a sensitive job often demands years of working with domain knowledge and professional field training. Overcoming this drawback is an important milestone when we consider the significance of the quality image datasets in the medical imaging community [16,17]. Furthermore, primary radio-logical datasets (such as medical image datasets) have accumulated in many Picture Archiving and Communication System (PACS) hospitals for decades. The main challenge is how to convert those vacations' Potential radiological data into the form of automatic learn-able data.

However, greater depth leads to the problem of deteriorating disappearance. This problems was resolved by ResNets [17], FractalNets [18] when adding connections from primary layers to subsequent layer. DenseNets, another such dataset by Huang et al. simplifies how omission connections are added [19]. This contribution improves and supports the dense connections among

¹ COVID Stats <https://www.worldometers.info/coronavirus/>.

the other connected layers. The addition of these dense connections leads to fewer parameters than conventional CNN. Another advantage of the sequence of these maps is the advantage of a smoother flow of gradient across the neural network, allowing the Machine Learning experts to train using deeper neural networks. Sabour et al. [20] have given that CNN has done well in many machine learning-based computer vision scenarios with some of the underlying drawbacks. Among the most prominent is that CNN's have not proven to be the strongest for the relevant transformations. Any slight change in the object's position causes CNN to change its prediction.

Although this problem can be reduced to a certain extent by increasing data during training, this does not make the network robust for any new mode or format present in the test data. Another essential drawback is that public CNNs do not consider spatial relationships between things in a picture while deciding. In short, CNN networks use only the presence of a couple of particular localized objects in the whole decision-making image processing task. In contrast, the object's context spatial in reality is equally significant.

The main reason is the final assembly process in the neural network, which gives significance to the existence of properties while ignoring the topical properties information used mainly for parameter reduction with the growth of the network. For overcoming such constraints, Sabour et al. [20] proposed basic yet significant network structure called CapsNet. CapsNet's trained model keeps the weighted information at the vector level rather than numerical form (as found normally in the most basic and simple neural networks). In this case, these groups of neurons have to be called together, thus getting the name capsule for the Patient and others entities. The concept of guidance was used by agreement and crushing on a stratified basis to achieve advanced precision in the ChestXray dataset. Data collection and detection of overlapping numbers in a rather better way as compared to traditional methods which are leveraging reconstructional organizations [21].

Though very powerful, CapsNets have so much improvement to be done in the complexity, Sabour et al. because the network currently uses only one layer of twisting and capsule nets. DenseNets, on the other hand, can achieve higher performance than the CapsNet by aggregating these features. In this work, we extend this idea from Sequence functions through the layers of DenseNets [19] because it can learn a variety of properties, which otherwise may require a somewhat deeper neural network, this work can be considered as an extension to previous work.

CNN has a remarkable ability to learn and classify the images without any knowledge related to the problem, making it an adaptable way to classify images. Neural networks and CNN networks are used with different pre-processing steps, including increased data (Augmentation). It has been shown that CNN networks without any pre-processing outweigh the other methods in classifying radiological images [16,22,23]. Despite the success of CNN in overcoming many ways to handle images, they still suffer from some flaws. For example, they are not transformed and do not consider the spatial relationships within the image. To improve its outcome, CNNs must have training data of all types of rotation and transformation.

CNN has performed better in many computer vision and learning tasks, but it has some flaws highlighted by Sabour et al. [20]. One is that CNN networks are not decisive for relevant transformations, meaning that a slight change in the object's position causes CNN to change its expectations. Although this problem can be reduced to some extent by increasing data during training, this does not make the network robust for any new situation or format that might be present in the test data. Another critical flaw is that CNN's networks do not consider spatial relationships

between objects in the picture when making a decision. Simply put, CNN networks use only the presence of particular local objects in a decision-making image, while the spatial context of objects, in reality, is equally important. The reason is mainly the pooling operation in the network, which gives importance to the existence of properties and ignores the spatial information in images, which is the primary purpose of reducing the parameters with the network growth.

To overcome these constraints, Sabour et al. [20] a technique Capsule Networks CapsNet is proposed. This model stores the information at the vector level instead of the numerical (as in simple neural networks). This group of neurons is called together as a capsule. They used the concept of routing by agreement and layer-based squashing to achieve advanced precision in the Mixed National Institute of Standards and Technology (MNIST) dataset and better expose the overlapping numbers by reorganizing the reconstruction. The CapsNets system is very efficient; however, at the same time, there is room for improvement in complexity since the authors did not use the layers or the depth of assembly, where the network currently uses only one layer of capsules.

1.1. Motivation

In our work, the motivation is to improve CapsNet performance in complex data set Chest X-ray to citewang2017chestx. In addition, we follow the intuition behind DCNET++ [24] to design a modified decoding network with dense sequential layers. In addition, we presented the intuition behind the selection of high-density networks and then improved them to improve performance in complex data sets. Capsule networks are powerful enough to work on images that have complex data sets. Our goal is to enable capsules using DenseNet. We are trying to customize the capsule network into frames, explained in the following sub-sections.

1.2. Contributions

We proposed modification of the capsule network, Dense Capsule Network (DNet), which replaces the standard embedding layers in CapsNet with a tightly connected torsion. The addition of direct links between two successive layers helps to learn better maps of the features, which in turn helps to form better quality capsules [24,25]. The effectiveness of this proposed model is reflected in the results we achieved, as our model with depth levels five has produced better results. Furthermore, we trained our model with seven levels of depth and compared the results with our five levels trained model, which need a higher computational environment because of its complexity but performed better than the five levels of DNet.

1.3. Organization

The rest of the paper has been organized as follows: Section 2 highlights state of the art. The proposed model is explained in Section 3. Simulation and its results are discussed in Section 4. Finally, the results and overall discussion is explained in Section 5 followed by Section 6 where the conclusive remarks and future directions are presented.

2. Related work

Computer-aided diagnosis (CADx) and detection (CADE) have always been the main areas of research in medical image processing [26]. In past years, deep learning models have begun to overcome traditional methods of statistical learning in different tasks, such as automated skin lesion classification [27], detection

of liver injury, and detection of pathological results [28]. However, current CAde methods generally target a particular disease or injury, such as pulmonary nodules, benign tumors, or lymph nodes [29]. A recent and prominent exception was introduced by Wang et al. [22], as the chest x-ray data collection was presented on a large scale Processing of associated radiological images and reports (extracted from the PACS institutional database) using natural language processing (NLP) techniques. The publicly available dataset contains 112000 chest radiographs with a front display of 30,805 unique patients.

In the era of machine learning, deep learning, and computer vision research efforts in the construction of many annotated data sets [30–35] with different characteristics to play an important role to be dispensed within the best definition of future problems, And then, potential technological advances. In particular, here, we focus on joint learning and the relationship between images (x-ray) and text (X-ray reports). The work of generating the titles of previous images [36–38] uses data sets Flickr8K, Flickr30K [39] and MS COCO [32] containing 123,000, 31,000, and 8000 images respectively, and each image is described in five sentences by (AMT) Amazon Mechanical Turk.

The text usually describes the scorer's attention to objects and activities that occur in a direct image. Detect-based neural network trunking ImageNet pre-training (CNN) is used regionally to analyze input images and create a list of attributes or "high-level concepts with visual ground connection" (including objects, Actions, scenes, etc.) in [37,38]. The VQA requires a more detailed analysis and a complex conclusion in the image content to answer the associated natural language questions. A new dataset containing 250,000 raw images, 760,000 questions, and 10 million text responses are presented [34]. In the face of this new challenge. In addition, databases such as "Flickr30k entities" [33], "Visual7W" [35] and "optical genome" [40,41] are presented (as shown in 94,000 images and 4100,000 address areas) to construct and learn spatially. Dense and increasingly difficult Deep symbolical links between text descriptions and image areas through grounding at the object level. Although one can argue that there is a significant similarity between generating image labels, visual responses to questions, and diagnosis of disease-based imaging [42,43]

Three factors make the diagnosis based on large-scale medical imaging 1, you cannot get general, open signs of anatomy, and level of pathology through many sources, such as AMT, which an Unpredictable for medically untrained practitioners. Therefore, we exploited the extraction of common thoracic pathology markers (possibly several images) of chest X-ray images attached to images using natural language processing techniques (NLP). Radiologists tend to write more logical, abstract logical sentences than simple-textual texts. 2, the spatial dimensions of radiology are generally 2000 x 3000 pixels. Regions of local pathogens may show significantly variable volumes or extensions but are very small compared to the full-scale image.

In Fig. 1, we have illustrated eight examples using the pathological result of actual images, which are often much smaller and hardest for detection. We formulate and verify the classification of multi-tag images and the framework of the disease site with little supervision to meet this difficulty. 3, so far, all the image translation techniques and VQA in com The computer vision depends heavily on the already-trained and well-rehearsed CNN models that work well in many object classes and serve as a good baseline for a more precise configuration model. However, this position does not apply to diagnostic medical imaging.

In synthetic vision, deep learning demonstrated its ability to accurately classify images [44–47]. In addition, the field of medical image processing deeply explores deep learning. However, a major medical problem is the availability of large data sets

with a reliable explanation of the underlying truth. Two larger sets of x-ray data are available recently: the CXR data collection from Open-i [48] and ChestX-ray14 from the National Institutes of Health (NIH) clinical center [22]. Due to its size, ChestX-ray14, which consists of 112200 front-end images of CXR from 30,805 unique patients, has attracted considerable interest in the deep learning community. Activated by Wang et al. [8] Through the use of neural networks (CNN) of computer vision, many search groups began to process the CNN application for the CXR classification. In [49], Yao et al. A combination of CNN and a frequent neural network was introduced to exploit the tag dependencies. Using the backbone of CNN, they used the DenseNet [19] model, which was fully adapted and trained in X-ray data. Recently, Rajpurkar et al. [16] It was proposed to transfer accurate education, using DenseNet-121 [19] and raise the AUC results in ChestX-ray14 for a higher label-label classification.

The proposed architectures using convolution techniques have increased significantly due to increased computational power. Literature for CNN is vast. CNN tries to learn the hierarchy from the bottom to the top, where the lower layers learn essential functions such as edges and upper layers learn sophisticated features by combining these low-level features [50]. Although deeper networks have improved performance, their training is much more difficult because of the large number of learn-able parameters [51]. The proposed modern structures aim to enhance performance while improving learn-able parameters in number at the same time [52]. The Highway-Network [53] was the first proposed structure in this direction to form a deeper network with a large number of layers. They added by-passing paths to the model to train efficiently. The ResNet [17] model improves training by adding residual connections. Another of these networks was proposed by Huang et al. [19]. He devised a new method to add the skip links by inserting links from the first layers into deeper layers and naming them a dense block.

Capsules are a group of neurons representing the properties of several entities present in the image. An image can have several properties that can be captured as position, size, and texture. Capsules use routing-by-agreement where the output is sent to all final capsules. Each capsule predicts the central capsule, which is then compared to the actual production of the central capsule. Suppose the outputs coincide, the coupling coefficient between the capsules increases. Capsule networks [20] have recently been introduced to overcome the disadvantages of CNN networks discussed. Dcnet++[24] an architecture is proposed which is based on capsule network and uses the idea of Densely connected Networks and shows improvement in results of MNIST data set as compared to the state of the art results of Hinton et al. [54].

In recent years, there has been significant growth in attention to deep learning methodologies with improved accuracy of X-ray classification of the chest. In particular, the architecture that uses the Convolution Neural Network (CNN) has produced better solutions for image classification functions and object recognition [22]. But CNN has some disadvantages: (1) it is not equivalent in translation, (2) it does not consider the spatial characteristics in the images because of the maximum aggregation, and it only learns numeric values. Capsules are a group of neurons that represent the properties of many of the entities in the picture. There may be many characteristics of the captured image, such as position, size, and texture, because the capsule grid learns vector values [20]. The results obtained by traditional CNN networks are not suitable for clinical trials.

3. Proposed solution

Inspired from dense connectivity of layers block carried out by Huang et al. [19] and dense convolution network for capsule network Dcnet++ [24].

Table 1
Related Work.

textbfTechniques	Achievements	Data Set	Limitations(s)	Results
DenseNet-121. [16]	Achieved better accuracy than [22].	Chest x-ray-14	Only flipped data horizontal.	Average AUC: 0.8413
AlexNet, GoogLeNet, VGGNet-16, ResNet-50. [22]	Presented a new chest X-ray database, namely "ChestX-ray8", Did initial Comparison of different CNN models.	Chest x-ray-8	Used only 50 layer deep CNN[Resnet50]. Could have achieved more accuracy by going deeper.	Avg AUC:0.6962
DenseNet-121 [52]	Used patient Wise split of Data Set.	Chest x-ray-14	Used larger data set with same convolution layers as [16].	Avg AUC: 0.841
CNN+LSTM[10] [23]	Proposed CNN+LSTM combined architecture to use history data along with images.	Chest x-ray-14	Did not improved results on images itself.	Avg AUCs = 0.992,0.722
Dynamic Routing on Deep Neural Network. [25]	Proposed a Dynamic routing Connection between capsule layers.	Chest x-ray-14	Improve disease localization by integrating .comlocation information provided in the dataset.	Avg AUC = 0.775
Capsule Networks [20].	Proposed a new approach which overcomes the Drawbacks in CNNs.	MNIST	Better approach but it probably requires a lot more small insights before it can out-perform a highly developed technology.	Accuracy:99.22
DCnet++ [24].	Achieve better accuracy than [20] on MNIST dataset , by proposing more dense architecture of Capsule Network.	MNIST	Resources limitation, could have achieved more accuracy by proposing more deeper architecture.	Accuracy:99.75

We proposed modification in the capsule network, Dense Capsule Network (DNet), which replaces the standard embedding layers in CapsNet with a tightly connected torsion. The formation of direct links between two successive layers allows the system to comprehend the maps of features better. Hence, this enables the system to form better quality capsules. The proposed system's effectiveness is reflected in the results we achieved, as our model with depth levels five has produced better results than [22]. Furthermore, we trained our model with seven levels of depth and compared the results with our five levels trained model, Which needed a higher computational environment because of its complexity but performed better than the five levels of DNet.

3.1. Dense block

As, the maps of features – comprehended by the first link CapsNet baseline model – recognize rudimentary features only. Thus, the maps mentioned above may not be sufficient to create capsules for complex data sets. Therefore, we attempted to increase the torsion layers up to two and eight in the first layer for improvement. We perceived that this methodology did not yield better results, as detailed in Table 1. In DNet, the capsule network was modified to establish a deeper structure, which consisted of eight layers.

Skipping connections form the basis of a dense subnet. The layers are arranged in a sequence so that as different layers progress, they combine to make the final wrap layer. This methodology provides better results – in the form of good gradient flow – compared to directly stacked torsion layers. The input sample consists of eight levels of combinations. These rounding levels create their own thirty-two distinct new maps and a sequence of properties of all previous attributes. Layers that produce the maps feature 257 (the input image is included). These characteristics are varied. The capsule layer takes the maps as an input and applies a nine-by-nine wrapping with one step to the distinctive maps obtained: capsules, essential capsules the work from, and

others. [20] focuses primarily on equivalence rather than the consistency we see fully.

We did not utilize the average and assembly layers employed in DenseNet (DNet). Hence, this resulted in the loss of spatial information. It is pivotal to keep in mind that Sabour et al. [20] created initial capsules of 256 distinct maps produced by the same level of complexity.

From torsion, DNet central capsules are created by integrating all properties as shown in Fig. 1. Different levels of complexity further improve the classification. The feature mentioned above maps function as thirty-two 8D capsules. These capsules are then advanced to squash activity. Then, the training layer is succeeded by the routing algorithm. The final of the ten categories (numbers) is obtained. Next, for the above final, 16D capsules are created. These generate 10D output vectors from a single-style heat. A network of conventional capsules inspired by the intensive communication carried out by Huang et al.. The reconstruction model was modified for the capsule network. The decoder is a four-layered model with shine properties. These properties of the first and second layer take the XrayCaps layer as an input (masked by the exit sign during training), which result in a superior rebuild. The size of this image is greater than 32 x 32, then the number of neurons is switched from 512 to 600 and 1024 to 1200.

3.2. Capsule block

Adding omission connections to implications was not enough to improve performance. This may be due to simple essential capsules that are not sufficient to encode information in such complex images.

DNet model activates a part of the picture. We envision which portion of the picture is operated by a DNet 8D core capsule through forwarding propagation. Note that a tiny spatial territory creates every first capsule of the image, which then works together to decide, but this is not the case as it was pretty complex to overcome this. Hence, we have implemented a new method

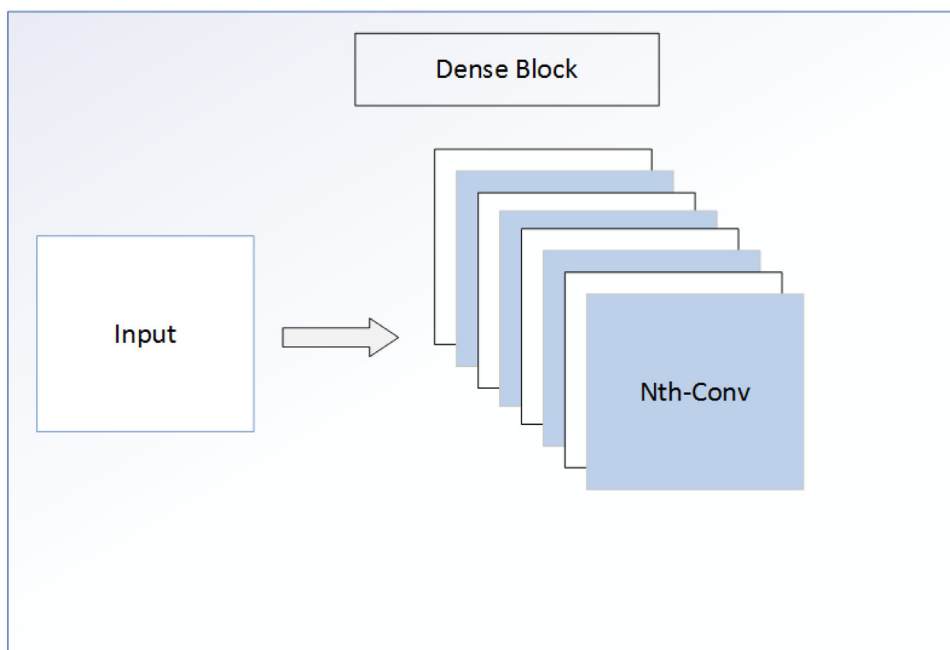


Fig. 1. Dense block.

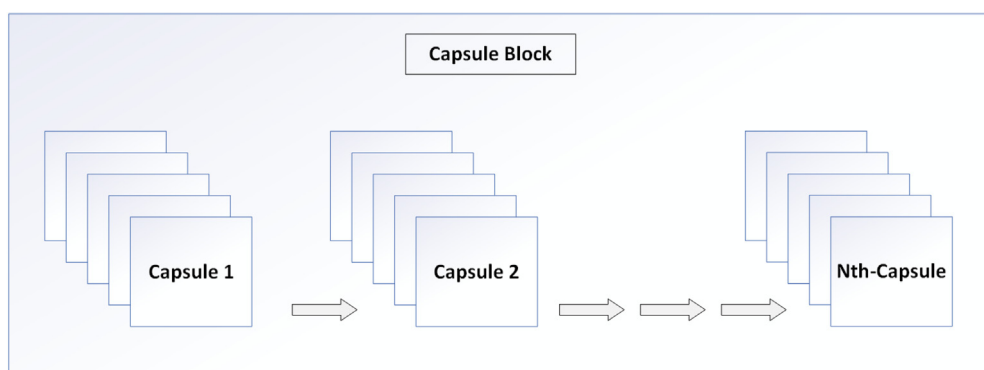


Fig. 2. Capsule block.

to create essential capsules that carry the information of several picture measurements, thus varying the capsules. This assisted in establishing links between the primary capsules of various levels of images. Considering only one level, this will be like picture activation operations in DNet and CapsNet from the baseline. Besides, the activation process will not have a spread external to the square zone of the capsule network.

The activated zone is spread in DNet because of the “same” padding combinations in the high-intensity torsion layers. A set of basic models of CapsNet (from Sabour et al. [20]) provides accuracy in the chest-X-ray-8 data set, possibly because the patches are utilized as inputs, which actuates a more extensive spatial zone to create each capsule. Even though it accomplishes a sensibly better resolution in a solitary capsule model, a set of seven models of this sort significantly increases the number of parameters that can be diminished for little size pictures. We center around lessening the number of parameters used to model information and learning better data by making numerous levels of capsules. A three-layer DNet module with 13.4M parameters checks a test resolution. An elaborate pipeline of the DNet model is used to learn the chest-x-ray-8 dataset.

It is a progressive model where a DNet form is made, and its proxy characterization is utilized as an input to the following

DNet. As a result, this produces a feed representation for the subsequent layer of DNet. There are twelve capsules per DNet. As shown in Fig. 2, Repentance is applied in steps (size nine by 9 and 2 steps) which diminishes the size of the picture forwarded to the following levels.

This is also similar to brainwork that separates data into channels. For instance, there are independent ways for high and low spatial frequency content and color data. Besides the three layers above of XrayCaps (output), we made an output layer of XrayCaps in addition to a three-layer series directive of PrimaryCaps. The purpose behind adding this other level is to permit the model to grasp the properties collected for several levels of capsules. Note that in the case of the simple accumulation of XrayCaps and common genetic reproduction. Others are dominated by the losses of the last level of PrimaryCaps. This prompts low-quality learning and initial levels pretending as basic repentance classes. Accordingly, the model was prepared together to prevent any poor learning, yet four-layer losses were published separately.

XrayCaps generated from the PrimaryCaps series play an essential role in reconstruction, an additional effect for different layers of capsules. During the test, the four layers of XrayCaps are combined to form the final 54D capsule for each of the ten categories, and reconstructions of one image channel were created

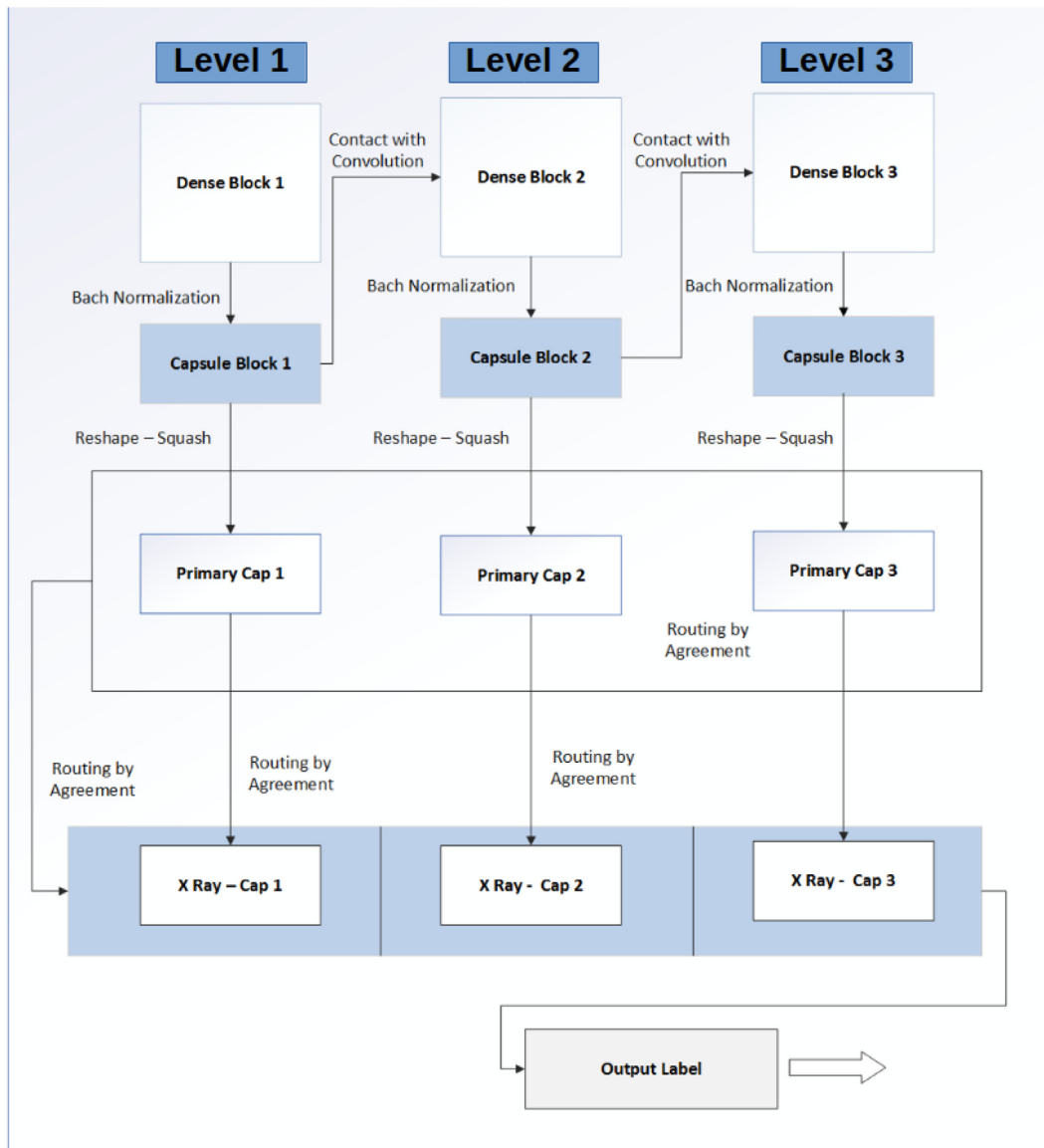


Fig. 3. 3-levels of proposed model.

using these four capsules. The Xray-8 reconstruction of the chest was not very good, which we believe is due to the dominance of the background noise in the samples and the presence of complex information that the decoder is not strong enough to reproduce. Interestingly, we observed the effect of noise in XrayCaps from different levels on the reconstruction output on the chest-Xray-8 data developed by subtracting 0.2 from each x-ray one by one in XrayCaps 54D. It was noted that the impact of reconstruction processes falls from level 1 to the last level of capsules.

3.3. DNet

We have created convolution blocks as shown in Fig. 4, first convolution block takes an input image of 1080×1080 in size, the input image and all feature maps generated by eight convolution layers are then be used as an input to capsule block after Batch normalization, which has 12 capsule layer and obtained capsules output is used as an input for second convolution block than this same process repeated for all convolution blocks.

The feature maps obtained from the capsule blocks after squash function used PrimaryCaps will then be used as final

XrayCaps carried out by “Routing by agreement” algorithm [20], PrimaryCaps which are feature maps of capsule blocks, are concatenated and used as one of the final XrayCaps. The 3 level of DNet is shown in Fig. 3 and with nth level is shown in 4

The Data set we used is chest x-rays data set recently published by Wang et al. [22]. We will further do more experiments by increasing the number of dense blocks and capsule blocks and try to achieve improvements in results.

Deep learning methods include multilayered processing with less time and more satisfactory performance. Sub-layers give a better result through the use of CNN and autoencoders. With the increase in the number of automatic encoders, image resolution increases. Such as increasing the number of subsampling, you also get better [17,19,50]. So we experimented with different levels of depth and achieved results improvement.

4. Simulation results

In this work we have performed different experiments of our proposed model DNet, and we have also experimented with Resnet50 model, which was implemented by [22] on chest x-ray-14 data set. And we have also implemented the architecture

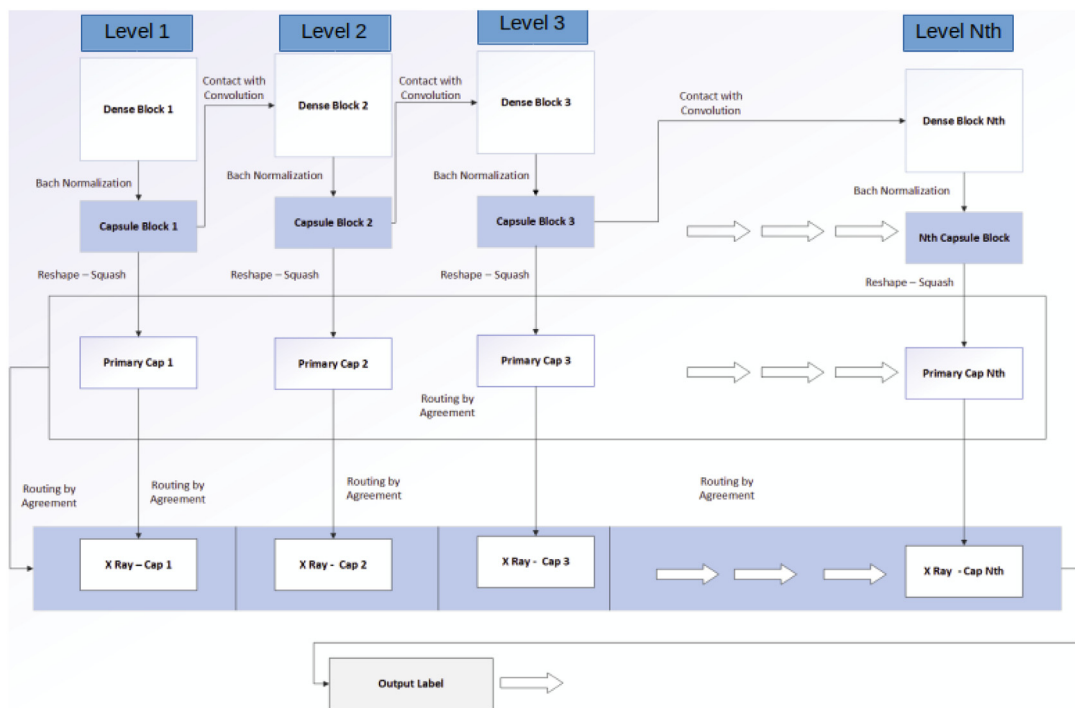


Fig. 4. Architecture of the proposed DNet model.

proposed by [25] and compared both papers' results with our model's results.

4.1. Data

We utilize the chest-x-ray set14 set by Wang et al. (2017). This set contains 112,120 x-rays in a frontal view of 30,805 distinct patients. Wang et al. [22] utilized automatic extraction methods in radiology reports and recorded each picture with up to 14 varying signs of thoracic pathology. Some sample pictures are shown below in Fig. 5 and the frequency of examples are shown in Fig. 6

The main block of a convolution neural network (CNN) is the torsion process. In the deep network, the essential work of the twist (and weights) is basically "detecting" the basic features. When we train a deep network, we adjust the weights of our CNN taps to detect or "activate" certain features in the image. It is essential to understand that top-level functions combine lower-level functions as a weighted addition: the activation of a previous layer is multiplied by the weights following the added layer trap before moving to nonlinear activation. There is no place in this flow of information that is the relationship between the features taken into account. Therefore, we can say that the primary failure of CNN is that it does not carry any information about the relative relations between characteristics. This is just a failure in the central design of CNN because it is based on the basic torsion process applied to numerical values.

Hinton et al. mention that to identify the image correctly, it is essential to maintain hierarchical relationships between the picture's properties. When these relationships are integrated into the network, it is elementary to understand that what you see is just another view of what you have seen before since it is not only based on independent properties; now, it is joining these properties to form a complete representation of "knowledge". The key to this representation is a richer feature using vectors rather than climbers. Capsules provide the basis for building a better model for the relationships above within the network in the deep learning domain.

The change in capsule networks can be divided simply by using vectors instead of scaling. It helps vectors because it helps us write down more information, not just any kind of relational and relational information. Imagine that instead of taking the standard activation of a characteristic property only, we consider its vector to contain something such as that containing [probability, orientation, size]. An original numerical version like the ones on the left can work. Detects the face, although the eyes and lips are prominent features for a face (95 per cent chance)!

This wealth presents new challenges: a more significant number means more calculation and complexity. On the other hand, because of the wealthier vector information, capsules notice that different properties' sizes are different and generate less probability of detection. With the rich collection of information supplied by capsules, they show great potential to become the main engine of our deep future network.

CNN is a black box learning algorithm, The best part of CNN is that it learns features with itself like in its first layer it will learn edges in the next layer may be it learns derivatives. In its higher layer, it will learn more elevated levels of feature maps like shapes [50]. So it did not need any noise reduction technique before feeding data to CNN. The most used pre-processing methods for CNN are data augmentation, like a flip of data rotation of data as CNN in transnational invariant that is why for better training it should be trained with all variation in data. Still, capsule networks overcome this drawback of CNN [20] that is why we have experimented with our model without data augmentation.

4.2. Experiment environment

We have performed this experiment on a machine with 8 CPUs 24 GB of RAM. The GPU we used for the experiment is Tesla Nvidia P100 which has 16 GB of memory for 1st and 2nd experiments. While for the 3rd experiment, we needed a more significant memory, so we used 2 GPUs through parallelism the training with the same number of CPUs but more significant RAM of 52 GB (see Table 2).

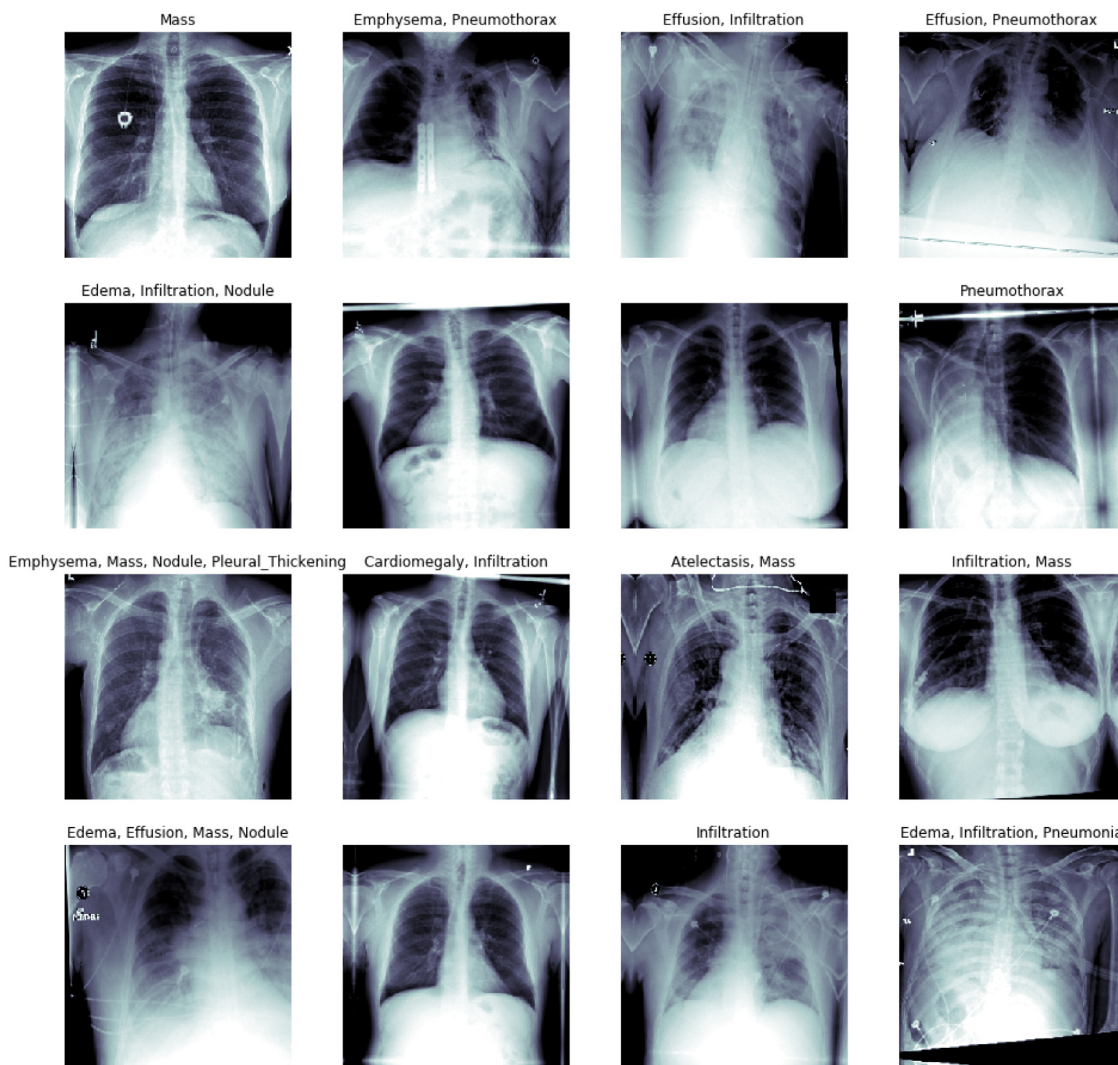


Fig. 5. Data samples images.

Table 2
Experimental Environments.

	CPUs	Ram Memory	Libraries	GPU(s)
Exp no. 1	8 x skylake	24 GB	Python	NvidiaTeslaP100 x 1 = 16GB
Exp no. 2	8 x skylake	24 GB	Python	NvidiaTeslaP100 x 1 = 16GB
Exp no. 3	8 x skylake	52 GB	Python	NvidiaTeslaP100 x 2 = 32GB

4.3. Experiments

We have done three experiments with different depth levels of the proposed model in 4, one with three levels of depth 2nd one with five levels of depth, and 3rd one is seven levels of depth.

4.3.1. ResNet50

ResNet is the short name for Red Residual. As shown in the network name, the new terms introduced by this network are the remaining learning. Several advances in image classification were due to deep subliminal neural networks. Several image recognition tasks can leverage the benefits of deep learning and machine learning models. Thus to enhance the accuracy, research has been done in this domain to improve the overall results. When the deeper networks begin to converge, the degradation problem is exposed: by increasing the depth of the grid, accuracy becomes saturated and then degrades quickly. Let us take a

shallow network and its deep counterpart by adding more layers. A shallow grid can replace the first layers of the deeper model, and the remaining layers can function simply as an identity function. In the deepest grid, the additional layers are closer to the assignment to the shallow meter portion and reduce the error by a large margin. In the worst case, both the surface network and its deepest forms must give the same precision. In the case of the reward scenario, the deeper model should provide a better resolution than its more superficial counterpart. But the experiences with our current customers reveal that the deeper models do not work well. Therefore, the use of deeper networks corrupts the performance of the model.

The problem with going deeper in layers is the increase in complexity and saturation of results. Resnet50 attempts to resolve this issue using the Deep Learning Framework. Generally, an efficient way in profoundly adaptive neural networks is to stack and use the network to do a complex task. While doing

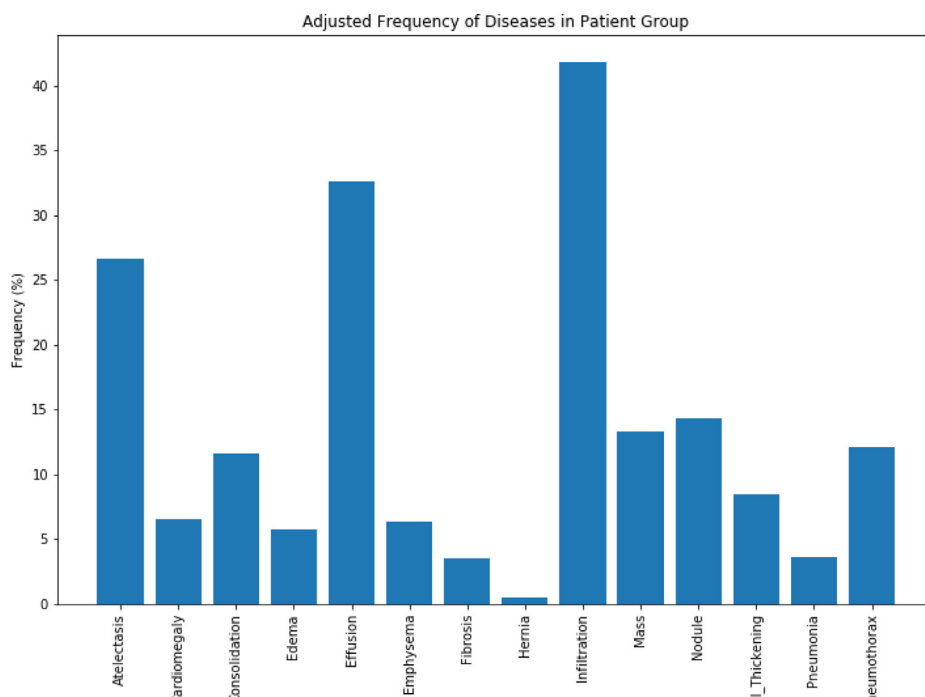


Fig. 6. Frequency of data.

so, the network tries to recognize and solve these several layers. In the remaining learning, instead of learning some properties, we try to learn something that remains. The rest can simply be understood as a characteristic of the inputs of that class. ResNet does this via the shortcut links (connect the input layer directly to Na layer n (n + x)). It has been proved that it is easier to train a network, and the problem is also solved by degrading precision ResNet50, a residual network of 50 layers. There are other types such as ResNet101 and ResNet152 too. The results we achieved by implementing ResNet50 are shown in Fig. 7.

4.3.2. Dynamic routing connection

In our networks, chest-Xrays are computed in advance and afterwards forwarded to a descending sampling block from Conv-Pool-Conv-Pool. The second layer has the size of step 7 and steps 2. The second assembly layer uses the maximum assembly of volume 3 and steps 2. After the top assembly layer, we utilized a layer of 1 and 2, and finally, we utilized a medium layer of sizes 2 and 2 before feeding a thick layer. Dynamic routing between dense layers Each dense mass consists of 8 layers of composite functions. Dynamic routing is used to update the one by one symbiotic layer. Dense blocks are used after sampling. The proposed dense pattern is tracked in [19] except for the one by one metaphysical layer. The dense layer comprises successive layers of composite functions that take the serial output created in the past layers. Each composite function comprises six successive batch normalization processes.

The network and dynamic orientation are included between 1 by one torsion layer connections. A dense block of dynamic routing is indicated. After the thick masses, we utilized a more prominent layer of step 9 and step 1. Afterwards, we utilized a medium assembly layer of size four and step 4. Then, we used a capsule layer that is fully connected to class classifications. The characteristic map shape was changed to the primary capsules in the entirely interconnected capsule layer. This task was done by taking eight maps of the properties of each pixel as a capsule. Afterwards, we direct the entire linked layer between the primary capsules and the capsule of the disease marker after the guidance

mechanism by agreement. Finally, we take the L2 standard for each vector in digital capsules as a number for each disease marker. Results achieved by this model is shown in 8.

4.3.3. Experiment : 1(DNet-3)

In our First Experiment, we have used 70000 images for training and 10000 images for testing without data augmentation. We have taken 14 classes from the dataset [22]. Image input size we used 1080×1080 , we conditioned the model for 500 epochs with a batch size of 256 images in the dataset. The loss function we used is margin loss [20], and optimizer Adam et al. [55]. The ROC Curve for the predicted test. First, we have done experiments with three levels of depth and achieved results shown in Fig. 9. The average of AUCs are shown in Table 3.

We detail the capability of our model in numerous depth levels and compare them with the Resnet50 [17] model actualized in [22] structure. The first test was performed utilizing the Nvidia Tesla P100 with 12 GB of RAM. We carry out all our models 50 to 100 times. The introductory learning rate was set at 0.001, and 0.9 decay rate with Adam [55] as the optimizer. We change the reproduction property to reduce the reconstruction loss according to the image's size so that the loss of margins does not dominate. To create our 2 sample test errors once we calculate each model. Three methods were utilized for all trials.

To make just comparisons, we essentially retain the DNet model parameters proposed after the PrimaryCaps layer, just like traditional CapsNet. The following details the corresponding implementation of the data set used. A collection of chest x-ray-14 manuscript data and data collection will have the training and test images of 80000 size each image 1024 by 1024 in size. We did not use any system to increase the data, and we repeated the experiment 3 times. DNet can identify the differences in data entry speeds compared to CapsNet, i.e. it is capable of achieving an accuracy test in chest x-ray-14 data set with 20 times fewer overall duplicates.

The step in the PrimaryCaps layer of the second level of DNet was modified from two to one. This modification was done to adjust the picture. No improvement was noticed in the execution

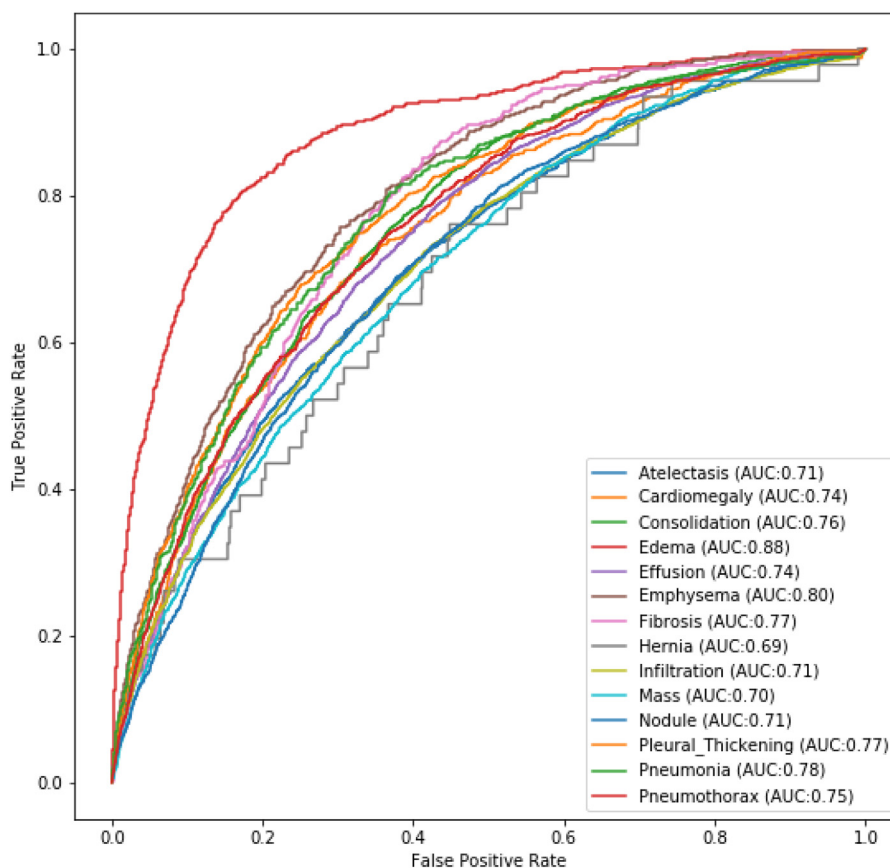


Fig. 7. Resnet50 ROC curves.

of DNet 3 levels in both chest x-ray-14, which is expected data because we capture fine properties at the rough level. The training set will gain each slight difference in writing a certain number. During the testing phase, a problem can occur due to these “thin” properties. More time needs to be invested in improvement. We discover that the loss of the modified decoder in DNet is significantly reduced, with the same loss multiplex as CapsNet. Also, we illustrate the comparison of the test of DNet, which we can undoubtedly conclude that DNet has a quicker convergence rate.

4.3.4. Experiment : 2(DNet-5)

In our 2nd experiment, we have utilized 70000 pictures for training and 10000 pictures for testing without data augmentation. We have taken 14 classes in the dataset [22]. Image input size we used 1080×1080 . We taught our model for 500 epochs with a batch size of 256 images. The loss function we used is margin loss [20], and optimizer Adam [55]. The ROC Curve for the predicted test. First, we have done experiments with three levels of depth and achieved results shown in Fig. 10. The average of AUCs is shown in Table 3.

We detail the capability of our model at multiple levels of depth and compare it to the improved Resnet50 model in the [22] structure. The first test was performed using the Nvidia Tesla P100 with 12 GB of RAM. We carry out all our models from 50 to 100 times. The first learning rate was set at 0.001 and 0.9 decay rate with Adam as the optimizer. We modified the reproduction property to reduce the reconstruction loss according to the image's size so that the loss of margins does not dominate. Once each model is calculated, CapsNet² and DenseNet³ to

generate our sample test errors. All trials used three methods. To make just comparisons, we essentially maintain the standards for the proposed DNet model after the PrimaryCaps layer, just like traditional CapsNet. The corresponding implementation of the set of data used is shown below. A set of data from the chest x-ray-14 manuscript and data collection will contain training and testing of 80000 images per 1024×1024 image. We did not utilize any system to increase data and executed the experiment three times.

DNet can identify differences in data entry rates compared to CapsNet, i.e. it is capable of accurately testing a chest x-ray-14 data set with 20-fold fewer overall duplicates. The step in the PrimaryCaps layer of the second level of DNet was modified from two to one. This modification was done to adjust the picture. No improvement was noticed in the execution of DNet 3 levels in chest x-ray-14, which is expected because we get fundamental properties at the approximate level. Each slight difference will be captured in writing a specific number from the training group. During the testing phase, a problem can occur due to these “thin” properties. More time needs to be invested in improvement. We feel that the loss of the modified decoder in the DNet has dropped significantly, with the same loss multiplier that happened to CapsNet. In addition, we illustrate the comparison between the DNet test, which we can undoubtedly conclude that DNet has a quicker convergence rate.

4.3.5. Experiment : 3(DNet-7)

In 3rd, we have utilized 70000 pictures for training and 10000 pictures for testing without data augmentation; we have taken 14 classes in the dataset [22]. Image input size we used 1080×1080 . We taught our model for 500 epochs with a batch size of 256 images. The loss function we used is margin loss [20], and optimizer Adam [55]. The ROC Curve for the predicted test (see Fig. 12).

² CapsNet <https://github.com/XifengGuo/CapsNet-Keras>.

³ DenseNet <https://github.com/liuzhuang13/DenseNet>.

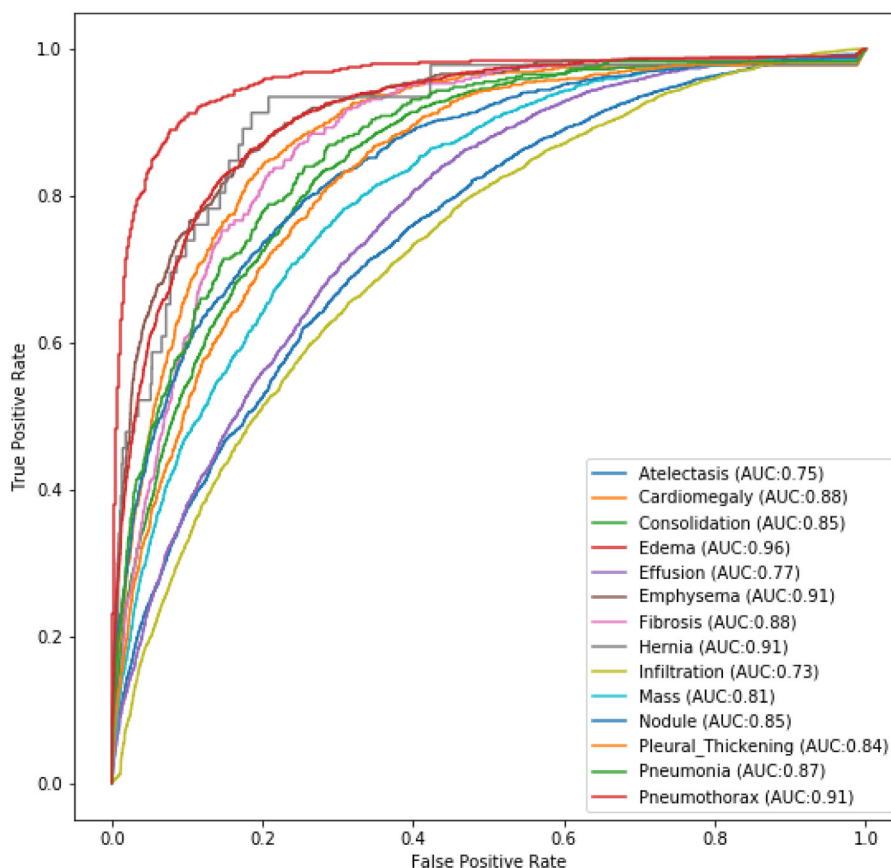


Fig. 8. Dynamic routing ROC.

First, we have done experiments with three levels of depth and achieved results shown in Fig. 11. The average of AUCs is shown in Table 3.

We detail the capability of our model in numerous depth levels and compare them with the Resnet50 [44] model actualized in [22] structure. The first test was performed utilizing the Nvidia Tesla P100 with 12 GB of RAM. We carry out all our models 50 to 100 times. The introductory learning rate was set at 0.001, and 0.9 decay rate with Adam [55] as the optimizer. We change the reproduction property to reduce the reconstruction loss according to the image's size so that the loss of margins does not dominate. Once we train each of our models for DenseNet and CapsNet to create our 2 sample test errors. Three methods were utilized for all trials. To make just comparisons, we essentially retain the DNet model parameters proposed after the PrimaryCaps layer, just like traditional CapsNet.

The following details show the corresponding implementation of the data set used. A collection of chest x-ray-14 manuscript data and data collection will have the training and test images of 80000 size, each image 1024 by 1024 in size. We did not use any system to increase the data, and we repeated the experiment 3 times. DNet can identify the differences in data entry speeds compared to CapsNet. It can achieve n accuracy test in chest x-ray-14 data set with a 20 times drop in all of the duplicates. We have shifted the step from two to one in the PrimaryCaps layer of DNet's second layer to adjust the image.

No improvement was observed in the performance of DNet 3 levels in both chest x-ray-14, which is expected Data because we capture fine properties at the rough level. Thus, each slight difference is captured from the training set while writing a certain number. Such "thin" properties have the probability of causing a problem in the testing phase. Thus, we require to devote more

Table 3

Average AUCs.

	Wang et al. [22]	DNet-3	DNet-5	DNet-7
Atelectasis	0.70	0.77	0.75	0.80
Cardiomegaly	0.810	0.72	0.82	0.86
Consolidation	0.70	0.73	0.80	0.84
Edema	0.80	0.77	0.91	0.92
Effusion	0.75	0.71	0.76	0.80
Emphysema	0.83	0.69	0.86	0.91
Fibrosis	0.78	0.65	0.83	0.86
Hernia	0.87	0.66	0.84	0.90
Infiltration	0.66	0.75	0.73	0.77
Mass	0.69	0.71	0.77	0.80
Nodule	0.66	0.67	0.82	0.87
Plueral Thickening	0.685	0.69	0.77	0.82
Pneumonia	0.65	0.70	0.81	0.85
Pneumothorax	0.79	0.74	0.87	0.91
AVG	0.745	0.711	0.810	0.867

time to improve. It was found that the loss of the modified decoder in DNet is significantly reduced, with the same loss multiplex as CapsNet. In addition, we demonstrate the comparison of the test of DNet, which concludes that DNet has a more rapid convergence rate.

We have used 70000 training images and 10000 testing images without data augmentation. We have taken 14 classes in the dataset [22]. We used image input size 1080 × 1080, and our model was trained for 500 epochs with 256 image-batch sizes. The loss function we used is margin loss [20], and optimizer Adam [55]. The ROC Curve for all predicted tests. First, we have done experiments with three levels of depth and achieved results shown in Fig. 9. Then we have done experiments with five levels of depth and achieved results are in Fig. 11 in this experiment,

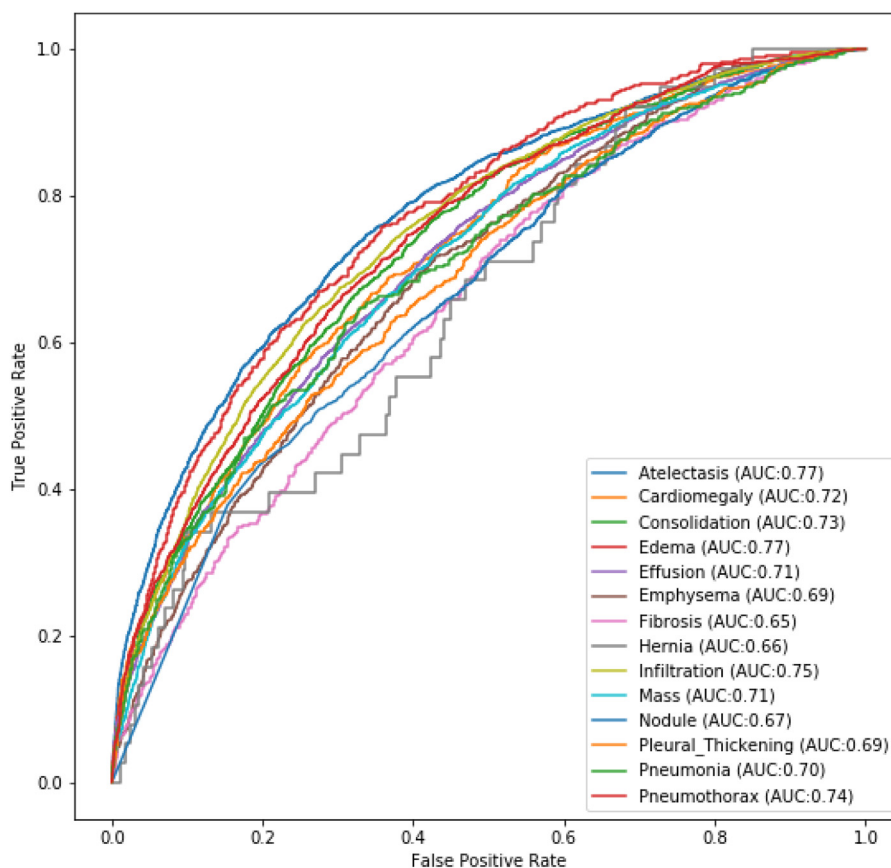


Fig. 9. ROC-DNet-3.

we have achieved better results than Wang et al. [22]. In the 3rd experiment, we have taken our model deeper than five levels, we have experimented with seven levels of depth, and it performed better than five levels of DNet, achieved results are shown in Fig. 11. The average of AUCs is shown in Table 3.

4.4. Complexity of DNet

Neural synapses rely on the simple fact that the vision system needs to use the same knowledge in all places of the image. This is achieved by linking the weights of feature detectors so that the features learned are available in one location at other locations. Stretch capsules knowledge sharing between sites includes knowledge about partial relationships that characterize a familiar form. The changes in the view have complex effects on pixel density but simple linear effects on the matrix situation that represent the relationship between an object or a part of the object and the viewer. The goal of capsules is to make good use of this latent line, deal with differences in perspective and improve partitioning decisions.

Capsules are used to filter a high-dimensional match: A familiar object can be detected by seeking to reach an agreement between the sounds of the modulation matrix. These sounds come from parts that have already been discovered. One of the parties produces a vote by multiplying a matrix of its positions by an acquired conversion matrix that represents the fixed relationship between the point of view between the segment and the whole. As the view changes, the position matrices for Parties and the group will change in a coordinated manner so that any agreement between the voices of the different parties will continue. Finding narrow sets of high-dimensional sounds that coincide with the noise of irrelevant voices is one way to solve

the problem of appointing parties to all. This is not trivial because we cannot put the space of high-dimensional space in a way that squares the area of low-dimensional translation to facilitate the twists.

We use a quick iterative process called “guide agreements” to solve this challenge. This group is a powerful retailing principle that allows familiarity in familiar ways to derive fragmentation, rather than simply using low-level signals such as proximity or agreement in color or speed. It determines the probability of allocating part to segment based on the proximity of the voices from that segment to those from other parts. The crucial difference between capsules and standard neural networks is that the activation of the capsule is based on the comparison of multiple incoming position predictions. While in the traditional neural network and based on the comparison of the incoming single vector activity and the vector of weight learned, this complexity can be reduced by using EM Matrix [54].

5. Discussion

In the present work, a modified version of two models capsule network and Deeper DNet has been proposed, which replaces the standard embedding layers in CapsNet with a tightly connected torsion. The addition of direct links between two successive layers helps learn better maps of the features. This, in turn, helps to form better quality capsules, and the effectiveness of this proposed model is reflected in the results. In our 1st experiment with three levels of depth, we achieved 0.711 average AUC, which is not better than [22], but in our 2nd experiment, our model with depth levels five has produced better results than [22] which has 0.811 average AUC level. Furthermore, we did 3rd experiment with seven levels of depth and compared the results with our

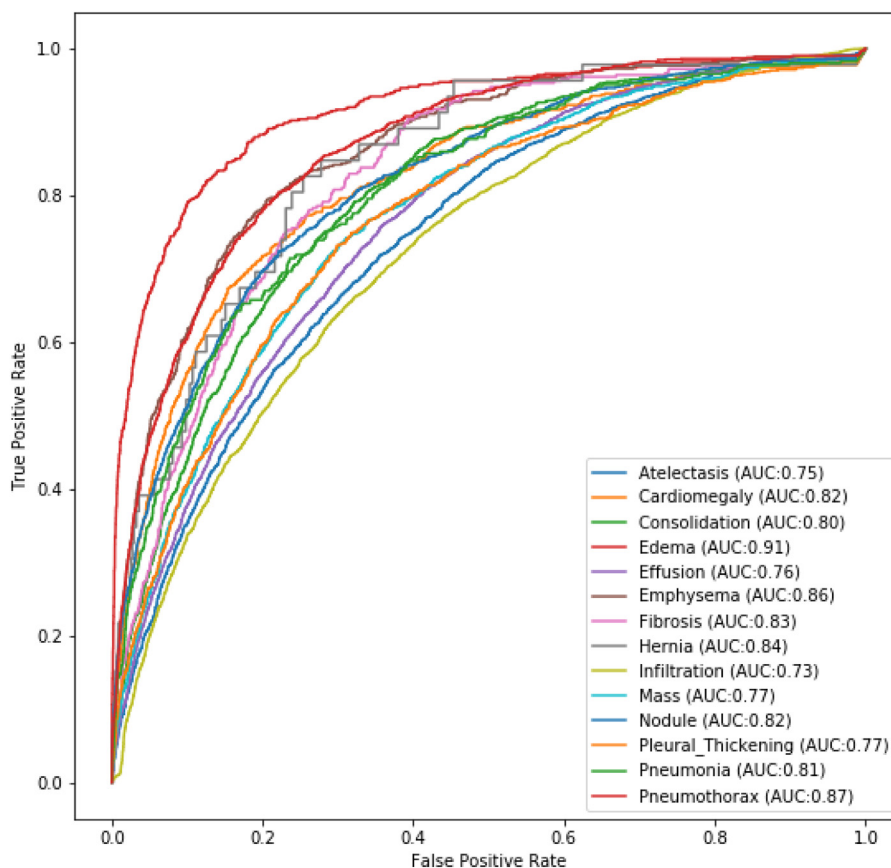


Fig. 10. ROC-DNet-5.

previous trained model. This surely needs a higher computational environment, but we achieved better results than 5 level depth model with a 0.867 average AUC level. The network which we proposed shows an improvement in average AUC compared to the previously implemented model ResNet50 [22]. However, this experiment leads to an increase in complexity as we go deep with our model.

Consequently, It is recommended to improve the accuracy of the proposed model by modifying it with em routing based on x-ray data set. This can help reduce the complexity and better diagnose the health status of patients against COVID-19.

As we have achieved better results with 5 and 7 levels of depth from our model DNet, but as we go deep with our model, the complexity increases as CapsNet vectorizes the model and data. To handle this complexity, we can in the future use the EM matrix proposed by [54]. By modifying our model with em routing, we can significantly reduce complexity and achieve better results. In the future, we plan to integrate EM directive [54] into DNet And further reduce the computational complexity of the model and improve the accuracy of our model.

The novel coronavirus pandemic has created a unique as well as demanding challenge for not only the overburdened healthcare system but all the connected domains and, most importantly, Information Communication and Technology (ICT) [56]. Most countries cannot have a pandemic-ready emergency response system to tackle such global situations. Effective use of Artificial Intelligence and especially using big data for creating Machine Learning models for precision medical care to cope with a future pandemic as well as give recommendations about arranging and effectively managing healthcare resources [57,58].

In this work proposed a modified capsule network named Deeper DNet, which replaces the standard embedding layers in

CapsNet with a tightly connected torsion. The addition of direct links between two successive layers helps to learn better maps of the features, which helps form better quality capsules. The effectiveness of this proposed model is reflected in the results. In our 1st experiment with three levels of depth, we achieved 0.711 average AUC, which is not better than [22] but in 2nd experiment, our model with five levels of depth has produced better results than [22] with an average AUC level of 0.811. Furthermore, we did 3rd experiment with seven levels of depth and compared the results with our previous trained model. This needed a higher computational environment, but we achieved better results than our 5 level depth model 0.867. The network which we proposed shows an improvement in average AUC compared to the previously implemented model ResNet50 [22].

6. Conclusions and future work

Capsule networks (CapsNet) are strong for rotation and hard translation and require much less training information, which applies to the processing of data sets from medical images, including radiological images of the chest x-rays.

In this research, we proposed a deep learning model for the problem of x-ray classification based on the technique of CapsNet. Capsule network and Deeper DNet are modified to replace the standard embedding layers in CapsNet with a tightly connected torsion. The addition of direct links between two successive layers helps to learn better maps of the features, which helps form better quality capsules. Different experiments were conducted on chest x-rays data sets with different experimental settings. The performance of our proposed model was evaluated by comparing it with the previous state of art CNN techniques. The results showed better performance in terms of area under curve (AUC) average as compared to previous approaches.

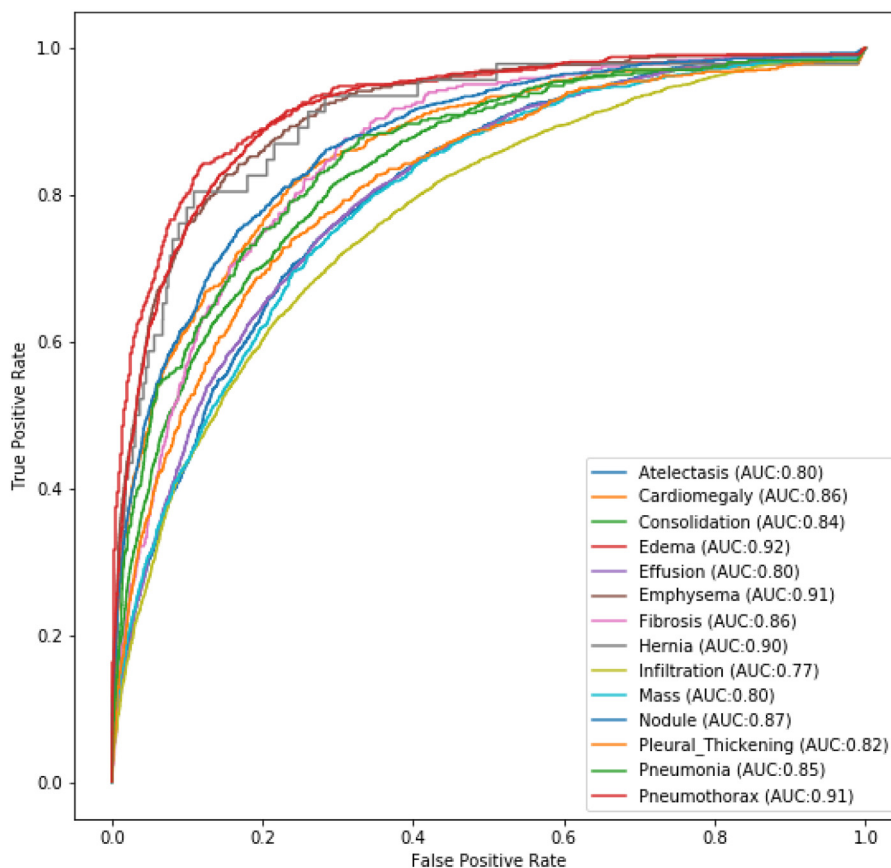


Fig. 11. ROC-DNet-7.

The output of the proposed system not only contributes to improving the intelligent health care system but also opens opportunities for the rest of the research world. A possible extension in this domain can be in improving optimization code and reducing the computational complexity of the model. It can be done using the EM matrix proposed by [54]. By modifying our model with em routing, we can significantly reduce complexity and achieve better results. In the future, we plan to integrate EM directive [54] into DNet And further reduce the computational complexity of the model and improve the accuracy of our model.

A capsule is a group of neurons representing the vector of their activity ionization process parameters for a particular type of entity, such as an object or part of the object. It uses the length of the activity vector to represent the probability of an entity's existence and direction to represent instance creation parameters.

We have achieved better results with 5 and 7 levels of depth from our model DNet, but as we go deep with our model, the complexity increases as CapsNet vectorizes the model and data. To handle this complexity, we can, in the future, use the EM matrix proposed by [54]. We can significantly reduce complexity and achieve better results by modifying our model with em routing. In the future, we plan to integrate EM directive [54] into DNet And further reduce the computational complexity of the model and improve the accuracy of our model.

The active capsules in the standard make predictions, through shift matrices, for the parameters of creating cases of capsules at a superior level. When many predictions coincide, a higher-level capsule is activated. We proposed Dense Capsule Networks (DNet) on a more depth level. The proposed frames allocate CapsNet by replacing the standard tiling layers with densely

connected sequences. This helps to integrate the characteristics maps learned by the different layers to form the primary capsules. DNet essentially adds a deeper network of communication, leading to learning maps with discriminatory characteristics. DNet uses a hierarchical structure to learn capsules that represent more or less spatial information, making them more effective in learning complex data. DNet checks the best performance in the ChestXray-14 data set on traditional CNN. In addition, DNet performs better with a higher level of total depth. Experiments on image classification tasks using reference data sets illustrate the effectiveness of proposed architectures.

CRedit authorship contribution statement

Faizan Karim: Investigation, Software, Visualization. **Munam Ali Shah:** Conceptualization, Project administration. **Hasan Ali Khattak:** Conceptualization, Investigation, Writing – original draft. **Zoobia Ameer:** Validation, Formal analysis. **Umar Shoaib:** Methodology, Project administration, Software. **Hafiz Tayyab Rauf:** Writing – review & editing, Project administration, Resources. **Fadi Al-Turjman:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

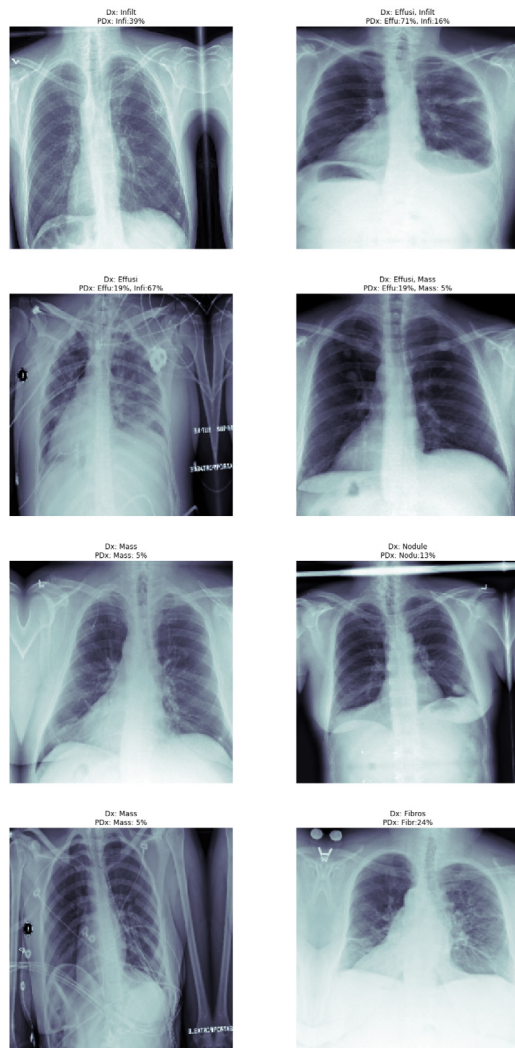


Fig. 12. Sample of predicted cases.

References

- [1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, et al., A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* (2020).
- [2] M. Ahmadi, A. Sharifi, M. Jafarian Fard, N. Soleimani, Detection of brain lesion location in MRI images using convolutional neural network and robust PCA, *Int. J. Neurosci.* (2021) 1–12.
- [3] R.M. Anderson, H. Heesterbeek, D. Klinkenberg, T.D. Hollingsworth, How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 395 (10228) (2020) 931–934.
- [4] R. Bashshur, C.R. Doarn, J.M. Frenk, J.C. Kvedar, J.O. Woolliscroft, Telemedicine and the COVID-19 pandemic, lessons for the future, 2020, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New.
- [5] J. Yang, Y. Zheng, X. Gou, K. Pu, Z. Chen, Q. Guo, R. Ji, H. Wang, Y. Wang, Y. Zhou, Prevalence of comorbidities in the novel wuhan coronavirus (COVID-19) infection: A systematic review and meta-analysis, *Int. J. Infect. Dis.* 10 (10.1016) (2020).
- [6] J. Balilla, Assessment of COVID-19 mass testing: the case of South Korea, 2020, Available At SSRN 3556346.
- [7] S. Hassantabar, M. Ahmadi, A. Sharifi, Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches, *Chaos Solitons Fractals* 140 (2020) 110170.
- [8] C.J. Wang, C.Y. Ng, R.H. Brook, Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing, *JAMA* 323 (14) (2020) 1341–1342.
- [9] S. Boccia, W. Ricciardi, J.P. Ioannidis, What other countries can learn from Italy during the COVID-19 pandemic, *JAMA Intern. Med.* 180 (7) (2020) 927–928.
- [10] M. Ienca, E. Vayena, On the responsible use of digital data to tackle the COVID-19 pandemic, *Nat. Med.* 26 (4) (2020) 463–464.
- [11] S. Rahmatizadeh, S. Valizadeh-Haghi, A. Dabbagh, The role of artificial intelligence in management of critical COVID-19 patients, *J. Cell. Mol. Anesth.* 5 (1) (2020) 16–22.
- [12] A.S.S. Rao, J.A. Vazquez, Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine, *Infect. Control Hosp. Epidemiol.* 41 (7) (2020) 826–830.
- [13] V. Chang, Data analytics and visualization for inspecting cancers and genes, *Multimedia Tools Appl.* 77 (14) (2018) 17693–17707.
- [14] V. Chang, Computational intelligence for medical imaging simulations, *J. Med. Syst.* 42 (1) (2018) 1–12.
- [15] M. Ahmadi, A. Sharifi, S. Khalili, Presentation of a developed sub-epidemic model for estimation of the COVID-19 pandemic and assessment of travel-related risks in Iran, *Environ. Sci. Pollut. Res.* 28 (12) (2021) 14521–14529.
- [16] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning, 2017, arXiv preprint arXiv:1711.05225.
- [17] H. Wang, Y. Xia, Chestnet: A deep neural network for classification of thoracic diseases on chest radiography, 2018, arXiv preprint arXiv:1807.03058.
- [18] G. Larsson, M. Maire, G. Shakhnarovich, Fractalnet: Ultra-deep neural networks without residuals, 2016, arXiv preprint arXiv:1605.07648.
- [19] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [20] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, 2017, arXiv preprint arXiv:1710.09829.
- [21] M. Ahmadi, A. Sharifi, S. Hassantabar, S. Enayati, QAIS-DSNN: tumor area segmentation of MRI image with optimized quantum matched-filter technique and deep spiking neural network, *BioMed Res. Int.* 2021 (2021).
- [22] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2097–2106.
- [23] X. Wang, Y. Peng, L. Lu, Z. Lu, R.M. Summers, Tienet: Text-image embedding network for common thorax disease classification and reporting in chest X-rays, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9049–9058.
- [24] S.S.R. Phaye, A. Sikka, A. Dhall, D. Bathula, Dense and diverse capsule networks: Making the capsules learn better, 2018, arXiv preprint arXiv:1805.04001.
- [25] Y. Shen, M. Gao, Dynamic routing on deep neural network for thoracic disease classification and sensitive area localization, in: *International Workshop on Machine Learning in Medical Imaging*, Springer, 2018, pp. 389–397.
- [26] G. Chartrand, P.M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C.J. Pal, S. Kadoury, A. Tang, Deep learning: A primer for radiologists, *Radiographics* 37 (7) (2017) 2113–2131.
- [27] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (7639) (2017) 115–118.
- [28] Z. Zhang, P. Chen, M. Sapkota, L. Yang, Tandemnet: Distilling knowledge from medical images using diagnostic reports as optional semantic references, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 320–328.
- [29] J. Liu, D. Wang, L. Lu, Z. Wei, L. Kim, E.B. Turkbey, B. Sahiner, N.A. Petrick, R.M. Summers, Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks, *Med. Phys.* 44 (9) (2017) 4630–4642.
- [30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
- [31] M. Everingham, S.A. Eslami, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: A retrospective, *Int. J. Comput. Vis.* 111 (1) (2015) 98–136.
- [32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: *European Conference on Computer Vision*, Springer, 2014, pp. 740–755.
- [33] B.A. Plummer, L. Wang, C.M. Cervantes, J.C. Caicedo, J. Hockenmaier, S. Lazebnik, Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2641–2649.
- [34] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C.L. Zitnick, D. Parikh, Vqa: Visual question answering, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2425–2433.

- [35] Y. Zhu, O. Groth, M. Bernstein, L. Fei-Fei, Visual7w: Grounded question answering in images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4995–5004.
- [36] O. Vinyals, A. Toshev, S. Bengio, D. Erhan, Show and tell: A neural image caption generator, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3156–3164.
- [37] A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3128–3137.
- [38] Q. Wu, P. Wang, C. Shen, A. Dick, A. Van Den Hengel, Ask me anything: Free-form visual question answering based on knowledge from external sources, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4622–4630.
- [39] P. Young, A. Lai, M. Hodosh, J. Hockenmaier, From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions, *Trans. Assoc. Comput. Linguist.* 2 (2014) 67–78.
- [40] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D.A. Shamma, et al., Visual genome: Connecting language and vision using crowdsourced dense image annotations, *Int. J. Comput. Vis.* 123 (1) (2017) 32–73.
- [41] J. Johnson, A. Karpathy, L. Fei-Fei, Densecap: Fully convolutional localization networks for dense captioning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4565–4574.
- [42] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, R.M. Summers, Interleaved text/image deep mining on a very large-scale radiology database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1090–1099.
- [43] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, R.M. Summers, Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2497–2506.
- [44] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 25 (2012) 1097–1105.
- [45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
- [46] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
- [47] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [48] D. Demner-Fushman, M.D. Kohli, M.B. Rosenman, S.E. Shooshan, L. Rodriguez, S. Antani, G.R. Thoma, C.J. McDonald, Preparing a collection of radiology examinations for distribution and retrieval, *J. Am. Med. Inform. Assoc.* 23 (2) (2016) 304–310.
- [49] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, K. Lyman, Learning to diagnose from scratch by exploiting dependencies among labels, 2017, arXiv preprint arXiv:1710.10501.
- [50] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [51] I.M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, A. Saalbach, Comparison of deep learning approaches for multi-label chest X-ray classification, *Sci. Rep.* 9 (1) (2019) 1–10.
- [52] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, D. Comaniciu, Learning to recognize abnormalities in chest X-rays with location-aware dense networks, in: Iberoamerican Congress on Pattern Recognition, Springer, 2018, pp. 757–765.
- [53] R.K. Srivastava, K. Greff, J. Schmidhuber, Highway networks, 2015, arXiv preprint arXiv:1505.00387.
- [54] G.E. Hinton, S. Sabour, N. Frosst, Matrix capsules with EM routing, in: International Conference on Learning Representations, 2018.
- [55] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
- [56] A. Alimadadi, S. Aryal, I. Manandhar, P.B. Munroe, B. Joe, X. Cheng, Artificial intelligence and machine learning to fight COVID-19, 2020.
- [57] Z. Ahmed, K. Mohamed, S. Zeeshan, X. Dong, Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine, *Database* 2020 (2020).
- [58] M. Afzal, M. Hussain, S. Lee, H.A. Khattak, Redesign of clinical decision systems to support precision medicine, in: TENCON 2018-2018 IEEE Region 10 Conference, IEEE, 2018, pp. 2259–2263.