2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) | 978-1-6654-1296-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/AINIT54228.2021.00076

A Lung Cancer Detection System Based on Convolutional Neural Networks and Natural Language Processing

Jiahao Chen[1,*,†]
[1]The University of Melbourne
Melbourne, Australia
[*]Corresponding author's e-mail:
jiahchen4@student.unimelb.edu.au

Qianli Ma[2,*,†]
[2]University of Massachusetts Boston
Boston, USA
[*]Corresponding author's e-mail:
qianli.ma002@umb.edu

Weixin Wang[3,*,†]
[3]University of California, Irvine
Irvine, USA
[*]Corresponding author's e-mail:
weixinw1@uci.edu
[†]These authors contributed equally.

Abstract- **Lung Cancer has long been regarded as one of the most threatened diseases to human beings, and detecting early malignant tumors is of vital importance for treatment. Contemporarily, Radiology departments in hospitals usually have to deal with multiple CT images to carry out the detection, which is a huge workload for doctors. Here, we propose a novel system to help with lung cancer detection. Specifically, deep feature based convolutional neural networks (CNN) is applied to classify lung cancer tumors, realizing an accuracy of 88%. Moreover, a chatbot based on natural language processing (NLP) technology is embedded into the system to provide immediate knowledge and information. These results shed light on how doctors' workload might be reduced to a considerable extent.**

Keywords- Lung cancer; convolutional neural networks; natural language processing

I. Introduction

Lung cancer is a common type of cancer serving as a leading cause of cancer death. In 2020, there were 2.2 million new cases and 1.8 million deaths of lung cancer. There are a variety of ways to assist physicians making accurate lung cancer diagnosis, including sputum cytology, flexible bronchoscopy, and transthoracic needle aspiration. Sputum cytology is a noninvasive test that is useful in identifying centrally located tumors. Generally， flexible bronchoscopy is the test of choice in patients with central tumors, with a combined sensitivity of 88 percent in these patients. Transthoracic needle aspiration has been shown to be more sensitive than bronchoscopy in patients with peripheral lung tumors, which may be used when transbronchial needle aspiration is inconclusive or in patients who are not surgical candidates. After the metastatic evaluation is complete, the staging classification can be determined based on the type of tumor identified [1].

Deep learning is a popular method in computer vision areas. The growth of computational power makes it possible to train deep neural networks. Neural networks may perform better at classifying pictures because they can extract and organize features and do not require prior knowledge. Convolutional neural network (CNN) is a class of neural networks, based on shared weights of convolution filters. In the past decade, we've seen many CNNs with promising results in the field of cancer diagnosis. Alakwaa *et al.* [2] suggested a way to detect lung cancer from 3D CT images. They first used a U-Net to label the nodules in lungs. They passed the data to 3D CNN to detect the actual malignancy. With a combination of U-net and 3D CNN, they reached 86.6% accuracy. Welch *et al.* [3] implemented AlexNet for this lung cancer classification problem. AlexNet has 4 convolutional layers, 4 max pooling layers and 2 fully connected layers, which reached an accuracy of 93.55%.

So far, the research about lung cancer diagnosis has prevailed over the deep learning domain. However, the focus of cancer diagnosis using deep learning is vastly on image classification, for images have an abundant amount of data storage over the world. Nevertheless, Medical knowledge base has long been an important source of information for both doctors and patients. For instance, community-driven question answering (QA) websites, e.g., Quora, Yahoo-Answers, and Answers.com are accumulating millions of users and hundreds of millions of questions. A large portion of the questions are about facts or trivia. It has been a long pursuit to enable machines to answer such questions automatically [4]. Harnessing the potential of clinical knowledge base will require strategies for efficient and automated information extraction [5]. In this term, a system of chatbot answering the questions will provide medical practitioners with a range of knowledge extraction tasks and real-time, personalized feedback [6]. Deep learning, specific to neural networks, is the most popular model related to Chatbot. For instance, Bi-Long Short-Term Memory (Bi-LSTM) is a variant of Recurrent neural network (RNN), which could efficiently capture distant time-ranged information in the Chatbot domain [7]. There are some other existing medical tasks such as identifying the sentences that involve clinically important recommendation information [8]. Meanwhile, an efficacious chatbot has the ability to reduce time and financial cost for patients to a large extent.

The overall goal of this work is to develop a system which helps doctors make diagnosis and communicate with others. CNN is implemented to detect malignant tumors from CT images of lungs. In addition, a chatbot based on Hierarchical Bi-Long Short-Term Memory Attention Model (HBAM) is built to provide a function which can answer questions based on the description of the patient's circumstances. Experimentation on training the models and design of the system are two main focuses in this work.

The remainder of this article is organized as follows. The structure and mechanism of the whole system are described in

978-1-6654-1296-4/21/$31.00 ©2021 IEEE
DOI 10.1109/AINIT54228.2021.00076

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.

Section II. The image recognition tool using deep feature based CNN, as well as its experimental results, is presented in Section III. Then in Section IV, the chatbot based on natural language processing (NLP) technology is set. Finally, concluding remarks are contained in the Section V.

II. Design of The System

A. Overall Structure

The system is built based on the python flask framework. Users of the system are currently only doctors, and they can log in to check new announcements or diagnosing records at any time. Figure 1 presents the main functions in our system.

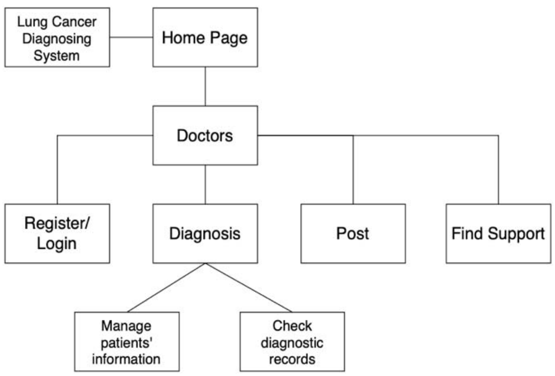

Figure 1. System structure diagram

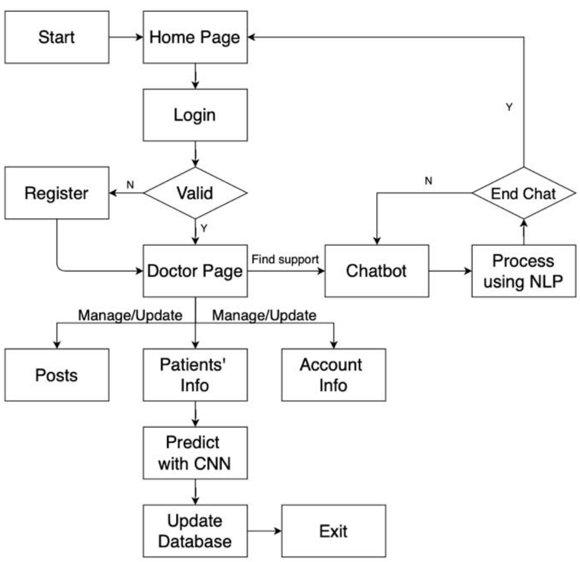

Figure 2. System flow chart

Besides making predictions of the cancer, doctors can also create posts if they encounter some problems or have some new thoughts, aiming at providing a collaborative platform for doctors to share their ideas. Moreover, a chatbot called Jarvis is built, which provides immediate knowledge and information. Figure 2 explains the design of our system flow.

B. Database

A basic database is also constructed using MySQL. The database mainly includes Doctors (User), Posts and Patients. Users' account information, including username, email, password, and profile image, are stored. For patients, the diagnosing results are generated and stored in the database once patients' information is submitted by the doctor. Date and time for each activity are kept track of, i.e., they can be viewed chronologically. Most importantly, sensitive data is stored using hash to guarantee privacy security. Figure 3 shows table view of the database.

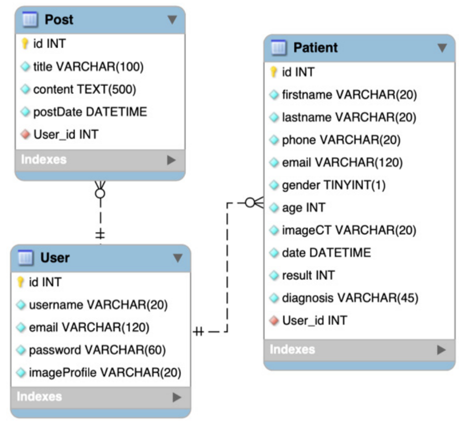

Figure 3. Database table view

C. User Interface

The user interface is designed to be user-friendly. Figure 4 illustrates the home page of the system, which contains a navigation bar both on the top and the right sides. The shortcuts on the right side are kept for future extension since the system's functionality is relatively simple at this stage. Users can simply modify their profiles and post new announcements through the shortcuts on the top bar. Figures 5 and 6 demonstrate the path users add a new patient and check diagnosing records.

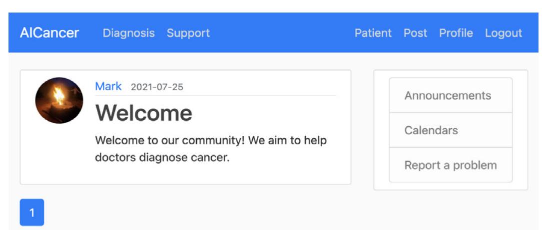

Figure 4. Home page

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.

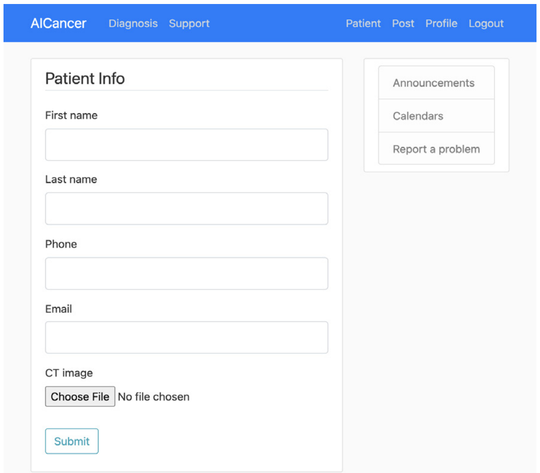

Figure 5. New patient page

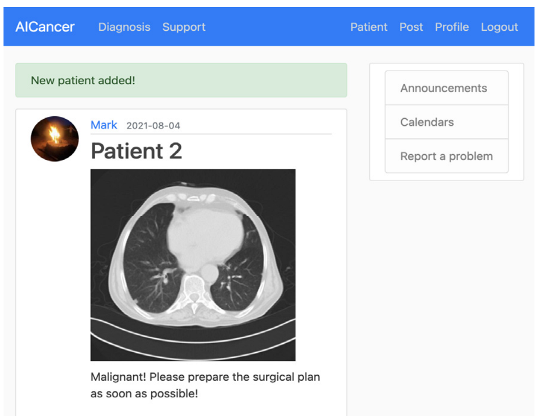

Figure 6. Diagnosing records

The chatbot, named Jarvis, is embedded in the system as well, i.e., users can get support conveniently. It handles users' messages using the pre-trained NLP model and generates corresponding responses in real time. Figure 7 depicts the chatbot interface.

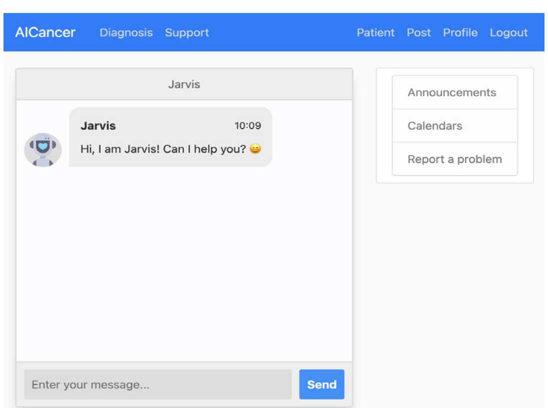

Figure 7. Chatbot interface

III. Image Recognition via Convolutional Neural Networks

A. Dataset and preprocessing

The IQ-OTH/NCCD lung cancer dataset from Kaggle is used to train our models, which contains 1097 CT images of lungs. All the images have been already transferred to jpg from DICOM, with size of 512*512*3 initially. They are labelled as healthy, benign, and malignant originally. However, the test results of the models are poor. In this case, we decide to split them into two categories: malignant and non-malignant, which makes sense to some extent since detecting malignant tumors is the most important aim for us. For preprocessing, pixels are normalized to be zero-centered since all the pixels can be regarded as a value between 0 and 255. Finally, 985 images are treated as training dataset while the rest becomes the testing one.

B. Experimental results

For our first trial, VGG16 is an appropriate choice for it is shown that this model generalizes well to a wide range of tasks and datasets, matching or outperforming more complex recognition pipelines built around less deep image representations [9]. Nevertheless, this model didn't perform well during our classification task for lung cancer diagnosis. Its validation accuracy vacillated frequently during the training process and didn't have a convincing accuracy as shown in figure 8. Besides, the Inception model invented by Google became an alternative model because of its new idea for optimizing local construction and to repeat it spatially [10]. On the contrary to our conjecture, the Inception model performed even worse than the VGG16 model and fluctuated intensely as Figure 9 below shows.

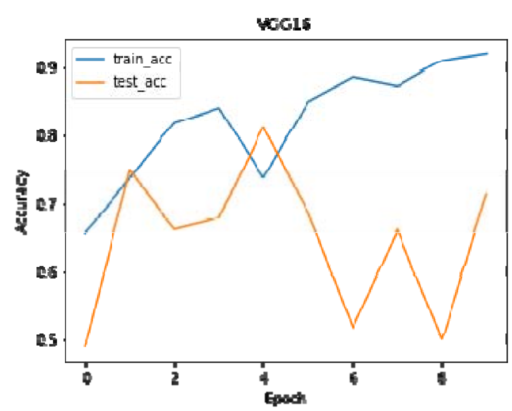

Figure 8. Result of VGG16

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.

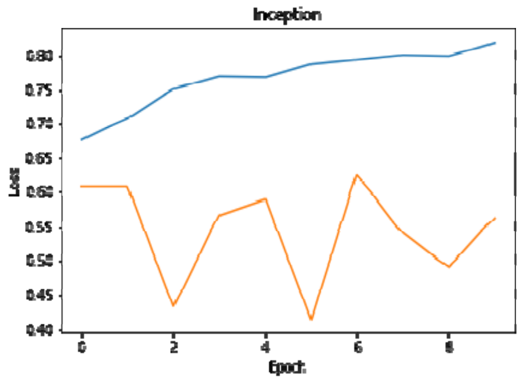

Figure 9. Result of Inception

Since the above models did not generate satisfying results for us, we found another two promising models. The first one is a densely connected CNN proposed by Huang *et al.* [11], which required huge memory space and computation power. Pleiss *et al.* [12] came up with a solution to implement this network in an efficient way. However, we did not manage to train this model multiple times to generate a decent result due to limitations in our devices. The final validation accuracy achieves 85%, which is still an obvious progress.

AlexNet proposed by Krizhevsky *et al.* [13] became another option, which showed the best performance and reached 85% accuracy in the contest of classifying ImageNet dataset into 1000 classes. We decided to follow the same structure and apply it on our lung cancer dataset. The validation accuracy reached 83.2% after 50 epochs as Figure 10 shows, which was also a decent result.

Moreover, Jiao *et al.* [14] proposed a deep feature based CNN (DCNN) for breast classification. Compared to the dense net, this one had much lower requirements on our devices and produced satisfying results. We trained the model for 10 epochs and finally achieved validation accuracy of nearly 88%, which was the highest among all models. Table I presents the CNN framework and Figure 11 shows our training results.

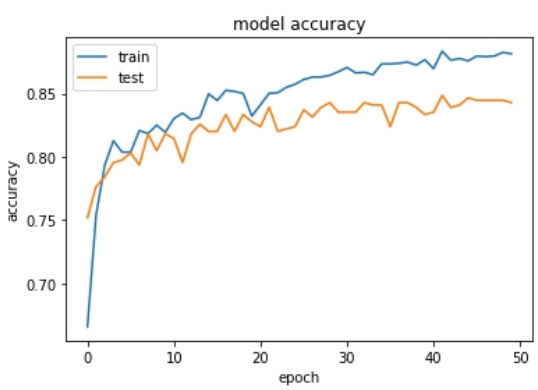

Figure 10. Result of AlexNet

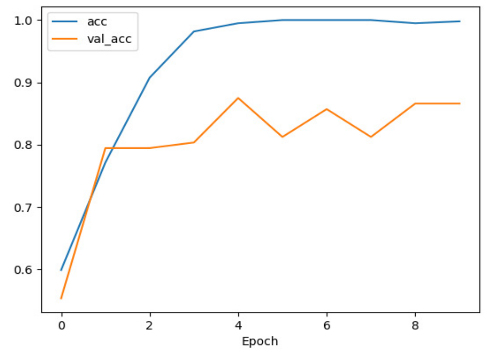

Figure 11. Result of DCNN

TABLE I. DCNN STRUCTURE [14]

Name	Filter size	Filter dimension	Stride	Padding
Conv1	7	1	2	0
ReLU1	1	/	1	0
Pooling1	3	/	3	2
Conv2	5	96	1	2
Pooling2	2	/	2	1
Conv3	3	256	1	1
ReLU3	1	/	1	0
Conv4	3	384	1	1
ReLU5	1	/	1	0
Pooling5	3	/	3	1
Fc6	1	256	1	0
ReLU6	1	/	1	0
Fc7	1	2048	1	0
ReLU7	1	/	1	0
Fc8	1	2	1	0

C. *Comparision*

TABLE II. VALIDATION ACCURACY OF DIFFERENT MODELS

Model	Validation Accuracy
VGG 16	72%
Inception	61%
AlexNet	83%
Dense net	85%
DCNN	88%

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.

Table II lists the validation accuracy of all the four models we implemented. It is straightforward to see that the last two obtained better results.

Comparing Dense net and DCNN, we plot the confusion matrices respectively as summarized Tables III and IV. The number of false positives is relatively low for both models, i.e., the real threats are reduced. Since the final validation accuracy of DCNN is the highest, we decided to implement it in our system.

TABLE III. CONFUSION MATRIX FOR DENSE NET

Actual \ Predict	Malignant	Non-Malignant
Malignant	44	6
Non-malignant	10	49

TABLE IV. CONFUSION MATRIX FOR DENSE NET

Actual \ Predict	Malignant	Non-Malignant
Malignant	49	4
Non-malignant	8	51

IV. CHATBOT BASED ON NATURAL LANGUAGE PROCESSING TECHNOLOGY

A. *Function*

Our chatbot can answer questions based on the similarity between the user's input sentence and questions in databases. The framework is a retrieve-based model using HBAM. The main innovation is to use the Siamese_LSTM [15] + word attention + Manhattan distance to compare with the normal Siamese_LSTM + normal attention + Manhattan distance. The main reason that we need to model is to retrieve the top k answer from the medical question answer pair dataset according to the medical semantic similarity model.

B. *Data preprocess*

There are 2 parts of datasets, one is from Quora, an open-source platform, consisting of 10000 questions similarity pairs using as train and test datasets. These datasets are stored by id1, id2, Question1, Question2, and similarity label with value 0 or 1. The other one is our retrieval databases accumulated from eHealth Forum, iCliniq, Question Doctors and WebMD which in total have about 29000 pairs. Meanwhile, word-embedding is also needed for training deep learning models, for neural network models only accept vector-like data. For simplicity and efficiency, google word2vec model is implemented as a pre-trained model during word-embedding process.

C. *Training progress*

There are three kinds of training models as tentative candidates before implementation, HBAM, MaLSTM [16], and Bi-LSTM + attention. Primarily, we define a shared model for both sentence1 and sentence2. After using the shared model, Manhattan distance is computed by the hidden layer and can concatenate to get similarity between two sentences. Figure 12 shows a sample training curve for these kinds of work.

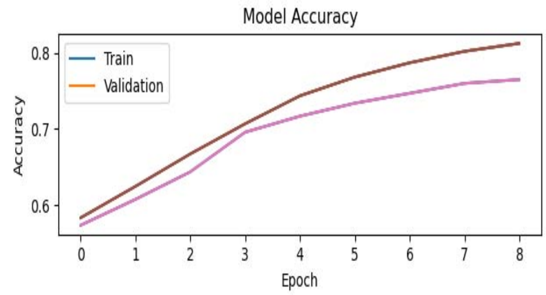

Figure 12. Result of MaLSTM

After training three models and comparing their results in Table V, we find that HBAM has the highest efficiency. Hence, HBAM is determined to be the final model that harnesses the backbone of chatbot.

TABLE V. COMPARISON OF DIFFERENT MODELS

Methods	Average Accuracy	Range of change by 10 times experiments
MaLSTM	76.2%	[-1.8%,0.9%]
Bi-LSTM+Attention	77.9%	[-2.1%,2.3%]
HBAM	80.1%	[-2,5%,2.1%]

D. *Evaluation*

In the steps above, only sentence similarity can be computed by HBAM. Nevertheless, it is unable to address the questions of the user. The process to choose the answer is using the top K method that computes the similarity of all sentences in datasets and choose top k sentences as possible answers.

V. CONCLUSION

In summary, this paper mainly focuses on designing an application for lung cancer diagnosis, which manages to efficiently detect malignant tumors and create a platform for medical practitioners. Image classification is conducted by deep feature based CNN after several experiments and comparisons, which reaches an accuracy of 88%. Besides, HBAM was implemented for the chatbot, which comprehensively creates a human interaction system helping doctors answering medical related questions. All the features are finally merged into a web application based on the flask framework, which has a high standard of user experience.

In the future, a more advanced model is required to reach higher accuracy, which is currently far below the point that can replace doctors. It can be improved through model stacking. Additionally, the size of the dataset used here is limited, which is slightly over 1000 images. To further improve our models and verify the validation, more data is required. On the other hand, other types of cancers are well worth adding to our system.

The system can be deployed to the radiology department to reduce doctors' workload. Moreover, it may help improve

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.

accuracy of lung cancer diagnosis since doctors can make decisions through combining several dimensions.

REFERENCES

[1] Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. Am Fam Physician. 2007 Jan 1;75(1):56-63. PMID: 17225705.

[2] Wafaa Alakwaa, Mohammad Nassef and Amr Badr, “Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)” International Journal of Advanced Computer Science and Applications(IJACSA), 8(8), 2017.

[3] Welch H G, Schwartz L M and Woloshin S 2000 Are increasing 5-year survival rates evidence of success against cancer? Jama. 283 2975-8

[4] Dai, Z., Li, L., & Xu, W. (2016). CFO: Conditional Focused Neural Question Answering with Large-scale Knowledge Bases. ArXiv, abs/1606.01994.

[5] Pons, E., Braun, L. M., Hunink, M. G., & Kors, J. A. (2016). Natural Language Processing in Radiology: A Systematic Review. Radiology, 279(2), 329–343.

[6] Bao, Q., Ni, L., & Liu, J. (2020). HHH: An Online Medical Chatbot System based on Knowledge Graph and Hierarchical Bi-Directional Attention. Proceedings of the Australasian Computer Science Week Multiconference.

[7] Liu, H., Xu, Y., Zhang, Z., Wang, N., Huang, Y., Hu, Y., Yang, Z., Jiang, R., & Chen, H. (2020). A Natural Language Processing Pipeline of Chinese Free-Text Radiology Reports for Liver Cancer Diagnosis. IEEE Access, 8, 159110-159119.

[8] Yetisgen-Yildiz, M., Gunn, M., Xia, F., & Payne, T. (2013). A text processing pipeline to extract recommendations from radiology reports. Journal of biomedical informatics, 46 2, 354-62 .

[9] Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

[10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9.

[11] Huang, G., Liu, Z., & Weinberger, K.Q. (2017). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261-2269.

[12] Pleiss, G., Chen, D., Huang, G., Li, T., Maaten, L.V., & Weinberger, K.Q. (2017). Memory-Efficient Implementation of DenseNets. ArXiv, abs/1707.06990.

[13] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*, 84 - 90.

[14] Jiao, Z., Gao, X., Wang, Y., & Li, J. (2016). A deep feature based framework for breast masses classification. Neurocomputing, 197, 221-231.

[15] Mueller, J., & Thyagarajan, A. (2016). Siamese Recurrent Architectures for Learning Sentence Similarity. AAAI.

[16] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9, 1735-1780.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on May 25,2022 at 04:09:56 UTC from IEEE Xplore. Restrictions apply.