Article

Deep Learning Approaches to Automatic Chronic Venous Disease Classification

Marina Barulina [1,2]*, Ashat Sanbaev [2,3], Sergey Okunkov[1,2], Ivan Ulitin[1,2] and Ivan Okoneshnikov [2]

1 Institute of Precision Mechanics and Control of the Russian Academy of Sciences; 24, ul. Rabochaya, Saratov 410028, Russia
2 R&D department, TOO Fle, office 20, 24, st. Panfilova, Almaty, 050000, Kazakhstan
3 Omega clinic; 46, ul. Komsomolskaia, Saratov 410031, Russia
* Correspondence: marina@barulina.ru

Abstract: Chronic venous disease (CVD) occurs in a substantial proportion of the world's population. If the onset of CVD looks like a cosmetic defect, then over time, it can develop into serious problems that require surgical intervention. The aim of the work is to use deep learning (DL) methods for automatic classification of the stage of CVD for self-diagnosis of a patient by using the image of the patient's legs. The required for DL algorithms images of legs with CVD were obtained by using Internet Data Mining. For images preprocessing, the binary classification problem "legs - no legs" was solved based on Resnet50 with accuracy 0.998. The application of this filter made it possible to collect a data set of 11,118 good quality leg images with various stages of CVD. For classification of various stages of CVD according to CEAP classification, the multi classification problem was set and resolved by using two neural networks with completely different architecture - Resnet50 and DeiT. The model based on DeiT without any tuning shows better results than the model based on Resnet50 (precision = 0.770 (DeiT) and 0.615 (Resnet50)). To demonstrate the results of the work, a telegram bot was developed, in which fully functioning DL algorithms are implemented. This bot allows evaluating the condition of the patient's legs with a fairly good accuracy for the CVD classification.

Keywords: chronic venous disease; deep leaning; data mining; Resnet50; DeiT; automatic classification; automatic CEAP classification

---

1. Introduction

Chronic Venous Disease (CVD) is a disease that affects a substantial number of the population [1, 2]. In the initial stages of CVD, patients do not take this disease seriously and prefer to think of it as a cosmetic defect. Whereas the further development of CVD may require surgical intervention, the cost of which can be quite high [3, 4]. At the same time, it is possible to slow down the development of the disease in the early stages using compression or pharmacological therapy, or by reducing risk factors such as overweight and obesity, prolonged standing or sitting [5].

The problem of using neural networks in the tasks of diagnosing the stage of varicose veins has long been of interest to researchers. The work [6] of 1995 year describes the process of splitting the stages of varicose disease into classes for a neural network and one of the first methods for applying a neural network to the problem of classifying images of legs with varicose veins. But the level of technological and algorithmic progress of that time did not allow transferring the ideas expressed in this work to life.

Further, with the development of technologies and developments in the field of machine and deep learning, more and more advanced methods for processing images appeared. The current level of progress in artificial intelligence and deep learning algorithms makes it possible to develop applications for self-diagnosis of patients, so that they themselves can determine the level of CVD and consult a doctor on time.

Thus, a number of works have already been published on determining the severity of CVD disease using artificial intelligence. In the work [7], the architecture of a model inspired by popular neural networks such as Google-Net and VGG is described. This model, with a detailed description of the mathematical functionality of each of the layers, according to the study, shows not bad results in diagnosing the stage of varicose veins using an image. In addition, this work demonstrates the result of working on a problem of various popular machine learning algorithms (for example, k-means) and making comparative analysis with them. In the work [8], a binary classification problem was solved to determine whether the chronic venous insufficiency exist or not. Authors studied 50 patients and their 5200 images from magnetic resonance venograms. These images were proceeded with the Google collaboratory network to train a 121-layer dense-net convolutional neural network. CEAP (clinical – etiology – anatomy – pathophysiology) classification was used as globally accepted and widespread. CEAP classification contains seven stages of CVD [9]: C0 (normal legs with no visible signs of CVD), C1 (spider and reticular veins), C2 (varicose veins, which have a diameter of 3 mm or more), C3 (Edema), C4 (Skin and subcutaneous tissue changes secondary to CVD), C5 (Healed venous ulcer), C6 (Active venous ulcer).

The accuracy of distinguishing between normal (C0 - C2) and abnormal (C3-C6) images was reached about 97%. Classes C3-C6 of CVD can be considered as chronic venous insufficiency according to the CEAP classification [9].

In the work [10] the results of developing of an automated system for classifying CVD conditions as mild, moderate and severe is described. The system was trained on 271 photos taken by certified doctors of vascular surgery, the background of the photographs was uniform. The Bag of Visual Words [11] was used as a neural network. The authors claim very good accuracy of the resulting model (up to 0.95 for mild, 0.89 for moderate, and 0.95 for severe). But the requirements for images, like certain position of leg on the photo or uniform background, make it difficult to use it in practical life for self-diagnosis of patients.

Therefore, the creation of an automated system for self-diagnosis of patients, which does not have strict requirements for the quality of foot photographs, remains an urgent problem.

The aim of this work is to develop a system for predicting the degree of CVD for self-diagnosis of patients using deep learning algorithms.

To achieve this goal, the following tasks were solved:

1. Data mining of images of men's and women's legs with varying degrees of CVD, including those with wounds and tattoos.
2. Filter “legs - not legs”, which is a necessary step so that only legs get into the main classification algorithm.
3. CVD degree classifier according to CEAP classification.

2. Materials and Methods

2.1 Data mining

One of the main problems for solving classification problems using deep learning algorithms is the problem of collecting enough high-quality images for each of the predicted classes. In addition, it is desirable that each class contains approximately the same number of images.

To solve some problems, we can use open databases with verified medical images, or conduct a study in medical institutions and thus collect the image dataset. But for many medical problems, there are no datasets with needed images, or the gathering the necessary quantity of images takes too long time.

In this case, data mining of images from open Internet sources, such as Instagram or Google images, can help to collect the required number of images.

Our data mining process included several stages that differed specifically in image's collecting ways. So, two ‘spider’ scripts were developed. The first one was developed using Scrapy, which is a fast, open-source web crawling framework written in Python for

 doi:10.20944/preprints202208.0495.v1

extracting the data from the web pages [12]. This script was used for getting images from Instagram accounts. The second spider was developed using Selenium and was used in the Chrome browser for getting images from Google. Selenium is a Python web-driver which allows an emulation of real user behavior on web pages [13].

2.1.1 Scrapy data mining

In the first stage, about 300 Instagram accounts with a big quantity of legs with CVD were selected for data mining. 67 000 images were collected, then no-legs and low-quality images were removed by hand. The number of remaining images was 10 618, these images showed legs with varying degrees of CVD. Thus, we got our dataset "legs with CVD".

In order to prevent manual work on separating images with and without legs in the future, a filter “leg - no leg” was trained using Resnet50 [14]. Since the no-legs images contained too many objects and some of them were poor quality, to solve the binary classification problem and train the “legs - not legs” filter, an open Flickr Image dataset containing more than 30,000 annotated images was used. From the Flickr Image dataset, 11,000 images were taken for the “not legs” category.
Thus, a balanced dataset “legs - no legs”, consisted of 10,618 images of legs and 11,000 images of “non-legs” was formed. The dataset “legs - no legs” was used for solving the problem of binary classification based on Resnet50.

2.1.2 Selenium data mining

The dataset "legs with CVD" was classified by a certified phlebologist into 7 CEAP classes - from C0 to C6. And the class C0 contained practically no images. Therefore, it was necessary to find and add more images with healthy legs.

So, we developed and ran the Selenium spider in the browser Chrome in order to find images of healthy legs through search queries in the Google search engine to fill the C0 class. We collected about 90 000 images, but only 10% of them were unique.

The unique images were passed through the “legs - no legs” filter, and we got only 400 images with healthy legs. These images were added to the C0 class of the dataset “legs with CVD".

Thus, we have collected the dataset “legs with CVD" of 11,118 images with different degrees of CVD.

2.2 Neural networks

2.2.1 Filter “legs - no legs”

The dataset “legs - no legs” were split into training and validation parts in the proportion 80/20 %.

As baseline for this binary classification problem, the pre-trained model ResNet50 (a residual neural network with 50 layers) [15] from the PyTorch library [16] was chosen.

ResNet is one of the most cited and most popular artificial neural networks for images classification problems. The fundamental difference between ResNet and classical convolutional neural networks is shortcut connections. That means that some images process through all layers of ResNet, but for the rest of the images, some layers are skipped. So ResNet is not fully connected networks, and large numbers of layers do not overfit the model, and do not reduce accuracy, unlike classical CNNs.

As the dataset “legs - no legs” contains two well-balanced class “legs” and “no-legs”, the quality of prediction was evaluated by accuracy metric, which is the ratio of the correct predictions number to the total number of predictions - correct and incorrect.

2.2.2 Multi classification problem

To solve the CVD grade classification task, 11,118 foot images were categorized by a certified phlebologist into 7 CEAP classes (C0 - C6). After classification of the images, the following distribution of images by CEAP classes was got: C0 - 872, C1 - 2810, C2 - 1493, C3 - 3743, C4 - 1530, C5 - 403, C6 - 267. The ratio of the number of images in each class to the total number of images in the dataset is shown in Fig. 2.

To solve the problem of classification according to the degree of CEAP, two neural networks were chosen - Resnet50 and DeiT [17], with completely different architecture. So, if ResNet50 is a variant of a convolutional network, then DeiT can be considered as a variant of Graph Neural Networks. The metrics of models based on Resnet50 and DeiT were compared to choose the best solution.

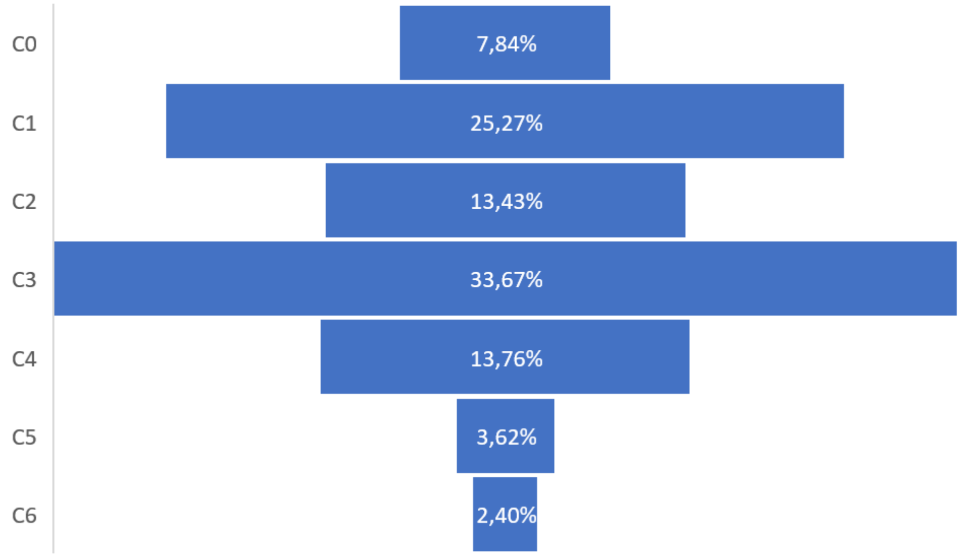


Figure 2. Images by CEAP classes (the ratio in percent of the total).

Since the imbalance of classes is clearly traced, as it can be seen in Fig.2, so the accuracy can be misleading. So, according to many works, for example [18], the following metrics have to be used for the evaluation of the quality of classification models for the imbalanced dataset:

- Precision = TruePositive / (TruePositive + FalsePositive)
- Recall = TruePositive / (TruePositive + FalseNegative)
- F-Measure = (2 * Precision * Recall) / (Precision + Recall)
- Logistic Loss curve.

A rated confusion matrix was also built according to the scheme:

		Predicted	
		Positive	Negative
Actual	Positive	Rated TP = TruePositive/ActualPositive	Rated FN = FalseNegative/ActualPositive
	Negative	Rated FP = FalsePositive/ActualNegative	Rated TN = TrueNegative/ActualNegative

where TruePositive - the number of images which were classified correctly; FalseNegative - the number of images that should have been assigned to this class but were not; FalsePositive - the number of images that were erroneously assigned to this class; TrueNegative - the number of images correctly marked as not being in this class; ActualPositive - the number of all images, which are actually in this class; ActualNegative - the number of all images, which are actually not in this class.

The Logistic Loss (LogLoss) curve is defined as:

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} a_{ij}\log p_{ij}$$

$$a_{ij} = \begin{cases} 1, & if\ object\ i\ belongs\ to\ class\ j \\ 0, & otherwise \end{cases}$$

where N – the number of images, C – the number of classes, p_{ij} – the probability of classifying *i* object to class *j*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0495.v1

3. Results

3.1 Resnet50 for filter "legs - no legs"

The deep leaning model based on Resnet50 for the binary classification problem were trained for the one epoch. The LogLoss value reached a constant value, closed to zero, already at about the 5th step (Fig. 4). The accuracy was 0.998, which is a good result for binary classification problems on a well-balanced dataset.

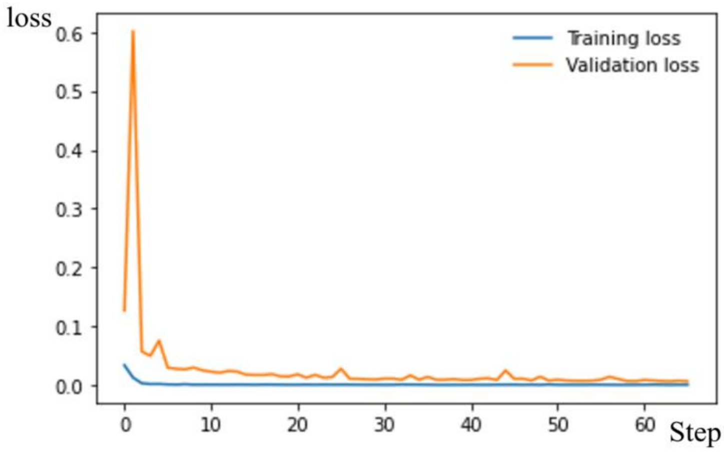


Figure 4. The Logistic Loss curve for the binary classification model "legs - no legs".

3.2 Resnet50 for the multi classification problem

For this problem, the model based on Resnet50 were trained for the twenty epochs.

Although the learning curve of the logistic loss is slowly decreasing, further training is not advisable since the validation LogLoss curve has reached a nearly constant value (Fig.5).

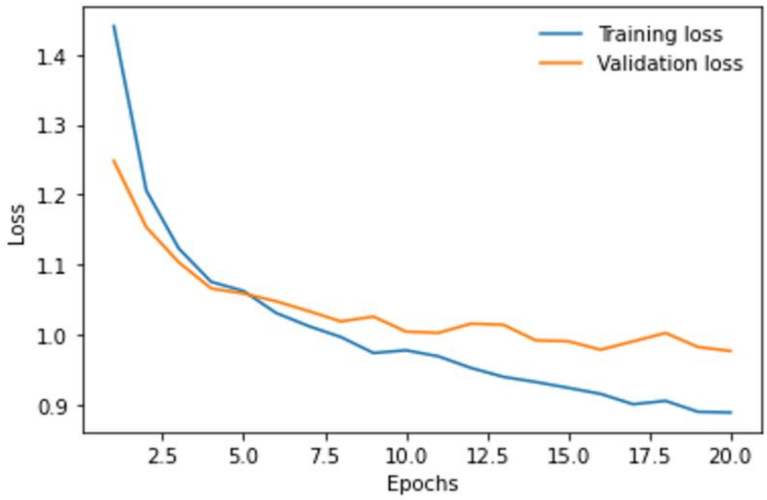


Figure 5. The Logistic Loss curve for the multi classification model based on Resnet50.

The metrics were not good enough, the precision was 0.615, the recall - 0.614, the F1 metric - 0.611. To identify problems in classification, the rated confusion matrices were constructed for each CEAP class (Appendix A).

As can be seen in Figures A1-A4, the model based on Resnet50 predicts an image does NOT fit into a certain class well enough, with a probability of at least 0.85 (for C3), and with a probability of over 0.97 for classes C0, C5, C6.

But the model does not perform well on predicting that an image belongs to a certain class. So, we have the probability near 0.8 that the image will be correctly classified only for classes C0 and C1. This is a remarkably interesting result, since C0 contains only 7.84% of the total number of images in the dataset.

The probabilities of correctly assigning an image to class C2 or C3, are almost equal to the probabilities of making a Type I error (for C2: 0.52 and 0.48; for C3: 0.6 and 0.4). Although it should be noted that the probability of correct classification equal to 0.6 for the multi classification problem can be considered as a fairly good result.

The probability of making a Type I error for C4, C5 and C6 classes is higher than the probability of correctly classifying an image in these classes. The worst case is for C5, where the probability of Type I error is 0.71, and the probability of the correct classification is only 0.29.

The probability of a Type II error does not exceed 0.1 for all classes, except for C3, for which it is equal to 0.15.

Thus, the quality of the classification could not be considered acceptable. But since there were no ways to significantly increase the classification accuracy based on the constructed model, it was decided to use a different, more complex neural network.

3.3 DeiT multi classification problem

The deep leaning model based on DeiT for the multi classification problem were trained for the twenty epochs as the same for the model based on Resnet50. The LogLoss curves for training and validation sets are shown in Figure 6.

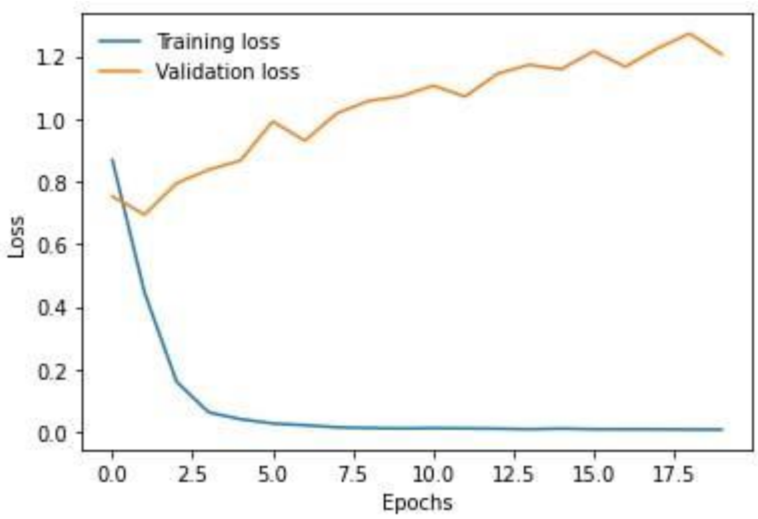


Figure 6. The Logistic Loss curve for the multi classification model based on DeiT.

As can be seen in Fig.6, the model is overtrained after 2 epochs. It was taken the weights of the model trained on 2 epochs only.

The metrics were much better than for the model based on Resnet50. So, the precision was 0.770, the recall - 0.766, the F1 metric - 0.768.

The rated confusion matrices which were constructed for each CEAP class for this model are shown in Appendix B.

As can be seen in Table 1 and Figures B1-B4, the classification accuracy has improved significantly for almost all classes. The probability of correct classification that an image does not fit the class is not less, than 0.95 for all CEAP classes. This probability is equal to 0.99 for C0, C5, C6.

Table 1. The probabilities of classification for each CEAP class.

Model	Model	Rated TP	Rated TN	Rated FP	Rated FN
C0	Resnet50	0.80	0.97	0.20	0.029
	DeiT	0.76	0.99	0.24	0.001
C1	Resnet50	0.79	0.91	0.21	0.086
	DeiT	0.86	0.94	0.14	0.055
C2	Resnet50	0.52	0.90	0.48	0.098
	DeiT	0.63	0.95	0.37	0.055
C3	Resnet50	0.60	0.85	0.40	0.150
	DeiT	0.83	0.90	0.17	0.099
C4	Resnet50	0.47	0.91	0.53	0.085
	DeiT	0.70	0.94	0.30	0.058
C5	Resnet50	0.29	0.91	0.71	0.030
	DeiT	0.40	0.99	0.60	0.014
C6	Resnet50	0.40	0.99	0.60	0.012
	DeiT	0.55	0.99	0.45	0.009

So the model based on DeiT can very well classify an image as C0, C1, C3, quite well as C2 and C4, and not good as C5 and C6. So further improvement of the model is needed. This model improvement can be made by increasing of the quantity of images in the training dataset for classes C5 and C6, and by tuning the model based on the DeiT.

4. Discussion

The model based on DeiT showed better results than the model based on Resnet50 for multi classification problem of CVD into C0-C6 CEAP classes. On the other hand, the filter "legs - no legs" were trained by Resnet50, and the accuracy of this model is 0.998. The advantages of Resnet50 are simple architecture, quick training even on CPU, easy implementation. Whereas DeiT is difficult to implement and its training takes a long time. So, the training of one epoch for the model based on DeiT takes about 10 hours on the CPU for the considered dataset. Thus, for practical use, the combinations of these models can be used. For example, the filter "legs - no legs" can be used as the first step, so images without legs or bad quality images will be filtered out and will not be passed to the main model - the multi classification model based on DeiT.

But the current quality of the model based on DeiT cannot be acceptable for practical use, since the probability of correct prediction for C5 and C6 classes are 0.4 and 0.55, respectively.

Therefore, in further research, the following steps seem advisable. At first, the training and validation datasets must be enlarged, especially for C5 and C6 classes. At second, the model based on DeiT can be improved, since we used only direct approaches and did not use any tuning possibilities. Thirdly, as patients could have tattoos on legs, or wounds not associated with CVD, we must collect more such images for the training and validation datasets for the filter "legs - no legs" and for the multi classification model.

We ran the Telegram bot @VaricoseVeinsCheck_bot for testing developed deep learning models. Any user can send photo of his/her legs and get prediction about the health of their legs.

Supplementary Materials: The latest trained model can be tested via the Telegram bot https://t.me/VaricoseVeinsCheck_bot (or @VaricoseVeinsCheck_bot).

Author Contributions: Conceptualization, M.B. and A.S.; methodology, M.B.; software, S.O., I.U. and I.O; validation, S.O., I.U. and I.O; formal analysis, S.O.; investigation, S.O. and I.U.; resources, M.B.; data curation, I.U.; writing—original draft preparation, S.O., I.U. and I.O; writing—review

 doi:10.20944/preprints202208.0495.v1

and editing, M.B. and A.S.; visualization, S.O.; supervision, M.B.; project administration, M.B.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript."

Funding: This research received no external funding.

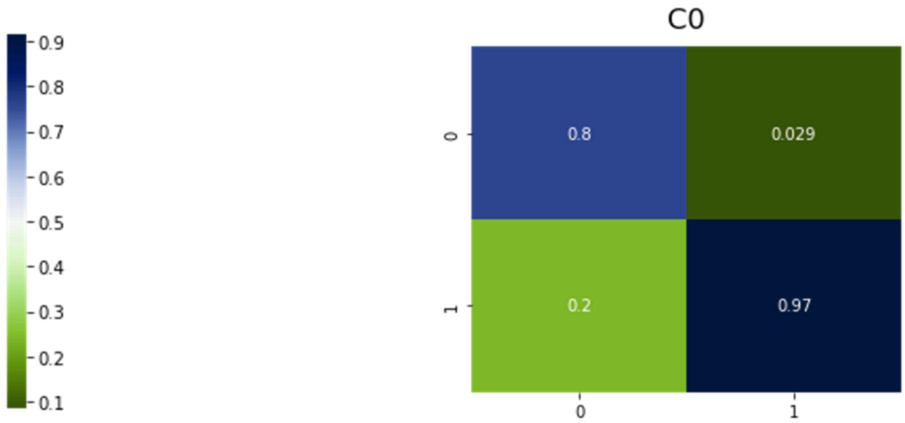
Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

 doi:10.20944/preprints202208.0495.v1

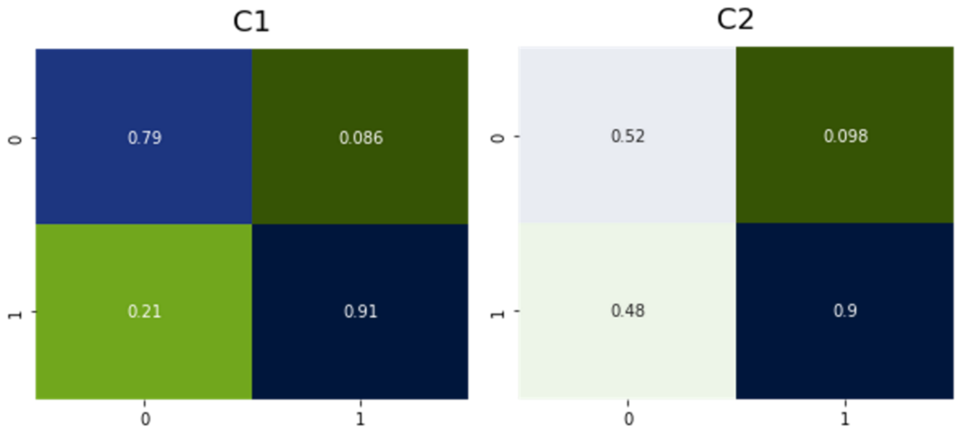
Appendix A

The rated confusion matrices for each CEAP class for the multi classification model based on Resnet50.



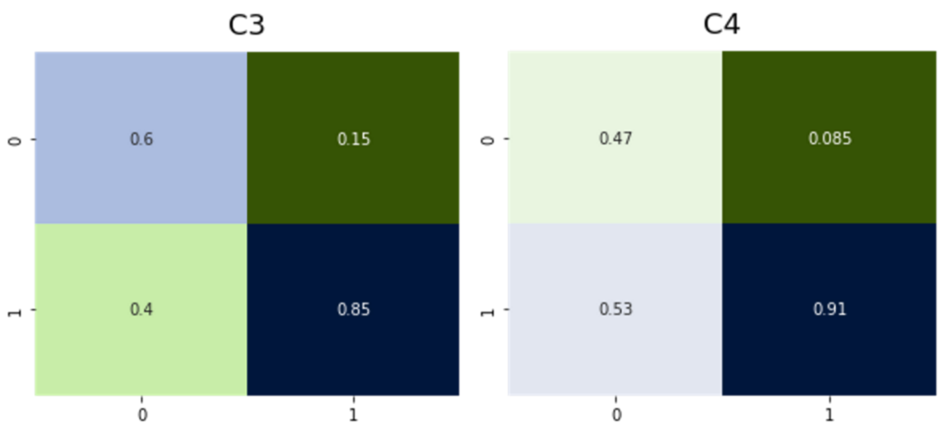
(**a**) (**b**)

Figure A1. The color scheme for confusion matrices for the multi classification model based on Resnet50 (a) and the confusion matrices for C0 class.



(**a**) (**b**)

Figure A2. The multi classification model based on Resnet50. The confusion matrices for C1 (a) and C2 (b) classes.



(**a**) (**b**)

Figure A3. The multi classification model based on Resnet50. The confusion matrices for C3 (a) and C4 (b) classes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 August 2022 doi:10.20944/preprints202208.0495.v1

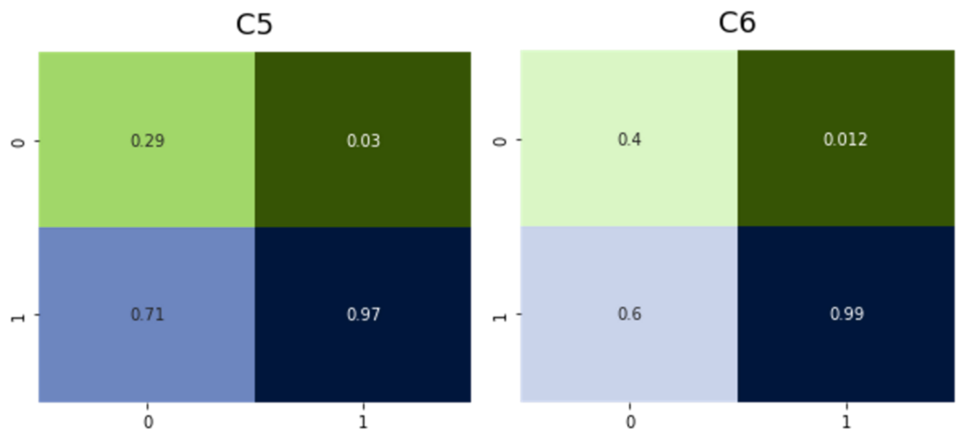
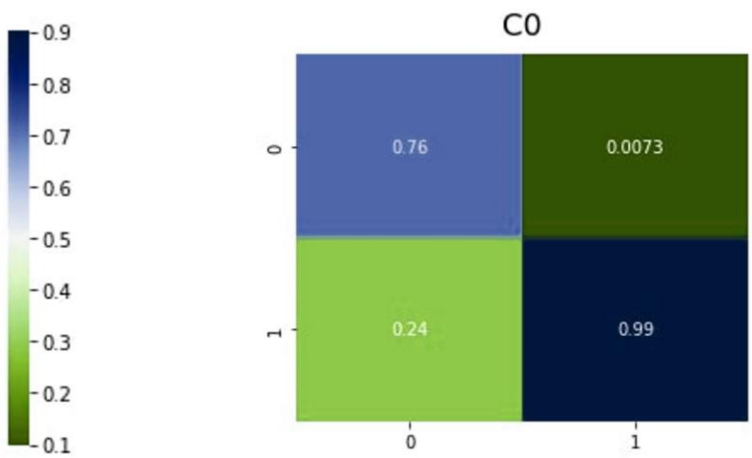


Figure A4. The multi classification model based on Resnet50. The confusion matrices for C5 (a) and C6 (b) classes.

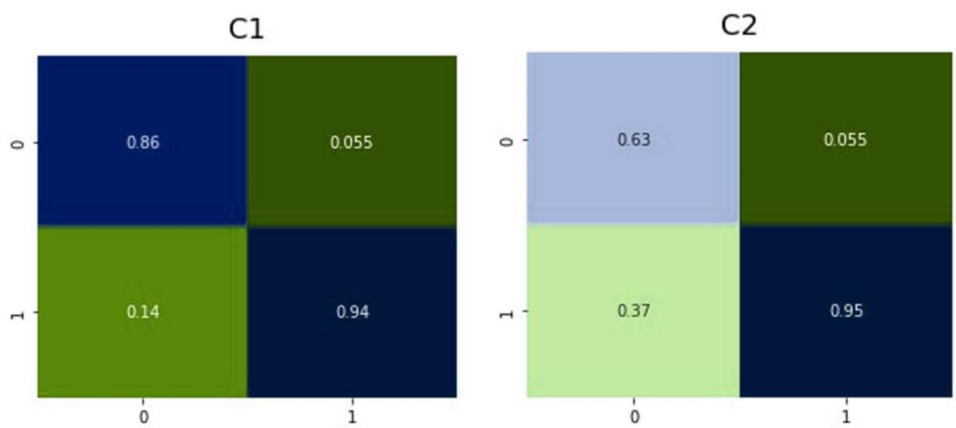
Appendix B

The rated confusion matrices for each CEAP class for multi classification model based on DeiT.



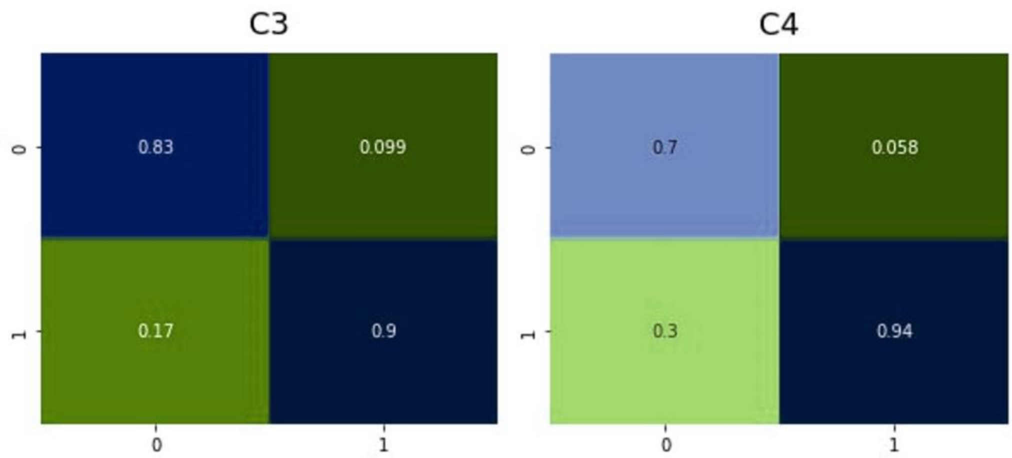
(a) (b)

Figure B1. The color scheme for confusion matrices for the multi classification model based on DeiT (a) and the confusion matrices for C0 class.



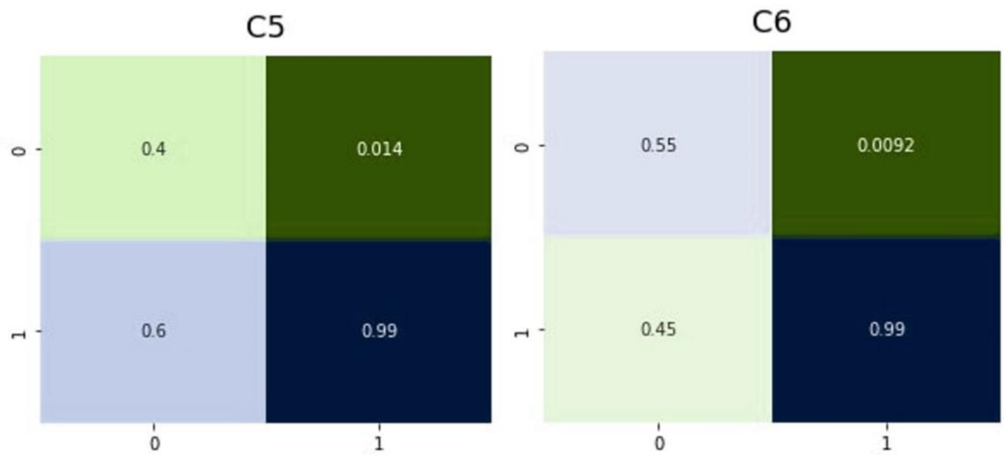
(a) (b)

Figure B2. The multi classification model based on DeiT. The confusion matrices for C1 (a) and C2 (b) classes.



(**a**) (**b**)

Figure B3. The multi classification model based on DeiT. The confusion matrices for C3 (a) and C4 (b) classes.



(**a**) (**b**)

Figure B4. The multi classification model based on DeiT. The confusion matrices for C5 (a) and C6 (b) classes.

References

1. Vuylsteke, M. E.; Colman, R.; Thomis, S.; Guillaume, G.; van Quickenborne, D.; Staelens, I. An Epidemiological Survey of Venous Disease Among General Practitioner Attendees in Different Geographical Regions on the Globe: The Final Results of the Vein Consult Program. *Angiology* **2018**, 69(9), 779–785.
2. Feodor, T.; Baila, S.; Mitea, I.-A.; Branisteanu, D.-E.; Vittos, O. Epidemiology and clinical characteristics of chronic venous disease in Romania. *Experimental and Therapeutic Medicine* **2019**, 17(2), 1097–1105.
3. Carlton, R.; Mallick, R.; Campbell, C.; Raju, A.; O'Donnell, T.; Eaddy, M. Evaluating the Expected Costs and Budget Impact of Interventional Therapies for the Treatment of Chronic Venous Disease. *American Health & Drug Benefits* **2015**, 8(7), 366.
4. Epstein, D.; Gohel, M.; Heatley, F.; Davies, A. H. Cost-effectiveness of treatments for superficial venous reflux in patients with chronic venous ulceration. *BJS Open* **2018**, 2(4), 203–212.
5. Ortega, M. A.; Fraile-Martínez, O.; García-Montero, C.; Álvarez-Mon, M. A.; Chaowen, C.; Ruiz-Grande, F.; Pekarek, L.; Monserrat, J.; Asúnsolo, A.; García-Honduvilla, N.; Álvarez-Mon, M.; Bujan, J. Understanding chronic venous disease: A critical overview of its pathophysiology and medical management. *Journal of Clinical Medicine* **2021**, 10(15), 3239.
6. Bailey, M.; Solomon, C.; Kasabov, N.; Greig, S. Hybrid systems for medical data analysis and decision making-a case study on varicose vein disorders. *Proceedings 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems* **1995**, 265-268
7. Zhu, R.; Niu, H.; Yin, N.; Wu T.; Zhao Y. Analysis of Varicose Veins of Lower Extremities Based on Vascular Endothelial Cell Inflammation Images and Multi-Scale Deep Learning. IEEE Access **2019**, 7, 174345-174358.
8. Atreyapurapu, V.; Vyakaranam, M. I.; Atreya III, S.; Gupta, P. I.; Atturu, G. V. Assessment of Anatomical Changes in Advanced Chronic Venous Insufficiency Using Artificial Intelligence) and Machine Learning Techniques. *Journal of Vascular Surgery: Venous and Lymphatic Disorders* **2022**, 10(2), 571–572.

9. Ortega, M. A.; Fraile-Martínez, O.; García-Montero, C.; Álvarez-Mon, M. A.; Chaowen, C.; Ruiz-Grande, F.; Pekarek, L.; Monserrat, J.; Asúnsolo, A.; García-Honduvilla, N.; Álvarez-Mon, M.; Bujan, J. Understanding chronic venous disease: A critical overview of its pathophysiology and medical management. *Journal of Clinical Medicine* 2021, 10(15).
10. Shi, Q.; Chen, W.; Pan, Y.; Yin, S.; Fu, Y.; Mei, J.; Xue, Z. An Automatic Classification Method on Chronic Venous Insufficiency Images. *Scientific Reports* **2018**, 8(1), 1–14.
11. Zhang, F.; Song, Y.; Cai, W.; Hauptmann, A.G.; Liu, S.; Pujol, S.; Kikinis, R.; Fulham, M.J.; Feng, D.D.; Chen, M. Dictionary pruning with visual word significance for medical image retrieval. *Neurocomputing* **2016**, 177, 75-88.
12. Scrapy. Available online: https://scrapy.org/ (accessed on 25 08 2022).
13. Selenium automates browsers. That's it! Available online: https://www.selenium.dev/ (accessed on 25 08 2022).
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016**, 770–778.
15. https://pytorch.org/hub/nvidia_deeplearningexamples_resnet50/
16. PyTorch. Available online: https://pytorch.org/ (accessed on 25 08 2022).
17. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint* **2020,** arXiv:2010.11929.
18. Hoens, T.R.; Chawla, N.V. Imbalanced Datasets: From Sampling to Classifiers. In *Book Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed., Yunqian Ma, Haibo He., Wiley-IEEE Press; 2013; pp. 43 - 60