



Augmented Reality Aided Pre-Diagnosis Environment For Telemedicine: Superficial Vein Surveillance System

Huseyin A. Erdem^{1*}, Semih Utku²

^{1*} Dokuz Eylül University, The Graduate School of Natural and Applied Sciences, Department of Computer Engineering, İzmir, Turkey, (ORCID: 0000-0002-5720-0017), huseyinaerdem@gmail.com

² Dokuz Eylül University, Faculty of Engineering, Department of Computer Engineering, İzmir, Turkey, (ORCID: 0000-0002-8786-560X), semih@cs.deu.edu.tr

(First received 22 April 2022 and in final form 2 August 2022)

(DOI: 10.31590/ejosat.1107531)

ATIF/REFERENCE: Erdem, H. A. & Utku, S. (2022). Augmented Reality Aided Pre-Diagnosis Environment for Telemedicine: Superficial Vein Surveillance System, *European Journal of Science and Technology*, (38), 376-385.

Abstract

The proposed system creates a virtual pre-diagnosis environment that can detect narrowings in superficial veins by using the near-infrared video images. In the study, the near-infrared video recordings of the tissue to be followed are taken by the user in the home environment via the smart device. The images obtained by improving the discontinuous structures in the vein images undergone the image pre-processing phase are classified by using a hybrid decision-making algorithm that evaluates two separate convolutional neural network models together. According to the results of the hybrid decision-making algorithm, the imaged regions could be classified with Model-1 (Accuracy Rate (0.872), Misclassification Rate (0.128), Precision (0.372), Prevalence (0.500) and F-Score (0.496)) and Model-2 (Accuracy Rate (0.816), Misclassification Rate (0.184), Precision (0.407), Prevalence (0.500) ve F-Score (0.543)) without the need for large amounts of training dataset. In the study, the detected vein narrowings in the vein images are marked on the relevant location. The marked images are superimposed on the real images and the narrowing progress process is presented to the user and his/her physician as a telemedicine application in the form of a video-based indirect augmented reality environment representing a long time interval (week, month, year).

Keywords: Near-Infrared Light, Vein Narrowing Detection, Convolutional Neural Network, Yolov3, Augmented Reality, Telemedicine.

Teletıp İçin Artırılmış Gerçeklik Destekli Ön-Teşhis Ortamı: Yüzeysel Damar Takip Sistemi

Öz

Önerilen sistem, yakın-kızılötesi video görüntülerini kullanarak yüzeysel damarlardaki daralmaları tespit edebilen sanal bir ön-teşhis ortamı oluşturmaktadır. Çalışmada, takip edilecek dokunun yakın kızıl-ötesi video kayıtları akıllı cihaz aracılığıyla kullanıcı tarafından ev ortamında alınmaktadır. Görüntü ön-işleme aşamasından geçirilen damar görüntülerindeki kesikli yapılar giderilerek elde edilen görüntüler, iki ayrı evrişimsel sinir ağı modelini birlikte değerlendiren hibrit karar verme algoritması kullanılarak sınıflandırılmaktadır. Hibrit karar verme algoritması sonuçlarına göre, görüntülenen bölgeler, Model-1 (Doğruluk Oranı (0.872), Yanlış Sınıflandırma Oranı (0.128), Kesinlik (0.372), Yaygınlık (0.500) ve F-Skoru (0.496)) ve Model-2 ile (Doğruluk Oranı (0.816), Yanlış Sınıflandırma Oranı (0.184), Kesinlik (0.407), Yaygınlık (0.500) ve F- Skoru (0.543)) büyük miktarda eğitim verisine ihtiyaç duyulmadan sınıflandırılmıştır. Çalışmada, damar görüntülerinde tespit edilen damar daralmaları, ilgili konum üzerine işaretlenmektedir. İşaretli görüntüler, gerçek görüntüler üzerine bindirilmekte ve daralma gelişim süreci, uzun bir zaman aralığı (hafta, ay, yıl) temsil eden video-tabanlı dolaylı artırılmış gerçeklik ortamı şeklindeki bir uzaktıp uygulaması olarak kullanıcıya ve hekimine sunulmaktadır.

Anahtar Kelimeler: Yakın Kızıl-Ötesi Işık, Damar Daralma Tespiti, Evrişimli Sinir Ağı, Yolov3, Artırılmış Gerçeklik, Uzaktıp.

* Sorumlu Yazar: huseyinaerdem@gmail.com

1. Introduction

Image-based applications currently used for smart devices (smart phones, personal digital assistants, tablet computers etc.) address areas such as gaming, health, banking and social media. The Near-Infrared Radiation (NIR, the part of the infrared region in the wavelength range of 700 to 900 nm (Ferrari et al., 2004; Rao et al., 2017; Şayli et al., 2004)) light region of the electromagnetic spectrum is used as an imaging technique in image processing studies carried out in the healthcare field. NIR light has been currently used in hospitals as a method of illuminating the tissue to be imaged in various applications. These applications include vein visualization in order to facilitate vascular access, determining the amount of oxygen in the blood, or patient registration with palm vein recognition.

Cholesterol accumulated on the walls of the veins results in narrowing in the vein over time and prevents the passage for blood flow, and may cause occlusion in the veins in the following periods (weeks, months, years). Sometimes, blood clots can block blood flow by blocking blood veins. In this study, a virtual pre-diagnosis environment that can detect vascular narrowings within the context of telemedicine is presented.

Telemedicine can be defined as a communication system that enables patients in remote locations to exchange health care information (Craig & Petterson, 2005) with specialists (physician, physiotherapist) in order to receive health services through information systems. In particular, it seems that the need for patient follow-up with telemedicine has increased even more during the Coronavirus Disease 2019 (Covid-19) pandemic (Chunara et al., 2020). In the study which is carried within the scope of the ongoing doctoral thesis, in the frame of telemedicine, it will be possible to prevent potential delays in treatment by providing early diagnosis of vascular narrowings with the help of the proposed Augmented Reality (AR) based virtual environment. Digital image processing, classification or object detection required to create such an environment are processes which necessitate high performance and long training periods. For this reason, it is not possible to perform these operations on smart devices that appeal to general use. Performing these operations on a server with high computing and hardware capabilities will enable the system to work more effectively. In the system, the user is only responsible for taking the NIR video recordings with the NIR camera connected to his/her smart device (preferably via USB connection) and for uploading them to the server. After the video recordings are uploaded to the server, first the vascular structures are obtained on the server, the images are combined in order to increase the vein integrity, then it is determined whether the veins in the images belong to the same tissue region or not, and finally, narrowing is detected. As the last step, a video-based indirect AR environment is created by superimposing virtual veins on real tissue (right or left forearm in the study) images in such a way that corresponds to their real positions. Additionally in the study, in cases where the image quality is insufficient and the number

of images is low, the contribution of both the proposed video-based image acquisition system and the hybrid decision-making algorithm to the classification accuracy has been evaluated.

2. Material and Method

The proposed system shown in Figure 1 is discussed in six stages as hardware, image pre-processing, image post-processing, classification, object detection and AR.

2.1. Hardware Phase

The spectroscopy technique examines the interaction of matter and light. In this technique, matter (in other words, molecules in matter) is illuminated by photons of light in a specific region of the electromagnetic spectrum. By measuring and interpreting the interaction of illuminated molecules with photons, information about the structure of the material is obtained. In non-invasive superficial vein imaging studies, NIR spectroscopy technique (Alwazzan, 2020; Elnasir & Shamsuddin, 2014; Huda et al., 2021; Shrotri et al., 2010; Yılmaz, 2014) was used in terms of hardware. Hemoglobin molecules are sensitive to NIR photons. In the first NIR optical window (wavelength range of 700-900 nanometers) of the electromagnetic spectrum, the absorption coefficient of NIR photons is higher for Hemoglobin molecules in the 700-800 nm range in the veins (which has transferred its oxygen to the surrounding tissues, Hb), in the 800-900 nm range in the arteries (oxygen carrying, HbO₂), respectively (Sordillo et al., 2014; Wadhvani et al., 2015; Wang & Leedham, 2006). NIR light in the range of 740 to 940 nm can penetrate the skin up to 3-5 mm (Meng et al., 2015; Wadhvani et al., 2015). In the studies, NIR light with a wavelength of 750 nm (Crisan et al., 2007), 850 nm (Rao et al., 2017; Mangold et al., 2013; Meng et al., 2015; Tien et al., 2015; Wang & Leedham, 2006) or 960 nm (Francisco et al., 2021) was used for vein imaging. Among them, 850 nm wavelength was preferred more in terms of providing better NIR images. In terms of the physical placement of the NIR Light Emitting Diodes (LEDs), the best light distribution in vein imaging applications was stated to be provided when the camera is in the center and the LEDs are arranged circularly around it (Shrotri et al., 2010; Şeker & Engin, 2017; Wadhvani et al., 2015). There are also studies where disc-shaped placements (Şeker & Engin, 2017) or sequential LED placement (Meng et al., 2015) on head-mounted display glasses (within the scope of wearable vein visualization system offering AR) are used.

In the hardware phase of this study, a USB-supported external NIR camera with a resolution of 2 megapixels, in which 5 NIR LEDs with a wavelength of 850 nanometers are arranged circularly around the camera lens, was used to illuminate the veins. Smart device cameras are equipped with NIR-blocking (IR-cut off) filter to capture clearer images. There is a high probability of damaging of the camera lens during removal of these filters within the scope of NIR imaging. Therefore, in this study, it is preferred to use an external NIR camera that can be easily integrated with the user's smart device via USB connection.

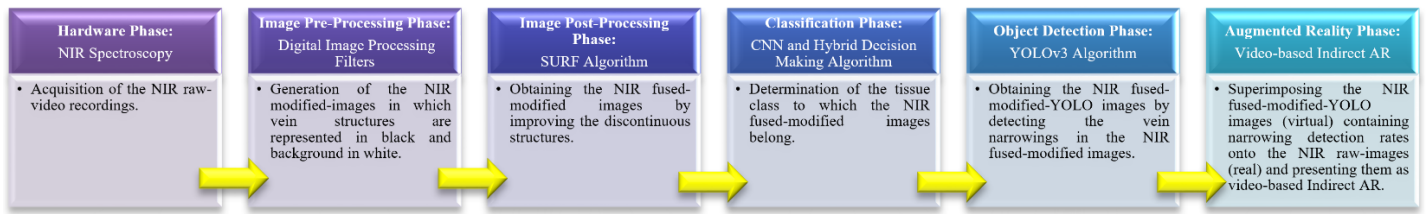


Figure 1. Flow chart of the proposed system phases.

Although processing only one image (taken directly as photo frame) per week reduces the processing load during the image processing phase, it makes it difficult to detect the entire vein from this one image in low camera resolution. Furthermore, it will not meet the needs of hundreds of images to be required during classification phase. On the other hand, the need for asking the user to take more than one photo frame for fixing the tissue region to be visioned hinders the ease of use of the application. For this reason, during the image taking part of the study, it was evaluated that instead of photography video shooting was performed to eliminate these problems.

2.2. Image Pre-Processing Phase

The 25-second videos recorded with the NIR camera are 30 fps (30 frames per second), each video consisting of 750 frames in total. Every video with 1920x1080 resolution was firstly converted into 1920x1080x3 uint8 type in the feature of RGB NIR raw-image files with bitmap extension. NIR raw-images were image pre-processed (In this section, the image processing steps used in Francis et.al.'s (2017) study are utilized. Unlike their study, extraction of Region of Interest (ROI) part, which expresses the part of the image where only the veins of interest are located, was applied before the image processing steps in order to reduce the processing load. This phase of the system, which was developed within the scope of the ongoing doctoral study, was introduced in (Erdem et al., 2020). Different from these two studies, in this study, the veins were obtained in black with a white background in the final images.) by using the MATLAB® R2017a (MathWorks, 1996) program and the resulting images are given in Figure 2.

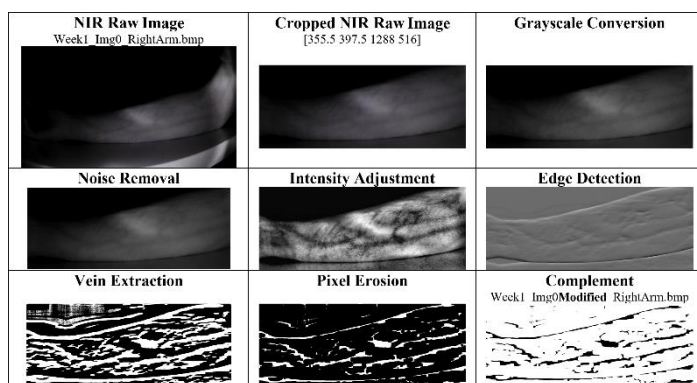


Figure 2. Image pre-processing steps (applied to image with code Week1_Img0_RightArm.bmp)

In order to extract the vein patterns with the image pre-processing phase, first of all, the ROI where the vein narrowing will be detected was determined in the three dimensional (with Red Green Blue-RGB values) NIR raw-image. Images were reduced to two dimensions (hue and saturation information in the image are deleted, luminance information is preserved

(MathWorks, 1996)) by grayscale conversion in order to apply image processing filters on the ROI (truecolor RGB image of 445x947x3 uint8 type) clipped from the parts that will not be used within the scope of narrowing detection. Then a median filter was applied to remove noise. For intensity adjustment in the images, adaptive histogram equalization was conducted consecutively and then the median filter was used again. Edge detection of veins was performed with the gabor filter which can be applied to two dimensional images. At a level of bitwise, vein extraction was applied to highlight vein patterns. By removing the small-sized structures that do not represent the veins through the morphological pixel erosion process, the veins in the images were further smoothed in shape. As the final step, the black and white colors in the images were inverted by complement and NIR modified-images in which the vein structures were represented in black were obtained.

2.3. Image Post-Processing Phase

When photographing the tissue area illuminated by NIR light, it was observed that some vascular structures which should be continuous were displayed intermittently and a complete integrity could not be obtained for the vascular structure. Among the main factors leading to this circumstance are the angle of illumination, camera position and resolution, vibration that may occur on the camera during shooting and so on. The aim of the image post-processing phase is to create an imaging system that will be affected as little as possible by such losses that may occur in the image due to external factors. In order to include such feature in the system, the fast working (Mistry & Banerjee, 2017) Speeded-Up Robust Features (SURF) local feature detector function (MathWorks, 1996) was used in this study. By using the SURF local feature detector function, matching points that provide similarity between two images are detected. With the help of these similarity points, the rotation angle or scale value (convergence/divergence) of these two images relative to each other can be determined. The detected rotation angle and scale value are used to bring the angle and/or scale of the second image to the previous one's. For this, matching points for the first and second images were determined as shown in Figure 3, and rotation and/or scaling corrections were made on the second image taking into account the calculated rotation and/or scaling values regarding these points (first image is taken as reference). According to the results calculated with the SURF local feature detector function, when the change in the second pre-processed image does not exceed ± 0.07 angularly and 0 to +0.1 in scale (values were determined by trial and error), correction was applied to the second image. After combining the first image and the corrected second image, all remaining NIR modified-images of the week were also combined in pairs to obtain NIR fused-modified images of the relevant week. In the study, this process was repeated for a period of three weeks (including the Week-1) trying to achieve an integrity in the vascular structure images.

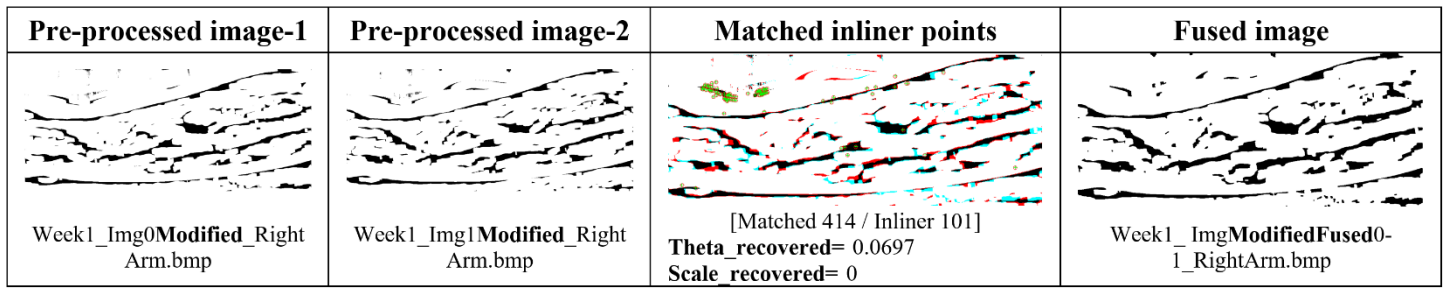


Figure 3. NIR fused-image obtained from two pre-processed images.

In this way, from 750 NIR modified-images generated by the pre-processing phase for Week-1 right forearm, 375 NIR fused-modified images were able to be obtained by the image post-processing phase. NIR fused-modified images created by combining the images of the relevant week were used both in the You Only Look Once version-3 (YOLOv3) (Farhadi & Redmon, 2018) algorithm which performs the narrowing detection process, and in the AR images to be presented to the user and the physician.

2.4. Classification Phase

The NIR fused-modified images obtained for the first week by the image post-processing phase were used in the classification process in order to detect if the same tissue (vascular network) region was displayed in the following weeks.

Since vascular structures are unique to the individual, biometric identification systems (Park, 2011; Wang & Leedham, 2006; Wu et al., 2019) can be developed with vascular imaging applications. In these systems, determination of whether the veins obtained from the imaged tissue region belong to the same person and/or tissue region, is conducted by classification process. In this phase, two models based on the Modified National Institute of Standards and Technology (MNIST) model (MatConvNet Team, 2017), which is a Convolutional Neural Network (CNN) model using the MNIST database (LeCun et al., n.d.), were proposed in order to establish the basic classification technique to determine whether the same tissue (vascular) region is imaged or not. MNIST model which accepts the MNIST database containing handwritten black and white digit images between 0 and 9 as input, can classify among 10 label values corresponding to 10 digits (28x28 sized gray image) (Wu et al., 2021). In the first model used in the study, instead of the digit dataset, the NIR fused-modified image dataset was applied to the input layer of the MNIST model, and the output layer deciding among 10 classes (numbers from 0 to 9) was changed in order to correspond to 2 classes (right forearm and left forearm). As indicated in the MATLAB® R2017a layer table given in Figure 4, in the input layer of the first CNN model, NIR fused-modified image dataset consisting of 28x28 images was used.

Layer:	input	conv	maxpool	conv	maxpool	conv	relu	conv	softmax
Input Data Size (Weights):	28	24	12	8	4	1	1	1	1
Filter Size (Support):	N/A	5	2	5	2	4	1	1	1
Padding:	N/A	0	0	0	0	0	0	0	0
Stride:	N/A	1	2	1	2	1	1	1	1

Figure 4. MNIST CNN Model-1 layers.

In Model-2, the MNIST CNN model has been modified to take 64x64 sized NIR fused-modified vein images as the dataset. In the second model, of which network layers are specified in Figure 5, the input layer data of the MNIST model was changed to correspond to NIR fused-modified image dataset instead of

digit dataset, and the output layer deciding among 10 classes (numbers from 0 to 9) was changed in order to correspond to 2 classes (right forearm and left forearm). Since one of the main objectives of the study is to perform the correct classification with a small number of training data, instead of adding new images to the dataset, a dropout layer (dropout rate: 0.5) was added to Model-2 just before the last convolution layer in order to increase the performance a little more (since, in the post-test trials, it was seen that the number of misclassifications was higher in Model-2). Also, the filter sizes and numbers of the convolution layers were changed to support the new input size. These values were used in order to determine the input data sizes of the next layers by using the formula (1) (Albawi et al., 2017) for 64x64 dataset sizes (NLIDS: Next Layer Input Data Size, PLIDS: Previous Layer Input Data Size, FS: Filter Size, P: Padding, S: Stride).

$$NLIDS = \frac{PLIDS - FS + (2 * P)}{S} + 1 \quad (1)$$

Layer:	input	conv	maxpool	conv	maxpool	conv	relu	dropout	conv	softmax
Input Data Size (Weights):	64	58	29	22	11	4	4	4	1	1
Filter Size (Support):	N/A	7	2	8	2	8	1	1	4	1
Padding:	N/A	0	0	0	0	0	0	0	0	0
Stride:	N/A	1	2	1	2	1	1	1	1	1

Figure 5. CNN Model-2 layers.

2.5. Object Detection Phase

Classical CNN models only detect which class the input images belong to but cannot visualize the position of the object in the image. Object detection algorithms (R-CNN (Girshick et al., 2014), Faster R-CNN (Ren et al., 2016), YOLO (Redmon et al., 2016), SSD (Anzueto-Rios et al., 2016)) are used for both detecting the object or objects in the image and for marking their positions on the image. In object-based detection studies, it has been stated that YOLOv3 algorithm is better than any other object detection algorithm (Faster R-CNN (Ren et al., 2016)) in terms of providing faster accurate classification performance (Abdulghani & Menekşe Dalveren, 2022; Dikbayır & Bülbül, 2020). YOLO algorithms are very popular nowadays, especially in real-time object detection (Tan et al., 2021).

As there is currently no public access dataset for superficial vein narrowing, the narrowing patterns were artificially generated on NIR fused-modified images in this study. Artificial narrowing patterns were created by examining narrowing patterns (Demir, 2019) found in the literature. As indicated in the orange boxes in Figure 6, 7 of the artificial narrowing patterns were manually placed on each NIR fused-modified image of the Week-1 right forearm (at random locations with different angles) and the training data were obtained artificially.

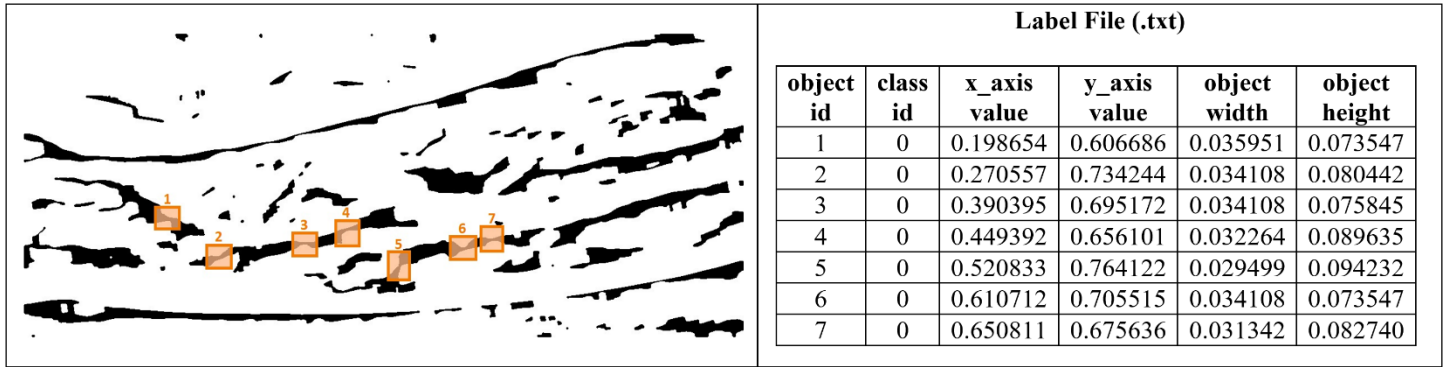


Figure 6. Marking of artificial narrowing locations on NIR fused-modified image.

The training dataset was prepared by using the coordinates and label files of the narrowing patterns created on each NIR fused-modified image to be used as training data for the YOLOv3 algorithm. Areas with created artificial narrowing in this context, were marked on NIR fused-modified images by using makesense.ai (www.makesense.ai) free (under GPLv3 License) web-based photo labelling application, and text-based label files containing the values specified in the Figure 6 were created for each image. A single class label named *stenosis_vein* was defined in the position marking process and the narrowing regions detected in the post-test operation of the algorithm were labelled only with this class.

2.6. Augmented Reality Phase

AR is a virtual environment that is used to create a more developed and detailed perception of reality by superimposing materials (images, information, shapes and so on) created on a computer on top of real images. In general, virtual veins are superimposed on real images within the scope of AR researches, similarly in this study, a virtual-real image combination was created as in (Ai et al., 2016; Francis et al., 2017; Meng et al., 2015). In AR applications, the user can also see the real world at the same time (Doğan et al., 2021). Within this framework, pre-processed virtual images containing only two dimensional vein images were combined with real raw images as in (Francis et al., 2017) and veins were visualized at their real locations. In general, real-time AR applications are designed with AR glasses. However, a real-time AR application cannot be presented because the processes used in this study are performed through the server. For this reason, the visualization of vein tracking was based on the use of a video in which virtual veins are placed on real images. In this way, the user and the physician can see the veins in a video-based indirect AR environment with the results of the narrowing added on the relevant tissue region. Indirect AR, which has been introduced in the literature for outdoor spaces, is effective when a tight match between real images and virtual materials is required (Wither et al., 2011). From this point of view, it has been evaluated that one-to-one matching of veins and tissues can be achieved with indirect AR. In the study, virtual (vein) - real (tissue) videos prepared for the following weeks were added to the videos of the previous week sequentially, and pre-diagnosis data were able to be created to follow the narrowing progress. Another reason why the system was designed as a video-based indirect AR application rather than a real-time AR application is the need to support a

visualization that will span weeks, months, or even years rather than moment-based. Thus, narrowing results were able to be visualized more clearly over time. The obtained AR images are given in the Results section.

3. Results and Discussion

Experimental results were interpreted for classification, object detection and AR phases.

3.1. Classification Phase Results

Experiments on CNN models were carried out using the matConvNet library (The MatConvNet Team, 2017) in the MATLAB® R2017a program. MatConvNet is a MATLAB® R2017a toolbox (Vedaldi & Lenc, 2015) that enables CNN operations to be performed.

The tissue region imaged and recorded in the first week represents the zero point for the recognition process. In order to create a surveillance system, the user is expected to record the same tissue area as the first week in the following weeks. For this reason, only Week-1 (right and left forearm) images were used in the training and pre-test processes, and the images for the following week were evaluated only within the scope of the post-test.

In the study, an application in which the number of training data cannot be increased much in order to ensure ease of use was developed for personal usage. Therefore, only three datasets were used in which only Week-1 right and left forearm NIR fused-modified images were used in different numbers to ensure correct classification with a smaller number of data. Whereas Dataset-1 (200 train/50 pre-test (validation)) and Dataset-2 (400 train/100 pre-test) were only created from Week-1 original NIR fused-modified images, Dataset-3 (augmented data: 600 train/200 pre-test) was built with the help of data augmentation techniques. The training was carried out in two classes, where the right forearm images represent the first class of the first week and the left forearm images represent the second class. In determining the epoch numbers and learning rate, the dropping condition of the objective function below 0.001 value was taken into account. In this context, while the learning rate was 0.0001, 8000 epochs were sufficient for the training of the Model-1 and 3200 epochs for the Model-2. The pre-testing process is performed during the training.

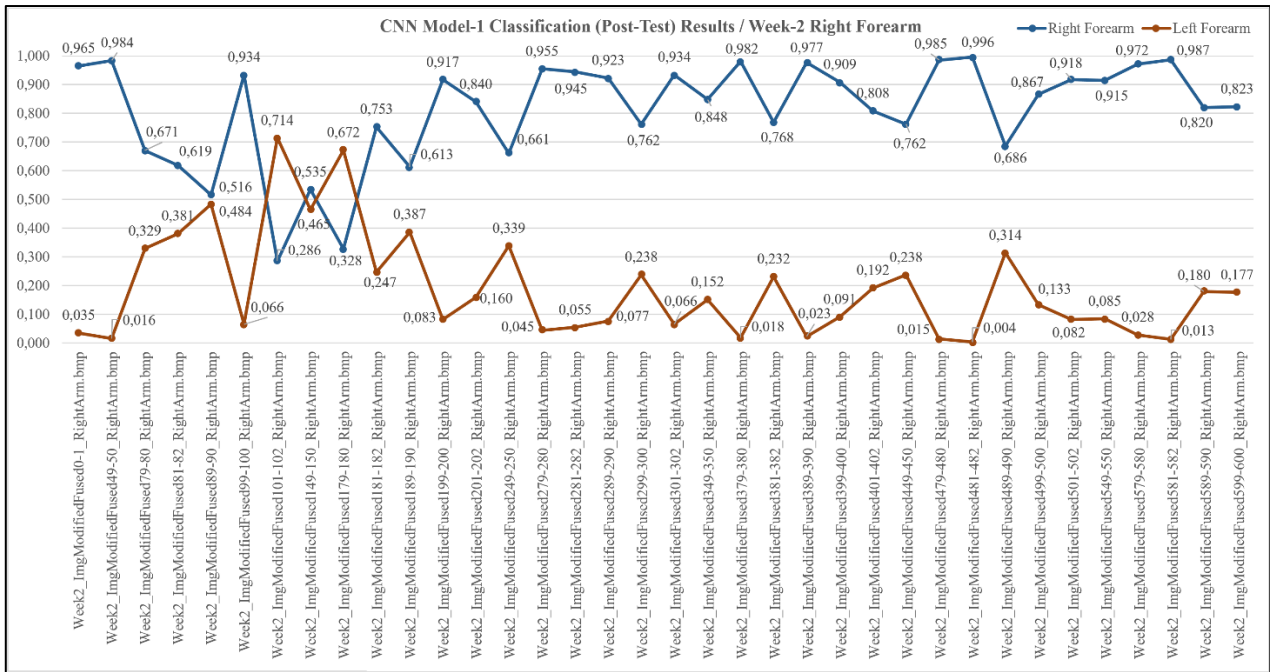


Figure 7. Classification (post-test) results obtained for CNN Model-1 (trained with Dataset-2) of 36 randomly selected right forearm images of Week-2.

After the training of the CNN models was completed, the models were post-tested. In the study, the predicted classes of the tissue images (right and left forearm) of the new weeks recorded as video by the user are determined by the post-test process. The post-test results indicate to which specified class in trainings (two classes were used as right and left forearm in the study) the tested image belongs to with a probabilistic result value between 0 and 1. For this purpose, 36 post-test images were randomly selected from the NIR fused-modified images of Week-2 and Week-3 in the study). The NIR fused-modified images of Week-2 and Week-3 were only used in the post-test process and were not included in the training and pre-test processes. In addition, 36 images of the right forearm and 36 images of the left forearm belonging to Week-1, which were not used during training, were included in the post-test trials. The results of randomly selected 36 right forearm images related with Week-2 were post-tested with Model-1 (trained with Dataset-2) as shown in Figure 7. According to the results, 34 of the 36 images (for those with a right forearm probability greater than 0.500 according to the post-test result, blue dots in Figure 7) of the right forearm were calculated more likely to be the right forearm, while the probability of 2 of them (for those with a right forearm probability less than 0.500 according to the post-test result, blue dots representing fused-images 101-102 and 179-180 in Figure 7) were calculated more likely to be the left forearm. When the misclassified fused-images 101-102 and 179-180 are examined (especially when attention is paid to fused-image 149-150, which is classified correctly with a low probability), it can be seen that the NIR illumination of the arm is not sufficient for correct classification during the video recording interval (approximately 2.67 seconds) corresponding to these three images. The 36 probabilistic results obtained in this way (representing right or left arm belonging and outnumbering) represent the predicted first-class of tissue displayed in the new week, as shown in Figure 8.

When all the results investigated, as can be seen in Figure 8, especially since Model-1 for Week-3 right forearm performed

e-ISSN: 2148-2683

misclassification in forecasting the first-class with both Dataset-1 and Dataset-2, a third dataset was prepared using data augmentation techniques. While creating Dataset-3, data augmentation techniques such as rotation (5 and 10 degrees), scaling (2 units zoom) and noise addition (salt and pepper type, 0.5 density) were used. Although misclassifications shown in Figure 8 were encountered according to the post-test class results obtained with the new dataset, the predicted first-class results were obtained correctly for all weeks.

The aim of the study, which is conducted for the use of a single person at home, is to make an accurate class estimation with less data. When looking at Dataset-1 and Dataset-2 in general, although the first-class results predicted in Model-2 were correct for all weeks, the rate of misclassification of post-test images of Model-2 (65 misclassifications in 432 post-test trials) compared to Model-1 (55 misclassifications in 432 post-test trials) is 2.315% higher, as shown in Figure 8. When looking at the misclassification cases in the model post-test class results obtained by using all three datasets, it is seen that the use of only one model or one dataset may not be sufficient alone to determine the predicted first-class for applications with low number of data. In addition, Week-3 right forearm images were classified incorrectly by Model-1 when using Dataset-1 and Dataset-2 for training, and classified correctly when using Dataset-3. When the models were trained with Dataset-3, although the predicted first-class results for all weeks were correct, the total number of incorrect post-test class results is higher than that of other datasets (for a total of 432 post-test trials, 61 misclassifications were made with Dataset-1, 59 misclassifications with Dataset-2, and 82 misclassifications with Dataset-3, respectively). In this context, the use of a hybrid decision-making algorithm that evaluates all the results of different dataset image numbers and different image sizes together in order to prevent an incorrect classification result in terms of the estimated ultimate-class has been put on the agenda. From this point of view, when Figure 8 is examined, when the model results for each week are evaluated together by using the hybrid decision-making algorithm, the predicted ultimate-classes

of the images for all weeks are correct. Confusion matrices and associated calculation values of the ultimate-class classification performances of Model-1 and Model-2 are given in Figure 9. Whereas Model-1 was found to be more successful in terms of Accuracy Rate and Specificity (True Negative Rate) calculated by using the values forming the confusion matrices, Model-2 is found to be more successful in terms of Precision and Sensitivity (True Positive Rate).

Week No.	Class Name	Dataset-1: 200 train/50 pre-test		Dataset-2: 400 train/100 pre-test		Dataset-3 "Augmented Data": 600 train/200 pre-test		Hybrid Decision Making Algorithm: Predicted Ultimate-Class
		Model-1	Model-2	Model-1	Model-2	Model-1	Model-2	
Week 1 (Right Forearm)	Right Forearm	36	36	36	36	33	34	RIGHT FOREARM
	Left Forearm	0	0	0	0	3	2	
	Predicted First-Class	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	
Week 1 (Left Forearm)	Right Forearm	0	0	0	0	0	0	LEFT FOREARM
	Left Forearm	36	36	36	36	36	36	
	Predicted First-Class	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	
Week 2 (Right Forearm)	Right Forearm	33	22	34	29	21	21	RIGHT FOREARM
	Left Forearm	3	14	2	7	15	15	
	Predicted First-Class	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	
Week 2 (Left Forearm)	Right Forearm	0	8	0	11	0	15	LEFT FOREARM
	Left Forearm	36	28	36	25	36	21	
	Predicted First-Class	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	
Week 3 (Right Forearm)	Right Forearm	11	28	11	30	26	28	RIGHT FOREARM
	Left Forearm	25	8	25	4	10	8	
	Predicted First-Class	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	Right Forearm	
Week 3 (Left Forearm)	Right Forearm	0	3	0	8	0	14	LEFT FOREARM
	Left Forearm	36	33	36	28	36	22	
	Predicted First-Class	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	Left Forearm	

Figure 8. Post-test results of classification phase.

Actual Class	Predicted Ultimate-Class n=648		Actual Class	Predicted Ultimate-Class n=648	
	Positive (Right Forearm)	Negative (Left Forearm)		Positive (Right Forearm)	Negative (Left Forearm)
Positive (Right Forearm)	TP=241	FN=83	Positive (Right Forearm)	TP=264	FN=60
Negative (Left Forearm)	FP=0	TN=324	Negative (Left Forearm)	FP=59	TN=265

(a) (b)

Performance Values Formulas			Model-1	Model-2
Accuracy Rate=	$(TP + TN) / (TP + TN + FP + FN)$		0.872	0.816
Misclassification Rate=	$(FP + FN) / (TP + TN + FP + FN)$		0.128	0.184
True Positive Rate=	$TP / (TP + FN)$		0.744	0.815
True Negative Rate=	$TN / (TN + FP)$		1.000	0.818
False Positive Rate=	$FP / (TN + FP)$		0.000	0.182
False Negative Rate=	$FN / (TP + FN)$		0.256	0.185
Precision=	$TP / (TP + FP)$		0.372	0.407
Prevalence=	$(TP + FN) / (TP + TN + FP + FN)$		0.500	0.500
F Score=	$(2 * Precision * True Positive Rate) / (Precision + True Positive Rate)$		0.496	0.543

(c)

Figure 9. Post-test trials' confusion matrix values. (a) Model-1 (28x28), (b) Model-2 (64x64), (c) Comparison of Model-1 and Model-2 performance values.

3.2. Object Detection Phase Results

There is currently no publicly available dataset of NIR images for training the YOLOv3 algorithm used for the object detection phase. Therefore, artificial narrowing patterns were added onto the Week-1 (right forearm) NIR fused-modified images having 512x512 resolution in order for the algorithm to be trained (at the same time by performing pre-testing, average errors are calculated as avgLoss). There are 100 train and 50 pre-test images in the artificial dataset consisting of NIR fused-modified images. By creating 7 narrowing patterns on each image, a total of 1050 artificial narrowing patterns are introduced to the system. The YOLOv3 training, which was carried out using the NIR fused-modified image dataset with artificial narrowing patterns for the object detection phase, was

completed in approximately eight and a half hours by utilizing the Google Colaboratory (colab.research.google.com) (which allows remote allocation of Google's GPU hardware) over the web browser. The training was continued until the average loss value obtained was less than 1. Different NIR fused-modified images (conforming to the narrowing figural format in the trainings, but with new artificial narrowing patterns added at different angles in random positions) that were not used in the trainings were used as YOLOv3 post-test data. In the trials, the locations and rates of possible narrowings that may occur on the vascular structure on these images was tried to be determined. For this, the YOLOv3 algorithm's feature of marking the location of the sought feature (narrowing detection in our study) on the image was used.

After completing the training and pre-testing parts with 150 NIR fused-modified images, which include artificial narrowings given in Figure 6, the algorithm was post-tested with different NIR fused-modified images having new artificial narrowing patterns similar to those in training. When looking at Figure 10, YOLOv3 narrowing result of 0.41 was found for the post-test process for the NIR fused-modified image created from the 76th and 77th images of Week-1, and 0.39 for the 78th and 79th NIR fused-modified images. For Week-3, narrowing was determined at the same location as in Week-1 with detection rates of 0.42 for the 110th and 111th NIR fused-modified image, and 0.45 for the 192nd and 193rd NIR fused-modified image. According to the YOLOv3 result for the Week-2 NIR fused-modified image, a narrowing detection at the same location as in the other weeks could not be obtained due to the flashing resulting from NIR illumination taking place on some parts of the image. This loss of detection experienced in Week-2 can be overcome by presenting the narrowing detections to be made in the same region of the new images to be taken in Week-3 and the following weeks, in the form of a video stream at the AR phase.

3.3. Augmented Reality Phase Results

Images containing narrowing detection rates are superimposed on NIR raw-images obtained from video and virtual (vein) - real (tissue) images are created. For this process, the image containing the YOLOv3 result of the NIR fused-modified image, as shown in Figure 10, is superimposed on the first NIR raw-image and the image to be used for the AR video infrastructure is obtained. For the second image frame of the video, the YOLOv3 result of the third and fourth NIR fused-modified images and the raw version of the third image are used. Similarly, both other images for the same week and images from other weeks are overlaid on the corresponding raw images, including the YOLOv3 results. A video file is created by adding images containing virtual vein structures and real tissue images as sequential image frames. In this way, video images in which narrowing detection rates are marked on the relevant vascular regions are presented to the user and the physician on the basis of a video-based indirect AR application. Veins are highlighted in white for better visibility in video images. The images obtained are indicated in Figure 10.

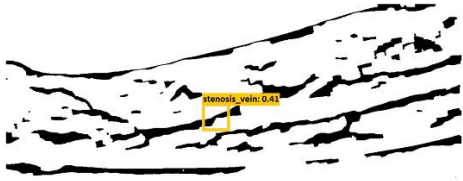

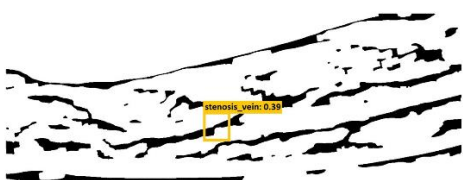

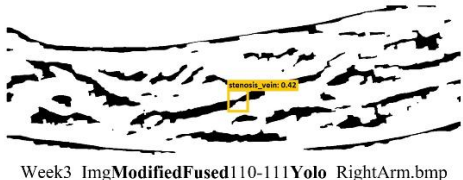

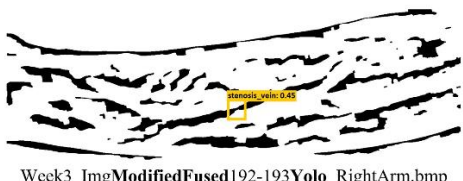

Week	NIR Fused-Modified Image Yolov3 Result	Augmented Reality Image with Narrowing Detection Rate
Week1	 Week1_ImgModifiedFused76-77Yolo_RightArm.bmp	 Week1_ImgModifiedFused76-77YoloAR76_RightArm.bmp
	 Week1_ImgModifiedFused78-79Yolo_RightArm.bmp	 Week1_ImgModifiedFused78-79YoloAR78_RightArm.bmp
Week3	 Week3_ImgModifiedFused110-111Yolo_RightArm.bmp	 Week3_ImgModifiedFused110-111YoloAR110_RightArm.bmp
	 Week3_ImgModifiedFused192-193Yolo_RightArm.bmp	 Week3_ImgModifiedFused192-193YoloAR192_RightArm.bmp

Figure 10. YOLOv3 results and AR video frames.

4. Conclusions and Recommendations

The system designed for personel use at home enables to perform superficial vein surveillance (by recording at weekly intervals, optionally) that can span a long period of time by using the NIR-based video images. In the study, the image post-processing phase was used to improve the discontinuous structures (caused by the insufficient NIR illumination) on the vein images. It has been determined that the interpretation of the results with hybrid decision-making algorithm used in the classification phase of the six phased system, can increase the ultimate classification accuracy (instead of individually evaluating the CNN models used) for a long-term personal follow-up system aimed at home use.

In the AR phase, the images containing the narrowing detection results obtained by the object detection phase were superimposed on the user's NIR raw-video images of the relevant week, and these images containing the virtual (vein) - real (tissue) structures were used to create a video-based indirect AR environment. By using these images, the user and his/her physician can see both the locations of possible narrowings and their detection rates as a follow-up system that can spread over a long period of time, thus delays in treatment will be prevented. By adding the virtual-real images obtained by superimposing the

narrowing results, especially week by week sequentially on the NIR raw-images, the locations where the narrowing is detected were able to be visualized much more clearly in the form of stream images. Thus, the possibility of pre-diagnosis can be provided regarding the rate at which the changes in a certain location progress in terms of time.

Within the scope of future studies, it is planned to structure the proposed system on different tissue regions, especially to include vascular enlargement patterns (progress surveillance of varicose veins in legs). In addition, the system will be developed in such a way that it will be possible to zoom in virtual-real video images to the veins or to examine information such as narrowing development statistics, via an interactive viewing application, preferably designed for smart devices, and will be detailed within the scope of home surveillance system on the basis of telemedicine.

5. Acknowledge

The authors would like to thank PhD Bora UZUN (Dokuz Eylül University, Department of Biomechanics, School of Medicine, İzmir, Turkey) for his contributions during the hardware phase of the study and to Işıl ERDEM (Civil Engineer, MSc) for making the final edits of the paper.

References

- Abdulghani, A. M. A., & Menekşe Dalveren, G. G. (2022). Moving object detection in video under different weather conditions using YOLO and faster R-CNN algorithms. *European Journal of Science and Technology*. (33): 40-54.
- Ai, D., Yang, J., Fan, J., Zhao, Y., Song, X., Shen, J., Shao, L., & Wang, Y. (2016). Augmented reality based real-time subcutaneous vein imaging system. *Biomedical Optics Express*. 7(7): 2565-2585.
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, (pp. 1-6). IEEE.
- Alwazzan, M. J. (2020). Low cost blood vein detection system based on near-infrared LEDs and image-processing techniques. *Polish Journal of Medical Physics and Engineering*. 26(2): 61-67.
- Anzueto-Rios, A., Hernandez-Gomez, L. E., & Hernandez-Santiago, K. A. (2016). Forearm and hand vein detection system for an infrared image database. *Res. Comput. Sci*. 127(1): 137-147.
- Chunara, R., Zhao, Y., Chen, J., Lawrence, K., Testa, P. A., Nov, O., & Mann, D. M. (2020). Telemedicine and healthcare disparities: A cohort study in a large healthcare system in New York City during COVID-19. *Journal of the American Medical Informatics Association*. 28(1): 33-41.
- Craig, J., & Petterson, V. (2005). Introduction to the practice of telemedicine. *Journal of Telemedicine and Telecare*. 11(1): 3-9.
- Crisan, S., Tarnovan, J. G., & Crisan, T.E. (2007). A low cost vein detection system using near infrared radiation. In *2007 IEEE Sensors Applications Symposium*, (pp. 1-6). IEEE.
- Demir, A. G. (2019). *Determination of vascular stenosis on angiography images using convolutional neural network method*. [Master's thesis, Başkent University, Ankara, Turkey]. "(Thesis in Turkish with an abstract in English)"
- Dikbayır, H. S., & Bülbül, H. İ. (2020). Real-time vehicle detection by using deep learning methods. *Tünav Bilim Dergisi*. 13(3): 1-14. "(Article in Turkish with an abstract in English)"
- Doğan, D., Erol, T., & Mendi, A. F. (2021). Sağlık alanında karma gerçeklik. *Avrupa Bilim ve Teknoloji Dergisi*. (29): 11-18. "(Article in Turkish with an abstract in English)"
- Elnasir, S., & Shamsuddin, S. M. (2014). Palm vein recognition based on 2D-discrete wavelet transform and linear discrimination analysis. *Int. J. Advance Soft Compu. Appl*. 6(3): 43-59.
- Erdem, H. A., Erdem, I., & Utku, S. (2020). Near-infrared mobile imaging systems for e-health: Lighting the veins, *The Twelfth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED)*, (pp. 80-84). IARIA.
- Ferrari, M., Mottola, L., & Quaresima, V. (2004). Principles, techniques, and limitations of near infrared spectroscopy. *Canadian Journal of Applied Physiology*. 29(4): 463-487.
- Francis, M., Jose, A., Devadhas, G. G., & Avinash, K. K. (2017). A novel technique for forearm blood vein detection and enhancement. *Biomedical Research*. 28(7): 2913-2919.
- Francisco, M. D., Chen, W. F., Pan, C. T., Lin, M. C., Wen, Z. H., Liao, C. F., & Shiue, Y. L. (2021). Competitive real-time near infrared (NIR) vein finder imaging device to improve peripheral subcutaneous vein selection in venipuncture for clinical laboratory testing. *Micromachines*. 12(4): 373.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 580-587). IEEE.
- Huda, A. N., Goh, C. M., Lim, C. H., Aluwee, S. S., Bajuri, M. N., & Wahab, N. H. A. (2021). Development of a near-infrared (NIR) forearm subcutaneous vein extraction using deep residual U-Net. *International Conference on Biomedical Engineering (ICoBE)*.
- LeCun, Y., Cortes, C., & Burges, C. J. C. (n.d.). *The MNIST Database of handwritten digits*. Retrieved July 02, 2022, from <http://yann.lecun.com/exdb/mnist/>
- Mangold, K., Shaw, J. A., & Vollmer, M. (2013). The physics of near-infrared photography. *European Journal of Physics*. 34(6): 51-57.
- MatConvNet Team (2017). *MatConvNet: CNNs for MATLAB*. Retrieved July 02, 2022, from <https://www.vlfeat.org/matconvnet/>
- MathWorks, Inc. (1996). MATLAB (R2017a): The language of technical computing, computation, visualization, programming, installation guide for UNIX version 5. Natick, Massachusetts.
- Meng, G. C., Shahzad, A., Saad, N. M., Malik, A. S., & Meriaudeau, F. (2015). Prototype design for wearable veins localization system using near infrared imaging technique. In *2015 IEEE 11th International Colloquium on Signal Processing and Its Applications (CSPA)*, (pp. 112-115). IEEE.
- Mistry, D., & Banerjee, A. (2017). Comparison of feature detection and matching approaches: SIFT and SURF. *GRD Journals Global Research and Development Journal for Engineering*. 2(4): 7-13.
- Park, K. R. (2011). Finger vein recognition by combining global and local features based on SVM. *Computing and Informatics*. 30(2): 295-309.
- Rao, H., Zhang, P., & Sun C. (2017). Contrast enhancement for the infrared vein image of leg based on the optical angular spectrum theory. *Signal, Image and Video Processing*. 11(3): 423-429.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 779-788). IEEE.
- Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. Arxiv preprint.
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39(6): 1137-1149.
- Shrotri, A., Rethrekar, S. C., Patil, M. H., & Kore, S. N. (2010). IR-webcam imaging and vascular pattern analysis towards hand vein authentication. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, (Vol. 5, pp. 876-880). IEEE.
- Sordillo, L. A., Pu, Y., Pratavieira, S., Budansky, Y., & Alfano, R. R. (2014). Deep optical imaging of tissue using the second and third near-infrared spectral windows. *Journal of Biomedical Optics*. 19(5): 056004.

- Şayli, Ö., Akdemir, A., Ataklı, Y., Emir, U. E., Çıtlak, P. Ö., Cengiz, L. S., & Akın, A. (2004). Kaslardaki oksidatif metabolizma farkının işlevsel yakın-kızılötesi spektroskopisi ile incelenmesi. *BIYOMUT*. İstanbul. “(Article in Turkish with an abstract in English)”
- Şeker, K., & Engin, M. (2017). Deep tissue near-infrared imaging for vascular network analysis. *Journal of Innovative Optical Health Sciences*. 10(3): 1650051.1-12.
- Tan, F. G., Yüksel, A. S., Aydemir, E., & Ersoy, M. (2021). Derin öğrenme teknikleri ile nesne tespiti ve takibi üzerine bir inceleme. *Avrupa Bilim ve Teknoloji Dergisi*. (25): 159-171. “(Article in Turkish with an abstract in English)”
- Tien, T. V., Mien, P. T., Dung, P. T., & Linh, H. Q. (2015). Using near-infrared technique for vein imaging. In *5th International Conference on Biomedical Engineering in Vietnam*, (pp. 190-193). Springer, Cham.
- Vedaldi, A., & Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM International Conference on Multimedia*, (pp. 689-692).
- Wadhvani, M., Sharma, A. D., Pillai, A., Pisal, N., & Bhowmick, M. (2015). Vein detection system using infrared light. *Int. J. Sci. Eng. Res.* 6(12): 780-786.
- Wang, L., & Leedham, G. (2006). Near-and far-infrared imaging for vein pattern biometrics. In *2006 IEEE International Conference on Video and Signal Based Surveillance*, (pp. 52-57). IEEE.
- Witther, J., Tsai, Y. T., & Azuma, R. (2011). Indirect augmented reality. *Computers & Graphics*. 35(4): 810-822.
- Wu, W., Elliott, S. J., Lin, S., & Yuan, W. (2019). Low-cost biometric recognition system based on NIR palm vein image. *IET Biometrics*. 8(3): 206-214.
- Wu, Y., Zhang, Z., & Wang, G. (2021). Training deep neural networks via branch-and-bound. ArXiv Preprint. arXiv:2104.01730.
- Yılmaz, K. (2014). *The design of a new portable ophthalmoscope and utilization of the device for diagnosis and identification*. [Doctoral thesis, Ege University, Izmir, Turkey]. “(Thesis in Turkish with an abstract in English)”