

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**INTEGRATING DETECTION OF VASCULAR
DEGENERATION INTO AUGMENTED REALITY
ENVIRONMENT: AN E-HEALTH APPLICATION
BASED ON NEAR-INFRARED SPECTROSCOPY
AND DEEP LEARNING**

by
Hüseyin Aşkın ERDEM

September, 2022
İZMİR

INTEGRATING DETECTION OF VASCULAR DEGENERATION INTO AUGMENTED REALITY ENVIRONMENT: AN E-HEALTH APPLICATION BASED ON NEAR-INFRARED SPECTROSCOPY AND DEEP LEARNING

**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Computer Engineering, Computer Engineering Program**

**by
Hüseyin Aşkın ERDEM**

**September, 2022
İZMİR**

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**INTEGRATING DETECTION OF VASCULAR DEGENERATION INTO AUGMENTED REALITY ENVIRONMENT: AN E-HEALTH APPLICATION BASED ON NEAR-INFRARED SPECTROSCOPY AND DEEP LEARNING**” completed by **HÜSEYİN AŞKIN ERDEM** under supervision of **ASSOC. PROF. DR. SEMİH UTKU** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

.....
Assoc. Prof. Dr. Semih UTKU

Supervisor

.....
Prof. Dr. Aylin KANTARCI

Thesis Committee Member

.....
Prof. Dr. Efendi NASİBOĞLU

Thesis Committee Member

.....
Assoc. Prof. Dr. Tuğba ÖZACAR ÖZTÜRK

Examining Committee Member

.....
Assoc. Prof. Dr. Övünç ÖZTÜRK

Examining Committee Member

.....
Prof. Dr. Okan FISTIKOĞLU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my sincere respect and gratitude to my advisor Assoc. Prof. Dr. Semih UTKU, whose trust I have always felt, for his insightful help, guidance and patience at all the stages of this thesis and for being an exemplary to me.

Also, I wish to express my deepest thanks to my thesis monitoring committee members Prof. Dr. Aylin KANTARCI and Prof. Dr. Efendi NASİBOĞLU for guiding me through all the stages of my thesis with their experience, support and kindness.

And special thanks to Res. Assist. Dr. Bora UZUN, who lit the first spark of this thesis with the idea of near-infrared imaging and received his unwavering support afterwards.

I am extremely grateful to Col. Murat KILIÇ and Assit. Prof. Dr. Col. Ersin CİVAN, for creating every kind of opportunity with all their indulgence, so that I could complete the work on my thesis.

Finally, my dearest father, mother and sister Civil Engineer (M.Sc) Işıl ERDEM, like throughout my whole life, once again you were there for supporting, motivating and above all believing in me... Endless thanks for you all.

Hüseyin Aşkın ERDEM



In loving memory of
Assist. Prof. Dr. Şen ÇAKIR

INTEGRATING DETECTION OF VASCULAR DEGENERATION INTO AUGMENTED REALITY ENVIRONMENT: AN E-HEALTH APPLICATION BASED ON NEAR-INFRARED SPECTROSCOPY AND DEEP LEARNING

ABSTRACT

Developing technology makes our lives easier in every area imaginable. Today, the integration of images obtained with medical imaging techniques, especially in the deep learning infrastructure, mediates the saving of countless lives within the scope of early diagnosis and follow-up. In this context, the star of e-health applications that especially appeal to home use is shining rapidly. The images obtained from the user at periodic intervals in the Imaging Technique Phase of this thesis, are enhanced on the server with the Digital Image Pre-Processing Phase. Visual deficiencies caused by visualization are corrected to a certain extent in the Digital Image Post-Processing Phase. If the Classification Phase confirms that the same tissue region is being viewed, vascular degeneration (narrowing/enlargement) in the relevant images is detected in the Object Detection Phase. Videos showing the degeneration detection results and their locations are prepared by the Augmented Reality Phase and shared with the user and the physician. In this way, possible degeneration can be detected at an early stage and its developmental stages can be followed. According to the experimental results, the most effective object detection was obtained for varicose_vein class with Accuracy Rate (1.0000), Misclassification Rate (0.0000) and Precision (1.0000).

In general, within the scope of this study, molecules that can be revealed by using near-infrared radiation that cannot be detected by the eye, computer systems that can create their own intelligence with deep learning, and augmented reality environments that can offer time-independent visualization have been examined.

Keywords: Near-infrared radiation spectroscopy, vascular degeneration, deep learning, convolutional neural networks, YOLOv3, indirect augmented reality.

DAMARSAL DEJENERASYON TESPİTİNİN ARTIRILMIŞ GERÇEKLİK ORTAMINA ENTEGRASYONU: YAKIN-KIZILÖTESİ SPEKTROSKOPİ VE DERİN ÖĞRENME TEMELLİ BİR E-SAĞLIK UYGULAMASI

ÖZ

Gelişen teknoloji akla gelebilecek her alanda hayatımızı kolaylaştırmaktadır. Bugün, medikal görüntüleme teknikleriyle elde edilen görüntülerin özellikle derin öğrenme altyapısına entegrasyonu, erken teşhis ve takip kapsamında sayısız hayatın kurtarılmasına aracılık etmektedir. Bu kapsamda özellikle ev kullanımına hitap eden e-sağlık uygulamalarının yıldızı ise hızla parlamaktadır. Bu tez çalışmasının Görüntüleme Tekniği Aşamasında kullanıcıdan periyodik aralıklarla elde edilen görüntüler, sunucuda Dijital Görüntü Ön-İşleme Aşaması ile iyileştirilmektedir. Görüntülemeden kaynaklı görsel eksiklikler, Dijital Görüntü Son-İşleme Aşamasında belirli bir düzeyde düzeltilmektedir. Sınıflandırma Aşaması aynı doku bölgesinin görüntülediği doğrularsa, ilgili görüntülerdeki damarsal dejenerasyonlar (damar daralması/damar genişlemesi) Nesne Tespiti Aşamasında tespit edilmektedir. Dejenerasyon tespit sonuçlarıyla konumlarını gösteren videolar, Artırılmış Gerçeklik Aşamasıyla hazırlanarak kullanıcı ve hekim ile paylaşılmaktadır. Bu sayede, olası dejenerasyonlar erken aşamada tespit edilebilmekte ve gelişim aşamalarının takibi gerçekleştirilebilmektedir. Deneysel sonuçlara göre Accuracy Rate (1.0000), Misclassification Rate (0.0000) ve Precision (1.0000) olarak en etkili nesne tespiti varicose_vein sınıfı için elde edilmiştir.

Genel olarak bu çalışma kapsamında, gözle algılanamayan yakın-kızılötesi radyasyon kullanılarak açığa çıkarılabilen moleküller, derin öğrenme ile kendi zekâsını yaratabilen bilgisayar sistemleri ve zamandan-bağımsız görsellik sunabilen artırılmış gerçeklik ortamları incelenmiştir.

Anahtar kelimeler: Yakın-kızılötesi radyasyon spektroskopisi, damarsal dejenerasyonlar, derin öğrenme, evrişimli sinir ağları, YOLOv3, dolaylı artırılmış gerçeklik.

CONTENTS

	Page
Ph.D. THESIS EXAMINATION RESULT FORM	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	v
ÖZ	vi
LIST OF FIGURES	x
LIST OF TABLES	xiii
ABBREVIATIONS	xv
 CHAPTER 1 - INTRODUCTION	 1
1.1 Overview	1
1.2 Problem Definition (Motivation)	4
1.3 Aim of the Thesis	6
1.4 Importance of the Thesis	7
1.5 Literature Overview	9
1.6 Organization of the Thesis	10
 CHAPTER 2 - METHODS AND LITERATURE REVIEW	 12
2.1 The Hardware Block	13
2.1.1 The Imaging Technique Phase: Obtaining Near-Infrared Images	14
2.1.1.1 Electromagnetic Spectrum and Near-Infrared Radiation	15
2.1.1.2 Spectroscopy	17
2.1.1.3 Near-Infrared Radiation Spectroscopy	18
2.1.1.4 Literature Review of Imaging Technique	26
2.1.1.5 The Proposed System's Imaging Technique Phase	33
2.2 The Software Block	36
2.2.1 The Digital Image Pre-Processing Phase: Enhancing Near-Infrared Images	 36
2.2.1.1 Digital Image Processing	37

2.2.1.2 Literature Review of Digital Image Processing	40
2.2.1.3 The Proposed System's Digital Image Pre-Processing Phase	47
2.2.2 The Digital Image Post-Processing Phase: Finding the Missing Pieces	53
2.2.2.1 Digital Image Matching	54
2.2.2.2 The SURF Algorithm	56
2.2.2.3 Literature Review of Digital Image Matching	58
2.2.2.4 The Proposed System's Digital Image Post-Processing Phase	61
2.2.3 The Classification Phase: Recognizing Near-Infrared Images	65
2.2.3.1 Deep Learning	65
2.2.3.2 Convolutional Neural Networks	70
2.2.3.3 Literature Review of Classification	82
2.2.3.4 The Proposed System's Classification Phase	94
2.2.4 The Object Detection Phase: Catching Vascular Degeneration	100
2.2.4.1 Object Detection from Digital Images	102
2.2.4.2 The YOLO Algorithm and Versions	105
2.2.4.3 Literature Review of Object Detection	107
2.2.4.4 The Proposed System's Object Detection Phase	110
2.2.5 The Augmented Reality Phase: Winding the Time Fast	116
2.2.5.1 Mixed Reality	117
2.2.5.2 Indirect Augmented Reality	119
2.2.5.3 Literature Review of Indirect Augmented Reality	125
2.2.5.4 The Proposed System's Augmented Reality Phase	129
CHAPTER 3 - EXPERIMENTAL RESULTS	132
3.1 Experimental Results of the Classification Phase	132
3.2 Experimental Results of the Object Detection Phase	140
CHAPTER 4 - CONCLUSIONS	151
REFERENCES	155

APPENDICES 174

Appendix A - General Overview of the Proposed System's Six Phases in 10-minutes.....	174
Appendix B - The Classification Phase Post-Test Results for Period-1	176
Appendix C - The Classification Phase Post-Test Results for Period-2	178
Appendix D - The Classification Phase Post-Test Results for Period-3	180
Appendix E - Quick Overview of the Literature	181
Appendix F - Publications Published within the Scope of the Thesis	183
Appendix G - Glossary	184



LIST OF FIGURES

	Page
Figure 2.1 General diagram of the client-server network architecture model of the proposed system	12
Figure 2.2 The proposed system's six phases in 1-minute	14
Figure 2.3 Electromagnetic spectrum	16
Figure 2.4 Absorbance graph of the NIR optical window range of the electromagnetic spectrum.....	21
Figure 2.5 Camera and LED placement techniques used in NIR imaging	22
Figure 2.6 The proposed system's Imaging Technique Phase in 1-minute.....	34
Figure 2.7 The proposed system's Digital Image Pre-Processing Phase in 1-minute	53
Figure 2.8 The proposed system's Digital Image Post-Processing Phase in 1-minute	64
Figure 2.9 Artificial intelligence, machine learning and deep learning	66
Figure 2.10 Single-layer perceptron	67
Figure 2.11 Stages of Machine Learning and Deep Learning	68
Figure 2.12 Architectural layers of a classical convolutional neural network model	71
Figure 2.13 The proposed system's Classification Phase in 1-minute	100
Figure 2.14 The proposed system's Object Detection Phase in 1-minute.....	116
Figure 2.15 Milgram's terms of the reality-virtual continuum	119
Figure 2.16 The proposed system's Augmented Reality Phase in 1-minute.....	131
Figure 3.1 20 filters and sizes of the 20 depth first layers of CNN Models using Dataset-1. (a) 5x5 size filter used in CNN Model-1, (b) 7x7 size filter used in CNN Model-2, and (c) 11x11 size filter used in CNN Model-3	133
Figure 3.2 Objective function and toplerror result graphs of CNN Models using Dataset-1. (a) CNN Model-1 training graph completed at 14,000 epochs, (b) CNN Model-2 training graph completed at 4,000 epochs, and (c) CNN Model-3 training graph completed at 1,000 epochs	134

- Figure 3.3 Stenosis_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period1_Img8-9_LeftArm.jpg is used as a representative). (a) Before post-test, and (b) After post-test..... 141
- Figure 3.4 Spider_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period1_Img2-3_LeftArm.jpg is used as representative). (a) Before post-test, and (b) After post-test..... 144
- Figure 3.5 Varicose_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period3_Img2-3_LeftArm.jpg is used as a representative). (a) Before post-test, and (b) After post-test..... 146



LIST OF TABLES

	Page
Table 1.1 The scope of the literature examined in the thesis phases.....	10
Table 2.1 Six phases and general descriptions of the proposed system	13
Table 2.2 Hardware criteria and explanations used in the Imaging Technique Phase of the proposed system.....	33
Table 2.3 Steps and explanations of the proposed system's Imaging Technique Phase.....	34
Table 2.4 Digital image processing steps and explanations.....	38
Table 2.5 Steps and explanations of the proposed system's Digital Image Pre- Processing Phase	49
Table 2.6 Feature detection and matching algorithms and explanations	55
Table 2.7 SURF algorithm steps and explanations	57
Table 2.8 Steps and explanations of the proposed system's Digital Image Post- Processing Phase	62
Table 2.9 Architectural layers and explanations of a classical convolutional neural network model.....	72
Table 2.10 Architectural features of popular convolutional neural network models	75
Table 2.11 Popular datasets and features	76
Table 2.12 Data augmentation techniques and explanations	77
Table 2.13 Architectural layers of a classical convolutional neural network model and their effects on performance	80
Table 2.14 Training hyperparameters and performance effects of a classical convolutional neural network model.....	81
Table 2.15 Steps and explanations of the proposed system's Classification Phase..	96
Table 2.16 Data augmentation techniques and explanations of the proposed system's Classification Phase.....	99
Table 2.17 Object detection algorithms and explanations	103
Table 2.18 YOLO algorithm versions and explanations.....	105
Table 2.19 Steps and explanations of the proposed system's Object Detection Phase	111

Table 2.20	Data augmentation techniques and explanations of the proposed system's Object Detection Phase	115
Table 2.21	Steps and explanations of the proposed system's Augmented Reality Phase.....	130
Table 3.1	Post-test results for three separate periods according to the datasets and CNN Models of the proposed system's Classification Phase.....	136
Table 3.2	The complexity matrix and performance evaluation parameters of the post-test results of the proposed system's Classification Phase	139
Table 3.3	YOLOv3 confidence values determined as a result of post-test for stenosis_vein class artificial patterns (artificial pattern sizes are shown proportionally).....	142
Table 3.4	YOLOv3 confidence values determined as a result of post-test for spider_vein class artificial patterns (artificial pattern sizes are shown proportionally).....	145
Table 3.5	YOLOv3 confidence values determined as a result of post-test for varicose_vein class artificial patterns (artificial pattern sizes are shown proportionally).....	147
Table 3.6	The complexity matrix and performance evaluation parameters of the post-test results of proposed system's Object Detection Phase	149

ABBREVIATIONS

1D	: One-Dimensional
2D	: Two-Dimensional
3D	: Three-Dimensional
ANN	: Artificial Neural Network
AR	: Augmented Reality
CCD	: Charge Coupled Device Image Sensor
CLAHE	: Contrast Limited Adaptive Histogram Equalization
CMOS	: Complementary Metal Oxide Semiconductor Image Sensor
CNN	: Convolutional Neural Network
COVID-19	: Coronavirus Disease 2019
CPU	: Central Processing Unit
CT	: Computed Tomography
DIP	: Digital Image Processing
e-Health	: Electronic Health
FIR	: Far-Infrared Radiation
GPU	: Graphics Processing Unit
IAR	: Indirect Augmented Reality
IR	: Infrared Radiation
LED	: Light Emitting Diode
MLP	: Multi-Layer Perceptron
MNIST	: Modified National Institute of Standards and Technology
MRI	: Magnetic Resonance Imaging
NIR	: Near-Infrared Radiation
RAM	: Random Access Memory
RE	: Real Environment
ReLU	: Rectified Linear Unit
RGB	: Red-Green-Blue
ROI	: Region of Interest
SIFT	: Scale Invariant Feature Transform
SLP	: Single-Layer Perceptron

SURF	: Speeded Up Robust Features Local Feature Detector Function
USB	: Universal Serial Bus
VE	: Virtual Environment
YOLOv3	: You Only Look Once (version-3)



CHAPTER ONE

INTRODUCTION

In the First Chapter, general information about the thesis topic is given. In the Overview subheading, the importance of the concept of Electronic Health (e-health) during the Coronavirus Disease 2019 (Covid-19) pandemic was emphasized. In the Problem Definition subheading, medical imaging devices used in the scope of early diagnosis and vascular degeneration are discussed. The purpose of the system developed within the scope of the thesis is explained in the subheading of the Purpose of the Thesis in terms of creating early diagnosis data and a follow-up system. In the subheading of Importance of the Thesis, the technical contributions of the thesis to the fields of computer science, medicine and imaging techniques are examined. In the Literature Overview subheading, the literature researched within the scope of the six phases that make up the thesis study is summarized, and in the subheading of the Thesis Chapters, the chapters described in the thesis are indicated.

1.1 Overview

The internet, which is the most popular communication method today, has been needed more than ever, especially during the Covid-19 pandemic. Most people were able to maintain their communication with the world via the internet during this period when they were forced or voluntarily quarantined at home. In the pandemic, students attended classes from their locations with distance education, employees of companies located in different countries were able to maintain their common working environment by remote meetings during this period when international travel was restricted. In the health sector, the follow-up and treatment of patients without Covid-19 could be continued from their homes within the scope of e-health, without the need to go to hospitals, which pose a high risk of transmission, especially during the pandemic process.

E-Health (the use of communication and information technologies in the health sector (Singhal & Cowie, 2022)) is a method that was in use even before the

pandemic process. In this method, the patient and the doctor can communicate through communication tools from their locations without the need to go to the hospital. The patient, who is under follow-up, can keep in constant contact with his/her doctor by performing simple medical measurements (blood pressure measurement or beats per minute) about himself/herself at home or by reporting changes (mild non-routine complaints such as weakness or loss of balance) in his/her general condition within a certain period of time. In particular, e-Health offers highly effective methods in terms of time and cost in the follow-up of patients who have difficulty leaving their homes.

Since vascular imaging (containing sufficient discriminating data for classification) offers a non-invasive, non-contact and user-friendly method, it is preferred for health (hygienic and comfortable use) or safety reasons (vein patterns are individual) (Elnasir & Shamsuddin, 2014; Sierro, Ferrez, & Roduit, 2015).

Since bruises on the skin in some vascular diseases (varicose, deep vein thrombosis) reduce the visibility of the veins, it is very critical to visualize the veins first in order to initiate the treatment (Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015). Near-Infrared Radiation (NIR) spectroscopy can be used in the diagnosis of varicose veins (enlarged and twisted veins in the legs) (Rao, Zhang, & Sun, 2017). Commercially available NIR imaging devices, on the other hand, are designed for specific medical purposes only (vascular access, oxygen level detection etc.) (Sanchez-Morago et al., 2010).

Developing a vascular imaging system requires an interdisciplinary (electronics, computer science, optics, embedded systems) approach (Sierro, Ferrez, & Roduit, 2015). Adding classification and object recognition features to such a system (artificial intelligence, deep learning), and even designing a multifunctional system by integrating the system into Virtual Environments (VEs) is possible with a much wider workspace and information combination.

Basically, an NIR imaging system can be constructed with a light source that emits Infrared Radiation (IR) light of a certain wavelength and a filter (additionally, visible light blocking filter can be preferred for a clearer image) capable of transmitting this wavelength-sensitive IR light to the camera lens (Mahmood, Ahmed, & Shahid, 2016). Photons absorbed by deoxy-hemoglobin in veins viewed with NIR make the veins appear darker in the image. In order to obtain the vein patterns more clearly, Digital Image Processing (DIP) steps are applied on the NIR image. The processed images are stored in the dataset to represent a certain class and used for comparing with new images within the scope of classification operations. (Shrotri, Rethrekar, Patil, & Kore, 2010)

In this study, superficial vascular degeneration (narrowing/enlargement) that may occur in any tissue (hand, arm, leg, etc.) of the body can be detected by a low-cost NIR camera. Images were taken on the basis of NIR spectroscopy by NIR camera having 850 nanometer (nm) wavelength (λ) that can be easily integrated into the smart device with Universal Serial Bus (USB) support. Vein patterns were obtained by clarifying the veins in the obtained NIR images with image processing filters. Missing/cut structures caused by illumination in vein patterns are repaired at a certain level with the Speeded Up Robust Features (SURF) Local Feature Detector Function. Obtained patterns are classified by Convolutional Neural Networks (CNN). Artificially created degeneration (narrowing/enlargement) on vein patterns are detected with the object detection algorithm You Only Look Once version-3 (YOLOv3) (Redmon & Farhadi, 2018). The obtained results were presented to the user and his/her physician as Indirect Augmented Reality (IAR) environment. In this way, it is ensured that vascular degeneration follow-up processes can be carried out within the scope of e-health.

In this thesis, a superficial vein imaging and tracking system that can be used within the scope of e-Health has been developed. With the mobile application infrastructure of the developed system, the user can view his/her veins in the comfort of his/her home and upload them to the server as a video. The detection results of superficial vascular degeneration (narrowing/enlargement) obtained from the videos

processed on the server are shared with the patient and his/her doctor. In this way, degeneration can be detected at an early stage and its developmental stages can be followed.

1.2 Problem Definition (Motivation)

Modern medicine, with the support of technological developments, is now able to diagnose most diseases in the early stages. Devices with very different imaging techniques are currently used to detect diseases or monitor their developmental stages.

Traditional medical imaging systems (X-Ray, Computed Tomography (CT) Scan, MRI, electron/optical microscopy), most of which work with harmful rays to the body, unsuitable for home use include technologies that enable the visualization of structures (bone in the body, vascular network, internal organs) invisible to the naked eye (the human eye can see the wavelength range of 400-700 nm of the electromagnetic spectrum) for biomedical purposes by using different methods (photons, sound waves or magnetic fields of different wavelengths of the electromagnetic spectrum). In this way, very detailed measurements can be made to help the doctor in diagnosis and treatment. (Mahmood, Ahmed, & Shahid, 2016)

A wide variety of tissues, from bone structures to internal organs, can be viewed with these devices, which utilize light photons, magnetic fields or sound waves in different regions of the electromagnetic spectrum. In this way, most diseases can be detected in the early stages and treated easily. However, in most cases, routine controls are disrupted for various reasons (lack of awareness, use of harmful rays in imaging, fear of disease, indifference, etc.) and the opportunity for early diagnosis can be missed.

For example, some common eye diseases can be easily detected by correct interpretation of the retinal image. Inaccurate interpretation of optic disc abnormalities in these images due to factors such as tiredness or incomplete

understanding of the physician may lead to miss the chance of early diagnosis and to encounter a wide variety of dramatic results from visual disturbances to vision loss due to delay in treatment. (Şengür et al., 2018)

Systems that can combine the experience of physicians with remote patient data can pay the way for early diagnosis opportunities (Demir, 2019).

Conventional devices used for early diagnosis emit harmful rays to human health, interpretation of images obtained with some of them requires expertise and the costs of most of them are very high. However, advances in imaging techniques and DIP methods offer newer imaging options for different tissues.

Since intravenous access (also called venipuncture or infusion, it refers to direct and rapid injection of serum or liquid, etc. (Huang et al., 2021) into the vascular access, or opening a needle tip in the vein to take blood from the vein), which is routinely used in the medical context, is a painful/invasive method, therefore painless vein imaging devices are needed to visualize the veins with low visibility, both in a short time and less number of venipuncture attempts (which can cause swelling, bruising, nerve damage or blood clots in the affected area) without causing pain to the patient. (Mahmood, Ahmed, & Shahid, 2016; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015)

In particular, devices (NIR imaging devices such as AccuVein, VeinViewer, Vasculuminator currently in use (Mahmood, Ahmed, & Shahid, 2016)) that provide imaging with NIR light are extremely effective in visualizing superficial veins that are less visible to the naked eye due to various factors (ethnicity/dark skin, excess adipose tissue, burn scars, thin veins, skin damage (Mahmood, Ahmed, & Shahid, 2016)). These systems are used by nurses, phlebotomist and anesthesiologists in hospitals and clinics (Ai et al., 2016). On the other hand, these devices, which are commercially available and frequently used in hospitals, especially in vascular access, also have hand-held sizes. But their cost is quite high for personal use (Juric

& Zalik, 2014; Mahmood, Ahmed, & Shahid, 2016). Therefore, these devices are only suitable for use in hospitals and clinics and are not suitable for home use.

First of all, imaging devices with harmless rays as being portable and cost-effective are three of the most important criteria that will enable them to be used anywhere, anytime, especially within the scope of e-health. However, the measurements (in terms of processing power) and imaging techniques (in terms of providing a harmless imaging environment to the body) that these systems can offer for home use are limited.

In this study, the primary reason for choosing NIR light is that it provides a basis for home use as a type of light that is harmless to the human body. In addition, a simple webcam can be easily converted into a NIR camera. This is the second factor that lays the groundwork for home use by allowing NIR systems to be developed at low cost. On the other hand, mobility can be supported by the integration of the system into smart devices.

In this study, vascular imaging is provided on the basis of NIR spectroscopy, which is not harmful to the human body, the vascular patterns in the images are obtained, the patterns are classified, the vascular degeneration in the classified patterns are detected and the degeneration developmental stages can be followed over a wide period of time.

1.3 Aim of the Thesis

In this thesis, an imaging system suitable for mobile use has been created for superficial veins. The system focuses on the early diagnosis and follow-up of the degeneration (narrowing/enlargement) that may occur on these veins by using the vein images taken with the NIR imaging technique. These images are improved with DIP techniques and the vascular degeneration detected in the obtained vascular patterns are reported to the user and physician via IAR environment. Thanks to this system, which provides e-health service, the users will be able to detect and follow

the degeneration that may occur in their veins in their environment without the need to go to the hospital. With the data obtained, the user will be directed to the right physician without losing time, and early initiation of treatment will be supported effectively. By starting the treatment of the patient with the correct diagnosis as soon as possible, it will be ensured that possible pecuniary losses and intangible damages can be prevented by using the time well in critical situations.

The main purpose of this system is to create a decision support system in the field of medicine by presenting early diagnosis data on superficial vascular degeneration to the physician, so as to prevent delay in treatment. Among the other aims of the study:

- Creating a fast and effective **early diagnosis method** with a harmless, painless, non-contact and easy-to-use imaging technique,
- Providing **home monitoring** for individual use with a low-cost system design (especially effective in pandemic times),
- Presentation of images covering a long time interval as week-month-year as **IAR environment** (virtual-real video image),
- Performing a **long-term tracking system** for different veins such as superficial arm, leg, upper hand, palm or neck veins (which may be spread over months, years for the individual).

1.4 Importance of the Thesis

The technical contributions of this thesis to the fields of computer science, medicine and imaging techniques can be summarized with the following items:

- *Contribution to Dataset Preparation Time.* The dataset preparation time, which contains a large number of images required for training to be performed in classifying the tissue region (CNN Models) and providing object detection (YOLOv3 algorithm), has been reduced by the use of video (instead of a few images)
- ✓ Since the system was designed to be able to recognize different tissues of a single user, the dataset preparation time is reduced (a 35-second video

contains 1050 images, therefore the data acquisition time was reduced by video recording) in tissue region and degeneration detection by recording many small differences (by obtaining similar images with different light angles in the form of video recording, problems such as camera shaking and tissue movement during viewing of personal images by the person himself/herself can be eliminated to a certain extent) from the same person with video use.

- ✓ For the detection of degeneration, 2250 artificial pattern samples were obtained with a 150-image dataset created by adding and labelling 5 artificial degeneration patterns belonging to 3-classes on a single image.
- *Contribution to Classification Performance.* The effects of image sizes (training times and results for 28x28, 64x64, 128x128), hyperparameters (epoch), different image types (NIR-raw, NIR-fused) on model performance in classifying tissue region (convolutional neural network models) were investigated.
- ✓ According to the results obtained, it has been determined that a Hybrid Decision-Making Algorithm (this algorithm was first introduced in the study of Erdem & Utku (2022b)) that evaluates different results together (by evaluating 3 different CNN Models) instead of sticking to only a single model (determined by image size) can increase the accuracy of class predictions made with a dataset containing fewer images.
- ✓ CNN Models with fewer layers were used to shorten the training times.
- ✓ The final-class was decided using three different image datasets (of different sizes) and three different CNN Models.
- *Contribution to Increasing Vascular Integrity in the Image.* In cases where a complete vascular structure could not be obtained due to the illumination of the viewed area with different angles of NIR light, the integrity of the vascular structures was increased to a certain level by image fusion (SURF algorithm).
- *Development of Multi-Purpose Detection and Tracking System.* From a medical point of view, multi-purpose use has been achieved by integrating both vasoconstriction and vasodilation detection (it is possible to obtain information about the rate of progress of the process by detecting the veins in the spider

vein/talengiectasia structure, which is the initial level, and the varicose veins, which are the advanced form.) and monitoring into a single system.

- *Providing Indirect Augmented Reality Environment.* IAR environments have been introduced to the literature for outdoor use. The importance of using it in the thesis work is that it can ensure that virtual materials and real images overlap with zero error (in offline registration). With this thesis, the IAR method was integrated into a vein imaging system for the first time.
- *Different Imaging-Platforms Support.* When evaluated in terms of medical imaging systems, the system uses superficial vein images within the scope of e-health application. Instead, it will be possible to detect degeneration by converting the images in deeper veins obtained from different imaging platforms (X-Ray Imaging, Magnetic Resonance Imaging (MRI), Ultrasound etc.) to black and white format and giving them as ready input to the system (only excluding the Imaging Technique Phase of the proposed system).

1.5 Literature Overview

In the literature studies examined within the scope of this thesis, it has been seen that tissue imaging (vein, eye, ear, palm) with NIR light is mostly performed for the purpose of obtaining biometric data. In addition, there are studies in which the images obtained are presented to the user via their mobile device within the scope of e-health application. However, a study in which the detection of vascular degeneration was presented in the form of an IAR environment as early diagnosis data was not found.

The scope of the literature examined in this thesis study has been discussed in six-phases as shown in Table 1.1, in connection with the thesis phases.

Table 1.1 The scope of the literature examined in the thesis phases

Thesis Phase	Literature Coverage
Imaging Technique Phase	In terms of hardware, spectroscopy used within the scope of imaging technique, light types, wavelengths, number of Light Emitting Diodes (LEDs) and camera features.
Digital Image Pre-Processing Phase	In terms of software, filters and methods used within the scope of DIP and providing digital image enhancement.
Digital Image Post-Processing Phase	In terms of software, algorithms used in DIP that detect and match similar points in images.
Classification Phase	In terms of software, CNN Models used in classification and based on deep learning.
Object Detection Phase	In terms of software, object detection algorithms based on deep learning that are used in object detection and determine the class and position of objects in images.
Augmented Reality Phase	In terms of software, applications that create a VE by integrating video-based virtual content used within the scope of IAR into real images.

1.6 Organization of the Thesis

This thesis consists of four main chapters. In the first (introduction) chapter, the thesis study is discussed in general within the scope of the definition of the problem, the purpose of the thesis and the importance of the thesis. The second (methods and literature review) chapter is divided into two main blocks as hardware and software. In the hardware block, the Imaging Technique Phase has been studied specifically for electromagnetic spectrum and NIR spectroscopy. In the software block, in the Digital Image Pre-Processing Phase filters and algorithms used in image enhancement are explained, and in the Digital Image Post-processing Phase, the SURF Algorithm is discussed within the scope of image matching and merging. In the Classification Phase, the contribution of CNN Model architectures, datasets and hyperparameters to classification accuracy is explained. How to detect vascular degeneration was explained in the Object Detection Phase, and how the results could be visualized was detailed in the Augmented Reality Phase. In the third (experimental results) chapter, the experimental results obtained are interpreted in two sub-headings as Classification Phase results using the Hybrid Decision-Making Algorithm and Object

Detection Phase results. In the fourth (results) chapter, a general evaluation of the study was made and it was stated in which aspects it could shed light on future studies.



CHAPTER TWO

METHODS AND LITERATURE REVIEW

In the Second Chapter, the methods used in the proposed system and the related literature review are discussed under two main blocks as “The Hardware Block” and “The Software Block”. The hardware block consists of a single phase covering to the imaging technique. In the software block, processing, matching and classification of images and then presenting the object detection results to the user in the form of a VE are explained.

With the system proposed in this thesis, it is aimed to realize the early diagnosis and follow-up of degeneration that may occur in superficial veins within the scope of e-health (as a low-cost system for home use). In this context, a system with a client-server network architecture model, the general scheme of which is given in Figure 2.1, has been developed in the study.

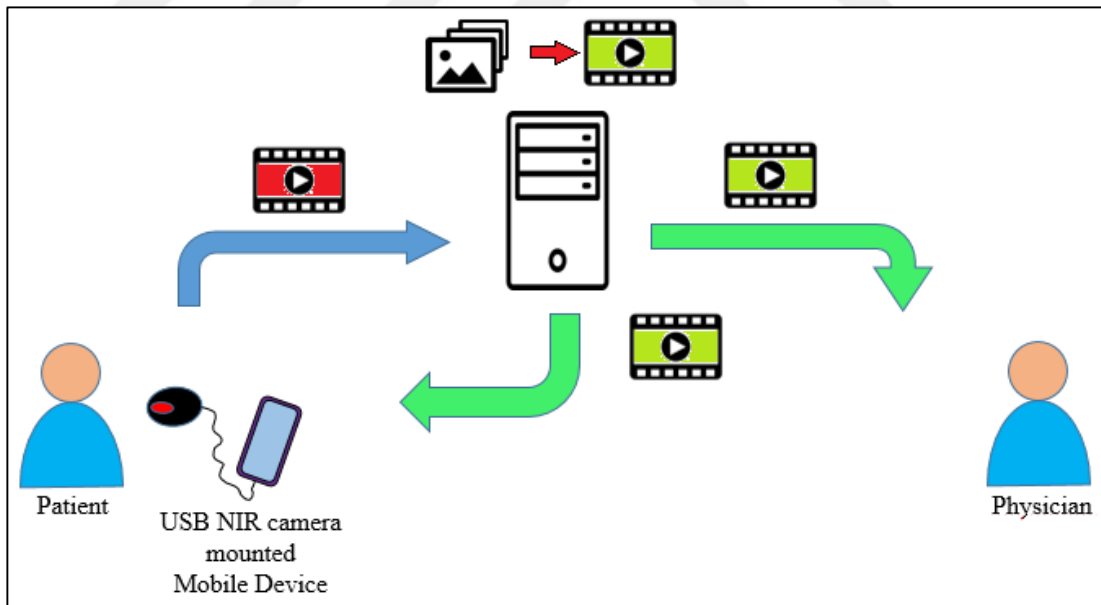


Figure 2.1 General diagram of the client-server network architecture model of the proposed system

The six phases that make up the architectural model of the proposed system as shown in Figure 2.1 and their general explanations are given in Table 2.1.

Table 2.1 Six phases and general descriptions of the proposed system

Thesis Phase	Explanation
The Imaging Technique Phase	In the home environment, the user takes a NIR video recording from the tissue region he/she wants to view or track with the NIRVeinTrackerApp application that he/she installs on his/her smart device connected to the NIR camera and uploads it to the server.
The Digital Image Pre-Processing Phase	The NIR video recording uploaded to the server is first converted to the NIR raw-images , by visually clarifying the vein structures in the images with DIP techniques, and the NIR modified-images are obtained.
The Digital Image Post-Processing Phase	The missing/discontinuous parts of the vein structures in the NIR modified-images due to the insufficient illumination are repaired to a certain extent by image matching and merging method, then the NIR fused-modified images are created.
The Classification Phase	Since the purpose of the system is to follow the same tissue region for a certain period (week, month, or year), the NIR fused-modified images are subjected to a classification process that considers the tissue region displayed in the first imaging period as the main class.
The Object Detection Phase	As a result of the classification, if it is confirmed that the same tissue area with the previous period is displayed in the new images, the degeneration (narrowing/widening) that may occur in the veins in these images is detected by the object detection algorithm and marked on the image, then the NIR yolo-fused-modified images are obtained.
The Augmented Reality Phase	As the last operation performed on the server, the NIR ar-yolo-fused-modified images created by superimposing the marked NIR fused-modified images on the NIR raw-images are converted into video format and sent to the user and the physician by using the NIRVeinTrackerApp application.

Summary of the proposed system's six phases in 1-minute is given in Figure 2.2. Also, general overview of the proposed system's six phases in 10-minute is given in Appendix-A.

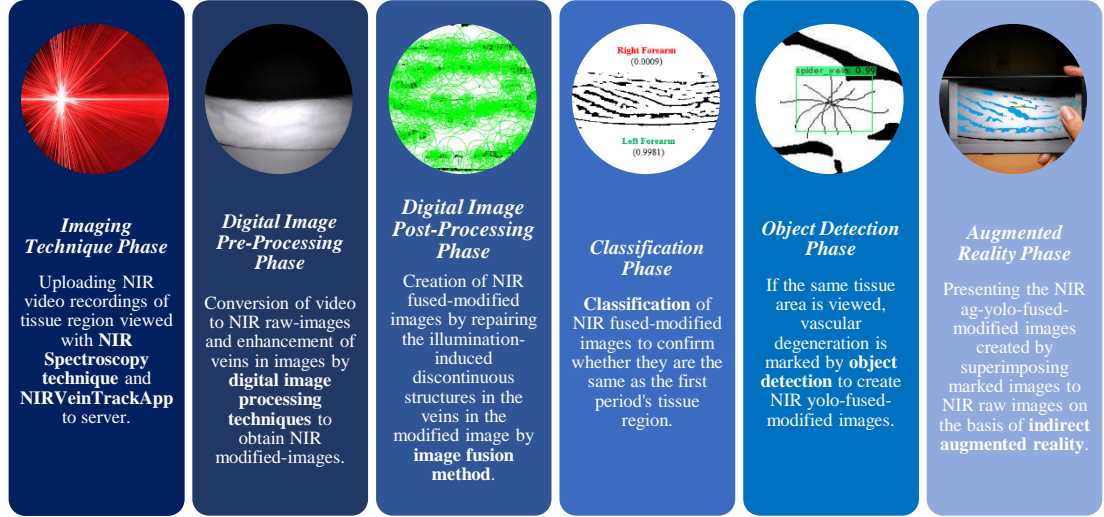


Figure 2.2 The proposed system's six phases in 1-minute

2.1 The Hardware Block

The Hardware Block consists of a single phase called “Imaging Technique Phase: Obtaining Near-Infrared Images”. The Imaging Technique Phase focuses on the NIR spectroscopy technique used to obtain vein images from the user.

2.1.1 The Imaging Technique Phase: Obtaining Near-Infrared Images

Today, there are a wide variety of imaging devices and DIP techniques in healthcare that offer physicians advantages in terms of time and cost in diagnosis and treatment (Osmanoğlu, 2016). These imaging techniques, which are frequently used in medical applications, enable the visualization of different tissues in the human body. Some of the imaging techniques perform imaging with light rays of different wavelengths (X-Ray Imaging, CT), some with magnetic waves (MRI), and some with sound waves (ultrasound). For example, X-Rays in the electromagnetic spectrum are used for imaging bone structure, MRI is used for imaging soft tissues with the magnetic field it creates, by using Hydrogen atoms in the tissues (Osmanoğlu, 2016), and Ultrasound is used for imaging deep soft tissues and deep veins (Lamperti & Pittiruti, 2013) by generating sound waves. Although some imaging techniques provide visualization without any surgical intervention, in most cases (especially in X-Ray and tomography imaging), the patient may be exposed to

extremely dangerous radiation. Unlike devices such as X-ray, harmless to the body, MRI (not recommended for use in pregnant women, prosthetic users or small children) is not a device suitable for home use in terms of size and cost (Osmanoğlu, 2016). Since ultrasound-based imaging systems prevent patient exposure to radiation, they can only be used in hospitals by pre-trained professionals, and the resulting images allow only a qualified physician to make sense of them in most cases (Mahmood, Ahmed, & Shahid, 2016; Şeker & Engin, 2017). What is important is how an imaging based on "do no harm", which is accepted as the first rule of medicine, can be provided within the scope of e-health application.

In the Imaging Technique Phase, visualization of the tissue region with superficial veins via the mobile application using the NIR spectroscopy technique and transferring the obtained NIR video recordings to the server are explained.

2.1.1.1 Electromagnetic Spectrum and Near-Infrared Radiation

Light (electromagnetic radiation) carries energy in the form of photon particles with discrete energy called quantum. In Figure 2.3, the diagram (electromagnetic spectrum) formed by the light showing wave properties with different frequencies and wavelengths is shown according to the sub-areas of the light. The spectrum includes X-ray and Gamma-rays with high radiation energy (short wavelength and high frequency), and microwave and radio waves with low radiation energy (long wavelength and low frequency). The visible region called white light (380-770 nanometers (Mangold, Shaw, & Vollmer, 2013; Royal Society of Chemistry, 2022), 400-700 nm (Lingyu & Leedham, 2006; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015)) is the only region that can be detected with the naked eye. (Royal Society of Chemistry, 2022)

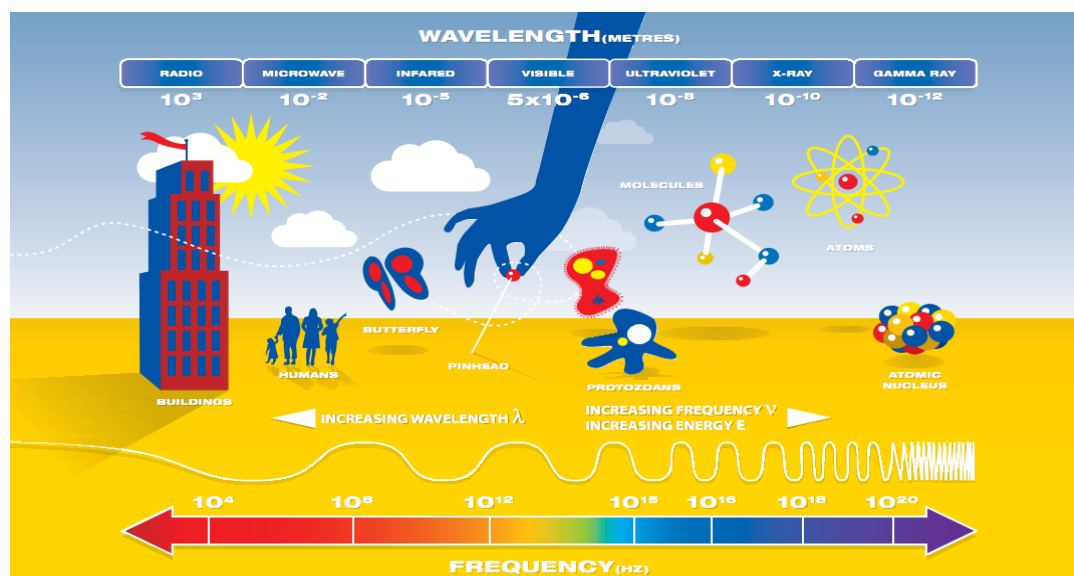


Figure 2.3 Electromagnetic spectrum (Royal Society of Chemistry, 2022)

IR light of which depth of penetration from the skin surface is greater than that of visible light (Francis, Jose, Devadhas, & Avinashe, 2017), is divided into three spectral regions (near, mid and far) useful in different applications where no definite agreement on the initial and final limit values has been reached (Mahmood, Ahmed, & Shahid, 2016; Xueyan & Shuxu 2008). The wavelength range of the spectrum (650-900 nm (Rao, Zhang, & Sun, 2017), 700-900 nm (Elnasir & Shamsuddin, 2014; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015), 760-870 nm (Crisan, 2017), 700-1400 nm (Xueyan & Shuxu, 2008)), that is close to the visible region (380-760 nm wavelength range (Yılmaz, 2014)) that the human eye can see is called the NIR light region. Imaging technique using NIR light is especially preferred in superficial (compared to visible light, it can infiltrate 1-3 mm (Mangold, Shaw, & Vollmer, 2013; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015), 3 mm (Cömert, 2015; Lingyu & Leedham, 2006; Sierro, Ferrez, & Roduit, 2015; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015), 3-5 mm (Fernández & Armada, 2017), 5 mm (Mahmood, Ahmed, & Shahid, 2016) or 5-8 mm (Chiao et al., 2013; Lamperti & Pittiruti, 2013) below the skin without deviating after the contact point in the human body) vein imaging studies. In health institutions, the NIR imaging technique is mostly used to visualize the superficial veins in the tissue to be accessed intravenously.

The use of IR light as a contactless, harmless, and non-invasive method (injection of an agent into the vein is not needed) in vascular imaging is performed with two different techniques as Far-Infrared Radiation (FIR) and NIR. FIR imaging (thermography) works on the basis that the superficial veins have higher temperature than the surrounding tissues. Based on the fact that most heated substances emit IR radiation, IR imaging passively relies on detecting only IR radiation emitted from the body (since the superficial veins have higher temperature than the surrounding tissues, NIR imaging can be provided for the veins) (Lingyu & Leedham, 2006). However, since it is thermal based (based on body temperature), it is extremely sensitive to ambient conditions (humidity and temperature). The temperature profile of the tissue viewed with the FIR camera is obtained. Since NIR imaging is not a thermal method, body temperature or the temperatures of surrounding tissues do not affect the imaging. In this method, it is based on the diffusion of light in the NIR Optical Window range into the blood and tissue in different ways. In NIR imaging, vein detection is performed on the basis of spectroscopy, according to the response of hemoglobin molecule in the blood to light. The veins appear darker when NIR light is detected by a Charge Coupled Device Image Sensor (CCD) camera with an IR-pass filter. Generally, FIR is used to visualize large veins on the hand, while NIR is more effective in imaging small veins (palm, wrist). However, since FIR images have low contrast, it is difficult to separate the veins from the background, while for NIR, defects on the tissue surface (hair, scars, skin folds) pose a disadvantage in imaging. (Lingyu & Leedham, 2006; Seymen, Boyraz, Çetin, & Yıldız, 2014; Şeker & Engin, 2017; Wadhvani, Sharma, Pillai, Pisal, & Bhowmick, 2015; Xueyan & Shuxu, 2008)

In this thesis study, vein images were obtained with NIR imaging, which provides better contrast between the vein and surrounding tissues (Sierro, Ferrez, & Roduit, 2015), in order to better capture the details in the superficial veins and to create a system independent of ambient conditions.

2.1.1.2 Spectroscopy

The interaction of light and matter is studied within the scope of spectroscopy. When light energy hits the material, studies that involve examining the reaction of the material to this light stimulus and obtaining information about the molecules that make up the matter (the structure and properties of the matter) are of interest to spectroscopy. (Royal Society of Chemistry, 2022)

IR spectroscopy is a method based on obtaining information about the substance (physical and chemical properties), independent of the state of matter (liquid, powder or solution) according to the level of absorption (in connection with the vibration of atoms in the molecule) when illuminated with IR light (780-2500 nm wavelength) (Klakegg, Luo, Goncalves, Hosio, & Kostakos, 2016). Imaging devices that require intravenous (i.v.) access to the patient for the visualization of veins are considered invasive (Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015). This technique, on the other hand, is a non-invasive and contact-less technique as it provides imaging without the need for intravenous agent administration (Lingyu & Leedham, 2006).

Since veins are inside the body, it is difficult in most cases for vascular structures to be seen under visible light. For this reason, different imaging techniques that provide imaging with X-ray, ultrasound and IR can be used to observe these structures. However, although radiological methods such as X-Ray provide high quality vascular imaging (sufficient contrast can be provided in imaging (Şeker & Engin, 2017)), they require the injection of special agents into the blood stream. Therefore, these methods, which are classified as invasive techniques and used for medical treatment, are not suitable for general-purpose vascular imaging applications where non-invasive imaging is required. In addition, the superficial vessels remain too shallow for ultrasound detection. (Rao, Zhang, & Sun, 2017). (Lingyu & Leedham, 2006; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015; Xueyan & Shuxu, 2008) The methods that can be used in vein or tissue imaging are either invasive or effective in observing deeper veins. For this, NIR spectroscopy, which is

not very new, but has been gaining in popularity recently especially in biometric applications, is used in non-invasive imaging of superficial veins.

2.1.1.3 Near-Infrared Radiation Spectroscopy

NIR spectroscopy is a method for non-invasive imaging of superficial veins, which uses the ability of the hemoglobin molecule (according to oxygen transport or not) to absorb NIR light of different wavelengths more than the surrounding tissues (Boezeman, Moll, Ünlü, & Vries, 2016; Lingyu & Leedham, 2006).

NIR spectroscopy (in other words NIR imaging technique) has been in use in medicine for over 20 years (Lamperti & Pittiruti, 2013). For example, Phlebology (the branch of medicine that deals with the vein) makes use of NIR imaging, especially in the treatment of veins (venepuncture or varicose vein) that cannot be seen with the naked eye and ultrasound (due to shallowness) (Rao, Zhang, & Sun, 2017). In vascular access, which can be a time-consuming and painful procedure for the patient, especially in hospitals, vascular access is extremely effective both in making the veins with low visibility (depending on factors such as dark skin or excess fat layer) visible in the relevant location and in providing sterile imaging without the need for patient contact. (Lamperti & Pittiruti, 2013; Sanchez-Morago et al., 2010).

VeinViewer (the first vein imaging device using NIR) and AccuVein (portable), which are among the commercially available NIR imaging devices, reflect the vascular structure back on the patient's skin, thus they make the processes of vein finding and vascular access (it can also detect varicose veins) easier. Small-sized or portable versions of these devices (VeinViewer Vision and VeinViewer Flex) are already in use. Also, plastic surgeons and dermatologists prefer the guidance of devices such as AccuVein in order to reduce complications like bruising after neurotoxin (Botox, Dysport) and filler injection. (Highton & Ekwobi 2011; Lamperti & Pittiruti, 2013; Wells, 2014)

NIR spectroscopy is also used within the scope of functional-NIR spectroscopy (the vein-based brain function imaging method (Quaresima & Ferrari, 2016; Quaresima & Ferrari, 2019) that non-invasively measures the oxygen changes in hemoglobin via multiple detectors placed on the scalp) to non-invasively examine brain activities in neuroscience researches (Scholkmann et al., 2014; Wahr, Tremper, Samra, & Delpy, 1996).

Different tissues absorb light which cannot be seen with the naked eye differently. While the veins are nearly invisible under normal white light, they are visible under NIR light. In biology, a specific range of the electromagnetic spectrum (700-900 nm (Elnasir & Shamsuddin, 2014; Lingyu & Leedham, 2006; Mahmood, Ahmed, & Shahid, 2016; Rao, Zhang, & Sun, 2017), 700-1000 nm (Sierro, Ferrez, & Roduit, 2015), 740-1100 nm (Xueyan & Shuxu, 2008), 760-1100 nm (Cömert, 2015)) is known as medical spectral window. In this region, the vascular patterns under the skin can be visualized with imaging techniques that use NIR light. The hemoglobin molecule in the blood in the arteries, veins and capillaries plays an important role in this visualization, providing oxygen transport to the tissues. Hemoglobin molecule in human blood combines with oxygen in the lungs and carries oxygen from the heart to the tissues in the structure of Oxy-hemoglobin (HbO_2), after giving its oxygen to the tissues, it returns to the heart and then to the lungs as Deoxy-hemoglobin (Hb) containing carbondioxide and continues its transport process (Mahmood, Ahmed, & Shahid, 2016). As seen in Figure 2.4, the NIR light in the NIR optical window range is absorbed by the oxygenated Hemoglobin (HbO_2) in the arteries and the deoxygenated Hemoglobin (Hb) in the veins at certain wavelengths, according to certain absorption coefficients. In addition, Hemoglobin molecule absorbs the light from NIR more than the surrounding tissues. (Crisan, 2017; Lingyu & Leedham, 2006; Rao, Zhang, & Sun, 2017; Wadhwani, Sharma, Pillai, Pisal, & Bhowmick, 2015; Xueyan & Shuxu, 2008)

Especially since the superficial veins in the hands, wrists, arms and legs are located just under the skin (1-3 mm deep), they can be easily visualized with NIR spectroscopy (Mangold, Shaw, & Vollmer, 2013; Wadhwani, Sharma, Pillai, Pisal, &

Bhowmick, 2015). Since the NIR absorption feature of hemoglobin occurs outside of the visible region, vein visualization can be achieved independently of skin colour in NIR imaging (Lingyu & Leedham, 2006; Zhong, Kanhere, & Chou, 2017).

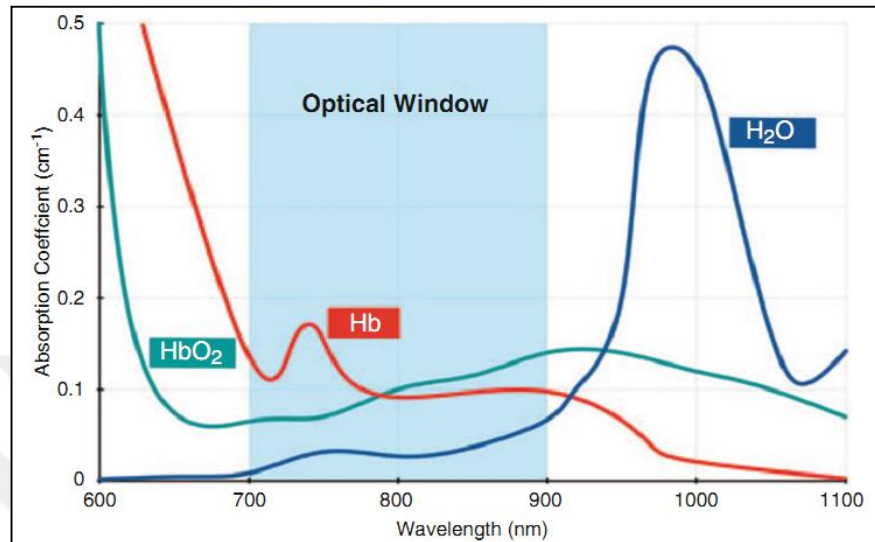


Figure 2.4 Absorbance graph of the NIR optical window range of the electromagnetic spectrum (Crisan, 2017)

Contrary to what is seen in Figure 2.4, it has been reported that photons are better absorbed and the contrast in the image reaches the most optimal level when vein imaging is performed with NIR light with wavelengths of 850 to 900 nm in most of the studies conducted in the literature. Since the veins absorb more NIR light at 850 nm wavelength (especially at 850 nm, the NIR light can penetrate up to 3.57 mm in the palm region) compared to the other wavelengths, they appear much darker than the surrounding tissues (Ai et al., 2016; Cömert, 2015; Elnasir & Shamsuddin, 2014). In addition, the noise effect created by the heat emitted by the human body at the wavelength of 850 nm and the ambient heat emitted by the environment (in other words, IR radiation at the wavelength of 3-14 μm) can be reduced. (Elnasir & Shamsuddin, 2014; Lingyu & Leedham, 2006; Mangold, Shaw, & Vollmer, 2013; Wadhvani, Sharma, Pillai, Pisal, & Bhowmick, 2015)

The placement of the camera and light source to be used when designing the non-invasive NIR system is extremely critical. Basically, in the system, the image of the

tissue region illuminated by the NIR light is captured with a CCD or Complementary Metal Oxide Semiconductor Image Sensor (CMOS) camera that is photosensitive (using NIR-permeable, visible light-blocking filters (Shrotri, Rethrekar, Patil, & Kore, 2010)). Here, both the NIR camera and the LEDs to be used as the NIR light source must support the same NIR wavelength. The NIR image is formed by the photons that come to the camera lens by reflection or transmission from the illuminated tissue region of the NIR light as in Figure 2.5. **In the visualization technique of NIR with Reflection;** the tissue is illuminated by scattering the light photons from the tissue placed under/over the camera and the light source. It is used to visualize the veins of the hand dorsum, palm and wrist. It is a more preferred method as it emits less radiation and provides a balanced distribution in contrast, but skin folds in the palm can interfere with the image (Cömert, 2015). **In the NIR imaging technique with Transmission;** the light photons of the NIR are passed through the tissue placed between the light source and the camera to illuminate the tissue. It is preferred on thinner limbs such as fingers. The rays coming from the source pass through the tissue and reach the camera lens on the other side. Although it contains a little more radiation, bone structures can cause darkening on the image (Cömert, 2015). The image of the tissue region illuminated by NIR LEDs can be obtained with CCD or CMOS camera with a NIR-pass filter. Two-dimensional (2D) single/double arrays, rectangular or circular arrays are used in the placement of NIR LEDs according to the lens, and it is stated that the most effective light distribution is obtained with a circular layout centered on the camera lens (Shrotri, Rethrekar, Patil, & Kore, 2010). (Rao, Zhang, & Sun, 2017; Sierro, Ferrez, & Roduit, 2015)

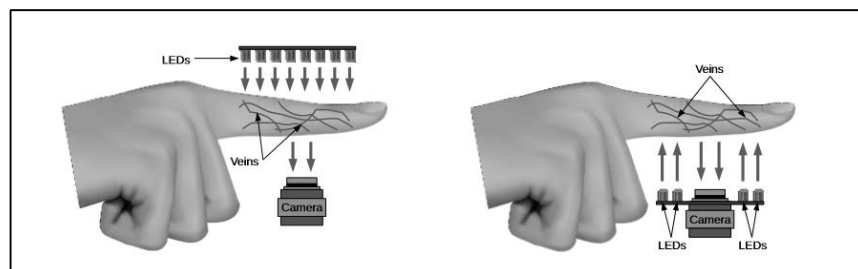


Figure 2.5 Camera and LED placement techniques used in NIR imaging (Sierro, Ferrez, & Roduit, 2015)

Since the tissue is thick, NIR images to be taken over the arm (antecubital fossa (Juric & Zalik, 2014)), leg or trunk are obtained with reflection NIR imaging technique instead of transmission (useful for slender limbs like fingers). In this thesis, since it is aimed to detect and follow-up the vascular degeneration in the arm and leg veins, the reflection NIR imaging technique is preferred.

With a portable NIR Spectroscopy system which can be integrated into smart phones, advanced analysis environments that will minimize the user's involvement can be created. In this way, the user will start scanning only for the object to be measured from his/her smart device, without the need to know the system infrastructure, and all the remaining operations (NIR spectroscopy, analysis, machine learning or interpretation of results, etc.) will be performed on the server and the results will be presented to the user in near real-time via the smart device. With such a system that will be developed for ordinary devices on the basis of NIR Spectroscopy, a NIR measurement system that can be used by expert personnel will be available to anyone who can use a mobile device. (Blakegg, Luo, Goncalves, Hosio, & Kostakos, 2016)

It is extremely easy to develop a low-cost NIR camera. In order to convert an ordinary webcam capable of detecting visible light (with a lens filter that passes photons in the visible region of the light and blocks those in the IR region, to detect a better quality image) into an IR camera in three simple steps: (1) The lens on the CCD or CMOS sensor is removed, (2) In order to prevent IR light from deteriorating the image, the IR-cut filter placed on the lens is removed and replaced with an IR-pass filter (although unexposed negative film can be used as the simplest IR-pass filter, it can cause blurring of the image as it does not completely block visible light (Cömert, 2015)), (3) For an IR camera that can also capture a higher quality IR image by blocking visible light, the lens is coated with several layers of 35mm film negative exposed to light (Lingyu & Leedham, 2006; Sawant, 2022; Shrotri, Rethrekar, Patil, & Kore, 2010). However, although the camera modification process can be performed in three simple steps, the slightest mistake that may be encountered during the modification can easily render the camera useless. In addition, for the IR

camera to be a NIR camera sensitive to only a certain wavelength, the IR-pass filter and LEDs must also have the same wavelength.

The target device of this thesis is ordinary users' mobile devices with ordinary camera (smartphone, tablet, etc.). However, the conversion of the cameras of these devices into a NIR camera by outside intervention is not a risk that every user can take. Therefore, in this thesis, a low-cost (\$65) commercial night vision camera having an 850 nm wavelength IR-pass filter and 850 nm LEDs and supporting USB-connection was utilized.

NIR images are basically acquired in 6-steps: (1) The tissue area to be imaged is illuminated with a NIR light source within the scope of NIR spectroscopy, (2) While some of the photons from the LEDs leak from the skin and the fat layer underneath (since the absorption coefficient of the skin and muscle tissue is low (Fernández & Armada, 2017)), reaching the hemoglobin molecules with high absorption coefficient in the veins, some are reflected back from skin surface, (3) NIR photons reaching the hemoglobin are absorbed, (4) Photons returning to the camera lens form the NIR image, (5) In the NIR image, the light-coloured parts containing more photons represent the surrounding tissues, and the darker parts (the result of greater absorption of the NIR photons by Hemoglobin) represent the blood (hemoglobin) inside the veins. (6) The obtained images are developed by image processing methods and vein patterns are obtained. (Lingyu & Leedham, 2006; Rao, Zhang, & Sun, 2017; Sierro, Ferrez, & Roduit, 2015; Şeker & Engin, 2017; Wadhvani, Sharma, Pillai, Pisal, & Bhowmick, 2015)

NIR spectroscopy is harmless to the body, provides portability (offers the advantage that the laboratory can be brought near the sample), provides fast and immediate results (instead of packaging the sample and sending it to the laboratory, measurement within seconds in the current environment) and can work with a small number of samples. In this context, applications that can open new horizons in the health, agriculture, textile, industrial production, food and pharmaceutical sectors can be developed with a portable NIR spectroscopy device that works integrated with a

mobile device. In this way, even a simple smartphone can be turned into a device that can make advanced analytical measurements for daily use by making use of different measurement and detection capabilities such as location, temperature, ambient light, and camera. (Klakegg, Luo, Goncalves, Hosio, & Kostakos, 2016)

When generally looked at some of the questions still waiting to be answered about NIR imaging within the scope of the thesis: (1) *How should the training of people who will use these systems be?* (Lamperti & Pittiruti, 2013) The developed system's usage in this thesis was designed to be extremely simple for an ordinary user. Users of all ages who are familiar with smart device usage will be able to easily view their veins and upload the images to the server with the NIRVeinTrackerApp mobile application, of which interface images are given in the Mobile Application Step of Table 2.3. In addition, vein patterns and degeneration detection results processed on the server will be shared with the user and the physician. Since the results are placed on the raw images of the relevant tissue region in a very understandable way, they can be interpreted without the need for additional training. In what technical aspects is the system more useful? (Lamperti & Pittiruti, 2013) The system offers a non-invasive imaging. In addition, it is effective in observing the progression of the disease in terms of detecting vascular degeneration and performing a long-term follow-up. (3) Will users be able to afford the system cost? (Lamperti & Pittiruti, 2013) Since the target audience is ordinary users and the target usage area is home, the system is designed to have the least cost as possible. It will be enough for the user to have a NIR camera that supports USB-connection and the free NIRVeinTrackerApp application. The physician, on the other hand, will be able to access the results on either his/her mobile device or personal computer. However, since the target audience is ordinary users and expert experience cannot be provided in imaging, special attention should be paid to the distance of the camera and the light source to the tissue (white flashes may occur at the center point if held too close), the lighting angle (photons from different angles can form discontinuous structures in veins), the cleanness of the tissue surface (hairs or scars can affect the accuracy of the vein patterns), the noise or different types of light (visible light negatively affects the quality of the NIR image), and the use of different

backgrounds (requires additional actions to be removed) (Klakegg, Luo, Goncalves, Hosio, & Kostakos, 2016) besides, depending on the situation, it is necessary to provide orientation through the application or to use image processing techniques (since the veins appear dark and the rest of the tissues appear light in the NIR raw-images, the colour distinction between the veins and the surrounding tissue may be more or less noticeable, but it may be difficult for non-experts to distinguish. For this reason, image enhancement is provided by DIP techniques for more distinct separation of veins. (Cömert, 2015; Francis, Jose, Devadhas, & Avinashe, 2017)) (4) *Are NIR imaging systems harmful to health?* The most important feature that paves the way for the use of NIR spectroscopy in the home environment is that it is approved by the U.S. Food & Drug Administration for non-invasive and continuous monitoring of oxygen status in body and cerebral tissues. (Boezeman, Moll, Ünlü, & Vries, 2016).

2.1.1.4 Literature Review of the Imaging Techniques

Most of the studies on harmless and non-invasive imaging techniques in the literature use the NIR spectroscopy method.

In the studies of Chiao et al. (2013), the relationship between patient characteristics (noting that ethnicity-skin colour, age and obesity may cause fewer veins to be detected) and finding vascular sites suitable for intravenous access was investigated. In this context, superficial veins were visualized using both the traditional method (naked eye) and the Veinsite (is VueTek's IR vein finding device and is the first study in which the device was used) method. Header-shaped Veinsite device consists of a portable NIR light source (700-900 nm), video recording and LCD display device. According to the results of the observation in the upper extremity veins (as the forearm and upper surface of the hand are the most commonly used body parts for vascular access and they offer practicality for taking image samples) performed on 384 people (768 observations made with the traditional method and device), it was determined that the number of veins that could be detected for intravascular access in patients with reduced vein visibility due to the

characteristics was higher when the device was used compared to the traditional method. In addition, it was underlined that the quality of clinical care could be improved, complications could be prevented, and treatment could be started without delay by non-invasively finding of suitable veins for intravenous access quickly.

Rao, Zhang, and Sun (2017) tried to improve the contrast with the filter in order to increase the image quality by eliminating the noise effect created by the hairs on the leg tissues in the NIR image. In this context, NIR LEDs (850 ± 30 nm) and CCD camera were used.

In the work of Şeker and Engin (2017), it was aimed to perform vascular network analysis (complexity of veins in the Region of Interest (ROI) or vein diameter/width measurements) from NIR images of deep tissues. In this context, an in-vitro environment was prepared. Liquid phantom (liquid-based) to be used as human blood and solid phantom (pork gelatin-based) to be used as skin tissue were created. Blood flow was imitated by passing the liquid phantom through plastic tubes (using as artificial vein) of 1-3 mm diameter placed inside the solid phantom at a depth of 3.5 mm from the surface. 17 pieces of NIR LEDs (875-940 nm) were used by arranging around the lens in both disc and ring shape. A single tube was used in experiments with in-vitro medium. In-vivo trials were performed with wrist and arm images of a female volunteer. The mean error (0.3) obtained from the experiments carried out with the 875 nm wavelength in the studies was lower than that obtained with the 940 nm wavelength (0.5). According to the results of the study, no difference was found in terms of spot size and irradiation uniformity in disc and ring-shaped placements. The proposed imaging system, with non-ionizing radiation, it was advantageous in terms of cost-effectiveness and in providing algorithmic simplicity in analysis.

In the work of Fernández and Armada (2017), a multi-sensor system was proposed for the detection of superficial veins that would be required within the scope of a system that can perform catheter placement robotically. In this context, a wavelength sensitive short-wave IR (SWIR) camera with a range of 900-1700 nm, a

time-of-flight (TOF) camera, and 96 NIR LEDs (at 940 nm wavelength and in two linear arrays positioned to the right and left of the SWIR camera) were used. While superficial vein images with 636x508 resolution were obtained with the SWIR camera, position information was obtained from the 176x144 resolution images with the TOF camera. According to the trials, it was stated that the system was found to be applicable. In particular, the suitability of the system for clinical use was ensured by its ability to support a real-time application with a response time, its affordable cost, and its small size and compactness.

In the studies of Ai et al. (2016), it was aimed to reduce the time and error rate required for intravenous injection. In this context, images were taken with two commercial mini USB cameras (1.3 megapixels). In addition, micro projector (used to back-project veins on the hand with a brightness of 200 lumens), NIR LEDs (850 nm wavelength), monitor, control circuit, support structure and base were used in the system. Cameras and LEDs were positioned 35 cm above the tissue to be viewed.

Watanabe, Sekine, Mizuno, & Miwa (2016) developed a smartphone or PC-controlled portable NIR spectroscopy system, which facilitates monitoring of changes in brain and arm muscle hemodynamics (blood circulation) and oxygenation (oxygen status/oxygenation) (in breath-hold and cuff-occlusion tests). The system measures the oxyhemoglobin, deoxyhemoglobin and total hemoglobin concentrations in the tissues. In addition, its ability to establish wireless data communication and its small and light weight allow this system to be used anywhere anytime (at home, in sports, etc.) as a healthcare application.

In the studies of Bazrafkan, Nedelcu, Costache, and Corcoran (2016), it was aimed to develop a biometric authentication system for handheld devices by the vascular structure map obtained from the finger's intermediate phalange. In this context, a small-sized system with a low-cost sensor was developed. In the study, in which 10 NIR LEDs (940 nm wavelength) were arranged circularly around a 5 MP monochrome camera (850 nm IR long-pass filter, in this filter type the visible/ambient light is blocked and contrast altered) with USB 3.0 connection, 10

images of the finger (640x480 resolution) illuminated from different angles were taken. According to the results of the study, it was shown that a biometric system, which would use the NIR images for verification purposes, could be developed with a simpler design especially for handheld devices.

In the work of Zhong, Kanhere and Chou (2017), it was aimed to protect devices from unauthorized access with a system that used vein patterns within the scope of smartphone security. In this context, the Kinect v2 depth sensor was used to capture images (raw images have 512x424 resolution while ROI images have 80x80 resolution) instead of the IR depth sensor used in three-dimensional (3D) imaging on smartphones. The hand was kept at a distance of 50-65 cm under the camera and before each image was taken, the hand was pulled away from the camera and placed back. However, this process caused slight differences, called jitter, in images of the same person. Within the scope of the trials, 40 images (the hand is made into a fist to bring the veins closer to the skin surface) of the upper surface of the hand (dorsum) of 20 users (12 men/8 women) were taken with the Kinect v2 depth sensor.

In Lingyu and Leedham's (2006) studies, how the FIR (thermography) and NIR hand vein patterns could be used in the design of a biometric system was investigated. In this context, in the NIR experimental set, dual array LEDs (850 nm wavelength) and NIR CCD camera (850 nm wavelength) were used, between which the hand was placed. Images (3 NIR and 3 FIR images of the upper surface of the hand, palm and wrist) were taken with the FIR camera placed 30 cm above the hand and the NIR camera placed 60 cm above the hand. In front of the camera lens, 3 NIR-pass filters (to use only one of the filters with different cut-off wavelengths as 720, 800 and 900 nm in each experiment) were placed respectively in order to block visible light. Within the scope of the study, it was explained that the most optimal images were obtained with an 800 nm NIR-pass filter. In addition, it was pointed out that performing vascular imaging with NIR could be more functional for biometric purposes, since it was effective in visualizing small vein details.

In Francis, Jose, Devadhas and Avinashe's (2017) studies, how to improve forearm vein images obtained with NIR image processing techniques was investigated. In addition, it was shown that NIR imaging, which was usually performed via commercially available IR cameras, could also be performed with modified digital cameras. In order to create such a camera, the NIR light blocking filter in front of the CCD in the digital camera was replaced with the visible light blocking filter. It was stated that the ROI could be illuminated with NIR LEDs in order to obtain a more effective NIR image. In the study, the NIR imaging technique with reflection was used. In addition, it was stated that increasing the visibility of the veins would provide convenience for doctors, nurses and medical students, and reduce the pain during injection for patients.

In Shrotri, Rethrekar, Patil and Korea's (2010) studies, it was aimed to develop a biometric system using hand vein patterns with a low-cost modified NIR-webcamera (\$25). In this context, a lighting system was designed in which NIR LEDs (due to their high contrast and cost-effectiveness) were arranged concentrically (indicating that it provides the most optimal lighting) around the camera. The NIR imaging system proposed in the study was stated to be suitable for use in banking transactions, personnel computer system access authorization, security systems, electronic records management, and identity control in the health sector.

In the studies of Elnasir and Shamsuddin (2014), it was aimed to develop a palm vein recognition system. In this context, images (PolyU hyper spectral palmprint database) with a wavelength range of 420-1100 nm were obtained in the system developed by using 1 liquid crystal tunable filter (LCTF), 1 CCD and 2 halogen lights. 380 different right and left palms belonging to 190 people were imaged in two separate sessions (7 images were taken in each session, 14 images in total, but only 10 were used for each palm due to missing images) over one-month period. Among the obtained images, 3800 images with a wavelength of 850 nm, in the range of 700-900 nm, which have the ability to penetrate the deepest into the tissues, were used.

In the studies of Sierro, Ferrez and Roduit (2015), it was aimed to develop non-contact imaging systems that would provide finger and palm NIR imaging. In this context, 2 palm vein imaging and 1 finger vein imaging prototypes (working with the reflection method) were developed in this study. In Prototype-1 (lighting with a single wavelength is used), 20 NIR LEDs of 940 nm wavelength (best contrast between 770, 880 and 940 nm) and 920 nm IR long-pass filter placed around the CCD camera were used. Images were taken from the tissue at a distance of 10-20 cm. A teflon plate with a thickness of 1 mm was placed on the LEDs in order to make the NIR illumination uniform. Prototype-2 (multi-wavelength illumination was used because systems using a single wavelength could be easily fooled by printing the vein structure with a standard laser printer) was designed with 4 blue LEDs of 460 nm wavelength, 4 FIR LEDs of 740 nm wavelength and 30 NIR LEDs of 940 nm wavelength placed circularly around the camera. While skin surface structure was enhanced with blue LEDs, vein structure was obtained with IR LEDs. Images were taken with a CCD camera (15 fps). In Prototype-3 (since the finger veins are deeper than the palm veins, it reduces the contrast in the image. Therefore, different lighting scheme was used on the basis of the reflection method, and thanks to this arrangement, the need for teflon plate was eliminated), which was developed to support mobile applications with USB-connection support, 12 NIR LEDs of 850 nm were used. Images were taken at a distance of at least 10 cm from the tissue.

In the study of Lee, Jung, and Kim (2011) a CCD webcam (the visible light passing filter, which blocks the NIR light, was removed and replaced with the NIR passing filter, which blocks the visible light and allows IR light with a wavelength greater than 750 nm.) with USB-connection was used in the proposed finger recognition device. As the light source, 5 NIR LEDs with a wavelength of 850 nm were connected. In addition, 1 hot mirror (reflecting NIR light and transmitting visible light) was added to the device, thus ensuring that the height of the device was low. In the study, a system with Intel Core i7 (2.67GHz) Central Processing Unit (CPU) and 6 GB Random Access Memory (RAM) was used. Since there was no additional alignment process for the finger displayed in the system, it was noted that the finger had to be kept in a certain position during image acquisition.

Moving from the necessity of a vein detection device that can clearly show the veins, Wadhwani, Sharma, Pillai, Pisal and Bhowmick (2015) developed a portable, cost-effective, painless superficial vessel imaging system based on NIR imaging in their work. In the study, only NIR images were taken and DIP was applied. In this context, a customized CCD web-camera with USB-connection and 8 NIR LEDs arranged circularly around the lens (arranged in this way to provide perfect illumination and also to create contrast in veins and surrounding tissues. However, the ringing effect light in the form of a light/dark band caused by this arrangement was corrected by sanding the LEDs with sandpaper and keeping them under water and therefore a uniform illumination was provided.) were used. The IR cut-off filter of the camera (which is in front of the sensors of all modern cameras in order to obtain the maximum amount of visible light) was removed.

In Yılmaz's (2014) study, a portable and cost-effective ophthalmoscope device (which can work by integrating with an ordinary camera) was designed that could be used to prevent possible vision loss by performing early diagnosis and follow-up of eye diseases by examination of the fundus (retina). In this context, with the retinal imaging device (using LED light source producing white colour light) to be used as a fundus machine, the retinal fundus image could be taken as desired in the experimental studies on the prototype eye. The position of the circular glows formed in the center of the image could be changed by the device being movable. It was evaluated that the proposed system can be used in the diagnosis and follow-up process of eye diseases. In addition, it was stated that the system would provide a more reliable working environment for the doctor by transferring the obtained images to the storage unit via cable (since the wireless medium is high-frequency, it has an ionization effect that adversely affects human health).

In Mahmood, Ahmed and Shahid's (2016) work, it was aimed to develop a NIR imaging system that enables the risk-free and cost-effective performance of intravenous access procedures to be used in the health sector. In this context, in the system proposed in the study, the ROI (arm) was illuminated with an 850 nm wavelength NIR light source. In addition, the image was captured with a modified

camera obtained by removing the IR-blocking filter in front of the CCD sensor of an ordinary camera and by replacing it with a visible light-blocking filter. According to the results obtained, the success rate of the system was 71.43% in obese individuals, whereas it was 100% in non-obese individuals.

2.1.1.5 The Proposed System's Imaging Technique Phase

In the Imaging Technique Phase in the Hardware Block of the study, an external NIR camera (USB 2.0 connection supported) with 5 NIR LEDs at 850 nm wavelength arranged circularly around the CMOS imaging sensor (can achieve higher frame rates than that of CCD, requires less power, but tends to have noisy images (Teledyne FLIR, 2022)) camera lens (2.0 Megapixel 3.6mm) with an IR-pass filter at 850 nm wavelength was used, taking into account the hardware criteria specified in Table 2.2.

Table 2.2 Hardware criteria and explanations used in the Imaging Technique Phase of the proposed system

Hardware Criteria Type	Explanation
Wavelength of the NIR Light	The wavelength of NIR light is extremely critical in making veins visible. NIR light can penetrate approximately 3 mm into the skin surface. Hemoglobin molecules in the blood in the superficial veins at this depth absorb the light in the range of 700-900 nm. According to the NIR optical window, more photons are absorbed by the oxygenated hemoglobin in the range of 700-800 nm, and by the deoxygenated hemoglobin in the range of 800-900 nm. Superficial vessels carry deoxygenated hemoglobin as they are usually veins. However, NIR light with a wavelength of 850 nm was preferred in the proposed system, since most studies in the literature that performed superficial vein imaging indicated that this wavelength provided better contrast.
Camera Filters	An IR-pass filter should be used on the lens so that the NIR photons (more generally, IR light) can reach the camera lens, thereby creating an image of the veins. In addition, in order for the camera to detect the NIR image, the wavelength of the NIR light that will illuminate the imaged tissue must be matched with the wavelength of the filter. Ordinary cameras on smart devices can be used to capture images in the home environment. However, ordinary smart device cameras have a visible light-passing filter and an IR-blocking filter on the lens to provide better quality images. Since it would not be very practical for the end-user to use these cameras by modifying them (by replacing camera filters), an external NIR camera (with USB-connection support) was used in this study.

Table 2.2 Continues

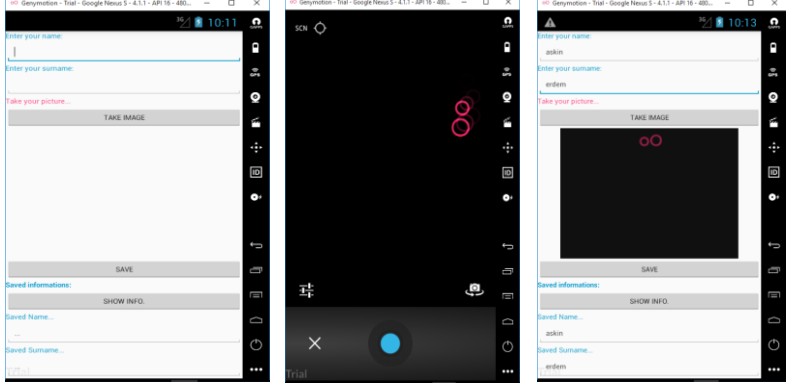
LED Positioning	The better the illumination of the tissue to be imaged, the less likely it is to form a discontinuous structure in the vein patterns. In this context, for the positioning of the LEDs that will realize NIR lighting, positioning in which the LEDs are arranged circularly around the lens, which is one of the most effective methods, as stated in most studies in the literature, was preferred.
Imaging Environment	Taking the image of the tissue in a dark environment (preferably) reduces the negative effects of visible light on the perception of the image.

The steps and explanations of the proposed system's Imaging Technique Phase are detailed in Table 2.3.

Table 2.3 Steps and explanations of the proposed system's Imaging Technique Phase

Imaging Technique Phase Step	Explanation
Illumination	The tissue region to be imaged is illuminated by NIR LEDs with a wavelength of 850 nm. LEDs are arranged circularly around the camera lens so that the light photons can be distributed evenly (homogeneously) on the tissue.
Spectroscopy	NIR spectroscopy technique is used during image acquisition from tissues. In this technique, the photons that illuminate the tissue reflect back from the surrounding tissues (skin and underlying adipose tissue) or transmitted to the lower layers. Photons transmitted to the lower layers are also absorbed according to certain properties by the hemoglobin in the veins.
Creating the Image on the Lens	The reflected photons pass through the 850 nm wavelength IR-pass filter in front of the camera lens and fall on the lens, forming the NIR raw-image. In the resulting raw-image, the photons reflected back from the surrounding tissues form the light-coloured parts, while the veins are seen as the dark-coloured parts.
Mobile Application	A wide variety of applications such as e-health, banking, games, fitness and social media can be developed for smart devices (Zhong, Kanhere, & Chou, 2017). As the proposed system targets mobile devices (smartphone, tablet computer, etc.), the NIRVeinTrackerApp mobile application has been developed within the scope of this study. With the NIRVeinTrackerApp, the user can easily perform the viewing operations after entering his/her information, and upload the NIR video recordings to the server through the same application. The application, of which interface screenshots are given in the Figures below, was developed in the Android Studio environment.

Table 2.3 Continues

	 <p>Screenshots of the NIRVeinTrackerApp mobile application interface of the proposed system</p> <p>Android Operating System, developed by Google targeting mobile devices, is an open source mobile operating system built on the Linux kernel. The Android Operating System was developed by Google and supported by the Open Handset Alliance (Open Handset Alliance, 2022) consortium of 84 technology and mobile companies (Narman, 2016; Önder & Mermerkaya, 2013). Android Studio, the official integrated development environment, is used as the code development environment. (Aliferi, 2016; Köse, 2022)</p> <p>Genymotion (Genymotion, 2022) emulator developed by Genymobile was used in order to run the developed Android application on the virtual device.</p> <p>With the Android-based NIRVeinTrackerApp application to be installed on the smart device to which the NIR camera is connected, the user takes a video recording of 30 fps (contains 30 frames per second) for 35 seconds of the tissue to be followed. The received 1920 (Horizontal) x 1080 (Vertical) pixel Full High Definition NIR raw-video recording (2,073,600 pixels) is uploaded to the server by the user via the mobile application.</p>
--	--

Summary of the proposed system's Imaging Technique Phase in 1-minute is given in Figure 2.6.

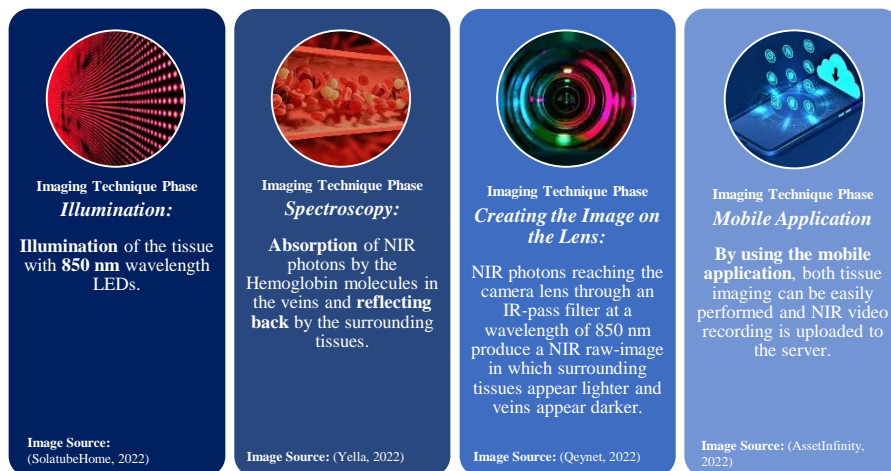


Figure 2.6 The proposed system's Imaging Technique Phase in 1-minute

2.2 The Software Block

The Software Block consists of five phases as “The Digital Image Pre-Processing Phase: Enhancing Near-Infrared Images”, “The Digital Image Post-Processing Phase: Finding the Missing Pieces”, “The Classification Phase: Recognizing Near-Infrared Images”, “Object Detection Phase: Catching Vascular Degeneration”, and “Augmented Reality Phase: Winding the Time Fast”. The Digital Image Pre-Processing Phase explains the filters and algorithms used in image enhancement, while the Digital Image Post-Processing Phase focuses on the SURF algorithm used in image matching and merging. In the Classification Phase, the designs of the CNN Model architectures, the contents of the datasets and the hyperparameters used are discussed. In the Object Detection Phase, vascular degeneration classes and their locations are detected using the YOLOv3 algorithm. The visualization of the detection results over a long period of time is provided by the Augmented Reality Phase.

2.2.1 The Digital Image Pre-Processing Phase: Enhancing NIR Images

An image is worth a thousand words. In particular, images obtained with medical imaging systems can contain very detailed information from the detection of diseases to their developmental stages. For example, X-Ray images provide effectiveness in the follow-up of bone tissue union processes, while ultrasound images can visualize ailments in soft tissues such as internal organs. Superficial veins (vascular structures appear darker in NIR raw images) can be detected via NIR spectroscopy. Although an image where the veins are not clear does not make the separation of veins and surrounding tissue difficult for an expert eye, but it causes a computer-based detection and tracking system to be very prone to errors in terms of classification or object detection. What matters is how to access the "hidden" detailed information contained in these raw images so that they can be used in an automated system.

In the Digital Image Pre-Processing Phase, the enhancement of the NIR video recordings uploaded to the system by the user, with various DIP filters and methods,

in order to use them more effectively in the next phases of the proposed system is explained.

2.2.1.1 Digital Image Processing

DIP, which is an interdisciplinary field (covering many fields such as computer science, optics, mathematics, artificial intelligence and pattern recognition), describes the modification of a digital image (the visual representation of the object, place or person obtained by an optical device such as a camera is used. This representation is a 2D image matrix containing a finite number of matrix elements called pixel, pel or picture element. Details in the image are directly proportional to the number of pixels that make up the resolution value), usually by automatic processing with digital information processing devices (using certain algorithms). (Gonzalez & Woods, 2002; Marques, 2011; Osmanoglu, 2016)

The most basic operation performed in DIP is to obtain a new and modified output image by changing the input image via operators (filtering, rotation, scaling, etc.). (Marques, 2011; Petrou & Petrou, 2010).

DIP is used for 4-main purposes as: (1) Enhancement (clarification of important parts by increasing the quality of the image by methods such as sharpening), (2) Compression (encoding the image with lossy coding, which describes the reduction of the number of bits representing the image while keeping the quality degradation at an acceptable level, or lossless coding, which describes the preservation of quality), (3) Restoration (improving clarity by reducing noise or blurring), and (4) Segmentation and Labelling (identification of image by obtaining image-specific characteristics) (Marques, 2011; Petrou & Petrou, 2010). The steps used for these processes are divided into three parts: (1) Low-Level Methods (contrast enhancement, noise reduction), (2) Mid-Level Methods (region or edge extraction) and (3) High-Level Methods (image classification, analysis or object detection) (Marques, 2011).

The steps and explanations used in DIP in the studies are given in Table 2.4 according to the order used in this thesis.

Table 2.4 Digital image processing steps and explanations

Digital Image Processing Step	Explanation
Image Acquisition	Taking an image or video recording with an imaging device.
Region of Interest (ROI) Extraction	Determining the parts (ROI) to be used in the image for the purpose and eliminating the other parts. No matter which device is used in vein imaging, unnecessary parts that will not be used in recognition/detection processes are also included in the image. In order to perform an effective and fast recognition without increasing the processing load, these parts must be separated from the image. (Liu, Li, Sun, Zhang, & Chen, 2017)
Grayscale Conversion	While a standard camera image is represented by three 8-bit numerical pixel values in the range 0-255 showing three separate channels as a colour image (Red, Green, Blue), when converted to a monochrome image (to grayscale) it will be represented by a single 8-bit (generally) pixel value. In the grayscale image, the lightness value can be represented by specifying black with a pixel value of 0, white with 255 and different levels of gray with all values in between. In this way, it is ensured that the processing load is reduced by using images containing less information. (Marques, 2011; Osmanoğlu, 2016)
Noise Removal	There are two types of filters, linear and non-linear , to remove noise (undesirable colour or brightness changes that are not in the Real Environment (RE) but are randomly mixed into the image from external sources) that reduces image quality. Different digital filters (Filters working in spatial-domain : Mean Filter (Arithmetic, Geometric or Harmonic) or Median or Order Statistic/Rank Filter (Median, Min and Max, Midpoint). Filters working in frequency-domain : Bandpass, Bandreject, Notch Filter) are used according to the noise type (Gaussian, Impulse/Salt and Pepper, Uniform, Gamma/Erlang, Rayleigh, Exponential, Periodic Noise). (Marques, 2011; Osmanoğlu, 2016; Verma & Dubey, 2015)
Contrast Enhancement	Increasing the image contrast means making the objects in the image more prominent. It is a fundamental method in facilitating edge detection (allows the separation of objects in the image). The sharpness of the image will be increased by making the edges and details more prominent (Marques, 2011). It also helps to perform an easier segmentation process (segmentation, which is the process of dividing the image into specific segments that are more meaningful, is usually achieved by the use of thresholding. In this way, the grayscale image is converted to a binary image) (Verma & Dubey, 2015). For example, the Histogram equalization (the normalization of pixel values according to frequency and because all pixel values are used, brightness differences can have a negative effect on the result), the Adaptive Histogram Equalization (used to improve local contrast to eliminate the negative effect (Benziane & Benyettou, 2015). By dividing the image matrix into blocks/segments, the histograms of these blocks are taken into account, and the lightness values are determined by performing local equalization (Benziane & Benyettou, 2015; Cömert, 2015). However, this time, cuts may occur due to noise and contrast imbalance between local blocks (Benziane & Benyettou, 2015;

Table 2.4 Continues

	Cömert, 2015). There is also the possibility of increasing the noise in homogeneously distributed regions of the image (Benziane & Benyettou, 2015)), or the Contrast Limited Adaptive Histogram Equalization (CLAHE) (distributes the discontinuities in local histogram equalization, the horizontal/vertical interpolation between blocks, and the maximum limits of contrast to the blocks. In order not to increase the noise-containing components in the CLAHE Algorithm, the image is developed into small regions by dividing it into user-defined sub-images (Şeker & Engin, 2017)) can be used. (Cömert, 2015)
Edge Detection	Identifying the most effective edges (defined as the border between two image parts that have differences in terms of features such as colour, pattern) to represent the objects in the image facilitates the segmentation of the image. In this step, the intensity values of neighbouring pixels are compared and the parts where a significant change in intensity values are observed are considered as edge regions. Performing edge detection in a noise-free image is extremely important for the effectiveness of the results. There are varieties of edge detection algorithms such as Canny Edge Detector (the most popular), Roberts Cross Operator, Sobel Operator and Gabor Filter. (Marques, 2011; Öztürk & Akdemir, 2015)
Thresholding/Binarization	Thresholding is the process of converting a grayscale image into a binary image containing only black and white, according to a certain threshold value (pixel values more than or equal to the threshold value is white-255, below is black-0), and thus eliminating unwanted parts in the image (Shrotri, Rethrekar, Patil, & Kore, 2010). In addition, the process can be accelerated by reducing the 256-level grayscale image to 2 levels. For this, two-level (only black and white) grayscale image transformation, which is usually called binarization, is used (Marques, 2011). In this way, segmentation can be easily performed. In addition, in binary images, colour transformation can be achieved by making 0's 1's and 1's 0's in the image pixels with the complement process (even if grayscale or Red-Green-Blue (RGB) images are used, the complement process can be applied again and the complement calculation can be made by subtracting the highest pixel value from each pixel value (Imcomplement, 2022)).
Morphological Operations	Dilation (used to thicken/enlarge objects in binary image) and Erosion (used to thin/shrink objects in binary image), which use a neighbourhood-based (matrix with a particular pixel in its center and "neighbouring" pixels in its immediate vicinity) structuring element (a matrix consisting of 1s and 0s that explores the size and value compatibility of the corresponding image region with itself by wandering over the image), are the most popular morphological operations (Marques, 2011). Another process, Skeletonization , defines the contours of objects with a single-pixel thickness and forms the basis for processing such as object recognition and matching (Ma, Ren, Tsviatkou, & Kanapelka, 2022).
Blurring/Deblurring	With blurring , patterns or details in the image are intentionally distorted (removal of unnecessary details for secrecy, number detection or shape recognition), while with deblurring , the blurred image is corrected (due to movement of camera/imaged object or focusing disturbances) (Marques, 2011). (This step was not used in the thesis study)
Segmentation and Labelling	DIP-based classification is one of the methods frequently used in clustering, comparison or object recognition studies (Marques, 2011). (This step was carried out by CNN within the scope of the thesis study.)

2.2.1.2 Literature Review of Digital Image Processing

In most of the studies on DIP in the literature, the steps in Table 2.4 are used.

The raw images obtained in the studies of Chiao et al. (2013) are first converted to greyscale images and the user is shown an enhanced video of the vascular system almost in real time.

In the study of Şeker and Engin (2017), in which they performed vascular network analysis from NIR images of deep tissues, the vascular structure was obtained with reflectance geometry. Grayscale transformation, contrast enhancement with CLAHE, and binary image transformation with thresholding were applied to the NIR raw RGB images obtained with the setup in the study.

In the studies of Bazrafkan, Nedelcu, Costache and Corcoran (2016), a biometric authentication system was developed for handheld devices. Each NIR image obtained was divided into three main regions, and the region-2 (since in region-1 the density is very high and in region-3 it is very low), which showed the best vein image and which had the most optimal density, was combined in all 10 images to form a final fused-image. While the details in the final image were made visible with a high-pass temporal median filter, Gabor filter was used for vascular structure extraction.

In the study of Lingyu and Leedham (2006), in which they designed a biometric system, only the images taken from the top of the hand forming the FIR database went through five different image processing stages as: (1) Hand image acquisition, (2) Image enhancement and ROI selection, (3) Vessel pattern segmentation, (4) Skeletonization, and (5) Matching.

In Shrotri, Rethrekar, Patil and Korea's (2010) studies, the OpenCV1.1 library was used on Microsoft Visual C++ 2003 Edition in Pentium Core2duo 2.3 GHz CPU environment for image processing within the scope of using hand vein patterns for

biometric purposes. Improvement of the image in the work was performed in 8-steps as: (1) Acquisition of NIR image, (2) Pre-processing, (3) Obtaining ROI, (4) Grayscale transformation, (5) Thresholding, (6) Edge Detection, (7) Removal of unwanted objects, and (8) Thinning. Within the scope of future studies, it was reported that the system would be expanded to include data comparison and validation phases.

In Francis, Jose, Devadhas and Avinashe's (2017) studies aiming to detect and improve vein images, raw images in jpeg format taken from the tissue by a modified NIR camera were passed through image processing phases via MATLAB® R2014a. The image processing steps in the study consist of 8-steps as: (1) **Grayscale conversion** (conversion of 3D RGB based NIR raw images to 2D grayscale), (2) **Noise reduction and smoothing** (removal of salt and pepper type small noises originating from the image acquisition mechanism via the 2D Median Filter and smoothing by the Gaussian Filter), (3) **Contrast enhancement** (CLAHE applied 2 times in succession to increase the image density so that the background can be separated from the foreground more easily, and the 2D Median Filter applied 2 times sequentially to make the separation a little clearer by removing minor imperfections in the image), (4) **Edge detection** (with the orientation-sensitive 2D Gabor Filter, obscure vein edges were slightly more pronounced according to a particular orientation, in the study, the angular theta value of the Gabor Filter was determined as the fixed value of 101.25 as the viewing angle of the forearm), (5) **Binarization** (determining the threshold value with Otsu's thresholding method, completely distinguishing the veins and surrounding tissues from each other by expressing them only with black and white colours), (6) **Morphological operations** (more correction of vein shapes with dilation/pixel insertion and erosion/pixel deletion operations, only erosion was used. By taking the complement of the obtained image, the veins were represented in white), (7) **Region of Interest** (selecting only the region representing the veins and deleting other parts from the image by thresholding the image with the median pixel value of 100), (8) **Real-time visibility** (superimposing veins in the obtained region of interest as red coloured to the corresponding location on the raw NIR image). Within the scope of future studies, it was explained that the

algorithm can be used for real-time vessel tracking by considering (if all frames were used system performance would reduce, therefore it was predicted that 16 frames per second would be sufficient) each frame of real-time video inputs as a separate image.

Wadhwani, Sharma, Pillai, Pisal, and Bhowmick (2015) used MATLAB® and Laboratory Virtual Instrument Engineering Workbench (LabVIEW, to take real-time images) in their studies where they developed a vein detection system. The algorithm used for image processing consisted of 4-steps as: (1) **image acquisition** (taking an image of the forearm as 8-bit real-time video snapshots of 640x480 pixels with a modified webcam), (2) **pre-processing** (increasing the image quality with ROI, noise removal via Gaussian and median filter, smoothing, and contrast enhancement processes so that vein patterns could be detected more easily in segmentation), (3) **segmentation** (performing global threshold and foreground/vein and background/surrounding tissue separation), and (4) **post-processing** (correction of other noise, shadows and indistinguished veins in the grayscale image obtained in the previous step). Real-time video snapshots were obtained in trials of the study where veins were visualized in anterior forearm images. Although the images obtained could show the vascular structures, it was stated that the system needed additional software adjustments for further improvements in the images.

In their study, Verma and Dubey (2015) aimed to obtain palm vein patterns by DIP (noise reduction and image enhancement) from NIR vein multispectral images. In this context, the study coded with MATLAB® had 4-steps: (1) **Raw image acquisition** (multispectral acquisition of palm and hand surface vein images with a CCD camera capable of detecting NIR light) (2) **Noise removing** (increasing clarity by improving image quality with filters) (3) **Contrast Enhancement** (providing more successful results with the Adjustment Method in the segmentation process) (4) **Segmentation** (obtaining ROI via separating the hand from the background by specifying the black pixels with 0 and the white pixels with 255 or 1 in the image). Five different filters (Median, Blind Convolution, Weiner, Regularised and Lucy Richardson filters) were used in the experiments, and it was found that Weiner and Median filters among them were able to remove salt and pepper style noise without

causing blurring at the edges of the image. In addition, Weiner and Median filters were found to be more effective than other filters in providing noise removal without blurring the edges. According to the results obtained, it was explained that palm patterns could be obtained from the multispectral image.

In Sierro, Ferrez and Roduit's (2015) studies aiming at finger and palm NIR imaging, 3-steps were used in order to obtain ROI from the raw palm vein image obtained (contrast is increased by applying a simple unsharp mask to make the vascular patterns clear) in Prototype-1, 2 and 3 in terms of image processing: (1) Application of thresholding (value=50) to 8-bit raw image (hand in foreground is white, background is black), (2) Extracting hand contours with standard edge detection (the part with the hand image, which was the region with the largest area, was selected, the others were deleted), (3) Determination of ROI representing palm and fingers. An image dataset was created using prototypes for the experiments. For Prototype-1, 5 palm images (for each session) of 110 people's right and left hands were taken in 2 recording sessions, and a total of 2200 images in PNG format with 640 x 480 resolution were obtained. For Prototype-2, 5 palm images of 84 people's right and left hands were taken once for each wavelength, and a total of 2520 images in PNG format with 656 x 490 resolution were obtained. For Prototype-3, 5 images were taken once from the right and left hands of 84 people, and a total of 840 images in PNG format with 640x480 resolution were taken. Within the scope of future studies, it was stated that it had been planned to develop the system by performing raw image processing, feature selection-extraction and classification on the images taken from the prototypes in terms of software. In addition, it was underlined that the integration (it is stated that such a system that will work with reflected light should be able to take good quality images in different ambient conditions, while also taking into account power constraints) of a multi-spectral low-cost mobile finger vein authentication system into smart devices would be investigated.

In the work of Lee, Jung, and Kim (2011) a modified Gaussian high-pass filter-based finger recognition method was proposed. The images detected by the camera were selected in the range 640x480 to 128x96 pixels to reduce time complexity and avoid pixel noise. The resulting images were converted to black and white format by

binarization (with 250 thresholding) and enhanced with a Modified Gaussian High-pass filter.

In the work of Rao, Zhang ve Sun (2017), a filter design based on Optical Angular Spectrum Theory (a method of expanding into a complex wave field with respect to spatial frequencies) was carried out. With the filter, the spatial frequency of thresholding was determined to separate the leg veins and hairs in the NIR images. Digital filtering algorithms were used to remove hair noise. In the study, 8-bit greyscale NIR images with 256x256 pixels were obtained. By stating that the veins in the NIR image were represented by low spatial frequency information (global information: orientation and rates), and the hairs were represented by high spatial frequency information (small/fine details, sudden spatial changes: edges), spatial frequencies in the image were obtained by Fast Fourier Transform (FFT). According to the spatial frequency information of the hairs, the transfer function (Lowpass, Highpass and Butterworth Annulus digital filters) and subsequently the inverse FFT were applied. In this way, noise-free, high-quality and contrast-enhanced NIR images were obtained. According to the trials performed on two men and one woman, the Butterworth Annulus digital filter was found to be more successful than the others in removing hair noise and increasing the contrast in images by the naked eye (noting that the human visual system worked like a band-pass filter operating in the spatial frequency domain) and image processing quality metrics. It was emphasized that further diagnosis could be made with the improved images.

In the studies of Fernández and Armada (2017), a system that performs catheter placement robotically was developed. The pre-processing part of the system consisted of 5-steps: (1) The SWIR image was divided into two regions (foreground: parts with veins, background: pixel parts that would not be used in the study) by applying segmentation, (2) By determining a global threshold value with Otsu's Method, the distinction between these two regions were clarified, (3) With two sequential dilatation procedures (with two square structuring elements of 10 and 15 pixels), the pixels that would remain in the background and foreground intermediate region and that would cause trouble in vessel detection were eliminated, (4) Fixed

non-uniform lightings with CLAHE, and (5) Foreground and background contrast was slightly improved by image density adjustment. Feature extraction was achieved by subjecting pre-processed SWIR images to maximum curvature method and k-means clustering. The registration of the veins in the SWIR image was made by using the range data obtained via the TOF camera and a common reference frame was created. The resulting vascular structures were superimposed on the relevant regions in 3D. According to the experiments, although both the maximum curvature method and k-means clustering provided the desired level of detection precision, better results were obtained with k-means clustering.

The images obtained with the system proposed in the studies of Mahmood, Ahmed and Shahid (2016) were processed with the OpenCV infrastructure as an embedded system application with Raspberry Pi. The system could detect 24 frames per second, supporting real-time, and the frames were processed and displayed separately. In addition, the image was projected onto the ROI with the Pocket Pico-Projector. Trials of the study included 4-separate methods: (1) Images taken with a normal camera were used without processing (veins could not be detected), (2) Images taken with a night vision camera (unlike an IR camera, since it contains no visible light and no IR light filter, it allows both types of light to reach the lens) were used without processing (veins could be detected), (3) Images were processed with Histogram Equalization or Window/Level (increasing the contrast, making the vessels appear darker), and (4) By adding the IR-pass filter between the camera lens and the CCD and processing the images with CLAHE. It was stated that a higher quality camera using 750 nm wavelength (it was been stated that this wavelength was more absorbed by the deoxy-hemoglobin in the vein) and a better projector would be used within the scope of future studies.

In the studies of Seymen, Boyraz, Çetin and Yıldız (2014), NIR images were taken by using NIR CCD camera, IR lens and IR LEDs and were processed by DIP techniques. The ROI parts of the obtained images were first reduced to grayscale and their contrasts were adjusted (to clear blur in the image of tissue regions that could not be illuminated homogeneously with NIR LEDs). In the contrast adjustment, all

pixels were updated by using the largest and smallest values of the pixels in the image. Noises were removed with the Weiner Low-Pass filter, and the image was converted to black and white with Adaptive Threshold. The noise in the processed images was removed with the Gaussian Low-pass filter. The processed image was also slightly clarified with the median filter (which gives better results in small areas). In the trials of the study, a dataset of 1575 images containing right and left hand images of 100 people was used. Before the imaging in order to physically clarify the veins, 3 kg bag was carried by the participants, the elastic ball was squeezed for 1 minute, and cold ice was passed over their hands. With the developed system, vein extraction could be performed clearly. Within the scope of future studies, it was stated that it had been aimed to develop a system that would facilitate vein detection in injection and laser applications by the NIR vein imaging system that reflects the processed image back with a projector. It was also underlined that the application was planned to be integrated into a mobile vein imaging device.

In Yılmaz's (2014) study, in which a portable and cost-effective ophthalmoscope was designed, an intelligent system design that could make both eye diseases' diagnosis and retinal identification was proposed. The retinal vascular network structure in the images was obtained by using MATLAB® image processing (eye structures such as retina, optic disc and macula were separated from each other). In the study, the best details in the 565x585 resolution colour images were taken with the G (green) component for the retina and macular vessels, and the B (blue) component for the optic disc, and accordingly, the relevant component of each tissue was expressed in gray scale (with the border range of gray colour 0-255) in order to better analyze the images. In retinal fundus images (images taken from the international DRIVE retina database and adjusted to 800x800 resolution for fast processing), image processing techniques used to highlight retinal vessels, optic disc (creating reference for veins) and other tissue components and to better feature retinal recognition were stated as: (1) Image Enhancement, (2) Filterings, (3) Conversion actions (H-minima was used), (4) Histogram equalization (CLAHE), (5) Morphological operations, and (6) Edge finding.

In Osmanoğlu's (2016) study, the usability of image processing and analysis method with MATLAB® as an auxiliary analysis technique in the field of health in intraocular lens measurement was examined. In this context, the intraocular lens power measurement parameters (optical parameters obtained with optical biometry, which is the gold standard in routine eye applications: Anterior chamber depth, lens thickness and axial length) were obtained by processing the raw MRI images with MATLAB® R2011b in 4-steps as; (1) Conversion of MRI image pixel values to mm (using *imtool*, it had been determined as 1 pixel = 0.257 mm), (2) Obtaining the ROI (using *imcrop*), (3) Converting the type of the cropped image to double and then to binary (by using *double*, *mat2gray*, *graythresh* and *im2bw*, the automatically calculated threshold value was used at this stage), (4) Calculation of optical parameters (using *imtool*). In the trials, 10 patients with Orbita MRI indication were studied after routine eye examination in Eskişehir Osmangazi University Health Practice and Research Hospital Ophthalmology Clinic. Two measurements were taken from each patient's left and right eyes, and optical biometry was applied by creating a dataset consisting of 20 images. Trials were conducted on a system with Intel Skylake Core i5 (4 cores) CPU 3.2 GHz and 8 GB DDR4 RAM. According to the test results of the study, it was stated that intraocular lens power measurement data obtained with MATLAB® application from MRI images taken for other purposes could be a helpful method for physicians in cases where it would be difficult to perform optical biometry (for example, excessive transparency of the lens).

2.2.1.3 The Proposed System's Digital Image Pre-Processing Phase

During the Imaging Technique Phase, the NIR video recording that the user uploads to the server is stored on the server as belonging to the person. At this phase of the thesis, the vascular structures in the NIR raw-images obtained from the video recording are more or less distinguishable with the naked eye (because it is darker than other parts of the image), but are not functional enough to achieve high levels of accuracy in the classification or object detection phase. For this reason, by clarifying the images with DIP techniques, removing unwanted parts or removing noise, the

veins will be more pronounced, increasing the accuracy in other phases and reducing the processing load. In the Digital Image Pre-Processing Phase, from the NIR video input, NIR modified-images in which vascular structures can be clearly selected are obtained in black and white format. The Digital Image Preprocessing Phase of the proposed system was coded with the MATLAB[®] R2017a (MathWorks Inc., 1996) program.

In DIP, the MATLAB[®] (Matrix Laboratory) program, which has toolboxes containing special functions and commands, is frequently used in many fields, especially in engineering applications. MATLAB[®] is preferred especially because of its programming language that provides particularly easy coding, easy-to-use development environment supporting matrix-based operations and graphical features (Marques, 2011).

In MATLAB[®], images are represented as a matrix of which elements are all pixel values. Processing a digital image consists of 5-basic steps: (1) **Image acquisition** (as photo frame or video with CCD or CMOS camera), (2) **Image pre-processing** (performing noise removal, contrast enhancement, sharpening and edge detection on the raw image), (3) **Segmentation** (separating the part of the image to be worked on from the other parts for a specific purpose), (4) **Feature Extraction** (identification of vertices, intersections, curves or gaps in the image that may be decisive in the ROI), (5) **Recognition** (assigning identifiers/labels to objects to be detected in the image according to the specified features) (McAndrew, 2022).

The basis of the vessel recognition system is DIP, which provides image enhancement (filtering, segmentation, noise removal, ROI extraction, etc.) (Verma & Dubey, 2015). In order to obtain better accuracy values in image classification and object detection processes to be carried out in the next phases of this thesis, it is extremely critical that the images contain information only for the purpose as much as possible.

The steps, MATLAB[®] commands and explanations of the proposed system's Digital Image Pre-Processing Phase are detailed in Table 2.5.

Table 2.5 Steps and explanations of the proposed system's Digital Image Pre-Processing Phase

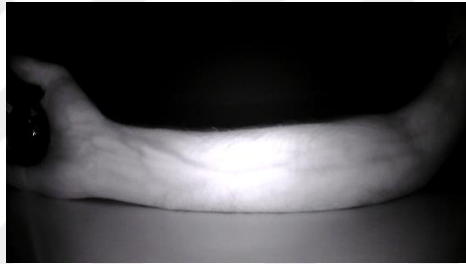
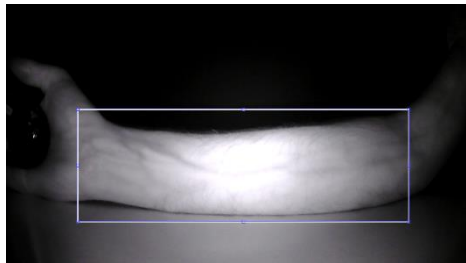
Digital Image Pre-Processing Phase Step	Explanation
Image Extraction from Video	<p>In the Image Extraction from Video Step, the 3D NIR video recording consisting of RGB components is converted to NIR raw-images (1080x1920x3 uint8). Each video is 30 fps, and a total of 1050 frames are obtained from a 35-second video. The NIR raw-image used as representative is shown in the Figure below. (<i>readFrame</i> command is used)</p>  <p>As representative, the NIR raw-image with the code Period1_Img0_RightArm.bmp is used</p>
Region of Interest (ROI) Extraction	<p>In the image processing steps, using only the region containing the veins (assuming the upper left corner of the image as the point (0, 0), the rectangular region represented by the coordinates [x_{min} y_{min} width height] = [500.5 314.5 1002 428] is used for Period-1 right forearm images) instead of using the whole image reduces the processing load. For this, first of all, the ROI in the image is determined as shown in the Figure below and extracted from the other parts, and the NIR cropped-raw image (447x1049x3 uint8) is obtained. (<i>imcrop</i> command is used)</p>  <p>Region of Interest identified in NIR raw-image</p>

Table 2.5 Continues



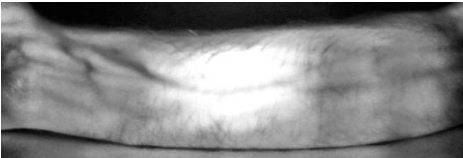
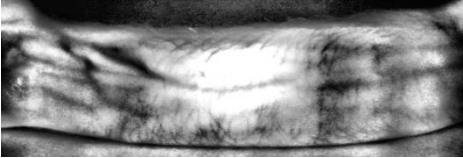
Grayscale Conversion	<p>In order to ease the processing load in the next steps and to use the 2D median filter (accepts 2D grayscale or binary image input), the cropped 3D input image is converted to the 2D grayscale image (447x1049 uint8) shown in the Figure below. (<i>rgb2gray</i> command is used)</p>  <p>Image converted to grayscale</p>
Noise Removal	<p>A 2D median filter is applied to remove impulse (salt and pepper) type noise (in this noise type, the noise pixel takes the maximum or minimum value of the gray value (Liang, Li, & Zhao, 2021)) in the grayscale image. The Median Filter is a non-linear filter and it can protect the edges while reducing the noise (salt and pepper) (Demir, 2019). For the median filter, the median of the 3x3 neighbourhood pixels is taken by default and the denoised (noise-free) grayscale image (447x1049 uint8) is obtained as in the Figure below. (<i>medfilt2</i> command is used)</p>  <p>Denoised grayscale image</p>
Contrast Enhancement	<p>The brightness and dark values of the objects in the grayscale image are adjusted by increasing the contrast, so that the veins in the image can be made clearer and better distinguishable from the surrounding tissue. In CLAHE, the contrast is increased on small parts of the image, called tiles, instead of the entire image, ensuring that the resulting histogram of each part matches the specified histogram distribution and is combined with neighbouring tiles (Adapthisteq, 2022). In this context, the CLAHE function is applied twice sequentially for a more distinctive appearance and the contrast corrected images (447x1049 uint8) are obtained as shown in the Figures below. (<i>adapthisteq</i> command is used)</p>  <p>Obtained grayscale image after applying CLAHE for the first time</p>  <p>Obtained grayscale image after applying CLAHE for the second time</p>

Table 2.5 Continues

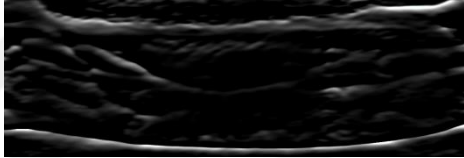


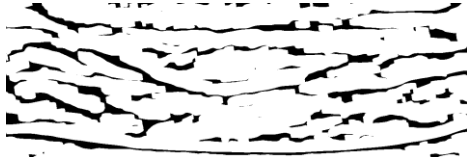
Edge Detection	<p>Gabor Filtering, which is one of the edge detection algorithms, is used to make the vein edges a little more specific. Since the Gabor Filter is more successful in detecting edges extending in a certain direction in a grayscale image, the tissue to be imaged is expected to be kept roughly in a certain direction. By the trial-and-error method, the most appropriate filter orientation value is determined as 95 degrees. However, in this thesis study, it is predicted that there will be angular shifts in the image, since a system aimed at home use has been developed (therefore, in the Digital Image Post-Processing Phase, the system has been adjusted to withstand the rotation difference in the range of $[-0.07, 0.07]$ and the scale difference in the range of $[0, +0.1]$). The Gabor filter produces two output values as magnitude (447x1049 double) and phase (447x1049 double) response (Imgaborfilt, 2022). The phase response where the detected edges are visualized is shown in the Figure below. (<i>imgaborfilt</i> command is used)</p>  <p>Gabor filtered grayscale image's phase response</p>
Thresholding/ Binarization	<p>Phase response from Gabor Filter outputs is used for binarization. With Binarization, the grayscale (multi-tone) image is converted to a binary image (447x1049 logical) consisting of only black and white (two tone) pixel values. In the study, all pixels above the global image threshold value determined automatically by Otsu's Thresholding are assigned a value of 1 and the others a value of 0 (Imbinarize, 2022). The resulting binary image is shown in the Figure below. (<i>imbinarize</i> command is used)</p>  <p>Phase response image converted to binary image</p>
Morphological Operations	<p>Binary erosion is used to clean up the unnecessary small pixel residues on the vein edges in the binary image. Since the veins are line-shaped patterns, the linear (line-shaped) morphological structuring element is determined as 95 degrees in accordance with the Gabor Filter angle and the eroded-image (447x1049 logical) is obtained as in the Figure given below. (<i>strel</i> and <i>imerode</i> commands are used)</p>  <p>Binary image with pixel erosion applied</p>

Table 2.5 Continues

<p>Colour Conversion</p>	<p>In the Colour Conversion Step, vein colours represented as white in binary images are inverted by applying the complement process (0 pixel values are changed to 1 and 1's to 0's) in order to adapt to the CNN Models (in the original Modified National Institute of Standards and Technology (MNIST) dataset used by the MNIST Model to be used in the Classification Phase, white background, black foreground format was used). Thus, the NIR modified-image (447x1049 logical), in which the veins are represented in black and the surrounding tissues in white as shown in the Figure below is created from the NIR raw-image. (<i>imcomplement</i> command is used)</p>  <p>Binary image with complement process applied (NIR modified-image with code Modified_Month1_Img0_RightArm.bmp is obtained)</p>
---------------------------------	---

Digital Image Pre-Processing Phase's Grayscale Conversion, Noise Removal, Contrast Enhancement, Edge Detection, Thresholding/Binarization and Morphological Operations Steps of this thesis are based on the work of Francis, Jose, Devadhas, & Avinashe (2017). Differently, the video recording is converted into image frames with the Image Extraction from Video Step, Region of Interest (ROI) Extraction Step is performed in the secondly before all other steps, and these steps of the thesis are introduced in the study of Erdem, Erdem & Utku (2020). In addition, as the last step, in the Colour Conversion Step, colour transformation is made in such a way that the veins are black and the surrounding tissues are white, and this step is explained in general terms in Erdem & Utku's (2022b) study together with the other phases of the thesis.

Summary of the proposed system's Digital Image Pre-Processing Phase in 1-minute is given in Figure 2.7.

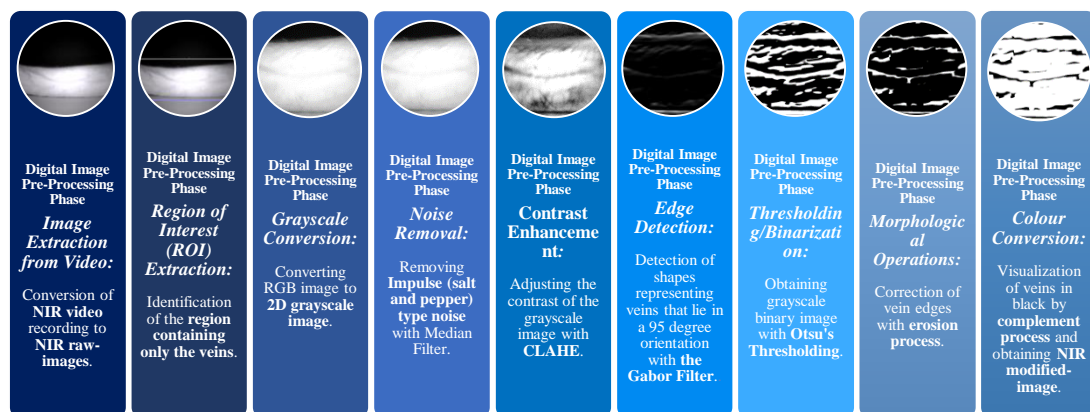


Figure 2.7 The proposed system's Digital Image Pre-Processing Phase in 1-minute

2.2.2 The Digital Image Post-Processing Phase: Finding the Missing Pieces

An effective recognition system should be able to recognize a known pattern regardless of its state (rotation and scale) or the environment in which it is viewed. The ambient lighting required for imaging, the angle of the camera, the location of the tissue to be viewed, or the proximity/distance from the camera may not always be perfect. It is extremely critical to consider these conditions, especially in an imaging system to be used within the scope of e-health, because the person receiving the image is just the user, not an expert in imaging. Although these problems can be overcome to a certain extent by video recording, there may still be differences between the previous frame and the next frame (due to slight shake of the camera or tissue not being fully illuminated) when the video recording is converted to an image. The vein patterns in the video recordings are similar in shape, but differences due to lighting may cause discontinuous structures. Therefore, the user is expected to move the tissue slightly during video recording, thereby illuminating the tissue from all directions. Thus, in the video recording, some parts of the veins that can be seen intermittently in the previous frames will be displayed properly in the next frames by changing the illumination angle (since the illumination angle changes, the veins that are imaged continuously in the first frames may also be seen intermittently in the next frames). The action to be performed is to complete the missing piece in the next frame with the full piece in the previous frame. What matters is how the missing parts can be completed independently of the effects of rotation and the environment.

In the Digital Image Post-Processing Phase, consecutive NIR modified-images obtained from NIR video recording are matched with feature detection and matching methods, and the rotation and scale differences between the two modified-images are detected, and the images are fused by aligning the second modified-image with the first one.

2.2.2.1 Digital Image Matching

Features in the images (specific characteristics such as point, line, blob or edge shapes (Mistry & Banerjee, 2017)) are used to determine the similarity between the images. By matching the properly extracted features, operations such as image recognition, object detection or tracking can be performed (Mistry & Banerjee, 2017; Çan, 2021).

Two methods are used in the literature to match vessel images. In the **Key Points method**, the similar points in the test image are matched with the image with a specific class by using vein endings and crossings, while in the **Similarity score method**, the morphological similarities between two vein patterns are calculated. (Zhong, Kanhere, & Chou, 2017)

For example, among the fingerprint (suitable to be used for biometric purposes due to its uniqueness and its properties that do not change over time) matching techniques (used for verification with an automated matching process between two fingerprints) there are 3-methods used within the scope of recognition: (1) **Correlation-based matching** (the two compared images are superimposed to determine the correlation between them), (2) **Minutiae-based matching** (compares new sample with minutiae which is obtained from two fingerprints and which is stored as a set of points in the 2D plane), and (3) **Pattern/image-based matching** (patterns between the saved sample and the queried sample are compared by aligning them to a specific central point) (Bana & Kaur, 2011).

Although the skelotonization process is generally used for feature extraction, skeletons can cause high False Acceptance Rates in trials. For this, methods that combine multiple features are also used. (Benziane & Benyettou, 2015)

Algorithms used in detecting and matching features are given in Table 2.6.

Table 2.6 Feature detection and matching algorithms and explanations (Çan, 2021; Devi & Kirubaharan, 2015; Mistry & Banerjee, 2017; Özgen & Sarıtaş, 2022)

Feature Detection and Matching Algorithm Name	Explanation
BRIEF (Binary Robust Independent Elementary Features)	Detection can be performed independently of the blur effect that causes deterioration in images (Çan, 2021). Since it only creates a key point identifier (descriptor operation), it should be used with an additional feature algorithm (Bayraktar & Boyraz, 2017; Çan, 2021).
FAST (Features from Accelerated Segment Test)	In feature extraction, key points determined by using corner points are used. (Çan, 2021). Since it only creates a key point identifier (descriptor operation), it should be used with an additional feature algorithm (Bayraktar & Boyraz, 2017; Çan, 2021).
ORB (Oriented FAST and Rotated BRIEF)	Detection can be performed independently of the light intensity effect that causes distortion in the images (Çan, 2021). While it provides the identification of key points with FAST, it also determines the attributes of key points with BRIEF (Çan, 2021; Işık, 2015). It is an alternative algorithm to Scale Invariant Feature Transform (SIFT) and SURF (Çan, 2021). However, it has less scale independence than SURF and does not work on blurry images (Bansal, Kumar, & Kumar, 2021). It performs both detector and descriptor operations (Bayraktar & Boyraz, 2017).
SIFT (Scale Invariant Feature Transform)	SIFT and SURF are the two most powerful methods for feature extraction and matching, regardless of image distortions (rotation, scale, blur, contrast or illumination) (Çan, 2021; Mistry & Banerjee, 2017; Özgen & Sarıtaş, 2022). Detects corner features in images (Çan, 2021). The working steps of the SIFT algorithm, which basically uses the Difference of Gaussians Edge Enhancement Algorithm calculation (Mistry & Banerjee, 2017): (1) Scale space $L(x, y, \alpha)$ is created by multiplying the $I(x,y)$ image matrix with the Gaussian Filter $G(x, y, \alpha)$ with different standard deviation (α) values, and the position of the potential key points that remain unchanged at each scale is determined. (2) With Difference of Gaussians, unstable (affected by image distortion) key points are eliminated and real key points are located. (3) The orientations of the determined key points are determined by ensuring their durability against 2D cyclic distortions. (4) Key point descriptors are detected by providing resistance against other corruptions. (Işık, 2015; Mistry & Banerjee, 2017; Oral & Gençal, 2016; Özgen & Sarıtaş, 2022)

Table 2.6 Continues

	<p>It is superior to SURF in being independent of scale and saturation effect (Mistry & Banerjee, 2017). There is no significant difference between SIFT and SURF in terms of lighting effect (Mistry & Banerjee, 2017). It performs both detector and descriptor operations (Bayraktar & Boyraz, 2017).</p> <p>It requires high computational power for feature description, detection and matching. For this reason, SURF is preferred for feature tracking in real-time Augmented Reality (AR) applications. (Billinghurst, Clark, & Lee, 2015)</p>
SURF (Speeded Up Robust Features) local feature detector function	<p>SIFT and SURF are the two most powerful methods for feature extraction and matching, regardless of image distortions (rotation, scale, blur, contrast or illumination) (Çan, 2021; Mistry & Banerjee, 2017). It detects the maximum and minimum points with the quadratic derivative of the images for scale and position selection (Çan, 2021). Basically, Laplacian of Gaussian Filter calculation is used (Mistry & Banerjee, 2017).</p> <p>Although its working logic is similar to SIFT, it uses Laplacian of Gaussian Filter with Box-Filter algorithm for corner detection and uses vertical and horizontal wavelet responses as feature descriptors (Bansal, Kumar, & Kumar, 2021).</p> <p>It is superior to SIFT in being independent of rotation, hue, distortion, RGB noise and blur effects (Mistry & Banerjee, 2017). It can also work faster (three times more (Mistry & Banerjee, 2017)) than SIFT and with equivalent accuracy performance (Devi & Kirubakaran, 2015). There is no significant difference between SIFT and SURF in terms of lighting effect (Mistry & Banerjee, 2017). It performs both detector and descriptor operations (Bayraktar & Boyraz, 2017).</p> <p>The computational load required by SURF for real-time applications is too high to be performed on mobile devices with low computing capacity (Billinghurst, Clark, & Lee, 2015).</p> <p>The operation of the SURF algorithm is detailed in the subheading 2.2.2.2 SURF Algorithm.</p>

BRISK (Binary Robust Invariant Scalable Keypoints), ORB, SIFT and SURF consist of 4-basic steps that perform both **detector** ((1) Representation scale-space, (2) Locating key points, and (3) Determining orientation) and **descriptor** ((4) Key point identifiers) (Bayraktar & Boyraz, 2017; Oral & Gençal, 2016).

2.2.2.2 The SURF Algorithm

In algorithms that perform feature detection and description, the main goal is to determine the features in images regardless of changing conditions (rotation, scale, illumination, perspective or affine transformation) (IGI Global, 2022).

SURF is an algorithm that can **detect** (feature detector/extractor, is the process of extracting a set of regions or in other words local features from the input image (IGI Global, 2022)) and **describe** (feature descriptor/computed over the image region

defined by the feature extractor and represents intensity in terms of colour and the like (IGI Global, 2022)) points of interest (notably different from neighbouring points (IGI Global, 2022)) by operating independently of image distortions (scale and rotation).

The working steps of the SURF algorithm are given in Table 2.7.

Table 2.7 SURF algorithm steps and explanations (Mistry & Banerjee, 2017 Muzaffer, 2016)

SURF Algorithm Step	Explanation
Feature Detection	<p>Operations performed in the Feature Detection Step;</p> <ul style="list-style-type: none"> In the SURF algorithm, the input image is represented by its integral (Mistry & Banerjee, 2017). The integral image is created by calculating the integral values of all pixels in the input image (the sum of the intensity values of the pixels in the rectangular region between the (0, 0) point of the image and any (x, y) point (Mistry & Banerjee, 2017; Muzaffer, 2016). $I_{\Sigma}(x, y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(x, y)$ <ul style="list-style-type: none"> A quick calculation is performed by applying the convolution operation to the integral image with the box filter (approximate version of the Gaussian filter). $I(x, y)$ represents the coordinates of the image and σ represents the scale. (Mistry & Banerjee, 2017). $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$ <ul style="list-style-type: none"> The convolution operation is integrated into the used Hessian Matrix (which contains information on regional variations). In this way, the keypoints in the image are detected. (Muzaffer, 2016) $H(x, y, \sigma) = [L_{xx}(x, y, \sigma), L_{xy}(x, y, \sigma), L_{xy}(x, y, \sigma), L_{yy}(x, y, \sigma)]$ <p>$L_{xx}(x, y, \sigma)$: Convolution of the 2nd-order Gaussian derivative of the Laplacian of Gaussian point $I(x, y)$.</p> <ul style="list-style-type: none"> By using both integral and convolution operations (especially since the convolution operation reduces the computational complexity), the computation time is reduced (Muzaffer, 2016). While creating the scale-space structure, box filters of different sizes (starting with 9x9 and increasing to 15x15, 21x21 and 27x27) and scales are applied to the integral image, and therefore scale-space is formed in a pyramid structure (consists of scaled images that increase in size recursively) (box filters and integral images used in SURF prevent scale-space rendering using the same filter as in SIFT) (Muzaffer, 2016). Feature points are extracted according to Hessian determinants in the scale-space structure (Feature point is the one with the highest gradient value out of a total of 27 pixels in 3x3 regions selected on three consecutive scales) (Muzaffer, 2016).

Table 2.7 Continues

Feature Description	<p>Operations performed in the Feature Description Step;</p> <ul style="list-style-type: none"> • By using keypoints as center, 20s dimensional quadratic regions are created (assuming the scale is s). • The 20s region is divided into 5s subregions (4x4 quadratic regions) and a 2s dimensional Haar wavelet filter is applied to each subregion horizontally and vertically (obtaining dx and dy) to calculate derivatives in the x and y direction. • Gaussian weighting with σ of 3.3s is applied to Wavelet filter responses (dx and dy) with the keypoint in the center (to allow independence to position and geometric distortions). • The first part of the descriptive vector is formed by summing the dx and dy values obtained for each subregion. • The sum of the absolute values of dx and dy is calculated to add the polar density change information to the descriptive vector. • Since the descriptor vector is a four-dimensional $v=(\sum dx, \sum dy, \sum dx , \sum dy)$ for every 4x4 dimensional subregion, the descriptor vector is eventually created in 64 dimensions. (Muzaffer, 2016)
----------------------------	---

2.2.2.3 Literature Review of Digital Image Matching

In the vast majority of studies on image matching in the literature, the steps in Table 2.6 are used.

In the work of Bana and Kaur (2011) a fingerprint biometric recognition system was developed by using the Minutiae-based matching method, which is a fingerprint comparison algorithm. It was stated that the ridge (black parts) and furrow (white parts) shapes, which were similar in terms of parallelism and average width in the fingerprint, were not sufficient to establish similarity, so the recognition process with the third feature, Minutiae (abnormal spots on the ridges) was used. In this context, the ending and bifurcation parts of the ridges were used. In the study using the MATLAB® program was used, the grayscale fingerprint images were first improved with the **Minutiae Extraction step** ((1) Image segmentation (histogram equalization, FFT, image binarization), (2) Image enhancement (ridge flow estimation, ROI Extraction), (3) Final extraction (ridge thinning, minutiae marking)) and brought into binary format to obtain minutiae points, and then with the **Minutiae Matching step** ((4) Minutiae alignment (reference minutiae pair finding, minutiae sets transforming), (5) Match (match score finding)), the fingerprints were compared in pairs by using the minutiae points and a match score was calculated. The FAR and

FRR values of the test results obtained were found to be between 30-35% and the verification rate of the system was found to be between 65-70%. It was underlined that the relatively low percentage of verification of the proposed method was compared to other biometric methods is due to the fact that the algorithm used was not very robust (vulnerable to the effects of scaling and elastic deformations). It was also stated that pre-processing of poor quality fingerprint images also reduced the verification rate.

In Cömert's (2015) master's thesis, in which the palm vein image was discussed in terms of biometric classification, an embedded system prototype was developed that includes a palm vein map reading sensor and a portable hand unit. In this context, predecessor operations (ROI extraction, histogram equalization, thresholding, and skeletonization) were applied to the original dataset obtained with the system run on a mini embedded computer with Linux operating system by coding as open source modules with Python (high-level programming language). Then, the feature vectors were extracted and the classification was made. For the database of the study, 3 palm vein images of the right hand (320x240 pixel resolution, 8-bit black/white, saved in png format) were taken from 100 people. While 2 images of 75 people were used as training data, 1 image from 75 people and 3 images from 25 people (not used during training) were used as test data. In the trials of the system (24x24, 32x32, 64x64 and 128x128 pixel values were used), the highest level of accuracy was obtained at 64x64 resolution, and as the image resolution enhanced, the accuracy level decreased with the increase in noise. In addition, the system was classified with both minimum Euclidean distance (based on the approach that the less linear distance between two feature vectors, the more similar they are) and cosine similarity (based on the approach that the similarity of two vectors on a scale between 0 and 1 increases as the measurement value gets closer to 0) at 64x64 pixel resolutions, and it was determined that the results obtained from cosine similarity were found to be more effective in comparison. The False Rejection Rate of the system is 94 out of 1000, and the False Acceptance Rate is 0 out of 1000.

The system developed by Zhong, Kanhere and Chou (2017) within the scope of smartphone security consists of 3-stages; (1) Adaptive threshold filter and depth threshold (assuming the hand is closer to the camera than the background) were used to extract the vein patterns by eliminating the background in the IR image, then perspective changes were removed by rotating the pixels in 3D with the depth map, (2) The key points, which are reference points, were extracted by determining the vein endings and intersections, (3) The similarity score (using Kernel Distance and Hausdorff Distance) was determined more quickly by comparing the vein patterns, in the reference and test images using key points. According to the results of the study, it has been revealed that the VeinDeep system was superior to the Kernel distance (0.9) and Hausdorff distance (0.5) methods with 0.98 precision. Recall value was determined as 0.83.

In their study Shrotri, Rethrekar, Patil, and Kore (2010) aimed to develop a method for better feature extraction for low-quality dorsal hand vein images. In this context, two separate datasets (SAB11 and BIOM14 datasets) consisting of grayscale images were created for the dorsal hand vein and the developed method (which allows to obtain vein patterns from low quality images) was tested with these datasets. With the proposed method, 5 images were taken from each hand of 34 people and tested with a dataset of 3400 images in total. Steps used in terms of software, (1) **Data acquisition** (to determine the Region of extraction, the mean of each row/column in the image is taken and compared with the threshold value), (2) **Image enhancement** (double adaptive equalization is applied to increase the contrast), (3) **Feature extraction** (2D discrete wavelet transform was used for hand vein pattern extraction. The purpose of wavelet transform is data compression), and (4) **Classification/Matching** (feature vector based classification is applied on binary image. Euclidean distance is used to calculate the distance between two fixed size matrices representing binary images). In conclusion, it has been pointed out that the use of wavelet transform is useful for the extraction and manipulation of contours for feature extraction even on low quality hand vein images.

2.2.2.4 The Proposed System's Digital Image Post-Processing Phase

In the Imaging Technique Phase, the user is asked to take a video recording of the tissue region he/she wants to follow. Because when the classification process is performed with one or more images, the small differences between the images affect the accuracy result. These differences are due to the user moving the camera or the imaged area during imaging (camera shakes, zooming in/out or blurring due to inability to focus may occur). Such an imaging causes the NIR light source to illuminate the tissue at different angles. As a result, deficiencies in vascular integrity (interruptions/breaks) are encountered in some imaged regions. In addition, the number of images the user can capture in a single viewing period is also limited. However, since CNN is used in the Classification Phase of the study, a large number of images (hundreds, thousands) are needed during the training of the network. For these reasons, the user is requested to take a video recording during the Imaging Technique Phase in order to obtain a more complete vein image and to meet the need for a large number of images. While the recordings are being taken, the user keeps the imaged tissue region fixed on the vertical axis and slightly rotates it around its own axis on the horizontal axis, thus contributing to the illumination of the entire tissue by NIR photons coming from different angles. In the Digital Image Pre-Processing Phase, these video recordings are processed and NIR modified-images are obtained in which the vascular structures can be clearly selected. At this phase, it is aimed to obtain a more complete vascular structure by correcting the discontinuous structures that cannot be resolved despite hardware (the Imaging Technique Phase) and software (the Digital Image Pre-Processing Phase) improvements in the veins in the obtained images.

The Digital Image Post-Processing Phase of the proposed system was coded with the MATLAB® R2017a program. With the SURF local feature detector function based on the similarities (blob features) between the images, sequential video images were processed and the rotation and scale differences between the two images were determined and the images were fused.

The steps, MATLAB® commands and explanations of the proposed system's Digital Image Post-Processing Phase are detailed in Table 2.8.

Table 2.8 Steps and explanations of the proposed system's Digital Image Post-Processing Phase (DetectSURFeatures, 2022; ExtractFeatures, 2022)

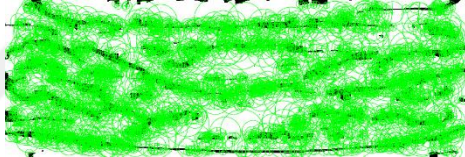
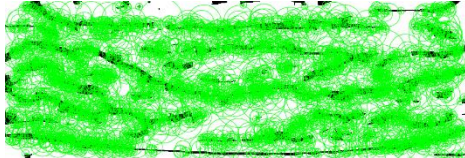
Digital Image Post-Processing Phase Step	Explanation
Feature Detection	<p>In the Feature Detection Step, the <i>detectSURFfeatures</i> which can detect SURF features (blob-shaped) in images command is used. When NIR grayscale images are taken as input, enough matching points can not be detected, therefore binary NIR modified-images (429x1003 logical) are used in this step. Skeletonization is generally used to represent vein patterns in the literature (Verma & Dubey, 2015). However, in this study, skeletonization is not used because sufficient number of matching points could not be determined when skeletonization was used.</p> <p>In this step, all consecutive NIR modified-images are subjected to <i>detectSURFfeatures</i> in pairs and the key (interest) points (as a SURF points object containing Scale, SignOfLaplacian, Orientation, Location, Metric and Count values) in both images are determined separately. Key points of the two sample images are outlined in the Figures below. (<i>detectSURFfeatures</i> command is used)</p>  <p>2674 points of interest (2674x1 SURFPoints) are detected for the NIR modified-image with the code Modified_Period1_Img0_RightArm.bmp</p>  <p>2721 points of interest (2721x1 SURFPoints) are detected for the NIR modified-image with the code Modified_Period1_Img1_RightArm.bmp</p> <p>In the Feature Description Step, the individual descriptors (also called feature vectors) and positions of key (interest) points in the two images are extracted. The descriptor is formed by pixels surrounding a point of interest. Descriptors of the two sample images are outlined in the Figures below. (<i>extractFeatures</i> command is used)</p>

Table 2.8 Continues

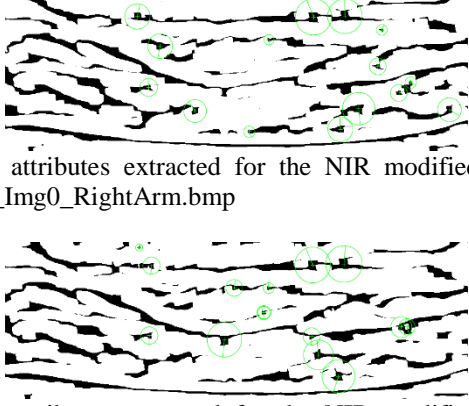

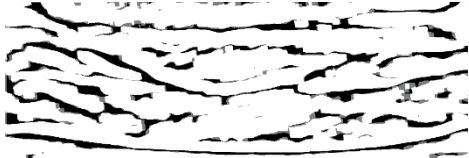
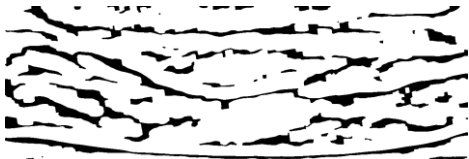
Feature Description	 <p>15 strongest valid attributes extracted for the NIR modified-image with code Modified_Period1_Img0_RightArm.bmp</p> <p>15 strongest valid attributes extracted for the NIR modified-image with code Modified_Period1_Img1_RightArm.bmp</p>
Feature Matching and Rotation/Scale Determination	<p>Matching features/key points (as a SURF points object containing Scale, SignOfLaplacian, Orientation, Location, Metric, and Count values) and locations are determined between the two consecutive images with different rotation and scale values. (<i>matchFeatures</i> command is used)</p> <p>In addition, the rotation and scale values between the matched point pairs detected in the NIR modified-images with Modified_Period1_Img0_RightArm.bmp and Modified_Period1_Img1_RightArm.bmp codes are determined as follows;</p> <ul style="list-style-type: none"> • Among 517 matching points (517x1 SURFPoints), 75 of them are inliner points (75x1 SURFPoints), • rotation value is -0.0264, • scale value is 0.9971. <p>The result image of the matching process of the two sample images is shown in the Figure below. (<i>estimateGeometricTransform</i> command is used)</p>  <p>The Image of matched inliner point pairs belonging to the NIR modified-images with codes Modified_Period1_Img0_RightArm.bmp and Modified_Period1_Img1_RightArm.bmp</p>
Image Fusion	<p>From the values detected between the matching point pairs;</p> <ul style="list-style-type: none"> • If the return value is in the range of [-0.07, 0.07] and, • The scale value is in the range of [0, 1.0], <p>the second image is transformed into the values of the first image and the two images are fused. The intervals are determined by trial-and-error. The NIR fused-modified image obtained as a result of the Digital Image Post-Processing Phase is converted to binary format (429x1003 logical). The result image of the two sample images fused by correcting the rotation and scale values and the binary form of this image to be used in the next phases of the proposed system are shown in the Figures below. (<i>imfuse</i> and <i>imbinarize</i> commands are used)</p>

Table 2.8 Continues

	 <p>The NIR fused-modified image with code Fused_Modified_Period1_Img0-1_RightArm.bmp</p>  <p>Binary version of the NIR fused-modified image with code Fused_Modified_Period1_Img0-1_RightArm.bmp</p>
--	---

Summary of the proposed system's Digital Image Post-Processing Phase in 1-minute is given in Figure 2.8.

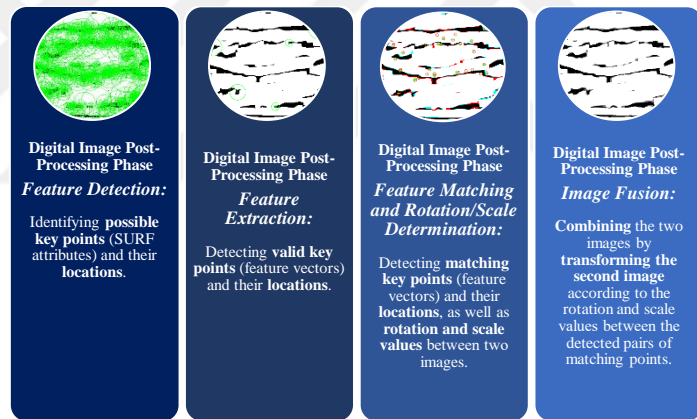


Figure 2.8 The proposed system's Digital Image Post-Processing Phase in 1-minute

2.2.3 The Classification Phase: Recognizing NIR Images

Digital data is the gold of our age. Just as today's naming of our world as a plastic planet (Gilbert, 2022) would be the most accurate description today, it would be best for someone who will visit a technological information processing device on a micro scale to define this world as a data world (rather than the world of cables or circuits). Especially information processing devices, whose dimensions are small enough to be carried in hand, represent different worlds consisting of data, thanks to their increasing data processing and storage capacities. Considering that NASA's Apollo

Guidance and Navigation System used to go to the Moon in Apollo missions uses a read-only fixed memory of 74 KB (Australian Broadcasting Corporation, 2022), today only the user/programmer's imagination will determine the limit of what can be done even with an ordinary smart device with 8 GB memory. These data worlds, when viewed from the eyes of a normal user, are an application where he shares holiday photos with his friends from social media or a storage area where he saves recipe videos, while from the eyes of a programmer, it is a data mine where the user can obtain all kinds of data from his personality profile to his habits, from places he frequents to his circle of friends. What matters is how the data will be extracted from this mine and for what purposes it will be used after processing.

In the Classification Phase, the tissue-specific features (the features that make the data different) in the first week's NIR fused-modified images were determined by deep learning techniques, and class predictions were made based on these features in order to determine whether the same tissue was viewed in the following weeks.

2.2.3.1 Deep Learning

The computer uses ordinary algorithms for ordinary operations. All operations such as addition, multiplication, mode calculation, integral calculation, graphing, word processing or playing music are uniform for the computer. Each process is run according to defined algorithms and the results are determinable without any guesswork or decision making.

The ability to make decisions in nature is a characteristic unique to living things. Every living thing can make a decision in the face of a new situation by using their old experiences. Decision takes place through the learning process using the acquired knowledge and requires intelligence. Well, for example, how can a carnivorous/herbivorous/omnivore distinction be made for an animal that is not defined in the entries of a computer application that classifies animals? The answer to this question might be the artificial creation of an intelligence that can imitate the human learning process, or more popularly, Artificial Intelligence.

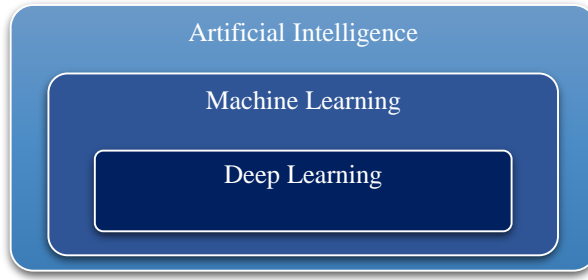


Figure 2.9 Artificial intelligence, machine learning and deep learning (Li, Deng, Zhu, Hua, & Tao, 2021; Violante, 2022)

Artificial Intelligence, the sub-concepts of which are stated in Figure 2.9, is the ability of computer-based systems to imitate abilities such as learning, decision making, and thinking performed by the brain in biological creatures (Kutlu, 2018; Kayaalp & Süzen, 2018; Yılmaz & Kaya, 2020). Imitating abilities is possible by modeling the biological brain (Yılmaz & Kaya, 2020). With the structure of **the Artificial Neural Network (ANN)**, the structure and functioning of the living brain are modeled within the scope of biomimicry, so that the learning process carried out by biological nervous system cells (neurons) in living things can be imitated by artificial neurons on the computer (Kayaalp & Süzen, 2018; Yılmaz & Kaya, 2020). Figure 2.10 shows a single-layer ANN structure consisting of a single artificial neuron, called the perceptron. The Single-Layer Perceptron (SLP) consists of the input neuron/layer (the layer where data enters the network) and the output neuron/layer (the last layer where the network's result data is located). In addition, in the calculations to be performed, weights (w) representing the importance of the input neuron, bias/deviation value (b), addition operation and activation function are used (I2tutorials, 2022; Javatpoint, 2022). In this structure, the output of a neuron (which is also the input of the next neuron (Hasan, 2019)) is calculated using the formula (1). (Kutlu, 2018)

$$y(t) = f(\sum_{i=1}^n (x_i(t) * w_i(t) + b)) \quad (1)$$

The data (electrical signals) received from the sense organs are converted into information in the human brain with a multi-layered biological neural network structure, each layer of which contains more than one biological neuron. Similarly, in computer systems that will perform complex tasks by creating artificial intelligence,

instead of a single artificial neuron/nerve cell, **multilayer ANNs** (the Multi-Layer Perceptron (MLP)) with multiple layers consisting of many artificial neurons are used. (Kutlu, 2018)

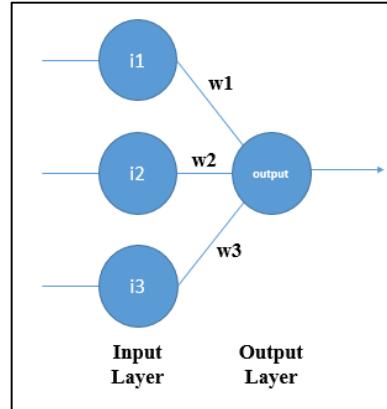


Figure 2.10 Single-layer perceptron (Mohammad, Alwada'n, & Al-Momani, 2016)

The MLP structures are created by connecting the SLPs, which can only perform linear calculations, in layers. MLP is composed of three basic layers as input layer, hidden layer (intermediate layers, layer or layers that apply certain operations to the data, network complexity increases as the number increases, more effective results are obtained) and output layer (Yılmaz & Kaya, 2020). With MLP, non-linear calculations (mathematical calculations such as logarithmic, trigonometric, exponential operations that can produce output values that are not proportional to the input) can be performed. A non-linear system will be no different from SLP.

In artificial intelligence, machine learning is used to perform the learning process (updating the weight and bias parameters until they reach the most appropriate value, taking into account the errors in the estimation (Demir, 2019; Kutlu, 2018)) (Kayaalp & Süzen, 2018).

Machine Learning is a sub-branch of artificial intelligence used within the scope of automatic pattern recognition from information (meaningful patterns) obtained using **data mining** (the process of extracting useful information and patterns from big data) from **big data** (a very large volume and variety of data obtained from various sources, that can be converted into information by special technological and analytical methods (De Mauro, Greco, & Grimaldi, 2016)), with rule sets or by

processes such as classification/clustering (Buyya, Calheiros, & Dastjerdi, 2016; Shalev-Shwartz & Ben-David, 2014). It is frequently used in fields such as astronomy (galaxy classification), medicine (pre-diagnosis by symptoms), engineering (airport passenger forecasting), and bioinformatics (drug development) (Shalev-Shwartz & Ben-David, 2014).

In machine learning 4-basic learning methods are used; (1) **Unsupervised** (training input data is unlabeled (classes are unclear). The model separates this data by categorizing it according to common points and relationships, and assigns the new queried data to the category it most closely resembles. For example, Clustering and Association), (2) **Supervised** (training input data is labeled (classes are clear). The model assigns the new queried data to the relevant category by using these labels. For example, Classification and Regression), (3) **Semi-supervised** (if the training input data contains many unlabeled data and few labeled data, unlabeled data is tried to be determined by using labeled data in model training), and (4) **Reinforcement learning** (agents are used as purpose-oriented in training with the logic of reward and punishment). (Hosseini, Lu, Kamaraj, Slowikowski, & Venkatesh, 2020; Io, 2022)

The problem of requiring human intervention (expert support and preprocessing) to learn from raw data in ML has been overcome with Deep Learning, which can perform self-learning from raw data (by creating features in different layers), as shown in Figure 2.11 (Kayaalp & Süzen, 2018; Kutlu, 2018).

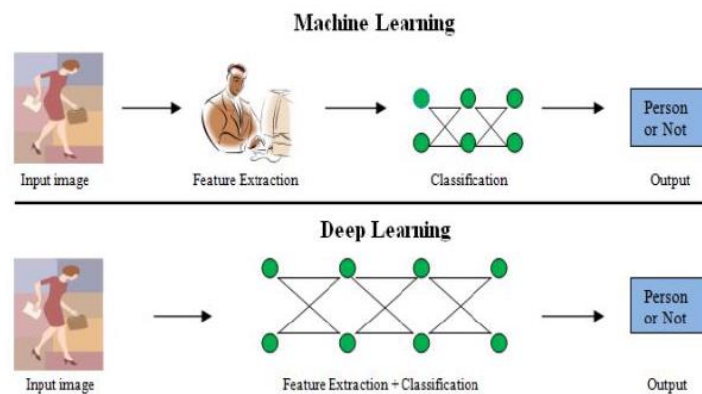


Figure 2.11 Stages of Machine Learning and Deep Learning (Gupta, Pathak, & Kumar, 2022)

Deep Learning is a sub-branch of machine learning that can perform classification by self-extracting features which are difficult for humans to detect, without the need for human intervention (unlike traditional machine learning methods such as classical ANN or SVM (Demir, 2019)), as a result of various learning processes that are not dependent on rule sets from labeled big data (Demir, 2019; Kayaalp & Süzen, 2018; Tükel, 2019). DL is called “deep” because it makes use of multi-layer ANNs with a large number of layers (hundreds) (Tükel, 2019). Low-level features are obtained in the lower layers, while deeper features are obtained in the higher-level layers. It is aimed that the success rate (correct determination of the class) obtained as a result of the DL process will reach a certain level (Kayaalp & Süzen, 2018). The larger the dataset and the more compatible the model used, the higher the prediction accuracy will be. (Tükel, 2019; Yılmaz & Kaya, 2020)

Traditional vein recognition systems consist of 4-steps as; (1) vein image acquisition, (2) image pre-processing, (3) feature extraction, and (4) image matching/recognition. However, in DL algorithms (for instance CNN) which is an advanced approach of ANN, on the other hand, is often preferred in recognition, detection, and classification applications, since education and learning can be performed directly without the need for an additional process involving human intervention. In addition, DL increases efficiency in terms of time and workload as it performs feature extraction automatically. One of the biggest advantages of deep learning methods is that it learns discriminative features from visual raw data (such as image, video) and is the most successful and flexible method in performing classification by matching the output according to the desired results (class). (Doğan & Türkoğlu, 2018; Karaali & Eminağaoğlu, 2021; Liu, Li, Sun, Zhang, & Chen, 2017; Reasat & Shahnaz, 2017).

The main DL architectures created specifically for various problems are; Deep Neural Networks, Deep Auto-Encoders, Restricted Boltzmann Machines, Deep Belief Networks, Fully Connected ANN (Dense Neural Network), Long/Short-Term

Memory with Recurrent Neural Networks (RNN), CNNs, and Hybrid CNN-LSTM (Kayaalp & Süzen, 2018; Kutlu, 2018; Memiş, 2019; Yılmaz & Kaya, 2020).

In the field of health, various heart diseases (coronary artery disease, arrhythmia or interior myocardial infarction) can be classified with CNN Models. Myocardial infarction (popularly known as heart attack), which is one of the main causes of death worldwide and whose survival rates increase when treatment is started within the first one or a few hours, can also be detected by changes in electrocardiography signals. Acute myocardial infarction detections, which cardiologists perform manually at the rates of 91% (sensitivity) and 51% (specificity) by looking at these changes, can be increased automatically with computer aided systems (CNN Models that process electrocardiography signals), these rates can be increased and support can be provided to the doctor in making a decision. (Reasat & Shahnaz, 2017)

2.2.3.2 Convolutional Neural Networks

CNN is a feedforward (**Forward Propagation** is the one-way and forward execution of calculations from the input layer to the output layer through hidden layers (Yılmaz & Kaya, 2020)) ANN structure consisting of neurons with learnable weights and biases. A classical CNN is divided into two parts as **Feature Extraction Stage** (the trainable part where CNN layers are used (İnik & Ülker, 2017)) and **Classification Stage** (the part containing fully-connected layers used in traditional ANNs) as shown in Figure 2.12. Convolution is the feature extraction process from the input data (2D matrix) using filters (Kutlu, 2018). Without the need for a manual intervention, it can learn the features in the data that are difficult to detect by humans (using filters) on its own. In CNN, the image is divided into small pieces and the similarity of these pieces to each other is checked. The accuracy of the classification result is highly correlated with the number of training data applied to the algorithm. The main goal in traditional CNN is to make classification with the highest accuracy (closest to actual values (Yılmaz & Kaya, 2020)) by minimizing the error (the loss function, which expresses how much the estimation differs from the expected value,

is determined by minimizing the cost/loss function) in learning (Tükel, 2019). (Demir, 2019)

The error in the obtained result (the difference between the desired result) is transferred to all weights with the back propagation (**Backward Propagation**; the network processes forward with randomly assigned weight and bias values in the first epoch (the number of iterations/repetition/rounds to be used while training the network) and the error value in the output layer is calculated with the loss function (minimum cost point is determined with the slope descent algorithm to minimize the cost). As the training progresses, the weights are updated by back propagation to the hidden layers using the slope of the cost function so that the error value can be minimized. (Yılmaz & Kaya, 2020)) algorithm, and the error is reduced by updating the weights (İnik & Ülker, 2017).

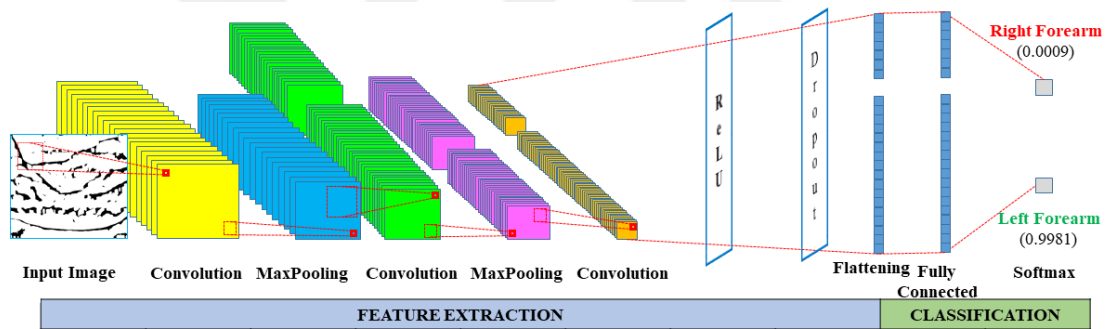


Figure 2.12 Architectural layers of a classical convolutional neural network model

The CNN Model architecture detailed in Table 2.9 can be summarized in 4-layers; (1) **Convolution layer** (obtaining feature maps and patterns of the input with filters), (2) **Pooling layer** (recognition of subsets of patterns representing different kinds of input), (3) **Flattening layer** (reducing multi-dimensional features to one-dimensional (1D)), (4) **Fully-connected layer** (classification of 1D inputs in one or more hidden layers working with MLP logic, by making an estimation with the softmax function (providing multi-classification with an output value in the range [0-1]) in the last layer, and updating the weights by back propagation in case of error) and updating the weights by backpropagation in case of error). In addition to these layers, the

activation function (Rectified Linear Unit (ReLU)) can be used in the hidden layers of the convolution layer and fully connected layer. (Karaali & Eminağaoğlu, 2021)

Table 2.9 Architectural layers and explanations of a classical convolutional neural network model (Bhaskar, 2022; Demir, 2019; Doğan & Pramoditha, 2022; Doğan & Türkoğlu, 2018; İnik & Ülker, 2017; Rubik's Code, 2022; Tükel, 2019; Karaali & Eminağaoğlu, 2021; Kutlu, 2018; Sharma, 2022; Yılmaz & Kaya, 2020)

CNN Model Architecture Layer	Explanation
Input Layer	The size of the input data (data such as image, video or audio, usually unprocessed or minimally processed (Yılmaz & Kaya, 2020)) to be applied to the first layer, the Input Layer, is highly decisive on the duration and success of the training. As the size increases, the network will become deeper and more features will be extracted, although this can result in long test times, but also increase performance. (İnik & Ülker, 2017).
Convolution Layer	<p>The Convolution Layer is the base layer of the model. On the input data (image) expressed as a matrix, the image is filtered (by performing the dot product of two matrices representing the input data and the filter (Demir, 2019)) by shifting the filter matrix (also called kernel, the filter size, also called window size, which specifies the width of the area to focus on in the input matrix, can be selected in various sizes such as 3x3, 5x5, 7x7 or 9x9) according to a certain stride value (the number of units to which the filter will be shifted horizontally or vertically on the input matrix, as the number of steps increases, the output matrix size will decrease), and then different feature maps (also called an activation map and shows which regions of the data are important) are created by padding (preserving the initial dimensions by adding preferably zero to the edge values of the reduced dimensions in the output matrix so that the output matrix and the input matrix are the same size). In the convolution process, as the filter moves over the image (horizontal or vertical), the results are collected by multiplying the coefficient values in the filter with the pixel values in the relevant region of the input image at each step (Tükel, 2019). When there is a similarity between the input image and the filter, the dot product result is higher, in other words, the features are detected by the filters (Demir, 2019). The filters are the neurons of the ANN, and the values in the filter are the weights of the neurons. Filter coefficients are updated during training. The large number of convolution layers is directly proportional to the ability to learn more complex features. As a result of the convolution operation, some information in the input matrix will be lost and if the padding is not applied, the size of the output feature matrix will decrease. The size of the output feature matrix can be calculated with the formula (2) (Kutlu, 2018). (Demir, 2019; Doğan & Türkoğlu, 2018; İnik & Ülker, 2017; Karaali & Eminağaoğlu, 2021; Kutlu, 2018; Rubik's Code, 2022)</p> <p>Size of output feature matrix = $(((\text{Input matrix size}) - (\text{Filter matrix size}) + (2 * \text{Padding})) / \text{Stride}) + 1 \quad (2)$</p>

Table 2.9 Continues

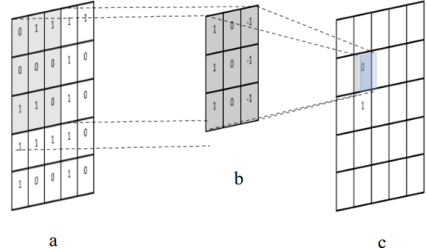
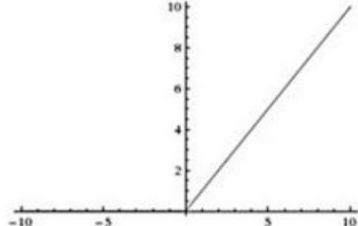
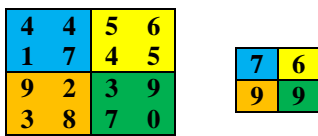
	 <p>Convolution operation. a) Image matrix b) Filter 3x3 c) New image matrix formed as a result of filtering (Doğan & Pramoditha, 2022; Türkoğlu, 2018)</p>
Activation Function	<p>Activation Function resets the negative values in the output feature matrix from the previous layer (the convolution layer that makes the network linear with the mathematical operations in the formula (2)) and makes the model non-linear (İnik & Ülker, 2017). ReLU, Sigmoid, Softmax, Hyperbolic Tangent (tanh), Sine or Step functions can be applied to the input of this layer (Doğan & Türkoğlu, 2018). ReLU (the x value in the $f_{ReLU}(x)=\text{maximum}(0, x)$ function represents the input value of the node) is one of the most used activation functions in CNN because it is not a linear function, can work fast for forward propagation and back propagation, and its first-order derivative can be expressed in terms of itself. (Karaali & Eminağaoğlu, 2021; Tükel, 2019).</p>  <p>ReLU activation function (Sharma, 2022)</p>
Pooling Layer	<p>An empty square matrix is shifted over the output feature matrix (calculates the maximum/maxpooling, minimum/minpooling or average/avgpooling of the values in the corresponding matrix section) from the previous layer (usually convolution layer or ReLU layer if applicable) according to a certain step value. The aim is to reduce the input size for the next convolution layer without affecting the depth, (downsampling, by preserving the pixel parts that contain the most important information (Bhaskar, 2022)) and to enable different variants of the image to be recognized. The resulting information loss is effective both in reducing the processing load on the network and in preventing overfitting (memorization). The output is the smaller sized (reduces the amount of data to process) input feature matrix. If the same dimensional output is desired, the input matrix dimensions can be preserved by padding. (Doğan & Türkoğlu, 2018; İnik & Ülker, 2017; Karaali & Eminağaoğlu, 2021; Kutlu, 2018; Rubik's Code, 2022; Tükel, 2019)</p>  <p>Pooling process (maxpooling)</p>

Table 2.9 Continues

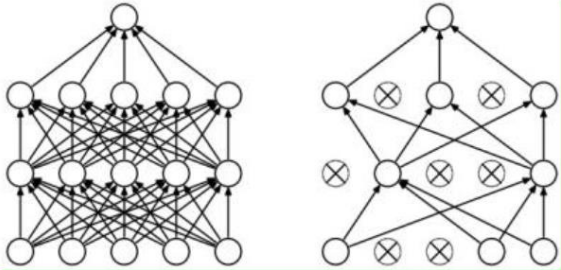
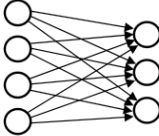
Dropout Layer	<p>In order to prevent overfitting (memorization) in the model, some nodes are randomly removed in this layer and the network performance is increased. It is used only during training, not during validation (called as pre-test in this thesis) and testing (called as post-test in this thesis). (Doğan & Türkoğlu, 2018; İnik & Ülker, 2017)</p>  <p>Dropout process. a) Fully-connected neural network b) Dropout neural network (Doğan & Türkoğlu, 2018)</p>
Flattening Layer	<p>The data to be the input to the Fully-Connected layer (which accepts only 1D data input) is prepared. The feature maps in the form of matrix (multi-dimensional array) in the previous layers are converted into a 1D array (vector) in this layer. Karaali & Eminağaoğlu, 2021; Rubik's Code, 2022)</p>
Fully-Connected Layer	<p>This layer, which is used for classification after all layers, is in the traditional multi layer neural network/MLP structure consisting of one or more hidden layers. Several FC layers can be used (Tükel, 2019). It uses feature maps as input, which are converted to 1D form in the flattening layer. In the model, all neurons in this layer have a direct connection (fully-connected) to all neurons in the next layer. Some features that cannot be detected in the convolution layer can be detected in this layer (Tükel, 2019). In this layer, using the feature maps, classification is performed according to the classes in the training dataset (by converting the data from the previous layer to the number of defined classes (Tükel, 2019)) and if there is an error, the filter weights are updated by back propagation. (Bhaskar, 2022; Doğan & Türkoğlu, 2018; İnik & Ülker, 2017; Karaali & Eminağaoğlu, 2021)</p>  <p>Fully-connected neuron layer (Doğan & Türkoğlu, 2018)</p>
Normalization Layer	<p>During training, the output values of neurons increase exponentially with operations such as addition and multiplication (can be very large or very small values), increasing the computational load in the next layers. To prevent this, the performance of the network is increased by normalizing the input data to a certain range. (Doğan & Türkoğlu, 2018; Kutlu, 2018)</p>

Table 2.9 Continues

Output Layer	<p>Probabilistic class estimation is performed. There are as many output values as the number of defined classes (so that their sum will be 1). The softmax function (which is a non-linear function) is preferred as the output layer activation function in (multiple) classification operations involving two or more classes. The belonging of the queried data to all defined classes is indicated by a prediction value (a probabilistic closeness/classification probability value calculated by cross entropy) normalized to the interval [0, 1]. (Demir, 2019; Doğan & Türkoğlu, 2018; Karaali & Eminağaoğlu, 2021; Tükel, 2019)</p> <p>(a) (b)</p> <p>Output Layer functions. (a) Sigmoid, and (b) Softmax (Kaya, Tuncer, & Baran, 2022)</p>
---------------------	--

LeNet (the first CNN Model), and the models developed afterward (competitions were organized) are summarized in Table 2.10.

Table 2.10 Architectural features of popular convolutional neural network models (Doğan & Türkoğlu, 2018; Fu & Rui, 2017; İnik & Ülker, 2017; Yılmaz & Kaya, 2020)

Deep Learning Model	Version	Network Depth/Total Number of Layers	Input Data Type and Size	Layers	Size
LeNet	N/A	5	32x32 grayscale image. Digit classification.	Input, 21 Conv + 1 FC layers Maxpool, FC, ReLU, Output (softmax)	60 K
AlexNet	N/A	8/25	Image (227x227x3)	Input, 5 Conv, 3 MaxPool, 2 DropOut, 3 FC, 7 ReLU, 2 Normalization, Output (softmax, with 1000 classes).	60 M
ZfNet	N/A	N/A / 19	224x224		60 M
GoogLeNet	N/A	22/144	Image (224x224x3)	Input, 21 Conv + 1 FC layers Maxpool, ReLU, Output (softmax) Batch Normalization, image degradation and RMSprop optimization	4M

Table 2.10 Continues

VGG	16	16/41	Image (224x224x3)	13 conv + 3 FC layers Maxpool, Dropout, ReLU, Output (softmax)	138 M
	19	19/47	Image (224x224x3)	16 conv + 3 FC layers Maxpool, Dropout, ReLU, Output (softmax)	N/A
ResNet	50	50/177	Image (224x224x3)	49 Conv + 1 FC layers	N/A
	110	110 / N/A	N/A	N/A	1.7 M
	152	152 / N/A	N/A	151 conv + 1 FC layers	25.6 M

A large amount of data is needed to train CNN Models. The more examples the dataset used in the model contains, the higher the classification accuracy of the model will be. The most popular datasets with open-source image content used in CNN training are listed in Table 2.11.

Table 2.11 Popular datasets and features (Dar, 2022; Tan, 2019)

Dataset Name	Number of Data	Number of Classes	Size	Explanation
CIFAR-10	50K (train-data) 10K (test-data)	10	170 MB	CIFAR-100 version includes 32x32 images, too. However, it contains 100 classes. (Tan, 2019)
ImageNet	1.5M	1000	150 GB	N/A
MNIST (Modified National Institute of Standards and Technology)	60K (train-data) 10K (test-data)	10 (all digits from 0 to 9)	50 MB	Contains 28x28 size images of handwritten numbers. The numbers are represented in black and the background in white.
MS-COCO (Microsoft Common Objects in Context)	330K (200K of them are labelled)	80	25 GB	N/A
Open Images Dataset	9M (5K of them are labelled) 125K (train-data) 40K (validation-data)	N/A	500 GB	N/A

In CNN Models, model performance increases in direct proportion to the number of data (sample) used. In order to achieve a good classification accuracy level with these models, a lot of data (hundreds, thousands) is needed during the training phase.

Although the publicly accessible datasets given in Table 2.11 are sufficient for most general-purpose applications, for some applications (disease diagnosis, biometric recognition, etc.) it is necessary to create special datasets containing images with specific features and content. In this context, either a large number of application-specific data (hundreds, thousands) will have to be collected or the number of data will be increased to a certain extent with DIP techniques, in a way that preserves the meaning of the image, with special methods (data augmentation techniques) specified in Table 2.12. However, while increasing the number of data will increase the correct classification performance, it will also increase the training time required for the operations (in order to overcome this situation, it is possible to complete the trainings in a very short time by preferring the use of Graphics Processing Unit (GPU) instead of a training that takes days or weeks with the CPU) (Doğan & Türkoğlu, 2018).

Table 2.12 Data augmentation techniques and explanations



Data Augmentation Process	Explanation
Raw-image	<p>A raw-image is an image that has not undergone any DIP.</p>  <p>Raw-image used as a representative (image source: (Daily Sabah, 2022))</p>
Rotation By Angle/Mirroring	<p>Rotation by Angle is the rotation of the image with a small angle value and/or relative to a certain axis in a way that preserves its meaning. Mirroring is the creation of a mirror image by making a 180-degree rotation on the x or y-axis, preserving the meaning of the image. Both rotating by angle and mirroring operations can be used at the same time.</p>  <p>Angle rotation and mirroring applied image</p>

Table 2.12 Continues



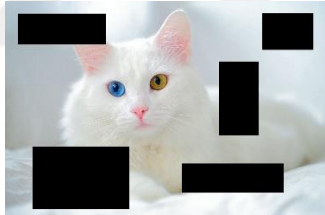
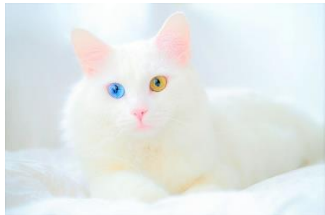

Noise Addition	<p>Adding certain types of noise (salt and pepper, etc.) to the image to allow for quality changes.</p>  <p>Noise (salt and pepper type) applied image</p>
Scale Change	<p>Zooming in (focusing on the details, randomly cropping a certain part of the image and discarding the other parts also called Random Cropping)/Zooming out (focusing on the overall image) according to a certain scale value.</p>  <p>Scale change (zoom in) applied image</p>
Information Loss	<p>Deletion of information (image part) in some random parts of the image.</p>  <p>Information loss applied image</p>
Contrast Change	<p>Implementation of changes that can be created by daylight (day/night or weather condition clear/cloudy, etc.).</p>  <p>Contrast change applied image</p>

Table 2.12 Continues

Colour Change	Changing the colour components (RGB) of the image.  Colour change (green) applied image
----------------------	---

In order to include CNN Models in the study within the scope of classification processes, two basic methods are used in the literature; (1) **Building a CNN Model architecture from scratch** or (2) **Transfer learning**.

(1) **Building a CNN Model architecture from scratch:** By changing the layers (convolution, pooling) and their order, the number of filters and their sizes used in the CNN Model, very different models can be created (as the convolution layers increase, the features that the filters can extract increase, but the training time gets longer) (Demir, 2019). To build a CNN from scratch, first of all, at least 2 convolution and activation (preferably ReLU) layers (if a single convolution layer is used, the desired accuracy rate may not be achieved in the model since sufficient number of feature maps cannot be created, if too many layers are added, the training time may increase while the accuracy increases, or the weight update by back propagation may not reach the first layers with the desired efficiency (Demir, 2019; Tükel, 2019)) are added in succession. Subsequently, the pooling layer can be added to this duo and repeated several times in the form of triple combinations. Also, a flattening layer must be present before the traditional neural network. The model should be finalized with the addition of the fully-connected layer (ReLU can be used afterwards) and the output layer (the softmax function should be used for a multiple classification). Operations can be briefly listed as *Input Image* \rightarrow *[[Convolution \rightarrow ReLU]*n \rightarrow Pooling?]*n \rightarrow Flattening \rightarrow [Fully-Connected \rightarrow ReLU]*k \rightarrow Fully-Connected* (where n, m, and k are the number of repetitions). (Rubik's Code, 2022)

After the model architecture is designed, changes can be applied according to the purpose (by trial-and-error method if necessary) to increase the classification accuracy (Demir, 2019);

- Considering the performance effects of the CNN layers given in Table 2.13, new layers can be added to and removed from the model as application-specific,
- Considering the effects of the training hyperparameters given in Table 2.14 on the performance, the way of training the network can be adjusted as application-specific or,
- The size of the dataset can be increased with the data augmentation techniques given in Table 2.12 (the number of data is directly proportional to the classification accuracy).

Table 2.13 Architectural layers of a classical convolutional neural network model and their effects on performance (İnik & Ülker, 2017; Karaali & Eminağaoğlu, 2021; Kutlu, 2018; Rubik's Code, 2022; Tükel, 2019; Yılmaz & Kaya, 2020)

Layer	Main Purpose of Use	Function	Effect on Performance
Input Layer	Data input	Transmission of raw or pre-processed input data to the network.	The size of the input data is directly proportional to the memory requirement, training time and performance.
Convolution Layer	Feature extraction	Generating feature maps (containing various high- and low-level features found in the input data) using filters.	Parameters such as filter size, stride, and padding change the size of the output features map, thus directly related to the processing load and thus the memory usage.
Activation Function	Feature extraction	Making the network non-linear (zero assignment to negative values preferably with ReLU).	It allows the network to perform learning faster.
Pooling Layer	Feature extraction	Reducing the size of the output feature maps without changing the depth (resulting in information loss).	Reducing the size of the output features maps by preserving important parts reduces the processing load. It is effective in preventing overfitting and increasing network performance by reducing the number of parameters.

Table 2.13 Continues

Dropout Layer	Feature extraction	It allows the network to have different combinations of neuron connections by randomly removing some nodes in the hidden layers.	It is effective in preventing overfitting (memorization) and increasing network performance.
Flattening Layer	Classification pre-processing	Converting feature maps to vector.	Preparation of data for the fully-connected layer.
Fully-Connected Layer	Classification	Classification is performed with hidden layer neurons.	If there is an error or the desired success criterion is not met (number of epoch or accuracy), the filter weights are updated by backpropagation.
Output Layer	Classification results	The belonging of the data to all defined classes is determined to make the sum 1.	Obtaining probabilistic estimation results is provided.

Table 2.14 Training hyperparameters and performance effects of a classical convolutional neural network model (Tükel, 2019; Yılmaz & Kaya, 2020)

Parameter	Main Purpose of Use	Properties	Effect on Performance
Activation Function	Converting linear values to non-linear values.	ReLU, tanh, Sigmoid, Softmax, Swish, Step functions are used.	It allows the network to perform learning faster.
Mini-batch Size	Processing the input data not all at once, but by dividing it into independent smaller sub-batches (mini-batches).	If mini-batch = 1, only one data is processed in each iteration. If mini-batch = maximum, all data is processed simultaneously.	Although the use of mini-batches increases the likelihood of misclassification, it reduces training time.
Batch Normalization	Ensuring that the data distribution between the layers is the same.	The data is normalized and sent to the next layer.	N/A
Number of Epochs	It is the number of iteration/repetition/cycle/round to be used to minimize the cost function while the network is being trained.	In the first epoch, the network is trained with the first mini-batch, the classification accuracy is tested, the weights are updated by backpropagation if necessary (if there is an error), and the corresponding epoch is terminated. These steps are repeated with other mini-batches until optimal weight values are reached in the second and subsequent epochs.	While the performance is low in the first epochs, the performance will increase as the training continues. The training of the network can be considered complete when the performance rate increase becomes too slow. If the number of epochs is large, overfitting can also be observed while the accuracy values of the mesh increase.

Table 2.14 Continues

Learning Rate	It is a coefficient used during parameter update by back propagation.	It can be selected as a fixed value or a progressively increasing/decreasing value.	If the learning rate is chosen too large, the dependency on the training data will increase, while if it is chosen too small, the training time will increase.
Momentum Coefficient	The new value formed after the noise generating operations is included in the ratio of the momentum coefficient of the previous value.	It is usually chosen as 0.9.	N/A
Optimization Methods	Solving nonlinear problems with optimum results. It is used to determine the most suitable parameters between the actual outputs and the estimated ones.	Stochastic Gradient Descent, Adam, Adamax optimization algorithms can be used.	The main goal in deep learning is to minimize the error. The lower the error, the more accurate the prediction will be. The network that produces the most optimized outputs is the network with the loss function minimized.

(2) Transfer learning. It is the realization of new class predictions by retraining the information (filter weights) obtained from an CNN Model trained using specific images for a specific purpose, with a different image dataset (by extracting new high-level features from the fully-connected layers of the pre-trained CNN Model) for another purpose related to the first problem. In some of the studies using the CNN Model, pre-trained models (low-level features extraction such as blob and edge is performed with the first layers, while task-specific higher-level features to be used in classification are extracted with the last/deeper fully-connected layers) may be preferred instead of building a Model from scratch (Al-johania & Elrefaei, 2019). Transfer learning is likened to human learning method in the scope of using previously-learned information for new inferences. (Al-johania & Elrefaei, 2019; Fırıldak & Talu, 2019; Tükel, 2019)

2.2.3.3 Literature Review of the Classification

Most of the studies on classification of vascular patterns in the literature have been carried out for biometric purposes.

Biometrics is the science of identifying a person according to their unique (personal) physical and behavioral characteristics and is a necessity in today's modern society, especially for secure applications based on the recognition of people. (Elnasir & Shamsuddin, 2014; Lingyu & Leedham, 2006; Shrotri, Rethrekar, Patil, & Kore, 2010). Biometric characteristics are divided into two as **physiological** (hand, finger geometry, vein, palm skin fold marks/vascular network patterns, fingerprint, face, iris, retinal vasculature, DNA, odor) and **behavioral** (voice, signature, gait, hand gestures, lip movements, keypress styles) characteristics and offer a good alternative to traditional methods that provide security with password (Al-johania & Elrefaei, 2019; Cömert, 2015; Lee, Jung, & Kim, 2011; Shrotri, Rethrekar, Patil, & Kore, 2010). By using these unique biometric characteristics, individuals can be recognized/verified (to develop a strong recognition system, a good classification should be made based on feature extraction) and granted access for certain transactions (personal data, banking transactions, hospital records) (Benziane & Benyettou, 2015).

Since studies in embryology show that both deterministic and random factors are effective in the formation of vascular tissue, it is stated that vessel patterns can be used as individual biometric data (Bazrafkan, Nedelcu, Costache, & Corcoran, 2016). Another reason for using vascular patterns (individual-specific reticulate structure formed by blood vessels under the skin (Shrotri, Rethrekar, Patil, & Kore, 2010)) within the scope of biometric systems is that these patterns do not change over the years, except in extreme cases (surgical intervention, burns) (Al-johania & Elrefaei, 2019; Lingyu & Leedham, 2006; Shrotri, Rethrekar, Patil, & Kore, 2010). Vein patterns are different in each individual, even the vein patterns on both hands of an individual are different from each other (Liu, Li, Sun, Zhang, & Chen, 2017). In addition, these structures are frequently preferred within the scope of secure applications as they do not have security weaknesses such as fingerprints leaving traces on surfaces or recording the human face in public, as they do not leave traces and are not easily visible from the outside due to their presence under the skin (Lingyu & Leedham, 2006; Zhong, Kanhere, & Chou, 2017). The fact that the

vascular patterns are unchanging and unique, and that they are also very difficult to copy, form the basis for their use as biometrics (Lingyu & Leedham, 2006).

A vascular biometrics-based recognition system can be performed in 4-basic steps: (1) **Data collection/NIR Imaging** (taking advantage of the reaction of the hemoglobin molecule in the blood inside the veins against NIR light (hemoglobin can absorb NIR light in the wavelength range of 760-1100 nm)), (2) **Image enhancement** (obtaining specific data with DIP techniques from raw data), (3) **Feature extraction** (*the first method*, increasing the contrast of pixels in vascular network images (histogram equalization/normalization), separation of foreground veins and other irrelevant tissues in the background by thresholding, and application of skeletonization based on geometric feature extraction by thinning. *The second method* is to obtain the template features with wavelet transform, Gabor filter and cosine transform instead of the previous operations), and (4) **Classification/Matching** (comparing the similarities in the vascular structures registered and queried in the database with correlation, vector distance calculation or ANNs by mapping the obtained features). (Benziane & Benyettou, 2015; Cömert, 2015)

As there is currently no publicly available vascular pattern dataset (for example, containing dorsal hand veins), researchers must first generate their own datasets to test NIR imaging systems. Benziane & Benyettou, 2015; Lingyu & Leedham, 2006; Shrotri, Rethrekar, Patil, & Kore, 2010).

In the biometric system developed by Lingyu and Leedham (2006), the similarity of the registered image and the test image was verified with the similarity measure obtained with Hausdorff Distance (the smaller the more similar). In the study, where it is stated that there is no public hand vein pattern database at present, FIR and NIR vein image database was created. Within the scope of the trials, hand tissue (hand dorsum, palm and wrist) images of participants of different ethnic origins (Indian, Chinese, Caucasian, Thai and Vietnamese) were taken with FIR and NIR. The FIR trials were carried out indoors with a temperature of 10⁰C and outdoors at a

temperature of 30-34°C (humidity is less than 80%). The first dataset, which was prepared by taking 9 images from the right and left upper tissues of 30 volunteers with the FIR, includes 540 images (320x240 resolution greyscale image files with bmp extension). The second dataset, which was prepared by taking a total of 9 images from the right and left hand dorsum, palm and wrist, from 150 volunteers with NIR, includes 2700 images (644x492 resolution grayscale image files with tiff extension). According to the results of the experiment, with FIR, only the main vein patterns on the dorsum of the hand, which were taken indoors, could be visualized in such a way that meaningful data could be obtained (as it does not contain details, it is a problem to use as biometric data in applications that require high security), while with NIR, both the main and small veins on the palm and wrist could be visualized (providing data that can highlight details and differences within the scope of biometrics) independently of the ambient conditions. Both the false acceptance rate and false rejection rate values obtained in the trials (with left hand FIR images, 180 of which is for train and 90 as test) were found to be 0%. However, it was stated that these values will increase as the number of volunteers increases, and therefore it is more important to use palm images. As another solution to increase the system performance, it was underlined that a hybrid system could be designed in which hand dorsum, palm and wrist images could be evaluated simultaneously.

In the studies of Elnasir and Shamsuddin (2014), it is aimed to improve the feature extraction in palm vein recognition systems. In this context, a palm vein recognition system based on two different methods, 2D Discrete Wavelet Transform (it separates images into high frequency components representing noise and low frequency components representing the part to be used in the system, this method was used first to avoid overfitting in Linear Discrimination Analysis) and Linear Discrimination Analysis (to make feature extraction more reliable, it keeps the inter-class distribution the largest and the within-class distribution the smallest), is proposed in the study. The MATLAB® 2012a program used in the proposed system was run on an i5 core 3.10 GHz Intel Core CPU and 4 GB RAM. The middle parts of the images were cropped from the other parts, and an ROI of 128x128 was obtained. In order to prepare the image dataset, firstly, the noise in the images was separated

using 2D Discrete Wavelet Transform (image quality is improved by using only low-frequency parts of the images that do not contain noise), the train (4 images) and test (6 images) data were determined, the features required for classification were extracted using LDA and the classification process (performed with nearest neighbour using cosine distance) was performed (binary vein images in black and white were not used in this study). In the system, which works according to two different modes, decision-making is performed with 1 to 1 comparison in the verification mode (if the similarity distance between the live image to be compared and the recorded image is less than the specified threshold, the tested image is considered to belong to a registered user), while 1 to many comparisons are made in the comparison mode (all images saved in the database of the user with the highest similarity value for the queried image are compared). According to the test results, the system was found to be successful in terms of identification rate (99.74%) and verification rate (100%).

In the finger recognition system proposed by Lee, Jung and Kim (2011) binary feature extraction and matching of images with Hamming Distance (which measures the differences between binary patterns) were applied. For the trials of the proposed system, a total of 2400 finger images were obtained with 8 finger images (not taken because the thumbs are short) taken 10 times from 30 people. According to the results of the study, it was stated that recognition accuracy could be increased (as both finger vein pattern and finger geometry are used), processing time could be reduced (using the entire image of the finger instead of using ROI), finger vein and finger geometry components could be enhanced better (according to Gabor filter) with the modified Gaussian high pass filter used, and a good recognition performance could be achieved with the adequate finger image capturing scheme used (without using any finger alignment algorithm). In addition, according to the experimental results, the system has an error rate of 0.13%.

The images obtained with the computer-based detection and decision support system developed in Demir's (2019) study were used in CNN for vasoconstriction detection, and it was aimed to assist physicians during angiography. In this context,

the CNN architecture used in the study (inspired by good models in the literature) includes an input layer (this layer is $50 \times 50 \times 1$ because the images have 50×50 resolution and are binary), 3 convolution layers with 8 filters of 6×6 (after each layer there is a pooling layer with a stride of 2 and a filter size of 2×2) and a fully-connected layer (a standard ANN structure) with 3 outputs (corresponding to 3 classes as healthy, stenosis and background). The vein segmentation process, which is used as pre-processing in order to increase the accuracy in CNN, consists of 4 steps as: (1) Noise removal with Weiner filter, (2) Detection of vein or tube-like structures with Frangi filter, (3) Binary image acquisition with Otsu's Thresholding Method, (4) Elimination of salt and pepper type noise with Median filter (3×3). In the study, it was argued that deep learning methods can only be trained correctly with a lot of training data and consent can be obtained after a long process within the scope of the Personal Data Protection Law (patients' health data is in the definition of "qualified personal data" and an ethical committee decision is required for their use in studies). For this reason, in the study, instead of this process, more data was produced from the existing data with data augmentation techniques (adding noise, loss of information, focusing on a random part, rotating at a certain angle, rotating according to a fixed axis so that the image does not lose its meaning were applied. However, since the dataset consists of binary images, colour change and contrast change processes are not applied).

In a portable and cost-effective ophthalmoscope design study proposed by Yılmaz (2014), the features (coefficients) extracted from the orbital (retina, optic disc and macula) images with the Zernike Moment Technique were used in learning with ANNs for retinal diagnosis (checking whether the person is registered in the system), and in comparing the coefficients of images taken at different times in disease diagnosis and follow-up (if the coefficients have changed, it is assumed that there is a health problem and it is stated that these coefficients will guide the physician for further controls). For retinal recognition, training was carried out by applying features to the input of the MLP ANN (consisting of an input layer, a 60-node hidden layer and a 20-node output layer). Since the fundus images obtained in the study were not sufficient for retinal recognition, the software developed was tested with

samples from the DRIVE international retinal database. It has been evaluated that the proposed system can be used within the scope of biometric system based on retinal scanning.

In the study of Doğan and Türkoğlu (2018), the performance of deep learning algorithms (AlexNet, VGG16, VGG19, ResNet50 and GoogleNet) in classifying plant leaves was examined. In this context, the classification process was carried out with the supervised-learning method with MATLAB[®] deep learning algorithms. A 1900-image dataset (retrieved from the UCI-Machine Learning Repository) consisting of 32 classes (each class has an average of 60 leaf images) was used in the study. For a better classification performance, the number of images in the dataset has been increased to 7628 (each class has an average of 240 leaf images) with data augmentation techniques (reflection and reversing). Raw images of 1600x1200 dimensions were reduced to 224x224 and 227x227 dimensions, and randomly selected 80% of these images were used for training and 20% for testing. In the study, the number of iterations in deep learning algorithms was kept low and the training process was completed more quickly (1-3 hours with GPU, more than 24 hours with CPU). Trials were conducted with a system with i7 2.8 GHz CPU and GTX 960m 6GB GPU and 24 MB RAM. According to the results of the study, the accuracy rates of deep learning algorithms in classifying plant leaves are AlexNet (99.72%), VGG16 (99.12%), VGG19 (98.36%), ResNet50 (99.15%) and GoogleNet (97.77%) was detected.

In the studies of Karaali and Eminağaoğlu (2021), a system has been developed in which different types of marble qualities, which are defined and classified manually by the expert (in an error-prone and time-consuming way), are automatically classified with a deep learning model. In this context, plate marble images were processed with the CNN Model specially developed with the Python programming language and classified according to six different quality types. The CNN Model in the study was developed as a result of testing different models and hyperparameters (by trial-and-error method). ReLU activation function is used at the end of both convolution layers (3x3 Conv1 with 64 filters and Conv2 with 128 filters of 3x3)

(dropout ratio of 2nd ReLU function was chosen as 0.2). No pooling process (max or avg) was used in the model since it reduces the classification accuracy of 2 marble classes. With the smoothing layer, a 1D array of 1x139392 and the 139392-node input layer of the FC1 layer (dropout ratio 0.5 and ReLU activation function were used) were created and connected to a 512-node hidden layer with 512x139392 links. The 512-node input layer of the FC2 layer is fully connected to 6 output nodes (representing 6 marble classes). The softmax function was used as the activation function in the output layer. In the trainings, batch normalization (batch size as 16) was applied to the 300x300x3 (RGB) sized data in the input layer. The training epoch was 30, the initial learning rate was 0.001, and the learning rate was determined using the Adam adaptive learning rate optimization algorithm. In the dataset used in the experiments, there are 2100 images cropped to 300x300 pixels, consisting of 6 marble quality classes, each containing 350 RGB (colour) images to 2400x2400 pixels. When the different model, different hyperparameter, different batch sizes, different dropout rates and traditional data augmentation methods used to eliminate the overfitting were also ineffective, a unique data augmentation was carried out. The total number of images was increased to 6300, with 1050 images in each of the 6 classes in the dataset. 5400 of 6300 data were used for training and 900 for testing. In the study, it was determined that with data augmentation, classification success and accuracy were increased and overfitting was prevented. Within the scope of future studies, it was stated that the integration of the software into smart phones would be provided as a single application package.

In the studies of Liu, Li, Sun, Zhang, and Chen (2017), a deep learning model that can be used in finger vein recognition has been proposed. In this context, a 7-layer CNN Model with 5 Conv layers and 2 FC layers has been developed. The recognition system proposed in the study consists of 2 parts: (1) **ROI extraction** (obtaining only finger image information) and (2) **CNN training** (The AlexNet CNN Model was used for the recognition process by changing the network structure and parameters. Grayscale images of 60x175 pixels, which are the input data of the CNN Model, were processed in 5 Conv and 2 FC layers. ReLU, 2x2 Maxpooling and Local Reponse Normalization were used in the Conv1 layer which had 64 filters of

5x5, while ReLU and 2x2 Maxpooling were used in the Conv2, Conv3, Conv4 and Conv5 layers which had 64 filters of 5x5. Feature extraction with 0.5 dropout is done from 150-dimensional inputs coming to FC1 fully-connected layer. In the FC2 fully-connected layer, the class probabilities of each feature are calculated with the softmax function and their belongings are determined for 198 classes). In order to carry out the trials of the study, a dataset of 2970 images was prepared with 15 finger images (13 train, 2 test) (except for the thumb) taken from 198 individuals in the laboratory environment, since there are only a few finger vein datasets currently open to public access. In addition, SDUMLA-FV dataset (3816-image dataset created with 6 finger images (3 trains, 3 tests) taken from 636 individuals) was used in the trials. In the study, it was noted that CNN is an end-to-end system and it provides high robustness while working faster than traditional algorithms for finger vein recognition processes.

In Tükel's (2019) study, it is aimed to develop a hybrid approach that will automatically detect the optic disc in fundus images, using a pre-trained CNN Model (for feature extraction) and k-NN classifier (the data that class will be determined with k-NN is included in the class with the minimum distance (Manhattan Distance, Minkowski Distance or Euclidean Distance can be used for calculations) from the already defined classes), thus helping to diagnose and treat retinal diseases. In this context, selected parts from fundus images were given as input to the pre-trained AlexNet CNN Model (the model includes 8 trainable layers, 5 Conv and 3 FC, and consists of 25 layers in total. Also, after Conv layers using different filter sizes, max pooling is used) via ImageNet, and the feature matrix was created by obtaining 4096-dimensional (i.e. containing 4096 features) feature vectors from the FC6 layer within the scope of transfer learning (feature extraction from fully-connected layers in pre-trained CNN architecture). The locations of the regions with/without the optic disc in the test images were determined by the k-NN classifier using the supervised learning method (Euclidean distance was used for distance measurement and the k value was determined as 11 by trial-and-error). In order to determine the performance in the experiments, training data was created by selecting 500 images with optical disc and 1565 image sections without optical discs from 280x280 sized images (scaled to size

227x227 to fit the CNN Model) in ROC, Drive and Messidor retinal datasets. The training matrix dimensions were obtained as 4096x500 for 500 images with optical disc and 4096x1565 for 1565 images without. For the test, 169 retinal images were selected, and 169 image sections with optic disc and 676 image sections without optic disc (from each image containing an optical disc, 4 parts without an optical disc are selected) were obtained. According to the results of the study, the accuracy value of the proposed approach was 92.66%, the sensitivity value was 79.88%, and the specificity value was 95.86%, among the performance evaluation results. It was determined that the reason for the low sensitivity value was due to the large retinal image dimensions and that the optic disc could not be fully covered by a 280x280 frame. Therefore, the dimensions of the image portions selected for the test were increased to 500x500, and in the new results, the accuracy was determined as 95.74%, the sensitivity 84.46% and the specificity 99.08%.

In the work of Reasat and Shahnaz (2017), a CNN architecture was proposed that uses raw electrocardiography signals for detection of inferior myocardial infarction. In this context, as pre-processing, 4-steps were used: (1) Raw ECG signals of 1000 Hz were reduced to 250 Hz, (2) Low-frequency components that affect the accuracy of the signal (baseline wandering) were removed with the Median Filter and the noise was removed with the Savitzky-Golay Smoothing filter, (3) The noise-free signal was reduced to 64 Hz, and (4) the signal was divided into segments of 3.072 seconds, with 196 signal samples in each segment. By grouping the segments obtained in the study, a total of 3222 IMI samples and 3055 Healthy Control samples were created. The raw ECG signal (each signal consists of Lead II, Lead III and AVF signal segments) was applied to the input layer of the model developed using the Keras Neural Network library. Each lead was given as input to inception blocks (Convolution, Batch normalization, ReLU, Maxpooling) with 7-parallel paths, and the feature maps created as output were concatenated and sent to the global average pooling layer. Categorical probabilities were determined with the Softmax activation layer. For comparison, another method that extracts features from ECG signals with Stationary Wavelet Transform (SWT) was used. Among the methods compared (SWT (classification with k-NN), SWT (classification with SVM) and CNN with raw

ECG), it was seen that the best results were obtained with CNN in terms of average accuracy (84.54%), sensitivity (85.33%), and specificity (84.09%). It was stated that within the scope of future studies, different infarction classifications will be examined according to their locations.

In the work of Emlek and Peker (2020), a CNN Model (used Content-CNN network in the literature) that corrects faulty depth map regions was proposed by using the features obtained from only the left images (image to correct the depth value in the dataset) of stereo image pairs obtained from stereo cameras to improve the depth map (image to be corrected from Content-CNN network). In this context, it was stated that traditional stereo matching (process of detecting the matching pixels in the right and left images obtained from the camera pairs on the same horizontal plane. The horizontal distance between these pixels gives the depth information) algorithms have 4-stages (match cost calculation, sum of costs, calculation/optimization of depth map and depth map correction operations), they work fast, they do not need training, but it is underlined that they cannot prevent errors (inter-object overlap, large unpatterned areas, and repetitive patterns in the image) in the stereo image. Instead, it was stated that CNN-based methods that operate in 3-stages (calculation of matching costs, detection of outliers and correction of the depth map, and obtaining a depth map with end-to-end learning) can be used in cases where traditional methods are insufficient. Depth map (obtained by applying Winner-take-all method to raw match costs from Content-CNN network) and stereo left image (the 9*9 piece taken from the left image is given as an input to the first of the Siamese networks that calculate the matching costs. The other is the set of images taken from the right image and containing all the possibilities along the search space, given as input. Correlation is applied to the right and left feature vectors at the output of the Siamese network and the matching costs are calculated across the search space for the given input. Next, the depth map, which will be the input of the network, is obtained with the Winner-take-all method) were applied to the input of the CNN network structure proposed in the study. Convolution layer (5*5 filter and feature map size: 64, Xavier method was used to determine the initial value of the training parameters, epoch: 100, learning rate: 0.001, beta values 0.9 and

0.999, Adam method was used to minimize the loss function) and ReLU layer (except the last layer) were applied to the left image sent as input to the network, and the feature map with size 1 was created in the last layer, and the correction map was obtained and combined with the initial depth map. As a result, the final depth map was obtained. KITTI 2012 (includes 194 images) and KITTI 2015 (includes 200 images) datasets containing corrected outdoor images for autonomous driving applications were used for the trials of the study and 80% of the data was reserved for training, 10% validation and 10% for testing. When the depth maps produced using the proposed three, four and five layer CNNs were compared with the reference depth maps (obtained with Content-CNN), it was observed that the error rates up to 3-pixel obtained were the least when using the 5-layer network (8.83% for KITTI 2012, 5.87% for KITTI 2015 and 7.52% for KITTI 2012+KITTI 2015). According to the results obtained, it has been determined that the proposed network models, having a small number of layers, can improve the accuracy of the entrance depth map without increasing the processing load.

In the studies of Al-johania and Elrefaei (2019), it was aimed to develop a dorsal hand vein recognition system with the use of CNN. Within this scope two CNN-based approaches were used as: (1) **pre-trained CNN** (features were extracted with FC6, FC7 and FC8 layers of AlexNet, VGG16 and VGG19 CNN Models, classification was performed with k-NN (neighbour: 5, euclidean distance) and Support Vector Machines using Error-Correcting Output Codes) and (2) **transfer learning (fine tuning) CNN** (both feature extraction and classification were performed with AlexNet, VGG16 and VGG19 models. Epoch numbers: 10, 20, 30, 40, 50, mini-batch size: 10 and initial learning rate: 0.0001 were selected. The current number of 1000 classes in the FC8 layer has been reduced to 100 classes in the study). In the study, the proposed deep learning-based system automatically learns to extract features from the original image without the need for any pre-processing. Dr.Badawi hand veins dataset (5 right, 5 left hand images taken from 50 people) and BOSPPHORUS dorsal vein dataset (1575 left hand dorsal images taken from 100 people) were used for the trials of the system. Trials were conducted with Intel Core i7-7700 HQ CPU @ 2.80 GHz 16 GB RAM and NVIDIA GTX 1060

GPU system specifications. For the approaches implanted with MATLAB[®] R2018a deep learning toolbox, the image sizes were arranged as 227x227 to be used in the CNN network. The data in the dataset were divided as train-test (80%-20%, 70%-30% and 60%-40%). According to the results obtained, the best value for recognition accuracy in the first approach was obtained as 99% in the FC6 layer (for split ratio: 80%-20% and 70%-30%) in AlexNet model for the Dr.Badawi dataset in the ECOC classification with k-NN, and 90% in the FC6 layer (for split ratio: 80%-20%) in the VGG16 model for the BOSPPHORUS dataset. In the ECOC classification with SVM, it was obtained as 100% in the FC6 layer (for split ratio: 80%-20% and 70%-30%) in the VGG16 and VGG19 models for the Dr.Badawi dataset, and 98.50% in the FC6 layer (for split ratio: 80%-20%) in the VGG19 model for the BOSPPHORUS dataset. In general, the system achieved the best recognition accuracy values in this approach with feature extraction in the FC6 layer for all models. In the second approach (transfer learning method), the best value for recognition accuracy was obtained for 50 epochs, 100% for the VGG16 and AlexNet models for the Dr.Badawi dataset, and 99.25% for the VGG19 model for the BOSPPHORUS dataset. As a result, it was stated that the use of pre-trained CNN Models in the scope of transfer learning gives better accuracy than using only for feature extraction.

2.2.3.4 The Proposed System's Classification Phase

The class of the tissue region viewed by the user is determined in the Classification Phase. In this context, NIR fused-modified images obtained in the Digital Image Post-Processing Phase were used. In the follow-up of vascular degeneration, it is extremely critical to classify the vascular pattern in images taken from the same tissue region (ROI) at certain periods (week, month, or year). The patterns obtained from the ROI in the Period-1 are used in the training and pre-testing (validation) of the CNN Model and they form the basis for determining the tissue classes. The belonging of the new NIR fused-modified images obtained from the videos recorded by the user in the second and subsequent periods for the purpose of degeneration monitoring, to the classes determined in this first period is

established by the CNN post-test process. If the class (belonging to one of the classes in the first period) of the viewed region can be determined according to the post-test result, in the next phase, the Object Detection Phase, objects are detected in these images, and in the Augmented Reality Phase, the images are added to the video of the determined class as a new period. During the Imaging Technique Phase (since the number of images that can be captured in a single viewing period is limited), the user is requested to take a video recording, so the need for a large number (hundreds, thousands) of images required for the training of CNN is met and it is ensured that the lighting or tissue movement will not affect the recognition process. It is aimed to determine that the tissue region viewed in the Classification Phase is the same as in the first period.

The Classification Phase of the proposed system is coded with the MATLAB[®] R2017a program and the MatConvNet (matconvnet-1.0-beta24 version (MatConvNet Team, 2022)) library (containing various CNN application examples such as segmentation, classification, recognition for computer vision) is used to perform CNN operations over MATLAB[®]. In the study, the MNIST CNN Model (a CNN Model that uses the MNIST dataset of handwritten digits), which is a CNN Model in the matconvnet library, is used to determine the class probabilities of the vein images.

The steps and explanations of the proposed system's Classification Phase are detailed in Table 2.15.

Table 2.15 Steps and explanations of the proposed system's Classification Phase

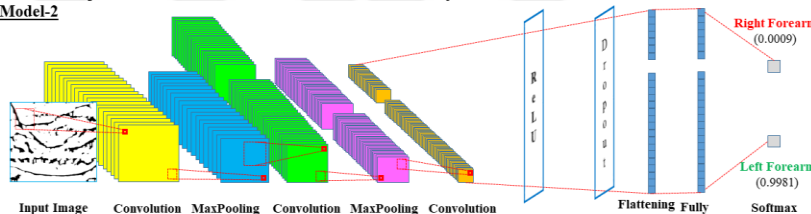
Classification Phase Step	Explanation																																																																																																																																																																																																																																							
Designing CNN Model Architectures (Effect of input image dimensions on classification)	<p>In order to create CNN Models, MNIST CNN Model, which was prepared for the MNIST dataset (containing 10 classes corresponding to 0-9 handwritten digits) and defined in the matConvNet library of the MATLAB® program, was used in this thesis study. In the Classification Phase, three different CNN Model architectures were used to observe the effect of the input image sizes on the classification;</p> <ul style="list-style-type: none">CNN Model-1: The original MNIST CNN Model supporting 28x28 input sizes (MatConvNet Team, 2022). The original output values of this model, corresponding to 10 classes, were reduced to 2 classes, recognizing the right and left forearms, by making changes only in the last layer within the scope of the thesis study. NIR fused-modified images containing vein patterns were used instead of the numerical dataset. <table><tr><th rowspan="2"></th><th rowspan="2">Input</th><th colspan="6">FEATURE EXTRACTION</th><th colspan="2">CLASSIFICATION</th></tr><tr><th>Conv-1</th><th>MaxPool-1</th><th>Conv-2</th><th>MaxPool-2</th><th>Conv-3</th><th>Rel_U</th><th>FullyCon-1</th><th>Softmax</th></tr><tr><td>Depth:</td><td>1</td><td>20</td><td>20</td><td>50</td><td>50</td><td>500</td><td>500</td><td>2</td><td>1</td></tr><tr><td>Data Size:</td><td>28 x 28</td><td>24 x 24</td><td>12 x 12</td><td>8 x 8</td><td>4 x 4</td><td>1 x 1</td><td>1 x 1</td><td>1 x 1</td><td>1 x 1</td></tr><tr><td>Filter Size:</td><td>N/A</td><td>5 x 5</td><td>2 x 2</td><td>5 x 5</td><td>2 x 2</td><td>4 x 4</td><td>1 x 1</td><td>1 x 1</td><td>1 x 1</td></tr><tr><td>Padding:</td><td>N/A</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Stride:</td><td>N/A</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table> <ul style="list-style-type: none">CNN Model-2: The CNN Model in which the filter values of the first model are recalculated and updated with the formula (2) to support 64x64 pixel input sizes, and the dropout layer (added after overfitting in the experiments) is added just before the last convolution layer. <p><u>CNN Model-2</u></p>  <table><tr><th rowspan="2"></th><th rowspan="2">Input</th><th colspan="7">FEATURE EXTRACTION</th><th colspan="2">CLASSIFICATION</th></tr><tr><th>Conv-1</th><th>MaxPool-1</th><th>Conv-2</th><th>MaxPool-2</th><th>Conv-3</th><th>Rel_U</th><th>Dropout</th><th>FullyCon-1</th><th>Softmax</th></tr><tr><td>Depth:</td><td>1</td><td>20</td><td>20</td><td>50</td><td>50</td><td>100</td><td>100</td><td>100</td><td>2</td><td>1</td></tr><tr><td>Data Size:</td><td>64 x 64</td><td>58 x 58</td><td>29 x 29</td><td>22 x 22</td><td>11 x 11</td><td>4 x 4</td><td>4 x 4</td><td>4 x 4</td><td>1 x 1</td><td>1 x 1</td></tr><tr><td>Filter Size:</td><td>N/A</td><td>7 x 7</td><td>2 x 2</td><td>8 x 8</td><td>2 x 2</td><td>8 x 8</td><td>1 x 1</td><td>1 x 1</td><td>4 x 4</td><td>1 x 1</td></tr><tr><td>Padding:</td><td>N/A</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>N/A</td><td>N/A</td><td>0</td><td>N/A</td></tr><tr><td>Stride:</td><td>N/A</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td><td>N/A</td><td>N/A</td><td>1</td><td>N/A</td></tr></table> <ul style="list-style-type: none">CNN Model-3: The CNN Model in which the filter values of the first model are recalculated and updated with the formula (2) to support 128x128 pixel input sizes, and a convolution layer is added before the ReLU layer. <table><tr><th rowspan="2"></th><th rowspan="2">Input</th><th colspan="8">FEATURE EXTRACTION</th><th colspan="2">CLASSIFICATION</th></tr><tr><th>Conv-1</th><th>MaxPool-1</th><th>Conv-2</th><th>MaxPool-2</th><th>Conv-3</th><th>MaxPool-3</th><th>Conv-4</th><th>Rel_U</th><th>Dropout</th><th>FullyCon-1</th><th>Softmax</th></tr><tr><td>Depth:</td><td>1</td><td>20</td><td>20</td><td>100</td><td>100</td><td>500</td><td>500</td><td>1000</td><td>1000</td><td>1000</td><td>2</td><td>1</td></tr><tr><td>Data Size:</td><td>128 x 128</td><td>118 x 118</td><td>58 x 58</td><td>50 x 50</td><td>24 x 24</td><td>18 x 18</td><td>8 x 8</td><td>5 x 5</td><td>5 x 5</td><td>5 x 5</td><td>1 x 1</td><td>1 x 1</td></tr><tr><td>Filter Size:</td><td>N/A</td><td>11 x 11</td><td>4 x 4</td><td>9 x 9</td><td>4 x 4</td><td>7 x 7</td><td>4 x 4</td><td>4 x 4</td><td>1 x 1</td><td>1 x 1</td><td>5 x 5</td><td>1 x 1</td></tr><tr><td>Padding:</td><td>N/A</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Stride:</td><td>N/A</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td><td>2</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table> <p>Since no previously-trained weights are used in any of the models used and new weights are created by starting all the trainings from scratch, the CNN Models in the Classification Phase are not considered within the scope of transfer learning.</p>		Input	FEATURE EXTRACTION						CLASSIFICATION		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	Rel_U	FullyCon-1	Softmax	Depth:	1	20	20	50	50	500	500	2	1	Data Size:	28 x 28	24 x 24	12 x 12	8 x 8	4 x 4	1 x 1	1 x 1	1 x 1	1 x 1	Filter Size:	N/A	5 x 5	2 x 2	5 x 5	2 x 2	4 x 4	1 x 1	1 x 1	1 x 1	Padding:	N/A	0	0	0	0	0	0	0	0	Stride:	N/A	1	2	1	2	1	1	1	1		Input	FEATURE EXTRACTION							CLASSIFICATION		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	Rel_U	Dropout	FullyCon-1	Softmax	Depth:	1	20	20	50	50	100	100	100	2	1	Data Size:	64 x 64	58 x 58	29 x 29	22 x 22	11 x 11	4 x 4	4 x 4	4 x 4	1 x 1	1 x 1	Filter Size:	N/A	7 x 7	2 x 2	8 x 8	2 x 2	8 x 8	1 x 1	1 x 1	4 x 4	1 x 1	Padding:	N/A	0	0	0	0	0	N/A	N/A	0	N/A	Stride:	N/A	1	2	1	2	1	N/A	N/A	1	N/A		Input	FEATURE EXTRACTION								CLASSIFICATION		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	MaxPool-3	Conv-4	Rel_U	Dropout	FullyCon-1	Softmax	Depth:	1	20	20	100	100	500	500	1000	1000	1000	2	1	Data Size:	128 x 128	118 x 118	58 x 58	50 x 50	24 x 24	18 x 18	8 x 8	5 x 5	5 x 5	5 x 5	1 x 1	1 x 1	Filter Size:	N/A	11 x 11	4 x 4	9 x 9	4 x 4	7 x 7	4 x 4	4 x 4	1 x 1	1 x 1	5 x 5	1 x 1	Padding:	N/A	0	0	0	0	0	0	0	0	0	0	0	Stride:	N/A	1	2	1	2	1	2	1	1	1	1	1
	Input			FEATURE EXTRACTION						CLASSIFICATION																																																																																																																																																																																																																														
		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	Rel_U	FullyCon-1	Softmax																																																																																																																																																																																																																															
Depth:	1	20	20	50	50	500	500	2	1																																																																																																																																																																																																																															
Data Size:	28 x 28	24 x 24	12 x 12	8 x 8	4 x 4	1 x 1	1 x 1	1 x 1	1 x 1																																																																																																																																																																																																																															
Filter Size:	N/A	5 x 5	2 x 2	5 x 5	2 x 2	4 x 4	1 x 1	1 x 1	1 x 1																																																																																																																																																																																																																															
Padding:	N/A	0	0	0	0	0	0	0	0																																																																																																																																																																																																																															
Stride:	N/A	1	2	1	2	1	1	1	1																																																																																																																																																																																																																															
	Input	FEATURE EXTRACTION							CLASSIFICATION																																																																																																																																																																																																																															
		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	Rel_U	Dropout	FullyCon-1	Softmax																																																																																																																																																																																																																														
Depth:	1	20	20	50	50	100	100	100	2	1																																																																																																																																																																																																																														
Data Size:	64 x 64	58 x 58	29 x 29	22 x 22	11 x 11	4 x 4	4 x 4	4 x 4	1 x 1	1 x 1																																																																																																																																																																																																																														
Filter Size:	N/A	7 x 7	2 x 2	8 x 8	2 x 2	8 x 8	1 x 1	1 x 1	4 x 4	1 x 1																																																																																																																																																																																																																														
Padding:	N/A	0	0	0	0	0	N/A	N/A	0	N/A																																																																																																																																																																																																																														
Stride:	N/A	1	2	1	2	1	N/A	N/A	1	N/A																																																																																																																																																																																																																														
	Input	FEATURE EXTRACTION								CLASSIFICATION																																																																																																																																																																																																																														
		Conv-1	MaxPool-1	Conv-2	MaxPool-2	Conv-3	MaxPool-3	Conv-4	Rel_U	Dropout	FullyCon-1	Softmax																																																																																																																																																																																																																												
Depth:	1	20	20	100	100	500	500	1000	1000	1000	2	1																																																																																																																																																																																																																												
Data Size:	128 x 128	118 x 118	58 x 58	50 x 50	24 x 24	18 x 18	8 x 8	5 x 5	5 x 5	5 x 5	1 x 1	1 x 1																																																																																																																																																																																																																												
Filter Size:	N/A	11 x 11	4 x 4	9 x 9	4 x 4	7 x 7	4 x 4	4 x 4	1 x 1	1 x 1	5 x 5	1 x 1																																																																																																																																																																																																																												
Padding:	N/A	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																												
Stride:	N/A	1	2	1	2	1	2	1	1	1	1	1																																																																																																																																																																																																																												

Table 2.15 Continues

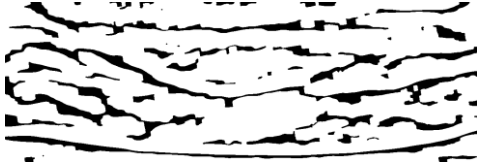
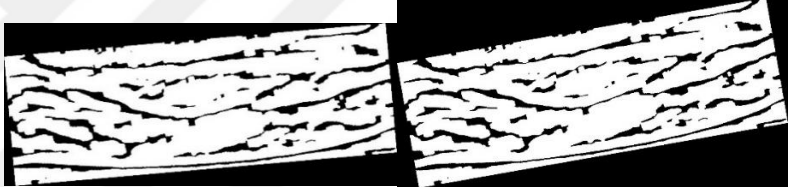
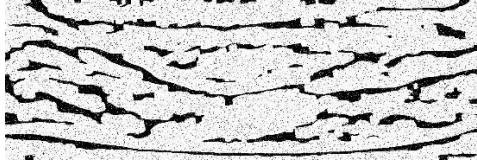
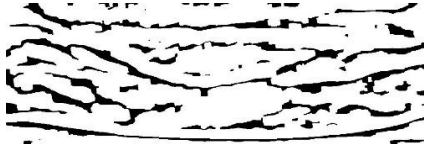
<p>Preparing Datasets (Effect of data numbers on classification)</p>	<p>Since there is no vascular dataset consisting of NIR images open to public access, data belonging to the user directly were used in this thesis study. The data was obtained not with high-end equipment, but with a cost-effective display device for home use. In the classification training, a video recording was requested from the user, since a few photo images to be taken by the user were not enough. In addition, data augmentation techniques (the technique of obtaining new data with different properties from existing data with different DIP techniques) were used in order to increase the classification accuracy by using more data than the results obtained from the experiments.</p> <p>Imaging was performed on the right and left forearm tissue (preferred both in terms of having less subcutaneous adipose tissue and providing practicality within the scope of image supply) on a monthly basis, specific to a single user. In Period-1, video recordings of 35 seconds (1050 right forearm and 1050 left forearm images in total with 30 fps) were taken separately from the right forearm and left forearm. The same process was continued for Period-2. In Period-3, only the left forearm was visualized (representing the absence of a condition requiring follow-up in the right forearm). In this context, a dataset of 5250 images in total with 2 classes (right forearm and left forearm) was prepared.</p> <p>In the training and pre-test (validation) step, only the NIR fused-modified images (containing 525 images each) of the right forearm and left forearm images of only the Period-1 obtained by the Digital Image Post-Processing Phase were used. In order to observe the effect of the data numbers on the classification, five datasets were included in the Classification Phase;</p> <ul style="list-style-type: none"> • Dataset-1: Contains 200 train and 50 pre-test images. • Dataset-2: Contains 400 train and 100 pre-test images. • Dataset-3: Contains 720 train and 180 pre-test images. Images in Dataset-2 were used. In addition, 320 train and 80 pre-test images obtained with data augmentation techniques (Rotation, Noise and Scale operations, examples of which are shown in Table 2.16) were added. • Within the scope of the study, Dataset-4 (200 trains and 50 pre-tests) and Dataset-5 (400 trains and 100 pre-tests) containing only NIR raw-images were also prepared. However, when these datasets were used, both the training times increased excessively (due to the average size of the NIR raw-images of 5MB) and the desired accuracy levels could not be reached (as the images do not contain enough detail) in the post-test step. To reduce training time and improve accuracy, the same datasets were reconstructed to include NIR cropped-raw images (average 2 MB in size) containing only the ROI. However, since the desired accuracy levels could not be reached with these datasets (using low file size images reduced training times), the trainings were continued with the first three datasets.
<p>Train and Pre-Test (Validation)</p>	<p>For CNN trainings, MATLAB[®] program matConvNet library studies were used. All proposed CNN Models were trained separately with all three datasets in this step. In the Train and Pre-Test Step;</p> <ul style="list-style-type: none"> • Placing the images in folders named right forearm and left forearm (created under the folders named train and test) so that the Datasets can be used by the program (CNN Models include data in the classification process by folder names only. In this context, instead of naming the images one by one, it is sufficient to simply copy them under the relevant class folder. Image naming was used in this thesis study for convenience, but the image names are not important at this phase),

Table 2.15 Continues

	<ul style="list-style-type: none"> • Resizing the images so that the datasets can be used by the program (images of which sizes have been changed according to the CNN Model should be stored in imdb.mat file), • Determining the Learning Rate value as 0.0001 (the desired accuracy levels could not be reached in the trainings performed with 0.01 and 0.001), • Changing the number of epochs for each model (the trainings were terminated by following the objective function value obtained for each epoch below 0.001 and the top1err value below 0.000 in the trainings), • Determining the batch size as 100 (the trainings are performed on a computer with 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz, 8 GB RAM and Windows 10 Operating System installed, without using GPU, so it is preferred to shorten training times), operations have been carried out. <p>In the trainings, the weight values obtained after each epoch are saved in files with mat extension, which increase sequentially as net-epoch-1.mat, net-epoch-2.mat etc. In long-term trainings, training can be continued by using the last file if needed.</p>
Post-Test	<p>The MATLAB® program matConvNet library was used for CNN post-test experiments. After the completion of the trainings, the evaluation of the accuracy performances of the CNN Models in the classification by using the data with and without the training dataset is called the “post-test” process within the scope of this thesis. In the scope of post-test, NIR raw-images, NIR cropped-raw images, and NIR fused-modified images were tested. However, the desired accuracy levels were achieved when only datasets (numbered 1, 2 and 3) containing NIR fused-modified images were used. In this context, Classification Phase results were obtained using NIR fused-modified images. For the Post-Test Step;</p> <ul style="list-style-type: none"> • Among the images obtained in Period-1, 21 images were selected for the right forearm and 21 images for the left forearm. • Among the images obtained in Period-2, 21 images were selected for the right forearm and 21 images for the left forearm. • In Period-3, where only the left forearm was displayed, 21 images were selected for the left forearm. <p>In addition, Period-1 images used in the training (10 train images were taken from Dataset-1 and all 21 train images were taken from Dataset-2 and Dataset-3) were used to confirm vein recognizability. Images with the same code numbers were selected from each dataset and used.</p> <p>The results obtained are interpreted in the subheading 3.1 Experimental Results of the Classification Phase.</p>

Data augmentation techniques of the proposed system's Classification Phase are given in Table 2.16.

Table 2.16 Data augmentation techniques and explanations of the proposed system's Classification Phase

Data Augmentation Process	Explanation
Raw-image	<p>For data augmentation, binary NIR fused-modified images (517x1289x3 uint8) obtained in the Digital Image Post-Processing Phase and an example of which is shown in the Figure below, were used.</p>  <p>As a representative, the binary NIR fused-modified image with the code Fused_Modified_Period1_Img0-1_RightArm.bmp was used</p>
Rotation By Angle	<p>In order to tolerate user-induced vibrations or tissue movement in the camera in the classification, 5 and 10 degrees of rotation were used as shown in the Figures below. Mirroring was not used as it will cause the image to lose its meaning (right and left forearm contours will interfere). (<i>imrotate</i> command was used)</p>  <p>(a) (b)</p> <p>Rotation by angle. (a) 5 degree rotation applied image, and (b) 10 degree rotation applied image</p>
Noise Addition	<p>In order to tolerate the noise that may occur in the image in classification, salt and pepper type noise with 0.1 intensity parameter was used as shown in the Figure below. (<i>imnoise</i> command was used)</p>  <p>Noise (0.1 intensity valued salt and pepper type) applied image</p>
Scale Change	<p>A downscaling of 0.8 was used in order to tolerate the camera being held close/far from the tissue by the user in the classification as shown in the Figure below. (<i>imresize</i> command was used)</p>  <p>Scale change (0.8 zoom out valued) applied image (the size of the sample image is shown proportionally)</p>

Summary of the proposed system's Classification Phase in 1-minute is given in Figure 2.13.

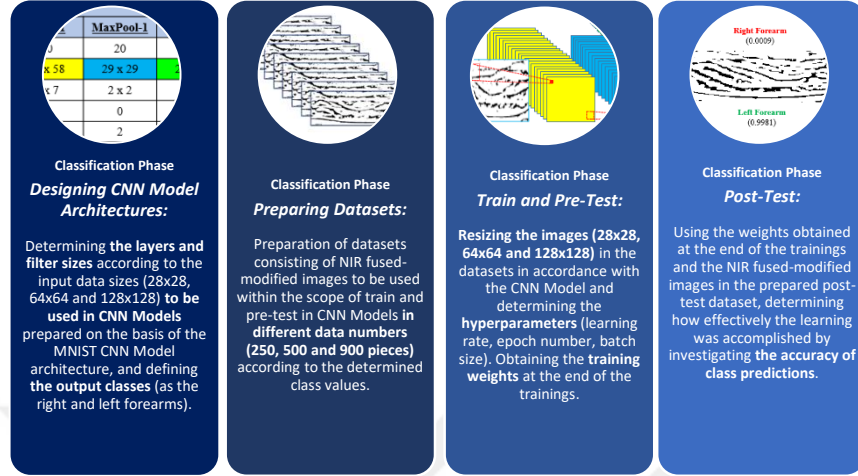


Figure 2.13 The proposed system's Classification Phase in 1-minute

2.2.4 The Object Detection Phase: Catching Vascular Degeneration

Early diagnosis saves lives. Vascular disorders, which are evaluated within the broad medical framework, are considered within the scope of the nomenclature of "vascular degeneration", indicating only vasoconstriction (stenosis vein) and vascular enlargement (spider vein and varicose vein). Vascular degeneration is functional disorders in the vascular structure due to various factors (usually environmental conditions such as diet and lifestyle, rarely congenital) (Roizen & Oz, 2005). These functional disorders can cause changes in the vascular structure by preventing normal blood flow over time, and even turn into dramatic results that are much more difficult to treat in later stages.

The arteries that distribute the blood pumped from the heart (Hemoglobin molecules in the blood carry out the task of carrying oxygen) to the body, the capillaries that deliver the blood to the deepest parts of all tissues and organs, and the veins that collect blood from the body (containing oxygen-free Hemoglobin molecules that have given their oxygen to tissues and organs) and bring it to the heart are the most basic components of the circulatory system. The valves, which can be

called the gates through which the blood passes into the vessels, have an extremely important role in maintaining the functionality of this system. Over time, cholesterol accumulation occurs in the damaged areas that occur due to deterioration in the inner walls of the vessels (intima layer) due to reasons such as diet (excessive fat, sugar consumption) or excessive blood pressure, and subsequently, narrowing and blockages in the internal volume of the vessels can occur with the accumulation of blood clots on it (Roizen & Oz, 2005). These narrowings, which can be encountered in the whole body, increase over time and cause a decrease in blood flow to the target tissues and organs. In extreme cases, blood clots can block the vessels and suddenly reduce or completely block blood flow. On the other hand, another factor that can lead to degeneration of the vessels is the lifestyle (a sedentary lifestyle due to the profession and long working hours at the desk). Standing or sitting for too long can lead to dysfunctions in the valves in the veins. Especially the leg veins, which carry the whole body weight, are exposed to a lot of pressure during the day. The blood, which must be carried up from the legs, escapes from the incompletely closed valves in the veins due to the pressure, causing accumulations. These accumulations also increase over time, resulting in morphological deformation (widening by twisting) of the veins. In fact, if left untreated, this condition may progress and lead to ulcers in the lower extremities (legs).

In case of early detection, vasoconstriction can be slowed down by changing dietary habits (avoiding refined and fatty foods, junk food) and vascular enlargement by changing the lifestyle (not sitting/standing for more than an hour). What is important is how early the detection of degeneration can be accomplished.

In the Object Detection Phase, vascular degeneration patterns found in the NIR fused-modified images are detected using the YOLOv3 algorithm within the scope of early-diagnosis.

2.2.4.1 Object Detection from Digital Images

Objects on the image can be detected together with their location information by using object detection algorithms. In this respect, in this thesis, the YOLOv3 object detection algorithm is used to detect the degeneration in the vein patterns obtained with the image processing steps, together with their locations. In the Object Detection Phase, classes of vascular degeneration (contraction/enlargement) patterns artificially added to NIR fused-modified images are determined and their locations are specified by marking them with bounding boxes.

If it is desired to obtain location information in addition to the class prediction, R-CNN (not preferred in real-time applications because the training period is too long), Fast R-CNN (input is the entire input image, speed could be increased by performing convolution only once per image) or Faster R-CNN (the RPN-Regional Proposal Network is used to estimate regional offers), which provide box and label information, can be used (Demir, 2019). In fact, SSD or YOLO algorithms may be preferred if faster detection is desired that can also support real-time applications.

Object Detection Algorithms, one of the Deep Learning Algorithms (Mesci, 2022), are basically examined under two headings: (1) **One-Stage Detectors** (an approach that does not offer a region proposal. Bounding box predictions and classification are performed in a single run with an end-to-end neural network. SSD, YOLO can be given as examples) and (2) **Two-Stage Detectors** (region proposal approach. Two separate processes are carried out as: (1) Suggesting the regions where the objects are likely to be found and (2) Classifying the objects in the proposed regions. R-CNN, Fast R-CNN, and Faster R-CNN can be given as examples).

Object detection algorithms and their explanations are given in Table 2.17.

Table 2.17 Object detection algorithms and explanations (Bandyopadhyay, 2022; Geeks for Geeks, 2022; İpek, 2019; Mesci, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020)

Object Detection Algorithm Name	Explanation
Sliding Window Approach	<p>A fixed size window is shifted starting from the top left corner of the image, right and then down. Relevant regions are processed in the model and the searched object is queried. (İpek, 2019)</p> <p>If the object is too far from the camera (too small in size) there will be a mismatch between the window size and the object size. Requires window size reduction (which will increase the number of processes) for more accurate detection. (İpek, 2019)</p>
R-CNN	<p>In R-CNN, which is a region-based object detection algorithm (Mesci, 2022), possible region suggestions (2000 different regions with Selective Search) where the searched object can be found in the image are detected first. Afterwards, CNN feature extraction and classification is performed for each proposed region one by one. Finally, operations such as box drawing and elimination of many regions representing the same object with confidence score are performed. (İpek, 2019)</p> <p>Since it requires running CNN about 2000 times for each image, the training time is very high (İpek, 2019).</p> <p>In addition, the processing load of these steps reduces the number of frames to be processed per second. The training process is complex. (İpek, 2019)</p>
Fast R-CNN	<p>In Fast-CNN, the image is first processed with CNN in one go. Although it provides high mAP, the training process is complex and contains too many iterations, so the detection speed is reduced. Therefore, it does not support real-time detection. (Bandyopadhyay, 2022; İpek, 2019)</p>
Faster R-CNN	<p>After the last convolution layer, Region Proposal Network (RPN) is applied to identify possible regions of interest and then these regions are individually recognized. It offers real-time object detection with an average accuracy of 7 fps/73.2% mAP (mean average precision) with the PascalVOC2007 dataset. (Bandyopadhyay, 2022; İpek, 2019; Mesci, 2022)</p>
SSD (Single Shot Multi-Detector)	<p>Regions with objects in the image can be detected at once. Bounding box estimation is performed using Anchor Boxes of various aspect ratios and sizes located at different locations in the image. For this, feature maps (with VGG16 backbone structure) are first extracted and then the classes of the objects are determined by applying VGG16 convolution filters to the detected objects. (İpek, 2019; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020)</p> <p>Small object size reduces detection performance slightly (Geeks for Geeks, 2022). With PascalVOC 2007 dataset, it can process with 59 fps/74.3% mAP accuracy rate (İpek, 2019). With SSD300, 8732 regions are detected per class for a 300x300 resolution image. The processing load is increasing (İpek, 2019).</p>

Table 2.17 Continues

YOLO (You Only Look Once)	<p>Based on GoogleNet, YOLO is based on applying a single neural network to the entire image (Bandyopadhyay, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020). The class and coordinates of all objects are estimated by passing the image through the end-to-end neural network in one go (taking it as a single regression problem by performing one step of determining the bounding boxes and classes). For this, the input image is divided into nxn grids (Bandyopadhyay, 2022; Mesci, 2022).</p> <p>It can work in real-time (with the use of GPU and PascalVOC2007 dataset with YOLOv1 45 fps/63.5% mAP accuracy) (İpek, 2019).</p> <p>In its more advanced versions, 150 fps processing speed has been achieved with the Anchor Box approach and similar additions (İpek, 2019).</p> <p>It gives better results than SSD when the object size is small (Geeks for Geeks, 2022).</p> <p>For an image with a resolution of 448x448, 98 regions are determined per class. The processing load will be less than the SSD300 algorithm. (İpek, 2019)</p>
--	--

2.2.4.2 The YOLO Algorithm and Versions

Launched as state-of-the-art (Bandyopadhyay, 2022), the YOLO Algorithm and its official versions (only the first 3 versions are accepted as part of the official YOLO study (Bandyopadhyay, 2022)) are an object detection algorithm that can form the basis for numerous studies in computer vision (vehicle detection, autonomous driving, etc. (İpek, 2019)), whether real-time or not.

The YOLO Algorithm performs two basic operations: (1) **Regression** (location detection with bounding boxes, by using regression) and (2) **Classification** (determining of the classes of detected objects) (İpek, 2019). The YOLO algorithms (including unofficial ones) are summarized in Table 2.18.

Table 2.18 YOLO algorithm versions and explanations (Bandyopadhyay, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020; Öksüz, 2022)

YOLO Version	Explanation
YOLOv1 2015	<p>YOLOv1 architecture (based on GoogleNet architecture) consists of 24 Conv and 2 FC layers (Bandyopadhyay, 2022).</p> <p>Operating Method: YOLOv1 is performed in 4-steps (Bandyopadhyay, 2022; İpek, 2019; Mesci, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020);</p> <ol style="list-style-type: none"> (1) The input image is divided into nxn grids (such as 3x3, 5x5, 7x7), where each grid can detect a single object (Mesci, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020). (2) For each grid: (i) it is checked whether there is any object in it, (ii) if there is an object, it is checked whether the middle (center) point is in the grid, and (iii) even if the middle point is within the grid, the width, height and class of the object are determined and a bounding box (B) is drawn around the object (Mesci, 2022). (3) A 6-element prediction vector of each grid is generated. The prediction vector consists of: (i) Confidence Score (expresses the predictive accuracy of whether there are any objects in the grid, indicated by a value in the range [0, 1]), (ii) B_{CX} (x-coordinate of the center of the detected object), (iii) B_{CY} (y-coordinate of the center of the detected object), (iv) B_{OW} (width of detected object), (v) B_{OH} (height of detected object), (vi) Affiliated Class Probability (prediction value as number of classes) (Mesci, 2022). <ol style="list-style-type: none"> a. Confidence Score = Box Confidence Score x Affiliated Class Probability = $(P(\text{object}) \times \text{IoU}) \times \text{Affiliated Class Probability}$. ($P(\text{object})$: The probability that the box contains an object or not. $\text{IoU} = \text{Area of Overlap} / \text{Area of Union}$. IoU: The criterion for evaluating the accuracy of location detection of object detection algorithms, obtained by dividing the intersection of the areas covered by the box whose contents were predicted with Ground Truth, by its combination. True Positive (or positive prediction, $\text{IoU}_{\text{predicted}} > \text{IoU}_{\text{threshold}}$) or False Positive (or negative prediction, $\text{IoU}_{\text{predicted}} < \text{IoU}_{\text{threshold}}$) evaluation is made according to the threshold value to be determined (0.5 is taken by default). (Bandyopadhyay, 2022; İpek, 2019; Mesci, 2022) b. If no objects are detected within the grid, the Confidence Score = Box Confidence Score x 0 = 0 (Mesci, 2022). <p>Deficiency: Although detecting a single object in a single grid (as only one class can be assigned to a single grid) provides a relatively simple structure, it cannot be detected if there are multiple objects in the grid. Detection of small objects and specifying their locations is difficult because only one object can be detected in each input. Localization accuracy (according to Fast R-CNN) is low (Bandyopadhyay, 2022).</p>
YOLOv2 YOLO9000 2016	<p>The YOLOv2 architecture backbone includes DarkNet-19 (containing 19 Convolution and 2 maxPool layers) (Bandyopadhyay, 2022).</p> <p>Accuracy has been increased in the YOLOv2 version without affecting the low detection time (Morera, Sánchez, Moreno, Sappa, & Vélez, 2020).</p> <p>With batch normalization, the mAP value could be increased by 2%.</p> <p>The YOLO9000 (with YOLOv2 architecture) has been trained with both ImageNet (22.000 classes) and COCO (80 classes) datasets to recognize approximately 9000 classes. Compared to YOLOv2, the mAP value is low.</p> <p>Operating Method: While it is assumed that there is only 1 object in each grid in YOLOv1, more than one object can be recognized in YOLOv2 (includes Anchor Box method) (Mesci, 2022).</p> <p>Anchor Box method is used which allows multiple bounding boxes (5 to balance model complexity and prediction performance) for multiple object detection in a</p>

Table 2.18 Continues

	<p>single grid.</p> <p>In YOLOv2, objects to be detected in a grid are also checked for similarity (horizontal or vertical boxes) to pre-determined Anchor-Boxes (calculated with IoU). Thus, the prediction vector obtained for each grid contains 6 elements as much as the number of Anchor Boxes. However, in this case, both many redundant box drawings and more than one box drawing for the same object can be realized. Boxes that do not contain objects can be suppressed directly, while boxes that contain a center point can be represented by only 1 bounding box with the Non-max Suppression Algorithm (consists of 3-stages: (1) Boxes with a Confidence Score below the threshold are eliminated, (2) Of the remaining boxes, all but the one with the highest confidence score are eliminated, and (3) All boxes with an IoU value above 0.5 with the remaining box are eliminated). (Mesci, 2022)</p> <p>Deficiency: Although the detection time of YOLOv2 and YOLO9000 is shorter compared to SSD, the accuracy value remains low. Detection of small objects is not completely resolved. (Bandyopadhyay, 2022)</p>
YOLOv3 2018	<p>The backbone of YOLOv3 is DarkNet-53 architecture (an open-source ANN framework designed with C programming language and CUDA that can support both CPU and GPU (https://pjreddie.com/darknet). It consists of 106 layers including Residual Connections and Upsampling) trained with ImageNet dataset (Bandyopadhyay, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020).</p> <p>Operating Method: 3 bounding boxes (total of 9 anchor boxes) are used at 3 different scales for multiple object detection in a single grid (Bandyopadhyay, 2022).</p> <p>With Residual Connections and Upsampling, feature maps extracted in the 82nd, 94th and 106th Convolution layers can be detected at 3 different scales. Thus, time and accuracy can be changed inversely without the need for retraining. (Bandyopadhyay, 2022; Morera, Sánchez, Moreno, Sappa, & Vélez, 2020; Öksüz, 2022)</p> <p>Since the output of upsampled layers can be concatenated with feature maps from previous layers, small-sized features can be preserved and therefore smaller objects can be detected. (Bandyopadhyay, 2022).</p> <p>Deficiency: As the algorithm complexity increases, the detection time increases (Bandyopadhyay, 2022).</p>
YOLOv4 2019 YOLOv4 Tiny (Not considered an official YOLO study)	<p>In YOLOv4, 65 fps and 43.5% mAP values were reached by making changes in the YOLOv3 architecture and training style (Weighted Residual Connections, Self Adversarial Training Cross Stage Partial Connections, Cross Mini Batch Normalization, and Mish Activation) (Bandyopadhyay, 2022).</p> <p>The YOLOv4 Tiny version, on the other hand, is a version where the accuracy is slightly reduced, but the prediction time can be reduced (Bandyopadhyay, 2022; İnaç, 2022).</p>
YOLOv5 2020 YOLACT (Not considered an official YOLO study)	<p>YOLOv5 is an open-source project containing a pre-trained YOLO model with the COCO dataset (Bandyopadhyay, 2022).</p> <p>In You Only Look At Coefficients (YOLACT), real-time instance segmentation based on YOLO was performed with 33.5 fps and 29.8% mAP values (Bandyopadhyay, 2022).</p>

2.2.4.3 Literature Review of Object Detection

A computer-based detection and decision support system which aimed to assist physicians during angiography was developed in Demir's (2019) study. For the trials of the study, a dataset consisting of binary images with 50x50 resolution containing real data (open source images consisting of anonymous images scored by physicians for vasoconstriction) and artificial data (created by changing the rotations of 3D artificial vessel models and by adding noise, also by generating new images simulating angiography images) were used. In the study, narrowing detection for heterogeneous/unprocessed (containing illumination differences or low resolution) images yielded 50% accuracy with R-CNN. Since the CNN Model changes could not increase the detection accuracy, the real images in the dataset were brought into a homogeneous format by applying the segmentation process. However, due to insufficient success, it was decided to experiment with classical CNN. 70% of the images in the dataset were used for training and 30% for pre-testing (validation), and an accuracy rate of 94.84% was obtained. Due to the high number of false positive values, misclassified images were re-given to the system as training data with the correct label name, and the accuracy of the system was increased. As a result, narrowed veins could be classified with 95.30% accuracy, healthy vessels 91.31%, and background with 100% accuracy.

In the work of Çağıl and Yıldırım (2020), it was aimed to identify the mounting accessory parts in the camera images and to determine the number of parts. By using the developed system in factory production lines, it was envisaged that productivity can be increased in production and that workers will be able to work without the need to memorize part codes. In this context, DIP and deep learning-based object detection were used in the study. The system design was carried out with tiny YOLOv3 algorithm (required trainings for part detection), Google Colaboratory (Colab) Platform (for GPU infrastructure that will accelerate the training of the developed system), Python Programming Language, OpenCV library (for determining the number of pieces in the image) and Darknet Neural Network Framework. The study consisted of 6-steps as: (1) **Creating the Dataset** (obtaining

images in a dataset consisting of 312 images having 600x600 pixels with different backgrounds and different angles for object detection independent of the environment. By reducing the size of the images to 416x416 pixels in order to shorten the training time, the program works with 173056 parameters instead of 360000 parameters for original size images), (2) **Labelling the Dataset** (creating text files in which the coordinates of each image are specified with YOLO format labelling process. During the process, unnecessary backgrounds were not included in the box that would cover the marked accessory in order to increase the possibility of object recognition), (3) **Integrating Google Drive into Google Colab** (exporting images and label files uploaded from Google Drive to Google Colab), (4) **Separating the Dataset as train and test** (random selection of images, 90% for train and 10% for pre-test. Train.txt and test.txt text documents containing image filenames were used), (5) **Integration of YOLO algorithm with Google Colab** (Darknet was installed to start the training by using the initialization coefficients found in the YOLO original system), and (6) **Realization of trainings on Google Colab** (the training was continued for 1000 iterations/3 hours and an error rate of 0.2 was reached. Also, by using OpenCV's Deep Neural Network Module, the number of parts detected in the images taken from the camera in real-time was printed on the image). In the test image showing the 7 pieces tested in the study, all 7 pieces could be detected correctly in 262.58 ms. In addition, in 25 trials performed with the camera, the parts in the image were determined 84% correctly. As a result, it was underlined that errors can be reduced (without the need for newly recruited workers to memorize accessory part codes and human intervention in counting) and therefore efficiency in production can be increased.

In the studies of Huang et al. (2021), which aimed to increase the accuracy of detecting the fluid level in the infusion bottle used in intravenous infusion, an image processing and deep learning-based system was developed. If the fluid level in the infusion bottle containing the serum is not detected in a time, when the liquid is depleted, blood can escape from the vein into the bottle by the effect of pressure, leading to swelling and pain at the injection site where the bottle is attached, and even the patient can go into shock. In the study, which draws attention to the fact that

the development of a system that can decide whether the fluid level is decreasing, will be extremely useful for both patients and healthcare professionals, two separate fluid level detection processes, which are based on **deep learning** ((1) real-time images of the infusion bottle are taken with the HD camera, (2) the bottle is detected in the image with the trained YOLOv3 target detection model, cropped according to the position size of the target frame and separated from the whole image, and (3) if the fluid finish line examined by the classification process is below a certain threshold value, the nurse is warned by means of an alarm, assuming “fluid is running out”), and **traditional image processing** ((1) the image goes through the stages of greyscale, filtering and pooling, (2) contouring is done with Canny edge and the fluid level is roughly determined with Projection positioning, and (3) if the fluid finish line is below a certain threshold value, the nurse is alerted by an alarm, assuming that the “fluid is running out”). The database used in the study was created from 280 images taken manually in a time-consuming manner from patients who received intravenous infusion treatment. For data augmentation, adding salt and pepper noise, adding gaussian noise, image blurring, perspective transformation, contrast enhancement and contrast reduction processes were applied respectively. In this way, a dataset of 1680 images was obtained and randomly selected 80% was reserved for train and 20% for pre-test. In the study, YOLOv3, R-CNN (Regions with CNN), Fast-RCNN and SURF algorithms were compared by using the dataset and it was seen that YOLOv3 gave better results (in terms of IoU (0.8943), Precision (0.9768), Recall (0.9688) and Test Time (2.9 s)) than the others. In addition, it was explained that the brand labels on the infusion bottle made the detection process difficult. According to the results of the study, it was emphasized that YOLOv3 showed the fluid level most clearly, while the lines obtained from other algorithms were blurred and dashed. It was stated that real-time fluid level detection from video would be provided within the scope of future studies. It was also suggested that the proposed system could also be used in other fluid level detection processes (milk bottle level, dam/lake level).

2.2.4.4 The Proposed System's Object Detection Phase

Vascular degeneration and locations in the classified images are detected in the Object Detection Phase. In this context, the YOLOv3 algorithm, which can perform object detection in a shorter time (according to SSD, R-CNN, Fast R-CNN, Faster R-CNN) and that can give better results in detecting small objects (compared to SSD, YOLOv1 and YOLOv2), has been preferred. In order to follow the progression of vascular degeneration, it is extremely important to detect the unusual shapes in the vessels together with their location. Artificially created degeneration on NIR fused-modified images of Period-1 are used in the training and pre-testing of the YOLOv3 Algorithm, and it forms the basis for introducing the vascular degeneration to be questioned to the system within the scope of object detection. For the purpose of degeneration detection, a certain degeneration pattern is artificially added at a certain location to the new sequential NIR fused-modified images obtained from the videos recorded by the user in subsequent periods to represent the progress. The belonging of the post-test patterns to the defined degeneration classes is determined by the YOLOv3 post-test process. NIR yolo-fused-modified images containing the information (bounding box showing the detected class name, the probability of belonging to the detected class and the relevant location) of the detected vascular degeneration are added as new period images to the video of the related tissue class in the Augmented Reality Phase.

In the Object Detection Phase of the proposed system, only the numbers of classes at the output of the Fully-Connected layer of the YOLOv3 Algorithm are changed and the algorithm is trained (in a web-based Google Colaboratory (Google Colaboratory, 2022) environment with remote GPU support) under transfer learning by using the vascular degeneration training dataset and pre-trained weights. As in the Classification Phase, pre-testing is also carried out simultaneously with the training. After the completion of the training and pre-test processes, the final object detection results are obtained with the post-test process.

The steps and explanations of the proposed system's Object Detection Phase are detailed in Table 2.19.

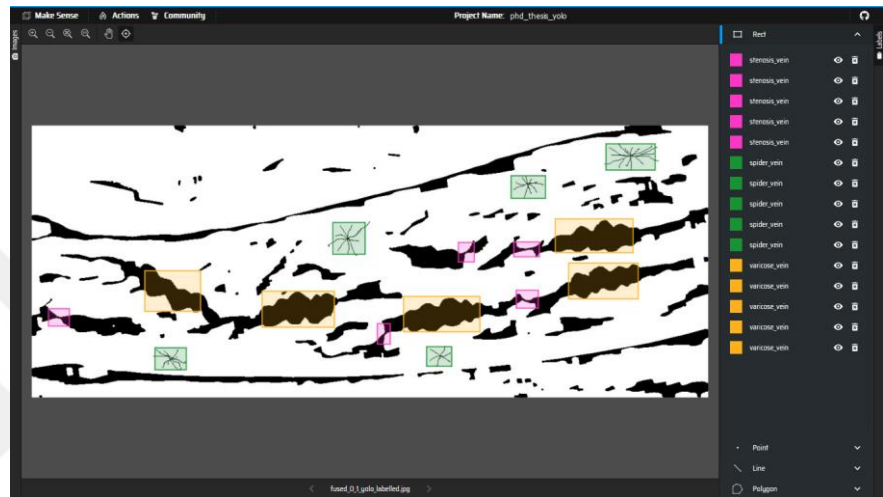
Table 2.19 Steps and explanations of the proposed system's Object Detection Phase

Object Detection Phase Step	Explanation
Preparing Dataset and Specifying Classes	<p>Since there is no publicly accessible vascular degeneration dataset obtained by NIR spectroscopy, a vascular degeneration dataset was prepared by adding artificial (contraction/enlargement) patterns on Period-1 NIR fused-modified images.</p> <p>The patterns were created in random rotations and scales, preserving a certain morphological structure. All images in the dataset consist of binary images in jpg format. The foreground (veins and patterns) are outlined in black.</p> <p>A total of 15 vascular degeneration artificial patterns (5 each of stenosis_vein, spider_vein and varicose_vein patterns) were added to each image. In this way, 2250 artificial degeneration patterns were obtained in the 150-image dataset. In addition, 50 new images obtained by data augmentation techniques (Rotation, Mirror, Noise and Scale operations, examples of which are shown in Table.20) were added to increase the detection accuracy. In this way, a dataset of 200 images (containing 3000 pattern samples) was created to be used in the trainings. Examples of vascular degeneration artificial patterns processed on images are shown in the Figures below.</p> <div data-bbox="730 1048 1197 1189" data-label="Image"> <p>The image contains three sub-images labeled (a), (b), and (c). Sub-image (a) shows a black, irregular, and somewhat jagged shape on a white background, representing a stenosis pattern. Sub-image (b) shows a black, star-like shape with multiple thin, radiating lines, representing a spider vein pattern. Sub-image (c) shows a black, irregular, and somewhat jagged shape on a white background, representing a varicose vein pattern.</p> </div> <p>Examples of vascular degeneration artificial patterns. (a) stenosis_vein, (b) spider_vein, and (c) varicose_vein</p> <p>Vascular narrowing (stenosis_vein) detection in this thesis study was introduced in Erdem & Utku's (2022b) study, and vascular dilation (spider_vein and varicose_vein) detection was introduced in Erdem & Utku's (2022a) study. In this study, multi-purpose use of the system within the scope of vascular degeneration detection was achieved by integrating both detections into the same system.</p>
Labelling	<p>While performing the classification process, the general pattern shapes of all defined classes should be introduced to the algorithms (during training and pre-testing).</p> <p>Traditional CNN Models operate on the assumption that the entire image belongs to a single class (classifying the image as a whole). In this context, only the categorization of images (which can be thought of as a simple variant of labelling) is sufficient to assign class names. In the Classification Phase of the study, the input data of the CNN Models are added under the relevant class folders (as right forearm and left forearm) and used in training and pre-test processes. The specific names of the image files are not taken into account in the CNN Models, only the folder names are processed.</p> <p>With object detection algorithms such as YOLO, one or more objects can be detected in an image. In addition, since a single image can contain more than one object, the information (locations) of these objects in the image should also be given to the algorithm during training and pre-testing. Since only categorizing the image with class information is not sufficient to define the object, labelling is used both in specifying the name and location. With labelling, which is a manual</p>

Table 2.19 Continues

process, all objects with defined classes in the image are marked with separate boxes one by one (by trying to include as much as possible only parts of the image that will describe the object of interest in the box), and their locations are indicated, and these boxes are labeled with the relevant class name.

In this study, data labelling was carried out with three classes named stenosis_vein, spider_vein and varicose_vein using makesense.ai (Skalski, 2022), a web-based application under the free-software (under General Public License version 3). Boxes were drawn to contain only the object as much as possible. The following Figure shows the labelling process of the image containing the vascular degeneration pattern using the makesene.ai web-based application.



Vascular degeneration labelling process (makesense.ai (Skalski, P., (2022) web-page screenshot).

In each image, a total of 15 artificial patterns were labeled, including 5 stenosis_vein, spider_vein and varicose_vein patterns. In this way, 3000 labeled objects representing vascular degeneration classes were obtained in a 200-image dataset.

As a result of the labelling process, an example of a label file suitable for YOLO format (containing the integer equivalent of the class and the coordinates of the objects), created for each image in its own name to be used in training and pre-test processes, is shown in the Figure below for 15 objects. The first value in the label file is the ID of the defined object classes, the second and third values are the values of the bounding box's center coordinates on the x-axis and y-axis divided by the width and height of the image (as normalized in the range [0, 1]), and the fourth and fifth values represent the values obtained by dividing the width and height of the bounding box by the width and height of the image (as normalized in the range [0, 1]) (Tiwari, 2022).

Table 2.19 Continues

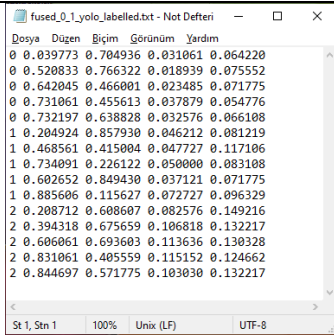
	 <p>Label file compatible with YOLO format</p>
Train and Pre-Test	<p>The Training and Pre-Test Step was carried out with the YOLOv3 algorithm. The algorithm was run on the web-based Google Colaboratory (Google Colaboratory, 2022) platform (providing free remote GPU support for deep learning studies (Çağıl & Yıldırım, 2020)) that works with the Google Drive storage infrastructure. Google Colaboratory (Colab for short) is a Jupyter Notebook-based (an open source, browser-based tool with various software languages and libraries) cloud service that provides an environment for machine learning and deep learning studies (Çağıl & Yıldırım, 2020). This step utilized the work of Tiwari (2022).</p> <p>The training of the YOLOv3 Algorithm was carried out within the scope of transfer learning. In other words, the number of layers of the DarkNet53 CNN Model, which is the backbone of the YOLOv3 Algorithm, were not changed (not a CNN Model architecture built from scratch) and the trainings were not started from scratch (the process of assigning random or fixed values to the filter weights/coefficients at the beginning of the training). A new training for vascular degeneration detection has been performed, using the weights obtained by pre-training the YOLOv3 Algorithm with the 80-class COCO dataset as initial coefficients for filters in the convolution layers (in the form of Transfer Learning). As in the Classification Phase, the pre-test (validation) process is carried out simultaneously with the training in the Object Detection Phase.</p> <p>In the trainings, the yolov3_new.cfg file containing the adjustments in the CNN Model architecture, obj.names (object names) and obj.data (containing the locations of the text files showing the train/pre-test belonging of the images and the number of classes) files containing the metadata are used. In the Train and Pre-Test Step;</p> <ul style="list-style-type: none"> • In the Darknet yolov3 config file, the classes value (representing the class numbers at the output of the Fully Connected layer) was changed to 3 according to the vascular degeneration class number (stenosis_vein, spider_vein and varicose_vein). • The filters value was calculated as 24 (according to the formula $(class_num+5)*3$). • Among the hyperparameter values, the Learning Rate was determined as 0.001, the Batch size was 64, and the Subdivision size was 16. • The max_batches value is calculated as 6000 (according to the formula $class_num*2000$). • The steps value is calculated as 4800 and 5400 (according to the formulas $max_batches*80\%$ and $max_batches*90\%$). • The images in the dataset were brought to the data sizes (416x416 pixel resolution) suitable for the YOLOv3 Algorithm.

Table 2.19 Continues

	<p>It is aimed to shorten the training times (by reducing the number of backpropagation and weight updates required in the learning process) by using pre-trained weights (in the context of transfer learning) instead of random values as filter coefficients (weights) in the convolution layers in training. In this way, random or fixed values are prevented from being assigned to the coefficients when the training first starts.</p> <p>The weight values obtained after every 100 epochs in the trainings are recorded in the yolov3_new_last.weights file. If there is a need for long-term training, training can be continued by using this file. The performance of the training is determined by the Average Loss value (which is the average value of how wrong the class estimation is) obtained for the most recently processed epoch (1 epoch takes about 10 seconds). The final weights obtained after the completion of the trainings are saved as the yolov3_new_final.weights file. In the study, the Avg Loss value could be reduced to 1.720302 at the end of the trainings that lasted for 16.5 hours and 6000 epochs.</p>
Post-Test	<p>The Post-Test Step was also carried out by running the YOLOv3 algorithm on the remote GPU using Google Colaboratory.</p> <p>In the Post-Test Step, a post-test dataset was prepared to be used for experiments. In the dataset of 45 images in total, there are 5 consecutive NIR fused-modified images for each period for each of the three different degeneration classes. Only a single pattern (patterns similar to the shape format used in training and pre-testing) was added on the consecutive NIR fused-modified images (images not used in training and pre-testing) of the Period-1, Period-2, and Period-3 left forearm. The artificial patterns superimposed on NIR fused-modified images were created to represent the degeneration progression process at beginner, intermediate and advanced levels.</p> <p>After the completion of the trainings, vascular degeneration detection results (as confidence values) and locations (marked with the bounding box) with the Post-Test Step were marked on the NIR fused-modified post-test images. In this context, the weights in the yolov3_new_final.weights file obtained at the end of the trainings were used.</p> <p>The results obtained are interpreted in the subheading 3.2 Experimental Results of the Object Detection Phase.</p>

Data augmentation techniques of the proposed system's Object Detection Phase are given in Table 2.20.

Table 2.20 Data augmentation techniques and explanations of the proposed system's Object Detection Phase



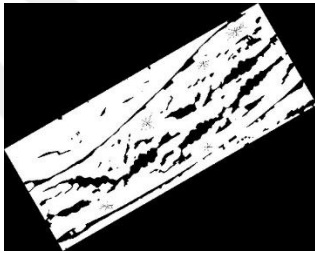



Data Augmentation Process	Explanation
Raw-image	<p>For data augmentation, binary NIR fused-modified images (517x1289x3 uint8) obtained by the Digital Image Post-Processing Phase and an example of which is shown in the Figure below, were used.</p>  <p>As representative, the binary NIR fused-modified image with the code Fused_Modified_Period1_Img0-1_RightArm.bmp is used as a representative</p>
Rotation By Angle/Mirroring	<p>In order to increase the number of data by including different formations (rotation by angle and mirroring) of existing shapes in the dataset, 10 and 30 degree rotation operations and matrix-row based mirroring operations (as it does not change the meaning of the vascular degeneration in the image) are used and examples are shown in the below Figures. (<i>imrotate</i> and <i>flip</i> commands are used.)</p> <div style="display: flex; justify-content: space-around; align-items: flex-end;"> <div style="text-align: center;">  <p>(a)</p> </div> <div style="text-align: center;">  <p>(b)</p> </div> </div> <p>Rotation by angle. (a) 10 degree rotation applied image, and (b) 30 degree rotation applied image</p> <div style="text-align: center;">  <p>Mirroring applied image</p> </div>
Noise Addition	<p>In order to tolerate the noise that may occur in the image in object detection, salt and pepper type noise with 0.1 intensity parameter was used as shown in the Figure below. (<i>imnoise</i> command is used)</p> <div style="text-align: center;">  <p>Noise (0.1 intensity value salt and pepper type) applied image</p> </div>

Table 2.20 Continues

Scale Change	<p>A downscaling of 0.8 was used in order to tolerate the reduction in vascular degeneration shape dimensions in the object detection as shown in the Figure below. (<i>imresize</i> command is used)</p>  <p>Scale change (0.8 zoom out valued) applied image (the size of the sample image is shown proportionally)</p>
---------------------	---

Summary of the proposed system's Object Detection Phase in 1-minute is given in Figure 2.14.

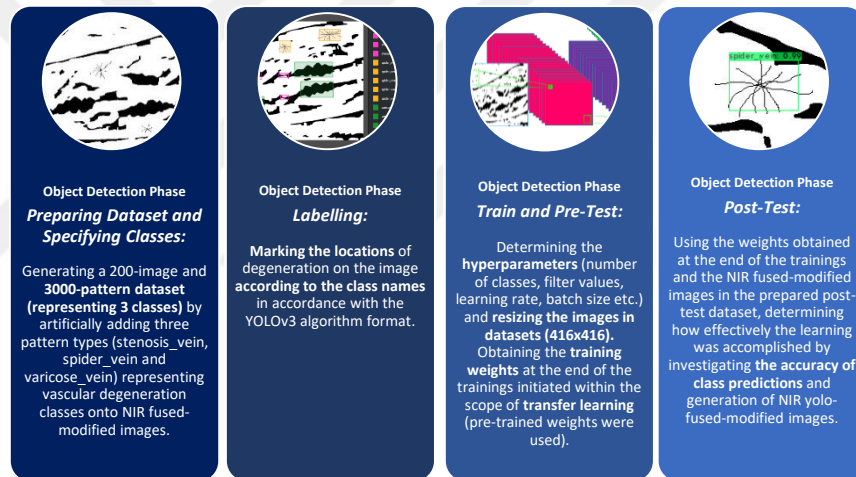


Figure 2.14 The proposed system's Object Detection Phase in 1-minute

2.2.5 The Augmented Reality Phase: Winding the Time Fast

Detection of the rate of progression of diseases plays a critical role in treatment as much as early detection of diseases. By determining the rate of progression of the disease and adapting the treatment according to the new situation, the chance of success of the treatment can be increased. For most detected diseases, the follow-up of the progress is carried out with routine controls in hospitals or clinics. In addition, different imaging techniques or medical analyses are used according to the type of disease. The results obtained from these devices and analyses are kept in an archive

of the patient as image or text-based records (including the information of the relevant diagnosis and the drugs used, if any). With the developing automation systems, these records belonging to the patient in a certain hospital or health institution are transferred to the e-pulse Personal Health System archives of the T.R. Ministry of Health. Both the patient himself/herself and his/her doctor, or when he/she applies to another health institution, the new doctor has the opportunity to access these records retrospectively. Although such systems, which have found widespread use especially in recent years, provide information about disease progression, recorded results generally need to be evaluated from a perspective that requires professional expertise. In some cases, the patient may not realize the seriousness of the situation and delay his/her treatment and follow-up because he/she cannot fully interpret the data. As time progresses, this may make the treatment more difficult. What is important is which data we will learn from the elapsed time in order to increase the chance of success of the treatment.

In this study, the follow-up of the progression of vascular degeneration was carried out by integrating the data obtained by NIR imaging systems and deep learning methods into the VE.

2.2.5.1 Mixed Reality

***Virtual Reality** can be shortly defined as a generated environment or technology which is three dimensionally formed with all sorts of existing/non-existent creatures, objects or places designed in computer environment similar to their origins and which is perceived as if it was real (Erdem, 2013, pg. 7).*

The term Virtual Reality aroused a lot of expectations, but it still could not fully meet the expectations due to the inability of the technological infrastructure to reach the desired level (Gutierrez, Vexo, & Thalmann, 2008; Mazuryk & Gervautz, 1996). Therefore, instead of this term, **Virtual Environments** denoting an environment consisting entirely of virtual material in which reality is presented at the level of

acceptable copies (Kantarıcı & Çakır, 1998; Gutierrez, Vexo, & Thalmann, 2008) is more preferred, especially in the academic field. (Erdem, 2013)

Although VEs are called virtual, their most basic goal is to create a perception of reality. There are 4-main criteria for virtual experience to create a perception of reality: **immersion** (covering the user's sensory areas with virtual material), **interactivity** (manipulation of the VE by the user), **sense of presence** (the user feels himself/herself in the VE), and **real-time** (simultaneous execution of feedback provided for the interaction) (Erdem, 2013). Depending on which area the application addresses, some criteria can be stretched to a point. While some applications (game industry) determine perception of reality as the most basic target, others focus on content (education industry) and ignore it (Erdem, 2013). For example, while a few seconds of delay might be reasonable in a VE that offers an educational application for primary schools, a millisecond of delay would be vital in an application that presents information to the surgeon in visualizing vessels during surgery or in repairing a space shuttle in outer space.

Since the technologies that create VE are still in the development stage, it is still not clear how some concepts in the terminology can be brought together.

Although a fully valid terminology for virtual concepts is still not established, Milgram's Reality - Virtuality Continuum (Milgram, Takemura, Utsumi, & Kishino, 1995) (simplified form of the continuity of the metaverse, which expresses a shared virtual space (Peddie, 2017)) shown in Figure 2.15 is the cornerstone in expressing these concepts in general. In this continuity, there is no concept of virtual reality. Instead, the concepts are discussed under the title of **Mixed (Hybrid) Reality** (or more precisely, Mixed Environment), indicating a certain mixture of VE and RE (Erdem, 2013).

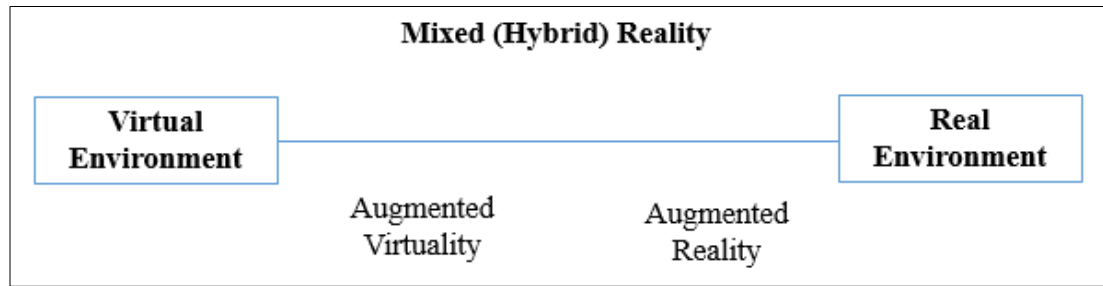


Figure 2.15 Milgram's terms of the reality-virtual continuum (Milgram, Takemura, Utsumi, & Kishino, 1995)

In the continuum, a naming has been made according to the ratio of virtual and real materials in the environment as mixed. Accordingly, at the ends of the continuum, there is a completely VE and a completely RE, while the concept of **Augmented Virtuality** is mentioned if the virtual material is denser in the environment, and **Augmented Reality**, if the real material is denser. Both of these concepts are within the scope of Mixed Environment. (Billinghurst, Clark, & Lee, 2015; Carmigniani & Furht, 2011)

Mixed Reality offers a perception of reality in which real-world objects and virtual objects are positioned as co-located (when approaching the object in the RE, the virtual object grows simultaneously, and its virtual counterpart can be manipulated by interacting with real objects), and in this way, both environments can coexist interactively (Peddie, 2017).

Today, touch screens, keyboard and mouse are widely used as data entry methods for interacting with materials in VEs, while natural user interfaces (hand, head, eye movement tracking or interaction with sound) are among the newly developed methods. Helmets, head-up displays, smart glasses, smartphones or monitors are used as output devices. (Peddie, 2017)

2.2.5.2 Indirect Augmented Reality

In VE (containing entirely artificial/virtual material (Craig, 2013)) applications, while the user is isolated from the real world, in AR, the real world is tried to be

detailed without being surrounded (as non-immersive) by virtual materials. In other words, while the goal in VE is to change reality, in AR, the goal is to increase reality. While the real world is replaced by an artificially simulated environment in the VE, the user is isolated from the RE (virtually created world replaces the real world) by completely covering his/her field of view with the virtual material displayed by the head mounted display. In AR, on the other hand, a real-time superimposition of virtual materials on specific locations on the real world are in question. The user can see the outer world. In short, while the user is taken out of the real world with the VE, the perception of reality is expanded in AR. (Billinghurst, Clark, & Lee, 2015; Peddie, 2017)

AR is like a visual system that enriches the sense of reality by developing and expanding what we see in the real world, by combining the RE and virtual materials (various digital information such as text, drawings, figures, 2D/3D visuals, audio, animation or step-by-step instruction videos) in such a way that increases the perception and interaction of the user (Alves, Marques, Ferreira, Paulo, & Beatriz, 2022; Carmigniani & Furht, 2011; Peddie, 2017). It is one of the latest developments in human and computer interaction, which aims to increase the user's real-world experience by making computer interfaces invisible (Billinghurst, Clark, & Lee, 2015). Moreover, by combining data from artificial intelligence, big data or sensors, it offers a brand new experience of reality that blurs the boundaries between the digital and physical worlds. By superimposing virtual materials on real-time images, an augmented and detailed real-time view of the physical world directly or indirectly (live-video stream) is obtained and projected into the user's field of view (Carmigniani & Furht, 2011; Peddie, 2017).

Although AR is one of the trendiest technologies of recent years, practical applications in terms of concept have been in use since the 17th century (Billinghurst, Clark, & Lee, 2015). The first practical application that can be considered within the scope of AR concept is Pepper's Ghost technique (an illusion technique used for entertainment purposes by using glass and light to project the image of an object or person in another room, usually a person dressed in a white

sheet, which is used as a ghost to attract attention, through the glass in the other room), which dates back to 1862 (Peddie, 2017).

Today, a wide variety of applications can be designed by using AR systems, such as engineering (in design and modelling), medicine (in anatomy education or visualization of medical data in the diagnosis, follow-up and treatment of diseases), education (as course materials or visualizing 3D concepts), military (in special trainings within the scope of equipment/vehicle use or battlefield) maintenance-repair (with step-by-step guidance as remote or checklist) and entertainment (in integrating computer games with the real world and providing a 3D movie experience).

As in VE, AR studies are designed as application-specific (Peddie, 2017). If the purpose of the application is to provide visuals, the use of hardware that requires high processing power is at the forefront, while devices with lower hardware specifications can be used when the purpose is to provide information about the relevant environment only (Liestoel, 2019).

The AR environment is presented to the user **without blocking the live view of the real world** (indoor/outdoor smart glasses, contact lenses, helmets, head-up displays, projectors etc. are used for visual see-through transparent AR), **by blocking the real world view** (head mounted display is used for video see-through or obstructed view AR. Augmented images are superimposed on video image), or **by reflecting it to the real world** (for projected AR, the virtual material is projected onto real-world objects). The devices used are divided into two as **wearable** (helmets, smart glasses, or headphones) and **non-wearable** (stationary devices - PC, TV, projector; mobile devices - smart phones, notebooks, tablets and head-up display devices). (Peddie, 2017)

VEs presented via screen or projector, although not fully immersive (i.e. a complete perception of reality cannot be created), are perfect for a glasses-free and

hands-free experience for some special occasions (education, medical treatment, maintenance-repair etc.).

It is thought that the importance of smartphones in our daily lives will also be in question for AR in the coming years. Until very recently, it is predicted that it will be an ordinary routine to experience the real world in daily life, at work, in shopping malls with AR glasses (perhaps as contact lenses), like using prescription glasses or sunglasses. In this context, VEs create a jumping-off point for AR, which is the main target, and lay the groundwork for adaptation to daily use. (Peddie, 2017)

AR is a real-time interactive environment where digital information (virtual materials) is added to the physical world by being recorded spatially and temporally (which can be manipulated simultaneously with the interaction). The basic criteria for AR can be summarized in 3-ways: (1) Augmenting the physical world view with digital materials (information, text, graphic, figure, video) superimposed on it, (2) Displaying the virtual material as registered to the physical world (rotating, scaling, manipulation in terms of viewpoint simultaneously with interaction), and (3) Positioning the virtual material in relation to the physical world. (Craig, 2013)

In AR studies, the virtual material needs to be registered (aligned) on real images in real-time and accurately, corresponding to its location in the real world (within a 3D coordinate plane). In addition, it is expected that virtual materials will be modified according to ambient lighting in order to increase the perception of AR. Alignment distortions (registration errors), which harm the AR experience occur, especially due to insufficient computational capacity of mobile devices or camera shakes in real live streaming images. With IAR, alignments can be ensured to be one-to-one with virtual materials (pre-recorded video footage is used in this thesis) superimposed offline to pre-recorded images. When the user views the relevant location in the live environment, the images (pre-recorded images with superimposed real and virtual material) aligned with zero errors in offline registration are visualized by roughly aligned in online registration. (Akaguma, Okura, Sato, & Yokoya, 2013)

The term IAR, introduced in the literature for outdoor, creates an AR environment by using videos in which pre-recorded panoramic images are combined with virtual material in applications where a tight spatial alignment/overlap between virtual material and real material is required (Wither, Tsai, & Azuma, 2011). With **offline registration** in IAR, virtual objects are superimposed on pre-taken panoramic omnidirectional images. Then, during **online registration**, these offline images are aligned according to the roughly estimated orientation on the live images to which they will correspond (inconsistencies that may occur are not critical). In this way, a precise match in geometric and photometric terms is ensured in offline images. Alignment (registration) in IAR is done precisely in offline superimposition, while it is done roughly in online superimposition. This may cause spatial and temporal inconsistency during online transaction. (Okura, Akaguma, Sato, & Yokoya, 2014)

IAR is used as a concept focused on outdoor, especially targeting mobile devices with low processing capacity (Okura, Akaguma, Sato, & Yokoya, 2014; Wither, Tsai, & Azuma, 2011). Advantages and disadvantages of IAR over traditional AR (where virtual materials are superimposed on real-time acquired images (Okura, Akaguma, Sato, & Yokoya, 2014)) (Alves, Marques, Ferreira, Paulo, & Beatriz, 2022; Liestoel, 2019; Okura, Akaguma, Sato, & Yokoya, 2014; Wither, Tsai, & Azuma, 2011):

- It is especially important if a tight spatial alignment/overlap with pixel precision is required between the real image and the virtual material (Wither, Tsai, & Azuma, 2011).
- In the IAR environment that the user experiences, he/she observes not live real-world images as in video-see-through, but observes **previously-recorded images** from the same real-world region (the real image is enhanced by superimposing with virtual materials and represented as video). In this way, the registration errors (the inability to obtain the alignment/overlap/position information due to shake and delay causes a shift between the real image and the virtual material) that may occur between the real image and the virtual material in superimposed images are reset. (Wither, Tsai, & Azuma, 2011)

- It provides **high accuracy** for **registration (alignment) errors** between live real image and superimposed image (the registration error observed in IAR only occurs between the device and its environment in a way that does not disturb the AR experience much) (Wither, Tsai, & Azuma, 2011).
- IAR is effective in **minimizing visual problems (alignment)** caused by the underlying tracking system (sensor, tissue matching processes) (Wither, Tsai, & Azuma, 2011).
- When **complexity and details** (if using multiple virtual components to be matched or attached to specific locations based on alignment, or if data from multiple cameras will be processed to create a particular panoramic live view) increase in real world images, IAR can produce very successful results in alignment by reducing **the computation and processing load** (Wither, Tsai, & Azuma, 2011).
- With IAR, a visualization can be presented within the scope of **before and after**, the **similarities and differences** between these situations can be examined, and **time-dependent changes** can be visualized to the user (Liestol, 2019).
- IAR is **cost-effective** as it does not require additional hardware (glasses, head mounted displays etc.) to visualize the environment (Alves, Marques, Ferreira, Paulo, & Beatriz, 2022).
- IAR is **dependent on the environment** from which offline images are taken. In online process, spatial inconsistencies may occur when the user wants to use the system from a different location or point of view. If live images differ from pre-taken images due to various factors (weather, lighting, day-night), this time temporal inconsistencies (which can distort the perception of reality more than spatial inconsistency) may be encountered. (Okura, Akaguma, Sato, & Yokoya, 2014)

In the Augmented Reality Phase of this thesis, the results of vascular degeneration are returned to the user and physician in video form. The aim here is to match the tissue region in the video (showing the results of the YOLOv3 algorithm) with the

associated vascular degeneration exactly, not with the real-time displayed tissue region. In this context, IAR is used in the study.

2.2.5.3 Literature Review of the Indirect Augmented Reality

Although there are numerous studies in the literature for AR, IAR is one of the least studied areas in the literature (Alves, Marques, Ferreira, Paulo, & Beatriz, 2022). The concept has been used in outdoor applications, including those who introduced the term IAR to the literature (Wither, Tsai, & Azuma, 2011).

In the study of Wither, Tsai, and Azuma (2011) a method using the IAR concept was proposed, stating that image processing-based (using defined patterns) and sensor-based (using compass and GPS data) tracking methods used in precision tracking showed poor tracking performance in an AR application to be prepared for mobile devices and outdoor environments. In this context, live-images used for video-see through AR were replaced with previously-recorded panoramic images within the scope of the proposed IAR method in cases where a perfect overlap between virtual material and real world is required. In the study, it is reported that for a panoramic scene viewed from a distance of 20m, an orientation error of 5 degrees will cause a registration error of 1.75m. For this, it is underlined that the IAR approach will provide a perfect alignment with the simultaneous movement of the previously-recorded real image and virtual material, and also prevent tracking errors due to jitter and delay. The study was tested with two case studies. In the first, nine users were asked to locate certain marked buildings in the real world (visible from their location) in the panorama images from different viewpoints shown to them, and their response times were recorded. According to the results, it was determined that the position and orientation of the image is important in finding the target. In the second, it was investigated with 18 people whether the IAR creates a traditional AR experience (immersion level) using different situations (cases where the panorama does not belong the user's location or sunny/cloudy weather). According to the results, it was stated that there was no significant difference between IAR and traditional AR, IAR was more preferred, but when the user's position and the

panorama were different, IAR reduced immersion. While a general AR experience was provided in the study, it was observed that the quality of follow-up was improved. Also, it was noted that IAR was a useful method in most outdoor applications.

In Liestoel's (2019) study, it was aimed to integrate the storytelling of the Omaha Beach landing (D-Day) during the Second World War into the IAR environment, enriching the experience on a location-based basis. In this context, the first hour of the D-Day was visualized, in which videos of battle scenes recreated with 3D graphics in the audiovisual narrative flow were shared. The system was designed as a mobile application and when it came to the relevant location (in situ) where the attack took place, the panoramic video image of the war moment from the user's perspective (the application also includes documents, images and audio) was displayed on the mobile device. In the experiments of the study carried out in the relevant location, since it was determined that the majority of the users activate the application when they were not in the spatial ROI (narrative starting point) (in this context, an additional user tracking application has been developed), information about the usage of the system was included in the application splash screen. It was also ensured that the application could only start if users were within 200 meters of the zone of interest for a more error-free experience.

Akaguma, Okura, Sato, and Yokoya's (2013) studies aimed to solve the alignment problem between virtual and real materials realized in real-time in mobile AR images. In this context, the study was composed of two separate parts as: **offline** (pre-superimposed alignment of virtual materials with real multi-directional images captured from fixed positions and then rendering for lighting) and **online** (alignment of superimposed images and live images from the user's perspective using position and orientation data of the device using GPS, compass, gyroscope and accelerometer sensors). Virtual materials were superimposed offline with precise alignment on pre-recorded multi-directional images and roughly aligned on live images online when the relevant location was viewed by the user. In this way, a perfectly aligned visuality, both geometric (spatial) and photometric (illumination), compatible with

mobile device capacities, can be created offline. In the trials of the study, Todai-ji Temple (Japan) was displayed in its real location and visualized as a virtual historic experience within the scope of mobile AR. Virtual sculpture models modeled on computer on omnidirectional images were superimposed offline. The AR mode of the system was activated when users came to the relevant checkpoints (when they were within 2-3 meters) while walking around the Hall of the Great Buddha with the navigation map containing 5 separate checkpoints in the application. A group of 46 participants evaluated the usefulness of the AR application with a 4-question questionnaire. According to the results, pre-captured omnidirectional images were evaluated and 90% of the participants stated that the historical experience was useful, 70% stated that they found the alignment in offline and online parts natural, while 40% stated that they did not find the illumination of virtual objects natural compared to the RE. It was reported that within the scope of future studies, pre-recorded images would be updated depending on the weather and time conditions.

In the studies of Okura, Akaguma, Sato and Yokoya (2014), it was aimed to eliminate the temporal inconsistencies that may occur due to lighting during online registration. In this context, the superimposed versions of omnidirectional video images (taken for 10 seconds from a fixed position) taken for different ambient lighting conditions as offline operation with virtual materials and the histograms of these superimpose images were recorded in the dataset. In the system, the most suitable image selection for the conditions of the live environment among different ambient lighting conditions for online registration was carried out by looking at the histogram similarity. Perspective adjustment was performed on the selected image and presented to the user. In the trials of the study, Todai-ji Temple (Japan) was displayed in its real location and visualized within the scope of a virtual historic experience within the scope of mobile AR. A group of 87 participants evaluated the usefulness of the AR application with a questionnaire. The results obtained from the trials of the proposed system showed that the method and the aesthetics of the scene could increase the realism in the IAR system. Within the scope of future studies, it was explained that instead of manually recording various lighting situations, it was considered to take images with the relighting technique.

In the studies of Alves, Marques, Ferreira, Paulo and Beatriz (2022), the effectiveness of three different AR-based visualization methods (handheld mobile AR, indirect AR projection on a monitor and AR using optical see-through head mounted display) in the assembly process was investigated. In this context in order to visualize the LEGO building blocks' assembly steps to the user, 3-methods were proposed as: (1) handheld mobile AR method via marker, (2) indirect AR projection on a monitor method via overhead view and marker, and (3) AR using optical see-through head mounted display method via virtual helmet. In the trials of the study, which was carried out with 30 people, users were asked to assemble 18 LEGO pieces using 3D step-by-step instructions with all three AR methods. Performances were evaluated in terms of assembly completion time (in seconds), number of errors, and number of errors corrected. According to the test results, all three methods were found to be sufficient in terms of performance and user preference. 14 of the users stated that they prefer to use IAR, 13 of them HMD AR and 3 of them prefer to use mobile AR. Within the scope of future studies, it was stated that an automatic verification option would be added to the systems in order to provide instant feedback to users.

In the study of Ai et al. (2016), which aimed to facilitate intravenous injection, a traditional AR-based system was developed in which 3D reconstructed vein patterns were superimposed on the tissues from which the images were taken. The low-contrast images obtained were segmented using the multi-feature segmentation method. The resulting vascular structures were matched and recreated in 3D and superimposed on the upper surface of the hand. It was determined that the performance of the system, in which real vein images are evaluated in terms of segmentation, vessel matching, vessel reconstruction, and enhanced display, gives effective results (in terms of providing high-speed quality imaging and AR experience). According to the results obtained, it was stated that the vein and skin surface reconstructed in 3D could be superimposed on the hand surface with high precision. In this way, it was underlined that the system can increase the success rate and efficiency of intravenous injection.

2.2.5.4 *The Proposed System's Augmented Reality Phase*

The system developed in this study offers a vascular degeneration follow-up that can be monitored both on the mobile device (by the user) and on the PC screen (by the doctor). However, in this study, a traditional AR application was not used as it was aimed for a long-term follow-up process by visualizing the ROI with certain periods (day, week, month, or year) instead of snapshots. For this, an IAR environment was created in the proposed system by using videos in which the vein patterns were superimposed on the images recorded at certain periods. In this context, the actual content was augmented in the form of video-based images. Although a high level of reality perception was not created in the study, the visualization of a wide time interval was carried out within the scope of IAR.

The system proposed in the thesis study is evaluated within the scope of IAR. In addition, in tracking vascular degeneration, the system can also be included in the scope of **Annotation Focused AR** by adding the information of the detected YOLO results on the image. In annotation focused AR systems, although the presence is not very critical, it is expected that the virtual material will be placed in the AR scene convincingly (in terms of spatial, morphological, dimensional, light) in terms of ensuring believability (Wither, Tsai, & Azuma, 2011). In addition, the developed system can also be evaluated within the scope of **Markerless AR** (the type of AR that uses only sensor data or object detection algorithms without using the Quick Response-QR code used for mapping the virtual material to the real-world location).

The steps and explanations of the proposed system's Augmented Reality Phase are detailed in Table 2.21.

Table 2.21 Steps and explanations of the proposed system's Augmented Reality Phase

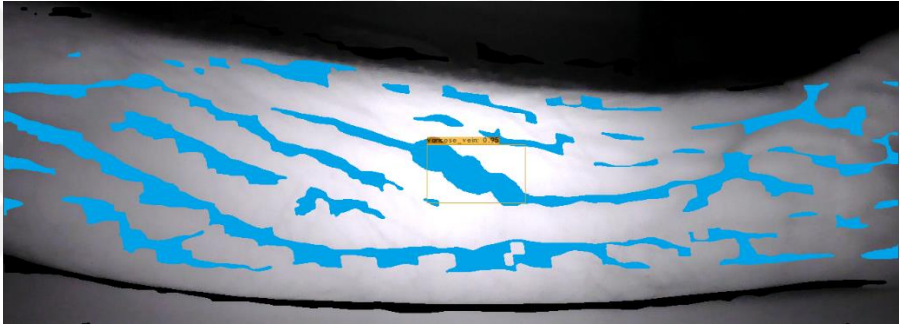
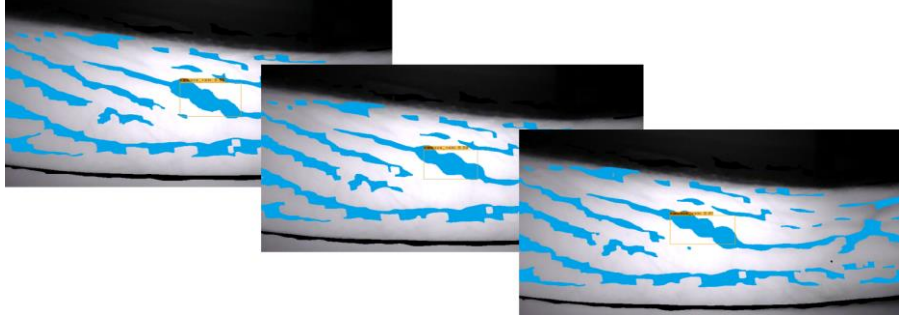

Augmented Reality Phase Step	Explanation
Offline Registration	<p>The NIR yolo-fused-modified images obtained in the Object Detection Phase are composed of the combination of two NIR fused-modified images. Therefore, these images (yolo-fused-modified-01) containing the results of vascular degeneration are superimposed on the raw form (raw-image) of the first of the two images, of which it is the corresponding combination, as in the Figure below (an IAR video of approximately 17.5 seconds can be created with 525 image frames from a total of 1050 raw image frames obtained from the 35-second NIR video recording). This ensures that the vascular degeneration is aligned with zero error at the corresponding position on the raw-image. In the Offline Registration Step, all yolo-fused-modified images of the relevant period are subjected to offline registration and superimpose images are obtained. In order to accurately visualize the location of the degeneration, this step requires a very tight match (offline registration) between the virtual material (yolo-fused-modified images) and the real material (NIR raw-images).</p>  <p>Superimposed image obtained as a result of offline registration</p>
Creating the Video	<p>In the Creating the Video Step, a video containing the results of vascular degeneration is prepared to be presented to both the user and his/her doctor. First of all, the results of Period-1 are added to the video. Since two classes (right and left forearm) are used by default in the thesis, two separate videos of Period-1 are created. As the process progresses, all images obtained in the new periods are sequentially added after the relevant Period-1 images according to their classes determined in the Classification Phase. One of the reasons why the user is asked to record video in the Imaging Technique Phase is that video footage of a certain length (approximately 17.5-second video of IAR offline registration can be created from 35-second video of NIR) can be created in this step. In this way, a video recording is created that includes the vascular degeneration results of all periods belonging to related tissue region, covering a certain time interval from Period-1.</p>  <p>Frames of the video image created with offline registration images</p>

Table 2.21 Continues

<p>Online Registration</p>	<p>The presentation of the video recording created from the images obtained within the scope of offline registration to the user is realized by recognizing the relevant tissue region in live view. In order for the Online Registration Step to be used, the tissue region whose results are to be examined must have been previously-viewed for at least one period and introduced to the system. The user is expected to perform a new NIR imaging to see the vascular degeneration results of the tissue that he/she has introduced to the system during the Imaging Technique Phase. This time, a 5-second viewing is sufficient as it will only be used for identification. New NIR video recording undergoes the Digital Image Pre-Processing and the Post-Processing Phases, resulting in new NIR fused-modified images. Of the 150 image frames obtained, 30 randomly corresponding to different durations are used to determine the tissue class of the Classification Phase Post-Test Step. At this phase, if it is verified that the vascular patterns found in the images are registered in the database, the offline video (including vascular degeneration results) of the determined class is superimposed in real-time, as shown in the Figure below, within the scope of online registration with a roughly positioning on related tissue region. In this step, there is no strict alignment requirement between the live view and the superimposed offline video.</p>  <p>Superimposed image obtained as a result of online registration.</p>
-----------------------------------	--

Summary of the proposed system's Augmented Reality Phase in 1-minute is given in Figure 2.16.

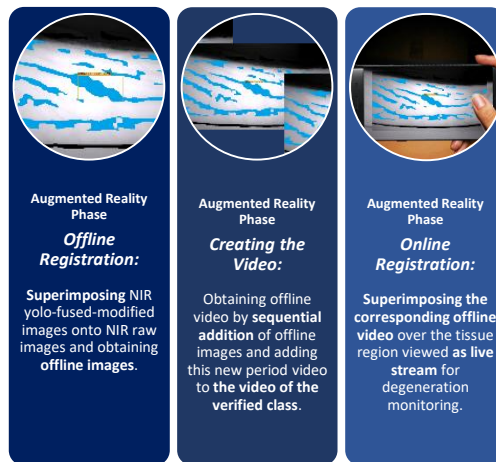


Figure 2.16 The proposed system's Augmented Reality Phase in 1-minute

CHAPTER THREE

EXPERIMENTAL RESULTS

In the Third Chapter, the experimental results obtained are interpreted in two sub-headings as “Classification Phase Experimental Results” in which Hybrid Decision-Making Algorithm is used and “Object Detection Phase Experimental Results”.

3.1 Experimental Results of the Classification Phase

The first step to detect and follow vascular degeneration is to determine whether the same tissue area is imaged. In this context, the Classification Phase was used in this thesis study. At this phase, three separate datasets containing different data numbers were created with NIR fused-modified images obtained from the user's video recording to be used in three different CNN Models proposed. In the study, the contribution of different image sizes to the classification accuracy was examined by using CNN Models with three different input sizes (28x28, 64x64 and 128x128). In addition, the contribution of different image numbers (including 500, 1000 and 1800 images in total) to the classification accuracy was investigated by using three datasets. Among those targeted at this phase within the scope of classifying the tissue region;

- Reducing the training time with CNN Models with fewer layers,
- Achieving desired classification accuracy levels with less training/pre-test data and,
- Being able to perform a more accurate class determination with the Hybrid Decision-Making Algorithm (this algorithm was first introduced in Erdem & Utku's (2022b) study) without relying on a single model.
- In addition, the effects of different hyperparameters (learning rates 0.01, 0.001 and 0.0001) and different image types (NIR raw-image, NIR cropped-raw image and NIR fused-modified image) on CNN Model performance are also examined. However, in the experiments carried out, the desired accuracy levels were achieved when only the fused-modified images were used and the

learning rate was 0.0001. For this reason, the post-test were carried out with the NIR fused-modified images and the learning rate value of 0.0001.

The proposed CNN Models are designed to be able to recognize only 2 different classes. Accordingly, in Period-1, imaging of the right and left forearms was preferred in terms of both less subcutaneous adipose tissue and practicality within the scope of image acquisition. The trainings were started by creating datasets as described in Table 2.15 Preparing Datasets Step with the NIR fused-modified images obtained as a result of the Digital Image-Pre-Processing and Post-Processing Phases from the video recording uploaded by the user to the server.

The 20 filters and sizes of the first layers of CNN Models with 20 depths used in the trainings are shown in Figure 3.1, and the objective function and top1error result graphs obtained after the completion of the trainings are shown in Figure 3.2.

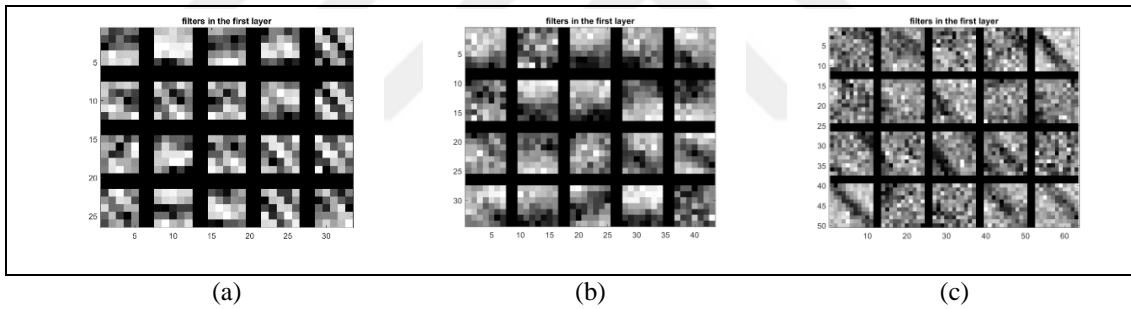


Figure 3.1 20 filters and sizes of the 20 depth first layers of CNN Models using Dataset-1. (a) 5x5 size filter used in CNN Model-1, (b) 7x7 size filter used in CNN Model-2, and (c) 11x11 size filter used in CNN Model-3

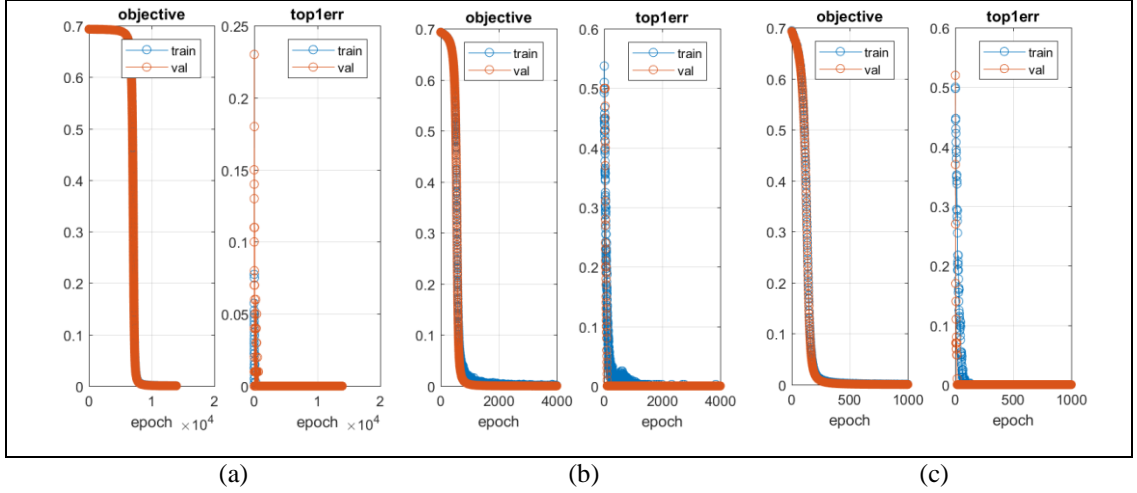


Figure 3.2 Objective function and top1error result graphs of CNN Models using Dataset-1. (a) CNN Model-1 training graph completed at 14,000 epochs, (b) CNN Model-2 training graph completed at 4,000 epochs, and (c) CNN Model-3 training graph completed at 1,000 epochs

After the trainings were completed, the images used in the post-test were selected as described in Table 2.15 Post-Test Step. Post-test results obtained using these images (predicted classes and prediction values of images) are given in Appendix-A for all periods. The summary of Appendix-A in terms of prediction count, prediction average, predicted first-class, predicted mid-class and predicted ultimate-class (determined according to the result of the Hybrid Decision-Making Algorithm) is given in Table 3.1.

When the results in Table 3.1 are examined within the scope of the **predicted first-class results**;

- First-class prediction was performed correctly in all CNN Model experiments for Period-1 right forearm, with high average. When Dataset-1 was used, the CNN Model-1 first-class prediction for left forearm results was incorrect, while the CNN Model-2 first-class prediction was correct, but the number of incorrect predictions is quite high.
- When Dataset-3 was used in Period-2, CNN Model-3's first-class prediction for right forearm results was correct, although it contained erroneous post-test results. When using Dataset-1, CNN Model-1 and CNN Model-2 performed erroneous first-class estimation with a very high number of erroneous

predictions for left forearm outcomes (this indicates that using only Dataset-1 for final-class estimation will result in misclassification).

- In addition, when Dataset-2 was used in Period-2, it was observed that there were a few incorrect predictions in CNN Model-2 and CNN Model-3 for the left forearm, but it performed the first-class prediction correctly, and a similar situation was also valid for the left forearm using Dataset-3.
- For the left forearm using Dataset-1 in Period-3, CNN Model-1 and CNN Model-2 performed incorrect first-class predictions with a very high number of incorrect predictions (this situation, which is also encountered in Period-2 left forearm, shows that using only Dataset-1 for final-class estimation will lead to misclassification).
- In addition, when Dataset-2 is used in Period-3, it was determined that there was an incorrect prediction for the left forearm in CNN Model-1 and CNN Model-2, but it performed the first-class prediction correctly.

Table 3.1 Post-test results for three separate periods according to the datasets and CNN Models of the proposed system's Classification Phase

		Dataset-1				Dataset-2				Dataset-3			
Train (Right Forearm / Left Forearm):		200 / 200				400 / 400				720 / 720 (includes noise/rot5/rot10/scale)			
Pre-test (Right Forearm / Left Forearm):		50 / 50				100 / 100				180 / 180 (includes noise/rot5/rot10/scale)			
CNN Model:		Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-1	Model-2	Model-3	Model-3
Input Image Size:		28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	28x28	64x64	128x128	128x128
Learning Rate:		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Epoch:		14000	4000	1000	10204	3459	800	10204	3459	1000	1000	400	400
Training Duration (sec / 1 epoch):		1.458123 sec	3.062919 sec	38.534514 sec	1.955146 sec	4.664262 sec	75.446255 sec	1.955146 sec	4.664262 sec	3.355219 sec	18.929039 sec	147.922873 sec	147.922873 sec

Actual Class		Predicted Class				Predicted Class				Predicted Class			
		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm
PERIOD-1	Right Forearm	Prediction Count	21	0	21	0	21	0	21	0	21	0	21
		Prediction Average	1.0000	0	0.9999	0	0.9999	0	0.9999	0	1.0000	0	1.0000
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	1.0000	0
	Predicted Mid-Class	Right Forearm	0	0.9573	0	0.9919	0	0.9999	0	0.9999	0	1.0000	0
PERIOD-2	Left Forearm	Prediction Count	11	10	10	11	0	21	0	21	0	21	0
		Prediction Average	0.8230	0.8870	0.7452	0.8784	0	0.9997	0	0.9866	0	0.8559	0
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9997	0	0.9999	0	0.9999	0
	Predicted Mid-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9997	0	0.9999	0	0.9999	0
PERIOD-3	Right Forearm	Prediction Count	21	0	21	0	21	0	21	0	21	0	21
		Prediction Average	0.9940	0	0.9814	0	0.9999	0	0.9934	0	0.9993	0	0.9304
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9993	0	0.9999	0
	Predicted Mid-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9993	0	0.9999	0
PERIOD-3	Left Forearm	Prediction Count	21	0	21	0	21	0	21	0	21	0	21
		Prediction Average	0.9980	0	0.9874	0	0	0.9928	0	0.9730	0	0.9979	0
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0
	Predicted Mid-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0
PERIOD-3	Right Forearm	Prediction Count	21	0	21	0	21	0	21	0	21	0	21
		Prediction Average	0.9980	0	0.9874	0	0	0.9928	0	0.9730	0	0.9979	0
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0
	Predicted Mid-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0
PERIOD-3	Left Forearm	Prediction Count	21	0	21	0	21	0	21	0	21	0	21
		Prediction Average	0.9780	0	0.9987	0	0	0.9481	0	0.9530	0	0.9992	0
	Predicted First-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0
	Predicted Mid-Class	Right Forearm	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0	0.9999	0

Inaccurate estimations made in the predicted first-class context show that situations can be encountered in e-Health where a single dataset (consisting of video images that may contain lighting and imaging deficiencies as a result of being supplied with low-cost cameras) or a single CNN Model obtained from the ordinary user is not effective in making the ultimate-decision alone. This poses a risk of undermining the reliability of the proposed vascular imaging system. The inclusion of vascular degeneration to be detected in the right forearm, as a result of misclassification, in the images of the left forearm is unacceptable in a system that can be used for medical purposes. In addition, requesting new data from the user or increasing the video recording time are solutions that overshadow the ease of use of the proposed system. In this respect, a hybrid approach has been proposed that can simultaneously evaluate all the results obtained in order to prevent the increase in the number of data to be requested from the user and to ensure that the number of architectural layers of the existing simple CNN Model is preserved (to prevent the training time from being negatively affected) in this thesis. With the proposed **Hybrid Decision-Making Algorithm** (this algorithm was first introduced in the study of Erdem & Utku (2022b)), all CNN Models are trained separately with three datasets and then subjected to post-test. The post-test result of a particular CNN Model trained with a particular dataset constitutes the first-class predicted of the corresponding CNN Model for the corresponding dataset. Post-testing is performed for each CNN Model using all three Datasets, yielding nine predicted first-class results. These results are grouped specifically for datasets and form the predicted mid-class results.

When the results in Table 3.1 are examined within the scope of the **predicted mid-class results** (obtained by evaluating the results of all models using the relevant dataset in a certain period), it is seen that;

- All mid-classes were predicted correctly for Period-1 right forearm. When Dataset-1 was used, incorrect prediction of the CNN Model-1 predicted first-class result for the left forearm did not affect the mid-class outcome.
- All mid-classes were predicted correctly for Period-2 right forearm. When Dataset-1 was used, erroneous estimation of the predicted first-class results of

CNN Model-1 and CNN Model-2 for the left forearm causes the mid-class result determined within the scope of Dataset-1 to be determined incorrectly.

- When Dataset-1 was used in Period-3, incorrect estimation of the predicted first-class results of CNN Model-1 and CNN Model-2 for the left forearm caused the mid-class result determined within the scope of Dataset-1 to be determined incorrectly.

When these results are examined, it is seen that the estimated mid-class results are not effective alone in making the ultimate decision.

According to the Hybrid Decision-Making Algorithm, three different predicted mid-class results obtained for a particular tissue (class) of a certain period are used to determine the predicted ultimate-class result. In this context, the purpose of using three different CNN Models in the algorithm is to create a control value from the post-test result obtained with the last CNN Model when the first-class results obtained for any two models for a given dataset are different. Similarly, by using three separate datasets, it is aimed to create a control value from the predicted mid-class result obtained with the last dataset when the predicted mid-class results obtained for any two datasets are different. It is not recommended to use the results obtained with Dataset-3 alone in determining the ultimate-class, especially when looking at the post-test results of CNN Model-3 for the right forearm in Period-2, according to Table 3.1. Because although the predicted first-class can be determined correctly, the very stable correct predictions of the same CNN Model according to the post-test results performed with other datasets resulted in a large number of incorrect predictions (8 of them) with this dataset alone. In addition, when Table 3.1 is examined within the scope of Period-2 in which Dataset-3 is used, it is observed that the post-test correct prediction averages of CNN Model-3 are lower than those of Model-1 and Model-2. In this context, all results should be evaluated in an ordinary user-oriented system based on all these results.

When the results in Table 3.1 are examined within the scope of the **predicted ultimate-class** results, it has been determined that all ultimate-classes for all periods can be predicted correctly.

The complexity matrix and performance evaluation parameters obtained according to the post-test results performed in the Classification Phase are given in Table 3.2. According to the table, CNN Model-1 and CNN Model-2 were more effective in terms of True Positive Rate (1.0000) and False Negative Rate (0.0000). CNN Model-3 was more effective in terms of False Positive Rate (0.0212), True Negative Rate (0.9788), Accuracy Rate (0.9619), Misclassification Rate (0.0381), Precision (0.9672) and F-Score (0.9516). Prevalence (0.4000) value was calculated equally for all three CNN Models.

Table 3.2 The complexity matrix and performance evaluation parameters of the post-test results of the proposed system's Classification Phase

n=315 (per CNN Model)		CNN Model-1 (28x28) Predicted Ultimate-Class		CNN Model-2 (64x64) Predicted Ultimate-Class		CNN Model-3 (128x128) Predicted Ultimate-Class	
		Right Forearm (Positive)	Left Forearm (Negative)	Right Forearm (Positive)	Left Forearm (Negative)	Right Forearm (Positive)	Left Forearm (Negative)
Actual Class	Right Forearm (Positive)	126 <i>TP</i>	0 <i>FN</i>	126 <i>TP</i>	0 <i>FN</i>	118 <i>TP</i>	8 <i>FN</i>
	Left Forearm (Negative)	54 <i>FP</i>	135 <i>TN</i>	58 <i>FP</i>	131 <i>TN</i>	4 <i>FP</i>	185 <i>TN</i>
Performance Evaluation Parameters	True Positive Rate= $TP / (TP+FN)$	1.0000		1.0000		0.9365	
	False Negative Rate= $FN / (FN+TP)$	0.0000		0.0000		0.0635	
	False Positive Rate= $FP / (FP+TN)$	0.2857		0.3069		0.0212	
	True Negative Rate= $TN / (TN+FP)$	0.7143		0.6931		0.9788	
	Accuracy Rate= $(TP + TN) / (TP+TN+FP+FN)$	0.8286		0.8159		0.9619	
	Misclassification Rate= $(FP + FN) / (TP+TN+FP+FN)$	0.1714		0.1841		0.0381	
	Precision= $TP / (TP + FP)$	0.7000		0.6848		0.9672	
	Prevalence= $(TP+FN) / (TP+TN+FP+FN)$	0.4000		0.4000		0.4000	
	F Score= $(2 * Precision * True Positive Rate) / (Precision + True Positive Rate)$	0.8235		0.8129		0.9516	

3.2 Experimental Results of the Object Detection Phase

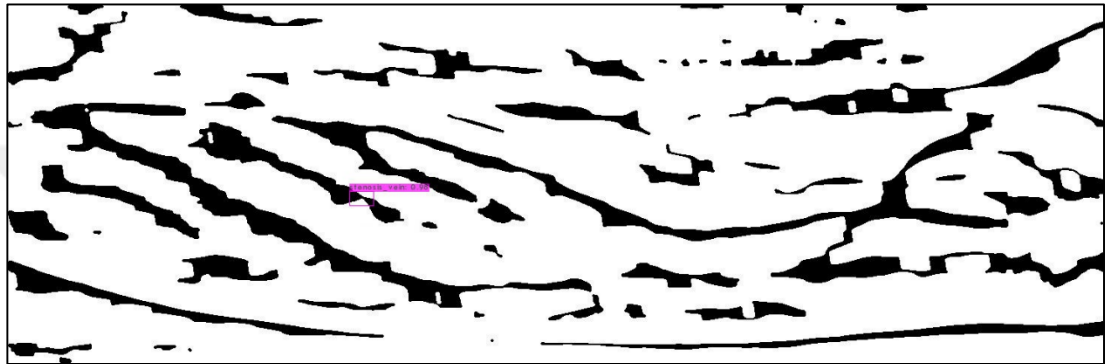
Images whose tissue classes are determined in the Classification Phase are subjected to vascular degeneration detection process in the Object Detection Phase. Since there is no vascular degeneration dataset available to the general public, a dataset containing artificial vascular degeneration patterns in which a certain template is preserved was created at this phase. The dataset to be used in the Train and Pre-test Step includes a total of 200 images (each image has 5 stenosis_vein, spider_vein and varicose_vein patterns), 50 of which were created with the data augmentation technique, belonging to the Period-1 right forearm, and a total of 3000 artificial patterns. In this way, it is ensured that the proposed system can detect both vasoconstriction (narrowing) and vasodilation (enlargement) at the same time. Following the completion of the Train and Pre-Test Step, the Post-Test Step was carried out. In the Post-Test Step, with the YOLOv3 algorithm, both the object detection rates (confidence value) were calculated in the range of [0.00, 1.00], and the position of the relevant object was indicated by marking with a bounding box.

In this study, vascular degeneration detection was performed with a total of 45 images in the post-test dataset (15 images representing each class) for all three classes using the YOLOv3 algorithm. Considering the degeneration progression process, the initial stage degeneration patterns (smaller sized spider_vein, less convoluted/cambered varicose-vein and less narrowed stenosis_vein) were used in the Period-1 images, while the advanced stage degeneration patterns (more branched spider_vein, more expanded varicose_vein showing, and more narrowed stenosis_vein) were used in the Period-3 images. Period-2 images, on the other hand, were shaped by considering as the transition phase between these two patterns.

The image containing the artificial pattern of the stenosis_vein class in the post-test dataset is shown in Figure 3.3 (a), and the YOLOv3 confidence value and bounding box representation of this image obtained after the post-test are shown in Figure 3.3 (b).



(a)





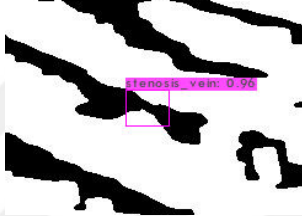
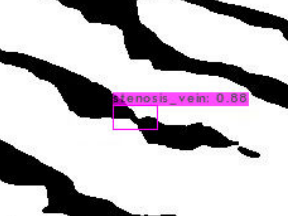


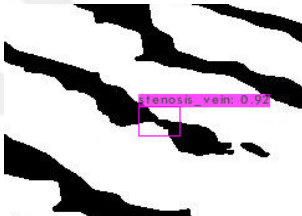





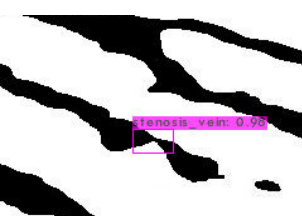


(b)

Figure 3.3 Stenosis_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period1_Img8-9_LeftArm.jpg is used as a representative). (a) Before post-test, and (b) After post-test

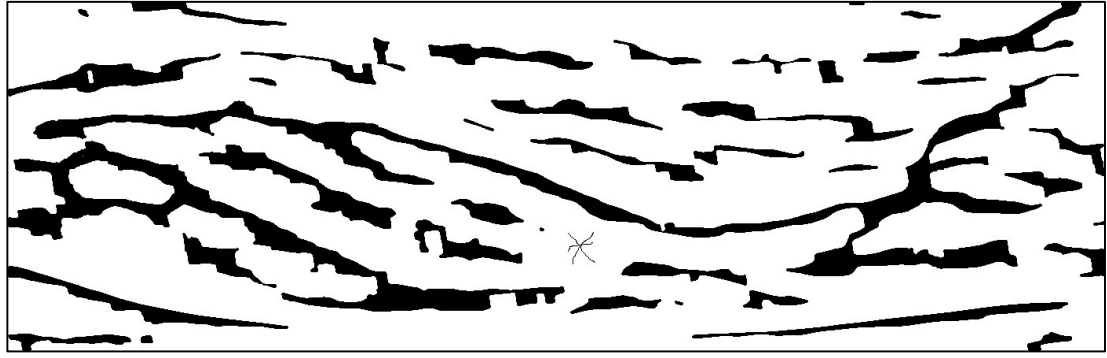
The YOLOv3 confidence values determined as a result of the post-test by adding the stenosis_vein class artificial patterns to the left forearm images obtained from three separate imaging periods are given in Table 3.3.

Table 3.3 YOLOv3 confidence values determined as a result of post-test for stenosis_vein class artificial patterns (artificial pattern sizes are shown proportionally)

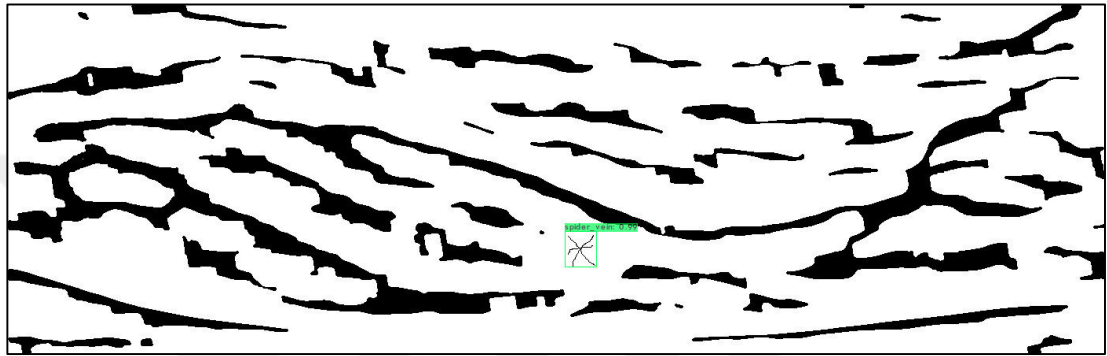
Period-1 (Month-1)	Period-2 (Month-2)	Period-3 (Month-3)
Image: Fused_Modified_0-1 Confidence value: 0.00 	Image: Fused_Modified_0-1 Confidence value: 0.41 	Image: Fused_Modified_0-1 Confidence value: 0.91 
Image: Fused_Modified_2-3 Confidence value: 0.00 	Image: Fused_Modified_2-3 Confidence value: 0.96 	Image: Fused_Modified_2-3 Confidence value: 0.88 
Image: Fused_Modified_4-5 Confidence value: 0.00 	Image: Fused_Modified_4-5 Confidence value: 0.68 	Image: Fused_Modified_4-5 Confidence value: 0.92 
Image: Fused_Modified_6-7 Confidence value: 0.00 	Image: Fused_Modified_6-7 Confidence value: 0.46 	Image: Fused_Modified_6-7 Confidence value: 0.91 
Image: Fused_Modified_8-9 Confidence value: 0.00 	Image: Fused_Modified_8-9 Confidence value: 0.71 	Image: Fused_Modified_8-9 Confidence value: 0.98 
Average confidence value of Period-1: 0.000	Average confidence value of Period-2: 0.644	Average confidence value of Period-3: 0.920

Accordingly, none of the patterns could be detected detected in Period-1, which represents the initial phase (average confidence value: 0.000). This is because contraction formats that are slightly thicker than the defined patterns are used for training. For its solution, additions with similar patterns to the dataset should be performed. Since patterns close to the training format were used in Period-2 and Period-3, determinations were made. However, although the narrowings with a slightly thick middle part in Period-2 reduced the confidence value, the relevant class could be determined (average confidence value: 0.644). In Period-3, on the other hand, since patterns closer to the training dataset format were used, object detections were achieved with high confidence values (average confidence value: 0.920). Since one of the purposes of this thesis study is the follow-up of the detected degeneration, the increase in the average confidence value as the periods progresses is extremely critical in terms of being an aiding factor for the physician in decision making.

The image containing the artificial pattern of the spider_vein class in the post-test dataset is shown in Figure 3.4 (a), and the YOLOv3 confidence value and bounding box representation of this image obtained after the post-test are shown in Figure 3.4 (b).



(a)

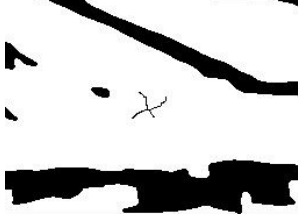
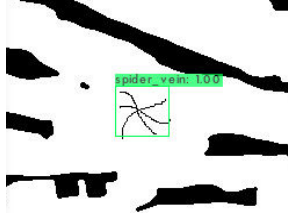
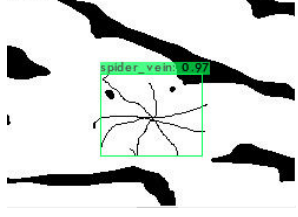
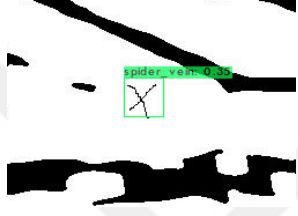
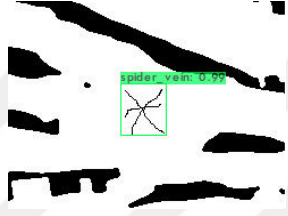
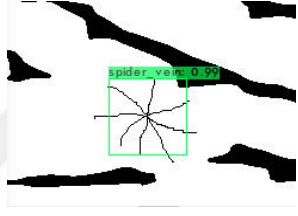
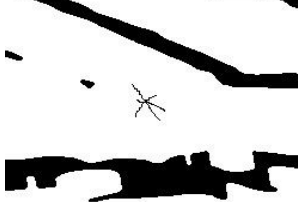
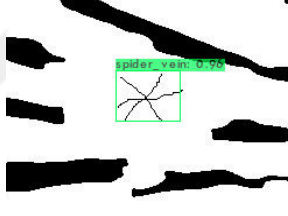
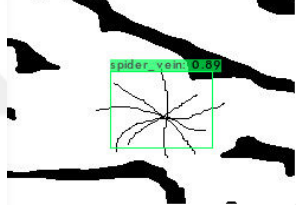
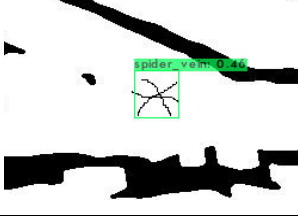
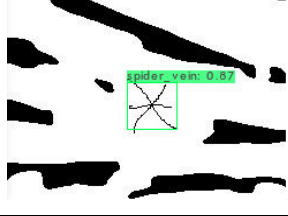
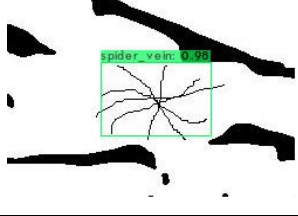
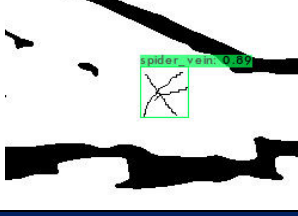
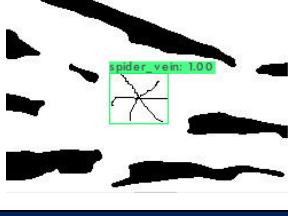
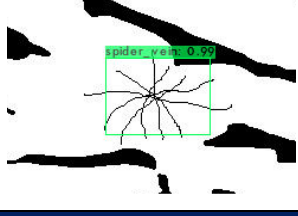


(b)

Figure 3.4 Spider_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period1_Img2-3_LeftArm.jpg is used as representative). (a) Before post-test, and (b) After post-test

The YOLOv3 confidence values determined as a result of the post-test by adding the spider_vein class artificial patterns to the left forearm images obtained from three separate imaging periods are given in Table 3.4. Accordingly, two of the patterns found in Period-1, which represents the initial phase, could not be detected (average confidence value: 0.340). This is because the slightly smaller and less branched spider_vein formats are used for training than the defined patterns. For its solution, additions with similar patterns to the dataset should be performed. Since patterns close to the training format were used in Period-2 (average confidence value: 0.964) and Period-3 (average confidence value: 0.964), all of the determinations were made.

Table 3.4 YOLOv3 confidence values determined as a result of post-test for spider_vein class artificial patterns (artificial pattern sizes are shown proportionally)

Period-1 (Month-1)	Period-2 (Month-2)	Period-3 (Month-3)
Image: Fused_Modified_0-1 Confidence value: 0.00 	Image: Fused_Modified_0-1 Confidence value: 1.00 	Image: Fused_Modified_0-1 Confidence value: 0.97 
Image: Fused_Modified_2-3 Confidence value: 0.35 	Image: Fused_Modified_2-3 Confidence value: 0.99 	Image: Fused_Modified_2-3 Confidence value: 0.99 
Image: Fused_Modified_4-5 Confidence value: 0.00 	Image: Fused_Modified_4-5 Confidence value: 0.96 	Image: Fused_Modified_4-5 Confidence value: 0.89 
Image: Fused_Modified_6-7 Confidence value: 0.46 	Image: Fused_Modified_6-7 Confidence value: 0.87 	Image: Fused_Modified_6-7 Confidence value: 0.98 
Image: Fused_Modified_8-9 Confidence value: 0.89 	Image: Fused_Modified_8-9 Confidence value: 1.00 	Image: Fused_Modified_8-9 Confidence value: 0.99 
Average confidence value of Period-1: 0.340	Average confidence value of Period-2: 0.964	Average confidence value of Period-3: 0.964

The image containing the artificial pattern of the varicose_vein class in the post-test dataset is shown in Figure 3.5 (a), and the YOLOv3 confidence value and bounding box representation of this image obtained after the post-test are shown in Figure 3.5 (b).

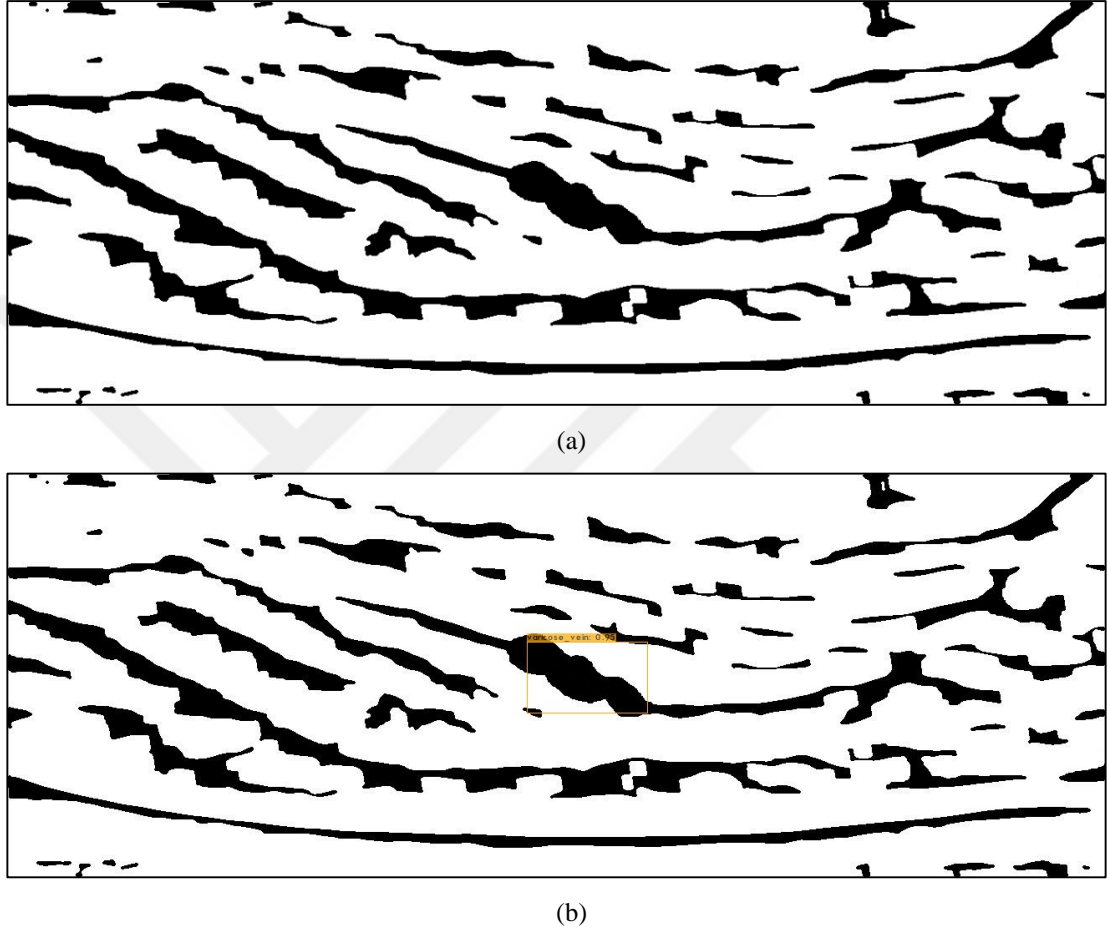




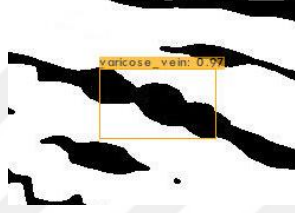
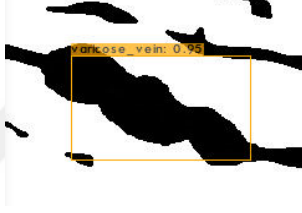
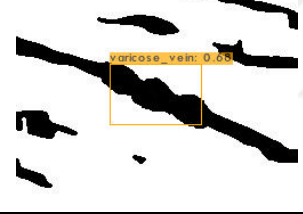
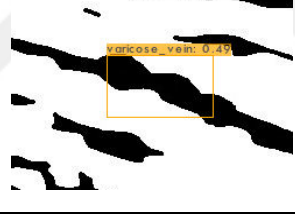
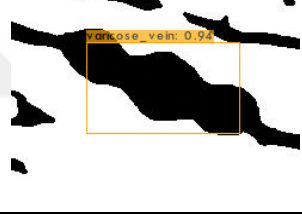


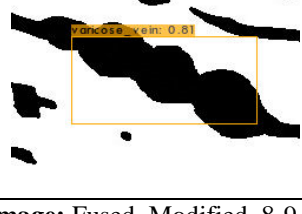
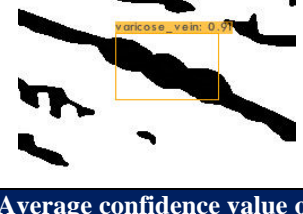

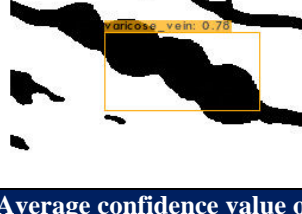


Figure 3.5 Varicose_vein class artificial pattern (binary NIR fused-modified image with the code Fused_Modified_Period3_Img2-3_LeftArm.jpg is used as a representative). (a) Before post-test, and (b) After post-test

The YOLOv3 confidence values determined as a result of the post-test by adding the varicose_vein class artificial patterns to the left forearm images obtained from three separate imaging periods are given in Table 3.5.

Table 3.5 YOLOv3 confidence values determined as a result of post-test for varicose_vein class artificial patterns (artificial pattern sizes are shown proportionally)

Period-1 (Month-1)	Period-2 (Month-2)	Period-3 (Month-3)
Image: Fused_Modified_0-1 Confidence value: 0.71 	Image: Fused_Modified_0-1 Confidence value: 0.84 	Image: Fused_Modified_0-1 Confidence value: 0.94 
Image: Fused_Modified_2-3 Confidence value: 0.98 	Image: Fused_Modified_2-3 Confidence value: 0.97 	Image: Fused_Modified_2-3 Confidence value: 0.95 
Image: Fused_Modified_4-5 Confidence value: 0.68 	Image: Fused_Modified_4-5 Confidence value: 0.49 	Image: Fused_Modified_4-5 Confidence value: 0.94 
Image: Fused_Modified_6-7 Confidence value: 0.69 	Image: Fused_Modified_6-7 Confidence value: 0.97 	Image: Fused_Modified_6-7 Confidence value: 0.81 
Image: Fused_Modified_8-9 Confidence value: 0.91 	Image: Fused_Modified_8-9 Confidence value: 0.64 	Image: Fused_Modified_8-9 Confidence value: 0.78 
Average confidence value of Period-1: 0.794	Average confidence value of Period-2: 0.782	Average confidence value of Period-3: 0.884

Accordingly, all of the patterns representing all phases of degeneration could be detected (average confidence value: 0.794, 0.782 and 0.884). However, Fused_Modified_4-5 (average confidence value: 0.49) and Fused_Modified_8-9 (average confidence value: 0.64) images with low confidence values in Period-2 draw attention. When both images are examined, it is thought that this is due to the fact that the cambered structures in the varicose veins (as defined in the training dataset) are not visualized in the lower part of these veins and represent a relatively flat structure. In addition, these two images caused the mean confidence value for Period-2 to be lower than Period-1. Another important point here is the possibility of encountering such flat structures in real patterns due to lighting. Instead of introducing a partially flat varicose_vein pattern to the training dataset, it is envisaged that this problem can be solved with the image fusion method in the Digital Image Post-Processing Phase used in this thesis in the real case.

The complexity matrix and performance evaluation parameters obtained according to the post-test results performed in the Object Detection Phase are given in Table 3.6. In order to calculate the performance evaluation parameters $TP_{stenosis_vein}=(\#1)$, $TP_{spider_vein}=(\#6)$, $TP_{varicose_vein}=(\#11)$ cell values used as True Positive (TP), $FN_{stenosis_vein}=(\#2)+(\#3)+(\#4)$, $FN_{spider_vein}=(\#5)+(\#7)+(\#8)$, $FN_{varicose_vein}=(\#9)+(\#10)+(\#12)$ values used as False Negative (FN), $FP_{stenosis_vein}=(\#5)+(\#9)$, $FP_{spider_vein}=(\#2)+(\#10)$, $FP_{varicose_vein}=(\#3)+(\#7)$ values used as False Positive (FP), and $TN_{stenosis_vein}=(\#6)+(\#7)+(\#8)+(\#10)+(\#11)+(\#12)$, $TN_{spider_vein}=(\#1)+(\#3)+(\#4)+(\#9)+(\#11)+(\#12)$, $TN_{varicose_vein}=(\#1)+(\#2)+(\#4)+(\#5)+(\#6)+(\#8)$ values used as True Negative (TN).

Table 3.6 The complexity matrix and performance evaluation parameters of the post-test results of proposed system's Object Detection Phase

n=15 (per class)		YOLOv3 Algorithm (416x416) Predicted Class			
		stenosis_vein	spider_vein	varicose_vein	Unclassified
Actual Class	stenosis_vein	10 #1	0 #2	0 #3	5 #4
	spider_vein	0 #5	13 #6	0 #7	2 #8
	varicose_vein	0 #9	0 #10	15 #11	0 #12
Performance Evaluation Parameters	True Positive Rate= TP / (TP+FN)	0.6667	0.8667	1.0000	
	False Negative Rate= FN / (FN+TP)	0.3333	0.1333	0.0000	
	False Positive Rate= FP / (FP+TN)	0.0000	0.0000	0.0000	
	True Negative Rate= TN / (TN+FP)	1.0000	1.0000	1.0000	
	Accuracy Rate= (TP + TN) / (TP+TN+FP+FN)	0.8889	0.9556	1.0000	
	Misclassification Rate= (FP + FN) / (TP+TN+FP+FN)	0.1111	0.0444	0.0000	
	Precision= TP / (TP + FP)	1.0000	1.0000	1.0000	
	Prevalence= (TP+FN) / (TP+TN+FP+FN)	0.3333	0.3333	0.3333	
	F Score= (2 * Precision * True Positive Rate) / (Precision + True Positive Rate)	0.8000	0.9286	1.0000	

According to Table 3.6, True Positive Rate (1.0000), False Negative Rate (0.0000), Accuracy Rate (1.0000), Misclassification Rate (0.0000), and F-Score (1.0000) performance evaluation parameters were more effective for varicose_vein class. The second most effective class prediction evaluation for the same parameters was determined for the spider_vein class (True Positive Rate (0.8667), False Negative Rate (0.1333), Accuracy Rate (0.9556), Misclassification Rate (0.0444), and F-Score (0.9286)). The lowest values in terms of performance evaluation parameters were obtained for the stenosis_vein class due to the narrowings represented as bold in Period-1 (True Positive Rate (0.6667), False Negative Rate (0.3333), Accuracy Rate (0.8889), Misclassification Rate (0.1111), and F-Score (0.8000)). In other words, the proposed system could not detect the initial stages of

contraction type degeneration due to the pattern formats used in training. Repetition of YOLOv3 trainings with the addition of patterns representing the initial stages of contraction to the training dataset will provide a solution to this situation. Another remarkable point according to the Table 3.6 is that although the system could not classify some patterns within the scope of stenosis_vein (5 unclassified) and spider_vein (2 unclassified) classes, it did not perform any misclassification of patterns representing all three classes. Avoidance of misclassification is one of the most important cornerstones of health practices. False Positive Rate (0.0000), True Negative Rate (1.0000), Precision (1.0000) and Prevalence (0.3333) performance evaluation parameters of YOLOv3 post-test results were calculated equally for all three classes.

CHAPTER FOUR

CONCLUSIONS

In the Fourth Chapter, a general evaluation of the study is made and in which aspects it can shed light on future studies.

Basically, this doctoral dissertation study, which aims to provide early diagnosis of vascular degeneration in the home environment and to follow the progress, consists of six phases. The primary reason for the preference of NIR spectroscopy used in **the Imaging Technique Phase** is that it provides harmless, non-invasive and contactless imaging to the body. The use of a low-cost NIR camera for image acquisition makes it possible to develop a system that will appeal to everyday use. The user takes the video recordings of the tissue region he/she wants to view with the NIR camera at certain periods and uploads them to the server. The video-based realization of imaging provides ergonomics for the user in imaging. In addition, by converting the video to sequential images, the deficiencies caused by the lighting are eliminated and the need for a large number of images to be used in tissue classification will be met. Within the scope of e-health, since the system prepared for home use addresses ordinary users, the user is only included in the image acquisition phase, all the other phases are performed on the server and the results are reported to the user and his/her physician. In this way, it is ensured that calculations which require high processing power can be easily performed on the server, and the user is provided with ease of use. Especially since the superficial veins are located under the skin, they can be seen more or less with the naked eye. However, this is not sufficient to obtain clear vascular patterns. For this reason, vein patterns are enhanced in **the Digital Image Pre-Processing Phase**. In order to eliminate the factors caused by imaging in the home environment (change of light angle, shake or displacement of the imaged area or camera, ambient conditions or low-spec camera usage, etc.) to a certain extent, image fusion processes are applied in **the Digital Image Post-Processing Phase**. It is aimed to detect and long-term follow-up of vascular degeneration in the system. In this context, imaging should be performed on a certain tissue area at certain periods. Convolutional neural networks are used in **the**

Classification Phase for the recognition of the tissue region. Since the number of images that can be requested from the ordinary user is limited, tissue images obtained in video format are used in datasets. The tissue in the first period is used as the basis for the introduction of tissue classes. With a 35-second imaging, 1050 similar tissue patterns can be obtained and the tissue classes to be followed can be introduced to the system. In this study, two classes as right and left forearm are introduced to the system. However, both the fact that the system works with low-spec imaging hardware in order to offer cost-effectiveness and that it uses non-professional images taken in the home environment lead to incorrect classifications. For this reason, high accuracy rates in the predicted ultimate-class results has been achieved by using a Hybrid Decision-Making Algorithm in which the results of three convolutional neural networks and three separate datasets are evaluated together. After the tissues are classified, the vascular degeneration that may occur in these tissue regions is detected by **the Object Detection Phase**. At this phase, detection is performed according to three different vascular degeneration classes defined within the scope of vascular narrowing and enlargement in the system. With the object detection algorithm used in this context, the detection rates of the related degeneration are calculated with a certain confidence value and their locations are marked on the image. **The Augmented Reality Phase** is used as the last phase of the system in order to present the images containing both the detection rate and the locations in the form of easy-to-understand visuals to the user and his/her physician. At this phase, images of the relevant period are prepared by superimposing the vein images including object detection results (virtual material) on the raw images (real material) containing the tissue region and they are converted to video format. The video image obtained for each new period belonging to a certain tissue region is added to the previous period videos sequentially. When the user runs the system in augmented reality mode and displays the relevant tissue region, these images are presented as superimposed on live stream images. In this way, both early detection and long-term follow-up of degeneration in a particular tissue area can be performed.

The convolutional neural network ultimate-class results obtained according to the results of the Hybrid Decision-Making Algorithm used in the Classification Phase of

the proposed system were able to correctly perform the classification for all imaging periods. According to the experimental results, Accuracy Rate (Model-1: 0.8286, Model-2: 0.8159, Model-3: 0.9619), Misclassification Rate (Model-1: 0.1714, Model-2: 0.1841, Model-3: 0.0381) and Precision (Model-1: 0.7000, Model-2: 0.6848, Model-3: 0.9672) were obtained. Hybrid evaluation of the results without relying on a single model and a single dataset increases the classification reliability of an e-health application that appeals to non-professional users. Since there is no relevant dataset open to public access, degeneration detection and follow-up could be achieved with the experiments carried out by creating artificial vascular degeneration patterns in the Object Detection Phase. Especially for the varicose_vein class, all of the patterns representing both the initial stages and the advanced stages could be detected. For the spider_vein class, only 2 patterns at the initial stage could not be detected. These two cases show that the system is suitable for both degeneration detection and follow-up of initial, middle and advanced stages of the progress for varicose_vein and spider_vein classes. Although no detection could be made in the first period, which includes the initial stage patterns for the stenosis_vein class, the second and the third period detections could be made. The reason for this is that there are many initial stage patterns in the training dataset. By expanding the dataset, it will be possible to recognize the initial stage for the stenosis_vein class. In addition, it was observed that 5 of 15 stenosis_vein patterns and 2 of 15 spider_vein patterns could not be classified in the system. However, no pattern was misclassified. Absence of misclassification is one of the basic criteria sought especially in applications to be used within the scope of health. According to the experimental results, Accuracy Rate (stenosis_vein: 0.8889, spider_vein: 0.9556, varicose_vein: 1.0000), Misclassification Rate (stenosis_vein: 0.1111, spider_vein: 0.0444, varicose_vein: 0.0000) and Precision (stenosis_vein: 1.0000, spider_vein: 1.0000, varicose_vein: 1.0000) were obtained. In addition, there is a need for case studies in which the developed system will be used by specialist doctors in a long-term follow-up process.

Within the scope of future studies, by the development of the mobile application infrastructure of the proposed system, it is planned to include the zoom in/out feature

in the visual interface in order to visualize various analyzes related to the vascular structure and to add images related to the representations of the advanced stages of degeneration. It is also considered to develop a version of the system that can be used within the scope of whole body scanning for the purpose of detecting only degeneration (an introduction process will not be required in the first period for classification). Although the proposed system is based on near-infrared radiation spectroscopy images provided within the scope of e-health, it can also be used for other detection and follow-up in the health field. The system can be easily modified for other image formats (x-ray, tomography, magnetic resonance etc.) and ailments (recovery times in bone fractures, organ development conditions, eye veins tracking, etc.) only by simply changing the training classes in order to detect and follow-up from the health data brought into black-and-white binary format. From this point of view, it is evaluated that the proposed multi-purpose detection and tracking system can shed light on future studies within the scope of providing support for different imaging-platforms.

REFERENCES

- Adapthisteq. (2022, September 30). *Mathworks Help Center*.
<https://www.mathworks.com/help/images/ref/adapthisteq.html>
- 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.
<https://doi.org/10.1364/BOE.7.002565>
- Akaguma, T., Okura, F., Sato, T., & Yokoya, N. (2013, November). *Mobile AR using pre-captured omnidirectional images* [Conference presentation]. In Proceedings of the SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, pp. 1–4. <https://doi.org/10.1145/2543651.2543666>
- Al-johania, N. A., & Elrefaei, L. A. (2019). Dorsal hand vein recognition by convolutional neural networks: feature learning and transfer learning approaches. *International Journal of Intelligent Engineering and Systems*, 12(3), 178-191. <https://doi.org/10.22266/ijies2019.0630.19>
- Aliferi, C. (2016). *Android programming cookbook: Kick-start your Android projects*. Exelixis Media P. C.
- Alves, J. B., Marques, B., Ferreira, C., Dias, P., & Santos, B. S. (2022). Comparing augmented reality visualization methods for assembly procedures. *Virtual Reality*, 26, 235-248. <https://doi.org/10.1007/s10055-021-00557-8>
- Asrar, M., Al-Habaibeh, A., & Houda, M. (2016). Innovative algorithm to evaluate the capabilities of visual, near infrared, and infrared technologies for the detection of veins for intravenous cannulation. *Applied Optics*, 55(34), D67-D75.
<https://doi.org/10.1364/AO.55.000D67>
- AssetInfinity. (2022, September 30). *Why Use Mobile Application for Cloud-Based Asset Management Software?* <https://www.assetinfinity.com/blog/mobile-application-for-asset-management-software>
- Australian Broadcasting Corporation. (2022, September 30). *The Lunar Module Computer*. <https://www.abc.net.au/science/moon/computer.htm>

- Bana, S., & Kaur, D. (2011). Fingerprint recognition using image segmentation. *International Journal of Advanced Engineering Sciences and Technologies*, 5(1), 12-23. Retrieved September 02, 2022, from <https://www.cs.sjtu.edu.cn/~shengbin/course/SE/Fingerprint%20Recognition%20using%20Image%20segmentation.pdf>
- Bandyopadhyay, H. (2022, September 30). *YOLO: Real-Time Object Detection Explained*. <https://www.v7labs.com/blog/yolo-object-detection>
- Bansal, M., Kumar, M., & Kumar, M. (2021). 2D object recognition: A comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimedia Tools and Applications*, 80, 18839–18857.
<https://doi.org/10.1007/s11042-021-10646-0>
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In: A. Leonardis, H. Bischof, A. Pinz (Eds.) *Computer vision-European conference on computer vision* (pp. 404-417). Lecture Notes in Computer Science, vol 3951. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/11744023_32
- Bayraktar, E., & Boyraz, P. (2017). Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics. *Turkish Journal of Electrical Engineering and Computer Sciences*, 25(3), 66, 2444-2454.
<https://doi.org/10.3906/elk-1602-225>
- Bazrafkan, S., Nedelcu, T., Costache, C., & Corcoran, P. (2016, January 07-11). *Finger vein biometric: Smartphone footprint prototype with vein map extraction using computational imaging techniques* [Conference presentation]. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), pp. 512-513. IEEE. Las Vegas, NV, United States.
<https://doi.org/10.1109/ICCE.2016.7430709>.
- Benziane, S., & Benyettou, A. (2015). Hand vein authentication based wavelet feature extraction. *Recent Advances on Electrosience and Computers*, 242-249.

- Retrieved September 02, 2022, from <http://www.inase.org/library/2015/barcelona/bypaper/ELECTR/ELECTR-38.pdf>
- Bhaskar R. (2022, September 30). *Understanding convolutional neural networks in 2 minutes*. Society of Robotics and Automation. <http://www.sra.vjti.info/blog/machine-learning/understanding-convolutional-neural-networks-in-2-minutes>
- Billinghurst, M., Clark, A., & Lee G. (2015). A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction*, 8(2-3), 73-272. <https://doi.org/10.1561/11000000049>
- Boezeman, R. P., Moll, F. L., Ünlü, Ç., & de Vries, J. P. P. (2016). Systematic review of clinical applications of monitoring muscle tissue oxygenation with near-infrared spectroscopy in vascular disease. *Microvascular Research*, 104, 11-22. <https://doi.org/10.1016/j.mvr.2015.11.004>
- Buyya, R., Calheiros, R. N., & Dastjerdi, A. V. (2016). *Big Data Principles and Paradigms*. Elsevier Inc. <https://doi.org/10.1016/C2015-0-04136-3>
- Carmigniani, J., & Furht, B. (2011). Augmented reality: An overview. In: B. Furht (Ed.) *Handbook of augmented reality* (pp. 3-46). Springer Science+Business Media, LLC. <https://doi.org/10.1007/978-1-4614-0064-6>
- Chiao, F. B., Resta-Flarer, F., Lesser, J., Ng, J., Ganz, A., Pino-Luey, D., Bennett, H., Perkins Jr, C., & Witek, B. (2013). Vein visualization: Patient characteristic factors and efficacy of a new infrared vein finder technology. *British Journal of Anaesthesia*, 110(6), 966-971. <https://doi.org/10.1093/bja/aet003>
- Cömert, M. (2015). *Avuç içi damar izi ile biyometrik kimlik doğrulama arayüzünün gömülü sistem üzerinde tasarlanması (Designing biometric identification interface with palm vein structure on embedded system)* [Master's Degree Thesis]. Mersin University.

- Craig, A. B., (2013). *Understanding augmented reality: Concepts and applications*. Elsevier Inc.
<https://doi.org/10.1016/C2011-0-07249-6>
- Crisan, S. (2017). A novel perspective on hand vein patterns for biometric recognition: Problems, challenges, and implementations. In: R., Jiang, , S., Al-Maadeed, A., Bouridane, P., Crookes, & A., Beghdadi (Eds.) *Biometric security and privacy* (pp. 21-49). *Signal Processing for Security Technologies*. Springer International Publishing, Cham.
https://doi.org/10.1007/978-3-319-47301-7_2
- Çağl, G., & Yıldırım, B. (2020). Bir montaj parçasının derin öğrenme ve görüntü işleme ile tespiti (Detection of an assembly part with deep learning and image processing). *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 3(2), 31-37.
<https://doi.org/10.38016/jista.710144>
- Çan, Z. (2021). *Otomasyon Sistemlerinde Görüntü İşleme Tekniklerini Kullanan Ürün Tanımı Uygulaması (Product Definition Application Using Image Processing Techniques In Automation Systems)* [Master's Degree Thesis]. Sakarya University.
- Daily Sabah. (2022, September 30). *World's most popular cat breeds*.
<https://www.dailysabah.com/gallery/worlds-most-popular-cat-breeds/images>
- Dar, P. (2022, September 30). *25 open datasets for deep learning every data scientist must work with*. Analytics Vidhya.
<https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/>
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features, *Library Review*, 65(3), 122-135.
<https://doi.org/10.1108/LR-06-2015-0061>
- Demir, A. G. (2019). *Anjiyografi görüntülerinde damar daralmalarının evrışimsel sinir ağı yöntemi kullanılarak belirlenmesi (Determination of vascular stenosis on angiography images using convolutional neural network method)* [Master's Degree Thesis]. Başkent University.

- DetectSURFFeatures. (2022, September 30). Mathworks Help Center. <https://www.mathworks.com/help/vision/ref/detectsurffeatures.html>
- Devi, M. C., & Kirubaharan, G. D., (2015). A comparative study of SIFT and SURF approaches. *International Journal Of Engineering Research & Technology*, 3(28), 1-4. Retrieved September 02, 2022, from <https://www.ijert.org/a-comparative-study-of-sift-and-surf-approaches>
- Doğan, F., & Türkoğlu, İ. (2018). Derin öğrenme algoritmalarının yaprak sınıflandırma başarımlarının karşılaştırılması. *Sakarya University Journal of Computer and Information Sciences*, 1(1), 10-21. Retrieved September 02, 2022, from <http://saucis.sakarya.edu.tr/tr/download/article-file/479189>
- Elnasir, S., & Shamsuddin, S. M. (2014). Palm vein recognition based on 2D-discrete wavelet transform and linear discrimination analysis. *International Journal of Advances in Soft Computing and Its Applications*, 6(3), 43-59. Retrieved September 02, 2022, from https://www.researchgate.net/publication/288225959_Palm_vein_recognition_based_on_2D-discrete_wavelet_transform_and_linear_discrimination_analysis
- Emlek, A., & Peker, M. (2020). Evrişimli sinir ağı tabanlı derinlik haritası iyileştirme (Improving depth map based on convolutional neural network). *Niğde Ömer Halisdemir University Journal of Engineering Sciences*, 9(2), 722-729. <https://doi.org/10.28948/ngumuh.709159>
- Erdem, H. A. (2013). *Utilization of virtual reality environment as an interactive visual learning tool in primary school education system* [Master's Degree Thesis]. Dokuz Eylül University.
- Erdem, H. A., Erdem, I., & Utku, S. (2020, November 21-25). *Near-infrared mobile imaging systems for e-health: Lighting the veins* [Conference presentation]. In Proceedings of the Twelfth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED), pp. 80-84, IARIA. Valencia, Spain. Retrieved September 02, 2022, from https://www.thinkmind.org/index.php?view=article&articleid=etelemed_2020_3_90_40039

- Erdem, H. A., & Utku, S. (2022a, July 22-24). *Detecting Venous Disorders via Near-Infrared Imaging: Observation of Varicose Vein Development* [Conference presentation]. In Proceedings of the IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications (IARIA Congresses), pp. 65-68, IARIA. Nice, France. Retrieved September 02, 2022, from https://www.thinkmind.org/index.php?view=article&articleid=iaria_congress_2022_1_130_50088
- Erdem, H. A., & Utku, S. (2022b). Augmented reality aided pre-diagnosis environment for telemedicine: Superficial vein surveillance system. *European Journal of Science and Technology*, 38, 376-385.
<https://doi.org/10.31590/ejosat>
- ExtractFeatures. (2022, September 30). Mathworks Help Center. <https://www.mathworks.com/help/vision/ref/extractfeatures.html>
- Fernández, R., & Armada, M. (2017). Multisensory system for the detection and localization of peripheral subcutaneous veins. *Sensors*, 17(4), 897.
<https://doi.org/10.3390/s17040897>
- Francis, M., Jose, A., Devadhas, G. G., & Avinashe, K. K. (2017). A novel technique for forearm blood vein detection and enhancement. *Biomedical Research*, 28(7), 2913-2919. Retrieved September 02, 2022, from <https://www.alliedacademies.org/articles/a-novel-technique-for-forearm-blood-vein-detection-and-enhancement.html>
- Fırıldak, K., & Talu, M. F. (2019). Evrişimsel sinir ağlarında kullanılan transfer öğrenme yaklaşımlarının incelenmesi (Investigation of Transfer Learning Approaches Used Convolution Neural Networks), *Computer Science*, 4(2), 88-95. Retrieved September 02, 2022, from <https://dergipark.org.tr/tr/pub/bbd/issue/49546/527863>
- Fu, J., & Rui, Y. (2017). Advances in deep learning approaches for image tagging. *APSIPA Transactions on Signal and Information Processing*, 6(11), 1-13.
doi:10.1017/ATSIP.2017.12

- Geeks for Geeks (2022, September 30). *Difference between YOLO and SSD*.
<https://www.geeksforgeeks.org/difference-between-yolo-and-ssd>.
- Genymotion. (2022, September 30). <https://www.genymotion.com/>
- Gilbert, L. (2022, September 30). Plastik gezegeni: Mikroplastiklerin küresel çapta takibi. *Popular Science Türkiye*. <https://popsci.com.tr/plastik-gezegeni-mikroplastiklerin-kuresel-capta-takibi/>
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Prentice Hall, New Jersey.
- Google Colaboratory. (2022, September 30). <https://colab.research.google.com>
- Gupta, J., Pathak, S., & Kumar, G. (2022). Deep Learning (CNN) and Transfer Learning: A Review. *Journal of Physics: Conference Series*, 2273, 1-10.
<https://doi.org/10.1088/1742-6596/2273/1/012029>
- Gutiérrez, M. A., Vexo, F., & Thalmann, D. (2008). *Stepping into virtual reality*. England: Springer-Verlag London.
- Hasan, S. N. (2019). *Derin Öğrenme Yöntemleri ile Cilt Lezyon Bölütlemesi (Skin Lesion Segmentation with Deep Learning Techniques)* [Master's Degree Thesis]. İstanbul University.
- Highton, L., & Ekwobi, C. (2011). Use of the AccuVein device to map the superficial venous system. *European Journal of Plastic Surgery*, 34(4), 305-306.
<https://doi.org/10.1007/s00238-011-0583-0>
- Hosseini, M. P., Lu, S., Kamaraj, K., Slowikowski, A., & Venkatesh, H. C. (2020). Deep learning architectures. In: W. Pedrycz, S. M. Chen (Eds). *Deep learning: Concepts and architectures* (pp. 1-24). Studies in Computational Intelligence, vol 866. Springer International Publishing, Cham.
https://doi.org/10.1007/978-3-030-31756-0_1
- Huang, Z., Li, Y., Zhao, T., Ying, P., Fan, Y., & Li, J. (2021). Infusion port level detection for intravenous infusion based on yolo v3 neural network. *Mathematical Biosciences and Engineering*, 18(4), 3491-3501.

<https://doi.org/10.3934/mbe.2021175>

I2Tutorials. (2022, September 30). *Understanding single layer perceptron and difference between single layer vs multilayer perceptron*. <https://www.i2tutorials.com/what-is-single-layer-perceptron-and-difference-between-single-layer-vs-multilayer-perceptron/>

IGI Global, (2022, September 30). *What is interest points detection and description for structure from motion*. Publisher of Timely Knowledge. <https://www.igi-global.com/dictionary/interest-points-detection-and-description-for-structure-from-motion/48703>

Imcomplement. (2022, September 30). *Mathworks Help Center* <https://www.mathworks.com/help/images/ref/imcomplement.html>

Imbinarize. (2022, September 30). *Mathworks Help Center*. <https://www.mathworks.com/help/images/ref/imbinarize.html>

Imgaborfilt. (2022, September 30). *Mathworks Help Center*. <https://www.mathworks.com/help/images/ref/imgaborfilt.html>

Io, R. (2022, September 30). *Supervised vs unsupervised learning: Key differences*. Medium. <https://medium.com/@recrosoft.io/supervised-vs-unsupervised-learning-key-differences-cdd46206cdcb>

Işık, Ş., (2015). A comparative evaluation of well-known feature detectors and descriptors. *International Journal of Applied Mathematics Electronics and Computers*, 3(1), 1-6. <https://doi.org/10.18100/ijamec.60004>

İnaç, E. (2022, September 30). *Sıfırdan Yolo v4 Tiny Özelleşmiş Nesne Tespiti (Custom Object Detection)*. <https://yazilimkaravani.net/sifirdan-yolo-v4-tiny-ozellesmis-nesne-tespiti-custom-object-detection/>

İnik, Ö., & Ülker, E. (2017). Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri (Deep Learning and Deep Learning Models Used in Image Analysis). *Gaziosmanpaşa Journal of Scientific Research*, 6(3), 85-104.

- Retrieved September 02, 2022, from <https://dergipark.org.tr/tr/pub/gbad/issue/31228/330663>
- İpek, B. (2019). *Derin öğrenme temelli nesne tespiti algoritmalarının insansız hava araçlarında uygulanması* [Bachelor's Degree Thesis]. Sakarya University.
- JavaTPoint. (2022, September 30). *Single Layer Perceptron in TensorFlow*. <https://www.javatpoint.com/single-layer-perceptron-in-tensorflow>
- Juric, S., & Zalik, B. (2014). An innovative approach to near-infrared spectroscopy using a standard mobile device and its clinical application in the real-time visualization of peripheral veins. *BMC Medical Informatics and Decision Making*, 14, 1-9. <https://doi.org/10.1186/s12911-014-0100-z>
- Kantarcı, A., & Çakır, Ş. (1998, November 06-09). *Sanal gerçeklik ve turizm* [Conference presentation]. In Proceedings of the 3rd International and the 9th National Tourism Congress, pp. 17-21, Kuşadası, Aydın.
- Karaali, İ., & Eminağaoğlu, M. (2021). Mermer işlemede kalite sınıflandırması için evrişimsel sinir ağı modeli. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 36(1), 347-358. <https://doi.org/10.17341/gazimmfd.720041>
- Kaya, V., Tuncer, S., & Baran, A. (2022, September 30). *Derin Öğrenme Yöntemleri Kullanılarak Nesne Tanıma*. <http://earsiv.ersincan.edu.tr/xmlui/handle/20.500.12432/4092>
- Kayaalp, K., & Süzen, A. A. (2018). *Derin öğrenme ve Türkiye'deki uygulamaları*. İksad Publishing House. Retrieved September 02, 2022, from <https://iksadyayinevi.com/home/derin-ogrenme-ve-turkiyedeki-uygulamaları/>
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A., S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53, 5455-5516. <https://doi.org/10.1007/s10462-020-09825-6>

- Khor, W. S., Baker, B., Amin, K., Chan, A., Patel, K., & Wong, J. (2016). Augmented and virtual reality in surgery - the digital surgical environment: Applications, limitations and legal pitfalls. *Annals of Translational Medicine*, 4(23), 1-10.
<https://doi.org/10.21037/atm.2016.12.23>
- Klakegg, S., Luo, C., Goncalves, J., Hosio, S., & Kostakos, V. (2016, September). *Instrumenting smartphones with portable NIRS* [Conference presentation]. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp), pp. 618-623. ACM.
<https://doi.org/10.1145/2968219.2971590>
- Köse, Ü. (2022, September 30). *Android Studio Dersleri*. <http://umiitkose.com/wp-content/uploads/2015/08/AndroidStudio.pdf>
- Kurt, F. (2018). *Evrişimli Sinir Ağlarında Hiper Parametrelerin Etkisinin İncelenmesi (Analysis of the Effects of Hyperparameters in Convolutional Neural Networks)* [Master's Degree Thesis]. Hacettepe University.
- Kutlu, H. (2018). *Biyoistatistik Temelli Bilimsel Araştırmalara Derin Öğrenme Uygulamaları (Deep Learning Applications in Biostatistics Based Scientific Research)* [Master's Degree Thesis]. Yakındoğu University.
- Lamperti, M. & Pittiruti, M. (2013). Difficult peripheral veins: Turn on the lights. *British Journal of Anaesthesia*, 110(6), 888-891.
<https://doi.org/10.1093/bja/aet078>
- Lee, E. C., Jung, H., & Kim, D. (2011). New finger biometric method using near infrared imaging. *Sensors*, 11(3), 2319-2333.
<https://doi.org/10.3390/s110302319>
- Li, S., Deng, Y. Q., Zhu, Z. L., Hua, H. L., & Tao, Z. Z. (2021). A comprehensive review on radiomics and deep learning for nasopharyngeal carcinoma imaging. *Diagnostics (Basel)*, 11(9), 1523, 1-32.
doi: 10.3390/diagnostics11091523.

- Liang, H., Li, N., & Zhao, S. (2021). Salt and pepper noise removal method based on a detail-aware filter. *Symmetry*, 13(3), 515, 1-24.
<https://doi.org/10.3390/sym13030515>
- Liestoel, G. (2019). Augmented reality storytelling: Narrative design and reconstruction of a historical event in situ. *International Journal of Interactive Mobile Technologies*, 13(12), 96-209.
<https://doi.org/10.3991/ijim.v13i12.11560>
- Lingyu, W., & Leedham, G. (2006, November 22-24). *Near-and far-infrared imaging for vein pattern biometrics* [Conference presentation]. In Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS), pp. 52-52. IEEE. Sydney, NSW, Australia.
<https://doi.org/10.1109/AVSS.2006.80>
- Liu, W., Li, W., Sun, L., Zhang, L., & Chen, P. (2017, June 18-20). *Finger vein recognition based on deep learning* [Conference presentation]. In Proceedings of the 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 205-210. IEEE. Siem Reap, Cambodia.
<https://doi.org/10.1109/ICIEA.2017.8282842>
- Ma, J., Ren, X., Tsviatkou, V.Y., & Kanapelka, V. K. (2022). A novel fully parallel skeletonization algorithm. *Pattern Analysis and Applications*, 25, 169-188.
<https://doi.org/10.1007/s10044-021-01039-y>
- Mahmood, H. U., Ahmed, A., & Shahid, M. T. (2016). *PRECIS'VEINÉ non-invasive vein visualization device* [Bachelor's Degree Thesis]. National University of Sciences and Technology.
- Mangold, K., Shaw, J. A., & Vollmer, M. (2013). The physics of near-infrared photography. *European Journal of Physics*, 34(6), S51-S71.
<https://doi.org/10.1088/0143-0807/34/6/S51>
- Marques, O. (2011). *Practical image and video processing using MATLAB®*. John Wiley and Sons, Inc.

- MatConvNet Team. (2022, September 30). *MatConvNet: CNNs for MATLAB*.
<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.
- Mazuryk, T., & Gervautz, M. (1996). *Virtual reality history, applications, technology and future*. Vienna: Vienna University of Technology.
- McAndrew, A. (2022, September 30). *An introduction to digital image processing with Matlab* [Notes for SCM2511 Image Processing-1, School of Computer Science and Mathematics, Victoria University of Technology].
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.707.7827>
- Memiş, G. (2019). *İnsan Sağlığı Takibi İçin Giyilebilir Sensör Verilerinin Çok Kipli ve Füzyon Tabanlı Analizi (Multimodal and Fusion Based Analysis of Wearable Sensor Data for Human Health Monitoring)* [Doctoral Dissertation]. Başkent University.
- Mesci, Y. (2022, September 30). *YOLO Algoritmasını Anlamak*.
<https://medium.com/deep-learning-turkiye/yolo-algoritmas%C4%B1n%C4%B1-anlamak-290f2152808f>
- Milgram, P., Takemura H., Utsumi A., & Kishino F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*. Society of Photo-Optical Instrumentation Engineers (SPIE) Digital Library. Vol.2351, 282-292.
<https://doi.org/10.1117/12.197321>
- Mistry, D., & Banerjee A. (2017). Comparison of feature detection and matching approaches: SIFT and SURF. *Global Research and Development Journal for Engineering*, 2(4), 7-13.
https://www.researchgate.net/publication/314285930_Comparison_of_Feature_Detection_and_Matching_Approaches_SIFT_and_SURF

- Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image and Vision Computing*, 104, 104046. <https://doi.org/10.1016/j.imavis.2020.104046>
- Mohammad, A. H., Alwada'n, T., & Al-Momani, O. (2016). Arabic text categorization using support vector machine, naïve bayes and neural network. *GSTF Journal on Computing*, 5(1), 108-115. https://doi.org/10.5176/2251-3043_4.4.360
- Morera, Á., Sánchez, Á., Moreno, A. B., Sappa, Á. D., & Vélez, J. F. (2020). SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities. *Sensors*, 20(16), 4587,1-23. <https://doi.org/10.3390/s20164587>
- Muzaffer, G. (2016). *Anahtar Noktası Tabanlı Kopyala Yapıştır Sahteciliği Tespiti (Keypoint Based Copy Move Forgery Detection Methods)* [Master's Degree Thesis]. Karadeniz Technical University.
- Narman, A. E. (2016). *Android Studio ile Programlama* (13th ed.). KODLAB Yayın Dağıtım Yazılım ve Eğitim Hizmetleri San. ve Tic. Ltd. Şti.
- Okura, F., Akaguma, T., Sato, T., & Yokoya, N. (2014, September 10-12). Indirect augmented reality considering real-world illumination change [Poster]. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 287-288, IEEE. Munich, Germany. <https://doi.org/10.1109/ISMAR.2014.6948453>.
- Open Handset Alliance. (2022, September 30). *FAQ*. https://www.openhandsetalliance.com/oha_faq.html
- Oral, M., & Gençal, M. C. (2016, May 11-13). *SIFT Yöntemini Kullanarak Madeni Para Tanıma* [Conference presentation]. In Proceedings of the Electrical-Electronics and Computer Symposium (EEC), pp. 55-58, Tokat, Turkey.
- Osmanoğlu, U. Ö. (2016). *Görüntü işleme ve analizinin tıpta kullanımı ve bir uygulama (Use of Image Processing and Analysis in Medicine and an Application)* [Master's Degree Thesis]. Eskişehir Osmangazi University.

- Önder, M., & Mermerkaya, A. O. (2013). *Merhaba Android* (4th ed.). Pusula Yayıncılık İletişim Yazılım İmalat San.Tic.Ltd.Şti.
- Öksüz, A. (2022, September 30). *You Only Look Once (YOLO) Nedir?*
<https://alioksuz.medium.com/you-only-look-once-yolo-nedir-eab14530484d>
- Özgen, N., & Sarıtaş, M. (2022, September 30). *SIFT Metodu ile Hedef Takibi*.
https://www.emo.org.tr/ekler/77ce700a091f3d6_ek.pdf
- Öztürk, Ş., & Akdemir, B. (2015). Comparison of edge detection algorithms for texture analysis on glass production. *Procedia-Social and Behavioral Sciences*, 195, 2675-2682.
<https://doi.org/10.1016/j.sbspro.2015.06.477>
- Peddie, J. (2017). *Augmented reality: Where we will all live*. International Publishing Springer, Cham.
<https://doi.org/10.1007/978-3-319-54502-8>
- Perry, A. M., Caviness, A. C., & Hsu, D. C. (2011). Efficacy of a near-infrared light device in pediatric intravenous cannulation: A randomized controlled trial. *Pediatric Emergency Care*, 27(1), 5-10.
<https://doi.org/10.1097/PEC.0b013e3182037caf>
- Petrou, M., & Petrou, C. (2010). *Image processing: The fundamentals* (2nd ed.). John Wiley and Sons, Inc.
- Pramoditha, R. (2022, September 30). *How to Choose the Right Activation Function for Neural Networks*. Medium. <https://towardsdatascience.com/how-to-choose-the-right-activation-function-for-neural-networks-3941ff0e6f9c>
- Qeynet. (2022, September 30). *Unconditionally Secure Key Exchange In The Post-Quantum Era*. <https://www.qeynet.com/>
- Quaresima, V., & Ferrari, M. (2016). Medical near infrared spectroscopy: A prestigious history and a bright future. *NIR News*, 27(1), 10-13.
<https://doi.org/10.1255/nirn.1575>

- Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods*, 22(1), 46-68.
<https://doi.org/10.1177/1094428116658959>
- 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.
<https://doi.org/10.1007/s11760-016-0977-3>
- Reasat, T., & Shahnaz, C. (2017, December 21-23). *Detection of inferior myocardial infarction using shallow convolutional neural networks* [Conference presentation]. In Proceedings of the IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 718-721. IEEE. Dhaka, Bangladesh.
<https://doi.org/10.1109/R10-HTC.2017.8289058>.
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *Arxiv preprint*.
<https://doi.org/10.48550/arXiv.1804.02767>
- Roizen, M. F. & Oz, M. C. (2005). *You-The Owner's Manual*. Beyaz Balina Yayınları.
- Royal Society of Chemistry. (2022, September 30). *Introduction to Spectroscopy*.
<https://edu.rsc.org/download?ac=11384>
- Rubik's Code. (2022, September 30). *Introduction to Convolutional Neural Networks*. <https://rubikscodene.net/2018/02/26/introduction-to-convolutional-neural-networks/>
- Sanchez-Morago, G. V., Sánchez, C. M., Villafranca, C. A., Cantero, A. J., Migallón, B. M., & Carrero, C. M. (2010). Viewing veins with AccuVein AV300. *Revista de Enfermería (Nursing Journal)*, 33(1), 33-38. Retrieved September 02, 2022, from <https://europepmc.org/article/med/20201197>
- Sawant, S. (2022, September 30). *Turn webcam to IR sensitive camera*.
<http://www.instructables.com/id/turn-webcam-to-IR-sensitive-camera/>

- Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, 85(1), 6-27.
<https://doi.org/10.1016/j.neuroimage.2013.05.004>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press, New York.
- Sharma, S. (2022, September 30). *Activation functions in neural networks: Sigmoid, tanh, softmax, ReLU, leaky ReLU explained!!!*. Medium.
<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
- Shrotri, A., Rethrekar, S. C., Patil, M. H., & Kore, S. N. (2010, February). *IR-webcam imaging and vascular pattern analysis towards hand vein authentication* [Conference presentation]. In Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE), pp. 876-880. IEEE. Singapore.
<https://doi.org/10.1109/ICCAE.2010.5451897>.
- Seymen, V., Boyraz, Ö. F., Çetin, Ö., & Yıldız, M. Z. (2014, June). *Kızılötesi görüntülerden el-damar örüntüsü tespiti* [Conference presentation]. In Proceedings of the 2nd International Symposium on Innovative Technologies in Engineering and Science (ISITES), pp. 722-729, Karabük. Retrieved September 02, 2022, from https://www.researchgate.net/publication/262600278_Kizilotesi_Goruntulerden_El-Damar_Oruntusu_Tespiti
- Sierro, A., Ferrez, P., & Roduit, P. (2015, September 09-11). *Contact-less palm/finger vein biometrics* [Conference presentation]. In Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1-12. IEEE. Darmstadt, Germany.
<https://doi.org/10.1109/BIOSIG.2015.7314596>.

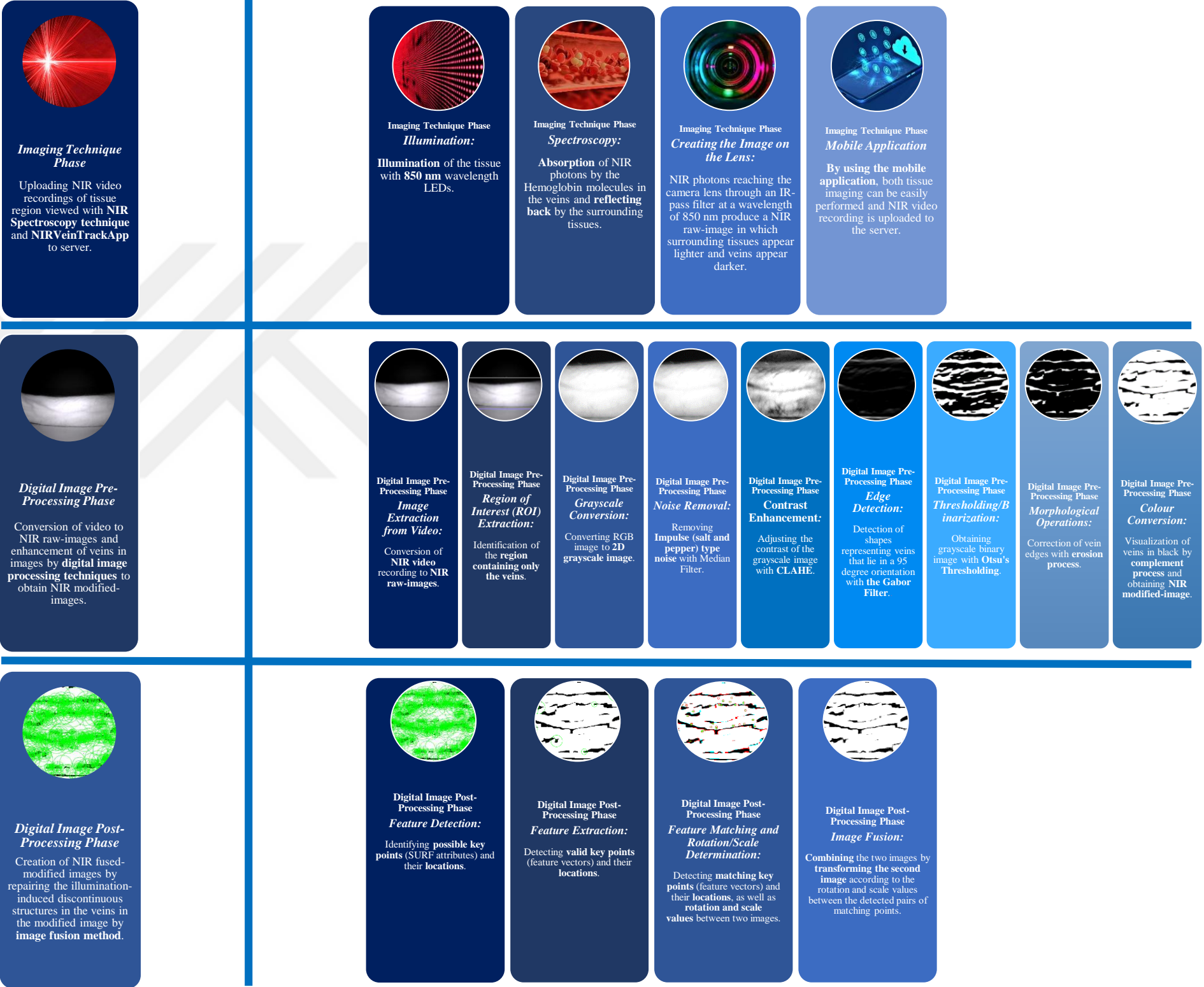
- Singhal A. & Cowie, M. (2022, September 30). *What is e-Health?* ESC European Society of Cardiology. <https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-18/what-is-e-health>
- Skalski, P. (2022, September 30). *Free To Use Online Tool For Labelling Photos.* Makesense, Alpha. <https://www.makesense.ai/>
- SolatubeHome. (2022, September 30). <https://solatubehome.com/the-truth-about-infrared-light>
- Şahin, Y. H. (2017). *Vascular Segmentation of Brain MR Angiography Images Using Convolutional Neural Networks* [Master's Degree Thesis]. İstanbul Technical University.
- Şeker, K., & Engin, M. (2017). Deep tissue near-infrared imaging for vascular network analysis. *Journal of Innovative Optical Health Sciences, Journal of Innovative Optical Health Sciences*, 10(3), 1650051. <https://doi.org/10.1142/S1793545816500516>
- Şengür, A., Akılotu, B. N., Tuncer, S. A., Kadiroğlu, Z., Yavuzkılıç, S., Budak, Ü., & Deniz, E. (2018, May 02-05). *Optic disc determination in retinal images with deep features* [Conference presentation]. In *Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4. IEEE. İzmir, Turkey. <https://doi.org/10.1109/SIU.2018.8404339>.
- Tan, Z. (2019). *Derin Öğrenme Yardımıyla Araç Sınıflandırma (Vehicle Classification with Deep Learning)* [Master's Degree Thesis]. Fırat University.
- Teledyne FLIR. (2022, September 30). *Key differences between CCD and CMOS imaging sensors.* <https://www.flir.com/support-center/iis/machine-vision/knowledge-base/key-differences-between-ccd-and-cmos-imaging-sensors/>
- Tiwari, N. (2022, September 30). *YOLOv3 custom object detection with Transfer Learning.* Medium. <https://tiwarinitin1999.medium.com/yolov3-custom-object-detection-with-transfer-learning-47186c8f166d>

- Tükel, T. (2019). *Görüntü İşleme ve Evrişimsel Sinir Ağları Kullanılarak Diyabetik Retinopati Hastalığının Tespiti (Evaluation of Retinal Images and Detection of Diabetic Retinopathy Disease Using Convolutional Neural Networks)* [Master's Degree Thesis]. Süleyman Demirel University.
- Verma, D., & Dubey, S. (2015). Processing and enhancement of palm vein image in vein pattern recognition system. *International Journal of Computer Science and Mobile Computing*, 4(4), 137-141. Retrieved September 02, 2022, from <https://www.ijcsmc.com/docs/papers/April2015/V4I4201502.pdf>
- Violante, A. (2022, September 30). *Forget the robot apocalypse, focus on building more useful AI: A Q&A with Gary Marcus*. Behavioral Scientist. <https://behavioralscientist.org/forget-the-robot-apocalypse-focus-on-building-more-useful-ai-a-qa-with-gary-marcus/>
- Wadhwani, M., Sharma, A. D., Pillai, A., Pisal, N., & Bhowmick, M. (2015). Vein detection system using infrared light. *International Journal of Scientific and Engineering Research*, 6(12), 780-786. Retrieved September 02, 2022, from <https://www.ijser.org/researchpaper/Vein-Detection-System-using-Infrared-Light.pdf>
- Wahr, J. A., Tremper, K. K., Samra, S., & Delpy, D. T. (1996). Near-infrared spectroscopy: Theory and applications. *Journal of Cardiothoracic and Vascular Anesthesia*, 10(3), 406-418. [https://doi.org/10.1016/S1053-0770\(96\)80107-8](https://doi.org/10.1016/S1053-0770(96)80107-8)
- Watanabe, T., Sekine, R., Mizuno, T., Miwa, M. (2016). Development of portable, wireless and smartphone controllable near-infrared spectroscopy system. In: Q., Luo, L., Li, D., Harrison, H., Shi, & D. Bruley (Eds.) *Oxygen transport to tissue XXXVIII* (pp. 385-392). Advances in Experimental Medicine and Biology, vol 923. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-38810-6_50
- Wells, S. (2014, August 27). *Keeping on point, a real botox game changer*. 25A Magazine. <https://www.accuvein.com/assets/Screen-Shot-2014-09-10-at-1.28.11-PM.jpg>

- Wither, J., Tsai, Y. T., & Azuma, R. (2011). Indirect Augmented Reality. *Computers & Graphics*, 35(4), 810-822.
<https://doi.org/10.1016/j.cag.2011.04.010>
- Xueyan, L., & Shuxu, G. (2008). The fourth biometric-vein recognition. In: P. Y., Yin (Ed.) *Pattern recognition techniques, technology and applications* (pp. 537-546). InTech. Retrieved September 02, 2022, from <https://www.intechopen.com/chapters/5801>
- Yella, H. M. (2022, September 30). *Why is blood red?* Medical News Today.
<https://www.medicalnewstoday.com/articles/318171>
- Yılmaz, A., & Kaya, U. (2020). *Derin öğrenme* (2nd ed.). Kodlab Yayın Dağıtım Yazılım ve Eğitim Hizmetleri San. ve Tic. Ltd. Şti.
- Yılmaz, K. (2014). *Taşıyabilir bir oftalmoskop tasarımı ve bu cihazın teşhis ve tanılamada kullanımı (The design of a new portable ophtalmoscope and utilization of the device for diagnosis and identification)* [Doctoral Dissertation]. Ege University.
- Zhong, H., Kanhere, S. S., & Chou, C. T. (2017, March 13-17). *VeinDeep: Smartphone unlock using vein patterns* [Conference presentation]. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 2-10. IEEE. Kona, HI, United States.
<https://doi.org/10.1109/PERCOM.2017.7917845>.

APPENDICES

Appendix A – General Overview of the Proposed System’s Six Phases in 10-minutes



Appendix A – Continues



Appendix B – Classification Phase Post-Test Results for Period-1

		Dataset-1						Dataset-2						Dataset-3							
		200 / 200						400 / 400						720 / 720 (includes noise/rot5/rot10/scale)							
		50 / 50						100 / 100						180 / 180 (includes noise/rot5/rot10/scale)							
		Model-1		Model-2		Model-3		Model-1		Model-2		Model-3		Model-1		Model-2		Model-3			
Input Image Size:		28x28		64x64		128x128		28x28		64x64		128x128		28x28		64x64		128x128			
Learning Rate:		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001			
Epoch:		14000		4000		1000		10204		3459		800		1000		1000		400			
Training Duration (sec/1 epoch):		1.4581 sec		3.0629 sec		38.5345 sec		1.9551 sec		4.6642 sec		75.4462 sec		3.3552 sec		18.9290 sec		147.9228 sec			
The NIR Fused-Modified Image Name		Predicted Class				Predicted Class				Predicted Class				Predicted Class				Predicted Class			
		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm		
1	Fused_Modified_Period1_Img0-1_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9996	0.0003	0.9371	0.0628	0.9998	0.0001	0.9991	0.0008	0.9850	0.0149	0.9998	0.0001	0.9996	0.0003		
2	Fused_Modified_Period1_Img50-51_Rightforearm.bmp	0.9999	0.0000	1.0000	0.0000	0.9999	0.0000	0.9998	0.0001	1.0000	0.0000	0.9999	0.0000	0.9998	0.0001	1.0000	0.0000	0.9999	0.0000		
3	Fused_Modified_Period1_Img100-101_Rightforearm.bmp	0.9989	0.0010	0.9999	0.0000	0.9999	0.0000	0.9989	0.0010	1.0000	0.0000	0.9999	0.0000	0.9989	0.0010	1.0000	0.0000	0.9999	0.0000		
4	Fused_Modified_Period1_Img150-151_Rightforearm.bmp	0.9998	0.0001	1.0000	0.0000	0.9999	0.0000	0.9954	0.0045	1.0000	0.0000	0.9999	0.0000	0.9921	0.0078	1.0000	0.0000	0.9999	0.0000		
5	Fused_Modified_Period1_Img200-201_Rightforearm.bmp	0.9995	0.0004	0.9999	0.0000	0.9999	0.0000	0.9988	0.0011	1.0000	0.0000	0.9998	0.0001	0.9969	0.0030	1.0000	0.0000	0.9999	0.0000		
6	Fused_Modified_Period1_Img250-251_Rightforearm.bmp	0.9997	0.0002	0.9996	0.0003	0.9999	0.0000	0.9996	0.0003	1.0000	0.0000	0.9999	0.0000	0.9994	0.0005	0.9999	0.0000	0.9999	0.0000		
7	Fused_Modified_Period1_Img300-301_Rightforearm.bmp	0.9992	0.0007	0.9999	0.0000	0.9998	0.0001	0.9840	0.0159	1.0000	0.0000	0.9996	0.0003	0.9795	0.0204	0.9999	0.0000	0.9998	0.0001		
8	Fused_Modified_Period1_Img350-351_Rightforearm.bmp	0.9981	0.0018	0.9995	0.0004	0.9999	0.0000	0.9427	0.0572	0.9999	0.0000	0.9998	0.0001	0.8901	0.1098	0.9999	0.0000	0.9999	0.0000		
9	Fused_Modified_Period1_Img400-401_Rightforearm.bmp	0.9946	0.0053	0.9998	0.0001	0.9999	0.0000	0.9944	0.0055	1.0000	0.0000	0.9999	0.0000	0.9879	0.0120	0.9999	0.0000	0.9999	0.0000		
10	Fused_Modified_Period1_Img450-451_Rightforearm.bmp	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9996	0.0003	0.9999	0.0000	0.9999	0.0000	0.9996	0.0003	0.9999	0.0000	0.9996	0.0003		
11	Fused_Modified_Period1_Img500-501_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9957	0.0042	0.9998	0.0001	0.9999	0.0000	0.9976	0.0023	0.9999	0.0000	0.9998	0.0001		
12	Fused_Modified_Period1_Img550-551_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9980	0.0019	0.9999	0.0000	0.9999	0.0000	0.9985	0.0014	0.9999	0.0000	0.9990	0.0009		
13	Fused_Modified_Period1_Img600-601_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9953	0.0046	0.9999	0.0000	0.9999	0.0000	0.9959	0.0040	0.9999	0.0000	0.9553	0.0446		
14	Fused_Modified_Period1_Img650-651_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9952	0.0047	0.9991	0.0008	0.9998	0.0001	0.9979	0.0020	0.9993	0.0006	0.9988	0.0011		
15	Fused_Modified_Period1_Img700-701_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9989	0.0010	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9125	0.0874		
16	Fused_Modified_Period1_Img750-751_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9962	0.0037	0.9999	0.0000	0.9997	0.0002	0.9996	0.0003	0.9999	0.0000	0.9278	0.0721		
17	Fused_Modified_Period1_Img800-801_Rightforearm.bmp	0.9999	0.0000	1.0000	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9455	0.0544		
18	Fused_Modified_Period1_Img850-851_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9032	0.0967		
19	Fused_Modified_Period1_Img900-901_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.7576	0.2423		
20	Fused_Modified_Period1_Img930-931_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9112	0.0887		
21	Fused_Modified_Period1_Img950-951_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.7932	0.2067		

Appendix B – Continues

		Dataset-1						Dataset-2						Dataset-3					
		200 / 200						400 / 400						720 / 720 (includes noise/rot5/rot10/scale)					
		50 / 50						100 / 100						180 / 180 (includes noise/rot5/rot10/scale)					
Train (Right Forearm/Left Forearm):		Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3						
Pre-test (Right Forearm/Left Forearm):		28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128						
CNN Model:		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001						
Input Image Size:		14000	4000	1000	10204	3459	800	1000	1000	400	1000	1000	400						
Learning Rate:		1.4581 sec	3.0629 sec	38.5345 sec	1.9551 sec	4.6642 sec	75.4462 sec	3.3552 sec	18.9290 sec	147.9228 sec	3.3552 sec	18.9290 sec	147.9228 sec						
Epoch:		Predicted Class						Predicted Class						Predicted Class					
Training Duration (sec/1 epoch):		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm
The NIR Fused-Modified Image Name		0.0033	0.9966	0.1223	0.8776	0.0000	0.9999	0.0000	0.9999	0.4904	0.5095	0.0000	0.9999	0.0000	0.9999	0.3682	0.6317	0.0000	0.9999
1	Fused_Modified_Period1_Img0-1_Leftforearm.bmp	0.2010	0.7989	0.5539	0.4460	0.0000	0.9999	0.0009	0.9990	0.0802	0.9197	0.0000	0.9999	0.0016	0.9983	0.4838	0.5161	0.0000	0.9999
2	Fused_Modified_Period1_Img50-51_Leftforearm.bmp	0.9974	0.0025	0.8815	0.1184	0.0000	0.9999	0.0008	0.9991	0.0000	0.9999	0.0000	0.9999	0.0023	0.9976	0.0013	0.9986	0.0001	0.9998
3	Fused_Modified_Period1_Img100-101_Leftforearm.bmp	0.5934	0.4065	0.5804	0.4195	0.0000	0.9999	0.0002	0.9997	0.0139	0.9860	0.0000	0.9999	0.0003	0.9996	0.1511	0.8488	0.0000	0.9999
4	Fused_Modified_Period1_Img150-151_Leftforearm.bmp	0.1088	0.8911	0.0324	0.9675	0.0000	0.9999	0.0001	0.9998	0.2623	0.7376	0.0000	0.9999	0.0004	0.9995	0.1200	0.8799	0.0000	0.9999
5	Fused_Modified_Period1_Img200-201_Leftforearm.bmp	0.5841	0.4158	0.0779	0.9220	0.0000	0.9999	0.0002	0.9997	0.0020	0.9979	0.0007	0.9992	0.0005	0.9994	0.1585	0.8414	0.0000	0.9999
6	Fused_Modified_Period1_Img250-251_Leftforearm.bmp	0.0833	0.9166	0.8523	0.1476	0.0002	0.9997	0.0004	0.9995	0.0128	0.9871	0.0034	0.9965	0.0002	0.9997	0.4197	0.5802	0.0000	0.9999
7	Fused_Modified_Period1_Img300-301_Leftforearm.bmp	0.0436	0.9563	0.0589	0.9410	0.0001	0.9998	0.0000	0.9999	0.0004	0.9995	0.0036	0.9963	0.0000	0.9999	0.0272	0.9727	0.0000	0.9999
8	Fused_Modified_Period1_Img350-351_Leftforearm.bmp	0.0536	0.9463	0.0495	0.9504	0.0003	0.9996	0.0000	0.9999	0.0177	0.9822	0.0057	0.9942	0.0000	0.9999	0.2782	0.7217	0.0000	0.9999
9	Fused_Modified_Period1_Img400-401_Leftforearm.bmp	0.0908	0.9091	0.0270	0.9729	0.0000	0.9999	0.0001	0.9998	0.0028	0.9971	0.0022	0.9977	0.0001	0.9998	0.4281	0.5718	0.0001	0.9998
10	Fused_Modified_Period1_Img450-451_Leftforearm.bmp	0.0210	0.9789	0.0070	0.9929	0.0000	0.9999	0.0004	0.9995	0.0105	0.9894	0.0008	0.9991	0.0001	0.9998	0.2833	0.7166	0.0000	0.9999
11	Fused_Modified_Period1_Img500-501_Leftforearm.bmp	0.2959	0.7040	0.0357	0.9642	0.0000	0.9999	0.0017	0.9982	0.0009	0.9990	0.0040	0.9959	0.0004	0.9995	0.2309	0.7690	0.0000	0.9999
12	Fused_Modified_Period1_Img550-551_Leftforearm.bmp	0.5737	0.4262	0.1307	0.8692	0.0011	0.9988	0.0002	0.9997	0.0003	0.9996	0.0315	0.9684	0.0000	0.9999	0.0446	0.9553	0.0000	0.9999
13	Fused_Modified_Period1_Img600-601_Leftforearm.bmp	0.2278	0.7721	0.6085	0.3914	0.0013	0.9986	0.0001	0.9998	0.0000	0.9999	0.0300	0.9699	0.0000	0.9999	0.0054	0.9945	0.0006	0.9993
14	Fused_Modified_Period1_Img650-651_Leftforearm.bmp	0.8230	0.1769	0.7443	0.2556	0.0004	0.9995	0.0027	0.9972	0.0000	0.9999	0.0236	0.9763	0.0007	0.9992	0.0060	0.9939	0.0000	0.9999
15	Fused_Modified_Period1_Img700-701_Leftforearm.bmp	0.9359	0.0640	0.8388	0.1611	0.0001	0.9998	0.0011	0.9988	0.0000	0.9999	0.0091	0.9908	0.0003	0.9996	0.0018	0.9981	0.0000	0.9999
16	Fused_Modified_Period1_Img750-751_Leftforearm.bmp	0.8966	0.1033	0.5488	0.4511	0.0000	0.9999	0.0002	0.9997	0.0000	0.9999	0.0052	0.9947	0.0000	0.9999	0.0068	0.9931	0.0000	0.9999
17	Fused_Modified_Period1_Img800-801_Leftforearm.bmp	0.8574	0.1425	0.3511	0.6488	0.0000	0.9999	0.0006	0.9993	0.0000	0.9999	0.0060	0.9939	0.0001	0.9998	0.0009	0.9990	0.0000	0.9999
18	Fused_Modified_Period1_Img850-851_Leftforearm.bmp	0.9604	0.0395	0.4444	0.5555	0.0001	0.9998	0.0113	0.9886	0.0000	0.9999	0.0153	0.9846	0.0026	0.9973	0.0037	0.9962	0.0002	0.9997
19	Fused_Modified_Period1_Img900-901_Leftforearm.bmp	0.9576	0.0423	0.8497	0.1502	0.0006	0.9993	0.0041	0.9958	0.0003	0.9996	0.0582	0.9417	0.0013	0.9986	0.0047	0.9952	0.0026	0.9973
20	Fused_Modified_Period1_Img930-931_Leftforearm.bmp	0.8707	0.1292	0.9936	0.0063	0.0017	0.9982	0.0158	0.9841	0.0000	0.9999	0.0802	0.9197	0.0065	0.9934	0.0007	0.9992	0.0013	0.9986
21	Fused_Modified_Period1_Img950-951_Leftforearm.bmp																		

Appendix C – Classification Phase Post-Test Results for Period -2

		Dataset-1						Dataset-2						Dataset-3					
		200 / 200						400 / 400						720 / 720 (includes noise/rot5/rot10/scale)					
		50 / 50						100 / 100						180 / 180 (includes noise/rot5/rot10/scale)					
		Model-1		Model-2		Model-3		Model-1		Model-2		Model-3		Model-1		Model-2		Model-3	
Input Image Size:		28x28		64x64		128x128		28x28		64x64		128x128		28x28		64x64		128x128	
Learning Rate:		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001		0.0001	
Epoch:		14000		4000		1000		10204		3459		800		1000		1000		400	
Training Duration (sec/1 epoch):		1.4581 sec		3.0629 sec		38.5345 sec		1.9551 sec		4.6642 sec		75.4462 sec		3.3552 sec		18.9290 sec		147.9228 sec	
The NIR Fused-Modified Image Name		Predicted Class						Predicted Class						Predicted Class					
		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm
1	Fused_Modified_Period2_Img0-1_Rightforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9996	0.0003
2	Fused_Modified_Period2_Img50-51_Rightforearm.bmp	0.9990	0.0009	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9995	0.0004
3	Fused_Modified_Period2_Img100-101_Rightforearm.bmp	0.9987	0.0012	0.9994	0.0005	0.9999	0.0000	0.9999	0.0000	0.9975	0.0024	0.9999	0.0000	0.9999	0.0000	0.9993	0.0006	0.9990	0.0009
4	Fused_Modified_Period2_Img150-151_Rightforearm.bmp	0.9996	0.0003	0.9994	0.0005	0.9999	0.0000	0.9910	0.0089	0.9999	0.0000	0.9998	0.0001	0.9946	0.0053	0.9998	0.0001	0.9976	0.0023
5	Fused_Modified_Period2_Img200-201_Rightforearm.bmp	0.9983	0.0016	0.9990	0.0009	0.9999	0.0000	0.9895	0.0104	0.9999	0.0000	0.9999	0.0000	0.9915	0.0084	0.9999	0.0000	0.9988	0.0011
6	Fused_Modified_Period2_Img250-251_Rightforearm.bmp	0.9997	0.0002	0.9998	0.0001	0.9999	0.0000	0.9991	0.0008	1.0000	0.0000	0.9998	0.0001	0.9995	0.0004	0.9999	0.0000	0.9854	0.0145
7	Fused_Modified_Period2_Img300-301_Rightforearm.bmp	0.9985	0.0014	0.9994	0.0005	0.9999	0.0000	0.9684	0.0315	0.9999	0.0000	0.9999	0.0000	0.9888	0.0111	0.9999	0.0000	0.9910	0.0089
8	Fused_Modified_Period2_Img350-351_Rightforearm.bmp	0.9995	0.0004	0.9997	0.0002	0.9999	0.0000	0.9911	0.0088	0.9999	0.0000	0.9997	0.0002	0.9976	0.0023	0.9999	0.0000	0.9894	0.0105
9	Fused_Modified_Period2_Img400-401_Rightforearm.bmp	0.9976	0.0023	0.9995	0.0004	0.9999	0.0000	0.9961	0.0038	0.9999	0.0000	0.9996	0.0003	0.9983	0.0016	0.9999	0.0000	0.9894	0.0105
10	Fused_Modified_Period2_Img450-451_Rightforearm.bmp	0.9778	0.0221	0.9634	0.0365	0.9999	0.0000	0.9885	0.0114	0.9999	0.0000	0.9995	0.0004	0.9768	0.0231	0.9999	0.0000	0.8624	0.1375
11	Fused_Modified_Period2_Img500-501_Rightforearm.bmp	0.9886	0.0113	0.9983	0.0016	0.9998	0.0001	0.9818	0.0181	0.9999	0.0000	0.9990	0.0009	0.9634	0.0365	0.9999	0.0000	0.7784	0.2215
12	Fused_Modified_Period2_Img550-551_Rightforearm.bmp	0.9949	0.0050	0.9996	0.0003	0.9999	0.0000	0.9956	0.0043	0.9999	0.0000	0.9993	0.0006	0.9915	0.0084	0.9999	0.0000	0.9155	0.0844
13	Fused_Modified_Period2_Img600-601_Rightforearm.bmp	0.9939	0.0060	0.9970	0.0029	0.9998	0.0001	0.9717	0.0282	0.9999	0.0000	0.9990	0.0009	0.9670	0.0329	0.9999	0.0000	0.5886	0.4113
14	Fused_Modified_Period2_Img650-651_Rightforearm.bmp	0.9857	0.0142	0.9955	0.0044	0.9996	0.0003	0.9988	0.0011	0.9999	0.0000	0.9976	0.0023	0.9981	0.0018	0.9999	0.0000	0.2996	0.7003
15	Fused_Modified_Period2_Img700-701_Rightforearm.bmp	0.9668	0.0331	0.7522	0.2477	0.9998	0.0001	0.9937	0.0062	0.9999	0.0000	0.9980	0.0019	0.9925	0.0074	0.9998	0.0001	0.1554	0.8445
16	Fused_Modified_Period2_Img750-751_Rightforearm.bmp	0.9942	0.0057	0.9624	0.0375	0.9997	0.0002	0.9975	0.0024	0.9999	0.0000	0.9985	0.0014	0.9963	0.0036	0.9999	0.0000	0.0519	0.9480
17	Fused_Modified_Period2_Img800-801_Rightforearm.bmp	0.9990	0.0009	0.9975	0.0024	0.9999	0.0000	0.9996	0.0003	0.9999	0.0000	0.9996	0.0003	0.9997	0.0002	0.9999	0.0000	0.0544	0.9455
18	Fused_Modified_Period2_Img850-851_Rightforearm.bmp	0.9945	0.0054	0.9934	0.0065	0.9999	0.0000	0.9996	0.0003	0.9999	0.0000	0.9997	0.0002	0.9993	0.0006	0.9999	0.0000	0.0301	0.9698
19	Fused_Modified_Period2_Img900-901_Rightforearm.bmp	0.9994	0.0005	0.9984	0.0015	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.0179	0.9820
20	Fused_Modified_Period2_Img930-931_Rightforearm.bmp	0.9990	0.0009	0.9870	0.0129	0.9999	0.0000	0.9997	0.0002	0.9999	0.0000	0.9992	0.0007	0.9998	0.0001	0.9993	0.0006	0.0119	0.9880
21	Fused_Modified_Period2_Img950-951_Rightforearm.bmp	0.9964	0.0035	0.9687	0.0312	0.9999	0.0000	0.9998	0.0001	0.9999	0.0000	0.9996	0.0003	0.9998	0.0001	0.9997	0.0002	0.0181	0.9818

Appendix C – Continues

		Dataset-1						Dataset-2						Dataset-3					
Train (Right Forearm/Left Forearm):		200 / 200						400 / 400						720 / 720 (includes noise/rot5/rot10/scale)					
Pre-test (Right Forearm/Left Forearm):		50 / 50						100 / 100						180 / 180 (includes noise/rot5/rot10/scale)					
CNN Model:		Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3
Input Image Size:		28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128
Learning Rate:		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Epoch:		14000	4000	1000	10204	3459	800	1000	1000	400	1000	1000	400	1000	1000	400	1000	1000	400
Training Duration (sec/1 epoch):		1.4581 sec	3.0629 sec	38.5345 sec	1.9551 sec	4.6642 sec	75.4462 sec	3.3552 sec	18.9290 sec	147.9228 sec	3.3552 sec	18.9290 sec	147.9228 sec	3.3552 sec	18.9290 sec	147.9228 sec	3.3552 sec	18.9290 sec	147.9228 sec
		Predicted Class						Predicted Class						Predicted Class					
The NIR Fused-Modified Image Name		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm
1	Fused_Modified_Period2_Img0-1_Leftforearm.bmp	0.9948	0.0051	0.9973	0.0026	0.0000	0.9999	0.0011	0.9988	0.0000	0.9999	0.0001	0.9998	0.0005	0.9994	0.0003	0.9996	0.0001	0.9998
2	Fused_Modified_Period2_Img50-51_Leftforearm.bmp	0.9993	0.0006	0.9858	0.0141	0.0000	0.9999	0.0159	0.9840	0.0000	0.9999	0.0007	0.9992	0.0050	0.9949	0.0001	0.9998	0.0007	0.9992
3	Fused_Modified_Period2_Img100-101_Leftforearm.bmp	0.9997	0.0002	0.9977	0.0022	0.0000	0.9999	0.0797	0.9202	0.0000	0.9999	0.0005	0.9994	0.0252	0.9747	0.0000	0.9999	0.0028	0.9971
4	Fused_Modified_Period2_Img150-151_Leftforearm.bmp	0.9975	0.0024	0.9997	0.0002	0.0000	0.9999	0.0105	0.9894	0.0000	0.9999	0.0015	0.9984	0.0021	0.9978	0.9999	0.0000	0.0053	0.9946
5	Fused_Modified_Period2_Img200-201_Leftforearm.bmp	0.9998	0.0001	0.9999	0.0000	0.0000	0.9999	0.0461	0.9538	0.0000	1.0000	0.0011	0.9988	0.0140	0.9859	0.0000	0.9999	0.0044	0.9955
6	Fused_Modified_Period2_Img250-251_Leftforearm.bmp	0.9996	0.0003	0.9998	0.0001	0.0000	0.9999	0.0018	0.9981	0.0000	1.0000	0.0005	0.9994	0.0003	0.9996	0.0000	0.9999	0.0022	0.9977
7	Fused_Modified_Period2_Img300-301_Leftforearm.bmp	0.9996	0.0003	0.9999	0.0000	0.0000	0.9999	0.0048	0.9951	0.0000	1.0000	0.0024	0.9975	0.0007	0.9992	0.0000	0.9999	0.0031	0.9968
8	Fused_Modified_Period2_Img350-351_Leftforearm.bmp	0.9993	0.0006	0.9999	0.0000	0.0000	0.9999	0.0569	0.9430	0.0000	0.9999	0.0025	0.9974	0.0091	0.9908	0.0000	0.9999	0.0150	0.9849
9	Fused_Modified_Period2_Img400-401_Leftforearm.bmp	0.9995	0.0004	0.9998	0.0001	0.0000	0.9999	0.0148	0.9851	0.0000	0.9999	0.0061	0.9938	0.0018	0.9981	0.0000	0.9999	0.0193	0.9806
10	Fused_Modified_Period2_Img450-451_Leftforearm.bmp	0.9995	0.0004	0.9992	0.0007	0.0002	0.9997	0.0191	0.9808	0.0000	0.9999	0.0183	0.9816	0.0051	0.9948	0.0001	0.9998	0.0074	0.9925
11	Fused_Modified_Period2_Img500-501_Leftforearm.bmp	0.9984	0.0015	0.9996	0.0003	0.0009	0.9990	0.3912	0.6087	0.0000	0.9999	0.0468	0.9531	0.0986	0.9013	0.0001	0.9998	0.0182	0.9817
12	Fused_Modified_Period2_Img550-551_Leftforearm.bmp	0.9973	0.0026	0.9999	0.0000	0.0004	0.9995	0.0202	0.9797	0.0000	0.9999	0.0233	0.9766	0.0079	0.9920	0.0000	0.9999	0.0228	0.9771
13	Fused_Modified_Period2_Img600-601_Leftforearm.bmp	0.9995	0.0004	0.9999	0.0000	0.0010	0.9989	0.0276	0.9723	0.0000	0.9999	0.0665	0.9334	0.0097	0.9902	0.0002	0.9997	0.0329	0.9670
14	Fused_Modified_Period2_Img650-651_Leftforearm.bmp	0.9996	0.0003	0.9999	0.0000	0.0033	0.9966	0.1037	0.8962	0.0000	0.9999	0.1209	0.8790	0.0330	0.9669	0.0002	0.9997	0.0333	0.9666
15	Fused_Modified_Period2_Img700-701_Leftforearm.bmp	0.9991	0.0008	0.9999	0.0000	0.0005	0.9994	0.1207	0.8792	0.0000	0.9999	0.0424	0.9575	0.0609	0.9390	0.0000	0.9999	0.0184	0.9815
16	Fused_Modified_Period2_Img750-751_Leftforearm.bmp	0.9976	0.0023	0.9998	0.0001	0.0007	0.9992	0.0026	0.9973	0.9999	0.0000	0.0654	0.9345	0.0007	0.9992	0.0000	0.9999	0.3096	0.6903
17	Fused_Modified_Period2_Img800-801_Leftforearm.bmp	0.9955	0.0044	0.9711	0.0288	0.0004	0.9995	0.1066	0.8933	0.0000	0.9999	0.0281	0.9718	0.0279	0.9720	0.9998	0.0001	0.0066	0.9933
18	Fused_Modified_Period2_Img850-851_Leftforearm.bmp	0.9986	0.0013	0.9703	0.0296	0.0483	0.9516	0.1086	0.8913	0.0000	0.9999	0.5899	0.4100	0.0771	0.9228	0.0001	0.9998	0.6888	0.3111
19	Fused_Modified_Period2_Img900-901_Leftforearm.bmp	0.9990	0.0009	0.9801	0.0198	0.0040	0.9959	0.1198	0.8801	0.9999	0.0000	0.1408	0.8591	0.0668	0.9331	0.0003	0.9996	0.0200	0.9799
20	Fused_Modified_Period2_Img930-931_Leftforearm.bmp	0.9927	0.0072	0.8345	0.1654	0.0147	0.9852	0.1187	0.8812	0.9956	0.0043	0.3353	0.6646	0.0376	0.9623	0.0363	0.9636	0.2043	0.7956
21	Fused_Modified_Period2_Img950-951_Leftforearm.bmp	0.9940	0.0059	0.9992	0.0007	0.0769	0.9230	0.2118	0.7881	0.0005	0.9994	0.7685	0.2314	0.0727	0.9272	0.0010	0.9989	0.8966	0.1033

Appendix D – Classification Phase Post-Test Results for Period -3

		Dataset-1						Dataset-2						Dataset-3					
		200 / 200						400 / 400						720 / 720 (includes noise/rot5/rot10/scale)					
		50 / 50						100 / 100						180 / 180 (includes noise/rot5/rot10/scale)					
Train (Right Forearm/Left Forearm):																			
Pre-test (Right Forearm/Left Forearm):																			
CNN Model:		Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3	Model-1	Model-2	Model-3						
Input Image Size:		28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128	28x28	64x64	128x128						
Learning Rate:		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001						
Epoch:		14000	4000	1000	10204	3459	800	1000	1000	400	1000	1000	400						
Training Duration (sec/1 epoch):		1.458123 sec	3.062919 sec	38.534514 sec	1.955146 sec	4.664262 sec	75.446255 sec	3.355219 sec	18.929039 sec	147.922873 sec	3.355219 sec	18.929039 sec	147.922873 sec						
		Predicted Class						Predicted Class						Predicted Class					
The NIR Fused-Modified Image Name		Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm	Right Forearm	Left Forearm
1	Fused_Modified_Period3_Img0-1_Leftforearm.bmp	0.8934	0.1065	0.9972	0.0027	0.0000	0.9999	0.0082	0.9917	0.0000	0.9999	0.0007	0.9992	0.0120	0.9879	0.0020	0.9979	0.0007	0.9992
2	Fused_Modified_Period3_Img50-51_Leftforearm.bmp	0.7992	0.2007	0.9904	0.0095	0.0002	0.9997	0.0004	0.9995	0.0000	0.9999	0.0042	0.9957	0.0002	0.9997	0.0057	0.9942	0.0002	0.9997
3	Fused_Modified_Period3_Img100-101_Leftforearm.bmp	0.9661	0.0338	0.9988	0.0011	0.0000	0.9999	0.0014	0.9985	0.0000	0.9999	0.0010	0.9989	0.0013	0.9986	0.0018	0.9981	0.0002	0.9997
4	Fused_Modified_Period3_Img150-151_Leftforearm.bmp	0.9679	0.0320	0.9961	0.0038	0.0002	0.9997	0.0094	0.9905	0.0000	0.9999	0.0044	0.9955	0.0082	0.9917	0.0018	0.9981	0.0011	0.9988
5	Fused_Modified_Period3_Img200-201_Leftforearm.bmp	0.9531	0.0468	0.9966	0.0033	0.0000	0.9999	0.0031	0.9968	0.0000	0.9999	0.0005	0.9994	0.0018	0.9981	0.0018	0.9981	0.0003	0.9996
6	Fused_Modified_Period3_Img250-251_Leftforearm.bmp	0.9970	0.0029	0.9998	0.0001	0.0000	0.9999	0.0159	0.9840	0.0000	0.9999	0.0009	0.9990	0.0133	0.9866	0.0016	0.9983	0.0002	0.9997
7	Fused_Modified_Period3_Img300-301_Leftforearm.bmp	0.9905	0.0094	0.9992	0.0007	0.0000	0.9999	0.0007	0.9992	0.0000	0.9999	0.0020	0.9979	0.0002	0.9997	0.0001	0.9998	0.0013	0.9986
8	Fused_Modified_Period3_Img350-351_Leftforearm.bmp	0.9959	0.0040	0.9958	0.0041	0.0000	0.9999	0.0057	0.9942	0.0000	0.9999	0.0006	0.9993	0.0015	0.9984	0.0014	0.9985	0.0002	0.9997
9	Fused_Modified_Period3_Img400-401_Leftforearm.bmp	0.9939	0.0060	0.9999	0.0000	0.0001	0.9998	0.0325	0.9674	0.0000	1.0000	0.0057	0.9942	0.0153	0.9846	0.0000	0.9999	0.0007	0.9992
10	Fused_Modified_Period3_Img450-451_Leftforearm.bmp	0.9974	0.0025	0.9999	0.0000	0.0000	0.9999	0.0202	0.9797	0.0000	0.9999	0.0009	0.9990	0.0109	0.9890	0.0000	0.9999	0.0024	0.9975
11	Fused_Modified_Period3_Img500-501_Leftforearm.bmp	0.9993	0.0006	0.9999	0.0000	0.0000	0.9999	0.0030	0.9969	0.0000	0.9999	0.0019	0.9980	0.0015	0.9984	0.0000	0.9999	0.0010	0.9989
12	Fused_Modified_Period3_Img550-551_Leftforearm.bmp	0.9942	0.0057	0.9999	0.0000	0.0000	0.9999	0.0014	0.9985	1.0000	0.0000	0.0058	0.9941	0.0001	0.9998	0.0000	0.9999	0.0153	0.9846
13	Fused_Modified_Period3_Img600-601_Leftforearm.bmp	0.9994	0.0005	0.9999	0.0000	0.0020	0.9979	0.0375	0.9624	0.0000	1.0000	0.0284	0.9715	0.0113	0.9886	0.0000	0.9999	0.0016	0.9983
14	Fused_Modified_Period3_Img650-651_Leftforearm.bmp	0.9996	0.0003	0.9996	0.0003	0.0032	0.9967	0.8008	0.1991	0.0000	0.9999	0.0611	0.9388	0.4043	0.5956	0.0001	0.9998	0.0120	0.9879
15	Fused_Modified_Period3_Img700-701_Leftforearm.bmp	0.9998	0.0001	0.9999	0.0000	0.0011	0.9988	0.3531	0.6468	0.0000	0.9999	0.0289	0.9710	0.2151	0.7848	0.0000	0.9999	0.0066	0.9933
16	Fused_Modified_Period3_Img750-751_Leftforearm.bmp	0.9935	0.0064	0.9999	0.0000	0.0015	0.9984	0.0215	0.9784	0.0000	0.9999	0.0568	0.9431	0.0056	0.9943	0.0000	0.9999	0.0447	0.9552
17	Fused_Modified_Period3_Img800-801_Leftforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.0238	0.9761	0.1661	0.8338	0.0000	0.9999	0.3389	0.6610	0.0804	0.9195	0.0000	0.9999	0.0703	0.9296
18	Fused_Modified_Period3_Img850-851_Leftforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.0052	0.9947	0.0354	0.9645	0.0000	0.9999	0.1487	0.8512	0.0168	0.9831	0.0000	0.9999	0.0740	0.9259
19	Fused_Modified_Period3_Img900-901_Leftforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.0078	0.9921	0.0323	0.9676	0.0000	0.9999	0.1779	0.8220	0.0497	0.9502	0.0000	0.9999	0.1995	0.8004
20	Fused_Modified_Period3_Img930-931_Leftforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.0105	0.9894	0.0201	0.9798	0.0000	0.9999	0.3078	0.6921	0.0124	0.9875	0.0000	0.9999	0.3220	0.6779
21	Fused_Modified_Period3_Img950-951_Leftforearm.bmp	0.9999	0.0000	0.9999	0.0000	0.0058	0.9941	0.0710	0.9289	0.0000	0.9999	0.1742	0.8257	0.1294	0.8705	0.0000	0.9999	0.3343	0.6656

Appendix E – Quick Reference of the Literature

No	Name	Aim	Method(S)
1	Veindeep: Smartphone unlock using vein patterns.	Protecting smart phones using vascular patterns from opportunistic access	Veindeep Application Using Smart Phone Infrared Depth Sensors.
2	Contrast enhancement for the infrared vein image of leg based on the optical angular spectrum theory.	Extracting the effect of the hair on the NIR image.	NIR Light, CCD Camera-Filtering Algorithm Based On Optical Angular Spectral Theory.
3	Multisensory System for the Detection and Localization of Peripheral Subcutaneous Veins.	Developing a guidance system for automatic robotic insertion of catheters.	NIR Light Source, ID Camera SWIR, 3D Camera TOF.
4	Development of portable, wireless and smartphone controllable near-infrared spectroscopy system.	Measuring oxyhaemoglobin, deoxyhaemoglobin, and total hemoglobin concentration in tissues.	Portable NIR Spectroscopy System.
5	Finger vein biometric: Smartphone footprint prototype with vein map extraction using computational imaging techniques.	Developing a system that uses the vascular structure map for biometric approach.	NIR Camera, Finger's Region-2 Images.
6	Instrumenting smartphones with portable NIRS.	Review (Portable NIRS)	Review (Portable Nirs)
7	Innovative algorithm to evaluate the capabilities of visual, near infrared, and infrared technologies for the detection of veins for intravenous cannulation.	Evaluating the suitability and capability for intravenous cannulation in vessel identification.	Visual, Infrared And Near-Infrared Cameras.
8	Deep tissue near-infrared imaging for vascular network analysis.	Developing a system for deep tissue NIR Imaging for vascular network analysis.	Liquid And Solid Phantoms.
9	Augmented reality based real-time subcutaneous vein imaging system.	Reducing the failure rate and time required for intravenous injection by performing back-project superimposed veins (as augmented vein structures) on the skin surface.	Augmented Reality
10	Systematic review of clinical applications of monitoring muscle tissue oxygenation with near-infrared spectroscopy in vascular disease.	Monitoring muscle tissue oxygenation for vascular disease.	Nir Spectroscopy
11	Vein visualization: Patient characteristic factors and efficacy of a new infrared vein finder technology.	Investigating relation between finding eligible vessels for cannulation and patient characteristics.	Infrared Vein Finding Technology (IR VF Device-Vuetek Scientific Veinsite).
12	Efficacy of a near-infrared light device in pediatric intravenous cannulation: A randomized controlled trial.	Investigating the improvement of success rate of the placement of intravenous catheters in the first-attempt.	NIR Light Venipuncture.
13	Use of the accuvein device to map the superficial venous system.	Investigating vein graft harvest and remove superficial vein maps to aid in other plastic surgery procedures.	Accuvein AV300
14	A convolutional neural network model for marble quality classification	Quality classification according to marble images.	Convolutional Neural Networks, Data Augmentation, Machine Learning.
15	Finger Vein Recognition Based on Deep Learning	A deep learning model that can be used in finger vein recognition.	Biometrics, Deep Learning, Convolutional Neural Network.

Appendix E – Continues

16	Detection of Inferior Myocardial Infarction using Shallow Convolutional Neural Networks	A CNN architecture that uses electrocardiography signals for detection of inferior myocardial infarction.	Convolutional Neural Network, Electrocardiography.
17	Dorsal Hand Vein Recognition by Convolutional Neural Networks: Feature Learning and Transfer Learning Approaches	Developing a hand vein recognition system using CNN.	Hand Vein, Patterns Recognition, CNN, SVM, K-NN.
18	Optic Disc Determination in Retinal Images with Deep Features	A hybrid approach that will assist in the diagnosis and treatment of retinal diseases in the classification of regions with/without optic disc.	Optic Disk Detection, Deep Features, K-NN Classifier.
19	Determination of Vascular Stenosis On Angiography Images Using Convolutional Neural Network Method	Analysis of images to be obtained by computer aided detection system using convolutional neural networks and detection of stenosis.	Convolutional Neural Network, Coronary Artery Disease, Stenosis.
20	Detection of an Assembly Part with Deep Learning and Image Processing	To determine the number of parts by recognizing the parts of the assembly accessories from the camera images.	Deep Learning, Opencv, YOLO Algorithm.
21	Infusion Port Level Detection for Intravenous Infusion Based On Yolo V3 Neural Network	To increase the detection accuracy of the liquid level in the infusion bottle used in the intravenous infusion application.	Deep Learning; Liquid Level Detection; YOLOv3.
22	Augmented Reality Storytelling: Narrative Design and Reconstruction of a Historical Event in situ	Location-based indirect AR environment to enrich the storytelling experience of the Omaha Beach landing (D-Day).	Mobile Augmented Reality, Mixed Reality, Situated Simulations, Sitsim, Storytelling.
23	Mobile ar using pre-captured omnidirectional images	Solve the alignment problem between virtual and real materials rendered in real-time on mobile ar images.	Augmented Reality, Mobile System, Omnidirectional Image.
24	Indirect augmented reality considering Real-world illumination change	Elimination of temporal inconsistencies that may occur due to lighting during online registration	Indirect Augmented Reality, Illumination Change.
25	Comparing Augmented Reality Visualization Methods for Assembly Procedures	The effectiveness of three different AR-based visualization methods in the assembly process.	Indirect Augmented Reality, Head-Mounted Display, Computer-Aided Manufacturing.

Appendix F – Publications Published within the Scope of the Thesis

- Erdem, H. A., & Utku, S. (2022b). Augmented reality aided pre-diagnosis environment for telemedicine: Superficial vein surveillance system [**Article**]. *European Journal of Science and Technology*, 38, 376-385.
<https://doi.org/10.31590/ejosat>
- Erdem, H. A., & Utku, S. (2022a, July 22-24). *Detecting Venous Disorders via Near-Infrared Imaging: Observation of Varicose Vein Development* [**Conference presentation/Short paper**]. In Proceedings of the IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications (IARIA Congresses), pp. 65-68, IARIA. Nice, France. Retrieved September 02, 2022, from https://www.thinkmind.org/index.php?view=article&articleid=iaria_congress_2022_1_130_50088
- Erdem, H. A., Erdem, I., Uzun, B. & Utku, S. (2021, March 26-27). General-purpose contactless biometric authentication systems: Vein recognition with convolutional neural networks [**Conference presentation/Abstract**]. In Proceedings of the Xth International Biomechanics Congress (BIOMECH'20), p. 38, Red and More Events. İzmir, Turkey.
- Erdem, H. A., Erdem, I., & Utku, S. (2020, November 21-25). *Near-infrared mobile imaging systems for e-health: Lighting the veins* [**Conference presentation/Short paper**]. In Proceedings of the Twelfth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED), pp. 80-84, IARIA. Valencia, Spain. Retrieved September 02, 2022, from https://www.thinkmind.org/index.php?view=article&articleid=etelemed_2020_3_90_40039

Appendix G – Glossary

* Retrieved September 30, 2022, from <https://en.oxforddictionaries.com/>, <http://dictionary.cambridge.org/>, and www.britannica.com websites.

A

Absorption: The process of taking something into another substance.

Accelerometer: A device that measures acceleration or G-force.

Acquisition: The process of getting something.

Authentication: (computing) The process or action of verifying the identity of a user or process.

B

Biometric: Relating to or involving the application of statistical analysis to biological data.

Biosensor: A device which uses a living organism or biological molecules, especially enzymes or antibodies, to detect the presence of chemicals.

Built-in: Forming an integral part of a structure.

C

Cerebral cortex: The outer layer of the cerebrum (the cerebral cortex), composed of folded grey matter and playing an important role in consciousness.

Cortical: Relating to the outer layer of the cerebrum.

Covid-19: An infectious disease caused by a coronavirus (= a type of virus), that usually causes fever, tiredness, a cough, and changes to the senses of smell and taste, and can lead to breathing problem and severe illness in some people

D

Degenerate: To become worse.

Dorsum (hand): A Latin word meaning "back", used in medical names and descriptions of the back or top surface of an organ.

F

fNIRS: A vascular-based neuroimaging method that measures non-invasively the changes in oxygenated (O₂Hb) and deoxygenated (HHb) haemoglobins by

multiple detectors placed on the head skin.

Fusion: An occasion when two or more things join or are combined.

G

Gamma Ray: Penetrating electromagnetic radiation of a kind arising from the radioactive decay of atomic nuclei.

GPS: Global Positioning System, an accurate worldwide navigation and surveying facility based on the reception of signals from an array of orbiting satellites.

Gyroscope: A device containing a wheel that spins freely within a frame, used on aircraft and ships to help keep them horizontal, and as a children's toy.

H

Haemodynamic: Relating to the forces involved in blood circulation.

Haemoglobin: A red protein responsible for transporting oxygen in the blood of vertebrates. Its molecule comprises four subunits, each containing an iron atom bound to a haem group.

Hyperscanning method: Measures simultaneous brain activities between two or more people.

I

Infrared: (of electromagnetic radiation) Having a wavelength just greater than that of the red end of the visible light spectrum but less than that of microwaves. Infrared radiation has a wavelength from about 800 nm to 1 mm, and is emitted particularly by heated objects.

Intravenous cannulation (intravenous Access): Administrative medication is the process of inserting cannula into vein for fluids or to take blood samples.

N

Neuroimaging: The process of producing images of the structure or activity of the brain or other part of the nervous system by techniques such as magnetic

resonance imaging or computerized tomography.

Neurovascular: ⁽¹⁾ Containing neural and vascular structures. ⁽²⁾ Of or relating to the nervous system and the vascular system or their interactions.

Near-Infrared (NIR): Electromagnetic spectrum's IR wavelengths of 700 to 900 nm are known as the near-infrared (NIR) window.

Non-invasive: (of medical procedures) not involving the introduction of instruments into the body.

S

Shoulder-surfing: The practice of spying on the user of a cash-dispensing machine or other electronic device in order to obtain their personal identification number, password, etc.

Spectroscopy: The study of the response of the substance to the light stimulus when the light energy hits the substance and the acquisition of information about the molecules forming the substance.

Spectrum (spectra/plural): The set of colours into which a beam of light can be separated, or a range of waves, such as light waves or radio waves.

O

On-board: Used to refer to things that are carried by a vehicle and form part of it.

Oxygenation: The process of adding oxygen to something.

Oxyhaemoglobin: A bright red substance formed by the combination of

haemoglobin with oxygen, present in oxygenated blood.

T

Tomography: A technique for displaying a representation of a cross section through a human body or other solid object using X-rays or ultrasound.

Topographical: Relating to or representing the physical distribution of parts or features on the surface of or within an organ or organism.

U

Ultrasound: Sound or other vibrations having an ultrasonic frequency, particularly as used in medical imaging.

Universal serial bus: Technology used to connect computers with peripheral devices.

V

Venipuncture: Medical progress of obtaining intravenous access.

W

Wavelength: The distance between successive crests of a wave, especially points in a sound wave or electromagnetic wave.

X

X-ray: An electromagnetic wave of high energy and very short wavelength, which is able to pass through many materials opaque to light.