Istanbul Medipol University

School of Engineering and Natural Sciences

Graduation Project

2022-2023

PROJECT TITLE
Hand Gesture Based Device Control with Applications in the Medical Domain
PROJECT ADVISOR
Prof. Dr. Reda Alhajj
TEAM MEMBERS
Osman Mersin İlknur Akçay Yunus Emre Gündüz

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

Project Code
Project Title: Hand Gesture Based Device Control with Applications in the Medical Domain
Project Advisor: Prof. Dr. Reda Alhajj
Project Team Members: Osman Mersin İlknur Akçay Yunus Emre Gündüz
Sponsor Company (if any) :

BUDGET (TL)	PROPOSED	CONSENTED
IMU FUNDING	0	0
SPONSOR COMPANY FUNDING	0	0
TOTAL	0	0

PROJECT PLAN	PROPOSED	CONSENTED
PROJECT PLAN Duration in Weeks	54 Weeks	28 Weeks
STARTING DATE	01.10.2022	30.06.2023

Istanbul Medipol University

School of Engineering and Natural Sciences

Graduation Project

Project Code

PROJECT ADVISOR	DEPARTMENT CHAIR
Name: Prof. Dr. Reda Alhajj	**Name:** Assoc. Prof. Mehmet Kemal Özdemir
Contact Information: Tel : E-mail : ralhajj@medipol.edu.tr	**Contact Information:** Tel : E-mail : mkozdemir@medipol.edu.tr
Signature:	**Signature:**

TEAM MEMBER	TEAM MEMBER
Name: Osman Mersin	**Name:** İlknur Akçay
Contact Information: Tel : +90 (554) 596 61 94 E-mail : osman.mersin@std.medipol.edu.tr	**Contact Information:** Tel : +90 (534) 916 09 07 E-mail : ilknur.akcay@std.medipol.edu.tr
Signature:	**Signature:**

TEAM MEMBER	Sponsor Company
Name: Yunus Emre Gündüz	**Name:**
Contact Information: Tel : +90 (545) 940 79 72 E-mail : yunus.gunduz@std.medipol.edu.tr	**Contact Information:** Tel : E-mail :
Signature:	**Signature:**

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

Project Title:

Hand Gesture Based Device Control with Applications in the Medical Domain

Project Advisor: Prof. Dr. Reda Alhajj

Team Members: Osman Mersin, Yunus Emre Gündüz, İlknur Akçay

Project Group Title: MedHands

PROJECT OVERVIEW/SUMMARY/ABSTRACT

Computers have become indispensable tools in the medical field thanks to the advancement in technology. So much so that, these devices have even entered the operating rooms. Currently, especially medical imaging systems are being used in surgeries to facilitate the operations. However, in critical situations like surgeries, it is difficult to control these systems. The aim of this project is basically solving this problem with a hand recognition system that uses webcam to control the device.

According to the literature, hand recognitions systems mostly used to control home appliances for challenged and old people. On the other hand, there is not many applications that uses hand gestures to control healthcare systems. So, considering the requirements of the usage areas, this project will come up with a new approach by creating a software based sterile human-machine interface in real-time. For this project, since the datasets in the literature are not thought to be fully efficient, its own dataset will be created. In addition, due to security issues and to determine which hands are dominant, it is planned to implement an authorization part to the project.

To achieve the aim of the project, firstly the literature reviewed to understand the similar projects. After that, standard hand gestures and their functions that is being used in other projects determined. According to the obtained results, 15 hand gestures chosen which means the first part of the project completed. To create a video stream dataset, 50 different people were asked to perform the selected hand gestures. In this part, it is planned for all the chosen hand gestures but, it was only accomplished for 5 hand gestures. These 5 hand gestures selected specifically to both check finger gestures and hand motion. It is also planned to use gloves to add the created video stream dataset since the project will be used in medical area. This step considered fulfilled with a small error percentage since the dataset contains 50 people with 5 different hand gestures. In the next step, for the created video stream dataset, a pre-processing algorithm was developed. This algorithm (uses MediaPipe library) was basically taking the videos and converting the hand gesture features into a numerical table. By using the coordinates of some keypoints and distance ratios between these keypoints, a numerical dataset prepared. To complete this step successfully, first the keypoints, then the dataset was checked with python code (checking keypoints via video for keypoints - accessing, storing data and retrieving data for dataset). After creating this, an ensemble learning model were created by combining machine learning algorithms for hand gesture classification.

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

After the hand movements were successfully classified, authentication was provided by comparing the information obtained from real time and the information in the dynamic dataset. As a result of the proposed algorithms, device management was provided by hand gestures.

After the hand movements were successfully classified and the SVM algorithm was created by normalizing the structural information obtained in the authentication part, authentication was provided by comparing the information obtained from real time and pretrained model. Then, at the end of testing process, it achieved to authorize correct users with higher than 90% accuracy. As a result of the proposed algorithms, device management was provided by hand gestures. In the end, prepared two parts (authentication & hand gesture recognition) were effectively integrated and adjusted to enable the system to generate appropriate responses based on recognized hand gestures. This integration enhances the system's responsiveness and usability, allowing users to interact with the system intuitively through their hand gestures.

In conclusion, the project demonstrated a satisfactory success rate of 95% according to the predefined criteria, confirming its effectiveness and reliability. This outcome also serves as the validation of the system's capability to accurately identify hand gestures.

Keywords: Hand Gesture Recognition, Gesture Recognition, Human-Computer Interaction, Healthcare Systems, Authentication, Standardization, Machine Learning

1. OBJECTIVE OF THE PROJECT:

With the advancement of technology, the field of health has gained a whole new momentum. In the medical field where people's lives are at stake, especially the medical imaging is very important for diagnosis and treatments. Currently, medical imaging systems are being used in surgeries. For example, MRI scan result of a patient can be displayed in a brain surgery to facilitate the operation. However, it is quite challenging to control the devices especially in critical conditions. With this motivation, the main objective of the project is developing a system that helps to control hospital devices with hand gestures. Because with this way, interventions at critical moments will be more effective. Also, 1) standardization of the meanings of hand gestures to create a common language in this field, 2) involving doctors to control devices which makes it more efficient for the doctor, 3) gaining authentication to control the devices in hospital environments can be listed as the sub-goals of this project. To sum up, with this project it is expected to obtain a hand gesture recognition system which will be used in the healthcare systems.

2. LITERATURE REVIEW:

2.1. Human Computer Interaction Technology

In recent days, the interaction of human beings with computing systems has accelerated thanks to dramatic increases in computer science and technology. In this sense, the researchers have started to focus mostly on the interfaces which support sufficient human-computer interaction (HCI) which is a multidisciplinary field of computer technology (Figure 1) (May, 2001). The main concern of HCI is design, assessment, and application of interactive computing systems for human use. In the beginning of computer technology, people started to manage computers by using a single mouse and keyboard (Gür vd., 2020). However, with the developments in the communication systems, hardware, and software technologies such as computer graphics, operating systems and programming languages, there has been occurred a need to discover more usable, reliable, safe and functional ways for fast communication and the most effective interaction with computer-like devices such as mobile phones, tablets, medical devices in hospitals and so on (Gulliksen, 2017). Because of the necessity, the researchers have sought innovative methods to expedite the interaction between human beings and computers and get instant real-time reactions from the computer-based devices (Paton vd., 2021). Because the more natural and lifelike interfaces for virtual environments have gained great importance. Therefore, the several approaches have been developed to serve those aims with the help of new technologies in the computer vision (CV), machine learning (ML), deep learning (DL), chip and sensor technologies, data science (data management and data analysis), and further on (Johnson vd., 2015). For instance, time in the robotic system for surgery is so crucial for surgical actions. Thereby, active, multi-mode and intelligent adaptive interfaces become remarkable for the reducing time of the teleoperation in the surgeries (Pelikan vd., 2018; Vendruscolo ve Martelli, 2001).

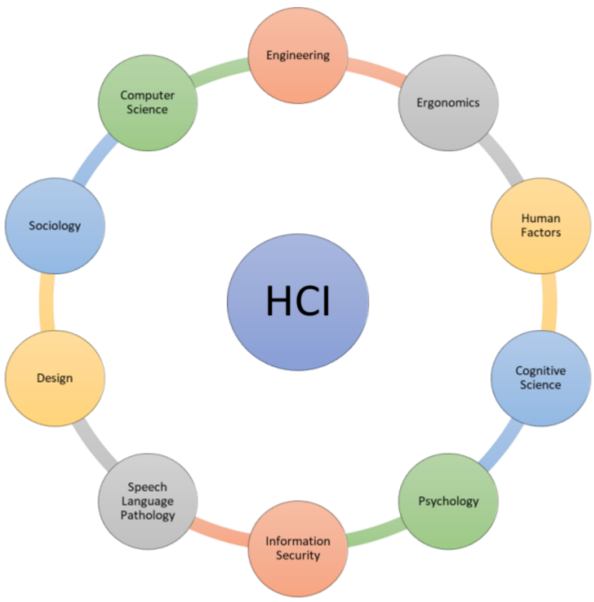


Figure 1: Research fields which usually use the human-computer interaction.

2.2. Human-Computer Interaction in Healthcare Systems

The user and machine interaction has great importance to control the medical devices in health system. Thus, one of the fields in which HCI is most widely used and tried to be implemented is healthcare systems. Because the HCI does not only contribute to the correct data management, image manipulation, it has also potential to figure out the sterility, limited space, and time pressure problems for the physicians. However, the current HCI tasks have some challenges in the operation routine since the joysticks, foot pedals, touchscreens and control panels are generally used to directly interact with the medical machines (Hübler vd., 2014). In this case, direct interaction means that the carrying several pathogens because of the difficulty in the effective sterilization of computers and their peripherals (Schultz vd., 2003). Although the machines and health care workers' hands are tried to be sterilized with anti-pathogenic gels and cleaned with several antiseptics, the surface of a device and skin cannot be completely sterilized (Changruenngam vd., 2022). Because of such problems in the healthcare environment, a need for a touchless HCI system has occurred (Blandford, 2019) and scientific research on the development of touchless HCI systems in medical domain has gained speed with the increment of improvements in the digital health technologies. According to the Cronin and Doherty's research, the benefits of touchless HCI systems are emphasized in terms of sterilization, three-dimensional (3D) data manipulation in 3D space, making easier the hands-busy interactions and speeding up the surgical tasks by removing the barriers (Cronin ve Doherty, 2018). The touchless HCI architecture is generally classified as unimodal and multimodal (Figure 2) (Jaimes ve Sebe, 2007). The unimodal HCI systems consist of audio-based, sensor-based, vision-based (video-based and image-based) while the multimodal HCI systems combination of the unimodal attentions (Sarma ve Bhuyan, 2021). The vision-based models are generally attributed to the body movement tracking and gesture identification (Mitra ve Acharya, 2007). The speech recognition and auditory emotion identification are in the category of audio-based models (Li vd., 2014).

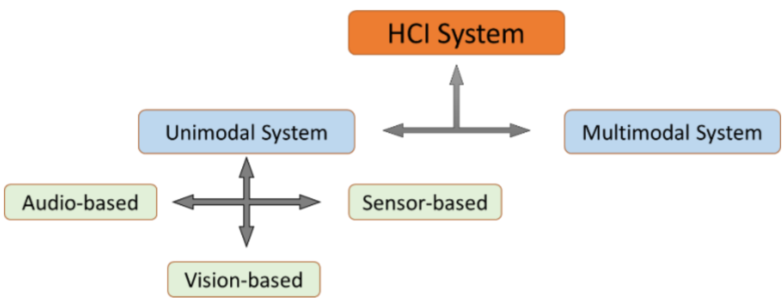


Figure 2: General scheme for the HCI architecture

2.3. Gesture Recognition-based on Human-Computer Interaction Systems

Body motion analysis and the gesture identification are grown rapidly by the increment in the interest of HCI and audio-based applications which makes the interfaces more natural. Besides the other fields, the healthcare systems also devote to have non-verbal communications. In this case, the gesture recognition enables clinicians to input commands using the movements of parts of body such has head (Tu vd., 2015), hand (Hasan ve Kareem, 2012) or whole-body posture (Alsaffar vd., 2021). As a result of many research carried out to find a best method for the gesture recognition from part of body method, the hand gesture identification have been pointed out as a more practical class among the gesture recognition classes due to its higher flexibility and being user-friendly (Maricchiolo vd., 2014). The figure 3 presents the gesture classifications in the literature and specifies the hand gesture subclasses such as static, dynamic and hybrid systems including both. The static hand gestures are the processing of a single hand image, and its computational cost is lower. Dynamic hand gestures are the trajectory-based gestures, which the image sequences are processed and requires more complex identification approaches (Pisharady ve Saerbeck, 2015). The static gestures refer the stable hand gestures while the dynamic gestures are the comparison of a sequence of hand motions.

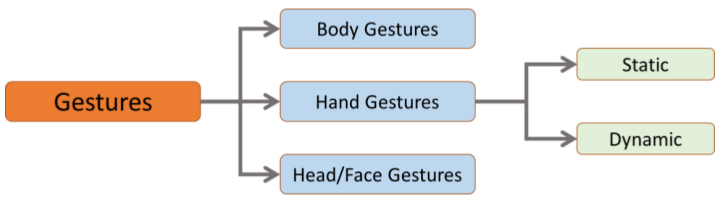


Figure 3: Gesture recognition scheme

2.4. Hand Gesture Recognition Methods

The hand recognition is categorized into the sensor approach and computer vision approach. In the beginning of the hand gesture recognition, the hand movements were succeeded to be recognized with the help of wearable sensors which can easily detect the hand movements directly or the finger bending, and the physical data is collected in the computer by using wire communication. Because of the microcontroller attached in the sensors, that method was accepted as the more portable method for the hand gesture recognition. In this technology, the flex, tactil (depended on the finger bending), fiber optic transducer-based, accelerometer and gyroscope sensors (based on the rate of rotation on an axis) are generally used. Since the wearable sensors are not cost-effective and discomfort because of hardware connection, cause the damage on the skin and infections (Lamberti ve Camastra, 2011), the vision-based technologies by using RGB, depth, time of flight (TOF), thermal cameras (Figure 4) have gained an importance for the hand gesture recognition in the health care systems, which are provides the touchless interaction between computers and people (Wachs vd., 2011). For the vision-based hand gesture recognition, the skin color analysis, appearance and skeleton of the hands, movements, depth and detections based on deep learning algorithms are the common algorithms for the detection and classification of the hand gestures.

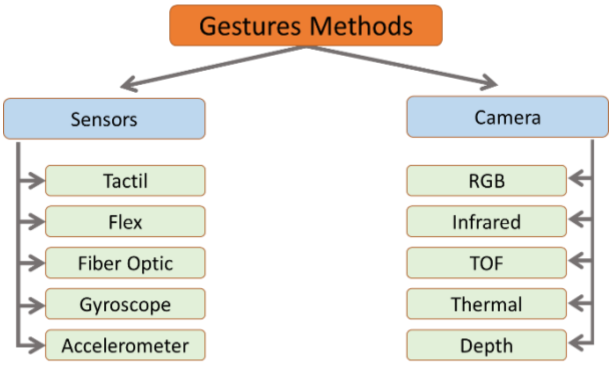


Figure 4: Different methods used in the hand gesture recognition

The hand gesture recognition based on the color of the captured images are classified into two categories: glove marker (Wang ve Popović, 2009) or skin color. The color-based hand gesture recognition is the recognition method which requires wearing a specifically colored a glove. According to the colors on the gloves, the algorithm based on the camera decides the geometric shape of the hand and classifies the hand gesture. The skin color is detected according to pixel distribution to detect the hand, intensity, and texture. Although the skin color method is divided into red-green-blue (RGB), hue/saturation (HSV) and luminance (Y-Cb-Cr) color spaces, the luminance is good for the recognition process since the complex color channels in RGB, insufficiency of HSC space for the multiple images. Therefore, the luminance space is more promising method for the hand gesture recognition based on skin color determination (Chai ve Bouzerdoum, 2000). Another machine learning algorithm that performs gesture recognition based on RGB values is MediaPipe. It provides segmentation and detection by using the RGB values of the image taken from the camera. MediaPipe hand framework takes visual or video as input and processes it in an algorithm consisting of 2 separate main structures. First, it detects the palm with BlazePalm Detector. After doing this, it finds 21 handmarks on hand thanks to regression. This algorithm, which reaches approximately 95% accuracy, can detect handmarks even on different backgrounds and different shapes of the hand thanks to its large training dataset (Zhang vd., 2020). Another method is based on the appearance of the hands, which is called as appearance-based hand gesture recognition. In the method, the two-dimensional (2D) image features of captured images are extracted in real-time, and the appearance of the hand are compared with the new parameters and the parameters stored in the database. For the static models, AdaBoost learning algorithm (Chen vd., 2007), OTSU and canny edge detection methods are the common methods. For the hand gesture recognition based on hands motion, the detected object is extracted as a sequence of images. The movement is characterized and modelled according to the image features. After the pattern of the motion is detected, the gesture is expected to be recognized. For the skeleton- based hand gesture recognition, the detection of the hand is attributed to the detection of the joints on the hand. The important criterion for the recognition is to determine the geometric attributes and the statistic features properly.

Orientations of the hand joints, distance between the joints, joint locations, rate of angle between the axis of joints and their trajectories are examples of the features which are evaluated in the skeleton-based gesture recognition. The depth-based hand gesture recognition method bases on the 3D geometric feature generation from the input taken from the depth camera. With comparison with the projection of color image, 3D image data points out the depth field of the recognized hand. Even though the lightness, shade and color variability in environment does not affect the recognition process, the depth camera is not cost-effective, and it has limited availability and size. The hands can also be detected by their posture estimation. According to the 3D kinematics of the 3D model hand gesture recognition strategy, the combination of volumetric and skeletal parameters presents the 3D model of the segmented hand. Basically, the features are extracted from the query image of the hand and the image descriptors are obtained. Then, with the matching process, the quantized model descriptors are generated and is associated to the 3D model. This method is processed on the large datasets, the complex mathematical formula for the shape characterization and might cause the time consuming for the matching process. The last one among the common recognition methods, the deep learning use in the hand gesture recognition. The integration of the artificial intelligence provides a better and more reliable predictions. The dataset quality is crucial for the deep learning process. Convolutional neural network (CNN), implementation of VGG Net architecture to the deep learning algorithm, correlation filters like kernel correlation filters, support vector machine (SVM) are some examples of the algorithms used in the hand gesture recognition using the deep learning (Oudah vd., 2020).

Against the single classification algorithms used in describing hand movements, ensemble modeling can also be used in which these algorithms are combined. Ensemble model can be formed by combining different algorithms such as Decision Trees, k-Nearst Neighbors, Multinomial Naïve Bayes (Pham vd., 2021). After using different algorithms in the classical ensemble model, the most accurate result is selected by voting. This type of ensemble model consists of different methods. In a study using the bagging and bootstrap technique, it was determined that the bagging technique alone gave better results (Nai-Arun ve Sittidech, 2014). In this way, a comparison is made between the outputs of different models and the result is decided. With the bootstrap technique, weak classifier models are tried to be strengthened by retraining them. In the hybrid ensemble model, the outputs feed each other unidirectionally. While the features of strong algorithms can be used, they can be overcome by identifying the weaknesses of other algorithms (Verma ve Hassan, 2011). In general, it has been observed that the ensemble model technique gives much more advanced results when compared to single machine learning algorithms.

In the recent technology, the static hand gesture recognition is accepted as the solved problem. However, this statement is not valid for the dynamic hand gesture recognition methods because of the several major problems which are still waiting to be answered such as spatial-temporal variation, the complex structure of the human hand which makes the detection difficult, and difficulties in the computer vision. The continuous segmentation of the hand gesture has the challenges in its segmentation because of the variations in the illumination, background complexity and occlusion. The illuminations exactly depend on the lightness in the environment. The change in the light intensity on skin affects the chrominance property of the skin color space and then, it causes the wrong detection of the skin color. Although the researchers have tried to figure out this problem, the developed methods are sufficient to tolerate the small changes in the illumination (Habili vd., 2004). Furthermore, the complexity in the backgrounds of the captured streams is the other problem faced in the gesture recognitions. Because of randomly variation of the texture and color, the complexity of the cluttered backgrounds is increased. Also, the background properties (color and texture) of the images in the streams can be dynamically changed in time. This requires the complex computational steps to detect the correct color space for the detection and precision values. Although the focus is on the skin color recognition from a stream, the background might have the skin-colored regions (Sigal vd., 2004). In that case, the hand gesture recognition is complicated. Moreover, in fact that two handed gesture recognition, one hand might not be recognized accurately because of the occlusion problem which are mainly encountered in the hand gesture recognition systems. In order to handle with stated problem, three approaches are presented: using multiple cameras, tracking based gesture recognition and combination of both. Although the first two methods are the efficient for the hand gesture recognition, the combination of these methods might have still some problems because of the quality of cameras. The researchers recommend the use of depth sensor to handle with such problems (Utsumi vd., 2002). The tracking of two hands at the same time has the still unanswered questions. Especially, if the hands have interacted each other, the change in the surface of hands because of the overlapping, the recognition of both hands might not be executed correctly (Oikonomidis vd., 2012). Size and shape of the hands causes difficulties in the detection and segmentation. Therefore, the recognition systems should not be variant to the shape and size.

As mentioned above, several methods might be used for the hand gesture recognition. However, the hand gestures can be varied from region to region and from application to application. For example, when the sign languages which have the potential to differ from location to location and person to person are considered, it is so clear to acquire that the American, British, Arab, and Turkish sign languages are all different. Similarly, the hand gestures differ from nation to nation. This type of situation can cause the lack of identification, make difficult the hand gesture recognition procedures and then, cause some crucial problems in the healthcare systems. According to the gestures referred in (Devineau vd., 2018; Kulkarni, 2010; Molina vd., 2017), although there are few common gestures, the major of them do not have the similarity and specialized for their use. Therefore, in the hand gesture recognition systems, the gestures should be standardized to acquire a worldwide application and use. Gestures and their purpose can be clearly understood. For example, in many studies, the opening of the hand represents the opening of the system, while the gesture of making a fist represents the closing of the system (Gonzalo ve Juan, 2015).

2.5. Authentication in Hand Gesture Recognition Systems

Authentication creates security by ensuring that the right user uses the system. By comparing a user's information to those stored in a database, authentication technology controls access to systems. Before today, these systems were only provided with people carrying and showing their own identity cards. These situations cause problems such as getting the ID card into someone else's hands or forgetting the material necessary for the person to be recognized by the system. Today, many companies have started to use biometric recognition systems to reduce vulnerability. Biometrics examines physiological or behavioral characteristics that are unique in humans. Biometric authentication-based criteria such as eye, face, fingerprint recognition exhibit high rates of verification (Sabhanayagam vd., 2018). The verification methods to be used in the projects are determined according to the requirements of the system. False Acceptance Rate lowest (more successful). The leading methods are fingerprint and iris recognition methods. However, in a situation where environmental conditions are bad, these two methods may not give good results. For this reason, the system requirements should be determined specifically and the method to be used should be decided.
Authentication methods working principle is based on the system taking the necessary information from the user with appropriate methods and transferring it to the computer environment. Then the data is analyzed and the right person is determined by comparing it among the users registered in the system. (Kumar ve Walia,2011).

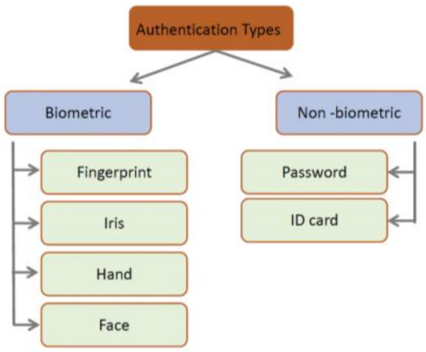


Figure 5: Different authentication methods

In an article that argues that the aforementioned biometric authentication methods do not work efficiently under all conditions, hand gesture authentication is adopted. However, in case it can be easily imitated by other people, they have prevented imitation by detecting fake movements (Terui ve Hosobe, 2022). In another study, it was used in authentication by arguing that people's hand movements are physiologically and behaviorally different from each other. By performing feature extraction with the 3D Temporal Difference Symbiotic Neural Network (3DTDS-Net), it was determined whether the hand movement belongs to the person or not (Song vd., 2022). Palm analysis is one of the most preferred methods in hand authentication. The lines on the palm can be used to distinguish people. In research where authentication can be provided in this way, the palm was detected using basic operations such as threshold,

segmentation and ROI. Feature extraction was performed on the detected palm by applying a number of morphological methods. Template creation and matching were performed, and hand authentication was provided using the PBNN method (Cheng vd. ,2003). Another article focused on the veins as well as the palm lines. Multiple authentications were made using each of the vein analysis, fingerprint and IR hand geometry and it was aimed to increase the accuracy . First of all, palm and finger segmentation was made from the IR hand image, and the distances between certain points of the hand were determined. Later, a single fingerprint was extracted from the hand image and all these methods were used simultaneously for authentication by fusion method (Gupta , 2018). Despite these studies, in a study that detects the hand by making visual inferences, the external image of the fingers and the internal vein images were recorded by the advanced hardware system, and the point clouds were processed with the PointNet and DGCNN algorithms (Yang vd., 2021). In the analyzes on the bare hand, which is understood from the projects carried out, usually trainings were made with segmentation-based deep learning models that provide visual discrimination such as veins and hand lines. Against these, there are also authentication methods using only hand geometry in the literature. In the article named is Using of Hand Geometry in Biometric Security Systems, It is argued that the physical dimensions of the human hand contain information that can confirm the identity of an individual. In this project, 21 different lengths on the hand were measured and feature extraction was performed. Calculation will be made between the features coming from here and the pre-recorded image and the test image for comparison with 3 different methods. Euclidian distance, Hamming distance, and Gaussian mixture models (GMM) were tried. Expectation-maximization algorithm was used to estimate the density parameters of the GMM model, and the best result was obtained from the Gaussian mixture model (Varchol, Levicky, 2007). Apart from the methods mentioned, there are also cases where it is argued that the behavioral movements of the hand are personal (Fong vd., 2013). Different hand movements and the postures of the hand while performing them are examined and recognized by the classification model. In order to obtain a clear hand image from the images, several different pre-processing methods such as background subtraction were applied. Then, 27 features thought to be distinctive in hand gesture recognition were identified and used in the classification algorithm. The promising results from here strengthened the thesis that behavioral movements are a distinctive feature.
As mentioned above, authentication aims to prevent systems from being used by the different people. It can be provided in different ways depending on the intended use. Biometric recognition systems aimed to eliminate the disadvantages of traditional systems and make authentication user friendly. The limitations encountered here can be resolved with the development of technology (Sabhanayagam vd. 2018).

In conclusion, this introduction has provided an insightful glimpse into the fascinating world of hand gesture recognition systems with authentication optionality. By highlighting the growing importance of such systems in various domains, outlining their potential benefits, and discussing the objectives of this report, we have set the stage for a comprehensive exploration of this cutting-edge technology. As we delve deeper into the subsequent sections, we will delve into the underlying principles, examine the state-of-the-art techniques, and discuss the challenges and opportunities that lie ahead. By the end, we hope to not only gain a deeper understanding of hand gesture recognition systems but also recognize the immense potential they hold in revolutionizing authentication mechanisms.

3. ORIGINALITY:

In recent times, innovative methods started to be used by following technological developments in the interaction between humans and computers. One of these methods and the method that has been selected to be used in this project, is using hand gestures to control and manipulate the computers. In the literature, this method is mostly used in systems that have been developed for older people or visually challenged people. In those projects, hand gestures are being used to usually control home appliances. Also, when the literature is scanned, it is acquired that there is no any sufficient and accepted application for the control of devices in healthcare systems. Therefore, the project focuses on controlling healthcare systems. In the healthcare systems, especially in surgeries, it is highly emphasized that the human-computer interaction systems have lack of sterilization, and the absence of sterilization causes the transmission and proliferation of several pathogens on the surface. Thus, the first originality of the project is to develop a sterile real-time human-machine interface. By this way, wasting time in sterilization and spending too much effort for it are aimed to be prevented. Generally, recognition and authorization of a person is not needed in the literature. However, in the healthcare systems, the security and user characterization should be at the forefront to prevent any chaotic situation because of unauthorized people. It is aimed to prevent the movements of unauthorized people by giving authorization to certain people and to prevent any confusion that may occur in the recognition of the hand gesture of the dominant person, despite the presence of more than one hand in the detection window of the program aimed to be developed in the project. The integration of authentication gains an originality to this project. Furthermore, in the literature, since the system is easier to prepare, to control the home appliances, generally hardware has been used. In this project, mostly a development software is purposed to be used instead of hardware establishment, which is other originality of the proposed project. This decreases the cost of the project and gives more flexible surface area to clinical staff in the surgeries since any integration of sensor-based application will restrict the comfortable movement of surgeon in the operation room. As is well known, the hand gestures vary from culture to culture and from country to country, and the standardization is another missing point in the literature. Thereby, it is necessary to standardize the hand gestures to make the hand gestures universal. In this project it is mainly aimed to create a dataset including standardized hand-gestures and this aim also increases the originality of the project. As a final originality of the recommended project, it is being expected to develop a system which dynamically recognizes two-hand motions which has not previously been included in this field in the literature. Additionally, the researchers have generally developed a dynamic dataset by using their systems' hardware and capability. This makes the updating dataset quickly and assigning a person as authorized in the information entrance and presents more reliable predictions.

4. SCOPE OF THE PROJECT AND EXPERIMENTS/METHODS:

The project called as “Hand Gesture Based Device Control with Applications in the Medical Domain” is going to be carried out in the scope of the Computer Engineering and Biomedical Engineering graduation project which will be carried out under the supervision of Prof. Dr. Reda Alhajj. The fundamental purpose of the project is to develop a hand gesture recognition system which will have a potential to be used in the healthcare systems since the current healthcare systems have lack of efficient human-computer interaction and there is an urgent need to integrate computer science and biomedical applications more. For this aim, the project consisted of mainly six targets such as development of sterile real-time hand gesture recognition, integration of authentication, software-based gesture identification model development, dynamic recognition of two-hand motions and storing numerical data in a dynamic dataset. In the project, firstly hand gestures were recorded from the different people and the recorded videos were pre-processed in order to create a standardized video streaming dataset. After that, the numerical dataset was created by using MediaPipe library in Python. In the numerical dataset, it was planned to keep the information related to specific keypoints of hand gestures, motion information and contour information of hand gestures. According to principles of the hybrid ensemble learning model, the dataset was arranged as input which is required to train the ML models and get the prediction from the ensemble learning model. After this step, the authentication part was successfully performed by considering the keypoints and contour information calculated from the video stream. The basic device management was achieved, and the efficiency of the recommended system was evaluated by this way. The total time for the completion of the project was 12 months. The obtained results during this period were evaluated and the proposed model developed in the project was compared with the current hand gesture recognition systems in order to prove the efficiency of the developed model. If the designed model is successful at the end of each testing technique (in both local and clinical applications), the proposed system might be commercialized because it will highly serve the sterilization in hospital environment, make it easier to use in the surgeries, have a high security and be cheaper than the current methods.

4.1. Hand Gesture Standardization

In the project, it is aimed to predict the hand gestures dynamically and real-time. In this regard, common hand gestures, which are mostly used for control purposes in human-computer interaction systems in the literature and include single-hand or double-handed identification in the detection field, were standardized first. New ones were continued to be added to the most frequently used hand gestures by the project researchers. Afterwards, standardized hand gestures were recorded from different people as video streams and the streaming manipulations were executed on the created dataset.
The literature review was performed for detection of most common hand gestures and their commands in literature. The common hand gestures were noted down with their assigned names and use purposes. For example, in the literature, hand open palm with five fingers was usually used as a system initiation and hand closed palm was used as a system shutdown. In

this project, those two common hand gestures were used with the same meanings. At least 15 different hand gestures including single and double hands were defined with new hand gestures to be added for the control of medical devices in the health system.

4.2. Dataset Manipulation

The interaction of hand gesture data and algorithm which was developed in the proposed project was planned to be established dynamic. Therefore, having a large dataset was necessary for the machine to learn more information and to give the best results in all conditions. It was aimed to create a large data set containing gestures in different lighting conditions, different finger sizes and physically at different angles in order to prevent the scarcity of the two-handed dataset. In addition, since the aim of the project was to be used in hospitals and medical centers, gloved hands were used in the recorded videos while creating the data set.

For the classification and authentication, the hand gesture features were extracted by using the MediaPipe library in Python environment. The extracted features would be recorded into a database if there was a need. This also might have helped to gain dynamic behavior to the proposed recognition system. However, since the trained model solved the current problems and the project does not have this kind of large dataset for the hand landmark extraction, the information of hand gestures was kept only in trained model instead of store in online database. Addition to this, as stated before, the project also includes authentication part. The structural information of person's hands was stored in dataset including with personal information and also extracted features of the hand characteristics.

4.2.1. Video Streaming Dataset Creation

As briefly mentioned in Section 4.1, standardized hand gestures made with gloved hands were recorded with different electronic device cameras with different resolutions. In the video recording process, it was planned to take at least 15 different hand gestures, which were standardized, from at least 50 different people. Right- and left-hand video streams was recorded for each hand gesture and recorded videos were going to be labeled to indicate right/left hand and gestures. All videos were recorded with a maximum time interval of 6 seconds.

4.2.1.1. Video Streaming Dataset Pre-processing

Pre-processing video streams in a dataset was an important step to ensure that the data was in a format that could be easily used by a machine learning model or streaming manipulations. One common approach was to first resize the video frames to a smaller resolution to reduce the amount of data needed for training by making them all the same size and format. Next, unnecessary parts of the video frames were removed through cropping to focus the model on the most important features. For this reason, also the starting and ending frames of hand gesture were determined for the standardization of video streams to 1s. Then, the pixel values of the video frames were normalized to a specific range in order to improve model performance. Data augmentation was applied by generating additional training data by applying various transformations like rotation, flipping, and cropping to the video frames which helped improve model robustness.

4.2.2. Streaming Based Data Retrieval and Hand Gesture Dataset Preparation

The MediaPipe library in the Python environment was used to create the hand gesture recognition system which was proposed in the project. MediaPipe's pre-built pipeline was used for capturing the video stream since it provided several pre-built pipelines for capturing processes from different sources such as webcams, file and so on. Pre-trained models were also used to detect the hand keypoints in the recorded video streams. MediaPipe provided several pre-trained models for hand keypoint detection in terms of the location of the fingers, palm, and wrist. In the project, a new pipeline was created for processing the recorded video streams and detecting the hand keypoints. After the pipeline was started, it processed the recorded video streams and detected the hand keypoints in real-time. The detected hand keypoints were utilized to create a numerical dataset of hand gestures from the video streaming dataset.

4.2.2.1. Data Pre-processing

A video processing library such as OpenCV was used to pre-process the dataset. First, the frame rate of each video was determined to standardize the frames according to the hand gestures. For this, a buffer was assigned, which was size of 5. To determine the equal 5 frames for each video, maximum frame number of each video was calculated and then, each frame number was divided by 5 to find the time interval between each frame. Furthermore, frames were extracted from the videos in the dataset. Next, the hand region was extracted from each frame. The extraction was done by using skin contour detection and object detection to isolate the hand in the frame and remove the background. The images were resized and normalized to make them all the same size and format for consistency in the dataset which included the numerical representation of each frame.

4.2.2.2. Feature Extraction, Scaling and Selection

Features of hand gestures were extracted from the frames such as hand keypoints, contours, or optical flow information for motion recognition. The hand keypoints and optical flow information were used in the classification while the information taken from hand keypoints and contours were being used in the authentication. After the keypoints were detected properly, the location of each keypoint in x-axis, y-axis and z-axis were stored in a DataFrame. The information of distance between keypoint in the tip of each finger and end of the finger were calculated at first and recorded in a dataset with its unique label. Then, the information of distance between keypoint in the tip of each finger and the palm were calculated and again extracted information were appended into the same dataset. The contours and motional

information were also stored in the same dataset to be used in the classification and authentication parts.
For the classification, the ratio of distance of finger's tip/end to distance of finger's tip to palm was calculated. A matrix form was created by having the relationship between frame to finger specific ratios. Then, the obtained array form was squeezed for each video and the squeezed data were appended into a new DataFrame which each rows represent the video numbers while the columns were representing the frame-based hand gesture specific ratios.

4.3. Classification Model Development with Hybrid Ensemble Learning

In the project, it was aimed to use the hybrid ensemble learning for the hand gesture classification. The main aim behind of this was to get the optimal accuracy from the training process to create a model which dynamically predicted the hand gestures correctly in real-time. There were several classification methods that were used for this purpose. The classification methods generally depend on factors such as the size and complexity of the dataset, and the resources available for training and inference. For this, the proposed models were established six main architectures for the training process.
In the development of ensemble learning, it was aimed to combine the deep-learning and machine learning classification algorithms. As a possible models, Random Forest (RF), Decision Trees (DT), Support Vector Machine (SVM), Logistic Regression (LR), k-Nearst Neighbors (kNN), Multinomial Naïve Bayes (MBN), Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and bidirectional LSTM (biLSTM) models were trained with the created dataset. After that, the best five models were selected to be used in the last version of the hybrid ensemble learning model. In order to select the best parameters for the classification models, Grid Search was used. According to the results from these techniques, the most related parameters were selected as parameters of classification models.

4.4. Model Evaluation and Tuning

The ensemble model was trained on the training set and then evaluated on the unknown testing set. This process was repeated multiple times with different splits of the data, and the results were averaged to give a better estimate of the model's performance on unseen data.
In the performance evaluation of classification model, each classification model was tested separately. Accuracy, precision, recall, F1-Score, area under the ROC curve (AUC) and K-Fold cross validation metrics which provided a more accurate estimate of the model's performance and helped to identify overfitting were used. According to the results obtained from the models' evaluation, the classification model was optimized by tuning the architecture and parameters of the hybrid ensemble learning model.

4.5. Authentication Based on Feature Scaling

The user was gained authorization according to structural characteristics of his/her hands. The authentication was provided by keypoints and distance-based ratios obtained on hand. In order to standardize the holding position of hands, the reference parameters were created such as distance from camera and targeting center of the window. By using hand landmarks, the distance from camera was calculated. Afterwards, a red dot in a green square were placed at the center of live window. With the standardization purpose, users were asked to keep their MIDDLE_FINGER_MCP point to red dot at the center and arrange the hand's distance from the camera. By following that strategy, the dataset was aimed to be created for the authentication. Images of 3 different people were taken at standardized distances (700-705 units) from the camera and at centralized positions. At the end, 21 images were recorded for each person and a dataset including 63 images were generated at total. In order to augment the dataset, each image was rotated between -5 degree and 5 degrees. Rotated versions of the images were saved with rotation strategy in data augmentation technique. After the dataset was ready to use, the keypoints were generated for each image and then, the Euclidian distances between keypoints which were decided by application were calculated. After the distances were normalized, ratios between the distances were generated. At this stage, three different ratios have been obtained for each finger in one hand: the ratio of distances between non-locally moving hand joints, the ratio of these ratios among themselves, and the overall ratio. Also, the structural information of palm was extracted by using the same strategy. Distance between the keypoints at knuckles and wrist were calculated and then, first two and other distances were proportioned each other. At the end, there were obtained two different ratios. To find more characteristic features which is deterministic for the hand structure, those two ratios had been proportioned each other again and the third ratio was obtained. At the end of those computations, 18 different ratios were obtained and stored in a structure of DataFrame. With those values, classification was made using the SVC algorithm. In the model training, training set included 70% of total dataset, testing was consisted of 30% of dataset and also, the system was tested on unknown image dataset. After best parameters were detected with the help of Grid Search algorithm, the model was recorded in .plk format to be used in real time. While the person was controlling the medical device, the hand information of the person in the real-time video stream being received from the camera was compared with the information determined in the model. In that case, the accuracy of the prediction was evaluated by looking at their probabilistic results in real time. In order to distinguish the hands that were not in the model during the test process, the threshold value was determined. If the person's hand characteristics are below the specified threshold, the person is not authorized to control the device. Otherwise, the system recognizes the person and allows the person to use the system.

4.6. Device Management with Estimated Hand Gestures

The developed gesture recognition system was integrated into the device management framework. The users who are authorized in the system can operate the device with predefined hand gestures. Once a predefined hand gesture is recognized by the system, it triggers an appropriate response. In the light of this purpose, after the authentication and hand gesture prediction algorithms were determined, the functionalities of hand gestures were introduced to system. Clicking, opening, closing, zoom in/out like hand gestures were selected and assigned to hand gestures. Besides, to design a user-friendly system, the virtual mouse integration was integrated to the system interface. For this integration, the PyAutoGUI module in Python environment was used. When the sliding hand gesture is recognized by the prediction system, the virtual mouse control is activated and then, the user is able to control the local device's pointer.

5. PROJECT TARGETS AND SUCCESS CRITERIA:

According to statements given in Section 4, the project contains eight general targets. However, these targets can be adjusted, detailed, and divided to sub-targets when there is a need.

5.1- Scanning literature for the current applications in the hand gesture identification systems
5.2- Standardizing the hand gestures according to common utilization of them in the literature and meaning of commands
5.3- Creating a pre-processed video stream dataset by recording standardized hand gestures from different people
5.4- Preparing numerical dataset of hand gestures for gesture classification, authentication and device management
5.5- Development of machine learning and deep learning-based classification model with hybrid ensemble model by using common available algorithms such as DT, RF, SVM, LR, k-NN, MBN, RNN, GRU, LSTM, and bi-LSTM
5.6- Analysis of results taken from trained models' algorithms by using Accuracy, precision, recall, F1-Score, area under the ROC curve (AUC) and k-Fold cross validation
5.7- Assigning authentication to specific people and validation of the authority
5.8- Managing simple functions of device after the hand gestures are correctly classified

The success of the project can be fundamentally measured with eight criteria. For Item 5.1 and 5.2, the standardization of gestures is planned to be carried out by deeply scanning the literature. It is the first criteria of the project, which determines the project success. The predicted success of this target is 15%. For reaching to this criterion, at least 50 articles including hand gesture assignments or relevant information will be reviewed and common hand gestures are noted down. For Item 5.3, the large dataset is being expected to be occurred by recording hand gesture videos from different people. For this criterion, the estimated success is 10%. When the video recording is pre-processed after being taken with at least 15 different hand gestures taken from at least 50 different people, the criterion is assumed as being achieved. The success criterion of 5.4 depends on several factors such as use of MediaPipe, feature selection, keypoints manipulation, data arrangement, efficient data storing. The predicted success criteria of this item is 15%. When the dataset is created by the videos that has been collected and considering at least 5 video and extracting at least 21 keypoints' features dedicating the keypoints of one hand, the item implementation is going to be accepted as successful. For Item 5.5, there is planned to establish 14 classification algorithms and together them in an ensemble model. In this part, the success criteria is determined as 20% because the complexity of the input arrangement of models. If at least 10 models are implemented separately and 5 best models are selected to implement in the architecture of ensemble learning, the estimated success criterion is reached successfully. For next item related with evaluation metrics, Also, when the training accuracies are taken larger than 60%, precision greater than 65%, F1-Score between 0.5 and 0.8, AUC of 0.7 to 0.8 and the cross validation is divided into at least 5 folds, the training processes will be thought as successful, and the other half of the estimated success criterion is assumed as successfully obtained. In the authorization phase, on the condition that at least 1 of the 3 hands is correctly identified, a 20% success criterion will be met. Final target of the project is to manage a device by using developed system in the project. WebCam of any computer is planned to use in scope of this target. With the proposed model, when basic functions such as opening, closing, double-clicking, zooming in/out are fulfilled, it is expected to reach 10% success criteria.

6. RISKS AND B PLANS:

In the real life, each project has risks. Therefore, these risks should be considered if it is expected to achieve the project with success. In this project, there were several risks which some of them were predicted and with their alternative methods and some of them were not able to be predicted because of the unexpected events in project period. In the below table, review of the encountered risks and the corresponding solutions during the project implemented.

Work Package #	Risk	B-Plan
WP#3	Enough hand gestures could not be obtained in the dataset acquisition.	Datasets and videos in the online databases used.
WP#3	Video dataset did not have enough single or two hand gesture videos for the hand gesture recognition.	More videos recorded and added into the video dataset from the volunteers.
WP#5	The classification algorithms had some weakness and did not work as expected.	Hybrid Ensemble Learning model used as classification method.

WP#5	Hybrid Ensemble Learning method was not sufficient for the integration of complex classification models.	Different types of ensemble learning have been tried. For example, stacking ensemble learning is one of them.
WP#6	During the training of the models, the obtained results of the evaluation metrics did not meet the desired expectations.	In order to find the best hperparameters and solve the underfitting problem, Grid Search Algorithm was used.
WP#7	The number of keypoints obtained to be inadequate for the prediction of authentication.	By using MediaPipe, the new keypoints' location and ratios established and calculated. These new features recorded to the dataset to be used in the authentication part.
WP#7	The authentication result and prediction of a person from their hand obtained with low accuracy due to that algorithm.	A new algorithm has been developed based on the results obtained from previous authentication attempts.
WP#8	While creating the interaction between the device and Webcam, the same console could not be used for the authentication, hand gesture recognition and virtual mouse parts.	For authentication, hand gesture recognition and virtual mouse parts, the code has been adjusted to dynamically create a new console whenever necessary.

7. WORK TIME PLAN OF THE PROJECT:

The work time plan is presented in Table 3. According to WP#1, each week in the Project II's semester, literature review was performed because updating dataset with the current information and gaining more knowledge about the project's topic. Hand gesture standardization, video streaming manipulations and streaming data retrieval and preparation was continued until the ends of this semester. Since the videos were continuously being collected and there were always new videos to be processed, the WP#3 was also pursued until the end of the semester. Because the WP#3 was continued and the new videos were added to the dataset, the WP#4 was also continued this semester to create a numerical dataset. In this semester, it was mainly planned to perform WP#5, WP#6, WP#7 and WP#8. For the classification model it was planned to complete the construction of classification during the semester. WP#6 was begun after the 5th week since the construction model architectures took time. After the 7th week, the project acquired authentication feature and the related work package was continued till the end of the semester. Finally, in the 11^{th} week, the integration of proposed model to device were actualized and the device was successfully tried by using hand gestures used in this project.

8. FINANCIAL EVALUATION:

In the project, a computer with the necessary computational power was required. Istanbul Medipol University was capable to provide the required computer. On the other hand, with this project, it was planned to apply TUBITAK 2209-A to get the support for online source purchase. However, because of the time availability and confliction with other' projects supported by TUBITAK of the researchers, the researchers were not able to apply the TUBITAK 2209-A project. Addition to this, the researchers have decided to utilize their own computers and computer equipment such as external storage devices to finish the project with everything in their power. At the end of the given time for project completion, the researchers completed the project as expected without paying any fee.

9. RESULTS:

In this project, as the first step, the studies that have been done before in the literature have been examined to determine standard hand gestures and their functionality as intended. Currently more than 15 different hand gestures were selected to be used in this project. The selected common hand gestures are listed in Figure 6.

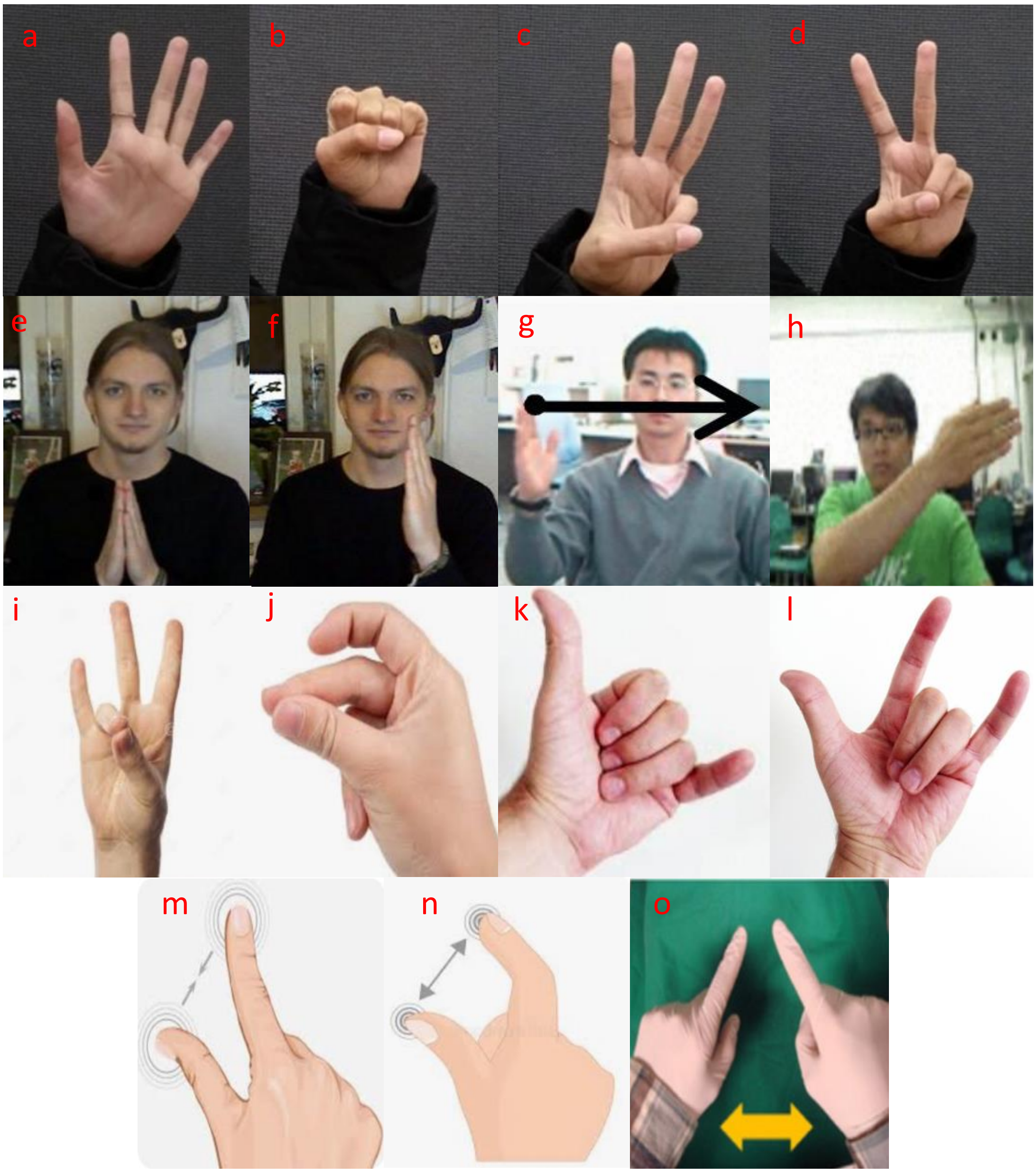


Figure 6: Standardized dynamic hand gestures **a)** System Open, **b)** System Close, **c)** Right Click, **d)** Double Click, **e)** Pause, **f)** Next, **g)** Forward, **h)** Moving Left **, i)** Close Virtual Control, **j)** Left Click, **k and l)** File manager, **m)** Zoom out, **o)** Zoom in

For the dataset creation part of the project, it was decided to use four hand gestures (Figure 7) that have been determined. The reason for choosing these four hand gestures was to make sure that both finger movements and the whole hand movement were considered.
In the Figure 7, the common hand gestures which were used in the implementation of the project can be seen with their movements. Opening hand gesture was selected to open a file

(Figure 7a) while the closing gesture was used to close the whole system (Figure 7b). In order to image management, basically zoom-in/out gestures were used (Figure 7c, e). Also, in order to activate the virtual mouse control, the sliding gestures were used (Figure 7d).

a) b) c) d) e)

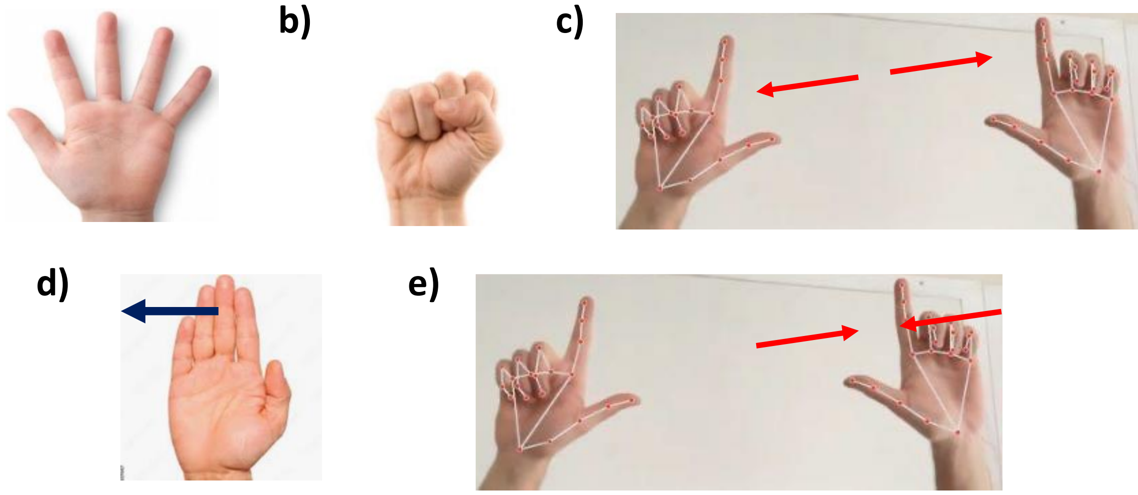


Figure 7: Selected 4 **a)** Opening, **b)** Closing, **c)** Zoom in **d)** Sliding (right to) Left, **e)** Zoom out Hand Gestures

In the next step, volunteers were asked to perform these five hand gestures while wearing and not wearing gloves. These hand gestures recorded, and a streaming dataset created from them. By using MediaPipe and OpenCV libraries in Python, an algorithm was used for detection of hand landmarks. The result of this algorithm with the input videos that have been taken from the volunteers can be seen from the Figure 8 and 9.

Figure 8: Obtained Result from one of the “Opening” Hand Gesture by Keypoint Detection Algorithm

Figure 9: Obtained Result from one of the “Sliding” Hand Gesture by Keypoint Detection Algorithm

In the following step, coordinates of the specified keypoints were found. With the help of calculus theorems, the length between the specific points and their ratios was calculated using Python (Figure 10). Since hand measurements varied from person to person, the proportions were taken into account when creating the dataset.

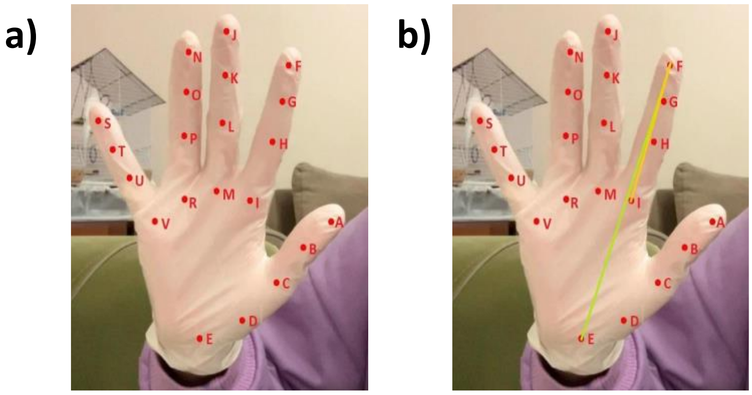


Figure 10: Labeled points found by **a)** Keypoint Detection Algorithm **b)** with FE and FI lengths

In the next step of the project, a dataset was created with the data obtained from the videos. To understand the whole hand movement, changes in the coordinates of the point E and to understand the finger motion, the ratio of lengths from the tip of the finger to the base of finger and to the base of the hand (point E) have been used. These proportions and the changes in the coordinates of point E were calculated for the frames taken from the videos. If these changes exceed a threshold, they were recorded into the dataset as 1 or -1 depending on the motion direction. Dataset where these data were recorded can be seen in the table below (Table 1). The dataset was created to be used in the ML model. For the initial frame and the frames that does not exceed the threshold, changes in the coordinates of point E were recorded as 0. The gesture codes given in the dataset were set to 1 for closing, 2 for opening, 3 for sliding, 4 for zoom-in and 5 for zoom-out gestures.

Table 1: Small part of the Created Dataset

Video Code	Gesture Code	E_{X_0}	E_{Y_0}	$\left(\frac{\overline{AE}}{\overline{AD}}\right)_0$	$\left(\frac{\overline{FE}}{\overline{FI}}\right)_0$	…		E_{X_4}	E_{Y_4}	$\left(\frac{\overline{AE}}{\overline{AD}}\right)_4$	$\left(\frac{\overline{FE}}{\overline{FI}}\right)_4$	
V1	0	0	0	1.357	2.2717	…		0	0	1.3569	2.2654	
V2	0	0	0	1.3816	2.2538	…	…	0	0	1.3882	2.2684	…
V3	0	0	0	1.3834	2.248	…		0	0	1.3861	2.2243	
V4	0	0	0	1.364	2.2279	…		0	0	1.3417	2.2065	
…	…	…	…	…	…	…		…	…	…	…	

The numerical dataset that obtained, was arranged according to the requirements of the planned training set. For this, a new dataset was created as frame-based. While the training dataset was being prepared, 5 frames of each video were selected with special frame intervals since each video has different frame rates. For each frame, the six parameters were considered such as distances between keypoints and direction of the location change of Point E (Figure 10). At the end, for one video, the shape of (1, 60) DataFrame was obtained. It is worth to note that the system might have either one hand or two hands at a time. For this, 42 keypoints' information were kept in the dataset and then, the DataFrame was created in the scope of this strategy. If the system has one hand, the keypoints belonging to that hand was filled and other cells in DataFrame was filled with 0. While the rows were representing the video index, columns were corresponding to the ratios and location change of Point E. After that, each row was labeled according to the hand gestures recorded on video. The created DataFrame was divided into training (70%), validation (30%) sets. Also, each model together with the ensemble model was tested on unknown datasets which was named as testing dataset on the project's dataset. At the end of the dataset creation and manipulation methods, each cell in DataFrame was evaluated manually and the correctness of the extraction technique was approved by researchers.

After the dataset was created properly, the ensemble model was established with RF, DT, SVM, LR, MBN, k-NN, RNN, GRU, LSTM, bi-LSTM and MLPC. Each classification model and ensemble model were separately trained on the dataset. In this stage, to get the best hyperparameters and tune the models, the Grid Search algorithm was used, the obtained results are presented at Table 2. In that case, since the DL algorithms were not compatible with streaming dataset for now and ML algorithms which generated good results during their self-training on the custom dataset, DL algorithms were eliminated from the model. After that time, the dataset was mostly trained on the RF, MBN, DT, LR, DVC, MLPC, and proposed hybrid ensemble model.

According to the metric analysis of established models (Table 2), the higher accuracies were obtained from RF as 0.90% and from SVC as 90% while the accuracy of the ensemble model was obtained as 0.90%. Also, other metric results of SVC and RF were obtained higher than the 0.85.

Table 2: Metric Analysis of ML Algorithms and Ensemble Model

Training Model	**Metrics**				
Name	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**MCC**
RF	0.90	0.94	0.89	0.90	0.85
MBN	0.80	0.62	0.67	0.62	0.72
DT	0.70	0.44	0.56	0.49	0.51
LR	0.80	0.78	0.82	0.79	0.70
SVC	0.90	0.92	0.93	0.92	0.86
MLPC	0.80	0.67	0.62	0.62	0.73
Ensemble	**0.90**	**0.89**	**0.89**	**0.87**	**0.85**

In order to get best understanding about the model behavior on dataset, the consistency of the models was tested with different dataset sizes. For this analysis, the size range of videos were changed between 3 and 33. According to results obtained from the consistency analysis (Figure 11), although the other models might have the higher metrics from the ensemble hybrid model, the consistency analysis pointed out that the ensemble model is more stable than other ML models.

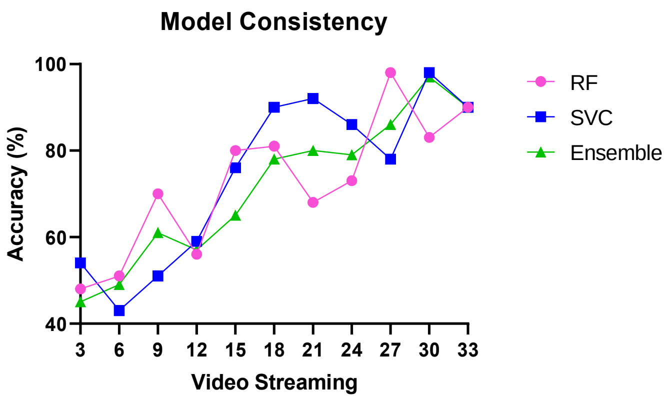


Figure 11: Model Consistency Analysis

In the next part, the OpenCV module was downloaded, and the webcam integration was used to test the trained model in real-time. According to the obtained results given in the Figure 12, the proposed hand gesture detection algorithm in the project, has been successfully achieved.

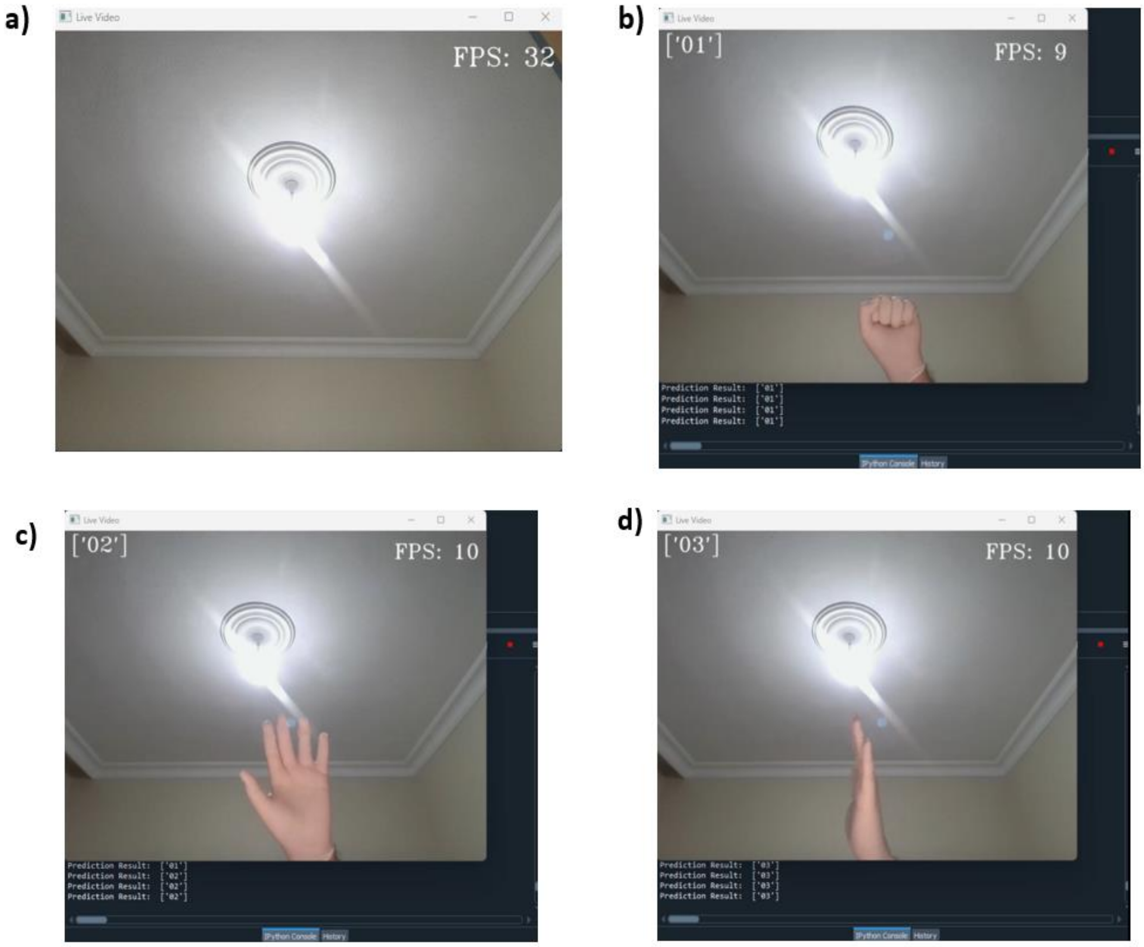


Figure 12: **a)** No-Gesture, **b)** Closing, **c)** Opening, **d)** Sliding (ex-gesture) hand gesture detection from Webcam with FPS values

Table 3: Real-time Testing Results of Gesture Prediction System

	Gesture 1	**Gesture 2**	**Gesture 3**
Gesture 1	0.87	0.08	0.05
Gesture 2	0.03	0.84	0.13
Gesture 3	0.06	0.18	0.78

For each prediction results, the system also returns the probability rate of correctness. Table 3 represents the prediction results taken from the system for each hand gesture. Gesture 1 was predicted with the 87% accuracy while the Gesture 2 was estimated with the 84% accuracy and Gesture 3% was predicted with the 78% accuracy. Here, Gesture 4 and Gesture 5 were not put into table because their optimization was still in progress. However, when they tested with the other gestures on the proposed system, the prediction accuracy was 65% for Gesture 4 and 58% for Gesture 5.

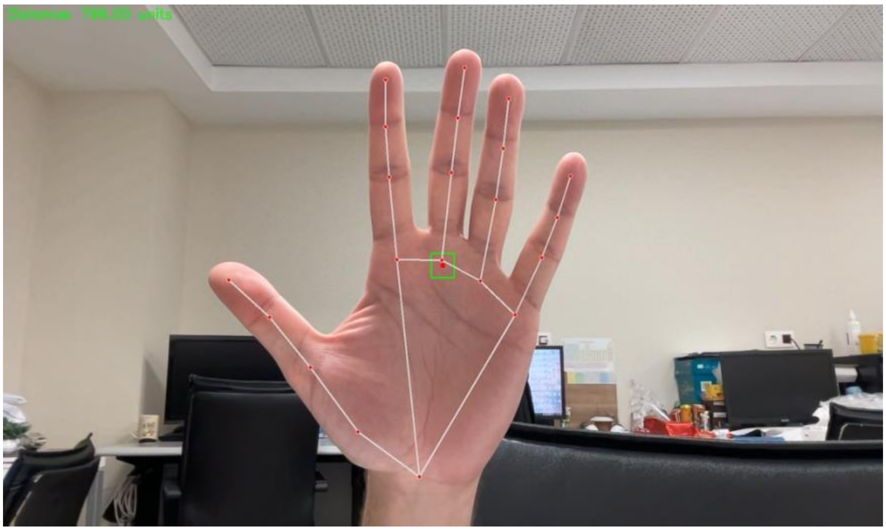
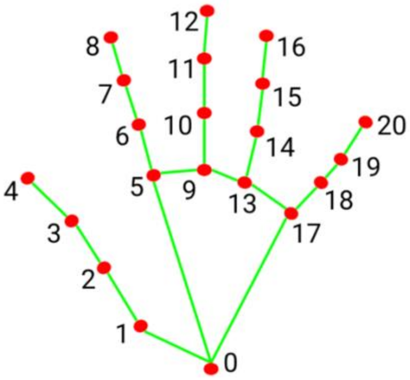


Figure 13: Representation of authentication.

While the prediction algorithm's optimization and dataset collection process were still carried out, the implementation of authentication part was started. In the initial authentication procedure, authentication was aimed through the coordinate of the hands of the individuals. Firstly, the photos were taken with the hands up of the people. In order to enlarge the data set, data augmentation was performed and the number of images was approximately 110. Thanks to the ImageDataGenerator in Tensorflow library of Python environment, the images rotated -15 and +15 degrees were saved. As a result of data augmentation, there were 36 images for each person. These images were taken at different distances and locations to the camera. Thanks to the MediaPipe python library, hand keypoints were determined and normalized. Here, in the normalization process, each finger was normalized within itself and all the last values were kept as data frame. The k-NN model was trained to classify by the created data frame. The model accuracy value was found to be 0.69. The accuracy value increased to 0.82 thanks to feature selection. The created model had 3 main classes. During the test process, a threshold value was determined in order to distinguish the hands that were not defined in the classes. While determining this threshold value, the probability values of each test image in classes were calculated. If the probability value of any class of the hand in the image remained below 0.67, it is understood that the hand did not belong to one of the classes in the model. However, this system was not able to predict the correct hands based on the structure. Therefore, the algorithm was changed.

In the updated algorithm for the authentication part, the researchers applied ratio-based approach. The system was tested for three people with 232 images for each person captured from directly people or produced from data augmentation technique. The system implementation and testing procedure for the authentication part is explained in Section 4.5 in detail. After the images were trained on SVC algorithm, the training accuracy was obtained as 92% and generated packages were tested in real-time. For the extraction of authentication features, the ratios were selected as stated in Figure 14. Also, the features are obtained by using Formula 1, 2, and 3.



$$Ratio\ 1 = \frac{dist_{9_{10}}}{dist_{10_{11}}} \quad (1)$$

$$Ratio\ 2 = \frac{dist_{9_{10}}}{dist_{0_5}} \quad (2)$$

$$Ratio\ 3 = \frac{Ratio\ 1}{Ratio\ 2} \quad (3)$$

Figure 14: Representation of Feature Extraction in Authentication

Table 4: Authentication Prediction Results

	Person 1	**Person 2**	**Person 3**
Person 1	0.9589	0.0152	0.026
Person 2	0.0245	0.9407	0.032
Person 3	0.0785	0.0186	0.903

In the authentication part, the people's hands characterized with an accuracy higher than 90%. Table 4 shows each prediction result in detail. To specifically say that the Person 1 was estimated with the 96% accuracy while the Person 2 was characterized with the 94% accuracy. For the last person, the authentication accuracy was obtained as 91%. Since the probabilistic results of predictions were investigated, each class's probabilities can be seen in the given table. Also, for the foreign user, the probabilistic results obtained with approximately 0.3 or 0.4 and since the threshold of the system is equal to 0.9 (or 90% accuracy), the foreign users could not enter the system.

Figure 15: Open virtual mouse mode

Figure 16: Click gesture

Figure 17: Double click gesture

Figure 18: Exit virtual mouse mode

For the next part, the authentication and hand gesture components were integrated to enable device control. Opening a file is triggered by the open hand gesture, closing a file by the close hand gesture, zooming in by closing up, zooming out by moving away, and enabling virtual mouse mode by sliding. The person whose verification is provided can take control of the device of the system by performing the gesture in Figure 15 (sliding his hand from right to left). When the virtual mouse mode is active, it also brings three actions as well. By touching middle finger with thumb clicking in performing in Figure 16, ring finger with thumb double clicking performing in Figure 17 and pinky with thumb disabling the virtual mouse mode actions performing in Figure 18 triggered. The system accurately responds to these gestures, enhancing user control and interaction. With this, the system accurately responded to these hand gestures, ensuring precise execution of the corresponding actions.

10. DISCUSSION:

In the “Hand Gesture Based Device Control with Applications in the Medical Domain” project, the main goal was developing a system that helps to control computers in the hospital or medical centers. To be able to achieve this goal, firstly a literature review was conducted for hand recognition systems regarding their purposes, requirements, and methods. According to the literature, hand recognition systems are mostly used to help visually challenged and old people to control home appliances. In those papers, for hand recognition, mostly devices and sensors were being used. However, in this project which is aimed to be used in medical area (surgeries etc.), using a hardware would create some problems such as sterilization issues and limitation of mobility for hands. Therefore, it was decided to use software to complete this project. After that, the literature review became a little bit more specific. Standard hand gestures and their functionalities were determined by collecting them in a file. In these hand gestures, the ones that are not compatible to be used in our project considering its aim and usage area were removed from the file. As stated in the result part, currently at least 15 hand gestures with their meanings were determined by being reviewed at least 50 articles. With this development, %15 of the project was considered to be completed successfully.

In the second part of the project, the researchers focused on how a hand gesture can be recognized in a digital environment. Initially, the solution considered was usage of a sensor. It was decided to develop the project based on using the Kinect sensor to detect hand gestures through infrared beams and determine the corresponding gestures. However, it was quite challenging to acquire this sensor. Therefore, researchers wanted to develop a project that could be easily used by everyone without the need for additional hardware. Thus, for hand gesture recognition, it was decided to gather the necessary data from commonly available WebCam. As a result, a video stream dataset was prepared by recording four different hand gestures from different people. While selecting these hand gestures, to be on the safe side, two different hand movement types were selected. With this way, both the whole hand movement and the finger movements were considered to understand if there was a problem while detecting the hand gestures. Also, while creating the video stream dataset, first of all, videos that contains the hands with gloves were used. With this way, the researchers became certain that using this system with a glove will not affect the system. Since the product is planned to be used in hospitals, it was important to obtain this result. As stated in the result part, 5 different hand gestures were taken from at least 50 different people. According to the proposed success criteria, %20 of the project was fulfilled without an error. For this part, %5 of the completion is still not met since there should be 7 more hand gestures (3 of them used in virtual mouse control) for the dataset. Unfortunately, as outlined in the work-time plan, addition of 7 more hand gestures could not be completed within the expected timeline. The challenges encountered during the optimization and debugging process of the algorithms hindered our progress in expanding the video dataset with 7 more hand gestures which was selected in the project's first part. This limitation affected the overall advancement of the project within this semester. It is important to acknowledge the complexities and challenges involved in the algorithm optimization and debugging, that can often require additional effort and time to overcome.

In the next part of the project, the video stream was converted into numerical values to create a numerical dataset. To convert the video streams into numerical values, keypoint algorithm has been used with the help of MediaPipe libraries. With this algorithm, 21 keypoints of a hand extracted as can be seen in the Figure 8 and 9 which is prepared with two different video streams. In this part, 5 frames from the each of the video has been used. The coordinates and the distances between the hand landmarks, and the ratio between these distances are calculated with a Python code. The Euclidean ratios were decided to understand the finger movements. So, the current ratio between the whole finger length from the base of the hand and the finger length was being used for each finger. In addition, for the whole hand movements, the coordinate of the point E (Figure 10) was recorded. Because changes in the point E are informative regarding the displacement of the hand during a gesture change. After the calculation, according to the direction of the motion, if the distance of the coordinate exceeds a threshold, 1 or -1 recorded to determine the hand's motion. After the calculations, a numerical dataset given in table 1 was created with the same code. Originally, it was planned to keep this dataset in a MySQL database. However, since there the dataset is no too large to handle, the information of hand gestures was kept only in trained model instead. The reason of the forming this dataset was mainly to create an ML model for prediction of the hand gestures. It will also give a chance to automatically re-train the model and quickly authorize the users by keeping characteristic biometrics in the dataset. With respect to the proposed success criteria, since more than 5 videos and 21 keypoints used to create a dataset successfully, %35 of the entire project is completed successfully.

The hybrid ensemble model was established with five classification algorithms in ML. For the model, RT, DT, LR, k-NN, SVM, MBN and MLPC were used to create the ensemble model. For each separate model and ensemble model, the optimum hyperparameters were detected by the Grid Search Algorithm. The models' structures were successfully established and connected each other to create the ensemble model. So, %20 of success was obtained from this part. At the first stage, 33 videos from 3 different left-single hand gestures were trained. For this, the numerical dataset was arranged according to equal frame intervals for each video. For each frame, six parameters for one hand on frame were kept on the dataset such as five distance ratios of hands and one the location change of point. Since each video has different frame rates, the project members selected a buffer which has a size of 5 and new training data were created with corresponding frame's numerical data. For the training process, this numerical dataset was divided into training (70%) and validation (30%) sets. Furthermore, the trained model was tested on unknown data. Accuracy, Precision, Recall, F1-Score and MCC metrics were used as model evaluation metrics. According to the results taken from the model analysis (Table 2), SVC performance showed better performance than other models. Also, RF was experienced as the second-best model. Meantime, accuracies of other models like MNB, LR and MLPC were taken approximately 0.80% while DT with 70% accuracy was showing worst performance among the other models. The accuracy of the ensemble model was computed as 0.90%. This result is good for a dataset of this size. It seems like the use of RF or SVC already solves the problem. However, the consistency of them is not stable and predictable for different sizes of dataset. Thus, in the project, it is claimed that the ensemble model shows more consistent behavior on different dataset configurations. When the matric results of the first trained model with low size of dataset was considered, this much of accuracy is in the acceptable range to simply prove the advantages of use of ensemble model in the hand gesture prediction. For the next stage, the model was trained with a larger dataset which should include at least 15 different hand gestures after the model was tuned with the new classification algorithms. After the analysis of the results of trained models, it can be stated that the %10 of the project is completed as well.

After the model was trained, the next step was to test the webcam and trained model configuration. For this purpose, the frames and model were implemented into Python script and the reaction taken from the webcam was tested. According to the result of the first trail, when the keypoints were detected, the FPS rate suddenly dropped to 1 from approximately 30. This created a problem to take late response from the detection model which means the model could not be used in real-time. Therefore, the multiprocessing technique was implemented in the prediction model to increase the FPS. With this method, FPS rate increased to approximately 10 while the gesture was being predicted. As a next step, the model's complexity is expected to be decreased more by considering the loop statements and also increasing the buffer size to get better performance at real-time applications. After the implementation of the double-hand gestures, the FPS was taken as approximately 3 which was accepted as low range again. However, for this, the reason to take this kind of result on real-time applications is mostly to not have sufficient computational power because the complexity of the algorithm is in the acceptable range. To get better understanding this problem, the algorithm model should be tested in computer systems which have more computational powers.

In the authentication part, the aim of this project is to gain the authority to use the system according to the hand characteristics of the person. When the literature is reviewed, palm is generally analyzed for hand authentication. Since methods such as skin color and palm lines are characteristic for individuals in palm analysis, separation between users could be achieved with high rates. Fingerprint is the most preferred method for authentication as it is unique for each person. When the hand is not focused on, face and iris recognition are also among the methods used today. Since this project will be used in a hospital environment, users will have white gloves. Fingerprint authentication application is not suitable for this project in the operating room environment where it is not possible for the user to be without gloves. For the same reason, it is not possible to detect the hand skin color of the users and the opaque glove color makes vein and hand line analysis impossible. Assuming that doctors wear masks, face recognition would not be an appropriate method within the scope of this project. For all these reasons, the project focused on hand's structural geometry initially. It was desired to use finger thickness and interdigital angles, but it was realized that it was not possible to do this in a gloved hand. Because the glove was too big for the hand or folded, the defect points could not be determined correctly. On the other hand, the end points of the fingers were determined and the ratios between these points were used. However, the fingertips could not be detected successfully. In addition to all these, the ratios between the points determined on the fingers were very close to each other and it was thought that these ratios were not a distinguishing feature for individuals. Although it is desired to make finger measurements by determining the reference on the video, it did not seem possible to do this in real time. For this reason, a new method has been proposed as authentication. In the new authentication method, the focus is on the coordinates of the key points at hand. These points were normalized and personalized mapping was created. The normalization part was initially done for all fingers on the hand, but the normalization method was changed considering that finger-based normalization would yield more accurate results instead of this method. Coordinate values (x and y values) in each finger were normalized in itself. Normalized values from 5 different fingers were finally combined into a single array. In this way, the dominance of the fingers was preserved and the accuracy of the studies carried out after this process increased. The SVM model was created with normalized values, but the model gave low accuracy due to the weak structure of the method. In this part, the alternative method, the k-NN method, was tried and better results were obtained. However, since the accuracy value of 70% taken from k-NN was not satisfactory for us, studies based on developing the model continued. Thanks to the future selection method, the features that affect the operation of the model well have been selected and the 42 features used in the data at the beginning have been reduced to 25 features. In this way, the accuracy value reached approximately 82%. When the authentication made by hand geometry is compared with the coordinate system-based authentication method, it can be concluded that the new method works much better than the first. In coordinate-based authentication, glove-related problems encountered in determining hand geometry have been overcome. Since key points from the MediaPipe are used in the coordinate-based authentication code, the margin of error in hand tracking is low. However, the code does not always work with a high success rate. When coordinate based authentication is run in real time, it could be said that this part works with 15% accuracy since it was not able to recognize the hand with high success. Therefore, the algorithm was optimized. In the optimized algorithm, it was recognized that use of coordinates cannot be guarantee that the detected features belong to specific class defined in the system. Therefore, the first thing to do against this problem to standardize and mapped the hand gestures in real-time. For this reason, on the center of the window, red dot in green square was placed and it was asked users to overlap their MIDDLE_FINGER_MCP point with that central red dot. Also, secondly the researchers though to arrange the hand position according to distance from the camera. Because it was recognized that the when the distance is changed, the extraction information from the video frames are changed and this situation creates a fluctuation in the prediction accuracy. Therefore, the optimal distance for the local computer systems was detected as 700- 705 units. From each user in dataset, it was asked to arrange their hand according to given instructions and then, their hands were captured. In order to avoid from overfitting and underfitting problems during training process, the data augmentation was applied on the current image dataset and, at the end of the process, 796 images were obtained and they were trained on SVC algorithm. The training model accuracy was obtained with accuracy of 87%. When the generated model was trained on real-time, in the detection of each class, the algorithm reached an accuracy which higher than 90%. Also, it can be truly said that the researchers created a better authentication system than before. Additionally, it is worth to say that the threshold to give authentication to users is arranged as 90% which is very high for prediction algorithm. Moreover, when the systems recognized foreign hands, the probability of each class were obtained in near values which were approximately 0.3 or 0.4. This results also shows that the authentication selects the correct user and the taken results from the models are so deterministic.

In the end, to achieve the predefined objectives and complete the project, the hand gesture recognition and the authorization parts were combined. This combination was adjusted so that a person first needed to be authorized before anything. After the user authorized, the system starts to check for the user’s hand gestures which exist in the collected data (videos). When a hand gesture recognized, the system has been set to perform corresponding response of the gesture. Currently, the system can recognize 5+3 hand gestures, enabling it to generate responses such as opening and closing a file, zooming in and out, and enabling virtual mouse mode. The virtual mouse mode is particularly important for device management, as it allows users to control computers without a physical mouse. For example, when opening a file, it is not sufficient to have the open code; the file also needs to be selected. In summary, the system's ability to perform these eight basic functions indicates that approximately 10% of the project has been completed. These functionalities form the foundation for further advancements in the project, allowing for more sophisticated and diverse interactions with the device.

As a result, the researchers achieved a satisfactory success rate of 95% based on their predefined criteria, validating the effectiveness and reliability of the integrated authentication and hand gesture recognition system.

11. CONCLUSION:

In the "Hand Gesture Based Device Control with Applications in the Medical Domain" project, as in our expectations given in the work-time plan and success criteria parts, %95 of the project completed successfully. To begin with, more than 50 articles that are related to the hand gestures recognition systems reviewed to understand which methodologies, requirements and challenges are being used in the literature. Then, a more specific literature search was conducted for standardization of the hand gestures and their commands. In the end, 15 hand gestures are selected with their meanings and stored in a file. By successfully determining these gestures, %15 of the project successfully completed. In the second part, a video stream dataset that contains hand gesture videos for 5 hand gestures from 50 different people with gloved and non-gloved hand recorded. This part was expected to completed in this semester. But, due to some problems in the algorithms, the researchers faced challenges that prevented them from allocating time to record the remaining 7 hand gestures. So, since there should be more hand gestures in the video stream dataset, %20 of the project is completed. In the next step, from the videos in the created video stream dataset, with the help of MediaPipe library and some algorithms that have been developed, the features extracted, scaled (Figure 7, 8, 9) and stored in a dataset (Table 1). Currently this dataset contains more than 5 videos and exact 21 keypoints which means additionally %15 of the project requirement is also fulfilled without facing an error. Then, the machine learning model was successfully developed with %20 of success. In the next step, the model is analyzed without an error. With this, %10 of the success criteria is fulfilled as well. For the following stage, users' hands are recognized by comparing them with the model. The model achieved an 87% accuracy, and this is a sufficient result to recognize the owner of the hand. So, the total percentage of the authentication part (20%) has been completed. In the end, the authentication and hand gesture recognition parts that have been developed combined and integrated into device management framework using python. With this implementation, the device successfully responds to the predefined hand gestures, contributing a 10% improvement to the project. In conclusion, despite encountering unexpected setbacks, 95% of the project has been completed, and the project objective has been achieved with satisfactory results.

12. ASSESSMENT OF ENGINEERING COURSES:

The project highly requires algorithmic information, good knowledge in selected programming languages, image processing techniques, and machine learning techniques. At Medipol University, Introduction to Machine Learning, Data Structure, Algorithm Analysis, Introduction to the Computer Vision lectures help to achieve each goal of the project successfully.

13. REFERENCES:

Alsaffar, M., Alshammari, A., Alshammari, G., Almurayziq, T. S., Aljaloud, S., Alshammari, D., … Koundal, D. 2021. "Human-Computer Interaction Using Manual Hand Gestures in Real Time". Intell. Neuroscience, *2021*.

Blandford, A. 2019. "HCI for health and wellbeing: Challenges and opportunities". International Journal of Human-Computer Studies, *131*, 41–51.

Chai, D., Bouzerdoum, A. 2000. "A Bayesian approach to skin color classification in YCbCr color space". 2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119), *2*, 421–424 c.2.

Changruenngam, S., Modchang, C., Bicout, D. J. 2022. "Modelling of the transmission dynamics of carbapenem-resistant Klebsiella pneumoniae in hospitals and design of control strategies". Scientific Reports, *12*(1), 3805.

Chen, Q., Georganas, N. D., Petriu, E. M. 2007. "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features". 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007, 1–6.

Cronin, S., Doherty, G. 2018. "Touchless computer interfaces in hospitals: A review". Health Informatics Journal, *25*(4), 1325–1342.

Devineau, G., Moutarde, F., Xi, W., Yang, J. 2018. "Deep Learning for Hand Gesture Recognition on Skeletal Data". 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 106–113.

Fong, S., Zhuang, Y., & Fister, I. (2013). A biometric authentication model using hand gesture images. *Biomedical engineering online*, *12*(1), 1-18.

Gonzalo, P.-J., Juan, A. H.-T. 2015. "Control of home devices based on hand gestures". Içinde *2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)* (ss. 510–514).

Gupta, P., & Gupta, P. (2018). Multibiometric authentication system using slap fingerprints, palm dorsal vein, and hand geometry. *IEEE Transactions on Industrial Electronics*, *65*(12), 9777-9784. (Gupta, 2018)

Gulliksen, J. 2017. "Institutionalizing human-computer interaction for global health". Global Health Action, *10*(sup3), 1344003.

Gür, D., Schäfer, N., Kupnik, M., Beckerle, P. 2020. "A Human–Computer Interface Replacing Mouse and Keyboard for Individuals with Limited Upper Limb Mobility". Multimodal Technologies and Interaction, *4*(4).

Habili, N., Lim, C. C., Moini, A. 2004. "Segmentation of the face and hands in sign language video sequences using color and motion cues". IEEE Transactions on Circuits and Systems for Video Technology, *14*(8), 1086–1097.

Han, C. C., Cheng, H. L., Lin, C. L., & Fan, K. C. (2003). Personal authentication using palm- print features. *Pattern recognition*, *36*(2), 371-381.

Hasan, H. S., Kareem, S. A. 2012. "Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey". Içinde *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)* (ss. 55–60).

Hübler, A., Hansen, C., Beuing, O., Skalej, M., Preim, B. 2014. "Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining". Içinde *CURAC*.

Jaimes, A., Sebe, N. 2007. "Multimodal human–computer interaction: A survey". Computer Vision and Image Understanding, *108*(1), 116–134.

Johnson, T. R., Thimbleby, H., Killoran, P., Diaz-Garelli, J. F. 2015. "Human Computer Interaction in Medical Devices". Içinde V. L. Patel, T. G. Kannampallil, & D. R. Kaufman (Ed.), *Cognitive Informatics for Biomedicine: Human Computer Interaction in Healthcare* (ss. 163–187). Cham: Springer International Publishing.

Kulkarni, V. 2010. "Appearance Based Recognition of American Sign Language Using Gesture Segmentation".

Kumar, S., & Walia, E. (2011). Analysis of various biometric techniques. *International Journal of Computer Science and Information Technologies*, *2*(4), 1595-1597.

Lamberti, L., Camastra, F. 2011. "Real-Time Hand Gesture Recognition Using a Color Glove". Içinde *ICIAP*.

Li, J., Deng, L., Gong, Y., Haeb-Umbach, R. 2014. "An Overview of Noise-Robust Automatic Speech Recognition". IEEE/ACM Trans. Audio, Speech and Lang. Proc., *22*(4), 745–777.

Maricchiolo, F., De Dominicis, S., Ganucci Cancellieri, U., Di Conza, A., Gnisci, A., Bonaiuto, M. 2014. "109. Co-speech gestures: Structures and functions". Içinde C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Ed.) (ss. 1461–1473). De Gruyter Mouton.

May, J. 2001. "Human–Computer Interaction". Içinde N. J. Smelser & P. B. Baltes (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (ss. 7031–7035). Oxford: Pergamon.

Mitra, S., Acharya, T. 2007. "Gesture Recognition: A Survey". Trans. Sys. Man Cyber Part C, *37*(3), 311–324.

Molina, J., Pajuelo, J. A., Mart\'{\i}nez, J. M. 2017. "Real-Time Motion-Based Hand Gestures Recognition from Time-of-Flight Video". J. Signal Process. Syst., *86*(1), 17–25.

Nai-Arun, N., Sittidech, P. 2014. "Ensemble Learning Model for Diabetes Classification". Advanced Materials Research, *931–932*, 1427–1431.

Oikonomidis, I., Kyriazis, N., Argyros, A. A. 2012. "Tracking the articulated motion of two strongly interacting hands". 2012 IEEE Conference on Computer Vision and Pattern Recognition, 1862–1869.

Oudah, M., Al-Naji, A., Chahl, J. 2020. "Hand Gesture Recognition Based on Computer

Vision: A Review of Techniques". Journal of Imaging, *6*(8).
Paton, C., Kushniruk, A. W., Borycki, E. M., English, M., Warren, J. 2021. "Improving the Usability and Safety of Digital Health Systems: The Role of Predictive Human-Computer Interaction Modeling". J Med Internet Res, *23*(5), e25281.
Pelikan, H. R. M., Cheatle, A., Jung, M. F., Jackson, S. J. 2018. "Operating at a Distance - How a Teleoperated Surgical Robot Reconfigures Teamwork in the Operating Room". Proc. ACM Hum.-Comput. Interact., *2*(CSCW).
Pham, K., Kim, D., Park, S., Choi, H. 2021. "Ensemble learning-based classification models for slope stability analysis". CATENA, *196*, 104886.
Pisharady, P. K., Saerbeck, M. 2015. "Recent methods and databases in vision-based hand gesture recognition: A review". Computer Vision and Image Understanding, *141*, 152–165.
Sabhanayagam, T., Venkatesan, V. P., & Senthamaraikannan, K. (2018). A comprehensive survey on various biometric systems. *International Journal of Applied Engineerin Research*, *13*(5), 2276-2297.
Sarma, D., Bhuyan, M. K. 2021. "Methods, Databases and Recent Advancement of Vision-Based Hand Gesture Recognition for HCI Systems: A Review". SN Computer Science, *2*(6), 436.
Schultz, M., Gill, J., Zubairi, S., Huber, R., Gordin, F. 2003. "Bacterial Contamination of Computer Keyboards in a Teaching Hospital". Infection Control & Hospital Epidemiology, *24*(4), 302–303.
Sigal, L., Sclaroff, S., Athitsos, V. 2004. "Skin color-based video segmentation under time-varying illumination". IEEE Transactions on Pattern Analysis and Machine Intelligence, *26*(7), 862–877.
Song, W., Kang, W., Wang, L., Lin, Z., Gan, M. 2022. "Video Understanding-Based Random Hand Gesture Authentication". IEEE Transactions on Biometrics, Behavior, and Identity Science, *4*(4), 453–470.
Terui, H., Hosobe, H. 2022. "A Hand Gesture-Based Authentication Method that Makes Forgery Difficult". Içinde A. Moallem (Ed.), *HCI for Cybersecurity, Privacy and Trust* (ss. 268–279). Cham: Springer International Publishing.
Tu, Y.-J., Kao, C.-C., Lin, H.-Y., Chang, C.-C. 2015. "Face and Gesture Based Human Computer Interaction". International Journal of Signal Processing, Image Processing and Pattern Recognition, *8*(9), 219–228.
Utsumi, A., Tetsutani, N., Igi, S. 2002. "Hand Detection and Tracking Using Pixel Value Distribution Model for Multiple-Camera-Based Gesture Interactions". Içinde *Proceedings of the IEEE Workshop on Knowledge Media Networking* (s. 31). USA: IEEE Computer Society.
Varchol, P., & Levicky, D. (2007). Using of hand geometry in biometric security systems. *Radioengineering-Prague-*, *16*(4), 82.
Vendruscolo, P., Martelli, S. 2001. "Interfaces for computer and robot assisted surgical systems". Information and Software Technology, *43*(2), 87–96.
Verma, B., Hassan, S. Z. 2011. "Hybrid ensemble approach for classification". Applied

Intelligence, *34*(2), 258–278.
Wachs, J. P., Kölsch, M., Stern, H., Edan, Y. 2011. "Vision-Based Hand-Gesture Applications". Commun. ACM, *54*(2), 60–71.
Wang, R. Y., Popović, J. 2009. "Real-Time Hand-Tracking with a Color Glove". ACM Trans. Graph., *28*(3).
Yang, W., Chen, Z., Huang, J., Wang, L., Kang, W. 2021. "LFMB-3DFB: A Large-scale Finger Multi-Biometric Database and Benchmark for 3D Finger Biometrics". Içinde *2021 IEEE International Joint Conference on Biometrics (IJCB)* (ss. 1–8).
Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., Grundmann, M. 2020. "MediaPipe Hands: On-device Real-time Hand Tracking". CoRR, *abs/2006.10214*. Tarihinde adresinden erişildi https://arxiv.org/abs/2006.10214

14. PROJECT ACTIVITIES AND WORK PLAN

Table 5: *The Work-Activity Plan for Project 1*

Work and Activity Project 1	Responsible Group Member	Timeline													
		1. week	2. week	3. week	4. week	5. week	6. week	7. week	8. week	9. week	10. week	11. week	12. week	13. week	14. week
1. Literature Review	Osman, Yunus, İlknur														
2. Hand Gesture Standardization	Osman, Yunus, İlknur														
3.Video Streaming Dataset Manipulation	Osman, Yunus, İlknur														
4.Streaming Data Retrieval & Preparation	Osman, Yunus,														
5.Construction of Classification Models	Osman, Yunus,														
6.Model Evaluation and Tuning	Osman, Yunus,														
7.Authentication Based on Feature Scaling	Osman, İlknur														
8.Device Management with Hand Gestures	Osman, Yunus, İlknur														

14.1. LIST OF WORK PACKAGES

***Table 6:** Detailed Definition of Work and Activity*

WP No	Detailed Definition of Work and Activity
1	The Literature Reviewed deeply. All articles including hand gesture recognition systems reviewed. In this step, also firstly hardware applications were searched because of understanding the insufficiency of it in healthcare systems. Then, each searching subject was focused on current software applications. The hand gestures used in the control of systems were noted down.
2	Hand Gesture Standardization was provided by comparing collected hand gesture data to each other. Based on the hand gesture, the frequency of it in its specific files such as home appliances was determined. According to their frequencies rate, the common hand gestures were detected. Also, the new hand gestures which have not been used before were added to the possible hand gestures which was used in this project.
3	Video streaming dataset manipulation included the dataset creation (video recording) and its pre-processing steps. The hand gestures of different people were recorded by using different cameras. Then, in order to make easier the feature extraction step, the videos were pre-processed by applying resizing, rescaling, data augmentation techniques (rotation, flipping, and cropping), removing unnecessary frames and so on.
4	Streaming data retrieval and preparation were expected to be performed by MediaPipe. In order to train a model (the ensemble learning), the models' input included the specific values. For this, the keypoints were extracted with the help of the MediaPipe library to represent hand gestures. In terms of classification, it appeared that using only the keypoints might be sufficient for predicting hand gestures. However, the contour (for authentication) and optical flow information were also extracted and preprocessed beforehand. The numerical data was initially extracted on a frame-by-frame basis, focusing on keypoint-specific information. Subsequently, the relationship between frame-keypoint distances was combined, and this relationship was converted into a video-hand gesture feature that encompassed both frame-based and keypoint-specific information.
5	It was thought that the efficiency of the classification algorithms might have been changed according to the input data. Therefore, there were expected to try several classification models. The deep learning and machine learning models were combined in the scope of ensemble learning model. According to the requirement of input data, the architecture of the ensemble model was opened to add new model or removed models from developed architecture.
6	In model evaluation, each DL and ML algorithm were trained and tested on the sample datasets. After that the hybrid ensemble learning model was trained and tested on the

	same sample set. Their accuracy, precision, F1-Score, AUC, fold number in cross validation were calculated. After that the model's had been compared each other in terms of their evaluation metrics. The best five model among ML and DL models were selected and tuning of ensemble learning model was executed.
7	For the authentication, it was expected to use the contours and keypoint information of hand gesture. Since those both released a hand specific information, the person whose hand's information was taken was given an authority and could be verified easily. For the contour detection, thresholding methods were used and then, combined with keypoints to provide authentication of technical staff in the hospital environment.
8	The developed system was tested on the real-time control management system such as computer. By using webcam, the different operations were done by the people using hand gestures specified in the project.

Table 7: *Work package targets, their assessment, and the contribution of each work package to the overall project success.*

Work package	Target	Measurable outcome	Contribution to overall success (%)	Achieved (%)
1 & 2	Standardization of the hand gestures according to common use of them in the literature and meaning of commands	At least 15 common hand gestures from more than 50 articles determined.	**15**	**15**
3	Creating a pre-processed video streaming dataset by recording standardized hand gestures from different people	A file containing videos from >15 standardized gestures from >50 people	**10**	**5**
4	Preparing numerical dataset of hand gestures for gesture classification, authentication and device management	A dataset created by considering ● >5 video of hand gestures ● 21 keypoints' features for one hand.	**15**	**15**
5	Development of machine learning and deep learning-	At least 10 implemented models and establish	**20**	**20**

	based classification model with hybrid ensemble model	ensemble model with 5 best models		
6	Analysis of results taken from trained models' algorithms	Obtained values for: ● Accuracy > 60%, ● Precision > 65%, ● F1-Score: 0.5 - 0.8, ● AUC: 0.7 - 0.8, ● The cross validation divided into at least 5 folds	**10**	**10**
7	Assigning authentication to specific people and validation of the authority	At least 1 of the 3 hands correctly identified and authorized.	**20**	**20**
8	Managing simple functions of device after the hand gestures are correctly classified	The basic functions performed with WebCam	**10**	**10**
			Total:100	**95**

Table 8: *The work package distribution to project team members: Who works on which work package? Specify the percentage contributions.*

WORK PACKAGE DISTRIBUTION								
Project Member	**WP1**	**WP2**	**WP3**	**WP4**	**WP5**	**WP6**	**WP7**	**WP8**
İlknur Akçay	33.33	33.33	33.33	0	0	0	50	33.33
Osman Mersin	33.33	33.33	33.33	50	50	50	50	33.33
Yunus Emre GÜNDÜZ	33.33	33.33	33.33	50	50	50	0	33.33
Total	100%	100%	100%	100%	100%	100%	100%	100%

15. BUDGET

Table 9: *Proposed Budget in TL*

	ITEMS				
	PEOPLE	MACHINE-INSTRUMENT	MATERIALS	SERVICE	TRAVEL
IMU FUND	0	0	0	0	0
SPONSOR COMPANY FUND	0	0	0	0	0
TOTAL	0	0	0	0	0

16. CURRICULUM VITAE

Osman MERSİN

Biomedical Engineer

Kavacık Mah. Yayabeyi Sok. No.3-Beykoz/İSTANBUL
Phone: +90 (554) 596 61 94
osman.mersin@std.medipol.e
du.tr
osmanmersin53@gmail.com

PERSONAL INFORMATION

Date of Birth: 18/07/1999
Nationality: T.C.
Marital Status: Single
Gender: Male
Driver's License: B (2018), A2 (2019)

PROFILE

To produce solutions to neurodegenerative diseases and cancer, which are common in our age, with different perspectives in the light of developing technology and changing scientific knowledge.

EDUCATION

2022-2024	**Graduate School of Engineering & Natural Sciences, Medipol University** Master's Degree Student, Biomedical Engineering and Bioinformatics (100% Scholarship, 100% English Program,)
2020-2023	**School of Engineering and Natural Sciences, Medipol University** Undergraduate Student, Computer Engineering, Double Major (100% Scholarship, 100% English Program, 3.41 GPA)
2017 – 2022	**School of Engineering and Natural Sciences, Medipol University** Graduate, Biomedical Engineering (100% Scholarship, 100% English Program, 3.52 GPA)
2017-2018	**Medipol University Language School**

Preparatory Student, 100% Scholarship, (89.02/100)

2013-2017 **Makbule Hasan Uçar Anatolian Teacher High School**
High School Student (90.95/100), Aydın, Turkey

EXPERIENCE

09/2022- Present **Graduate Teaching Assistant (Istanbul Medipol University)**
Introduction to Biomedical Engineering (BME1110769), Cellular and Molecular Biology (BME2133870), Object Oriented Programming (COE2113250)

09/2022-Present **Graduate Researcher (Medipol SABITA)**
Istanbul Medipol University, Medipol Remer Regenerative and Restorative Medicine Research Center, Innovative Polymer Nanotherapeutic (IPONT) Lab.
(WEB: https://ipont.medipol.edu.tr/) / Polymer Science& Nanotechnology

09/2022-12/2022 **Peer Mentor, GIP Medical Technology Innovation Course, Drexel University**
Mentoring teams under the scope of Global Classroom: Global Innovation Partnership Medical Technology Innovation course

06/2022- 10/2022 **Intern, Istanbul Medipol MEGA University Hospital, Radiology Department**
Fully automated medical image analysis

07/2022 -08/2022 **Intern (Teknokent İstanbul, Hyperion Advanced Technologies A.Ş.)**
Full Time/ R & D department, development of micro acoustic system for colon cancer detection with polymeric biomaterials and modeling & analyzing the system with machine learning algorithms

05/2021-09/2022 **Undergraduate Researcher (Medipol SABITA)- Volunteer**
Istanbul Medipol University, Medipol Remer Regenerative and Restorative Medicine Research Center, Innovative Polymer Nanotherapeutic (IPONT) Lab.
(WEB: https://ipont.medipol.edu.tr/) / Polymer Science& Nanotechnology
Under the supervision of Prof. Dr. Yasemin Yüksel Durmaz

03/2022 -06/2022 **Undergraduate Teaching Assistant (Istanbul Medipol University)**
Medical Biology (BME1210772)

10/2021 -02/2022 **Undergraduate Teaching Assistant (Istanbul Medipol University)**
Cellular and Molecular Biology (BME2133870)

01/2020-06/2022 **Undergraduate Researcher (Medipol SABITA)- Volunteer**

Istanbul Medipol University, Medipol Remer Regenerative and Restorative Medicine Research Center, KERMAN Lab. (WEB: http://bilalkermanlab.com) / Neuroscience, Under the supervision of Assoc. Prof. Bilal Ersen Kerman

08/2021-09/2021 **Intern (ASELSAN A.Ş.)**
Full-time / Transport, Security, Energy and Health Systems (UGES) Health Systems

09/2020-02/2021 **Undergraduate Teaching Assistant (Istanbul Medipol University)**
Full-time/General Physics 1 (BME1113160)

02/2020-05/2020 **Undergraduate Teaching Assistant (Istanbul Medipol University)**
Full-time/ General Physics 2 (BME1213230)

09/2019-01/2020 **Undergraduate Teaching Assistant (Istanbul Medipol University)**
Full-time/ General Physics 1 (BME1113160)

PROJECTS

06/2022-Present **TUBITAK BIDEB 2247-C Intern Researcher Scholarship Program**
“Development of siRNA Carrier “Smart” Ultrasound-active Nanoparticles and Investigation of Ultrasound Assistance to Gene Delivery**”**
Project ID: 118Z952
Department of Biomedical Engineering, Engineering Faculty, Erciyes University
Under the supervision of Assoc. Prof. Ömer Aydın

06/2022- 10/2022 **Istanbul Medipol MEGA University Hospital, Radiology Department**
“Development an algorithm and a software for prognosis of hypertensive cardiac diseases from radiological analysis**”**
Under the supervision of Prof. Dr. Cengiz Erol and Prof. Dr. Reda Alhajj

10/2021- 10/2022 **TUBITAK 2209-A Undergraduate Research Program**
“Development of Ultrasound Assisted Gene Delivery Agent”
Reference number: 1919B012101053
Under the supervision of Prof. Dr. Yasemin Yüksel Durmaz

08/2020-10/2021 **TUBITAK 2209-A Undergraduate Research Program**
“Analysis of Intracellular Cargo Traffic in an in vitro Pelizaeus-Merzbacher Disease Model”
Reference number: 1919B012000960

	Under the supervision of Assoc. Prof. Bilal Ersen Kerman
05/2017	**4006-TÜBİTAK Science Fairs Support Program** Eratosthenes Experiment (The measurement of circumference of our world)
03/2017	**International Eratosthenes Experiment and Competition** Group of the Photo Contest March 2017 in ODS Eratosthenes Experiment community
09/2016	**Erasmus/ Prague and Poland** Project Name: 'Beyond the Stars' Coordinator Country: Turkey Learning/ Teaching/ Training Activities
05/2015	**4006-TÜBİTAK Science Fairs Support Program** Design of Painter Robot

MEMBERSHIPS

07/2021	TMMOB-Chamber of Computer Engineers Student Membership
05/2019	IEEE Engineering in Medicine and Biology Society (EMBS) Membership
05/2019	IEEE Student Branch Membership
12/2017	EMO Foundation Membership (EMO-GENC/Chamber of Electrical and Biomedical Engineers)

CERTIFICATES

06/2022	Certificate of Animal Use in Experimental Research Qualification Score: 91, Confirmation Code/Date: 4894EC41XB/27.06.2022 Animal Research Local Ethics Committee, İstanbul Medipol University
03/2022	Medical Technology Innovation II Course, Drexel University
02/2022	Program Solving, Procter & Gamble (P&G) VIA pilot
09/2020	Andor Technology, Transmitted Light Microscopy-Lesson 2
09/2020	Andor Technology, The History of Microscopy-Lesson 1
05/2019	IEEE Medipol University Student Branch, Meducation Talks 4
06/2018	Pearson English Proficiency Certificate
05/2018	IEEE Medipol University Student Branch, Meducation Talks
12/2017	Medipol University Management and Economics Club, MECTALKS C-LEVEL

RESEARCH INTERESTS

Nanotechnology and Polymer Science, Biomaterials, Gene Therapy, Neurodegenerative diseases, Machine Learning, Development and Designing Medical Devices

COMPUTER SKILLS

Advanced: Python, MATLAB, JAVA, JavaScript, IMAGE/Fiji, Microsoft Excel, Adobe Photoshop

Good: C++, C Programming language, TensorFlow, Keras, Scikit-Learn, OpenCV, SPSS, Assembly Programming Language, HTML5, CSS3, 3D Slicer, Statistics

HOBBIES

Karate (Black Belt 1st Dan, Licensed Athlete), Turkish folk music instruments (Bağlama, Oud, Cümbüş- String instruments), mountain hiking, swimming and table tennis

LANGUAGES

English	**Reading:** Upper-intermediate,	**Writing:** Upper-intermediate,	**Speaking:** Intermediate
Deutch	**Reading:** Elementary,	**Writing:** Elementary,	**Speaking:** Elementary

YUNUS EMRE GÜNDÜZ

0545 940 79 72 yegunduz@st.medipol.edu.tr

BIOMEDICAL & COMPUTER ENGINEERING

WORK EXPERIENCE

ASELSAN | Intern
August 2021- September 2021
Assisting an engineer with her project about developing a medical device. Since the project is still ongoing, any detailed information about the project cannot be provided for confidentiality reasons.

Izmir Biomedicine & Genome Center | Intern
June 2021 - August 2021
Assisting a Ph.D. Student with her project., a code developed with Python for specific MDS result analysis.

Istanbul Medipol University
Undergrad Teaching Assistant/Physics I-II
Preparing and grading quizzes. Conducting prepared experiments in laboratory. Giving recitation classes to solve questions from the students.

Istanbul Medipol Uni. | Undergraduate Researcher
October 2020 - Present
'Advanced Computational Biophysics Lab for Designing Targeted and Safe Therapeutic Molecules Research Group'

Making a literature review about the subject chosen with Assist. Prof. Özge Şensoy & preparing a project.
Reading recent articles about computational biophysics and presenting it with the developments in our own project

PROFILE

I am an ambitious, biomedical engineer aiming to touch people lives in what I am going to do.
So, I want to work for Medical or Pharma companies to learn more and expand my knowledge where I can gain experience and better contribute to my purpose.
Also, as the world is tending to digitalization and robotics, I decided to have my second major in computer sciences which can contribute my purpose in a wider perspective and make me a more developing engineers in my journey in medical environment.

EDUCATION

Istanbul Medipol University
Engineering and Natural Sciences (Double Major)
2017 – Present

	GPA
Biomedical Engineering (100% English) Expected Graduation: July 2022	3.63
Computer Engineering (100% English) Expected Graduation: July 2023	3.65

References can be provided upon request

PROJECTS

TUBITAK 2209-A
"Identification of Small Therapeutic Molecules with the Potential to Inhibit Hsp70/BAG2 Interaction that Suppresses Apoptosis of Tumor Cells by Computational Methods"

INTERESTS

COMPUTER SKILLS

- WORD
- POWERPOINT
- EXCEL
- LtSPICE
- PYTHON
- MATLAB
- C/C++
- JAVA
- SPSS
- Arduino UNO

CERTIFICATES

Medipol University Management and Economics Club, MECTALKS C-LEVEL (2017)

Medipol Uni Student Branch, Meducation Talks (2019)

Pearson English Proficiency Certificate (2018)

İÜC, Kimya Mühendisliği Zirvesi 21 (2021)

Neuton AutoML– Machine Learning Made Simple (2021)

17. SUPPORT LETTERS (if any)

In the previous application, the project members were planning to apply for TUBITAK 2209-A. However, since each member has ongoing projects in the concept of TUBITAK 2209-A, they were not able to apply a new application.