**DETECTING CORRECT VESTIBULAR REHABILITATION EXERCISES THROUGH MACHINE LEARNING USING CONVOLUTIONAL NEURAL NETWORKS**

|  |  |  |
| --- | --- | --- |
| Prof. Dr. Kutluk ÖZGÜVEN  İstanbul Aydın University  [kutlukozguven@aydin.edu.tr](mailto:kutlukozguven@aydin.edu.tr) | Aydın TANRIVERDİ  İstanbul Aydın University  [aydintanriverdi@stu.aydin.edu.tr](mailto:aydintanriverdi@stu.aydin.edu.tr) | Görkem Genar AKKAYA  İstanbul Aydın University  [gorkemakkaya@stu.aydin.edu.tr](mailto:gorkemakkaya@stu.aydin.edu.tr) |
| Sümeyye Nur ÖZGENÇ  İstanbul Aydın University  [sumeyyeozgenc@stu.aydin.edu.tr](mailto:sumeyyeozgenc@stu.aydin.edu.tr) | Melek Başak ÖZKAN  İstanbul Aydın University  [mbasakozkan@stu.aydin.edu.tr](mailto:mbasakozkan@stu.aydin.edu.tr) | Buse HACIŞABANOĞLU  İstanbul Aydın University  busehacisabanoglu@stu.aydin.edu.tr |

**ABSTRACT**

In this project, it is aimed to help people who have problems with balance loss and cannot control themselves due to vestibular balance disorder to exercise at home without the need of physiotherapists and regain their health. Within the scope of the project, using convolutional neural networks, two pre-trained model movements are defined and compared with real-time camera images from the user. As a result of this comparison, feedback on both sound and text is received on the screen about accuracy or inaccuracy. Through this feedbacks, the user is aimed to make the exercises more accurate. The program is easy to use and to use this program, a computer with a camera is enough.

**Keywords:** Artificial intelligence, convolutional neural network, deep learning, motion recognize, vestibular rehabilitation exercises.

**EVRİŞİMLİ SİNİR AĞLARI KULLANILARAK MAKİNE ÖĞRENMESİ İLE DOĞRU VESTİBÜLER REHABİLİTASYON EGZERSİZLERİNİN BELİRLENMESİ**

|  |  |  |
| --- | --- | --- |
| Prof. Dr. Kutluk ÖZGÜVEN  İstanbul Aydın Üniversitesi  [kutlukozguven@aydin.edu.tr](mailto:kutlukozguven@aydin.edu.tr) | Aydın TANRIVERDİ  İstanbul Aydın Üniversitesi  [aydintanriverdi@stu.aydin.edu.tr](mailto:aydintanriverdi@stu.aydin.edu.tr) | Görkem Genar AKKAYA  İstanbul Aydın Üniversitesi  [gorkemakkaya@stu.aydin.edu.tr](mailto:gorkemakkaya@stu.aydin.edu.tr) |
| Sümeyye Nur ÖZGENÇ  İstanbul Aydın Üniversitesi  [sumeyyeozgenc@stu.aydin.edu.tr](mailto:sumeyyeozgenc@stu.aydin.edu.tr) | Melek Başak ÖZKAN  İstanbul Aydın Üniversitesi  [mbasakozkan@stu.aydin.edu.tr](mailto:mbasakozkan@stu.aydin.edu.tr) | Buse HACIŞABANOĞLU  İstanbul Aydın Üniversitesi  busehacisabanoglu@stu.aydin.edu.tr |

**ÖZET**

Bu projede, denge kaybı problemleri olan ve vestibüler denge bozukluğu nedeniyle kendilerini kontrol edemeyen kişilerin fizyoterapiste ihtiyaç duymadan evde egzersiz yapmalarına ve sağlıklarını yeniden kazanmalarına yardımcı olmak amaçlanmıştır. Proje kapsamında, evrişimli sinir ağları kullanılarak, önceden eğitilmiş iki model hareketi tanımlanmış ve kullanıcıdan alınan gerçek zamanlı kamera görüntüsü ile karşılaştırılmıştır. Bu karşılaştırmanın sonucunda, ekranda hem ses hem de metinle ilgili doğruluk veya yanlışlık hakkında geri bildirim alınır. Bu geri bildirimler sayesinde, kullanıcının egzersizleri daha doğru yapması hedeflenmiştir. Programın kullanımı kolaydır ve bu programı kullanmak için kameralı bir bilgisayar yeterlidir.

**Anahtar Kelimeler**: Yapay Zeka, Evrişimli Sinir Ağı, Derin Öğrenme, Hareket Tanıma, Vestibüler Rehabilitasyon Egzersizleri.

1. **INTRODUCTION**

Artificial intelligence has started to be included in many fields. In the health sector, artificial intelligence is used extensively. The aim of this program is to make technology useful for people's health. Human health is among our priorities and in order to improve the quality of human health, it is planned to help those who have balance problems known as a vestibular disorder in this project. There is a need for an appropriate program because people cannot control whether they are doing the right or wrong actions at home. So, we developed a project about it. As for the working logic of the project, after the program is started, motion analysis is performed, and feedback is received about its accuracy and inaccuracy. The distance to the camera is important. If you do not enter the camera angle or you do not catch the correct distance or you make the wrong move, you will see the wrong text on the screen. If you catch the right angle and make the right movements, you will see the correct text on the screen. As you can see, you must make these two conditions: the right angle that accurate camera vision and right movement.

1. **BACKGROUND**

In the Background section, information is given to explaining concepts, literature, research and discussing literature.

* 1. **Explaining Research**

This section contains the terminology, subject and some definitions related to the project.  
 **Artificial Intelligence:** The term “artificial intelligence” refers to a specific field of computer science that focuses on creating systems capable of gathering data and making decisions and/or solving problems.

**Deep Learning:** A subset of AI and Machine learning in which Neural networks are “layered”, combined with plenty of computing power, and given a large measure of training data to create extremely powerful learning models capable of processing data in new and exciting ways in some areas, e.g. advancing the field of computer vision.

**Convolutional Neural Network:** The name “convolutional neural network” indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.

**Audiology and Vestibular System:** The balance is achieved by interpreting the information received through the vestibular system, visual and proprioceptive system in the upper centers. With the adaptation excellence of the vestibular system that is not present in other sensory systems, it starts to work in a short time for the imbalance experienced especially in patients after peripheral vestibular losses and tries to create a situation where the patient will be less disturbed by the present problem. Thus, the brain tries to compensate for the patient’s imbalances. However, this adaptation cannot occur in patients with some peripheral vestibular loss as a result of unconscious movement avoidance strategies, long-term vestibular system suppressor drug use, and other unknown mechanisms.

The information obtained from all our sensory systems since birth is gradually collected in a center. This complex system, which is the center of integration and regulation of many senses; visual, proprioceptive and vestibular stimuli are under constant control; movement is affected by physiological factors such as learning or change. At subconscious levels; there is a continuous comparison with previously stored sensory models and appropriate muscle groups are stimulated to maintain balance. The creation of these new models is constantly present in everyday life and is called familiarization. (Glasscock III M.E., Cueva R.A., Thendinger B.A., 1990)

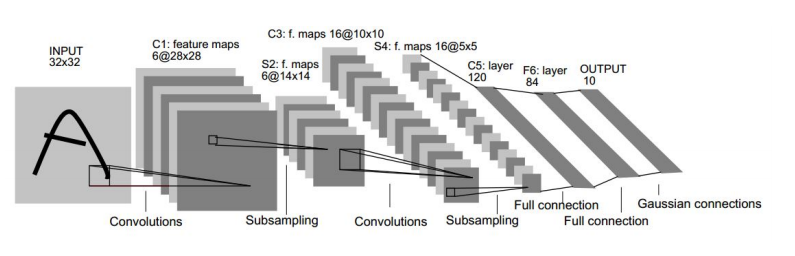
The balance system in humans and other organisms has three main functions: 1. Control of spinal reflexes, which allows the correction of body position and muscle activity to maintain the posture of the upright posture, 2. The control required to keep the image of the eye moving in the retinal layer of the eye during movement of the head, 3. Perception of motion and spatial orientation. Thus, the balance in the organism is maintained; vision, proprioceptive system, and vestibular system activity. In the petrosis part of the temporal bone in the inner ear, both the sensation of the sound and the changes in the condition of the head and trunk, receiving sensations; there are specific sensory cells called the vestibular system. (Guyton A.C., 1991)

Failure to achieve or posture of the body posture leads to impaired balance. Vertigo; It is defined as the subjective feeling of the deterioration in the relationship of the individual with his environment. Falling, turning or being pushed. Dizziness; gaps, light-headedness, light sleepiness. (Halmagyi G.M., Cremer P.D., 2000) The difference between the two terms is not clear, but one is due to peripheral disorders and the other is more evocative of central pathologies.  
 At the basis of vestibular rehabilitation, there are mechanisms associated with the neural plasticity of the central nervous system. Increasing visual stabilization with head movements, improving vestibular-visual interaction, increasing static and dynamic postural stabilization under conditions producing contradictory sensory information and reducing individual sensitivity to head movements are provided. In summary, the aim of vestibular rehabilitation is to activate the adaptive and compensatory mechanisms in the human brain. (Meli A, Zimatore G, Badaracco C, 2006)

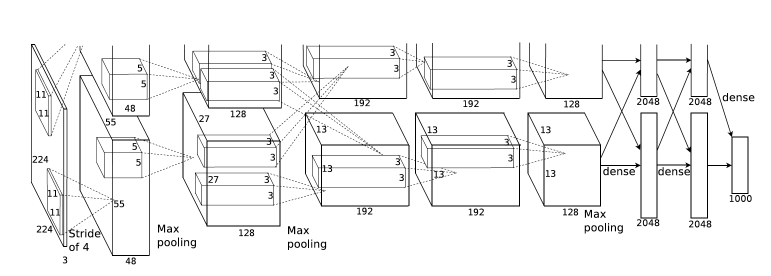
* 1. **Literature Research**

There is a lot of research about artificial intelligence and it is continuing to be done. In general, research is about the classification and training of convolutional neural networks. There are convolutional layers designed in various specifications. Some of those are LeNet-5, AlexNet, ZFNet.

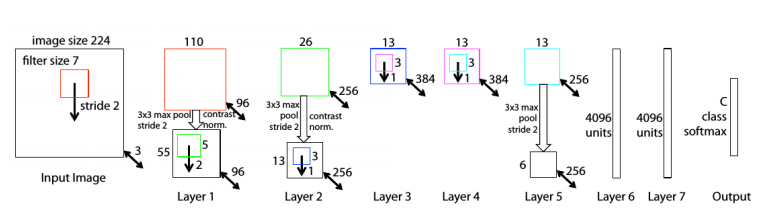
LeNet-5 as shown in figure one has seven weighted (trainable) layers. Among them, three (C1, C3, C5) convolutional layers, two (S2, S4) average pooling layers, one (F6) fully connected layer and one output layer. Sigmoid function was used to include nonlinearity before a pooling operation. The output layer used Euclidean Radial Basis Function units (RBF) to classify ten digits. (Sultana et al., 2018, p. 2)

Figure 1: Architecture of LeNet-5 (Sultana et al., 2018, p. 2)

The architecture of AlexNet is the same as LeNet-5 but much bigger as shown in figure two. It is made up of eight trainable layers. Among them, five convolutional layers (Conv layer) and three fully connected layers are there. Using rectified linear unit (ReLU) non-linearity after convolutional and FC layers helped their model to be trained faster than similar networks with tanh units. (Sultana et al., 2018, p. 3)

Figure 2: Architecture of AlexNet (Sultana et al., 2018, p. 3)

As shown figure three, The Architecture of AlexNet and ZFNet is almost similar except that the designers have reduced 1st layer filter size to 7×7 instead of 11×11 and used stride two convolutional layers in both first and second layers to retain more information in those layer's features. (Sultana et al., 2018, p. 3)

Figure 3: Architecture of ZFNet (Sultana et al., 2018, p. 3)

As seen in the figurines, different types of convolutional neural networks were used in each algorithm. Thanks to these different models, it has been seen that different convolutional neural network designs have reached different estimation values.

* 1. **Discussing Literature**

That has been shown in the research we did; the convolutional neural network layer number is very important for deep learning. If the data loaded into the model is analyzed in multiple layers, the model's estimate is so close to reality. However, a multi-layered model runs slower and requires more resources. In deep learning, data augmentation is as important as multilayer neural networks. Augmentation is achieved by performing some operations on the data loaded into the model. This allows the model to analyze the data contained in the dataset in very different ways. In this method, the model produces a better estimation. Below is an example of data augmentation.



Figure 4: Data Augmentation Example (Amaratunga, 2018)

1. **METHODOLOGY**

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analysing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing.

CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. CNNs take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme.

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is commonly a RELU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution. The final convolution, in turn, often involves backpropagation in order to more accurately weight the end product. Though the layers are colloquially referred to as convolutions, this is only by convention. Mathematically, it is technically a sliding dot product or cross-correlation. This has significance for the indices in the matrix, in that it affects how weight is determined at a specific index point.  
 **Convolutional:** When programming a CNN, the input is a tensor with shape (number of images) x (image width) x (image height) x (image depth). Then after passing through a convolutional layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map width) x (feature map height) x (feature map channels). A convolutional layer within a neural network should have the following attributes:   
 Convolutional kernels defined by a width and height (hyper-parameters).  
 The number of input channels and output channels (hyper-parameter).  
 The depth of the Convolution filter (the input channels) must be equal to the number channels (depth) of the input feature map.  
Convolutional layers convolve the input and pass its result to the next layer. Each convolutional neuron processes data only for its receptive field.  
 **Pooling:** Pooling layers reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, typically 2 x 2. Global pooling acts on all the neurons of the convolutional layer and may compute a max or an average. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Average pooling uses the average value from each of a cluster of neurons at the prior layer. **Fully connected:** Fully connected layers connect every neuron in one layer to every neuron in another layer. The flattened matrix goes through a fully connected layer to classify the images. **Receptive field:** In a fully connected layer, each neuron receives input from every element of the previous layer. In a convolutional layer, neurons receive input from only a restricted subarea of the previous layer. So, in a fully connected layer, the receptive field is the entire previous layer.  
 **Weights:** Each neuron in a neural network computes an output value by applying a specific function to the input values coming from the receptive field in the previous layer. The function that is applied to the input values is determined by a vector of weights and a bias (typically real numbers). Learning, in a neural network, progresses by making iterative adjustments to these biases and weights. The vector of weights and the bias are called filters and represent particular features of the input. A distinguishing feature of CNNs is that many neurons can share the same filter.  
In a CNN architecture, a few distinct types of layers are commonly used. **Convolutional layer:** The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the entries of the filter and the input and producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input. Stacking the activation maps for all filters along the depth dimension forms the full output volume of the convolution layer. Every entry in the output volume can thus also be interpreted as an output of a neuron that looks at a small region in the input and shares parameters with neurons in the same activation map.  
 **Pooling layer:** It partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. Intuitively, the exact location of a feature is less important than its rough location relative to other features. This is the idea behind the use of pooling in convolutional neural networks. The pooling layer serves to progressively reduce the spatial size of the representation, to reduce the number of parameters, memory footprint and amount of computation in the network, and hence to also control overfitting. The most common form is a pooling layer with filters of size 2×2 applied with a stride of 2 down samples at every depth slice in the input by 2 along both width and height, discarding 75% of the activations. In this case, every max operation is over 4 numbers. The depth dimension remains unchanged. Pooling is an important component of convolutional neural networks for object detection based on Fast R-CNN architecture.   
 **ReLU layer**: ReLU is the abbreviation of rectified linear unit. It effectively removes negative values from an activation map by setting them to zero. It increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer. Other functions are also used to increase nonlinearity, ReLU is often preferred to other functions because it trains the neural network several times faster (Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012).  
 **Fully connected layer:** Finally, after several convolutional and max pooling layers, the high-level reasoning in the neural network is done via fully connected layers. Neurons in a fully connected layer have connections to all activations in the previous layer, as seen in regular (non-convolutional) artificial neural networks. Their activations can thus be computed as an affine transformation, with matrix multiplication followed by a bias offset (vector addition of a learned or fixed bias term).   
 **Loss layer:** Loss function and Loss functions for classification the "loss layer" specifies how training penalizes the deviation between the predicted (output) and true labels and is normally the final layer of a neural network. Various loss functions appropriate for different tasks may be used.

* 1. **Hypotheses**

In the development of this project, it has been hypothesized that artificial intelligence may have positive effects on human health. The main goal is to reduce the time people reach their personal health by using this program. It is thought that this time will decrease thanks to this program controlling the exercises and directing the user. While starting this project, the research has started with the idea "Using the convolutional neural network to perform the correct movements used for treatment purposes in vestibular movement disorder”.

* 1. **Methods of Proof**

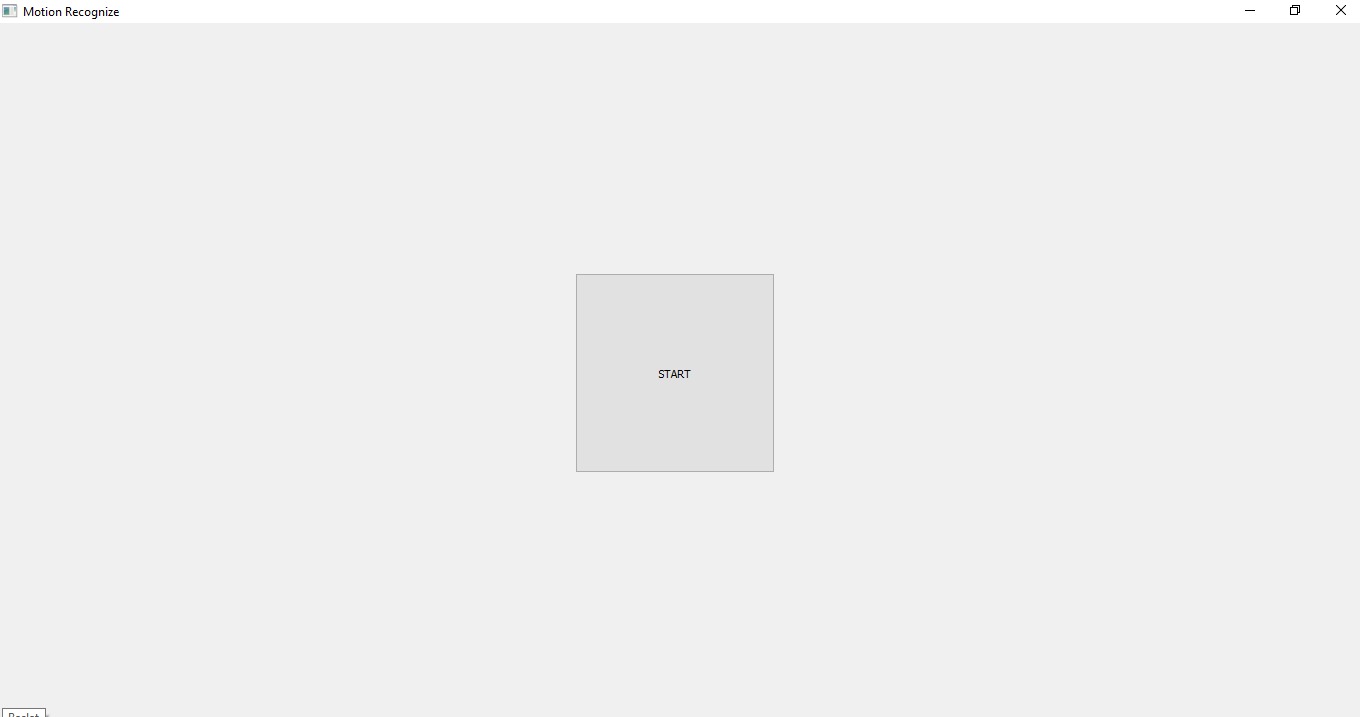
As a result of the experiment carried out on four people, the pictures used for 720 training data which is given from different people for every movement and the pictures used for 120 test data for every movement; after the 50 attempts, result was obtained 79,15% success on the accuracy of the movements. To prove the success of this hypothesis, the program needs to be tested on real patients. But, since the project, whose development process has been completed, has not yet been tested on real patients and we do not have statistical data.

* 1. **Research Environment**

While doing this project, a laptop with a camera (640\*480 Pixel), 4GB RAM, i5 4210n Processor, 2.6GHz Processor speed, Win10 x64 Operating System, Python 3.7 with PyCharm editor was used. Keras (v2.3.1), glob (v0.7), numpy (v1.18.1), cv2 (v4.1.2.30), os, playsound (v1.2.2), PyQt5 (v5.14.1), time, sys libraries were used while creating the project.  
**Keras (v2.3.1):** Keras is an open-source neural-network library written in Python. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing Deep Neural Network code. The project provided training on multiple images, creating a model, and layers, known as the building blocks used in CNN, were modelled through this library.  
**Glob (v0.7)**: Provides navigation between files. In the project, the snapshot taken from the camera is printed in the folder as a picture and compared with the trained module.  
**Numpy (v1.17.4):** is a Python package which stands for ‘Numerical Python’. It is the core library for scientific computing, which contains a powerful n-dimensional array object. Numpy array is a powerful N-dimensional array object which is in the form of rows and columns. We can initialize numpy arrays from nested Python lists and access it elements.  
**Cv2 (v4.1.2.30)**: OpenCV is used for all sorts of image and video analysis, like facial recognition and detection, license plate reading, image editing, advanced robotic vision, optical character recognition, and a whole lot more. In Python program called it cv2 instead of OpenCV.  
**Os**: Used in folder and file operations running on multiple operating systems.  
**Playsound (v1.2.2)**: Used for sound operations.  
**PyQt5 (v5.14.1)**: It uses to design for user interface.  
**Time**: Time uses to count down in the program to start to preparation and take correct movement position.  
**Sys:** This module allows us to get information about the version of Python we use and to do various operations with the version of Python we use.

1. **STUDY RESULTS**

The program is opened with a form and in this form there is a start button (Figure 5).

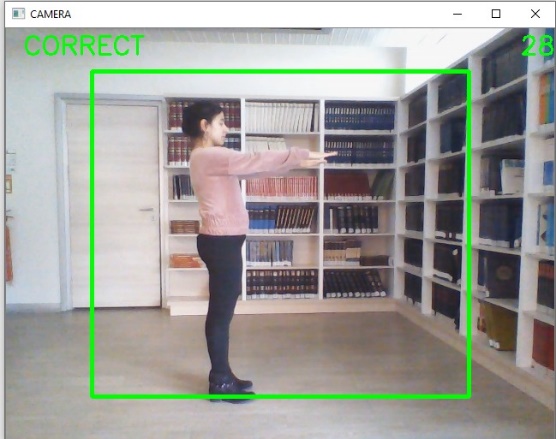
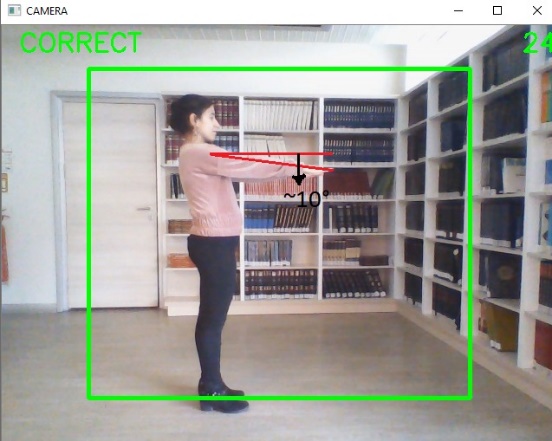
  
Figure 5: Inception Screen

After the start button is pressed, the images taken from the webcam are displayed on the screen. First, it is given a 15 second preparation time to user to move to the correct exercise position. In the next 5 seconds, the program will make some adjustments. During that time, the user must wait in the correct position. Any instant movement change or even a slight change in the background varies in terms of finding the accuracy of the movement. If this happens, the program will start towards giving wrong results (Figure 6).

Figure 6: Preparation and Position Time

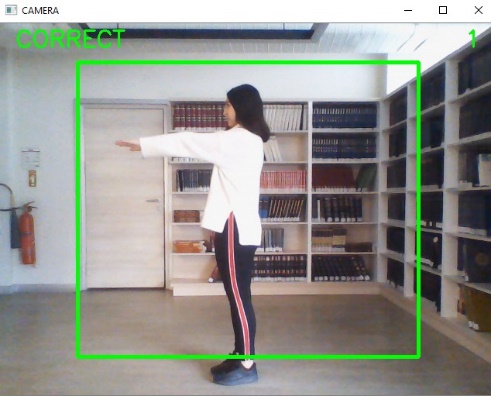
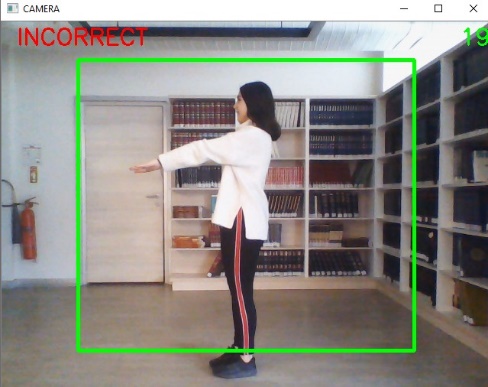
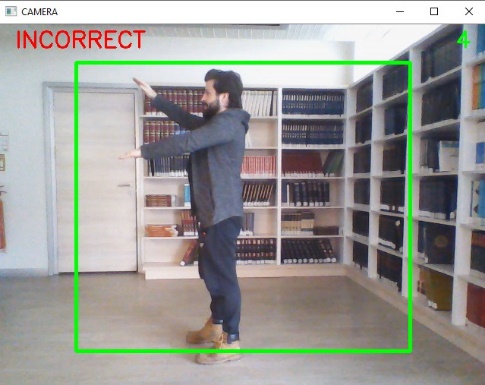
After adjustments are made, the user continues to exercise for 30 seconds for each movement. While the exercise is in progress, the user is informed whether the movement is correct and on the screen there will be writing written and heard a voice. As you can see in figure 7, accuracy rate up to 10 degrees is given in 2A movement. If any more angles increase or decrease, the movement will go to the wrong state while it is correct rate and give a warning accordingly. (Figure 7).

   
Figure 7: 2A Movement

As you can see on the result screen, the accuracy rate changes as it can be seen in Figure 8, since the motion change close to the end of the program is made and the wrong motion position is changed. Thus, the values ​​of the correct movements performed in 30 seconds and the wrong movements performed are taken as a result and displayed as "how many percent correct movements were performed" and "how many percent wrong movements were performed" on the screen (Figure 8).

  
Figure 8: 2A Movement Result Screen

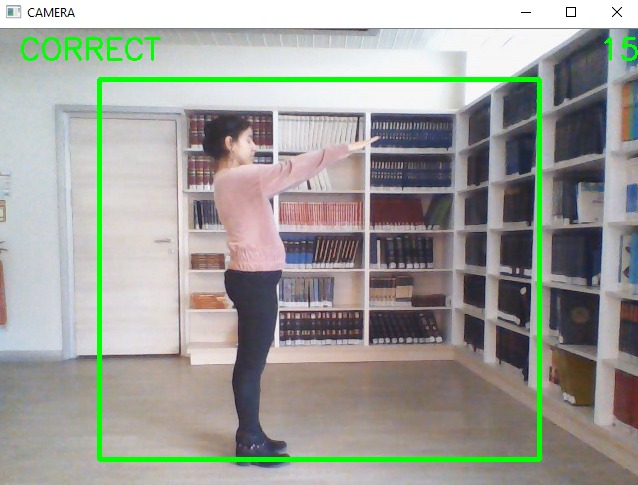
As a result of the trial performed by different people, if the movement is done correctly, it is written correctly and an audible warning is displayed on the screen (Figure 9). If the position of the arms exceeds the angle of 10 degrees, it will be Incorrect written on the screen as it will be seen in Figure 10 and it gives an audible warning. If you want to go out of the correct position and want to be take an image in a different position, the result will be false and this will affect the accuracy and inaccuracy rate of the movement. As can be seen in figure 10, two of the arms were recorded incorrectly as a result of being of different heights and lows. (Figure 10).

    
Figure 9: 2A Correct Position Figure 10: Degree Error and Different Position

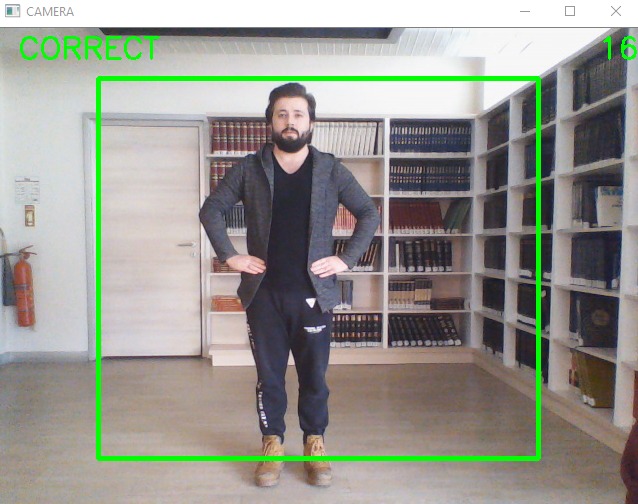
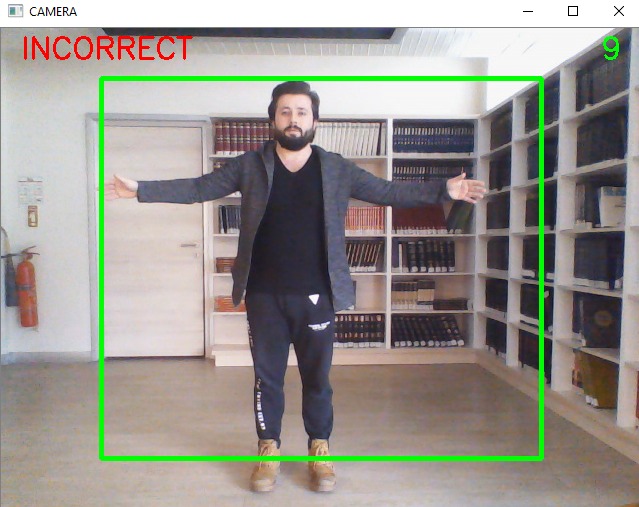
As can be seen on the result screen, the percentage of the position was rotated based on the results of the movement that the person made right or wrong. Since the person entered different positions within 30 seconds during the program, the error rate increased and 60.606% was found, while the accuracy rate was found 39.3939% and showed on the screen. (Figure 11).

  
Figure 11: 2A Movement Result Screen

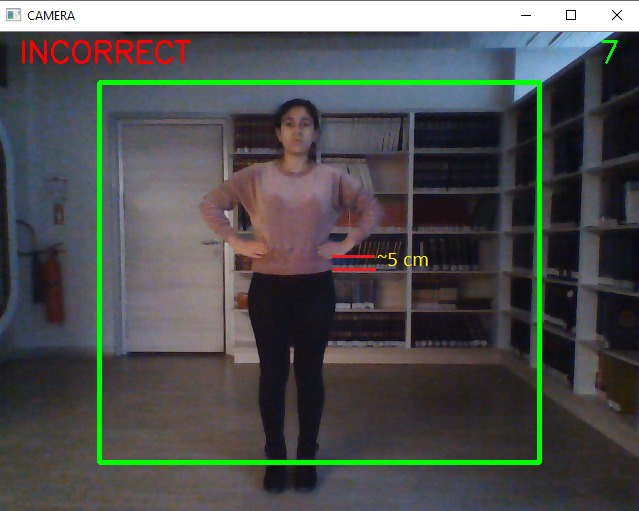
In the program, when the angle of 10 degrees is exceeded, in the trained model, in some cases, it may result in the screen correctly where it needs to be written incorrectly, or it may be incorrectly written in places where it should be written correctly and beeps accordingly (Figure 12).

  
Figure 12: Wrong Result

In the second move, the arms should be placed at the waist at an angle of 45 degrees which position known as 1B movement. As you can see in the pictures, in the first figure, the position is displayed correctly on the screen because it is done correctly, but in the second picture, it is written on the screen incorrectly because the arms are opened and staying a different position (Figure 13).

   
Figure 13: 1B Correct and Incorrect Position

As shown in the picture, the program has a few cm tolerance to almost 5 cm. The result of the movement made after 5 cm is written incorrectly on the screen (Figure 14).

  
Figure 14: 1B Wrong Position

1. **CONCLUSION**

As a result, it was determined whether the exercises were correct or not by analysing the exercises. A result screen is displayed so that the user can understand this image analysis. Therefore, the user is intended to see and understand the results easily. With these results, the user makes his exercises more careful. This indicates that the program has reached its target. The artificial intelligence model in the project was trained for two different exercises. As a result, 79.15% success was achieved (Appendix 3). The following days, it can be provided to recognize more vestibular rehabilitation exercises by the studied model. Thanks to that, it can be gotten much more efficient.

1. **REFERENCES**

Aggrawal J. K., Ryoo M. S., (2011), Human Activity Analysis: A review, The University of Texas, <http://cvrc.ece.utexas.edu/mryoo/papers/review_ryoo_hdr.pdf>, 01.02.2020

Wei L., Shah S.K., (2017), Human Activity Recognition Deep Neural Network with Contextual Information, The University of Houston, <http://scitepress.org/papers/2017/60995/60995.pdf>, 01.02.2020

Jiao L., Zhang F., Liu F., Yang S., Li L., Feng Z., Qu R., (2019), Deep Learning-Based Object Detection. In IEEE Access Volume-7, 128837-128868, <https://ieeexplore.ieee.org/document/8825470>, 01.02.2020

Shin J., Kim S., Kang S., Lee S., Paik J., Abidi B., Abidi M., (2005), Optical Flow-Based Real-Time Object Tracking Using Non-Prior Training Active Feature Model. Ching-Ang University, <http://sciencedirect.com/science/article/pii/S1077201405000215>, 01.02.2020

Sultana F., Sufian A., Dutta P., (2018), Advancements in Image Classification Using Convolutional Neural Network, The University of Gour Banga, <http://arxiv.org/pdf/1905.03288.pdf>, 01.02.2020

Giacinto G., (2000), Design of Effective Neural Network Ensembles for Image Classification Purposes. The University of Cagliari, <https://www.sciencedirect.com/science/article/pii/S0262885601000452>, 01.02.2020

Krishna M. M., Neelima M., Harshali M., Rao M. V. G., (2018), Image Classification Using Deep Learning, [http://researchgate.net/publication/325116934\_Image\_classification\_using\_Deep\_learning. 01.02.2020](http://researchgate.net/publication/325116934_Image_classification_using_Deep_learning.%2001.02.2020)

Doyle D. D., Jennings A. L., Black J. T., (2013), Optical Flow Background Estimation for Real-Time Pan/Tilt Camera Object Tracking, Air Force Institute of Technology, <http://sciencedirect.com/science/article/pii/S0263224113005241>, 01.02.2020.

Gangwar N., (2017), Image Classification Using Keras, viewed 09 Jan 2020, <<https://www.geeksforgeeks.org/python-image-classification-using-keras/>>

Mordvintsev A. & Abid K., (2013), Getting Started with Videos, viewed 09 Jan 2020, < [https://opencv-python-tutroals.readthedocs.io/en/latest/py\_tutorials/py\_gui/py\_video\_display/py\_video\_display.html#display-video](https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_gui/py_video_display/py_video_display.html%23display-video)>

Olafenwa M. & Olafenwa J., (2019), Prediction Classes, viewed 09 Jan 2020, <<https://imageai.readthedocs.io/en/latest/prediction/index.html>>

Rosebrock A., (2019), Human Activity Recognition with OpenCV and Deep Learning, viewed 09 Jan 2020, <<https://www.pyimagesearch.com/2019/11/25/human-activity-recognition-with-opencv-and-deep-learning/>>

Shaikh F., (2018), Understanding and Building an Object Detection Model from Scratch in Python, viewed 09 Jan 2020, <<https://www.analyticsvidhya.com/blog/2018/06/understanding-building-object-detection-model-python/>>

Amaratunga, (2018), Using Data Augmentations in Keras, viewed 09 Jan 2020, < <https://www.codesofinterest.com/2018/02/using-data-augmentations-in-keras.html>>

Glasscock III M.E., Cueva R.A., Thendinger B.A., (1990), Handbook of Vertigo, Raven Press, New York.

Guyton A.C., (1991), Textbook of Medical Physiology. 8th Edition, W.B. Saunders Company, Philadelphia, p: 609-619.

Halmagyi G.M., Cremer P.D., (2000), Assessment and Treatment of Dizziness. Journal of Neurology Neurosurgery and Psychiatry, 68(2): 129-34

Meli A, Zimatore G, Badaracco C, et al., (2006), Vestibular rehabilitation and 6-month follow-up using objective and subjective measures. Acta Otolaryngol, 126(3):259-66.

Krizhevsky, A., Sutskever, I., Hinton, G. E., (2012), Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems. 1: 1097–1105.