

CHAPTER 10

Analysis of Patient Emotions by Deep Learning Approach

Fatma Nur UZUN, Yusuf UZUN

Analysis of Patient Emotions by Deep Learning Approach

Fatma Nur UZUN

Konya Technical University

Yusuf UZUN

Necmettin Erbakan University

Introduction

Today, emotion detection is used in many fields. Emotions of users are used to reach objective evaluations on digital platforms. It has been observed that advertising channels or movie reviewers are starting to use real-time emotion detection to detect users' emotions. Real-time emotion detection was performed in projects developed using deep learning (DL) and artificial intelligence (AI) algorithms. In this way, it is possible to follow the emotional changes of the users simultaneously while watching a video or movie. As a result of the accurate results of emotion detection used in similar studies, its use in the field of health was also evaluated. The evaluation system of many hospitals is based on patient opinions as well as statistical values. Studies have shown that these assessments do not adequately detect the patient's emotions. Emotion detection studies using multiple datasets can explain most very complex patient reviews that indicate only a few emotions. In particular, it is of great importance to detect the emotions of patients who cannot express their emotions and to prevent erroneous emotion detections. This study aimed to address the problem of perception of multiple emotions as a result of emotion detection studies in the field of health.

Emotion detection is frequently used in applications such as human-computer interaction (HCI). Also, from emotion recognition systems, video games, animations, psychiatry, automobile safety, education, etc. it is used in many areas (Semerci et al., 2022). It has provided significant benefits in areas aimed at protecting human life. Human life has been significantly protected by detecting whether the driver is tired or not, thanks to emotion detection, and transferring it to the warning systems. In addition, it is predicted by emotion detection that patients or children with special needs will recover faster by taking their emotions into account during their rehabilitation training and following the right methods in treatment (Edo-Osagie et al., 2020). In addition, the convenience of emotion analysis has been used in the treatment of mental illness and in remote medical services. Similarly, with emotion analysis, it has been observed that in cases where human emotions cannot be clearly detected by the human factor, it is performed by means of a machine and gives us accurate results. For this reason, its use in the field of health has been frequently mentioned recently. The use of emotion detection has also been deemed necessary by hospital administrators to understand

patient satisfaction in hospitals. In this way, it is aimed to better understand the patient's expectations by analyzing the emotions of the patients.

Emotional change of the patient can also be understood by methods such as electrocardiograms, detection of psychological signals and skin temperature analysis (Serrano-Guerrero et al., 2022). However, such processes are both costly and time consuming. In the field of psychiatry, emotion detection is used to detect diseases such as stress and venous thromboembolism (Sabra et al., 2018). Especially during emotion detection, the large number of data sets will enable the detection of emotions that are normally difficult to understand from facial expressions. Machine learning and deep learning-based algorithms were used to detect multiple emotions. More than one experiment was carried out.

Project Purpose

With the developed project, it is aimed to detect emotion both via photo and video and real-time instant camera. In this way, emotional changes of individuals can be followed in many ways and information will be provided to the necessary places. Sensitive monitoring of emotional change is planned. It is aimed to reach a more objective result by using more than one data set.

Project Significance

Emotion analysis is a popularly studied field in computer science for emotion recognition. For the interaction between human and computer to be more natural, computers are aimed to recognize human emotions. Although sentiment analysis is mostly used in the field of text mining, it has recently been the subject of studies in the field of computer vision. In the studies in the field of text mining, three feelings of the person about any subject, namely positive, negative and neutral, are generally taken as reference. Sentiment analysis in the field of text mining can be used in advertising, product marketing, current events, etc. It is done to extract short and meaningful information in the fields. Sentiment analysis studies in the field of computer vision are called Automatic Facial Expression Recognition, shortly FER (Face Expression Recognition). It has been observed that studies in this field play a major role in detecting the emotions of individuals over time. In the project, using computer vision applications, more than one emotion that the individual has at the same time is determined. In this way, it is prevented that the individual is limited to only one emotion while describing his emotions. In many areas, erroneous results have been prevented by identifying the emotions of people more accurately.

In an article examined, while the face recognition system was being designed, it was made available to researchers as open source and a dataset was created from 4 different databases. While creating the dataset, the CelebA (large-scale CelebFaces), FFHQ (Flickr-Faces-HQ Dataset), LFW (Wild Home dataset) and Youtube Faces dataset used by the researchers were used. A total of 300,000 images from these datasets were used for educational purposes, while 60,000 images were used for testing purposes (Dehghan et al., 2020; Oliver et al., 2015).

In the research, it has been concluded that visual intelligence can also measure the emotional state with its tracking and observation abilities around the world. In the pilot project initiated by the Chinese government, it is planned to make primary school children's learning more efficient by measuring the concentration they show during class lectures from their facial expressions (Khanbhai et al., 2021).

Data Set

In the emotion detection project, Open-CV, TensorFlow and Keras libraries were used extensively. The appropriate dataset (fer-2013) was found from Kaggle. The training set consists of 48*48 pixel grayscale 35,888 samples. The data were divided into 7 categories in total, as in Figure 1. (0 = Angry, 1 = Disgust, 2 = Fear, 3 = Happy, 4 = Sad, 5 = Surprise, 6 = Neutral). In this way, the emotional change of the patient was determined by reaching a more objective result. The project was carried out in such a way that patients who could not express their emotions clearly could perceive more than one emotion at the same time. The training set consists of 35,888 examples. train.csv contains two columns "feel" and "pixel". The "Emotion" column contains a numerical code ranging from 0 to 6 for the emotion present in the image. The "Pixels" column contains a string in quotation marks for each image. A facial emotion recognition system was handled in a two-stage process. These processes are face detection (limited face) in the image, followed by emotion detection on the perceived limited face. Two techniques were used to carry out these processes. The Haar feature-based classifiers used provide good front-side detection in an image. It is real time and faster compared to other face detectors. The data set was separated according to the train test split method. The 80% training dataset is set to be the 20% test dataset. 28709 images were used for training. For the test, 7178 images were used.

Figure 1

Fer-2013 Data set (<https://www.kaggle.com/datasets/msambare/fer2013>)



Artificial Intelligence

Today, many studies are carried out in the fields of artificial intelligence. Artificial intelligence, which is the type of intelligence observed in machines, unlike other living things; It functions in areas such as image processing, data science, machine learning

(ML), computer vision and deep learning (Artificial Intelligence, 2022). Artificial intelligence is briefly defined as systems that imitate human intelligence and can regularly improve themselves in order to perform the tasks assigned to it. Chatbots, smart assistants, facial recognition systems, autonomous cars, smart home systems and recommendation engines are shown as examples of common usage areas of artificial intelligence. (AI, 2021). Artificial intelligence emerges in many areas that will make our work easier in our daily life. When it comes to Artificial Intelligence, although multifunctional robots that take over the world in the human mind come to life, artificial intelligence is not designed to replace humans. Rather, it is designed to make people's jobs easier and to improve their abilities significantly. For this reason, it is used as a facilitator in many sectors.

Reasons for the widespread use of artificial intelligence are shown to be affordable, high-performance computing, access to a lot of data for training, minimizing risks and creating a competitive environment. Artificial intelligence has also been used in the health sector for a long time. Using artificial intelligence algorithms, it was possible to detect liver, prostate and bowel cancer diseases with 94 percent accuracy (Artificial Intelligence Uses, 2020). In this way, the use of artificial intelligence in a disease where early diagnosis is important, such as cancer, is of great importance. Similarly, the work of healthcare professionals has been facilitated by medical imaging systems used in many areas. While detecting emotion, many technologies brought by artificial intelligence were used based on studies in which artificial intelligence was used in the field of health. Sub-branches of artificial intelligence such as deep learning, machine learning and computer vision have been used. Artificial intelligence, while detecting the emotions of patients, achieves more accurate results by comparing them with patients with similar emotions.

Machine Learning

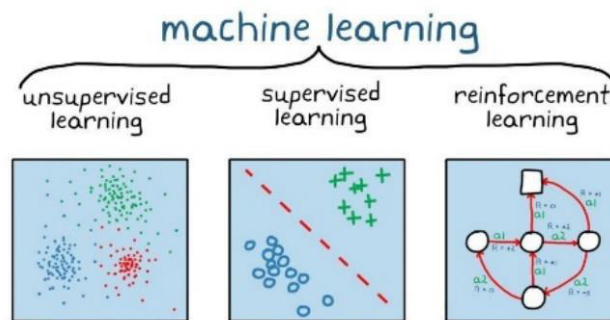
Machine learning is defined as a computer's use of mathematical models to directly enable learning (Machine Learning, 2021). It is used for processing and analyzing big data. It uses the patterns it has developed for this purpose to create a predictable data model. Similar to the development of people by gaining experience, more accurate results are obtained in machine learning as the amount of data and experience increases (Machine Learning, 2021). Machine learning is frequently preferred due to the fact that Bayesian analysis is highly preferred, the tendency towards machine learning is increasing and its use in data mining is widespread. Situations such as the increase in data types and volumes, performing computational operations that are both cheap and powerful, and storing data economically have led to an increase in the importance of machine learning (SAS, 2021). Machine learning is one of the subsets of artificial intelligence. A computer created with artificial intelligence thinks like a human and performs the tasks assigned to it. One way to train a computer to perform human-like activities is to use a neural network consisting of a set of algorithms modeled on the human brain. With machine learning, predictive analysis is performed using a static dataset. Since the static dataset is used, it needs to be refreshed to be updated.

Machine learning; it has advantages such as improving the user experience, predicting the behavior of individuals, ensuring and improving data integrity, reducing the risks that may occur and reducing the cost. In the emotion detection project, machine learning was used to increase efficiency and classify data more accurately. Machine learning is divided into a number of terms as supervised learning, unsupervised learning and reinforcement learning. In supervised learning, datasets are processed with labels. The processed data “trains” the machine, improving its predictive ability. Supervised

learning is a learning method in which the content of the dataset is known. Problems that may occur are solved by observing continuous outputs. In unsupervised learning, datasets are processed without labels. Unsupervised learning types are divided into two as clustering and association. There are relationships between the data by grouping the data in clusters. It is the method used when little or no information is available about the data. Therefore, there is no certainty in unsupervised learning. In reinforcement machine learning, artificial intelligence makes consecutive decisions and is not controlled by humans. Unlike supervised machine learning, data does not need to be labeled or corrected. The purpose of reinforcement machine learning is discovery.

Figure 2

Types of machine learning (Types of machine learning, 2022)



The machine learning process is based on collecting data, training the model, validating the model using the test set, and interpreting the results. Model training takes place by training data sets that are divided into training sets and test sets. To achieve the highest accuracy, some adjustments should be made, such as adjusting the amount of dataset, increasing the number of training sessions, and making necessary adjustments to the dataset. Factors such as predicting unexpected situations and values, determining the appropriate model for training and providing classification are realized with machine learning. Machine learning, which is used in the field of health as well as in many sectors, is also used in the emotion detection project.

Deep Learning

Deep learning is a type of machine learning and artificial intelligence method that enables digital systems to learn and make decisions based on unlabeled data. It is used when creating multi-layer artificial neural networks (ANN) in the fields of object, sound and image recognition and natural language processing. The difference between deep learning and machine learning is that coded rules are not used. Instead, it performs the learning process by using data such as pictures, videos, audio and text. Deep learning is flexible. The larger the size of the data used in deep learning, the higher the accuracy of the predictions made. Artificial neural networks are used during this process. As a result of training the data set, artificial neural networks can generalize about the data set and collect external information. Later, when he sees the examples he has never encountered, he can make a decision about those examples with the information he has learned (Luckey et al., 2020). At the same time, neural networks can process large amounts of nonlinear data to solve complex problems.

Artificial neural networks; Similar to the human nervous system, the neuron consists of weights, bias and activation function. Billions of neurons in the human brain process the necessary information through small electrical signals and enable us

to perform our daily activities. In artificial neural networks, each neuron processes the inputs that come to it and transfers the results to the next neuron for the continuation of the process (Artificial Neural Networks, 2019). Thus, it is ensured that the necessary reactions are given by obtaining instant results. Similar to a brain filled with digital neurons, artificial neural networks interpret and respond to situations faced by digital systems, similar to our nervous system. Weights are given according to the importance of the information. Therefore, seeing low weights in a place indicates the insignificance of the information there. Deviation, on the other hand, is the linear component applied to these weights, except for the weights applied to the inputs. The activation function, which has different types, converts input signals to output signals. It allows the model to learn nonlinear structures by breaking the linearity. Commonly used are the sigmoid, relu (rectified linear units) and softmax activation functions. Relu and softmax activation functions were used in the emotion detection project.

Convolutional Neural Network

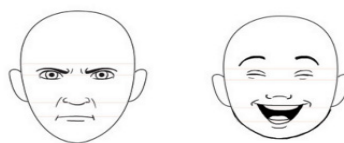
Convolutional neural networks (CNN) are a type of deep artificial neural network that is frequently used in image recognition and processing problems. Its use in pictures and video formats around the world gives very successful results. CNN first converts the images that come to it as input into a matrix format that can be defined by the computer. In this way, it is ensured that small details are captured. The system detects which label the picture has by considering the differences in the matrix and learns the effects of the differences on the label during the training. He then makes predictions for new pictures using the information he has learned. CNN has some layers that are responsible for performing these operations. In the convolution layer, the feature map of the picture is created by passing the picture through the filter. The sizes of the filters may vary according to the picture. Thanks to the attribute map, necessary operations can be performed by creating fewer connections.

Concept Mapping Model

In Figure 3, detection is provided by using matrix differences on different face types used as a data set during emotion detection. The training will be carried out by learning that the matrix differences in the mouth, eye and eyebrow circles of the model applied to the image have a great effect on the label. However, the forehead and ear regions appear to be the same for both images. Therefore, the effects of these values on training and forecasting are expected to be small. Thanks to the convolution layer, only the necessary and important features are determined and the features that do not make a difference are not taken into account.

Figure 3

Angry and smiling (Convolutional neural networks, 2020)

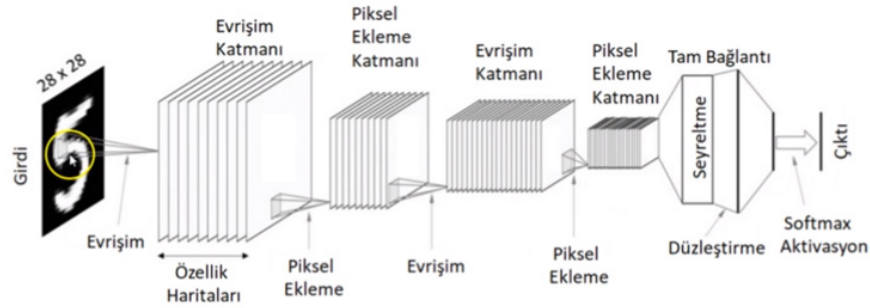


The pooling layer is applied to the feature matrices. The number of parameters is reduced by performing subsampling. It provides the perception of features that do not change according to changes in scale and direction. It provides control of overfitting by reducing the amount of computation in the network. Although the method

used in the pooling layer is max_pooling, methods such as mean_pooling and min_pooling are also used. In the mean_pooling method, the process is performed by taking the average. mean_pooling and max_pooling were used in the emotion detection project. The flattening layer, which is one of the CNN layers, transforms a two-dimensional data into a vector. It can be used more than once in the model. In the fill layer, pixels are added to preserve the original size of the picture as a result of decreasing the size of the picture as the convolution layers are applied. If the number of steps is set to 1, one-by-one navigation is provided between the pixels. Classification is performed at the full link layer. In the dilution process, randomly selected neurons are ignored during training. As the number of hidden layers and neurons increases, the depth increases. In other words, the perceptibility of complex structures increases. Figure 4 shows the convolutional neural network structure.

Figure 4

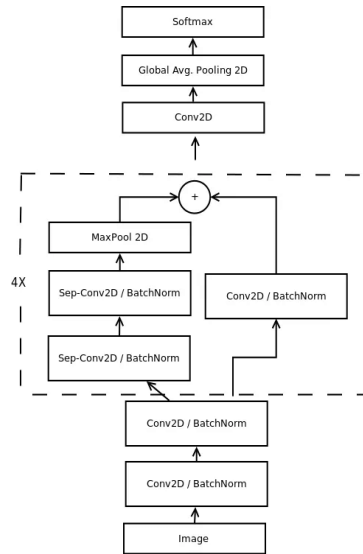
Convolutional neural networks (CNN, 2020)



Basically, there are continuous improvements on the CNN algorithm, which has a structure like Figure 4. LeNet, AlexNet, VGGNet, GoogLeNet, ResNet and ZFNet are examples of some well-known CNN architectures today.

Mini_Xception Model

The Xception CNN Model (Mini_Xception 2017) was used in the emotion detection project. The Mini_Xception model, whose general structure is shown in Figure 5, used in many projects similar to emotion detection, is different from the CNN model. Common architectures use fully connected layers at the end where most of the parameters reside. Also, standard convolutions are used. One of the modern CNN architectures, the Xception model benefits from a combination of two of the most successful experimental assumptions in CNN. These are now separable convolutions in terms of the use of modules and depth (Kim, 2014).

Figure 5*Mini_Xception model (Mini_Xception, 2020)*

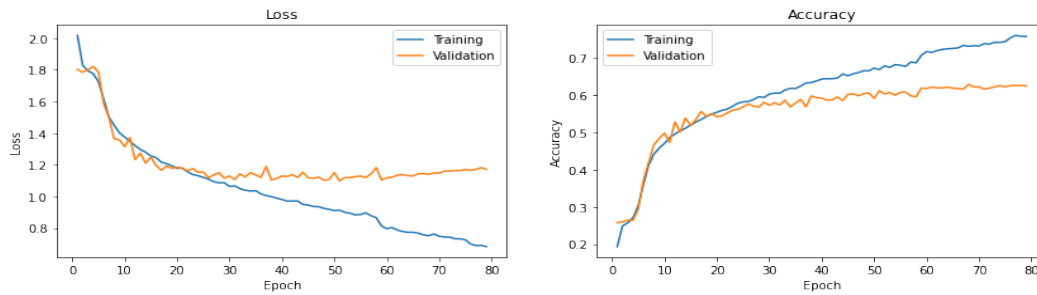
In the Mini_Xception model used in the project, it is noticed that the central block is repeated 4 times in the design, as seen in figure 5. Some of the techniques used while training the model are Data Augmentation, Kernel_regularizer, Batch Normalization, Global Average, Pooling, and Depth Separable Convolution. ReLU and Softmax functions are used as functions in the activation layer.

Education Result

As a result of the training process, the accuracy value generally increases, while the loss value decreases in general. Although there are changes in the increase and decrease values, this situation does not affect the general. The epochs value for training is set to 100. As a result of the training process, the training accuracy was approximately 91% and the test accuracy was approximately 84%. In order to increase these percentage values, necessary changes were made on the epochs value and retraining of the trained data was performed. As a result, an increase in the accuracy value was observed.

As a result of the training, the distribution of emotions resulting from the emotion detection of the patient and the responses given by the patient are compared. As a result, the accuracy of the model is compared objectively. Experiments on different CNN models are planned. According to the results obtained from the models and the answers given by the patient, it was deemed appropriate to use the model that gave the most accurate result. It has been observed that the Mini_Xception model gives the correct result as a result of the evaluations made for emotion detection. Figure 6 shows the graph of change in accuracy and loss values as a result of training the model.

Figure 6
Loss and accuracy change graph



According to Figure 6, while the loss value decreased during the training, the accuracy value increased. The error matrix, which is frequently used in machine learning, was used in the project. The error matrix allows us to evaluate the performance of classification models. Estimates of the target attribute and actual values are compared. The error matrix of the emotion detection project is given in Figure 7. According to the error matrix, there was more confusion between the expression of disgust with the expression of anger, the expression of frightened and the sad expression, and the neutral expression and the sad expression. The neutral expression was the most difficult to understand. In order to reduce the confusion in the error matrix, it is considered to increase the data set or make the necessary improvements.

Figure 7
Error matrix

true label	Angry	0.60	0.01	0.11	0.05	0.06	0.02	0.14
	Disgust	0.23	0.57	0.04	0.02	0.05	0.02	0.07
	Fear	0.09	0.00	0.50	0.04	0.13	0.09	0.16
	Happy	0.02	0.00	0.01	0.85	0.01	0.02	0.08
	Sad	0.13	0.00	0.11	0.06	0.44	0.02	0.24
	Surprise	0.02	0.00	0.06	0.05	0.01	0.84	0.02
	Neutral	0.06	0.00	0.05	0.10	0.08	0.01	0.70
		predicted label						
		Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral

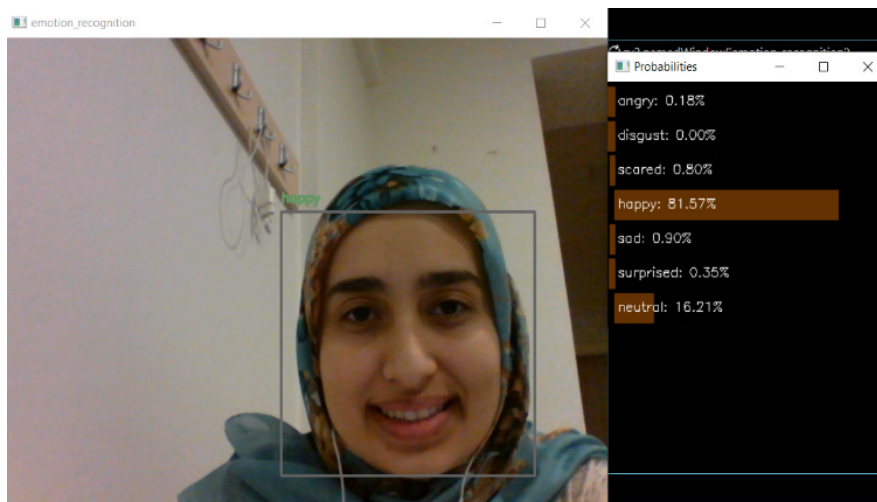
As seen in Figure 8, emotion detection is performed on more than one image depending on the coordinate location. The emotions of individuals are detected in real time via video. The simultaneous ability of the project to detect emotion enables the perception of instant emotional changes. Since emotional changes can change instantaneously in psychiatry clinics and private patients, simultaneous detection will provide more accurate results. It is planned to detect the instant emotions of individuals who have difficulty in moving their facial muscles. It was observed that the patients who came with their parents had difficulty in expressing their own feelings. In this way, the emotions of individuals who cannot express their emotions clearly will be detected by the machine and transmitted to the hospitals.

Figure 8
Image output



The data set used for the test was taken to include more than one emotion. Experiments have been made on sick individuals in real time and the desired results have been achieved. In Figure 9, the output with real-time emotion detection is given.

Figure 9
Real time code output



References

- Y. C. Semerci, G. Akgün, E. Toprak ve D. E. Barkana, A Comparative Analysis of Deep Learning Methods for Emotion Recognition Using Physiological Signals for Robot-Based Intervention Studies, 2022 Medical Technologies Congress (*TIPTEKNO*), Antalya, Türkiye, 2022, s. 1-4, doi: 10.1109/TIPTEKNO56568.2022.9960200.
- Edo-Osagie O, De La Iglesia B, Lake I, Edeghere O. A scoping review of the use of twitter for public health research. *Comp Bio Med* 2020;122:103770. <https://doi.org/10.1016/j.compbiomed.2020.103770>
- Serrano-Guerrero J, Bani-Doumi M, Romero F. P, Olivas J. A. Understanding what patients think about hospitals: A deep learning approach for detecting emotions in patient opinions. Elsevier B.V. 2022; 0933-3657. <https://doi.org/10.1016/j.artmed.2022.102298>
- Sabra S, Malik KM, Alobaidi M. Prediction of venous thromboembolism using semantic and sentiment analyses of clinical narratives. *Comp Bio Med* 2018;94: 1–10. <https://doi.org/10.1016/j.compbiomed.2018.05.010>

doi.org/10.1016/j.combiomed.2017.12.026.

Dehghan, Afshin, Enrique G Ortiz Guang, Shu Syed, and Zain Masood. DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Networks. <https://www.sighthound.com/products/cloud> (December 14, 2020).

Dürr, Oliver et al. 2015. "Deep Learning on a Raspberry Pi for Real Time Face Recognition TubeCam: A New System to Detect Small Mammals (Foremost Mustelids and Dormice) View Project Speaker Diarization View Project Deep Learning on a Raspberry Pi for Real Time Face Recognition." researchgate.net. <https://www.researchgate.net/publication/279537625> (December 18, 2020).

Khanbhai M, Anyadi P, Symons J, Flott K, Darzi A, Mayer E. Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health Care Inf* 2021;28(1):e100262. <https://doi.org/10.1136/bmjhci-2020-100262>.

Yapay Zeka. (2022) https://tr.wikipedia.org/wiki/Yapay_zek%C3%A2, (Date of access: 10.02.2023)

AI. (2021) <https://www.oracle.com/tr/artificial-intelligence/what-is-ai/>, (Date of access: 11.03.2023)

Yapay Zeka Kullanım Alanları. (2020) <https://www.indyturk.com/node/200346/bilim/yapay-zekan%C4%B1n-iyi-ama%C3%A7%C4%B1-kullan%C4%B1m%C4%B1na-10-%C3%B6rne%C4%B1>, (Date of access: 08.01.2023)

Makine Öğrenmesi. (2021) <https://azure.microsoft.com/tr-tr/resources/cloud-computing-dictionary/what-is-machine-learning-platform>, (Date of access: 15.01.2023)

Sas. (2021) https://www.sas.com/en_us/insights/analytics/machine-learning.html. (Date of access: 22.01.2023)

Types of machine learning (2022) <https://www.savasanadam.com/teknoloji/makine-ogrenmesi-ve-turleri-nelerdir/>, (Date of access: 22.02.2023)

Luckey, D., Fritz, H., Legatiuk, D., Dragos, K., Smarsly, K. (2020). Artificial Intelligence Techniques for Smart City Applications. Springer, 98, 3-15.

Yapay Sinir Ağları. (2019) [Derin Öğrenme \(Deep Learning\) Nedir ve Nasıl Çalışır? | by Nihat | Medium](https://www.medium.com/@nihatnihat/derin-ogrenme-deep-learning-nedir-ve-nasil-calisir?from=profile_article), (Date of access: 20.02.2023)

Convolutional neural networks. (2020) [Uygulamalı Evrimsel Sinir Ağları \(Convolutional Neural Network\) | by Adem Akdoğan | Bilişim Hareketi | Medium](https://www.medium.com/@ademakdogan/uygulamali-evrimsel-sinir-aglari-convolutional-neural-network-by-adem-akdogan-bilim-hareketi-medium), (Date of access: 20.02.2023)

CNN. (2019) [CNN Basit Uygulama | Kaggle](https://www.kaggle.com/datasets/msmbare/fer2013), (Date of access: 08.02.2023)

Kim Y. Convolutional neural networks for sentence classification. In: EMNLP 2014-proceedings of the conference on empirical methods in natural language processing; 2014. p. 1746–51. <https://doi.org/10.3115/v1/d14-1181>.

Mini_Xception. (2020) [Mini_Xception architecture for emotion classification | Download Scientific Diagram \(researchgate.net\)](https://www.researchgate.net/publication/354444444-Mini-Xception-architecture-for-emotion-classification), (Date of access: 08.02.2023)

Fer2013. (2020) <https://www.kaggle.com/datasets/msmbare/fer2013>, (Date of access: 08.02.2023)

About The Authors

Yusuf UZUN, PhD, is an Assistant Professor of Computer Engineering at Necmettin Erbakan University in Konya, Turkey. He holds a PhD in Mechanical Engineering from Necmettin Erbakan University. His main areas of interest are artificial intelligence, autonomous systems and augmented reality applications. He also works as the Rector's Advisor at Selcuk University.
E-mail: yuzun@erbakan.edu.tr, **ORCID:** 0000-0002-7061-8784

Fatma Nur UZUN is a Computer Engineer in Konya. She has a bachelor's degree in Computer Engineering from Konya Technical University.
E-mail: fnuruzun42@gmail.com, **ORCID:** 0000-0003-0153-3442

Similarity Index

The similarity index obtained from the plagiarism software for this book chapter is 15%.

To Cite This Chapter:

Uzun, Y. & Uzun, F.N. (2023). Analysis of Patient Emotions by Deep Learning Approach. In S.N. Guner, Y. Uzun & I. Cetin (Eds.), *Current Studies in Healthcare and Technology* (pp.126–138). ISRES Publishing.