All Sciences Proceedings
http://as-proceeding.com/

3rd International Conference on Innovative Academic Studies

September 26-28, 2023 : Konya, Turkey

https://www.icias.net/

Emotional Analysis with Deep Learning

İpek Şahin, Erdal Aydemir*

Dept. of Industrial Engineering, Faculty of Engineering and Natural Sciences, Süleyman Demirel University, Isparta, Türkiye

*(erdalaydemir@sdu.edu.tr)

Abstract **–** Image processing, one of the most exciting developments in deep learning, has made significant progress in various fields today. Suspicious behavior detection, social media user sentiment analysis, customer feedback, marketing and sales strategies, and many other areas can now analyze human reactions. Studies on the analysis of facial expressions date back to the 19th century. In 1872, Darwin proposed the universality of facial expressions. Therefore, dealing with facial expressions is not a new concept. Based on this, this project aims to integrate the concept of emotion analysis with the innovations of the modern era. According to a study conducted in 2012, Convolutional Neural Networks (CNN) are the most successful deep learning algorithm for object recognition. CNNs can distinguish features in images using numerous hidden layers. Filters that operate on matrices are used in each layer. These filters are the building blocks of the feature extraction process and play a crucial role in recognizing features such as edges, lines, colors, and other visual patterns. In this paper, a CNN model was built using the Python programming language and the Keras library. The model was trained using the FER-2013 dataset, and images of facial expressions were classified within the dataset. The performance of the model was evaluated using loss and validation metrics. Optimization efforts were made on the model to observe the changes in validation metrics due to training parameters. With an optimized model based on this information, simultaneous emotion analysis, age estimation, and gender prediction can be performed.

Keywords – Machine Learning, Deep Learning, Emotion Analysis, Cnn, Image Processing

I. INTRODUCTION

Facial expressions, perhaps the most effective way for humans to express themselves, are formed through various muscle movements such as narrowing of the eyes, raising of the eyebrows, and movements of the lips. They assist us in understanding people's emotions, not only for the purpose of communication in our daily lives but also in many other areas. Understanding people is a necessity not only in our everyday interactions but also in fields such as human-machine interaction, crime prediction systems, medical treatment, and interpreting consumers' emotional responses [1].

The increasing integration of human-machine interaction enhances user experience. The convergence of technologies such as artificial intelligence and machine learning with human-machine communication makes it easier for machines to understand the needs, preferences, and emotional states of humans. As a result of this learning process, it becomes possible to offer more personalized services by providing appropriate responses [2].

Research indicates that consumers tend to prefer brands that recognize their preferences and provide tailored recommendations. Many brands are aware of this and invest in analyzing consumer behavior and emotions to act accordingly. One familiar example of this is Netflix offering personalized content recommendations. By analyzing a user's past viewing preferences and considering factors like genre or actors that pique their interest, Netflix customizes the covers of recommended content accordingly. This prevents users from leaving the platform and increases brand loyalty. It's safe to say that machine learning algorithms are the creators behind all these processes [3].

When considering the preference for facial expressions in this context, we can emphasize the importance of universality. Regardless of their religion, language, or race, people can express their emotions in a common manner. It can be stated that this common means of expression is facial expressions [4]. Furthermore, it has been established that during communication with humans, facial expressions are effective at 55%, voice and intonation at 38%, and words at only 7% [5]. This indicates that data obtained from facial expressions will be effective.

The analysis of facial expressions typically involves a series of processes, starting with human detection within an image, followed by face detection and the extraction of facial features.

In recent years, one of the widely used methods in image processing and facial recognition is deep learning. Deep learning, a subset of machine learning, employs artificial neural networks and large datasets to perform tasks such as classification, recognition, and [6]. Unlike traditional machine learning methods that rely on encoded rules, deep learning can automatically learn from symbols of data related to images, videos, audio, and text [7].

II. Materials and Method

A. *Deep Learning*

One of the key distinguishing features of deep learning is its ability to create feature maps with its neural networks. Feature maps represent the intensity of pixels that represent different features in visual data. Each filter applied to an image creates a relevant feature map by emphasizing a specific feature in the input image. This is how deep learning acquires its predictive capability [8].

Deep learning generally excels in data-intensive problems and is particularly well-suited for processing complex structures like images, text, and sound. It relies on artificial neural networks inspired by the neurons in the human brain [9]. Deep learning algorithms achieve success by imparting complex and hierarchical features to the data through many layers of deep neural networks.

One of the deep learning methods, Convolutional Neural Networks (CNN), is highly effective in the field of computer vision and image processing. It is specially designed for processing and analyzing visual data [10].

In the formation of deep learning, data is the most crucial factor. With the increasing ease of data collection every day, thanks to growing data density, problem-solving with deep learning becomes more accessible.

In a Convolutional Neural Network model, datasets are grouped by labeling. Each pixel of labeled dataset images is used as input data for the model. Each pixel represents a color value at a specific position in the image. These color values are usually expressed in the RGB (Red, Green, Blue) format, ranging from 0 to 255 for each channel. In grayscale images, there is only one channel [11].

Additionally, another way to acquire data is through cameras and sensors. The captured images undergo preprocessing steps such as filtering and noise reduction, and are then segmented into relevant regions of interest.

As an example, consider lane detection systems in vehicles. In this case, the image obtained from the camera is processed only for the part of the road it covers. Subsequently, object detection and tracking take place. Feature extraction and classification follow for objects. In the detection of objects, human bodies, and faces, the use of the OpenCV library, which includes pre-trained models, is quite common. In this study, we will examine the use of the Keras deep learning library, known for its high performance, in creating a CNN model.

B. *Emotional Analysis with Deep Learning*

In this application that utilizes deep learning methods, the Visual Studio Code code editor and the

Python programming language have been used. A Convolutional Neural Network (CNN) model has been created using the Keras library. The aim of the model is to classify the emotional states of individuals in images using training and test datasets. The dataset used is the Facial Expression Recognition 2013 (FER-2013) dataset, which is commonly used in facial expression recognition studies. This dataset is a collection of images containing facial expressions and their corresponding labels.

FER-2013 was created by collecting images from various sources worldwide. This dataset includes facial images representing various emotional expressions of different individuals. The FER-2013 dataset consists of 35,887 grayscale images of size 48x48, labeled with 7 different expression categories [12].

Table 2.1 Number of Training and Test Data by Emotion Types.

Class	Train	Test
Angry	3995	958
Disgust	436	111
Fear	4097	1024
Happy	7215	1774
Neutral	4965	1233
Sadness	4830	1247
Suprise	3171	831

The data for the created CNN model was preferred to be in grayscale. The reason for this is that grayscale images process more quickly and efficiently compared to color images since they work with only one channel, while color images have three channels. Additionally, in grayscale images, filters are applied to only a single channel, which results in sharper and more distinct features. For operations like edge and corner detection and feature extraction, grayscale transformation is also preferred for the most accurate results [13].

In the model creation, the Keras library was used, which provides an intuitive API that allows models to be built and trained quickly. The choice of the Keras library is quite useful for several reasons, such as its ability to seamlessly integrate with popular deep learning libraries like TensorFlow, Theano, and CNTK, its capability to switch between different backend options, and its ability to perform high-performance computations on GPUs [14].

In the model architecture, convolutional and activation layers were added, followed by 3x3 max-pooling layers. A 32-filter layer with a 3x3 filter matrix was used. The purpose of this filter matrix is to allow the capture of different features through different filters. Filters take a small portion of the input image and extract the features in that area. A 3x3 filter matrix is commonly used because it generally produces good results due to its size. Afterward, these filters were increased to 64 and 128, respectively.

The training of the model was performed using the Fit Generator method. First, a training session lasting for 100 epochs was conducted, followed by another training session lasting for 2 epochs. The 100 epochs aim to allow the model to achieve higher accuracy and lower loss values. The 2-epoch training session allows the model to review the information it has learned so far and make one final update.

Table 2.2 Train Parametres

Parameters	Values
Batch Size	128
Dropout	0.5
Activation Functions	Relu and Softmax
Optimizer	Adam
Epoch	100 + 2
Conv 2D	32 Filter
	64 Filter
	128 filter
Maxpooling	2,2

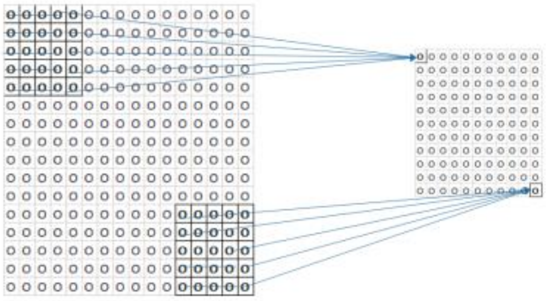


Figure 2.1 Extraction of Feature Maps with the Convolutional Layer [15]

The "ReLU (Rectified Linear Unit)" activation function was preferred as it accelerates the training process and has a high learning capacity. The ReLU function sets negative values to zero while leaving positive values unchanged. Because the function is linear, it makes it easier for the model to learn [8]. Additionally, the ReLU function helps mitigate the problem of "overfitting.".

The pooling layer reduces the size of feature maps while preserving important features, but it can also lead to data loss [10]. In the model, (2x2) maximum pooling was chosen. This operation divides the feature map into 2x2 regions and creates a new feature map by selecting the maximum value in each region. Taking the maximum values captures the most significant features in that area. As a result, the size is reduced by half. The purpose of this operation is to reduce the computational load in the model and also prevent overfitting. By making the features more generalized and prominent, overfitting is mitigated. These steps are repeated several times, which means the CNN model deepens. Each repetition narrows down the feature map and emphasizes more distinct features [16].

Next, a flattening (Flatten) layer is added. The Flatten layer takes the output from convolutional and pooling layers and converts all the data (multi-dimensional data structure) into a one-dimensional vector. The data size remains the same; only the shape changes. This allows the network to work with fully connected layers later on [17].

To create fully connected layers, the Dense function is used. In this model, fully connected layers with 1024 neurons are used, and these layers connect all units from the previous layer to all units in the next layer. In other words, they take feature maps as input and apply an activation function to make the feature maps suitable for the classification process [17].

The Softmax function is used in the output layer of the model. The Softmax function transforms the output values into a probability distribution, generating an estimated probability value for each emotion class [18]. Based on the given input, probability values are generated for each emotion class. The Softmax function used in the output layer normalizes the relationship between classes, ensuring that the total probability equals 1.

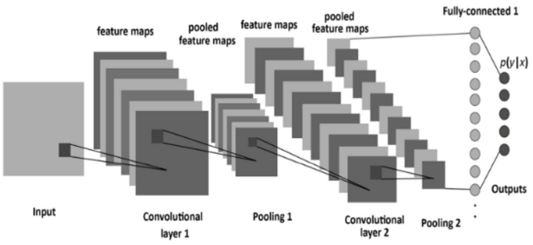
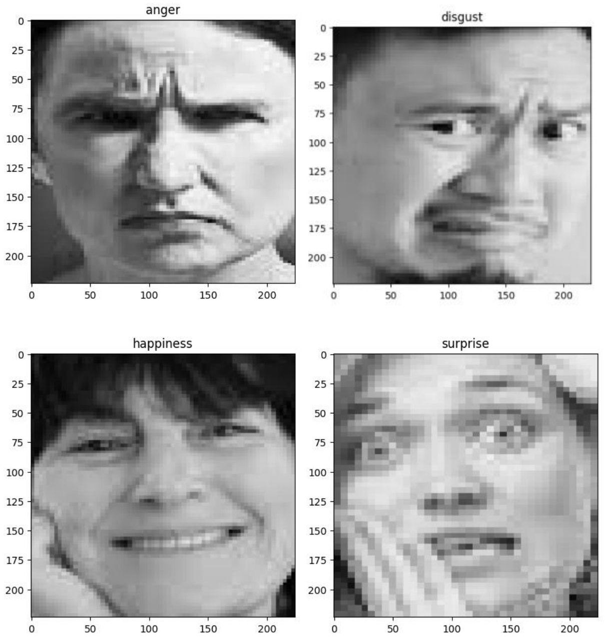


Figure 2.2 CNN Architecture [19]

The Image Data Generator class was used to apply various data augmentation techniques to image data, increasing diversity in the training and test datasets and reducing overfitting. For example, data can become recognizable from different angles through operations such as stretching and horizontal flipping. This process helps improve the model's generalization ability.

The model is compiled using the compile method. The usage of this method involves specifying the optimizer, loss function, and metrics. The ADAM algorithm regulates iterative weight updates during training.

III. Results



The task of predicting the emotional states of 4 different images selected from the training dataset was successfully accomplished. Obtaining accurate prediction results indicates that the model has been trained successfully and has achieved its intended purpose. The correct prediction of the emotional states of the images demonstrates that the model has learned to understand emotional expressions and recognize different emotional states.

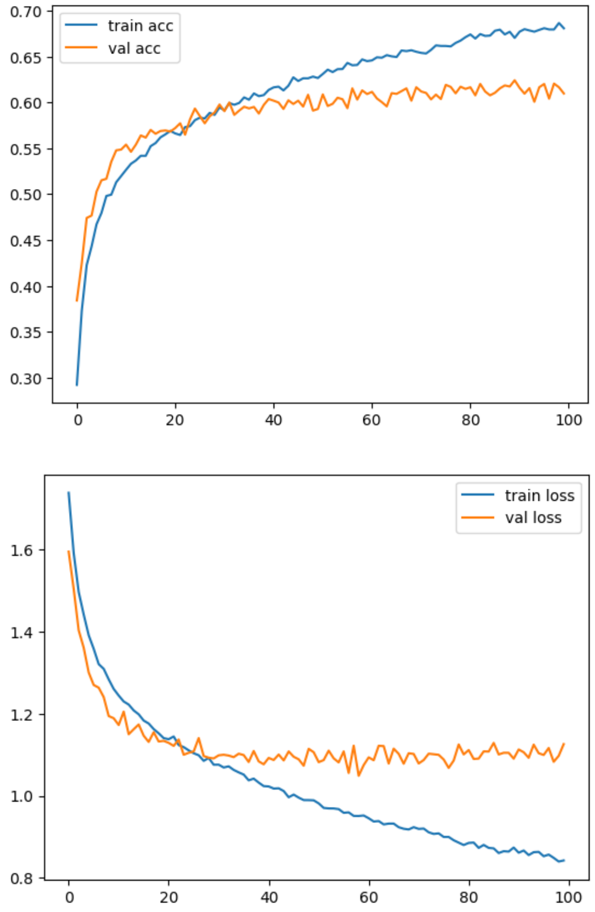


Figure 3.2: Graph of Model Performance Metrics

Performance metrics were used to evaluate the model. The metric called "Loss" measures the loss (error) in the training and validation sets. The metric called "Accuracy" measures the accuracy rate in the training and validation sets. These metrics are continuously updated during the training process and reported at the end of each epoch. A higher accuracy value and a lower loss value indicate that the model is performing better.

Table 3.1: Metric Values in the First Epoch

Loss	**Accuracy**	**Validation loss**	**Validation accuracy**
1.889	0.2923	1.5954	0.3842

When we look at the results of the metrics in the first epoch, we can see that the model has a high loss value and a low accuracy value at the beginning of training.

Table 3.2: Metric Values in the First Epoch

Loss	**Accuracy**	**Validation loss**	**Validation accuracy**
0.8172	0.6972	1.1233	0.6297

When we look at the results of the metrics in the last epoch, we can say that as the training of the model progressed, the loss decreased, and the accuracy increased.

Based on these results, it can be said that as the model's training advanced, its performance improved, and it started to produce better results after the initially low performance in the first epoch.

However, the validation dataset's loss (val_loss) is slightly higher, and the accuracy (val_accuracy) is slightly lower. This may indicate that the model has overfit to the training dataset or has the potential for further generalization.

IV. CONCLUSION

The optimization of the established CNN model involved exploring the importance of each parameter from the model's perspective. Work on the model is ongoing. Integration with a camera will enable real-time emotion classification on individuals, paving the way for real-time satisfaction analysis. In the future, it can be further enhanced with age and gender predictions, allowing for more comprehensive data collection.

ACKNOWLEDGMENT

The authors thank to the Scientific Research Projects Coordination Unit of Suleyman Demirel University for their support of this project. (Grant Number: FLP-2022-8973).

REFERENCES

[1] *Funda, A. K. A. R., & AKGÜL, İ. (2022). Derin Öğrenme Modeli ile Yüz İfadelerinden Duygu Tanıma. Journal of the Institute of Science and Technology, 12(1), 69-79.*

[2] *ŞENYÜZ, B. (2021). İletişim Çalışmalarında İnsan-Makine İletişimi (İmi): Paradigma Değişikliği Ve Temel Yaklaşımlar. Akdeniz Iletisim, (36).*

[3] *ZENGİN, F. (2021). Yapay Zekâ ve Kişiselleştirilmiş Seyir Kültürü: Netflix Örneği Üzerinden Sanat Eserinin Hiper Kişiselleştirilmesi. TRT Akademi, 6(13), 700-727.*

[4] *Ekman, P. (2007). Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. Macmillan*

[5] *A. Mehriban, “Communication without words”,Psychology Today, cilt 2, no. 4, pp. 53-56, 1968.*

[6] *Deng, L., & Yu, D. (2014). Deep learning: methods and applications. Foundations and trends® in signal processing, 7(3–4), 197-387.*

[7] *Yılmaz, A., & Kaya, U. (2020).Deep Learning. İstanbul: Kodlab Yayınevi.*

[8] *Tan, Z. (2019). Derin öğrenme yardımıyla araç sınıflandırma (Master's thesis, Fırat Üniversitesi, Fen Bilimleri Enstitüsü).*

[9] *LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.*

[10] *Metlek, Sedat ve Çetiner, Halit. Matlab Ortamında Derin Öğrenme Uygulamaları, IKSAD, 2021*

[11] *Peker, M. (2009). Görüntü işleme tekniği kullanılarak gerçek zamanlı hareketli görüntü tanıma (Doctoral dissertation, Sakarya Universitesi (Turkey)).*

[12] *Fabian Benitez-Quiroz, C., Srinivasan, R., & Martinez, A. M. (2016). Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5562-5570).*

[13] *Z. (2021). Otomasyon sistemlerinde görüntü işleme tekniklerini kullanan ürün tanımı uygulaması (Master's thesis, Sakarya Üniversitesi).*

[14] *Wazir, Muhammad., Irfan, Ullah., Mohammad, Ashfaq. (2020). An Introduction to Deep Convolutional Neural Networks With Keras. 231-272. doi: 10.4018/978-1-7998-3095-5.CH011*

[15] *Kayaalp, Kıyas ve Süzen, Ahmet Ali. Derin Öğrenme ve Türkiye’deki Uygulamaları. IKSAD, 2018*

[16] *Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., ... & Gambardella, L. M. (2011, November). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In 2011 IEEE international conference on signal and image processing applications (ICSIPA) (pp. 342-347). IEEE.*

[17] *YILDIRIM, M. (2022). Film Yorumları Kullanılarak Önerilen Yapay Zekâ Tabanlı Yöntemle Duygu Analizinin Gerçekleştirilmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 34(2), 751-760.*

[18] *Almurieb, H. A., & Bhaya, E. S. (2020, June). SoftMax neural best approximation. In IOP Conference Series: Materials Science and Engineering (Vol. 871, No. 1, p. 012040). IOP Publishing.*

[19] *Çiğdem, A. C. I., & ÇIRAK, A. (2019). Türkçe haber metinlerinin konvolüsyonel sinir ağları ve Word2Vec kullanılarak sınıflandırılması. Bilişim Teknolojileri Dergisi, 12(3), 219-228.*