Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1067-1082 YEAR: 2024
DOI: 10.17798/bitlisfen.1500558

# Optic Character Recognition Using Image Processing Techniques

Halit ÇETİNER[1*], Bayram CETİŞLİ[2]

[1]*Isparta University of Applied Sciences, Vocational School of Technical Sciences,Isparta, Türkiye*
[2]*Süleyman Demirel University, Electrical Electronics Engineer, Isparta, Türkiye*
*(ORCID: 0000-0001-7794-2555) (ORCID: 0000-0002-2600-3977)*

**Keywords:** Digit Recognition, Wavelet Transform, OCR, ANN, CNN, Deep Learning.

**Abstract**
In pattern recognition, the automatic identification of image-based input patterns by machine is an important problem. The need to convert images taken with tools such as cameras into formats usable by computers is increasing day by day. The output and efficiency of manual data entry processes are low and error rates are high. Outsourcing these processes to companies dealing with professional information entry is not preferred due to reasons such as security and lack of continuous service quality. For these and similar reasons, a system that enables automatic recognition of optical characters is aimed. With the help of this system, it is aimed to provide the opportunity to serve more people in a shorter time. In accordance with the stated objectives, a unique dataset has been created by performing identification card segmentation. This dataset was combined with the standard OCR dataset and classification was performed with ANN and proposed CNN methods on the extended dataset. The proposed CNN method, inspired by the Inception V3 model, is a deep learning model consisting of 15 layers. The ANN model uses features obtained from different wavelet types to increase discrimination. Competitive results were obtained from both ANN and proposed CNN models. In the proposed CNN model, in the extended version of the dataset, 99.49%, 98.87%, 99.48% and 99.50% values for F1 score, recall, precision, and accuracy metrics were obtained in the same order for training. Similarly, for validation, 99.13%, 98.43%, 99.21% and 99.27% values were obtained in the same order.

## 1. Introduction

Optical Character Recognition (OCR) poses a serious challenge in the field of computer vision as it plays an important role in the process of converting unstructured raw data into meaningful data. In order to overcome the challenges in OCR and to support ANN and CNN based artificial intelligence applications, algorithms that include text detection and text recognition steps in the OCR process are needed [1]. Text detection involves the identification and extraction of regions with meaningful data obtained using segmentation techniques. Text recognition involves the classification of the values in the extracted text regions with high performance criteria. In recent years, unlike machine learning based OCR approaches, deep learning based OCR approaches have made significant progress. These models are reported to offer high OCR performance even for images with complex backgrounds [1].

In OCR problems, there are a wide variety of scenarios for the detection of the text region and the recognition of meaningful characters in that region. Images, videos, scene images are among the materials used for OCR process. OCR is used in many application scenarios such as scene recognition, license plate recognition, machine translation, number recognition [1]. In the OCR process, it is aimed to provide automatic detection and recognition of text regions from video images consisting of an image or images.

[*]Corresponding author: *halitcetiner@isparta.edu.tr* *Received: 13.06.2024, Accepted: 21.10.2024*

Today, electronic government systems (EGS) have been spread too much. One of the most important reasons for the widespread offers great advantages. The biggest advantage of the EGS is provided with easily access to information about that person with identity. In addition to the person can obtain the information from institutions and organizations such as information services, integrated electronic services, payment transactions, short cuts to institutions and organizations, date information and announcements, message from the institutions to the citizens. This paper proposes a digit recognition system that can be integrated into systems such as EGS in the future. In this sense, both machine learning and deep learning based methods are tested for digit recognition and classification.

While the institutions employees logging to the EGS with identification card (IC), they are not entering as accurate and fast as systems of all the numbers on identity because of reasons such as eye strain, distractions. The ICRS to facilitate the use of EGS are realized. The ICRS is an optical character recognition system. Character recognition has found a wide range of applications in different fields such as recognition of card numbers belonging to various banks, directing of vehicles by reading road signs, advertising, banner, reading of market boards, automatic reading of bank checks, license plate recognition, reading of postal codes in the last century. More sophisticated methods are needed in order to meet the increasing request in this is a different application area. There are many different works related to the mentioned OCR based applications. Authors in the [2] presented number recognition system. The system uses pre-processing methods which are based on processes such thresholding, noise extraction. Then, each image of all images resized a standard size by the system. Finally, end points of numbers in thinned image are obtained by using chain code. Thus, they developed number recognition system. Authors in the [3] presented number recognition system. They use likelihood values between graphs which are created feature vectors. Then, they use KNN algorithm so as to classification of oriya numbers. Authors in the [4] presented license plate recognition system. The system consists of detection, respectively shredding according to character, recognition. Author in the [5] presented plate recognition system which uses thresholding, morphological operations, template matching. Authors in the [6] presented number recognition system which consists of segmentation and recognition. The system is a mobile application based on android operating system. Authors in the [7] presented plate recognition system which consists of thresholding, localization of plate with genetic algorithm. Authors in the [8] presented printed text recognition system which uses training with Hidden Markov model by using multiple fonts.

Authors in the [9] presented active learning methodology based on KNN. The methodology is used in the recognition of different script numbers. Author in the [10] presented character recognition system which consists of forms filled by writers, convert forms into images, creating of train and test sets, pre-processing with Otsu thresholding, segmentation with connected component analysis, feature extraction with framing feature, classification with KNN and SVM classifiers. The society began to development with information. It is difficult to control the information with increasing level of information. The circulation rate of the information is much increased according as development of computer and communications technology. This rapid growth makes necessary to make an efficient use of public and private institutions. The effective usage which can only be solved to make use of computer and internet technology. Computer and communications technology play key role in the data collection, storage, interpretation and the process of reaching the users. Advances in information technology has become facilitated access to information. The use of this information technology in public administration is suitable accordance with EGS for all private and public institutions. Thus, it will become faster and quality of public and private services. A character recognition system has been developed in Turkey due to the convenience of EGS in many different areas such as mobile electronic system integration, traffic information system, mobile information, earthquake emergency response network, lone pension operation, national judicial network. It is aimed to facilitate access to the system at a later point in the process, including facilitating the work. In Turkey, the verification process takes a lot of time as the employees of the organisation log in to the EGS with IC. In addition to this, institutional employees find the existence of individuals from EGS databases by manually logging in. During these processes, negative situations or delays may occur as a result of incorrect entries. For these reasons, ICRS has been developed.

As in the EGS system in Turkey, there are also applications in the world that automate the OCR process. In the United Kingdom, medical reports created by doctors are stored in the national health service database. Important information of these documents is identified and added to the patient's database. In this and similar ways, it is seen that studies have been carried out in the literature to automate the manual process by reducing the cognitive load [11]. The OCR helps to digitize printed materials such as valuable information, forms, etc. that are present in printed documents.

In line with the stated aims and objectives, the main contributions of this article to the literature are listed below.

- In this paper, a new segmentation algorithm is developed that automatically detects the digits in TC IC cards and separates the digits from each other.
- A new digit dataset, which is not defined in the literature, was created using the newly proposed segmentation algorithm.
- The new dataset was merged with the publicly available standard OCR dataset to extend the dataset.
- A new deep learning-based approach is presented for the OCR process, which consists of two main components: text recognition and text detection.
- Wavelet-based features were extracted on the extended dataset (Dataset-IV) and ANN classification was performed. In the Dataset-IV, the highest values achieved with ANN were obtained using Rbio 3.1 and Bior 2.2 wavelet types. These values were found to be 100% and 99.8% accuracy for training and testing, respectively. As can be seen from this result, the performance of the Rbio 3.1 wavelet type has also improved as a result of the Dataset-IV.
- At the same time, the proposed CNN method was used to classify the characters on the Dataset-IV. In the proposed CNN model, 99.24 %, 98.96 %, 99.77%, 99.77%, and 99.24% values were obtained for F1 score, recall, precision, and accuracy metrics in training for the Dataset-I, respectively. For validation, 99.12%, 98.50%, 99.43%, and 99.02% values were obtained in the same order. In the Dataset-IV, 99.49%, 98.87%, 99.48%, and 99.50% values were obtained in the same order for training. Similarly, for validation, values of 99.13%, 98.43%, 99.21%, and 99.27% were obtained in the same order.

The following sections of the paper are organized as follows. Section 2 provides a detailed summary of machine learning and deep learning based OCR approaches in the literature. Section 3 provides information about the datasets used in the paper and the algorithms that process these datasets. Section 4 presents the performance results obtained from the ANN and proposed CNN methods used in the article. In the last section, the article is concluded.

## 2. Related Works

There are researches on the number recognition application areas for many languages such as Arabic, Chinese, French, German, British, Indian, Italian, Persian [12]–[19]. Unfortunately, for number recognition of TR's identity card numbers was not found in any work in the literature. A study was carried out to improve the performance of the EGS system in similar structures and to overcome this deficiency in the literature.

Extracting text from images requires overcoming many challenges. The most prominent among these difficulties are background, font size, size of images, quality and format differences. In order to overcome these difficulties and to obtain accurate and meaningful texts, modern solutions are needed. Addressing these challenging problems in the right way is critical in terms of increasing the efficiency of OCR systems. When the OCR studies carried out to overcome these difficulties are analysed, it is seen that two different approaches, machine learning based approaches and deep learning based approaches, dominate the studies.

Inspired by the power of deep learning in different problems, it is thought that deep learning-based approaches can be used in OCR systems [20]–[23]. As a result of the research conducted in this context, deep learning-based approaches that tackle OCR problems in the literature are presented below. Niranjan and Kumar developed a hybrid method using BiLSTM and CNN architectures to improve text recognition performance [24]. They use a synthetic dataset called MJSynth in their study. Yang et al. introduce a deep learning based PSENet architecture for extracting meaningful textual information from natural scene images [25]. Kirthiga et al. carried out a study on automatic recognition of Tamil scripts using deep learning based character recognition algorithms. They applied various preprocessing techniques to remove noise in Tamil scripts and to improve images. The authors reported that obtaining strong features is important in character recognition and compared the segmentation of characters, recognition rates and accuracy [26]. Batra et al. discuss the benefits of the OCR approach in the healthcare system, especially OCR applications that enable the digitisation of medical laboratory records [27]. Tesseract, EasyOCR and DocTR, three widely used OCR applications and pre-processing techniques such as image thresholding, filtering and sharpening are examined in detail. Among these, DocTR is reported to be supported by Tensorflow 2 and Pytorch artificial intelligence libraries [27]. Low et al. evaluate the performance of pre-trained deep learning models for text detection and recognition on different datasets, especially for the detection and recognition of seven segment digits. According to different text detection algorithms, PaddleOCR's DBNet method gives a good result in text detection, while the PARSeq algorithm gives a good result in the recognition of seven segment digits [28]. Singh et al. developed an OCR model with CNN and BiLSTM based architectures using language models trained on the JFLEG dataset [29]. Gujjeti et al. develop a deep learning algorithm for detecting and classifying text in scene images [30]. In the developed approach, CNN architecture is used to extract features from each of the segmented scene images to be recognized. In the recognition and classification of characters, classification is performed with the SVM algorithm, one of the popular machine learning methods. CNN-based algorithms are reported to provide good results in extracting features of characters in scene text images. CNN architectures are very useful for recognizing characters in scene images due to their high robustness to changes in direction, scale, and noise [30]. Chandio et al. developed a CNN-based method for extracting features and combining multi-level features to recognize Urdu characters [31].

Machine learning based approaches that tackle OCR problems in the literature are presented below. Authors in the [32] presented a number recognition system by using principal component analysis and linear discriminate analysis to determine the distinct between person and machines. Thus, security communication connection between person and machines is provided by the classification and evaluation of electroencephalogram signals. Authors in the [33] convert ANN based on desktop character recognition application to mobile application with piecewise linear methodology. They have increased up to 60% the rate of speed of without any change in the rate of recognition. They have expressed to down half of the size of code in this made the system. Authors in the [34] presented text recognition system. The system uses images obtained from phone camera. Firstly, they detected regions as English texts by using DCT statistical values. Secondly, candidate text regions are merged by using morphological operations. Thirdly, thresholding is applied to increase the separation between foreground and background of text regions. Fourthly, they carried out text recognition by using Tesseract character engine. In this point, English characters obtained from fourth process are translated to Spanish characters. They search all characters of the character engine by using binary search algorithm in the translation process. Authors in the [35] presented number recognition system by using a classifier based on Euclidean and Mahalanobis distance. The system consists of thresholding, features obtained from statistical values, two classifiers. Authors in the [36] presented thresholding technique which is identified of broken Gujarati characters. Authors in the [37] presented character recognition system. The system consists of feature extraction with chain, distance and gradient based features, classification with linear discriminant, quadratic discriminant and KNN classifiers. Authors in the [38] presented number recognition system which uses six different moment features. They use multilayer perception classifier in the classification of the numbers. Authors in the [39] presented number recognition system which consists of pre-processing with DCT, matrices transform of values obtained from pre-processing, classification with SVM of these matrices.

OCR models based on deep learning are said to give better performance results than traditional methods [1]. When the studies in this field are analysed in the literature, it is seen that OCR is generally performed using ready-made datasets and deep learning models with pre-trained weights. In order to overcome this deficiency, a new OCR dataset was prepared. At the same time, a detailed comparison of the ANN method, which is a machine learning based method, and the deep learning based proposed CNN method was carried out.

## 3. Material and Method

### 3.1. Material

In this section of the study, the number dataset used in the study prepared by Çetiner in 2012 was used first [40]. In the study, this dataset is named as Dataset-I. There are 1160 images in Dataset-I, half of which are images of female and half of which are images of male TC IC cards. The images obtained within the scope of this study were obtained by preparing a special hardware setup. This setup is presented in detail in Sections 3.1.1 and 3.1.2.

#### 3.1.1. Identification Card Recognition System (ICRS)

In this paper, ICRS is introduced. ICRS has been tested on TR IC images. TR IC is a card with a blue or pink background colour. The drawn patterns in the background make it difficult to recognise the numbers in the foreground. The TR IC usually deforms because the level of light in environment affects the quality of the image taken from the camera. In addition to, it may be deformed due to many reasons such as identity plated, cold stamp. Figure 1 shows algorithm of the ICRS. All steps of this study are given in the Figure 1. The ICRS consists of hardware design, pre-processing, feature extraction, classification, recognition respectively.

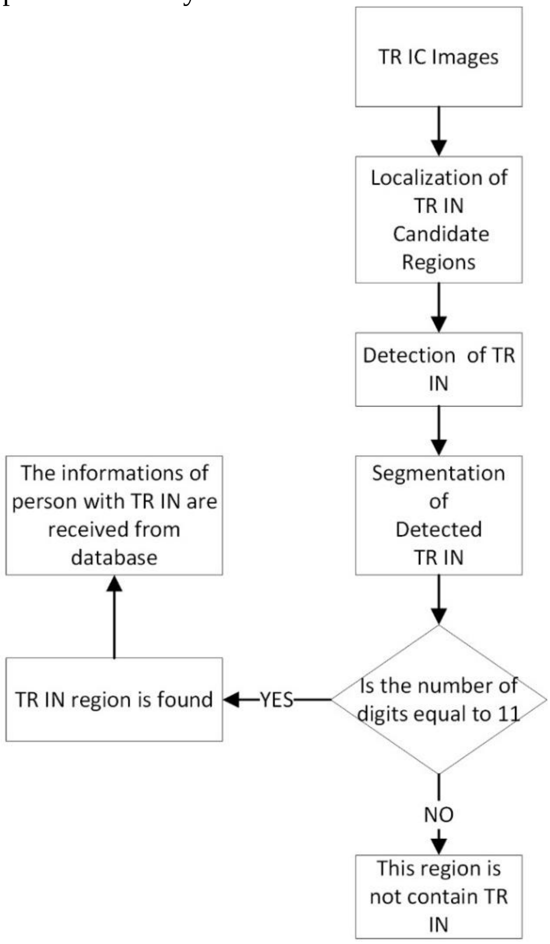


**Figure 1.** Identification card flow diagram [40]

### 3.1.2. Hardware Design

The hardware system is designed to prevent light changes. Otherwise, the images not taken under certain standards changes occur in many parameters such as quality, resolution, light level, lighting conditions. This condition makes difficult to localization of TR IN. The ICRS has been standardized to overcome all these difficulties by making closed cabinet. Figure 2 shows the designed model of cabinet. This cabinet consists of a camera, a camera lens, an illumination circuit, a computer that is connected via the USB port, an adjust power supply. An adjustable adapter is used as a power source. This adapter was used to provide power of the illumination circuit with LEDs (Light Emitting Diode).

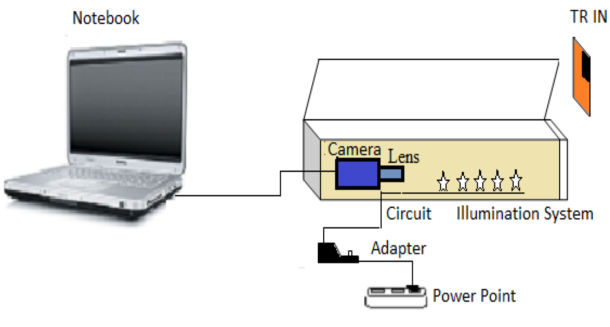


**Figure 2.** Test System Design [40]

The images obtained from Matrix Vision CCD BlueFox camera is send to computer via the USB port connection by placed within the cabin. There is place of IC on the other side and camera on one side of the cabin. Some statistical analysis is extracted to determine the best light values of the image obtained from the camera. Table 1 shows that voltage, brightness, contrast and recognition test results of images obtained from designed test system in the night environment. Illumination of camera within the closed cabin is provided with white LEDs placed on the side wall of the cabin.

**Table 1.** The experimental results in different voltages [40]

| Voltage (Volt) | Brightness Value | Contrast Value | Recognition |
|---|---|---|---|
| 1.5 | 0.00 | 0 | Unsuccessfully |
| 3 | 0.00 | 0.06 | Unsuccessfully |
| 4.5 | 0.00 | 0.06 | Unsuccessfully |
| 6 | 2.49 | 2.41 | Unsuccessfully |
| 7.5 | 88.34 | 45.01 | Successfully |

With the hardware and algorithm presented in Sections 3.1.1 and 3.1.2, 116 images of each class were obtained using the numeric images obtained from TR ICs. These images were created separately for each of the digits between 0 and 9. Some of the sample images of the dataset used in the study are presented in Figure 3.

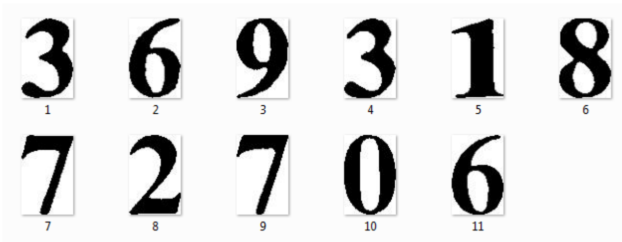


**Figure 3.** Images obtained with the prepared hardware [40]

In order to increase the number of images in the created dataset, images from the Standard OCR dataset were also included in the study [41]. In the study, the standard OCR dataset is named Dataset-II. Some of these images are also presented in detail in Figure 4.


**Figure 4.** Some of the images in the standard OCR dataset

In this way, a total of 2320 images were obtained by combined Dataset-I [40] to and the Dataset-II [41]. The dataset formed by the combined of these two datasets is named Dataset-III. When the studies on the subject in the literature are examined, it is seen that data augmentation is generally performed when the amount of data is low [42]. For this reason, the number of images was increased to 6960 by applying rotation, mirroring and inversion operations, which are the basic data augmentation operations. The new dataset obtained as a result of the data augmentation processes mentioned in the study is named Dataset-IV. The datasets used in the study are presented in detail in Table 2.

**Table 2.** Presented display of datasets

| Datasets | Description | Number of images used |
|---|---|---|
| Dataset-I ([40]) | The number dataset used in the study prepared by Çetiner in 2012 was used first [40]. | 1160 |
| Dataset-II ([41]) | Standard OCR dataset | 1160 |
| Dataset-III ([40] +[41]) | In order to increase the number of images in the created dataset, images from the Standard OCR dataset were also included in the study | 2320 |
| Dataset-IV ([40] +[41] with data augmented) | Extended dataset with data augmentation technique | 6960 |

### 3.2. Methods

In this study, classification performance results were obtained by using ANN and proposed CNN model on the generated dataset. For this purpose, firstly, basic preprocessing was performed on the images. In the second step, characters were detected by using horizontal and vertical histograms on the image. In the third step, features were obtained from the characters for the machine learning algorithms to be used in the study. At this stage, the features with the highest discrimination were used in the ANN algorithm. In the last step, which is the fourth step, the classification process was performed. At this stage, ANN and proposed CNN model were applied. The steps performed in the study are presented in detail in this section.

### 3.2.1. Preprocessing

Localization process of TR IN in TR IC is difficult due to the fact that TR IC has two different number regions. For this reason, the ICRS determines location of TR IN early. The dynamic thresholding of image is important to in real-time systems. The brightness and contrast adjustments were used for the transaction. The TR IC images have wear at different rates, various degrees of degradation. These cards printed by the machine may have different font, width and height values. The ICRS have determined that the density changes are caused by these changes. A separate threshold value for each image is determined due to these changes. To address this problem, the ICRS tried to set the appropriate brightness and contrast. After the experiment, binary images were obtained on the desired usefulness by using the average and standard deviation of gray level image.

The weighted average of the image gray values is an indicator of the image brightness. The weighted standard deviation of the image gray values is an indicator of the image contrast. Eq. 1 calculates brightness value of image.

$$IB_i{}^N = \sum_{i=0}^{N} \mu_i - I, 1,2, .... N \qquad (1)$$

$IB_i{}^N$ is insufficient brightness value of image with $i$ index. $i$ is the index number of the image. $N$ is the maximum index number image of the image database obtained from the ICRS with real time. $I$denotes the RGB image. $\mu$ denotes the mean of image. Eq. 2 shows that increasing the insufficient brightness value of the image pixel can be removed from a fixed $c$ value to adjust the brightness of the image.

$$Ind_i^N = IB_i^N < 50\{f_{imc}(I_{gray}) * (c - IB_i^N)\} \qquad (2)$$

Where $f_{imc}$ is a function that calculates the complement of the $I_{gray}$ which is grayscale image. Eq. 3 shows that increasing the insufficient brightness value of the image pixel can be multiplied from a fixed $c$ value to adjust the brightness of the image.

$$\begin{gathered} Ind_i^N = IB_i^N >= 50|then \\ Th_i^N = f_{ima}\left(\mu(I_{gray})\right) + \sigma(I_{gray}) * (0.3/c) \\ I_{gray} < Th_i^N is ... applied \end{gathered} \qquad (3)$$

Where $f_{ima}$ is a function that adjust image intensity values of the $I_{gray}$ by adjusting at minimum value 0.1 and maximum value 0.9. $\sigma$ denotes the standard deviation of image. $Th_i^N$ is thresholding value of image with $i$ index. The obtained thresholding method converts the two different RGB regions to binary format according to Eqs. 1, 2 and 3. The two different number regions convert to binary format with a simple thresholding.

### 3.2.2. The Detection of TR IN and feature extraction

In the detection stage, there are some difficulties due to the differences as font and size in TR IC. In addition to, there is more than one candidate region. The detection of TR IN is carried out in third steps. In the first step, pixel values in the vertical axis of the binary image were collected. In the second step, intersection points with the axis of greater places than total pixel value 0 have been appointed 1 according to the sum of the pixel values. In the third step, segmented regions are obtained with logic applied in the second step according to vertical projection in the vertical axis.

The ICRS is carried out connected component of each segmented region. Then, the number of characters detected with 8 connected objects in each region are kept. Then, horizontal projection of each segmented region is obtained to ensure the accuracy of the detected region and total pixel values of the region are calculated. The ICRS is calculated parameters such as the average distance between the characters, distance between characters, the number of characters for each candidate TR IN region. At this stage, wavelet transform is used for feature extraction.

Wavelet transforms use two functions that can extract local and global features of patterns by performing localised analysis. One of these two functions is the wavelet function determined by the high pass filter. The other is the scaling function, which is the low-pass filter that produces the wavelet approximations [43]. In image processing, discrete wavelet transform (DWT) is generally employed. In DWT, the mother wavelet $\psi(t)$ is scaled and translated at different decomposition level, and then the child wavelet $\psi_{m,n}(t)$is convolved with the image. The child wavelet is given as;

$$\psi_{m,n}(t) = \frac{1}{\sqrt{m}}\psi\left(\frac{t-n}{m}\right), m > 0, n \in \Re \quad (4)$$

In Eqs. 4, 5, and 6, $m$ and $n$ are scaling and shifting parameters in the time domain, respectively. In generally wavelet and scaling function transform are determined with multiples of two ([44]):

$$\begin{Bmatrix} \varphi_{m,n}(t) = 2^{-m/2}\varphi(2^{-m}t-n) \\ \psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t-n) \end{Bmatrix} \quad (5)$$

For the $S_t$ time series which are containing the samples of $N$ has an $i$ discrete time step, the discrete wavelet transform defined as in Eq. 6:

$$W_{m,n} = \int_{-\infty}^{\infty} f(t)\,\psi_{m,n}(t)dt \quad (6)$$

These coefficients measure the variations of the field $f(t)$ about the point $n$ with the scale given by $m$ and in Eq. 6. Also, $W_{m,n}$ has a time value that is the wavelet coefficients. Discrete wavelet transform separates the two main components of signals. These processes are done using filters. The first of these main components is low-frequency components, thus low-pass filter outputs. Other high-frequency components, thus high-pass filter outputs.

Approaches ($A$) are low-frequency components of the main components and Details ($D$) are high-frequency components of the main components in composed of discrete wavelet transformation. $S(t)$ signal is done parsing process by subjecting to discrete wavelet transform process. It is separated to equal the number of sub-frequency component, is called $A$and high-frequency component, is called $D$. The detail and approximation coefficients of the S(t) signal shown in Figure 6 are high-frequency low-scale components and low-frequency high-scale values, respectively. In Figure 6 observed that decomposition of a $S(t)$ signal in third level.

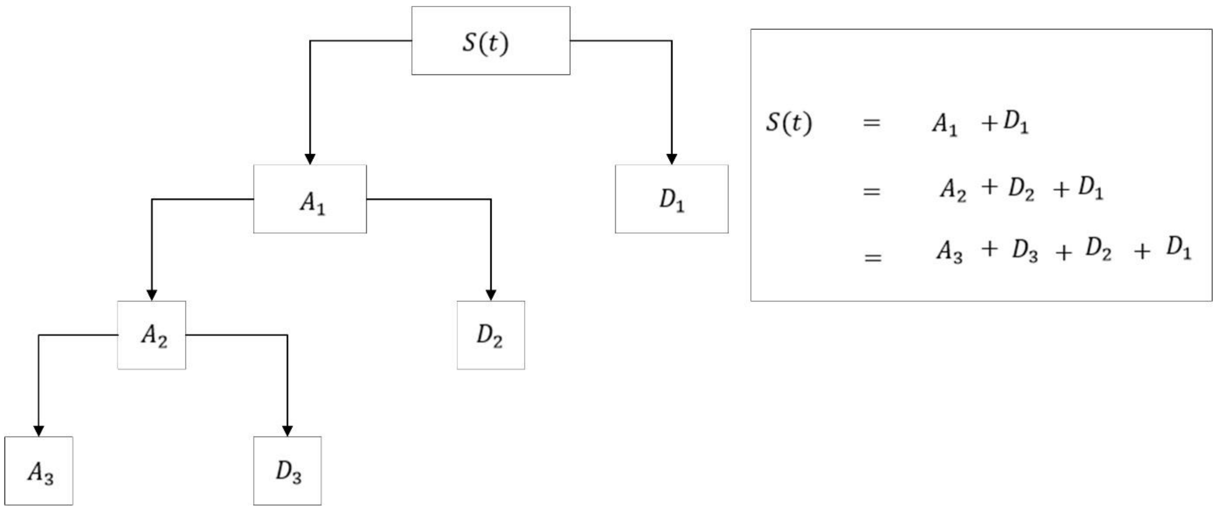


**Figure 5.** Filtering process in wavelet transformation [40], [43]

After leaving the sub-bands with the standard wavelet transform in Figure 5, approach component of signal is divided into sub-bands with same procedure again and this procedure continues until it reaches the desired resolution. As can be seen in Figure 5, the signal $S(t)$ subjected to discrete wavelet transformation. This signal does not any loss, only its components separated multiplied by main wavelet. In this here, $A_3$ approach component is basic component. Approaches are low frequency components with high weight. Details are high frequency components with low weight. As can be seen the purpose of wavelet transformation, signal are decomposed approach and detail components. The resulting approach is decomposed approach and detail components in second level again; this process can be repeated until the desired result is obtained. Signal can be separated in its components at desired resolution level by process of separating the components is sequential repeated [43]. Sequentially repeating the process is shown in Figure 5. Firstly, original $S(t)$ signal passes through high-pass $g(t)$ filter and a low-pass $h(t)$filter. This is a separation process with level and mathematically expressed as in Eq. 7 and Eq. 8:

$$D(k) = \sum_{t} S(t)\, g(2k - t) \quad (7)$$

$$A(k) = \sum_{t} S(t)\, h(2k - t) \quad (8)$$

Following of sub-sampling with two in Eq. 7 and Eq. 8, respectively high-pass and low-pass filter outputs. The k is the level of the filter. In this way, feature vector is obtained by TR IN images are passed through filters. After these operations, the two-dimensional discrete wavelet transform function, the approximation and other detail information of an image with a 3rd level filtering of Daubechies 'db2' type are shown in Figure 6.

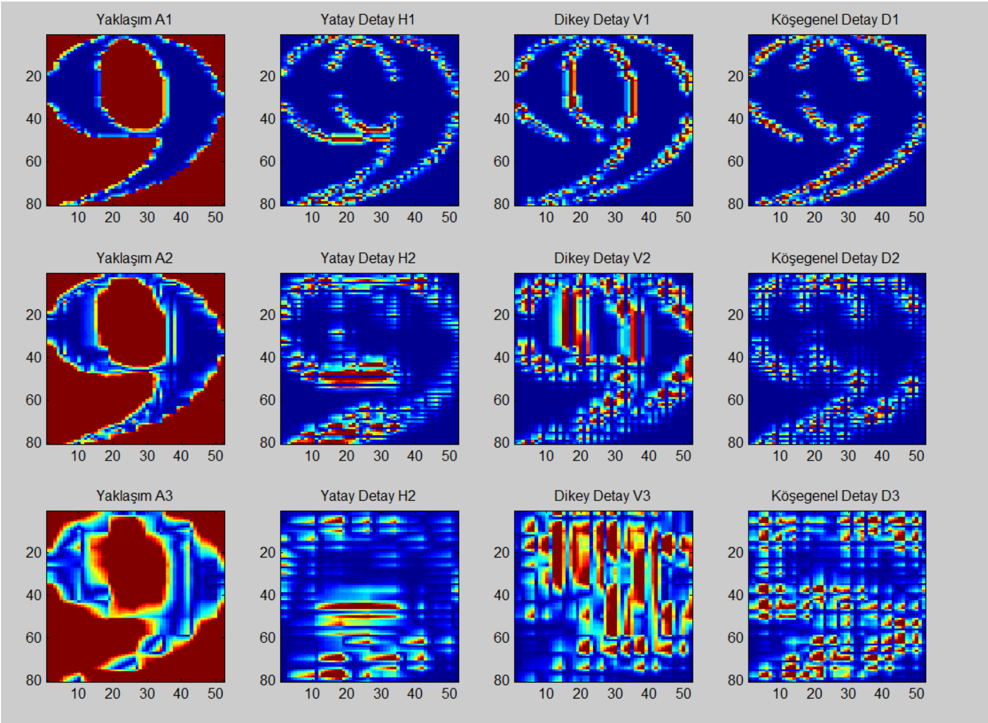


**Figure 6.** Third level approximation and detail features representation [40]

As shown in Figure 6, the wavelet-based feature with the most distinctive features is obtained from the 3rd level approximation coefficients. Although these features can be obtained automatically by convolution in CNN based algorithms, it is important to give highly discriminative features as input in classical machine learning based algorithms such as ANN.

### 3.2.3. TR IN Classification with ANN

Since the creation of human are living together with nature, and they have developed solutions for the problems faced by inspired from nature. These methods have made much progress the rapid development of numerical calculations is made using the computer. When some of the methods are developed, they are done by the example of living organism. ANN in this way is a method developed taking into account the structure of human biological neuron. In this study, the model ANN input is selected as 3480. This part is used for training purposes of 3480 feature vector. Network training is carried out by using in model ANN training of model input and output files are obtained.

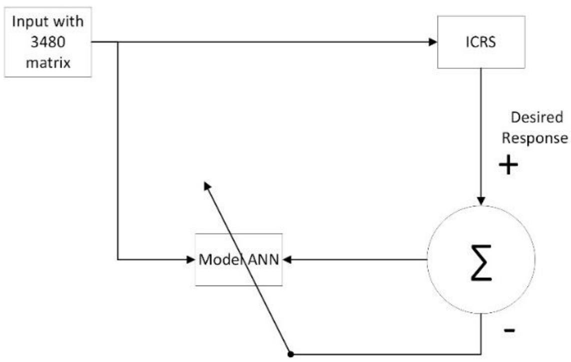


**Figure 7.** Training block diagram of the ANN model [40]

Each character is defined as a row matrix with 3480 inputs and 10 outputs for ANN training. The block diagram realized in this context is shown in

Figure 7. Thus, a model ANN that successfully recognizes a character given to the input has been obtained.

### 3.2.4. TR IN Classification with Proposed CNN

Unlike the ANN method used in Section 3.2.3, the methods commonly used in the current literature were investigated and as a result of this research, it was determined that there is a tendency towards Convolutional neural networks. For this reason, CNN-based methods were focused in this study. In this study, a new CNN model consisting of 15 layers is proposed. In the convolution layers of the proposed model, the ReLU activation function presented in Eq. 9 is used. The x value presented in Eq. 9 represents the input value.

$$f(x), = max(0, x) \quad (9)$$

Thus, the values obtained were kept within certain limits. In the study, unlike the literature, 3 new modules connected consecutively were developed. In this proposed module, inspired by the Inception module, the 1x1 convolution process was first applied. In the second step, 3x3, 5x5, 7x7, Mixed Pooling (2x2), Max Pooling (2x2) and Soft Pooling (2x2) operations were applied to the data obtained in parallel.

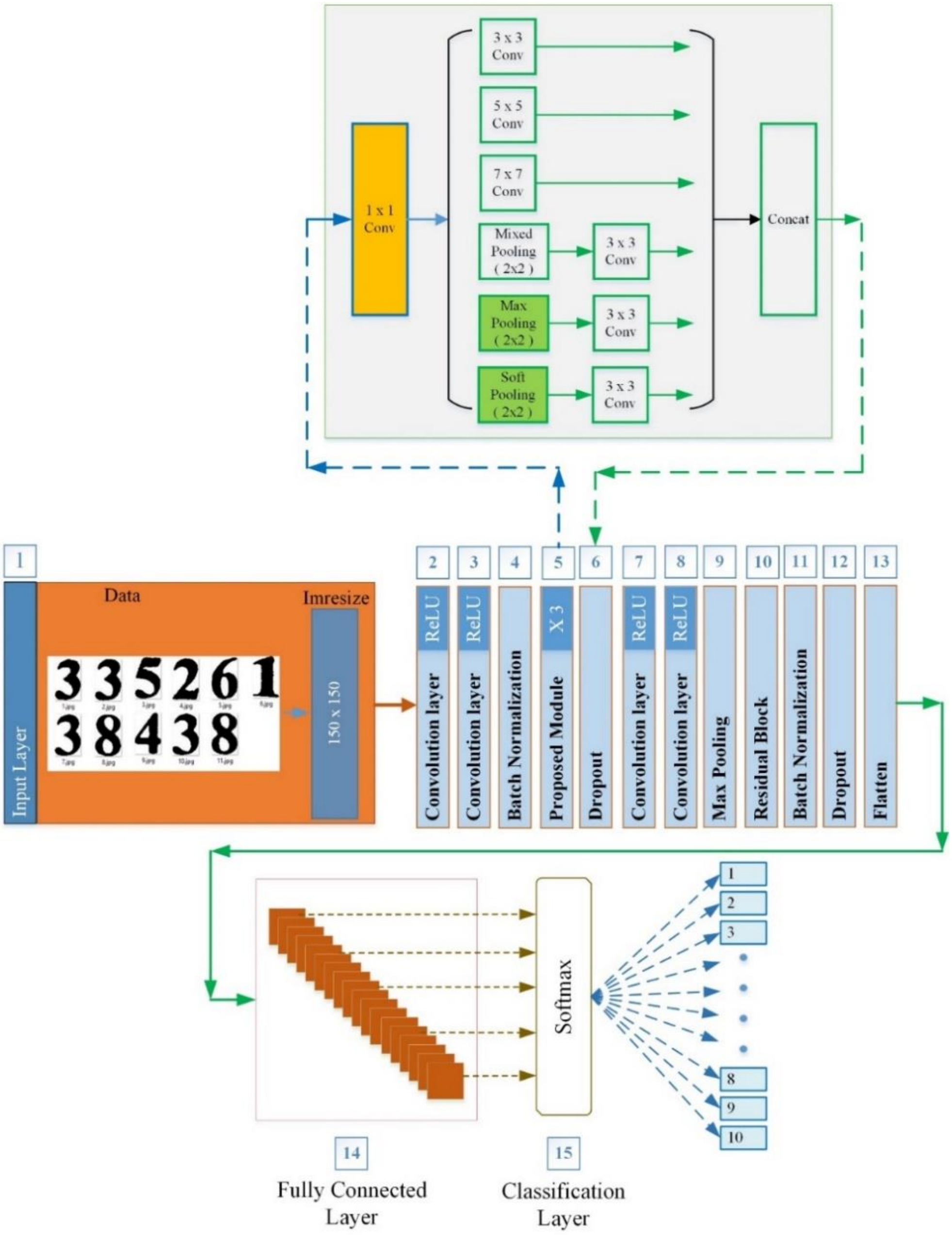


**Figure 8.** General structure of the proposed CNN model for OCR

The 3x3 convolution process was applied again to the data obtained from the pooling processes. As shown in Figure 8, new features were obtained by applying Concat process to the data obtained as a result of the mentioned processes. The data obtained with these methods were sent to the same module again in the next step and these processes were repeated 3 times in a row. Among the features obtained, there are very strong features as well as weak features. For this reason, Dropout layer was added to the proposed model. In the next step, convolution process was applied to the features obtained from the dropout layer. Max Pooling process was applied to the values obtained after these steps. Residual operation, which is the cornerstone of ResNet architecture, was applied to the values obtained after the pooling process. At this stage, dropout layer was applied due to too many features being obtained again and then Flatten operation was performed.

$$softmax(z_j) = \frac{e^{z}j}{\sum_{k=1}^{K} e^{z}k}, for\ j = 1, \dots, K \qquad (10)$$

In the last stages, fully connected layer and classification layer are added as in classical convolutional operations. In the classification layer, the SoftMax function presented in Eq. 10 is used. The expression $K$ in Eq. 10 expresses the number of classes. The $z_j$ value obtained as a result of these operations is the value that shows which class the data belongs to. The largest value obtained at which $K$ value means that the data belongs to that class.

## 4. Results and Discussion

A study is conducted to analyse the success rates in the classification of number features obtained with WT of ANN algorithm. This study is carried out in MATLAB 7.14 version. Experimental results have been obtained in Windows 10 operating system which is Intel Core i5 CPU (2.7 GHz) processor with 4GB RAM.

A classifier is trained to identify differences or types of samples among separated samples in classification. Task of the classifier demonstrates whether or not represented a specific object or area of interest in image. To achieve this, prior classifier is sufficient knowledge about the image. WT is used to feature extraction from images. WT is a feature extraction algorithm. 648 features are obtained with WT.

The numbers obtained from different TR IC are given training results using different types of wavelets and the different training algorithms of ANN in the previous my study. These results show that in Table 3 and Table 4. In addition to, even if oblique TR IN, the ICRS carries out successfully localization and detection of TR IN. The neural network training algorithms such as Scaled Conjugate Gradient classification (SCG), Flexible Back Propagation (FBG), Conjugate Gradient with Powell-Beale (CGPB), Fletcher-Reeves Conjugate Gradient (FRCG), Conjugate Gradient with Polak-Rbiere (PRCG) are used in classification. These training algorithms are used features which are extracted with different wavelet types of TR IC images. Results are obtained from using these training algorithms which are given in Table 3 and Table 4. In separation of high and low frequency components of the signal, Daubechies (1, 2) which is a type of mother wavelet, Symlet (3), Biortogonal (2.2, 3.1, 3.9), Reverse Biorthogonal (3.1), Meyer wavelet are wavelet types. These types are examined with each training algorithm and training/test times are given in Table 3.

**Table 3.** ANN training times of features obtained by the wavelet transform in different types and grades of the figures

| Method | Datasets | Training Times (second) of Dataset | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Db1* | *Db2* | *Bior 2.2* | *Bior 3.1* | *Bior 3.9* | *Sym3* | *dmey* | *Rbio 3.1* |
| SCG | Dataset-I | 65 | 68 | 70 | 65 | 80 | 70 | 76 | 67 |
| | Dataset-II | 66 | 69 | 71 | 66 | 81 | 71 | 77 | 68 |
| | Dataset-III | 131 | 137 | 141 | 131 | 161 | 141 | 153 | 135 |
| | Dataset-IV | 390 | 480 | 420 | 456 | 402 | 390 | 480 | 420 |
| FBG | Dataset-I | 100 | 102 | 106 | 99 | 108 | 113 | 105 | 109 |
| | Dataset-II | 101 | 101 | 107 | 98 | 109 | 112 | 106 | 108 |
| | Dataset-III | 201 | 203 | 213 | 197 | 217 | 225 | 211 | 217 |
| | Dataset-IV | 594 | 648 | 678 | 630 | 654 | 594 | 648 | 678 |
| CGPB | Dataset-I | 39 | 41 | 38 | 43 | 56 | 40 | 60 | 40 |
| | Dataset-II | 38 | 42 | 37 | 44 | 55 | 41 | 59 | 41 |
| | Dataset-III | 77 | 83 | 75 | 87 | 111 | 81 | 119 | 81 |
| | Dataset-IV | 258 | 336 | 240 | 360 | 240 | 258 | 336 | 240 |
| FRCG | Dataset-I | 98 | 103 | 80 | 103 | 120 | 103 | 125 | 117 |
| | Dataset-II | 98 | 103 | 80 | 103 | 120 | 103 | 125 | 117 |
| | Dataset-III | 196 | 206 | 160 | 206 | 240 | 206 | 250 | 234 |
| | Dataset-IV | 618 | 720 | 618 | 750 | 702 | 618 | 720 | 618 |
| PRCG | Dataset-I | 99 | 98 | 76 | 98 | 110 | 110 | 120 | 100 |
| | Dataset-II | 100 | 98 | 75 | 99 | 110 | 109 | 121 | 100 |
| | Dataset-III | 199 | 196 | 151 | 197 | 220 | 219 | 241 | 200 |
| | Dataset-IV | 588 | 660 | 660 | 720 | 600 | 588 | 660 | 660 |
| Average | Dataset-I | 80.2 | 82.4 | 74 | 81.6 | 94.8 | 87.2 | 97.2 | 86.6 |
| | Dataset-II | 80.6 | 82.6 | 74 | 82 | 95 | 87.2 | 97.6 | 86.8 |
| | Dataset-III | 160.8 | 165 | 148 | 163.6 | 189.8 | 174.4 | 194.8 | 173.4 |
| | Dataset-IV | 489.6 | 568.8 | 523.2 | 583.2 | 519.6 | 489.6 | 568.8 | 523.2 |

**Table 4.** ANN training performance results of features obtained by the wavelet transform in different types and grades of the figures

| Wavelet | Dataset-I | | Dataset-II | | Dataset-III | | Dataset-IV | |
|---|---|---|---|---|---|---|---|---|
| Type | TrRR(%) | TsRR(%) | TrRR(%) | TsRR(%) | TrRR(%) | TsRR(%) | TrRR(%) | TsRR(%) |
| Db 1 | 100 | 97.7 | 94.3 | 92.2 | 96.4 | 93.4 | 100 | 99.2 |
| Db 2 | 99.5 | 98.3 | 93.1 | 91.8 | 95.3 | 92.9 | 99.8 | 99.4 |
| Bior 2.2 | 100 | 99.4 | 93.7 | 93.4 | 95.3 | 96.2 | 100 | 99.8 |
| Bior 3.1 | 99.4 | 95.4 | 92.9 | 90.4 | 94.8 | 92.6 | 99.8 | 98.7 |
| Bior 3.9 | 98.6 | 96.2 | 91.4 | 91.4 | 93.7 | 93.5 | 99.7 | 99.3 |
| Sym 3 | 98.9 | 97.1 | 91.2 | 90.4 | 94.4 | 93.3 | 99.8 | 99.5 |
| Dmey | 99.3 | 94.8 | 92.8 | 89.5 | 95.7 | 93.4 | 99.8 | 98.4 |
| Rbio 3.1 | 99.7 | 98.2 | 93.2 | 92.4 | 96.4 | 95.3 | 100 | 99.8 |

While TrRR in Table 4 represents the training recognition rate, TsRR represents the Test Recognition Rate. The most preferable type of wavelets is determined according to the best training and testing rates in Table 3 and Table 4. This type is wavelet type which is called Bior 2.2.

**Table 5.** Proposed CNN performance results for different datasets

| Datasets | | F1 score | Recall | Precision | Accuracy |
|---|---|---|---|---|---|
| Dataset-I | Training | 0.9936 | 0.9896 | 0.9977 | 0.9924 |
| Dataset-I | Validation | 0.9896 | 0.9850 | 0.9943 | 0.9902 |
| Dataset-II | Training | 0.9511 | 0.9467 | 0.9556 | 0.9532 |
| Dataset-II | Validation | 0.9515 | 0.9435 | 0.9598 | 0.9586 |
| Dataset-III | Training | 0.9609 | 0.9552 | 0.9667 | 0.9688 |
| Dataset-III | Validation | 0.9607 | 0.9523 | 0.9694 | 0.9694 |
| Dataset-IV | Training | 0.9917 | 0.9887 | 0.9948 | 0.9950 |
| Dataset-IV | Validation | 0.9881 | 0.9843 | 0.9921 | 0.9927 |

In this study, CGPB training algorithm type is preferred to classify of features extracted with Bior 2.2 wavelet type, because it is completed the classification at the earliest time. Then, FDR is used in order to improvement these results obtained. Finally, selected features are classified using proposed CNN and ANN classification algorithms in real-time. The results of this classification are evaluated and compared.

When the data presented in Table 4 and Table 5 are evaluated together, it is observed that both models give good performance results. In accordance with these performance results, it is seen that both models are at a level that can in competition with the studies in the literature.

## 5. Conclusion and Suggestions

In the presented work, TR IN is recognised in real time by developing ICRS. The ICRS project based on real-time machine vision system to accelerate works of the institutions employees. The localization, detection and segmentation steps of the TR IN have an important role in the successful conclusion. CGPB training algorithm type is preferred to classify of features extracted with Bior 2.2 wavelet type, because it is completed the classification at the earliest time. In addition to, training recognition rate and test recognition rate with Bior 2.2. wavelet type achieves classification rate with 100% and 99.4% respectively for the Dataset-I.

In the Dataset-IV, the highest values achieved with ANN were obtained using Rbio 3.1 and Bior 2.2 wavelet types. These values were found to be 100% and 99.8% accuracy for training and testing, respectively. As can be seen from this result, the performance of the Rbio 3.1 wavelet type has also improved as a result of the expansion of the dataset. Similarly, performance improvement was also observed in other wavelet types. In the proposed CNN model, 99.24 %, 98.96 %, 99.77%, 99.77%, and 99.24% values were obtained for F1 score, recall, precision, and accuracy metrics in training for the Dataset-I, respectively. For validation, 99.12%, 98.50%, 99.43%, and 99.02% values were obtained in the same order. In the Dataset-IV, 99.49%, 98.87%, 99.48%, and 99.50% values were obtained in the same order for training. Similarly, for validation, values of 99.13%, 98.43%, 99.21%, and 99.27% were obtained in the same order.

An important limitation of the study is that max pooling, soft pooling, mixed pooling functions are used simultaneously. In the literature research, it is seen that there are different pooling functions. The contribution of the study to the literature is that soft pooling, mixed pooling and max pooling functions are used simultaneously. In future studies, different pooling functions will be tested and their effects on performance will be examined in detail.

As can be understood from these values, the performance of the proposed model in optical character recognition is promising. It is anticipated that the proposed model can also be applied in advanced embedded systems in future studies.

## Contributions of the authors

Çetiner: Literature review, Validation, Conceptualization, Formal analysis, Methodology, Software development, Visualization, Writing, review and editing
Cetişli: Supervision

## Conflict of Interest Statement

There is no conflict of interest between the authors.

## Statement of Research and Publication Ethics

- This study is derived from the master's thesis titled Optical Character Recognition Using Image Processing Techniques, completed at Süleyman Demirel University, Graduate School of Applied and Natural Sciences, Computer Engineering Department [40].
- A small part of this work was presented as a full-text paper titled "Real time recognition of identification cards of Turkish Republic with wavelet transforms" at the 21st Signal Processing and Communications Applications Conference (SIU) in 2013 [45])

## References


[1] X.-F. Wang, Z.-H. He, K. Wang, Y.-F. Wang, L. Zou, and Z.-Z. Wu, “A survey of text detection and recognition algorithms based on deep learning technology,” *Neurocomputing*, vol. 556, pp. 126702, 2023.

[2] R. Mohapatra, B. Majhi, and S. Jena, “Printed Odia Digit Recognition Using Finite Automaton,” in *Proc. 3rd Int. Conf. Adv. Comput. Netw. Informatics*, 2016, pp. 643–650.

[3] S. Ghosh, N. Das, M. Kundu, and M. Nasipuri, "Handwritten Oriya Digit Recognition Using Maximum Common Subgraph Based Similarity Measures," in *Inf. Syst. Des. Intell. Appl.*, 2016, pp. 165–173.

[4] A. Ghahnavieh and A. Raie, "A New Hierarchical Architecture Based on SVM for Persian License Plate Character Recognition," *J. Adv. Comput. Res.*, vol. 7, no. 1, pp. 49–66, 2016.

[5] S. Kaur, "An Automatic Number Plate Recognition System under Image Processing," *Int. J. Intell. Syst. Appl.*, vol. 8, no. 3, pp. 14, 2016.

[6] S. Babu and Z. Masood, "Android Based Optical Character Recognition for Noisy Document Images," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 1, pp. 34, 2016.

[7] G. Joseph and T. Singh, "Registration Plate Recognition Using Dynamic Image Processing and Genetic Algorithm," in *Innovations in Computer Science and Engineering*, 2016, pp. 37–43.

[8] I. Ahmad, S. Mahmoud, and G. Fink, "Open-vocabulary recognition of machine-printed Arabic text using hidden Markov models," *Pattern Recognit.*, pp. 97–111, 2016.

[9] H. Cecotti, "Active graph based semi-supervised learning using image matching: Application to handwritten digit recognition," *Pattern Recognit. Lett.*, pp. 76–82, 2016.

[10] H. Sajedi, "Handwriting recognition of digits, signs, and numerical strings in Persian," *Comput. Electr. Eng.*, vol. 49, pp. 52–65, 2016.

[11] S. Karthikeyan, A. G. S. de Herrera, F. Doctor, and A. Mirza, "An OCR Post-Correction Approach Using Deep Learning for Processing Medical Reports," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2574–2581, 2022.

[12] S. Naz, S. Ahmed, R. Ahmad, and M. Razzak, "Arabic Script based Digit Recognition Systems," in *Int. Conf. Recent Adv. Comput. Syst.*, 2016, pp. 67-73.

[13] V. Aradhya, G. Kumar, and S. Noushath, "Robust unconstrained handwritten digit recognition using radon transform," in *Signal Process. Commun. Netw.*, 2007, pp. 626–629.

[14] H. Boveiri, "Transformation-Invariant Classification of Persian Printed Digits," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 4, no. 3, pp. 153–164, 2011.

[15] R. Azad, F. Davami, and H. Shayegh, "Recognition of Handwritten Persian/Arabic Numerals Based on Robust Feature Set and K-NN Classifier," in *2013 First Int. Conf. Comput. Inf. Technol. Digit. Media (CITADIM Proceeding-Scientific)*, 2014.

[16] M. Ghaleb, L. George, and F. Mohammed, "Printed and Handwritten Hindi/Arabic Numeral Recognition Using Centralized Moments," *Int. J. Sci. Eng. Res.*, vol. 5, no. 3, pp. 140–144, 2014.

[17] C. Liu and C. Suen, "A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters," *Pattern Recognit.*, vol. 42, no. 12, 2009.

[18] I. Meymand, "Recognition of Handwritten Persian/Arabic Numerals Based on Robust Feature Set and K-NN Classifier," *Int. J. Comput. Inf. Technol.*, vol. 1, no. 3, pp. 220–230, 2013.

[19] D. Sharma, G. Lehal, and P. Kathuria, "Digit extraction and recognition from machine printed Gurmukhi documents," in *Proc. Int. Work. Multiling. OCR*, 2009.

[20] S. Metlek and H. Çetiner, "ResUNet+: A New Convolutional and Attention Block-Based Approach for Brain Tumor Segmentation," *IEEE Access*, vol. 11, pp. 69884–69902, 2023.

[21] S. Metlek, "CellSegUNet: an improved deep segmentation model for the cell segmentation based on UNet++ and residual UNet models," *Neural Comput. Appl.*, vol. 36, pp. 5799-5825, 2024.

[22] H. Çetiner and S. Metlek, "DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 8, pp. 101663, 2023.

[23] S. Metlek, "A new proposal for the prediction of an aircraft engine fuel consumption: a novel CNN-BiLSTM deep neural network model," *Aircr. Eng. Aerosp. Technol.*, vol. 95, no. 5, pp. 838–848, Jan. 2023.

[24] J. N. and S. K. A. M, "A Novel Text Recognition Using Deep Learning Technique," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6.

[25] L. Yang, D. Ergu, Y. Cai, F. Liu, and B. Ma, "A review of natural scene text detection methods," *Procedia Comput. Sci.*, vol. 199, pp. 1458–1465, 2022.

[26] R. Krithiga, S. Varsini, R. G. Joshua, and C. U. O. Kumar, "Ancient Character Recognition: A Comprehensive Review," *IEEE Access*, pp. 1, 2023.

[27] P. Batra, N. Phalnikar, D. Kurmi, J. Tembhurne, P. Sahare, and T. Diwan, "OCR-MRD: performance analysis of different optical character recognition engines for medical report digitization," *Int. J. Inf. Technol.*, vol. 16, no. 1, pp. 447–455, 2024.

[28] L. M. Low, F. H. M. Salleh, Y. F. Law, and N. Z. Zakaria, "Detecting and recognizing seven segment digits using a deep learning approach," in *ITM Web of Conferences*, 2024, pp. 1007.

[29] A. Singh, S. Jangra, and G. Aggarwal, "EnvisionText: Enhancing Text Recognition Accuracy through OCR Extraction and NLP-based Correction," in *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2024, pp. 47–52.

[30] S. Gujjeti, M. S., and V. G., "A Systematic Investigation on Different Scene Text Detection and Recognition Method using Deep Learning Techniques," in *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 2024, pp. 1–5.

[31] A. A. Chandio, M. Asikuzzaman, and M. R. Pickering, "Cursive character recognition in natural scene images using a multilevel convolutional neural network fusion," *IEEE Access*, vol. 8, pp. 109054–109070, 2020.

[32] R. Deore, B. Gawali, and S. Mehrotra, "Digit Recognition System Using EEG Signal," in *Brain-Computer Interfaces*, Cham: Springer Int. Publ., 2015, pp. 375–416.

[33] J. Park and Y.-B. Kwon, "An Embedded OCR: A Practical Case Study of Code Porting for a Mobile Platform," in *2009 Chinese Conference on Pattern Recognition*, 2009, pp. 1–5.

[34] A. Canedo-Rodriguez, S. Kim, J. H. Kim, and Y. Blanco-Fernandez, "English to Spanish translation of signboard images from mobile phone camera," in *IEEE Southeastcon 2009*, 2009, pp. 356–361.

[35] M. Nava-Ortiz, W. Gómez, and A. Díaz-Pérez, "Digit recognition system for camera mobile phones," in *8th International Conference on Electrical Engineering, Computing Science and Automatic Control*, Merida City, Mexico, 2011, pp. 1-5.

[36] R. Shah and T. Ratanpara, "A Mean-Based Thresholding Approach for Broken Character Segmentation from Printed Gujarati Documents," in *Proc. Second Int. Conf. Comput. Commun. Technol.*, 2016, pp. 487–496.

[37] P. Singh, A. Verma, and N. Chaudhari, "Performance Evaluation of Classifier Combination Techniques for the Handwritten Devanagari Character Recognition," *Inf. Syst. Des. Intell. Appl.*, pp. 651–662, 2016.

[38] P. Singh, R. Sarkar, and M. Nasipuri, "A Study of Moment Based Features on Handwritten Digit Recognition," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, pp. 1-17, 2016.

[39] G. C. Batista and W. L. S. Silva, "Application of support vector machines to recognize speech patterns of numeric digits," in *2015 11th International Conference on Natural Computation (ICNC)*, 2015, pp. 831–836.

[40] H. Çetiner, "Görüntü işleme teknikleri kullanarak optik karakter tanımlama." Yüksek lisans tezi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği AnaBilim Dalı, Isparta, Türkiye, 2012.

[41] A. Jaiswal, "Standard OCR Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/preatcher/standard-ocr-dataset. [Accessed: Jan. 1, 2024].

[42] A. F. de Sousa Neto, B. L. D. Bezerra, G. C. D. de Moura, and A. H. Toselli, "Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review," *SN Comput. Sci.*, vol. 5, no. 2, pp. 258, 2024.

[43] M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi, "Wavelet toolbox," Matlab User's Guid., 1997. [Online]. Available: https://feihu.eng.ua.edu/NSF_TUES/w7_1a.pdf.

[44] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[45] H. Çetiner and B. Cetişli, "Real time recognition of identification cards of Turkish Republic with wavelet transforms," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, Haspolat, Turkey, 2013, pp. 1-4.