Research Article/Araştırma Makalesi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

# Image Based Web Page Classification by Using Deep Learning

Muhammed Mutlu YAPICI*[a]

Submitted: 26.10.2023 Revised: 15.11.2023 Accepted: 22.11.2023 doi:10.30855/gmbd.0705N06

## ABSTRACT

The internet holds a significant role in all aspects of our lives, and its importance continues to grow each day. Therefore, the usability of the Internet holds great significance. Low data quality and disinformation severely impact the usability of the internet. Consequently, people face challenges in obtaining accurate and clear information. In the present day, websites predominantly feature image-based content like pictures and videos, as opposed to text-based content. The classification of such content holds immense importance for search engines. As a result, the classification of web pages stands as a crucial research area for scholars. This study focuses on the classification of image-based web pages. A deep learning-based approach is proposed to categorize web pages into four main groups: tourism, machinery, music, and sports. The suggested method yielded the most favourable outcomes when utilizing the Stochastic Gradient Descent (SGD) optimization method, achieving an accuracy of 0.9737, a recall of 0.9474, an F1 score of 0.9474, and an Area Under the ROC Curve (AUC) value of 0.9649. Furthermore, the utilization of Deep Learning (DL) led to achieving the most advanced results in web page classification within the existing literature, particularly on the WebScreenshots dataset.

**Keywords:** Deep learning, web page classification, DenseNet, optimization methods

[a,*] Ankara University, Elmadag Vocational School, Dept. Of Computer Technologies 06780 - Ankara, Türkiye
Orcid: 0000-0001-6171-1226
e mail: mutluyapici@ankara.edu.tr

*Corresponding author: mutluyapici@ankara.edu.tr

# Derin Öğrenme Kullanarak Görüntü Tabanlı Web Sayfası Sınıflandırma

## ÖZ

İnternet hayatımızın her alanında önemli bir yere sahip ve önemi her geçen gün artmaya devam ediyor. Bu nedenle internetin kullanılabilirliği büyük önem taşımaktadır. Düşük veri kalitesi ve dezenformasyon, internetin kullanılabilirliğini ciddi şekilde etkilemektedir. Bu nedenle insanlar doğru ve temiz bilgiye ulaşma konusunda zorluklarla karşılaşmaktadır. Günümüzde web sitelerinde metin tabanlı içerik yerine ağırlıklı olarak resim ve video gibi görsel tabanlı içerikler daha çok yer almaktadır. Bu tür içeriklerin sınıflandırılması arama motorları için büyük önem taşımaktadır. Sonuç olarak web sayfalarının sınıflandırılması bilim insanları için önemli bir araştırma alanı olarak karşımıza çıkmaktadır. Bu çalışma görsel tabanlı web sayfalarının sınıflandırılmasına odaklanmaktadır. Web sayfalarını turizm, makine, müzik ve spor olmak üzere dört ana grupta sınıflandırmak için derin öğrenmeye dayalı bir yöntem önerilmiştir. Önerilen yöntem, 0,9737 accuracy, 0,9474 recall, 0,9474 F1-score ve 0,9649 AUC değeriyle en iyi sonuçları Stokastik Gradyan İnişi (SGD) optimizasyon yöntemi ile elde etmiştir. Ayrıca, Derin Öğrenmenin (DL) kullanılması, web sayfası sınıflandırmasında, özellikle WebScreenshots veri kümesinde, mevcut literatürdeki en iyi sonuçların elde edilmesini sağlamıştır.

**Anahtar Kelimeler:** Derin öğrenme, Web sayfası sınıflandırma, Densenet,Optimizasyon yöntemleri

**To cite this article:** M. M. Yapici, "Image Based Web Page Classification by Using Deep Learning," *Gazi Journal of Engineering Sciences*, vol.10, no.1, pp. 72-83, 2024. doi:10.30855/gmbd.0705N06

## 1. Introduction

The field of the Internet has been profoundly impacted by the advancement of technology. Internet technology, officially introduced in 1968 and laid out in 1969 as ARPANET, reached a milestone with the integration of emails into our lives and rapidly expanded worldwide[1,2]. Currently, these advancements and expansions continue to grow exponentially, permeating nearly every aspect of human life[3]. Another pivotal moment in the history of the internet was the global proliferation of social media platforms, the integration of the internet into mobile devices, and the interconnection of all electronic devices through the Internet of Things (IoT). Presently, the internet holds vital importance across various domains, including health, education, commerce, and entertainment[4]. Furthermore, it is anticipated that the next significant breakthrough in the internet realm will involve the Internet of Humans (IOH), further solidifying its crucial role[5].

The usability of internet technology, which is progressively becoming more significant for individuals, is growing more challenging due to the escalating data volume and the prevalence of misinformation. Users often encounter difficulties when searching for specific information or products on the internet. Moreover, numerous companies aiming to establish a presence in the online market face challenges in attracting customers to their web pages. It is important to be able to access quickly the required web pages. Text based search approaches fail to list the web pages users are searching for. Therefore, image-based web page searching, and classification models are important to address the problem. Consequently, the classification of web pages has emerged as a complex research area. Although there are attempts within the literature to address this issue, the number of studies in this field remains relatively limited.

Web pages possess numerous attributes, including URL addresses, text content, hyperlinks, images, domain names, server information, HTML tags, and semantic web tags. Consequently, the classification methods developed by researchers are built upon these attributes. In the literature, research on web page classification appears to be divided primarily into two main categories based on the classification data type: 1) text-based classification and 2) image-based classification. Text-based classification methods utilize data such as website content, HTML tags, and domains as inputs for the classification process. Conversely, image-based methods employ screenshots captured from website pages as inputs [6]. Within these classification groups, three notable challenges pose difficulties for researchers. The primary challenge revolves around the rapid expansion of data, rendering analysis increasingly complex. Many classification techniques prove inadequate and slow in handling this data complexity. Another significant obstacle stems from the abundance of attributes present on web pages, particularly in text-based classification. Web pages are essentially text files composed of HTML tags, where both content and design are defined by these tags. The task of segregating content from design, deciphering the page's language, and eliminating extraneous words demands meticulous attention and intricate analysis. Inconsistencies between textual and raw data attributes further complicate the classification of web pages. Lastly, a critical issue pertains to extracting meaningful data for the classification process. On occasion, the content of a web page might not align with the site's purpose, posing a significant challenge for accurate classification. This discrepancy, whether intentional or inadvertent, contributes to the classification problem.

Text-based classifiers are subject to various limitations arising from data complexity and language-related problems. Given that images possess universal attributes, image-based methods offer a more effective and globally applicable solution. Additionally, in contemporary times, websites predominantly feature image-based content such as pictures and videos, as opposed to text-based content. The classification of these media elements bears immense significance for search engines. In light of the inadequacies of text-based classification methods, this study adopts a deep learning approach, a method with established success in various domains, to classify websites using image data. The proposed classification technique is rooted in a convolutional neural network (CNN) architecture for image analysis. The principal distinction of this novel model from existing methods in the literature lies in its training and testing using three distinct optimization

methods. This approach aims to address the challenges of feature extraction and data complexity, which represent the most formidable obstacles in web page classification. Contributions of the study are listed below:

1- Proposing a image based approach for web page classification,
2- Addressing to limitations of searching and classification of text based approaches,
3- It guides researchers to choose the best optimization method by comparing image-based classification methods,
4- It improves the literature results in the field of image-based web page classification.

## 2. Literature Review

Web page classification is the process of assigning a web page to one of the predetermined categories. These categories may be two categories such as malignant or benign, or they may consist of multiple categories such as e-commerce, education, entertainment etc. The increment of use of web pages and the desire to quickly access valuable data sought on the internet have made the problem of web page classification an important research topic. In order to overcome this problem, many researchers have tried to find solutions using different models and techniques. The studies about web page classification the literature going on two main groups according to the input data used by the researchers. These are 1) text-based, 2) image-based classification, as mentioned in the previous section. However, it is seen that text-based classification is also divided into 3 groups depending on the data type. First one is textual classification which is trying to classify the input data that consists of page contents, titles, HTML tags, domain name. Second one is graph-based classification, and it uses structural relationships of links and back linked web pages. Last one is named as other classification types and this types use server information and, semantic web data as input data[6-9].

When the studies are examined, it is seen that the most popular classification type for web page classification is text-based classification. The large number of features to be used has led researchers to prefer text-based classification methods. Many studies have been carried out to distinguish between adult and child-oriented sites by using text-based classification. Alvari et al.[10] used support vector machines to classify adults and children oriented web page with 12 different textual features. Likewise, Ahmadi et al.[11] used the decision tree structure for classification of pornographic web pages. Some researchers used both textual and graph-based data for web page classification. To address web page classification problem, Sun et al.[12], Qi et al.[13], Tian et al.[14] proposed a system by using support vector machines and, Kwon and Lee[15], Celado et al.[16] proposed other one by using K-nearest neighbor algorithms. Xia et al.[8] used HTML tag, header information and URL as input data for classification. They tried to classify five different web page categories which composed of health to sports. At last decades Deep Learning (DL) methods, which have been achieved state-of-arts results in many fields[17-22] were also used in web page classification. Lin[23] performed web page classification with combination of graph-based and text-based data by using recurrent neural networks (RNN) and deep residual neural networks (ResNet). Buber and Diri [24] reported that they achieved 85% success on RNNs. Evolution-based genetic algorithms and fuzzy logic methods are other robust methods which presented in the literature for text-based web page classification problem[25-28]. One of the biggest limitations of text-based classification methods is that the web page contents are differ according to the languages of the countries and even local regions. So, this makes language the main problem of the text-based classification.

Especially at last decades, image data is used as much as text data in web page classification approaches. In the literature it is seen that image features such as histogram, color information, edge detection and some other features obtained via image filter, are primarily preferred for classification process. De Boer et al. [29] used image features such as histogram, color beam, edge histogram information, Gabor filter and Tamura attributes to classify websites in two categories as beautiful and ugly according to their aesthetic appearance. They chose the Naive Bayes method as the classification method and reported that they achieved 83% success. Ugalde [30], which deals with the mobile-based web page classification problem in the thesis study conducted in 2015, used both text-based and image-based features. As image-based features, he also classified web pages

by analyzing histogram, color beam, edge histogram information, Gabor filter and Tamura features on WEKA [30]. In another study which used both text-based and image-based features, the researchers carried out classification by using deep learning method, on their own dataset that created by themselves. The authors encoded the text-based features as images and then classified the images. They used the bag of words method to encode textual data as images and, reported a success rate of 93.7% was achieved [31]. Li et al.[32] captured the web pages by using their own application, and then they used image features to classify web sites of gambling and sexual. The authors classified these features by using support vector machines. Susaki et al.[33] classified the screenshots of the tourism web sites' in order to create a tourism recommendation guide. They made a classification according to the features that obtained by performing color analysis. In other study, Abdali et al. [34] used more than 50 thousand images obtained from 500 different sites for classify web pages by using deep learning network. All this research in literature show that each researcher collected and created their own dataset independently from each and performed web page classification by using their own dataset. Moreover, it is seen that they haven't shared the datasets publicly. Thus, there is no possibility of comparing results of the methods with each other. It is not also possible for the studies carried out in this way to reveal their superior sides by comparing them with each other[7]. Only giving obtained results in the studies proves that they didn't compare the result with other results in the literature. However, Aydos et al.[6] created and shared a new web page dataset composed of 20 thousand sites that obtained from Google searches. Name of the published dataset is WebScreenshots. They tested the dataset by using deep learning methods. The authors reported that they achieved 94.90% classification success on 4 classes. So, in order to compare the success of the proposed method with other studies in the literature and then detect the state-of-art study, the WebScreenshots dataset was used in this study.

## 3. Proposed Method

In this study, an Artificial Neural Networks (ANN)-based classification system has been proposed for website classification. The system comprises two main components: feature extraction and the classification process from images. Initially, a deep neural network system based on Convolutional Neural Networks (CNN) is utilized to extract image features, followed by the execution of the classification process. While numerous CNN models exist in the literature, three models have emerged as particularly notable: VGG, ResNet, and DenseNet. These models have demonstrated considerable success and find applications in various domains. Research has consistently indicated that the DenseNet model outperforms other CNN models [35]. Therefore, for the classification phase of this study, the DenseNet121 model was chosen.

After the classification phase, in order to enhance the efficacy of the proposed method, three distinct optimization methods were employed during the fine-tuning stage. These optimization techniques encompass Stochastic Gradient Descent (SGD), Adam, and AdaGrad. The architecture of the proposed system is visually represented in Figure 1.

To best of our knowledge, there is no publicly available data set used by researchers in the literature in this field. It is seen that each researcher carries out his studies by creating his own database. In this study, a publicly shared dataset that created by Aydos et al.[6] was used to compare the obtained results and by the way clearly detect the best accurate web page classification system in the literature.

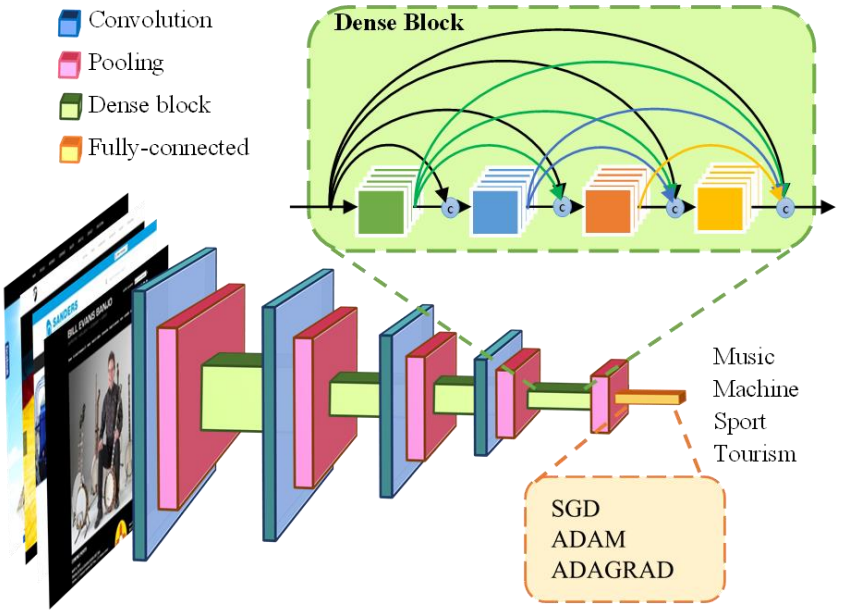


Figure 1. Architecture of proposed web page classification system

In the experiments, the success of the system was measured via 4 different metrics: Accuracy, Recall, F1 score and Area Under Curve (AUC). The classification results were obtained separately for each class. The definitions and equation of the metrics are shown in equation 1, equation 2, equation 3 and equation 4, respectively.

Accuracy: It refers to the proportion of correctly detected data in the entire test dataset. It is defined as in equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Recall: It is the ratio of positive samples predicted correctly by the model to all positive samples in the data set. It is defined as in equation 2.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1-Score: It is the harmonic mean of Precision and Recall values. The F1-Score formula is as follows in equation 3.

$$F1 - Score = 2\,x\,\left(\frac{Precision\,x\,Recall}{Precision+Recall}\right) \quad (3)$$

Area Under Curve (AUC): The area under the ROC curve is expressed as AUC. It is defined as in equation 4.

$$AUC = \int_0^1 \frac{TP}{TP+FN}\, d\,\frac{FP}{TN+FP} \quad (4)$$

### 3.1. Convolutional neural network

The Convolutional Neural Network (CNN), initially introduced by LeCun et al. [36], is an image processing technique characterized by two core attributes: 1) spatially shared weights and 2) spatial pooling. In 1998, the same research group advanced this concept by introducing the LeNet-5 architecture, a 7-layer CNN design, for recognizing handwritten digits on bank check images, which were resized to 32x32 digital images [37].

Presently, CNN stands as the most prevalent deep learning architecture for feature extraction, particularly in tasks such as image classification and object recognition. Fundamentally, CNNs [36] constitute a specialized and enhanced variant of neural networks, boasting multiple layers that have markedly revolutionized the realm of image processing.

A CNN architecture, illustrated in Figure 2, comprises three primary layers: the convolutional layer, the subsampling layer (also referred to as the pool layer), and the fully connected layer. This architecture has ushered in significant advancements in the domain of image processing.

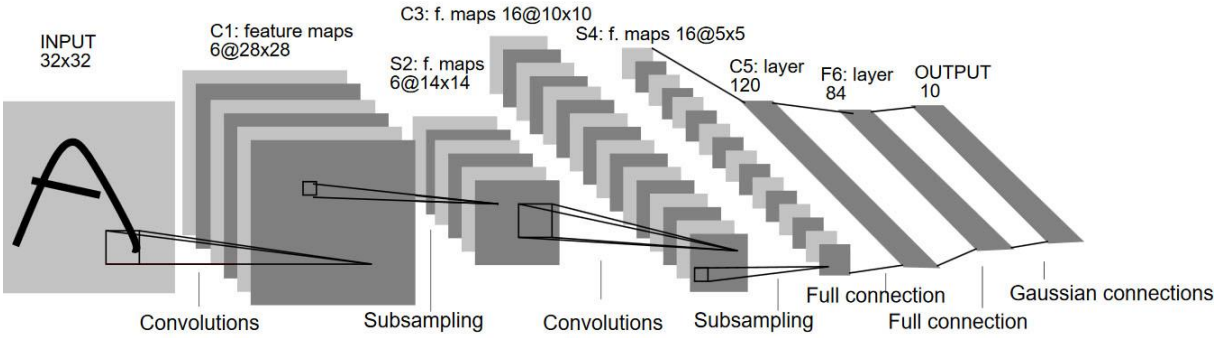


Figure 2. Basic CNN architecture [36]

CNN aims to learn abstract features by applying filters on the input image in the convolution layer and then subsampling the abstracted data in the pool layer. In the convolution layer, the convolution operation is performed by shifting the filter data matrix over the input data matrix. Thus, the features of the input data are obtained. In the first layer, these filters respond and sample edges or spots of color, while in the last layer they begin to sample shapes and parts of objects [38]. The convolution is carried out as a result of the multiplication of the input and output matrices and adding a bias term on the result. Figure 3 shows the basic convolution structure. And the basic formula that represent the convolution process is also given in equation 5. In the equation, the pixels of the filter (kernel), the pixels of the input image, the pixels of the output image, and the bias term are represented by w, x, y and b, respectively.

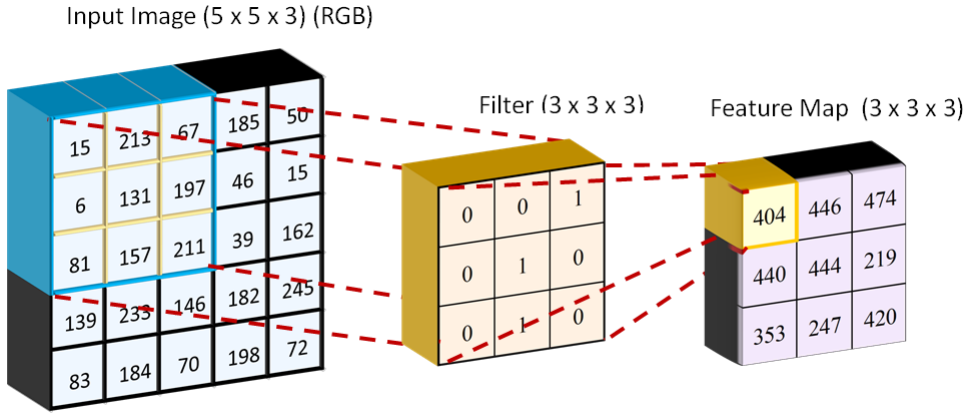


Figure 3. Basic convolution process, the convolution operation is obtained by multiplication of the input data matrix and the filter data matrix and then adding a bias term on it [39]

$$y_n = \sum_{n=1}^{9} (x_n . w_n + b_n) \quad (5)$$

Another layer that CNNs use is the pooling layer, also called the subsampling layer. Pooling [39] is used to reduce the size of feature obtained from the previous layer by moving into subsamples. The pooling layer activation function performs the size reduction process by getting maximum or average value of the obtained

features. The main purpose of the pooling layers is to gradually reduce the feature sizes of the representation and thus reduce the computational cost of the model by reducing the number of parameters [37].

An activation function is added at the end of each layer, and these functions may differ according to preference. Generally, Rectified Linear Unit (ReLU) activation is used at the end of the layer. Activation function aims to normalize result values. The implementation of the ReLU function is that the output of the function will be 0 if the input is less than 0, otherwise the input value will be sent as the output. That is, if the input is greater than 0, the output is equal to the input. The function of the basic process of ReLU is given in equation 6.

$$\mathrm{ReLU(x)} = \begin{cases} 0 & if\ x < 0 \\ x & if\ x \geq 0 \end{cases} \tag{6}$$

The last layer of a CNN is Fully Connected-(FC) layer. In the FC layer, each neuron at the previous layers is connected to every neuron in the next layers. FC layers are the basic building blocks of traditional neural networks. These layers are used to transform the activation feature maps into a one-dimensional feature vector where each value is associated with an object class. End of the FC layer, SoftMax activation function is applied to transform the FC layer results into a probability distribution indicating which class they belong to. By the way, fully connected layers transform the feature map of images into votes that represent the ratio of belongingness of every classes. These rates are expressed as weight or link strength between each value and each class [37,38].

### 3.2. Densenet

In the traditional CNN architecture, the feature map obtained from each convolutional layer is passed along to the subsequent layer as input. Consequently, every layer can solely glean knowledge from the feature map of the preceding layer. Unfortunately, this setup results in the latter layers being unable to directly access the raw features from the initial layers. This limitation poses challenges in effectively learning input data features and might lead to the loss of numerous features before they ultimately reach the classification layer.

In response to this issue, the ResNet [40] model introduces a novel approach by forwarding the outputs of the two previous convolutional layers as additional inputs to the current layer. By doing so, each layer gains insight from the feature maps of its two predecessors. Nevertheless, even within the ResNet framework, the features of prior layers still struggle to propagate deeply enough, contributing to ongoing feature losses.

In contrast, the DenseNet model refines the structure introduced by ResNet. In the DenseNet model, each layer possesses the capability to harness the feature maps of all preceding layers. This configuration signifies a notable improvement over the ResNet architecture, allowing for more efficient and comprehensive feature utilization across the network.

DenseNet, conceptualized by Huang et al. [41], arose from the insight that classical CNN models could achieve improved accuracy and efficiency through a reimagining of their design, specifically incorporating short connections between each preceding layer and the subsequent layers. Similar to ResNet, DenseNet incorporates these additional short connections, but with a notable distinction: these connections span not only between the current layer and the previous one but also encompass all layers in the network. This fundamental redesign aims to mitigate challenges such as the vanishing gradient problem during weight optimization, while simultaneously enhancing feature extraction, promoting feature reuse, and significantly curtailing the number of parameters involved.

In the classic N-layer CNN model, there exist N connections linking the layers, excluding the input layer. In contrast, the DenseNet model introduces a more intricate connectivity pattern, wherein each layer establishes

connections with all subsequent layers. As a result, every layer receives the feature maps from all preceding layers as input. This approach effectively tackles the vanishing gradient issue, facilitates robust feature extraction, and fosters feature sharing across the network. The DenseNet family encompasses several variations, including models with 121, 169, 201, and 264 layers. For this study, the DenseNet-121 model, comprising 121 dense layers, was selected. The architecture of the DenseNet model is visually presented in Figure 4.

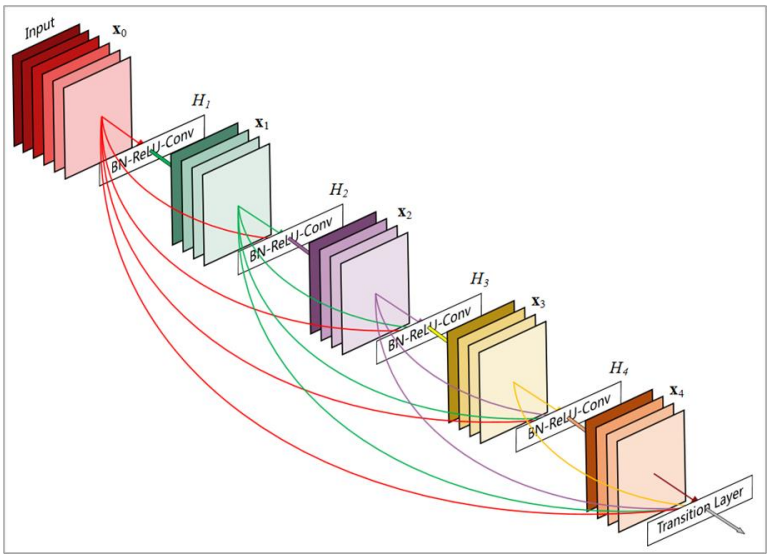


Figure 4. The architecture of the DenseNet [33,35]

## 4. Experiments and Discussion

This section describes the dataset, experimental setup, training and test phases. Subsequently, the results of the experiments are given and discussed. Finally, the obtained results are compared with the literature.

### 4.1. Dataset

The WebScreenshots database was developed by Aydos et al.[6] in 2019. The limitation of small databases in web page classification poses challenges in training complex systems like deep learning, which encompass a vast number of learning parameters. Additionally, the dynamic nature of web content over time complicates the comparison of classification results based on search engine-generated search results. To address these issues, researchers introduced the WebScreenshots database, which serves as an extensive collection of web pages.

This database comprises datasets for URL, content, and screenshots categorized into various classes. The screenshot dataset focuses on four distinct classes: machinery, music, sports, and tourism ("machinery," "music," "sport," "tourism"). The process involved listing the web pages from the DMOZ database into classes based on the screenshot dataset's categories and capturing screenshots of these pages. A total of 5000 web pages were captured for each class, culminating in a dataset containing 20,000 instances. Notably, the DMOZ database boasts universality and encompasses diverse web pages in different languages, totaling 44 different languages. Given that the study specifically pertains to image-based classification, solely the screenshots dataset was utilized for the research.

### 4.2. Results

As mentioned before, in the study, the categories of the web pages are estimated from the screenshots of the web sites. In the study, all experiments were carried out on the WebScreenshots Database, which consists of

20000 data belonging to 4 classes. There are 4 different subsets in the WebScreenshots database named Google10, Google20, Subset10 and Subset20. In this study, Google20, the most comprehensive data set, was used. The Google20 data group consists of 20000 data samples. Screenshots of each class are size of 224X224px. Each image is in RGB format. In the study, data augmentation methods such as rotation, mirroring, shifting, random cropping were used on all data. These methods were used only in the training phase. Data augmentation was not performed on the test data.

In the experiments, all data were split into two groups as 80% training and 20% testing. Each experiment carried out 100 epochs and each epoch consist of 32 batch sizes. The 5-fold cross validation (CV) method was applied in the experiments to reduce randomness and increase reliability. By this way we aimed to tackle of over fitting problem. Final results were obtained by getting average of the CV results. The study was carried out on the NVIDIA TITAN XP graphic card using Keras framework. As stated in the previous section, all experiments were performed on the DenseNet121 model. Experiments were repeated separately for Stochastic Gradient Descent-SGD, Adam and AdaGrad optimization methods. The success of the system was calculated separately for each class and each optimization method by using 4 different metrics: Accuracy, Recall, F1 score and AUC. The obtained results are given in table 1.

Table 1. Experimental results for each class and each optimization method

| | Classes | Accuracy | Recall | F1 Score | AUC |
|---|---|---|---|---|---|
| SGD | Machinery | 0,9754 | 0,9477 | 0,9507 | 0,9662 |
| | Music | 0,9685 | 0,9437 | 0,9374 | 0,9602 |
| | Sport | 0,9683 | 0,9329 | 0,9364 | 0,9565 |
| | Tourism | **0,9825** | 0,9652 | **0,9649** | **0,9767** |
| | Average of SGD | **0,9737** | **0,9474** | **0,9474** | **0,9649** |
| ADAM | Machinery | 0,9118 | 0,8272 | 0,8243 | 0,8836 |
| | Music | 0,8784 | 0,7403 | 0,7527 | 0,8323 |
| | Sport | 0,8731 | 0,7862 | 0,7560 | 0,8442 |
| | Tourism | 0,9361 | 0,8452 | 0,8686 | 0,9058 |
| | Average of ADAM | 0,8999 | 0,7997 | 0,8004 | 0,8665 |
| ADAGRAD | Machinery | 0,9776 | 0,9548 | 0,9552 | 0,9700 |
| | Music | 0,9679 | 0,9418 | 0,9361 | 0,9592 |
| | Sport | 0,9669 | 0,9266 | 0,9332 | 0,9534 |
| | Tourism | 0,9820 | **0,9654** | 0,9641 | 0,9765 |
| | Average of ADAGRAD | 0,9736 | 0,9471 | 0,9471 | 0,9648 |

Upon reviewing the obtained results, it becomes evident that the SGD (Stochastic Gradient Descent) and ADAGRAD optimization methods yielded highly similar outcomes. However, it's noteworthy that the SGD optimization method achieved the most successful result, albeit with a minor disparity. Specifically, the SGD method achieved an average AUC (Area Under the ROC Curve) of 0.9649, while ADAGRAD achieved an average AUC of 0.9648. In contrast, the ADAM optimization method demonstrated the least favorable performance within the WebScreenshots database.

Notably, the Tourism class emerged as the category with the most impressive results across the entire

database. This outcome can be attributed to the high distinctiveness observed within the samples belonging to this particular class. The inherent characteristics of the Tourism class samples appeared to facilitate more accurate and discerning classification outcomes.

Table 2. Comparison of the results with literature results (Elde edilen sonuçların literatür sonuçlarıyla karşılaştırılması)

| Paper | Model | Optimization Method | Accuracy | Recall | F1 Score | AUC |
|---|---|---|---|---|---|---|
| Aydos et. al.[6] | VGG16 | - | 0,9035 | - | - | - |
| Aydos et. al.[6] | DenseNet121 | - | 0,9300 | - | - | - |
| Aydos et. al. [6] | DenseNet169 | - | 0,9475 | - | - | - |
| Aydos et. al. [6] | DenseNet201 | - | 0,94450 | - | - | - |
| Proposed | DenseNet121 | Average of SGD | **0,9737** | **0,9474** | **0,9474** | **0,9649** |
| Proposed | DenseNet121 | Average of ADAM | 0,8999 | 0,7997 | 0,8004 | 0,8665 |
| Proposed | DenseNet121 | Average of ADAGRAD | 0,9736 | 0,9471 | 0,9471 | 0,9648 |

The comparison of the achieved results with those reported in the literature is outlined in Table 2. Upon analyzing the existing literature, it becomes evident that only one study has been conducted on the WebScreenshots database[6]. In this particular study, Aydos et al.[6]attained an accuracy rate of 0.9475 across the four classes. The authors conducted their experiments using a learning rate of 0.00001. However, they did not provide information regarding the optimization method employed in their experiments. Moreover, the authors exclusively presented results based on the accuracy metric, omitting any insights into other evaluation metrics.

Based on the results obtained in the current study, a clear distinction emerges: the proposed system, leveraging the SGD and ADAGRAD optimization methods, outperforms the previous study by Aydos et al. [6]. Notably, the success achieved through the proposed ADAM optimization method falls short of the performance recorded in the literature.

## 5. Conclusion

In the study, the classification of web page content was conducted using captured website images, totaling 20,000 images across four distinct classes. This classification task was executed employing a deep learning approach. Specifically, the DenseNet121 model, recognized for its success across diverse domains, was selected for the classification process, wherein it was paired with three different optimization methods: SGD, ADAM, and ADAGRAD. To ensure robustness and reliability, a 5-fold cross-validation (CV) methodology was implemented, and the final results were obtained by averaging the outcomes of the CV runs.

Upon analyzing the experimental results, a clear pattern emerges. The SGD optimization method emerged as the most successful, boasting impressive metrics: an Accuracy of 0.9737, a Recall of 0.9474, an F1 score of 0.9474, and an AUC value of 0.9649. These outcomes reflect the superior performance of the proposed approach, particularly with the SGD optimization method. Notably, the proposed system achieved state-of-the-art results within the existing literature concerning the WebScreenshots Database, not only with the SGD method but also with the ADAGRAD optimization method.

## Acknowledgments

This work has been supported by the NVIDIA Corporation. All experimental studies were carried out on the TITAN XP graphics card donated by NVIDIA. We sincerely thank NVIDIA Corporation for their supports.

## Conflict of Interest Statement

The authors declare that there is no conflict of interest

## References

[1] J. McQuillan, I. Richer and E. Rosen, "The New Routing Algorithm for the ARPANET" *IEEE Transactions on Communications*, vol. 28, no. 5, pp. 711-719, May 1980. doi:10.1109/TCOM.1980.1094721

[2] C. P. Berges and V. Schafer, "Arpanet (1969–2019)," *Internet Histories*, vol. 3, no. 1, pp. 1-14, 2019. doi:10.1080/24701475.2018.1560921

[3] M. T. Simsim, "Internet usage and user preferences in Saudi Arabia," *Journal of King Saud University-Engineering Sciences*, vol. 23, no. 2, pp. 101–107, 2011. doi:10.1016/j.jksues.2011.03.006

[4] A. Weinstein and M. Lejoyeux, "Internet Addiction or Excessive Internet Use," *The American Journal of Drug and Alcohol Abuse*, vol. 36, no. 5, pp. 277-283, 2010. doi:10.3109/00952990.2010.491880

[5] K. Chan and W. Fang, "Use of the internet and traditional media among young people," *Young Consumers*, vol. 8, no. 4, pp. 244–256, 2007. doi:10.1108/17473610710838608

[6] F. Aydos, A. M. Özbayoğlu, Y. Şirin, and M. F. Demirci, "Web page classification with Google Image Search results," *arXiv preprint arXiv:2006.00226*, 2020. [Online]. Available https://arxiv.org/abs/2006.00226

[7] X. Qi and B. D. Davison, "Web page classification: Features and algorithms," *ACM computing surveys (CSUR)*, vol. 41, no. 2, pp. 1–31, 2009. doi:10.1145/1459352.1459357

[8] C. Xia and X. Wang, "Graph-Based Web Query Classification," in *2015 12th Web Information System and Application Conference*, WISA, 11-13 Sept. 2015, Jinan, China [Online]. Available: IEEE Xplore, http://www.ieee.org. [Accessed: 04 February 2016].

[9] M. Hashemi, "Web page classification: a survey of perspectives, gaps, and future directions," *Multimed Tools Appl*, vol. 79, no. 17–18, pp. 11921–11945, 2020. doi:10.1007/s11042-019-08373-8

[10] H. Alvari, E. Shaabani, P. Shakarian, H. Alvari, E. Shaabani, and P. Shakarian, "Semi-Supervised Causal Inference for Identifying Pathogenic Social Media Accounts," *Identification of Pathogenic Social Media Accounts: From Data to Intelligence to Prediction*, pp. 51–61, 2021 doi:10.1007/978-3-030-61431-7_5

[11] A. Ahmadi, M. Fotouhi, and M. Khaleghi, "Intelligent classification of web pages using contextual and visual features," *Applied Soft Computting*, vol. 11, no. 2, pp. 1638–1647, 2011. doi:10.1016/j.asoc.2010.05.003

[12] A. Sun, E. P. Lim, and W. K. Ng, "Web classification using support vector machine," in *Proceedings of the 4th international workshop on Web information and data management*, New York, NY, USA, WIDM02, Association for Computing Machinery,2002, pp. 96–99.

[13] X. Qi and B. D. Davison, "Knowing a web page by the company it keeps," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, CIKM06, New York, NY, USA, Association for Computing Machinery, 2006, pp. 228–237.

[14] L. Tian, D. Zheng, and C. Zhu, "Image classification based on the combination of text features and visual features," *International journal of intelligent systems*, vol. 28, no. 3, pp. 242–256, 2013. doi:10.1002/int.21567

[15] O. W. Kwon and J. H. Lee, "Web page classification based on k-nearest neighbor approach," in *Proceedings of the fifth international workshop on Information retrieval with Asian languages*, IRAL00, New York, NY, USA, Association for Computing Machinery, 2000, pp. 9–15.

[16] P. Calado, M. Cristo, E. Moura, N. Ziviani, B. Ribeiro-Neto, and M. A. Gonçalves, "Combining link-based and content-based methods for web document classification," in *Proceedings of the twelfth international conference on Information and knowledge management*, CIKM03, New York, NY, USA , Association for Computing Machinery, 2003, pp. 394–401.

[17] K. Gürkahraman and R. Karakiş, "Brain tumors classification with deep learning using data augmentation," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 36, no. 2, pp. 997–1011, 2021. doi:10.17341/gazimmfd.762056

[18] A. Tekerek, "A novel architecture for web-based attack detection using convolutional neural network," *Computers & Security*, vol.

100, pp. 102096, 2021. doi:10.1016/j.cose.2020.102096

[19] R. Karakis, K. Gurkahraman, G. D. Mitsis, and M. H. Boudrias, "Deep learning prediction of motor performance in stroke individuals using neuroimaging data," *Journal of Biomedical Informatics*, vol. 141, pp. 104357, 2023. doi:10.1016/j.jbi.2023.104357

[20] S. Savaş, N. TOPALOĞLU, Ö. KAZCI, and P. KOŞAR, "Comparison of deep learning models in carotid artery Intima-Media thickness ultrasound images: CAIMTUSNet," *Bilişim Teknolojileri Dergisi*, vol. 15, no. 1, pp. 1–12, 2022, doi:10.17671/gazibtd.804617

[21] M. Kizilgul, R. Karakis, N. Dogan, H. Bostan, M. M. Yapici, U. Gul et al., "Real-time detection of acromegaly from facial images with artificial intelligence," *European journal of endocrinology*, vol. 188, no. 1, pp. 158-165, 2023. doi:10.1093/ejendo/lvad005

[22] S. Savaş, "Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2201–2218, 2022. doi:10.1007/s13369-021-06131-3

[23] Y. Lin, "RNN-Enhanced Deep Residual Neural Networks for Web Page Classification," Ph.D. dissertation, University of Calgary, Calgary, Canada, 2016.

[24] E. Buber and B. Diri, "Web page classification using RNN," *Procedia Computer Science*, vol. 154, pp. 62–72, 2019. doi:10.1016/j.procs.2019.06.011

[25] D. Alsaleh and S. Larabi-Marie-Sainte, "Arabic Text Classification Using Convolutional Neural Network and Genetic Algorithms," *IEEE Access*, vol. 9, pp. 91670-91685, 2021. doi:10.1109/ACCESS.2021.3091376

[26] A. P. García-Plaza, V. Fresno, R. M. Unanue and A. Zubiaga, "Using Fuzzy Logic to Leverage HTML Markup for Web Page Representation," in *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 919-933, Aug. 2017. doi:10.1109/TFUZZ.2016.2586971

[27] V. Petridis and V. G. Kaburlasos, "Clustering and classification in structured data domains using Fuzzy Lattice Neurocomputing (FLN)," *in IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 2, pp. 245-260, March-April 2001. doi:10.1109/69.917564

[28] C. Haruechaiyasak, Mei-Ling Shyu and Shu-Ching Chen, "Web document classification based on fuzzy association," *Proceedings 26th Annual International Computer Software and Applications*, Oxford, UK, 2002, pp. 487-492.

[29] V. de Boer, M. van Someren, and T. Lupascu, "Web page classification using image analysis features," in *Web Information Systems and Technologies: 6th International Conference*, WEBIST 2010, Valencia, Spain, April 7-10, 2011, pp. 272–285.

[30] D. S. Ugalde, "Android App for Automatic Web Page Classification: Analysis of Text and Visual Features," Ph.D. dissertation, Universidade de Coimbra, Portugal, 2015.

[31] T. Gogar, O. Hubacek, and J. Sedivy, "Deep neural networks for web page information extraction," in *Artificial Intelligence Applications and Innovations*, AIAI 2016, Thessaloniki, Greece, September 16-18 2016, Proceedings 12, 2016, pp. 154–163.

[32] L. Li, G. Gou, G. Xiong, Z. Cao, and Z. Li, "Identifying gambling and porn websites with image recognition," in *Advances in Multimedia Information Processing: 18th Pacific-Rim Conference on Multimedia*, PCM 2017, Harbin, China, September 28-29, 2017, pp. 488–497.

[33] J. Sasaki, S. Li, and E. Herrera-Viedma, "A classification method of photos in a tourism website by color analysis," in *Advances and Trends in Artificial Intelligence. From Theory to Practice: 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2019*, Graz, Austria, July 9–11, 2019, Proceedings 32, 2019, pp. 265–278.

[34] S. Abdali, R. Gurav, S. Menon, D. Fonseca, N. Entezari, N. Shah and E. E. Papalexakis, "Identifying misinformation from website screenshots," in *Proceedings of the International AAAI Conference on Web and social media*, ICWSM-21, California, USA, June 7-10, 2021, pp. 2–13.

[35] M. M. Yapıcı, A. Tekerek and N. Topaloğlu, "Performance Comparison of Convolutional Neural Network Models on GPU," *IEEE 13th International Conference on Application of Information and Communication Technologies (AICT), 23-25 October 2019*, Baku, Azerbaijan [Online]. Available: IEEE Xplore, http://www.ieee.org. [Accessed: 06 Feb. 2020].

[36] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard and L. Jackel, Advances in Neural Information Processing Systems: *Handwritten digit recognition with a back-propagation network*, Morgan-Kaufmann, 1990

[37] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, November 1998. doi:10.1109/5.726791

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, June 2017. doi:10.1145/3065386

[39] A. Tekerek and M. M. Yapici, "A novel malware classification and augmentation model based on convolutional neural network," *Computer & Security*, vol. 112, pp. 102515, January 2022, doi:10.1016/j.cose.2021.102515

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, CVPR 2016, 26-29 June 2016, Las Vegas, USA [Online]. Available: https://cvpr2016.thecvf.com/. [Accessed: 01 July. 2016].

[41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, CVPR 2017, 21-26 July 2016, Honolulu, Hawaii [Online]. Available: https://cvpr2017.thecvf.com/, [Accessed: 29 July. 2017].