Approaches to the recognition of handwritten letters of the Azerbaijani language and their analysis

R.B. Azimov

Institute of Control Systems, Baku, Azerbaijan

ARTICLE INFO

Article history:
Received 10.11.2022
Received in revised form 22.11.2022
Accepted 06.12.2022
Available online 05.04.2023

Keywords:
OCR
Machine learning
Ensemble learning
ANN
CNN
K-means

ABSTRACT

Optical character recognition (OCR) computer systems are widely used for recognition of car license plates, for enabling visually impaired people to work with written information, etc. One of the current areas of OCR problems is the recognition of handwritten characters. This study proposes approaches to the construction of combined models using simple and ensemble learning for recognition of handwritten letters of the Azerbaijani language. The models built using the proposed approaches utilized several classes of features. Based on the results of computer experiments, a comparative analysis of the recognition accuracy of the constructed models and features using the proposed recognition methods is conducted.

1. Introduction

OCR computer systems are widely used for license plate recognition, for enabling visually impaired people to work with written information, etc. At the same time, one of the important applications of OCR systems is facilitating the input of texts in Latin alphabets, as well as texts written in non-Latin alphabets such as Japanese, Chinese, Korean, Hindi, Arabic and other languages into electronic computing devices [1, 2].

There have been many studies on the development of OCR systems for the recognition of letters written in different languages. In [3] a comparative analysis of several studies in the field of recognition of letters of the Hindi alphabet is presented. A handwritten letter recognition system in English and French based on the "Hidden Markov Model" has been developed [4]. For the recognition of handwritten Chinese characters, a model was developed on the ICDAR-2013 competition dataset, giving results with high accuracy [5].

One of the topical areas among the OCR problems is the recognition of handwritten characters. In order to automatically process forms filled in the Azerbaijani language letters, work has been carried out on the recognition of handwritten letters in the Azerbaijani language using Artificial Neural Network (ANN) [6-11]. Recognition of handwritten letters in the Azerbaijani language using the Support Vector Machine (SVM) has been studied [12]. Along with this, a comparative analysis of the effectiveness of using ANN and Convolutional Neural Network (CNN) in the recognition of handwritten letters of the Azerbaijani language has been conducted based on the results of computer experiments [13].

E-mail address: rustemazimov1999@gmail.com (R.B. Azimov).

www.icp.az/2023/1-05.pdf https://doi.org/10.54381/icp.2023.1.05

In computer systems that recognize handwritten characters, character images are explicitly segmented. It is determined which character images these images represent. To do this, each character image is given a mathematical model that recognizes the character, that is, determining the class of the image - a classification model [14].

SVM [15], ANN [16], CNN [17] have been widely used worldwide as classification models in recent years. Along with this, K-Nearest Neighbors [18], Logistic Regression [19], Naive Bayes Classifier [20], Decision Tree (DT) [21] are used. Combined classification models using Ensemble learning have gained particular relevance in recent years. In the combined classification model trained with Ensemble learning, not one but several classification models are used. For example, when recognizing a given object class through a Random Forest model trained with ensemble learning, the object features is fed to the input of individual DT models, while the final conclusion of the combined Random Forest model is determined by voting between the results of the DT models. Along with this, combined models trained with Ensemble learning include Extra trees, AdaBoost, Gradient Boosting Machines, XGBoost and others [22-24].

This study proposes approaches to the recognition of handwritten letters in the Azerbaijani language, which differ from each other in the structure of the recognition model. The first of these approaches takes a unary classification model rather than a combined one. The second and third proposed approaches are based on the bagging type of ensemble learning, and the fourth is based on the stacking type. But the individual models included in the recognition apparatus in these proposed approaches are ANN models, not Decision Tree. The learning datasets used by the models in the recognition apparatus in the second and third approaches are determined as a result of clustering. Three different classes of features were used in the second, third, and fourth approaches. Two of these features were used in the ANN of the first approach, the others used CNN as an alternative.

In the study, we analyzed three different models that use the first approach. One of them is CNN, and the other two are ANN models, which use different features. In the second and third approaches, respectively, all observations in the training dataset and observations belonging to a certain class are clustered, and a separate model is constructed based on the observations falling into each cluster. With the recognition apparatus working on the second or third approach, the recognition first determines which cluster the object falls into - is close to, then the class of that object - here the classification model of the corresponding cluster determines which character the character image represents. Finally, the recognition apparatus in the fourth approach has as many models as the number of characters - classes. For each class, here, letter, a separate model is built, each model determining whether the image is the corresponding letter or not. Recognition is considered complete when only one of these models determines that the image of the letter is the right letter.

In this article, we conducted a comparative analysis of the accuracy of recognition models using the proposed recognition approaches and features based on the results of computer experiments.

2. Description of the proposed approaches to the construction of the classification model

This article proposes four approaches for the construction of a classification model. CNNs with different structure and ANNs with different structure, working with different features, were used as a classification model built using the first of these approaches for the recognition of handwritten letters in the Azerbaijani language. Each of the combined classification models built with the second and third approaches uses ANN models. However, in the fourth approach, a combined model is built consisting of ANN models that determine whether each class corresponds to itself, that is, whether the object belongs to the corresponding class. The second, third, and fourth approaches used 3 different classes of features.

Clustering was used in the second and third approaches. In the second approach, the objects in the training dataset are clustered, and in the third approach - the vectors denoting the classes are the numerical averages of the objects of the same class in the training dataset.

Description of the first approach. In the first of the proposed approaches the ANN, CNN classification models are built for recognition of handwritten letters of the Azerbaijani language. First, the training dataset is trained on these models. Then the accuracy of the model is evaluated by comparing the characters predicted by the model based on the test dataset with the real output data in the test dataset.

Fig. 1. Examples of images of handwritten letters

Description of the second and third approach. The 2nd and 3rd approaches used clustering. These approaches can be referred to the bagging type of ensemble learning. [21].

The construction of the combined classification model of the second approach consists of the following steps:

1. The image features in the training dataset are divided into clusters using the k-means method based on their proximity to each other.
2. A training dataset is created, consisting of images that fall into each cluster, and their responses in the training dataset - vectors indicating which class they belong to.
3. Each of the training dataset created in (2) is trained on a separate model.

The construction of the combined classification model of the third approach consists of the following steps:

1. The numerical average of the features of the images of each letter is calculated. Thus, for letter k a vector k is found, which "summarizes" these letters.
2. The vectors obtained in (1) are divided into clusters using the k-means method.
3. A training dataset is created, consisting of images in the training dataset, relating the letters in each cluster, and their responses in the training dataset - vectors indicating the characters.
4. Each of the training datasets created in (3) is trained on a separate model

Classification of a given image using the second and third approaches consists of the following steps:

1. Based on the features of the given image, it is determined which cluster it is close to.
2. Using the model corresponding to the cluster to which the given image is close, it is determined which character it represents.

Description of the fourth approach. This approach is close to the stacking type of ensemble learning. Thus, k number of models are built for class k, and a separate model is built for recognition of each class. All images in the training dataset are used to train each model.

The construction of the combined classification model of the fourth approach consists of the following steps.

The images in the training dataset are tagged separately for each model. When recognizing handwritten characters, each image in the training dataset is tagged with one of the two types of tags, e.g., presence or absence of the letter corresponding to a model. For example, when developing a training dataset for a model corresponding to the letter "A", the images of the letter "A" are marked as 1, and the images of other letters as 0. Similarly, for a model corresponding to the letter "B", the images of the letter "B" are marked as 1, and the images of other letters as 0.

Each model is trained on the corresponding training base.

Thus, in this approach, the multiclass classification model is transformed into binary classification models.

Determining the class of a given image using the fourth approach consists of the following steps:

1. Based on the features of a given image, all models predict which character is represented by that image.
2. Only when one of these letter models recognizes a given image as its own letter, e.g., when a given image is only recognized as the model of the letter "A", while the other models determine that it is not their corresponding letter, we say that it is the letter "A".

In general, the fourth approach results in one of the following:

1. The class of a given object is recognized by only one of the class models.
2. More than one class model recognizes the object class. When recognizing handwritten characters, for example, a given letter image is recognized by both the "B" model and the "P" model. Since this image cannot be both the letter "B" and the letter "P", the recognition is rejected here.
3. None of the class models recognizes the class of the object.

3. Some characteristics of the experiments

Features. In this work, ANN models are built that use the following features in all four approaches.

"Pixels" features. A bitmap image is in the form of a matrix consisting of pixels. In grayscale images, however, the pixel is represented by a number, that is, the image is a matrix of numbers. This matrix is converted into a vector with consecutive rows. A 20x20 grayscale image is expressed as a vector with 400 elements.

"PDC" (Peripheral directional contributivity) features. These features use the distance of black points relative to each other [25]. The algorithm used to extract PDC features is described below.

1. The image is approached from 4 directions: left to right, right to left, top to bottom and bottom to top.
2. Each of the four sides of the image is divided into 8 segments.
3. The first 2 black pixels are set, these two points are called the starting points.
4. From these starting points the search for black pixels in 4 directions begins, the steps are counted.

As a result, the number of elements of the vector of the PDC features are 4x8x2x4=256.

The first approach used CNN along with ANN. LeNet-5 was used as the CNN architecture [26]. CNNs with different architectures were built with changing the number of cores on the first and second "convolution" levels in the LeNet-5 architecture. However, the second and fourth approaches use ANNs working with the feature vector derived from CNNs with LeNet-5 architecture trained on the training dataset.

"Conv" features. As features of given image in this work we take the output of neurons in the last layer up to the fully connected layer of the CNN model with LeNet-5 architecture, trained on the training dataset. Thus, the features of each image are taken as a vector of 120 numbers

Dataset. A dataset consisting of handwritten samples of 28 of the 32 letters of the Azerbaijani alphabet was used in this work. All of these letters are capital letters.

Since recognition of the handwritten letters of the Azerbaijani alphabet {ĞİÖÜ} can be performed using simple algorithms by comparing characters with ascenders to their corresponding characters without ascenders, there are no examples of these letters in the dataset.

The images in the dataset are presented in black and white. Each of these images is resized to 20x20 and converted to grayscale format.

The dataset is divided into training and test datasets. The training dataset contains 500 samples of each letter, 14,000 in total, and the test dataset contains 100 samples of each letter, 2,800 images in total.

Methodology. In the first approach, the "Pixels" and "PDC" features were used in ANN, and the "Conv" features were used in CNN with LeNet-5 architecture. The models in the first approach were constructed at 5 starting points. In the first approach along with this augmentation, the size of the training dataset was multiplied by 2, 3, 5 times. In the second approach, the nontrivial numbers of clusters are 2, 3, ..., 13999. In the computer experiments, the number of clusters was taken to be 2, 3, ..., 50. The third approach used all non-trivial cluster numbers: 2, 3, ..., 27. For the comparative analysis of the approaches, the results in the starting point giving the highest accuracy in the first approach and the results in the number of clusters giving the highest accuracy in the second and third approaches were used.

Each of the four approaches was used with 3 different features. The results of the computer experiments show both the accuracy of the approaches by features and the accuracy of the use of the features by the approaches.

When recognizing which letter the given image represents according to the second and third approaches, the proximity of the features of the image to which cluster was initially determined by the proximity to the centers in the k-means method. However, in this way, in most of the images in the test dataset, erroneous clusters were selected - those that did not contain samples related to that letter. Therefore, in developing the recognition apparatus in Approaches 2 and 3, after dividing the training dataset into clusters, a training dataset consisting of images in the training dataset and the numbers indicating which cluster they fall into was prepared, and the ANN was trained using this training dataset. Determining to which cluster a given image is close is performed with the help of this trained ANN. In this case, it was possible to select in most of the images in the test dataset the correct cluster - one with examples relating to the right letter.

The computer experiments were conducted using the Keras and Scikit-learn libraries used in the Python language [27, 28].

4. Results of computer experiments

The accuracies of the ANN and CNN models built using the first approach on the test dataset are given in Fig. 2.

As can be seen from Fig. 2, CNN models showed more accurate recognition results than ANNs. ANNs working with "PDC" features showed more accurate results than ANNs working with "Pixels" features.

Figure 3 and 4 show the frequencies of selecting images in the test dataset into correct (with samples related to this letter) clusters, when the cluster to which the given image is close is determined based on the proximity to the centers in k-means, by the second and third approaches, respectively. An ANN was then trained, to be used to determine the cluster to which a given image is close in pairs, consisting of images in the training base and vectors indicating which cluster they should fall into when clustered. Thus, the proximity of a given image to which cluster is determined based on the prediction of this ANN.

The numbers of clusters giving high accuracy in the second and third approaches are given in Table 1.

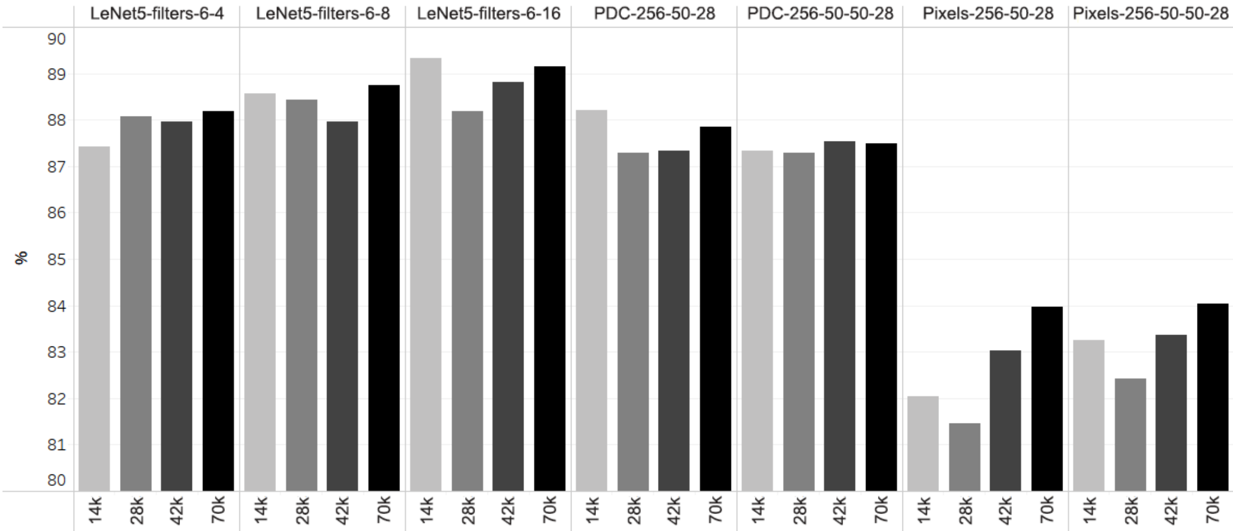

Fig. 2. Accuracy values in the first approach

Note: The parameter k in Fig. 2 is equal to 1000.

Table 1

The number of clusters giving high accuracy in the second and third approaches

	Pixels	PDC	Conv
Approach 2 (2,3,…,50)	32	50	26
Approach 3 (2,3,…,27)	15	25	24

The number of recognized and unrecognized observations in the test dataset with the fourth approach is given in Table 2.

Table 2

Number of recognized and unrecognized observations in the fourth approach

	Recognized by more than one model	Recognized by no model	Recognized
Pixels	268	383	2149
PDC	681	99	2020
Conv	351	155	2294

Table 3

Accuracy of the approaches for different features

	Pixels	PDC	Conv
Approach 1	84.04%	88.21%	89.32%
Approach 2	57.64%	72.03%	83.79%
Approach 3	71.92%	86.50%	85.64%
Approach 4	70.03%	67.21%	77.39%

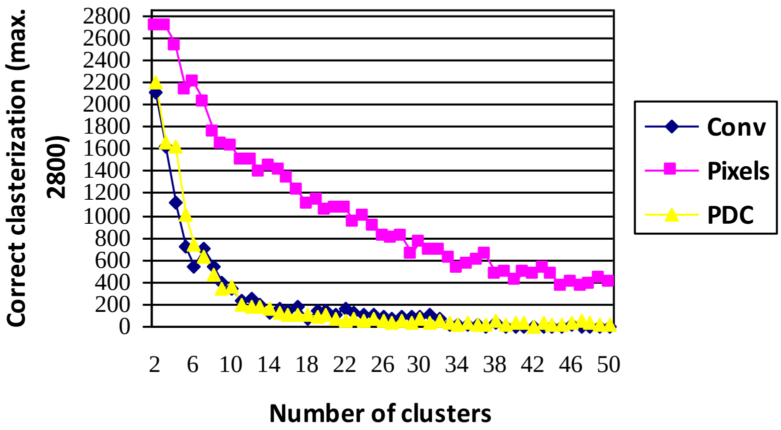

Fig. 3. The frequencies of a given image falling into a cluster, with which test observations can be recognized, if cluster detection is performed with centers in the k-means method, in the second approach.

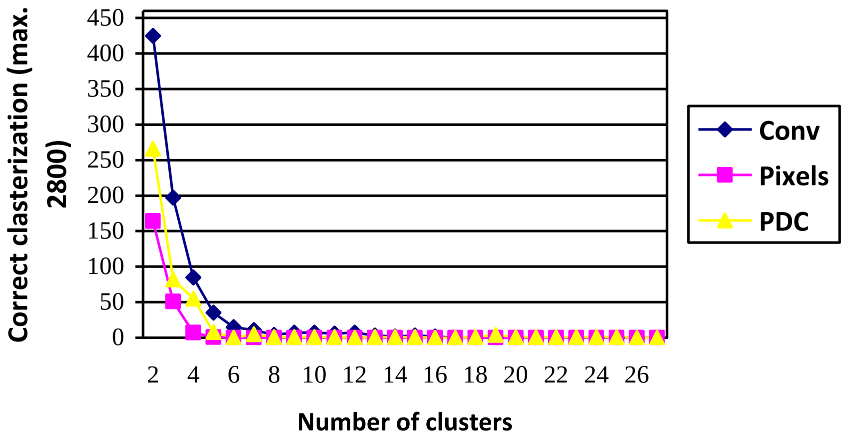

Fig. 4. The frequencies of a given image falling into a cluster, with which test observations can be recognized, if cluster detection is performed with centers in the k-means method, in the third approach

The accuracies of the tested approaches using different feature classes are shown in Table 3.

As shown in Table 3, Approaches 2-4 could not show on the test dataset on any of the features a higher recognition accuracy than Approach 1. The descending order of the approaches in terms of recognition accuracy in the test dataset for the features "PDC" and "Conv" is as follows:

1. Approach 1,
2. Approach 3,
3. Approach 2,
4. Approach 4.

For the "Pixels" features, Approach 4 showed a more accurate result in the test dataset than Approach 2, that is, the descending order of approaches in terms of recognition accuracy in the test database for the "Pixels" features is as follows:

1. Approach 1,
2. Approach 3,
3. Approach 4,
4. Approach 2.

5. Conclusion

The article proposes approaches for building combined models for recognition of handwritten letters of the Azerbaijani language, using simple and ensemble learning. The models built using the proposed approaches used several classes of features. No higher recognition accuracy for any feature classes was obtained in the proposed approaches of building combined models (Approaches 2, 3 and 4) than in the approach of building a simple classification model (Approach 1). The approach using clustering of classes - vectors denoting them (approach 3) showed higher recognition accuracy for all three features than the approach using clustering of objects (Approach 2). In descending order of accuracy of the models built by the proposed approaches for the features "peripheral directional contributivity" and "convolution", the approach, which builds a separate model for each class (Approach 4), ranks last. In descending order of accuracy of the models built using the proposed approaches, which use image pixels as features, the approach, which uses object clustering (Approach 2), ranks last.

References

[1] Najib Ali Mohamed Isheawy and Habibul Hasan, "Optical Character Recognition (OCR) System", IOSR Journal of Computer Engineering (IOSR-JCE). 17 No.2 Ver.II (Mar-Apr.2015) pp.22-26.

[2] S. Chandra, S. Sisodia & P. Gupta, Optical character recognition-a review, Int. Res. J. Eng. Technol. 7 No.4 (2020) pp.3037-3041.

[3] Pal Umapada & Wakabayashi Tetsushi & Kimura Fumitaka, Comparative study of devnagari handwritten character recognition using different feature and classifiers. (2009) pp.1111-1115.

[4] M. Kozielski, P. Doetsch, and H. Ney, ‘‘Improvements in RWTH’s system for off-line handwriting recognition,’’ in Proc. 12th Int. Conf. Document Anal. Recognit., Washington, DC, USA, Aug. 2013, pp.935-939. doi: 10.1109/icdar.2013.190.

[5] X-Y. Zhang, Y. Bengio, and C.-L. Liu, ‘‘Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark,’’ Pattern Recognit. 61 (2017) pp.348-360. doi: 10.1016/j.patcog.2016.08.005.

[6] К.Р. Айда-заде, Э.Э. Мустафаев, Интеллектуальная система распознавания рукопечатных форм азербайджанского языка, Труды Республиканской научной конференции «Современные проблемы информатизации, кибернетики и информационных технологий», Баку. 3 (2006) pp.85-88. [In Russian: K.R. Aida-zade, E.E. Mustafayev, Intelligent system of recognition of handwritten forms of the Azerbaijani language, Proceedings of the Republican Scientific Conference "Modern problems of informatization, cybernetics and information technology", Baku].

[7] К.Р. Айда-заде, Э.Э. Мустафаев, Об одной иерархической системе распознавания рукописных форм на основе нейронных сетей, Известия НАН Азербайджана, серия ф.т. и м.н. No.2-3 (2002) p.94-98. [In Russian: K.R. Aida-zade, E.E. Mustafayev, On one hierarchical system of recognition of handwritten forms based on neural

networks, Transactions of the Azerbaijan National Academy of Sciences, Physics, Engineering and Mathematics Series].
[8] К.Р. Айда-заде, Э.Э. Мустафаев, Дж.З. Гасанов, Об использовании баз знаний для повышения интеллектуальности систем распознавания, Доклады 11-й Всероссийской конференции «Математические методы распознавания образов», Москва. (2003) pp.6-8. [In Russian: K.R. Aida-zade, E.E. Mustafayev, J.Z. Hasanov, On the use of knowledge bases to increase the intelligence of recognition systems, Proceedings of the 11th All-Russian Conference "Mathematical Methods of Pattern Recognition", Moscow].
[9] Э.Э. Мустафаев, Многоуровневая иерархическая система распознавания рукописных форм, Материалы научной конференции «Современные проблемы прикладной математики», Баку. (2002) pp.154-157. [In Russian: E.E. Mustafayev, Multilevel hierarchical system of recognition of handwritten forms, Proceedings of the scientific conference "Modern problems of applied mathematics", Baku].
[10] K.R. Aida-zade, E.E. Mustafayev, Intelligent handwritten form recognition system based on artificial neural networks, Proceedings of the Intern. Conf. on Modeling and Simulation, 28-30 August, Konya, Turkey. (2006) pp. 609-613.
[11] К.Р. Айда-заде, Э.Э. Мустафаев, Интеллектуальная автоматизированная система обработки и оценивания результатов тестовых экзаменов, Прикладная математика и фундаментальная информатика, Омск. 5 No.2 (2018) pp.51-60. [In Russian: K.R. Aida-zade, E.E. Mustafayev, Intelligent automated system for processing and evaluating the results of test examinations, Prikladnaya matematika i fundamental'naya informatika, Omsk].
[12] Araz Aliev & Elviz Ismayilov, Selection of the optimal class of features for recognition of the Azerbaijani handprinted characters, (2020). 10.1007/978-3-030-68004-6_12.
[13] E.E. Mustafayev, R.B. Azimov, Comparative analysis of the application of multilayer and convolutional neural networks for recognition of handwritten letters of the Azerbaijani alphabet, Cybernetics and Computer Technologies. No.3 (2021) pp.65-73. https://doi.org/10.34229/2707-451X.21.3.6
[14] Ren, X., & Malik, J., Learning a classification model for segmentation, In Computer Vision, IEEE International Conference on Computer Society. 2 (2003) pp.10-10.
[15] Huang S., Cai N., Pacheco P.P., Narrandes S., Wang Y. & Xu W., Applications of Support Vector Machine (SVM) learning in cancer genomics, Cancer genomics & proteomics. 15 No.1 (2018) pp.41-51. https://doi.org/10.21873/cgp.20063
[16] Heriz Hazem & Salah Hamza & Abdu Sharief & Sbihi Mohammad & Abu-Naser Samy, English alphabet prediction using artificial neural networks, International Journal for Academic Development. 2 (2018) pp.8-14.
[17] Sen Maitra Durjoy & Bhattacharya Ujjwal & Parui Swapan, CNN based common approach to handwritten character recognition of multiple scripts. (2015) pp.1021-1025. 10.1109/ICDAR.2015.7333916.
[18] V.B. Prasath, H.A.A. Alfeilat, A. Hassanat, O. Lasassmeh, A.S. Tarawneh, M.B. Alhasanat & H.S.E. Salman, Distance and similarity measures effect on the performance of K-Nearest neighbor classifier--A Review, arXiv preprint arXiv: 1708.04321, (2017).
[19] E.Y. Boateng & D.A. Abaye, A review of the logistic regression model with emphasis on medical research, Journal of data analysis and information processing. 7 No.4 (2019) pp.190-207.
[20] M. Makhtar, H. Nawang & S.N. Wan Shamsuddin, Analysis on students performance using naïve bayes classifier, Journal of Theoretical & Applied Information Technology. 95 No.16 (2017).
[21] S. Tangirala, Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm, International Journal of Advanced Computer Science and Applications. 11 No.2 (2020) pp.612-619.
[22] Dong X., Yu Z., Cao W., Shi Y., & Ma Q., A survey on ensemble learning. Frontiers of Computer Science. 14 No.2 (2020) pp.241-258.
[23] Z.H. Zhou, Ensemble learning. In Machine learning. Springer, Singapore, (2021).
[24] N. Memon, S.B. Patel, & D.P. Patel, Comparative analysis of artificial neural network and XGBoost algorithm for PolSAR image classification, In International Conference on Pattern Recognition and Machine Intelligence, Springer, Cham. (2019) pp.452-460.
[25] Mori M., Wakahara T. & Ogura K. Measures for structural and global shape description in handwritten kanji character recognition. In Document Recognition V SPIE. 3305 (1998) pp.81-89.
[26] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[27] Keras library's website, https://keras.io/
[28] Scikit-learn library's website, https://scikit-learn.org/