Texture Feature Extraction and Classification Using Machine Learning Techniques

Rohini A. Bhusnurmath and Shaila Doddamani

Abstract Texture, a crucial aspect of an image, is something made up of components that are related to one another. Reliable feature extraction in image files requires the use of a texture-based categorization method, which is significant. This study proposes an effective method for classifying textures using machine learning (ML) approaches. Using these ML classifiers, which are in the form of artificial intelligence (AI), programmers can predict results exactly without providing instructed to do so explicitly. The proposed study focuses on the creation of own dataset in the form of CSV file, to do so Haralick features (contrast, dissimilarity homogeneity, energy, and correlation) extracted from the Brodatz texture dataset. Different ML algorithms are used like: K-Nearest Neighbor, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, and AdaBoost Classifier which are experimented on the created dataset to classify the texture of Brodatz dataset. Proposed approach exhibits better results with 100% accuracy with less computation time as compared to previous work in the literature.

Keywords Texture analysis · Brodatz · Classification · K-NN · Decision Tree · Random Forest · AdaBoost Classifier

1 Introduction

Image texture is characterized as a recurring pattern appears on an object's surface or structure. In computer vision and computer graphics, texture is still a crucial and essential issue with a verity of applications, consisting of synthesis, image comprehension, and picture content querying [1, 2]. Texture is a property that is used to divide visuals into regions of interest and to categorize those parts. The physical layout of shades or intensities in a vision is made clearer by texture. The spatial

R. A. Bhusnurmath · S. Doddamani (✉)
Department of Computer Science, Karnataka State Akkamahadevi Women's University, Vijayapura, Karnataka 586108, India
e-mail: dodamanishaila@gmail.com

patterns of intensity levels within a community define texture. Because of differences in visual appearance, orientation, or scale, textures in the actual world are not consistent, which presents a significant challenge for texture analysis [3].

A subclass of artificial intelligence [2, 4], machine learning, focuses on applying statistical methods to generate expert systems that can learn from databases that are already accessible. Machine learning algorithms use computer techniques to "learn" information directly from data, without requiring an existing equation as a model. The algorithms adjust to the performance of the samples when more are made available for learning. There are many classifiers used for classification in machine learning. The experimentation in the proposed paper considers different types of ML classifiers which are K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest Tree (RFT), etc.

2 Literature Review

Danuta et al. [1] presented the texture classification method and ML approach to identify image features of lightweight cementitious composites (LLC). The coating has changed with the nanocellulose and used them to address the strength of materials. Kapil et al. [2] show comparison result of the ML algorithms and chosen the K-Nearest Neighbor and the SVM classifiers to compare the results. Authors got the good accuracy with SVM classifier as compared to the k-NN. Daniel et al. [4] worked on most widely used machine learning algorithms in petroleum industry for reservoir properties. Authors used the ANN and SVM algorithms. ANN yields the better result and also worked with hybridization of multiple algorithms. Morshedul [5] worked with the machine learning algorithm to predict the Alzheimer's disease. These machine learning algorithms are used to identify the Dementia among various patients, and they have used the OASIS dataset. Among all the algorithms, SVM has given the good result for detecting the disease. Gregorius et al. [6] proposed the ratings of online review using the machine learning algorithms. Authors have used the text preprocessing and feature extraction methods. First worked with single and ensemble model. Next applied the best identified classifier for prediction. Finally applied the linear support vector classifier and got the good results.

Garpebring et al. [7] have worked on the Haralick texture features for image analysis. Authors have examined the effectiveness of density estimation techniques for GLCM approximation and subsequent identification of the related invariant features. Hiremath and Bhusnurmath [3] have worked structured approach for classification of the texture by using the local directional binary patterns and non-subsampled contourlet transform. Hiremath and Bhusnurmath [8] have worked on texture image classification based on novel color textures using the local directional pattern and anisotropic diffusion based on the RGB color space.

3   Methodology

Proposed approach for image classification follows the following steps:

Step 1: Reading the Brodatz texture image from the dataset.
Step 2: Divide each texture image into sub image of size 64 X 64.
Step 3: Extracted Haralick Features (contrast, dissimilarity homogeneity, energy, and correlation) from each sub image.
Step 4: Create the CSV file from the feature extracted in Step 3.
Step 5: Divide the dataset into training and testing sets in ratio of 80:20.
Step 6: Train the ML classifiers using training set.
Step 7: Test the ML classifiers using the testing set.
Step 8: Repeat the steps from Steps 6 to 7 using different ML classifiers.
Step 9: Select the best ML classifier in terms of accuracy.

3.1   Feature Extraction

Features are used to identify the characteristics of the image textures [9]. Proposed work is carried out using the following features.

3.1.1   Haralick Features

Haralick features are acquired from the Gray-Level Co-occurrence Matrix (GLCM). The GLCM describes how many times two gray-level pixels are adjacent to each other in an image [9]. For the proposed work, the five Haralick features extracted are explained below.

Contrast: Contrast is used to show the difference between amount of grayscale or color that exists in the images.
Dissimilarity: It shows the how data samples are different from one another.
Homogeneity: Is of type region of image that shows the changes of intensity that occurs in region.
Energy: Describe the changes in quality of image**.**
Correlation: Is the process of moving the mask over the image to compute the sum of product of each area.

3.2 Classification Algorithms

For the proposed approach, the different machine learning algorithms used are listed below:

- AdaBoost Classifier.
- Gradient Boosting Classifier.
- Random Forest Classifier.
- K-Neighbors Classifier.
- Decision Tree Classifier.

3.2.1 AdaBoost Classifier

AdaBoost Classifier [10] is a supervised machine learning algorithm which is primarily used for the classification as well as regression problems.

3.2.2 Gradient Boosting Classifier

A Gradient Boosting Classifier [4, 10] is also proven to the one of strong methods, and it is also used for the both classification purpose and regression problems. It is a group of machine learning ML that is used to combine the weak models together to come up with strong model.

3.2.3 Random Forest Classifier

Random Forest [3, 10] is a machine learning algorithm which is used for regression and classification problems. It is known as meta-estimator which is collection of many numbers of decision trees. It creates the set of decision trees from randomly chosen training set.

3.2.4 K-Neighbors' Classifier

K-Neighbor Classifier [8, 9] is a non-parametric supervised learning classifier used for the classification or prediction, which works based on the neighbor around the class. It works by finding the distance between the class and the examples of the data. Here, K defines the nearest value for the particular class.

3.2.5   Decision Tree Classifier

A classifier [15] creates the classification model by the decision trees. Each tree has
defined the attribute and each attribute is having the one possible prediction value
for the class. Here, the data will be continuously split by some parameters. The tree
can have mainly two parts that is decision nodes and the tree levels.

4   Data Collection

4.1   Dataset Preparation

Proposed approach is experimented on two Brodatz datasets of texture images. First
dataset consists of 1600 sub-images from 16 images from the Brodatz texture dataset,
and it is termed as Brodatz-1. The second dataset termed as Brodatz-2 consists of
11,100 sub-images derived from 111 images of the Brodatz texture album. The
images are grayscale images with format.gif and are without rotation. The Brodatz
texture dataset contains 111 texture images, each of size 640×640 pixels. For both
the datasets, sub-images are re-sampled into 100 non-overlapping sub-patch of size
64×64 pixels. The 16 chosen images for Brodatz-1 dataset are shown in Fig. 1.
Detailed description of the datasets is shown in Table 1.

Table 1 shows the detailed description of two Brodatz texture datasets used for
the proposed work.

4.2   Randomizations and Splitting the Data

The dataset which is mentioned in Sect. 4.1 is split into training and testing sets in
ratio of 80:20, respectively.

5   Results and Discussion

Proposed work is experimented on Intel core i3 processor running at 2.40 GHz speed
using 4 GB RAM, Windows 10 Operating System.

Proposed works focused on ML classifiers which are discussed in 3.2 are exper-
imented. The results are as follows: Figs. 2, 3, 4, 5, and 6 represent the results of
Brodatz-1 (1600 texture images) dataset in the form of confusion matrix for AdaBoost
Classifier, Gradient Boost Classifier, Random Forest Classifier (RFC), K-Nearest
Neighbor Classifier (KNN), and Decision Tree Classifier (DTC), respectively.
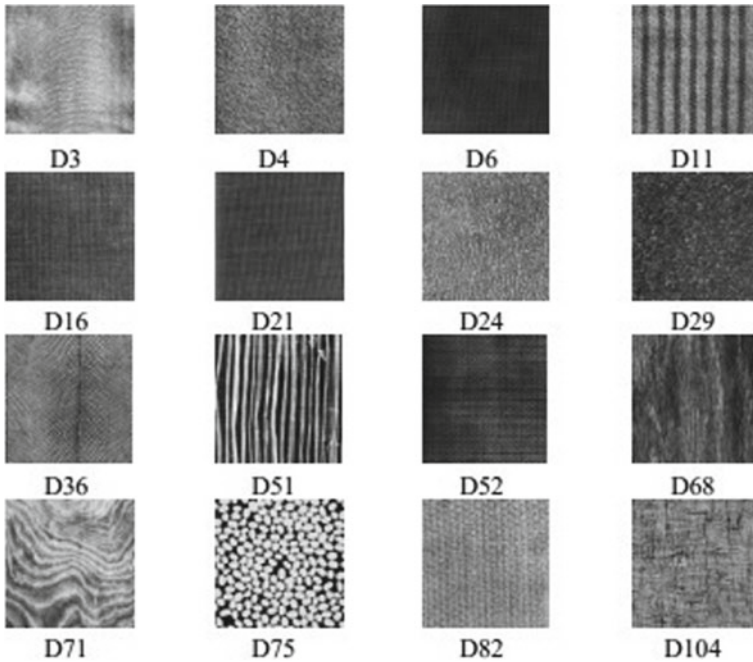Figure 7 represents the bar chart view of all five classifiers.

Fig. 1 Sixteen texture images from Brodatz-1 dataset

Table 1 Overview of the dataset used for the experiment

Dataset used	Brodatz-I	Brodatz-II
Total no. of images experimented	16	111
Original image size	640 × 640	640 × 640
Patchified sub-image size (pixel)	64 × 64	64 × 64
Total no. of images after patchifying	1600	11,100
Feature extracted	**Haralick features:** contrast, dissimilarity, homogeneity, energy, correlation	**Haralick features:** contrast, dissimilarity, homogeneity, energy, correlation

Table 2 shows the results of all created classifiers during this experiment.

Above figures show the details about the all the classifiers' prediction. From the results, we can conclude that the best result is obtained for K-Nearest Neighbor Classifier with classification accuracy of 100% and the Gradient Boost classifier classification accuracy of 99% which are best fit for the feature extracted from the Brodatz image texture dataset.
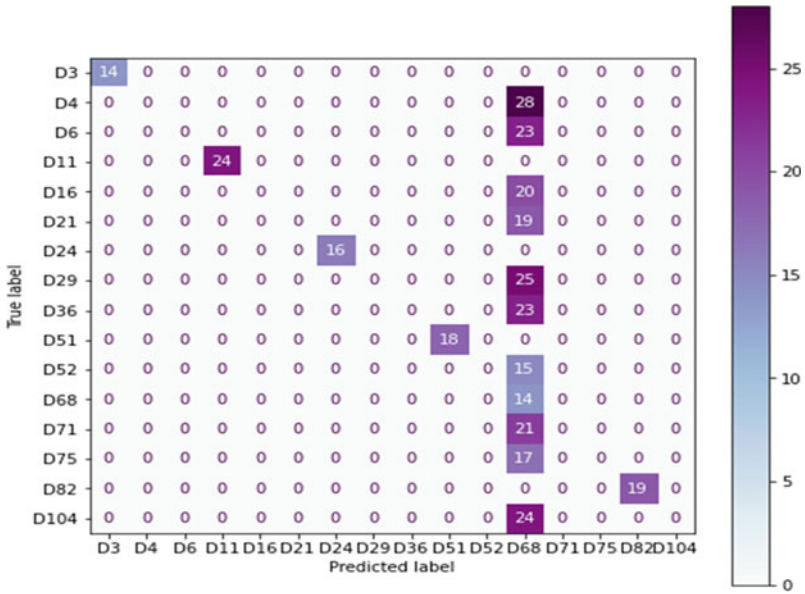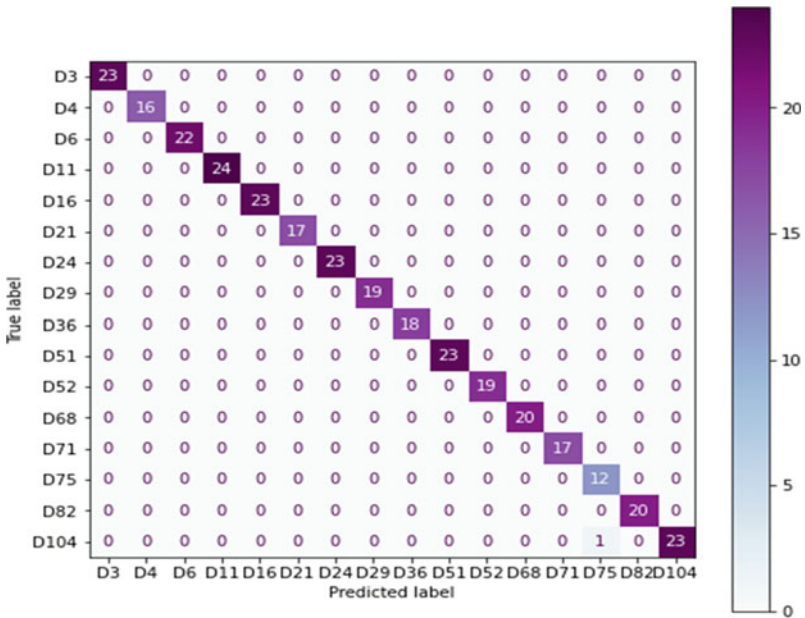
Fig. 2 AdaBoost classifier (ABC)

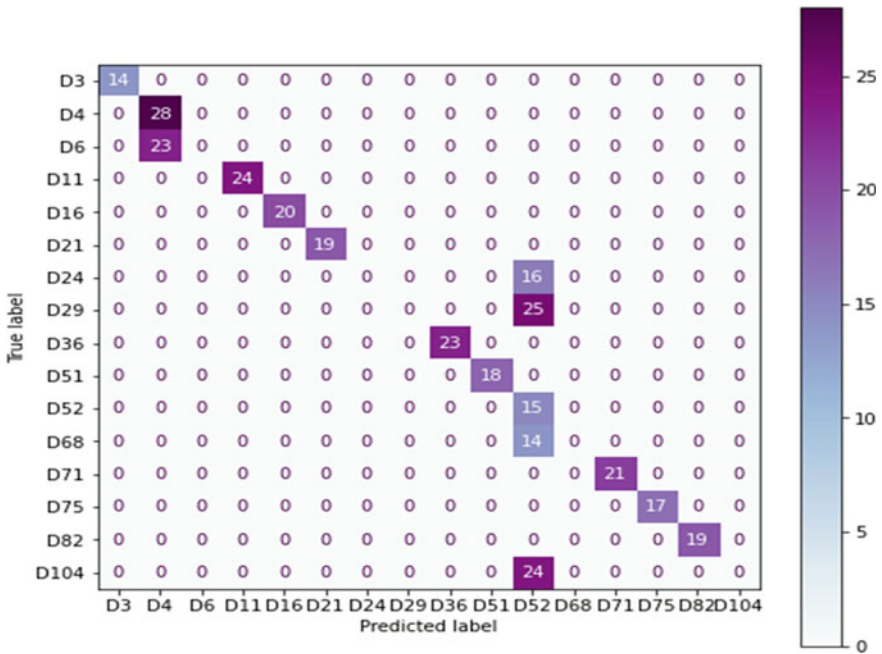**Fig. 3** Gradient boost classifier (GBC)

Fig. 4 Random forest classifier (RF)

Table 2 shows the details about the classification results' accuracy and the classification metrics precision, recall, and F1-score of each classifier of Brodatz-1 (1600 texture images) dataset. The train time and test time for all the classifiers are also tabulated.

From the Table 2, it is observed that experimented results are well suited on K-Nearest Neighbor Classifier with 100% accuracy and less computation time whereas Gradient Boost Classifier exhibits accuracy of 99%, Random Forest of 64%, Decision Tree Classifier of 65%.

Table 3 shows the results of Brodatz-2 (11,100 texture images) texture dataset.

It is observed from Table 3 that K-Neighbor Classifier and Gradient Boosting Classifier best fit for the proposed work with 100% accuracy.

The performance is evaluated on Brodatz dataset in the form of accuracy. Experimental results of the proposed experimentation have given better classification rate in comparison with some state-of-the-art approaches (Table 4).
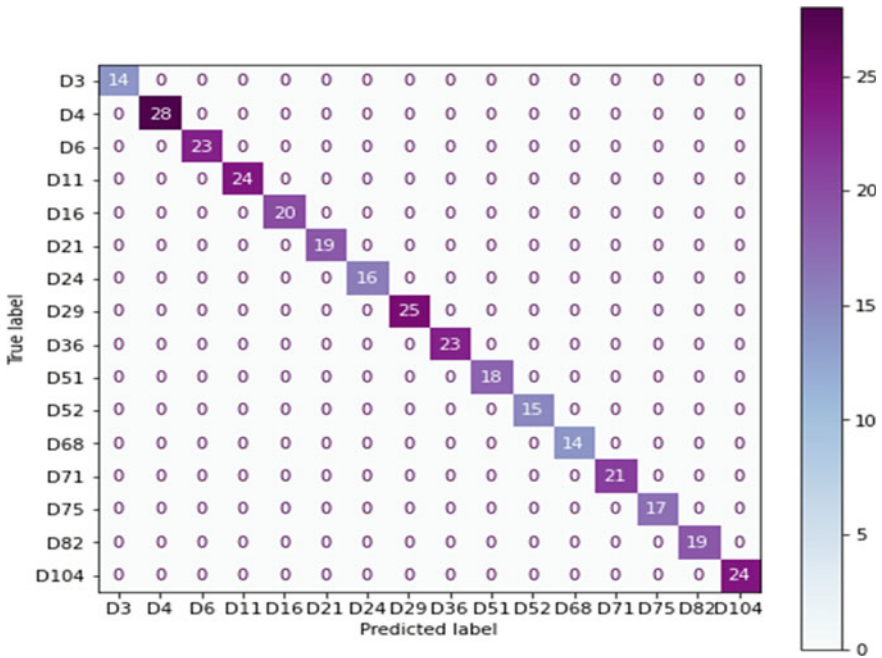
Fig. 5 K-nearest neighbor classifier (KNN)

6 Conclusion and Future Work

The proposed study mainly focuses on creation of new dataset in the form of CSV file using Brodatz texture dataset through Haralick feature extraction. Experiment is carried out on two datasets, one on 16 Brodatz image texture dataset and another one is on 111 Brodatz image texture dataset. Different machine learning classifiers (AdaBoost Classifier (ABC), Gradient Boosting Classifier (GBC), Random Forest (RF) Classifier, K-Nearest-Neighbors (KNN) Classifier, and Decision Tree (DT) Classifier) are experimented on the both datasets to classify the Brodatz textures. The proposed approach has performed better on created dataset. The future work can be done with deep learning techniques to obtain better accuracy.
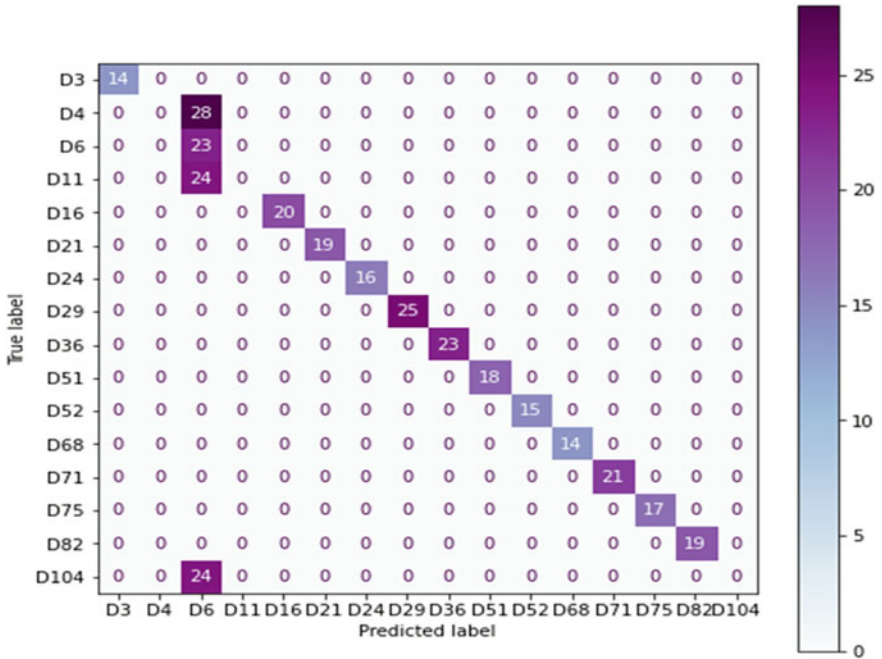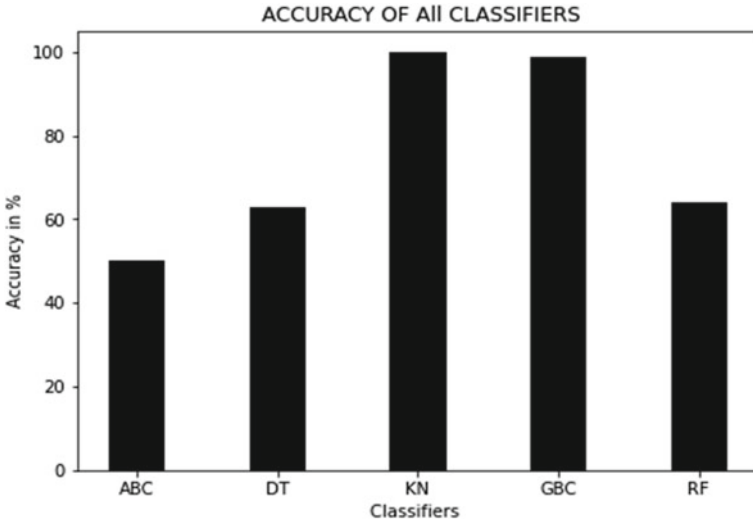
Fig. 6 Decision tree classifier (DT)

Fig. 7 Bar chart view of all classifiers

Table 2 Accuracy and classification report of Brodatz-1 (1600 texture images) texture dataset

Classifiers	Accuracy (%)	Tenfold validation (%)	Precision	Recall	F1-score	Train time (s)	Test time (s)
AdaBoost classifier	38	50	32	33	33	1.71	0.030
Gradient boosting classifier	**99**	**99**	**99**	**100**	**100**	**31.23**	**0.013**
Random forest classifier	82	64	81	76	76	0.113	0.03
K neighbors classifier	**100**	**100**	**100**	**100**	**100**	**0.016**	**0.094**
Decision tree classifier	63	65	68	63	63	0.09	0.009

Table 3 Accuracy and classification report of Brodatz-2 (11,100 texture images) texture dataset

Classifiers	Accuracy (%)	Tenfold validation (%)	Precision	Recall	F1-score	Train time (sec)	Test time (sec)
Gradient boosting classifier	**100**	**100**	**100**	**100**	**100**	**495.53**	**1.12**
Random forest classifier	62	64	57	64	62	0.31	0.093
K neighbors' classifier	**100**	**100**	**100**	**100**	**100**	**0.105**	**0.75**
Decision tree classifier	22	26	22	18	18	0.08	0.008

Table 4 Comparison of classification accuracy obtained with proposed method and other work on Brodatz image texture dataset

S. No.	Author	Proposed work
1	Alharan et al. [9]	GLCM and LBP feature extraction KNN = 99
2	Dhingra et al. [11]	Feature extraction using LBP KNN = 98%
3	Proposed method	Haralick Features (GLCM) KNN = 100%

References

1. Barnat-Hunek D, Omiotek Z, Szafraniec M, Dzierżak R (2021) An integrated texture analysis and machine learning approach for durability assessment of lightweight cement composites with hydrophobic coatings modified by nanocellulose. Measurement 179:109538
2. Sethi K, Gupta A, Gupta G, Jaiswal V (2019) Comparative analysis of machine learning algorithms on different datasets. In: Circulation in computer science international conference on innovations in computing (ICIC 2017), vol 87
3. Hiremath PS, Bhusnurmath RA (2014) A novel approach to texture classification using NSCT and LDBP. Int J Comput Appl 0975-8887
4. Otchere DA, Ganat TOA, Gholami R, Ridha S (2021) Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: comparative analysis of ANN and SVM models. J Petrol Sci Eng 200:108182
5. Bari Antor M, Jamil AHM, Mamtaz M, Monirujjaman Khan M, Aljahdali S, Kaur M, ... Masud M (2021) A comparative analysis of machine learning algorithms to predict Alzheimer's disease. J Healthc Eng 2021
6. Budhi GS, Chiong R, Pranata I, Hu Z (2021) Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis. Arch Comput Methods Eng 28(4):2543–2566
7. Garpebring A, Brynolfsson P, Kuess P, Georg D, Helbich TH, Nyholm T, Löfstedt T (2018) Density estimation of grey-level co-occurrence matrices for image texture analysis. Phys Med Biol 63(19):195017
8. Hiremath PS, Bhusnurmath RA (2014) Texture classification using anisotropic diffusion and local directional binary pattern co-occurrence matrix. In: Proceedings of 2nd International conference on emerging research in computing, information, communication and applications (ERCICA 2014), vol 2, pp 763–769
9. Alharan AF, Fatlawi HK, Ali NS (2019) A cluster-based feature selection method for image texture classification. Indonesian J Electr Eng Comput Sci 14(3):1433–1442. Patel DR, Vakharia V, Kiran MB (2019) Texture classification of machined surfaces using image processing and machine learning techniques. FME Trans 47(4):865–872
10. Armi L, Fekri-Ershad S (2019) Texture image analysis and texture classification methods—a review. arXiv preprint arXiv:1904.06554
11. Rao MS, Reddy BE, Kadiyala R, Prasanna K, Singh S (2021) Texture classification using Minkowski distance measure-based clustering for feature selection. J Electron Imaging 31(4):041204
12. Brodatz P (1966) Textures: a photographic album of artists and designers. Dover Publication, New York