Research Article GU J Sci 36(2): 693-703 (2023) DOI: 10.35378/gujs.919572

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

A Hybrid Deep Learning-Metaheuristic Model for Diagnosis of Diabetic Retinopathy

Omer Faruk GURCAN[1, *], Ugur ATICI[1], Omer Faruk BEYCA[2]

[1]*Sivas Cumhuriyet University, Department of Industrial Engineering, 58140, Sivas, Türkiye*
[2]*İstanbul Technical University, Department of Industrial Engineering, 3436, İstanbul, Türkiye*

Highlights

• This paper focuses on the diagnosis of diabetic retinopathy.
• A hybrid deep learning-metaheuristic model is proposed.
• The hybrid model gives an accuracy of 92.55% on the Messidor-2 dataset.

Article Info

Received: 18 Apr 2021
Accepted: 02 June 2022

Keywords

Deep learning
Metaheuristic
XGBoost
Diabetic Retinopathy
Optimization

Abstract

International Diabetes Federation (IDF) reports that diabetes is a rapidly growing illness. About 463 million adults between 20-79 years have diabetes. There are also millions of undiagnosed patients. It is estimated that there will be about 578 million diabetics by 2030 [1]. Diabetes reasons different eye diseases. Diabetic retinopathy (DR) is one of them and is also one of the most common vision loss or blindness worldwide. DR progresses slowly and has few indicators in the early stages. It makes the diagnosis of DR a problematic task. Automated systems promise to support the diagnosis of DR. Many deep learning-based models have been developed for DR classification. This study aims to support ophthalmologists in the diagnosis process and increase the diagnosis performance of DR through a hybrid model. A publicly available Messidor-2 dataset was used in this study, comprised of retinal images. In the proposed model, images were pre-processed, and a deep learning model, namely, InceptionV3, was used in feature extraction, where a transfer learning approach is applied. Next, the number of features in obtained feature vectors was decreased with feature selection by Simulated Annealing. Lastly, the best representation features were used in the XGBoost model. The XGBoost algorithm gives an accuracy of 92.55% in a binary classification task. This study shows that a pre-trained ConvNet with a metaheuristic algorithm for feature selection gives a satisfactory result in the diagnosis of DR.

1. INTRODUCTION

There are about 422 million diabetic people worldwide. The prevalence of diabetes has increased steadily over the last few years. Diabetes causes many health complications such as kidney failure, nerve damage, leg amputation, heart attacks, risk of fetal death, or vision loss. Diabetic retinopathy (DR) is known as diabetic eye disease, which is a common diabetes complication. Nearly one out of every three diabetic patients develops DR. DR is one of the leading causes of visual impairment or blindness worldwide. The retina is affected in DR. Blood vessels of the retina are injured and become blocked or leaky. Some unusual blood vessels can develop from the retina. These vessels can be the reason for retina scarring or bleeding. At the severe stage of illness, vision loss or blindness occurs [2, 3].

It is crucial to detect the DR before it progresses. Early diagnosing DR can prevent blindness. Diabetes patients should be monitored and checked regularly by ophthalmologists. There is a need for credible diagnosis technology to help ophthalmologists examine fundus images. Many successful tasks are carried out in medicine with artificial intelligence. One of these tasks is the diagnosis oDR. Artificial intelligence-based methods can overcome limitations in DR treatment, especially in low-income countries. Deep learning is a class of machine learning methods in artificial intelligence. Deep learning models have many

*Corresponding author, e-mail: ofgurcan@cumhuriyet.edu.tr

hidden layers and intend to specify the salient features in data, unlike traditional neural network (NN) based models [4-6].

The most applied deep learning models on image datasets are Convolutional Neural Networks (CNNs). Deep CNNs are similar to classic feed-forward NNs. Parameters of the network are adjusted using backpropagation. CNNs differ from the feed-forward networks in four properties: shared weights, local receptive fields, pooling operation, and different layers' aggregation. Deep CNNs start with convolutional layers and ends with fully connected layers. CNNs were originated from the visual cortex of the brain and have superior performance on several tasks, such as image recognition, natural language understanding, speech recognition, automatic video classification systems, or self-driving cars [7].

Heuristics and metaheuristics are approximation methods. These methods can solve NP (nondeterministic polynomial) challenging problems in an acceptable time and promise reasonable solutions. Heuristic algorithms are more task-dependent than metaheuristic algorithms. Metaheuristic algorithms offer a robust solution that combines two search schemas: exploitation and exploration. The exploration searches for the best solution areas. On the other hand, exploitation is responsible for finding new searching regions. Metaheuristics as generic algorithms can be applied to many optimization problems [8, 9].

This study proposed a hybrid deep learning-metaheuristic model for automated diagnosis of DR. A deep learning model, InceptionV3, is used in feature extraction from fundus images. Simulated Annealing is applied in the feature selection process. Overall, the proposed model aims to support ophthalmologists in DR evaluating process.

The paper is presented in six sections: Section 2 gives an overview of the related studies, Section 3 gives dataset information, and explains the proposed model. Section 4 outlines the details of the methods. Results and discussion are given in Section 5. Lastly, future work and conclusions are given in the last section.

2. LITERATURE REVIEW

Early studies in the diagnosis DR were based on feature extraction manually, such as Paranjpe and Kakatkar [10] applied contrast enhancement, top-hat transformation, and morphological filtering methods in blood vessel detection. Hard exudates were detected with image processing methods. Authors extracted texture features using grey level co-occurrence matrices (GLCMs) to classify DR. Harini, and Sheela [11] applied Fuzzy C-Means clustering and some image processing operations (morphological) to extract features from exudates, blood vessels, and microaneurysms which are used as inputs for Support Vector Machines. Punithavathi and Kumar [12] detected the area of microaneurysms through image transformation, top hat transformation, and Otsu's thresholding. The authors also calculated statistical texture properties like mean and entropy. Extreme Learning Machine was used in classification. Colomer et al. [13] locally computed (by dividing images in patches) granulometric profiles and local binary patterns to extract morphological and texture information from retinal images. Extracted features were fed Support Vector Machines, Random Forests, and Gaussian Processes classifiers.

CNN-based models, which have performed well in DR classification tasks, automatically extract features from retinal images. Training deep CNNs from scratch requires large datasets. So, transfer learning is especially suitable in DR classification tasks, where the number of retinal images is low.

Johari et al. [14] used AlexNet architecture and the Messidor dataset. The authors made a binary classification: normal images vs. exudates images. Takahashi et al. [15] trained a randomly initialized GoogLeNet model with a private dataset. Some authors used multiple networks in diagnosis DR. Such as Choi et al. [16] used AlexNet and VGG19 architectures. Zhang et al [17] used DenseNet169 & 201, and ResNet50 for feature extraction. Bodapati et al. [18] proposed a blended feature extraction model. Authors fused features from VGG16 and Xception. On the other hand, several studies have introduced new architectures [4], such as Pratt et al. [19] developed a new network with CNN architecture.

Besides deep learning-based feature extraction, feature selection is also essential in image classification tasks. There are many feature selection techniques to find the optimal subset of features from a given data in the literature. Some points are essential in deciding on a feature selection technique, such as contribution to the classifier's performance, reducing overfitting, or minimizing training time [20-23]. Metaheuristics have the potential to perform well in the feature selection process. Such as Canayaz [24] proposed a model that diagnoses COVID-19 from X-ray images. The author used deep learning models in feature extraction and applied two metaheuristic algorithms in feature selection. According to the results, BPSO decreased the number of features and increased classification performance.

In this study, features from retinal images are extracted by InceptionV3, and binary classification (non-referable DR: NRDR against referable DR: RDR) is made. To compare the proposed model's result with other studies in the literature, we summarized the studies that use the same testing dataset (Messidor-2), extracted features using the InceptionV3, and made a binary classification.

Voets et al. [25] used InceptionV3 for feature extraction. The authors applied ensemble learning to perform a binary classification (NRDR-RDR). A publicly available EyePACS dataset is used in training. The proposed model gives an AUC of 85.3% on Messidor-2. Li et al. [26] trained the last layers of Inception V3 networks and used a dataset from Chinese Hospitals, including 19,233 retinal images in training. Their model gives an accuracy of 93.49% on Messidor-2 in RDR diagnosis. Toledo-Cortés et al. [27] used InceptionV3 to diagnose RDR. The authors proposed a deep learning Gaussian Process (GP) model trained on the EyePACS dataset. A GP regressor gives an AUC of 87.87% on Messidor-2. Gurcan et al. [28] obtained feature representations from retina images using InceptionV3 pre-trained weights. The authors used Messidor-2 in training and testing. XGBoost was used as a classifier with an accuracy of 91.40% on Messidor-2.

3. CASE STUDY

3.1. Dataset

A publicly available Messidor-2 dataset [29-31] was used in this study. This dataset includes 1748 fundus images in JPG and PNG formats. The images are taken from patients with five grade levels: no, mild, moderate, severe, and proliferative DR. Seven images were excluded from the analysis. NRDR against RDR classification is made. This categorization is one of the most preferred distinctions in literature. The NRDR class comprises no DR and mild DR images. RDR class comprises moderate, severe, and proliferative DR images. There are 455 images in RDR and 1286 images in NRDR classes. Seven images were excluded from the study because of poor image quality.

3.2. Proposed Model

The proposed model has four steps. The dataset is comprised of different-sized images. The pixel values of each image should be scaled before feeding images as input to a deep learning model. So firstly, retinal images are pre-processed. Then, processed images are fed into a deep CNN network, and abstract features are obtained using the transfer learning approach. Extracted features are summarized by applying the Global Average Pooling operation. A metaheuristic algorithm is applied to reduce the number of features while obtaining the best potential features. Lastly, extracted and selected features are classified with an ensemble method which is a decision tree-based algorithm. The details of the methods are explained in the following section. The proposed model is presented in Figure 1.

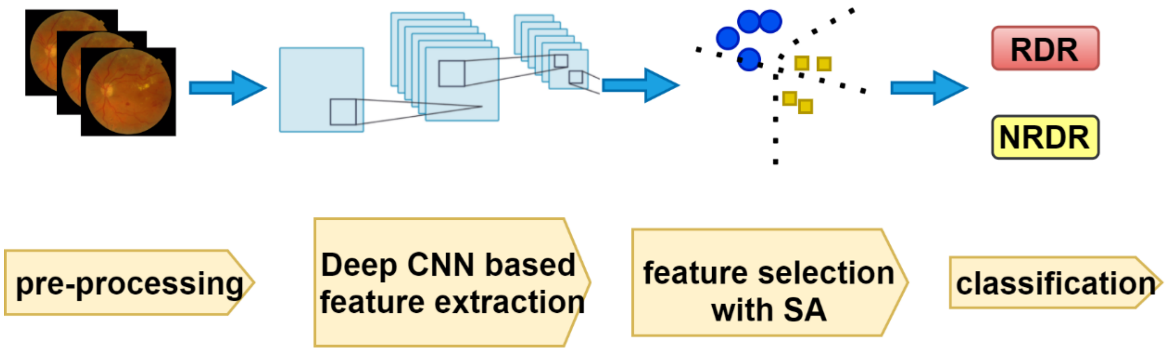

Figure 1. *The proposed model*

4. METHODS

This study comprises four processes: pre-processing, feature extraction, feature selection, and classification. Details of processes are explained in the following sections.

4.1. Pre-processing

The Messidor-2 dataset has fundus images in different sizes. So, images are resized to a resolution of 1200 (width) × 960 (height). The pixel value of each image is rescaled from the range of [0-255] to the range of [0-1] in red, green, and blue (RGB) channels by dividing pixel values 255 in the whole dataset. This operation is referred to as normalization, preferred in deep neural network model training. Then, pre-processed images are fed into the following process, feature extraction.

4.2. Feature Extraction

Deep learning models require much data to successfully learn the latent patterns of data. In some problems, it isn't easy to collect large datasets and label them, such as medical tasks. Limited data and costly labeling processes bring difficulties in developing well-performing deep learning models for medical tasks. The transfer learning approach helps to solve problems with an insufficient number of training data. Besides, in transfer learning, training and testing data are not necessarily independent and identically distributed [32]. A pre-trained network can be used as a starting point to learn a problem in a different domain.

There are deep learning models that are made available with pre-trained weights. These pre-trained models can be used for various tasks such as feature extraction, fine-tuning, or prediction. Fine-tuning a network is generally much easier and faster than training a network from scratch, which initializes network weights randomly. Learned features can be transferred to new problems using a small number of data.

The authors used InceptionV3 [33] pre-trained weights on ImageNet in feature extraction from fundus images. InceptionV3 is one of the most preferred models in image classification tasks. InceptionV3 has a depth of 159 layers and 23,851,784 parameters. The model uses building blocks that include convolution layers, inception modules, pooling, concatenation, batch norm, and fully connected layers. Inception modules help decrease computational expense and increase the network's depth by using stacked 1×1 convolutions. It uses multiple kernel filter sizes within the CNN and orders them to run on the same level [34]. The InceptionV3 model clusters similar sparse nodes into a dense structure. This operation increases the width and depth of the model and reduces computational expense [35].

The pre-processed images are fed into the InceptionV3, which has ten mixed (concatenate) layers through the model. Features are extracted from the mixed-3 layer, which is one of the initial layers of the InceptionV3. Pre-trained weights of InceptionV3 on ImageNet are used. In deep CNNs, spatial hierarchies of features are learned. Deeper layers encode high-level features (abstract features) such as a human nose, while initial layers work like edge detectors and detect simple shapes, lines, and edges from the data. Since pre-trained weights are used, features obtained from the mixed-3 layer give more valuable information than a deeper layer.

After feature extraction, 3,251,712 (73 × 58 × 768) features are obtained from each fundus image. That means 768 feature maps in 73 × 58 are obtained from each image. These features are summarized by the Global Average Pooling (GAP) operation. The GAP is a downsampling operation that downsamples an entire feature map into a single value by averaging the values. It summarizes the features in an image aggressively. The GAP application on extracted features gives 768 (1 × 1 × 768) features for each fundus image. The obtained features from the GAP operation are sent to feature selection and classification processes. The 73 × 58 values in each feature map are averaged and obtained 768 values are flattened in the subsequent processes.

4.3. Feature Selection

Simulated Annealing (SA) is one of the popular heuristic methods widely used to solve combinatorial optimization problems. The advantage of this approach is that it does not stick to the local optimum [36]. Random numbers with a uniform distribution between 0 and 1 were generated for the initial solution (*S*). If the randomly generated number is greater than 0.5, the relevant feature is included in the solution. The objective function value (*SObj)* of the initial solution is calculated using the k-Nearest Neighbor (*kNN*) Classifier. Neighbor solution (*N*) is created similar to the initial solution. The objective function value (*NObj*) of the neighboring solution is calculated. If the neighbor solution objective function N_{OBJ} is better than the solution objective function value ($\Delta OBJ=NObj\text{-}SObj,\ \Delta OBj<0$), the solution is replaced by the neighbor solution (S = N). If the neighbor solution does not improve the objective function’s value, the solution is not changed with a probability of $e^{(-\Delta Obj/T)}$ and the next iteration is passed. The annealing temperature is calculated using the geometric ratio. The annealing temperature ($T = T * Cr$) is updated by multiplying the annealing (*T*) with the cooling coefficient ($Cr = 0.999$) at each iteration. The iteration process is continued until $T=T_{end}$ [37]. The pseudo-code of the feature selection process is given in Table 1.

Table 1. *Pseudo-Code of feature selection with simulated annealing*

```
1.  Set parameters features (i), T, Cr, T_end
2.  Generate Initial Solution
        d← Dimension of features
        Repeat
        if rand(1,i) > 0.5 select feature
        Until i=d
3.  Calculate Obj by using kNN
4.  Repeat
        Generate Neighbor Solution
        Repeat
        if rand(1,i) > 0.5 select features(i)
        Until i=d
        Calculate  NObj by using kNN
        Calculate ΔOBJ ← NObj-SObj
        if  ΔOBJ<0 then S←N
        if  ΔC<0 or e^(-ΔC/T)>rand(0,1)  then
        S←N
        T←T*Cr
        Until T=T_end
5.  Output S
```

kNN classifier is applied in calculating the fitness value of SA. The classification methods generally seek boundaries that are linear or non-linear to separate data optimal way. These boundaries are then applied in predicting the classes of new instances. kNN algorithm has a different technique by using geographic neighborhood information of an instance to predict the instance’s class. kNN for a classification task predicts a new instance using the number of k closest instances from the training dataset. The closeness is calculated using a distance metric, like Minkowski and Euclidian. Deciding on metrics relayed to characteristics of problems [38].

In the most widely preferred application of kNN, uniform weights are used in weighting neighborhoods, where the classifier assigns a query point into a class using the majority vote of nearest neighbors. In some cases, weighting the neighbors according to their contribution is better. A frequently used weighting is assigning weights in direct proportion to the inverse of the distance from the related query point.

4.4. Classification

Tree boosting methods are very effective and widely applied in many machine learning challenges. XGBoost is short for "Extreme Gradient Boosting." It is a scalable end-to-end tree boosting system [39]. It is a supervised learning machine learning system that can handle tabular or structured datasets in regression and classification tasks.

XGBoost is developed based on a gradient boosting decision tree algorithm. Gradient boosting is a technique in which new models are generated that predict errors or residuals made by previous models and later added jointly to perform the final prediction. Its name comes from the gradient descent algorithm. When new models are added, the loss is minimized using a gradient descent algorithm [40].

Similar to other boosting algorithms, there is a dependence between the training of each model and the models already trained in gradient boosting. The learning process aims to build the base models that are correlated in maximum with the negative gradient of loss function related to the whole ensemble. In other words, regression models in a sequence are calculated, in which each successive model makes a prediction of pseudo residuals of antecedent models given a differentiable loss function. This loss function is needed to calculate the negative gradient. The loss function is minimized during the aggregation of predictions. The gradient boosting models have generally had many simple models in response to other ensemble methods with less but more complex models. During the training, deciding on the number of iterations (models) is critical. Too many numbers can reason for overfitting, while very few numbers can reason for underfitting. Validation methods help to choose the correct number of iterations. Gradient boosting uses decision trees as base learners generally [41].

The key success factor of XGBoost is its scalability. The system operates more than ten times faster than available solutions on a machine and scales to the excessive number of examples in memory-limited or distributed settings. The property of scalability results from some innovations, including specific optimization techniques and systems. The proposed tree learning algorithm can handle sparse data with nodes' default directions; the algorithm uses a weighted quantile sketch procedure that provides handling weights of instances in approximate tree learning. Parallel and distributed computing contribute to the learning speed and exploration process of the model. XGBoost benefits from the out-of-core computation. It uses a cache-aware structure that helps researchers process millions of instances on a desktop [38]. This study obtained features from feature extraction and feature selection processes classified with XGBoost. The algorithm made a binary classification of DR (NRDR vs. RDR).

4.5. Experimental Setup

For the experiments, the Messidor-2 dataset is randomly divided into training data (80%) and testing data (20%). The features are extracted from the mixed3 layer of the InceptionV3 network, which comprises ten mixed layers. The features are extracted from mixed10, mixed9, mixed8, etc. Because pre-trained weights are used in the study without layer-wise tuning, the initial layers are more valuable than later layers of InceptionV3. So in the experiments, features extracted from the mixed3 layer give the highest classification accuracy with XGBoost.

In the feature selection step, the kNN classifier is used as the fitness function of SA algorithm. For the kNN algorithm, euclidean distance is used as a distance metric, and k-value (the number of neighbors) is determined by grid search, and it is searched in the range of [2, 10]. A Uniform weight function is used where all points in each neighborhood are weighted equally. 5-fold cross-validation is made for kNN. The highest performance is obtained when the k-value of kNN equals 5.

The number of iterations in SA is controlled by annealing temperature (T), cooling coefficient (Cr), and end temperature (T_{end}) parameters. For SA, a grid search is carried out for T in the values of 50, 100, and 200. The parameters are chosen as follow: T=200; c_r=0.999; T_{end}=1. Experimental results for various T values are given in Table 2.

Table 2. *Experimental results for various T values*

T	Duration (seconds)	Number of extracted features	Accuracy performance with XGBoost (%)
50	235	374	89.68
100	639	379	90.54
200	731	367	92.55

In the classification step, grid search is carried out for XGBoost parameters as follows: learning rate in the range of [0.01, 0.1]; max depth in the range of [3, 9]; the number of estimators in the range of [100, 1500]. The other parameters of XGBoost are set to default ones. The highest accuracy performance is achieved when the learning rate = 0.06, maximum depth = 3, and the number of estimators = 900.

All models are built-in Python using Keras, Scikit-learn, and XGBoost Libraries. Experiments are conducted using NVIDIA GeForce RTX2070 8GB graphical process unit, 32 GB RAM, and AMD Ryzen 7 3700X 3.6 GHz processor.

5. RESULTS AND DISCUSSION

In this study, features are extracted from one of the initial layers of the InceptionV3 network. Pre-trained weights on ImageNet are used. Global Average Pooling operation is applied on extracted features, and an abstract feature vector, which comprises 768 features, is obtained. SA Algorithm is applied as a feature selection method on abstract feature vector to decrease the number of features and find the best potential features. The obtained feature number is decreased from 768 to 367 with SA.

Both feature vectors: extracted and selected, are classified with XGBoost, and results are compared in Table 3. The confusion matrix of the binary classification task (nRDR vs RDR) is visualized in Figure 2. According to the confusion matrix, 273 out of 280 nRDR cases are classified correctly, and 50 out of 69 RDR cases are classified correctly.

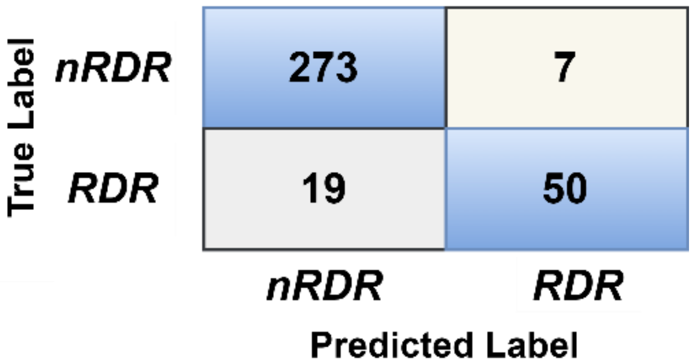

Figure 2. *Confusion matrix for XGBoost*

Table 3. *Analysis results of diagnosis RDR on Messidor2*

Methods	Number of features	Accuracy (%) of XGBoost
Extracted Features from InceptionV3	768	91.40
Selected Features with SA	367	92.55

Table 4 presents the performance comparison of InceptionV3 based studies in diagnosing RDR using Messidor-2. The proposed model splits Messidor-2 into train and test sets, while the studies used in the comparison [25-27] used different and considerably more image data in training. Additionally, because fine-tuning is not made, which increases training times considerably, the current study is not computational expensive against the other studies that made fine-tuning. So the current study offers competitive results.

Table 4. *Performance comparison on Messidor-2 in binary classification (RDR vs. NRDR)*

Study	Approach	Training Data	Accuracy (%)	AUC (%)
Voets et al. [25]	InceptionV3 CNN	EyePACS (45,717)	Not given	85.30
Li et al. [26]	InceptionV3 CNN	Custom (19,233)	93.49	99.05
Toledo-Cortés et al. [27]	InceptionV3 + GP Regressor	EyePACS (56,827)	Not given	87.87
Gurcan et al. [28]	InceptionV3 + XGBoost	Messidor-2	91.40	93.55
Proposed Model	InceptionV3 + SA + XGBoost	Messidor-2	92.55	94.32

SA feature selection enabled increased accuracy level in NRDR-RDR classification by decreasing the number of features by about 52.2%. The accuracy score of 92.55% is a competitive result on the Messidor-2 dataset. The proposed hybrid model offers some advantages:

- The model applies very few pre-processing steps.
- Training time is less, and the model doesn't require a high computational power because of transferring pre-trained weights.
- The proposed model gives a satisfactory performance with a limited number of images.
- Metaheuristic algorithm decreased feature number significantly while increasing accuracy performance.

6. CONCLUSIONS

DR is one of the most common vision loss or blindness worldwide. Future projections reveal that diabetic patients will increase. Recently, automatic models based on deep learning performed well in many tasks. Our study's primary aim is to help the diagnosis process of DR in low source settings: computational power, lack of retinal images, lack of experience, and the number of experts. For this purpose, a hybrid deep learning-metaheuristic model is proposed. A deep learning network, InceptionV3, is used in feature extraction, and the best potential features are selected with Simulated Annealing. XGBoost is used as a classifier. The hybrid model gives an accuracy of 92.55% in the binary classification task on the Messidor-2 dataset. This result is competitive with other studies in the literature.

For feature studies, other publicly available datasets can be used. The hybrid model can be extended by adding other networks and operations. Some other feature extraction methods can be applied, such as wavelet transform.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf. Access date: 02.03.2021

[2] https://apps.who.int/iris/bitstream/handle/10665/336660/9789289055321-eng.pdf. Access date: 02.03.2021

[3] Wong, T. Y., Sabanayagam, C., "Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence", Ophthalmologica, 243(1): 9-20, (2020).

[4] Kandel, I., Castelli, M., "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review", Applied Sciences, 10(6): 1-24, (2021).

[5] Yazici, I., Beyca, O. F., Gurcan, O. F., Zaim, H., Delen, D., Zaim, S., "A comparative analysis of machine learning techniques and fuzzy analytic hierarchy process to determine the tacit knowledge criteria", Annals of Operations Research, 1-24, (2020).

[6] Şenyiğit, E., Atici, U., "Artificial neural network models for lot-sizing problem: a case study", Neural Computing and Applications, 22(6): 1039-1047, (2013).

[7] Buyya, R., Calheiros, R. N., Dastjerdi, A. V., "Big data: principles and paradigms", Morgan Kaufmann, India, (2016).

[8] Kızılkaya Aydoğan, E., Delice, Y., Özcan, U., Gencer, C., Bali, O., "Balancing stochastic U-lines using particle swarm optimization", Journal of Intelligent Manufacturing, 30: 97-111, (2019).

[9] Abdel-Basset, M., Abdel-Fatah, L., Sangaiah, A. K., "Metaheuristic Algorithms: A Comprehensive Review", In A. K. Sangaiah, M. Sheng & Z. Zhang, Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications, Cham., (2018).

[10] Paranjpe, M. J., Kakatkar, M. N., "Automated diabetic retinopathy severity classification using support vector machine", International Journal for Research in Science & Advanced Technologies, 3(3): 86-91, (2013).

[11] Harini, R., Sheela, N., "Feature extraction and classification of retinal images for automated detection of Diabetic Retinopathy", Second International Conference on Cognitive Computing and Information Processing, Mysuru, India, (2016).

[12] Punithavathi, I. H., Kumar, P. G., "Severity grading of diabetic retinopathy using extreme learning machine", International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, Krishnankoil, India, (2017).

[13] Colomer, A., Igual, J., Naranjo, V., "Detection of early signs of diabetic retinopathy based on textural and morphological information in fundus images", Sensors, 20(4): 1-21, (2020).

[14] Johari, M. H., Hassan, H. A., Yassin, A. I. M., Tahir, N. M., Zabidi, A., Rizman, Z. I., Wahab, N. A., "Early detection of diabetic retinopathy by using deep learning neural network", International Journal of Engineering and Technology, 7(4): 198-201, (2018).

[15] Takahashi, H., Tampo, H., Arai, Y., Inoue, Y., Kawashima, H., "Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy", Plos One, 12(6): e0179790, (2017).

[16] Choi, J. Y., Yoo, T. K., Seo, J. G., Kwak, J., Um, T. T., Rim, T. H., "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database", Plos One, 12(11): e0187336, (2017).

[17] Zhang, W., Zhong, J., Yang, S., Gao, Z., Hu, J., Chen, Y., Yi, Z., "Automated identification and grading system of diabetic retinopathy using deep neural networks", Knowledge-Based Systems, 175: 12-25, (2019).

[18] Bodapati, J. D., Veeranjaneyulu, N., Shareef, S. N., Hakak, S., Bilal, M., Maddikunta, P. K. R., Jo, O., "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction", Electronics, 9(6): 914-930, (2020).

[19] Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., Zheng, Y., "Convolutional neural networks for diabetic retinopathy", Procedia Computer Science, 90: 200-205, (2016).

[20] Dener, M., Akcayol, M. A., Toklu, S., Bay, Ö. F., "Genetic algorithm based a new algorithm for time dynamic shortest path problem", Journal of The Faculty of Engineering and Architecture of Gazi University, 26(4): 915-928, (2011).

[21] Utku, A., Muhammet A. A., "Deep Learning Based Prediction Model for The Next Purchase", Advances in Electrical and Computer Engineering, 20: 35-44, (2020).

[22] Çerçioğlu, H., Özcan, U., Gökçen, H., Toklu, B., "A Simulated Annealing Approach for Parallel Assembly Line Balancing Problem", Journal of The Faculty of Engineering and Architecture of Gazi University, 24(2): 331-341, (2009).

[23] Moorthy, R. S., Pabitha, P. A., "Study on Meta Heuristic Algorithms for Feature Selection", International Conference on Intelligent Data Communication Technologies and Internet of Things, Springer, Cham., (2018).

[24] Canayaz, M., "MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images", Biomedical Signal Processing and Control, 64: 102257, (2021).

[25] Voets, M., Møllersen, K., Bongo, L. A., "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", Plos One, 14(6): 0217541, (2019).

[26] Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X., Wu, Z., "Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm", Translational Vision Science & Technology, 8(6): 4-4, (2019).

[27] Toledo-Cortés, S., De La Pava, M., Perdomo, O., González, F. A., "Hybrid Deep Learning Gaussian Process for Diabetic Retinopathy Diagnosis and Uncertainty Quantification", In International Workshop on Ophthalmic Medical Image Analysis, Springer, Cham., (2020).

[28] Gurcan, O. F., Beyca, O. F., Dogan, O., "A Comprehensive Study of Machine Learning Methods on Diabetic Retinopathy Classification", International Journal of Computational Intelligence Systems, 14(2): 1132-1141, (2021).

[29] Abràmoff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D. C., Quellec, G., & Niemeijer, M., "Automated analysis of retinal images for detection of referable diabetic retinopathy". JAMA Ophthalmology, 131(3): 351-357, (2013).

[30] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J. C., "Feedback on a publicly distributed image database: the messidor database", Image Analysis & Stereology, 33(3): 231-234, (2014).

[31] http://www.adcis.net/en/thirdparty/messidor2/. Access date: 01.02.2021

[32] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., "A survey on deep transfer learning", In International Conference on Artificial Neural Networks, Springer, Cham., 270-279, (2018).

[33] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., "Rethinking the inception architecture for computer vision", IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, (2016).

[34] https://deepai.org/machine-learning-glossary-and-terms/inception-module. Access date: 01.02.2021

[35] Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X., Wu, Z., "Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm", Translational Vision Science & Technology, 8(6): 4-4, (2019).

[36] Wu, C. C., Hsu, P. H., Lai, K., "Simulated-annealing heuristics for the single-machine scheduling problem with learning and unequal job release times", Journal of Manufacturing Systems, 30(1): 54-62, (2011).

[37] Kim, D. W., Kim K. H., Jang, W., Chen, F. F., "Unrelated parallel machine scheduling with setup times using simulated annealing", Robotics and Computer-Integrated Manufacturing, 18(3): 223-231, (2002).

[38] Kuhn, M., Johnson, K., Applied predictive modeling, Springer, New York, (2013).

[39] Chen, T., Guestrin, C., "Xgboost: A scalable tree boosting system", 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, (2016).

[40] Brownlee, J., "XGBoost With Python: Gradient Boosted Trees with XGBoost and Scikit-Learn". Machine Learning Mastery, (2016).

[41] Rokach, L., "Ensemble learning: Pattern classification using ensemble methods", World Scientific, (2019).