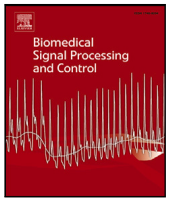
Biomedical Signal Processing and Control 81 (2023) 104534

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Traditional machine learning algorithms for breast cancer image classification with optimized deep features

Furkan Atban *, Ekin Ekinci, Zeynep Garip

Sakarya University of Applied Sciences, Faculty of Technology, Computer Engineering Department, Sakarya, Turkey

ARTICLE INFO

Keywords:
Histopathological image classification
Transfer learning
Meta-heuristic algorithms
Machine learning

ABSTRACT

For breast cancer diagnosis, computer-aided classification of histopathological images is of critical importance for correct and early diagnosis. Transfer learning approaches for feature extraction have made significant progress in recent years and are now widely used. To select the best representative features to classify breast cancer pathological images and avoid the curse of dimensionality, this work uses optimized deep features. The proposed approach firstly employs ResNet18 architecture for feature extraction to achieve deep features. Then, meta-heuristic algorithms namely Particle Swarm Optimization (PSO), Atom Search Optimization (ASO) and Equilibrium Optimizer (EO) algorithms, are employed to provide more representative features of breast cancer pathological images. To understand the effect of optimized deep features on classification, traditional machine learning (ML) algorithms are used. The experimental analysis of the proposed approach has been done on the public benchmark dataset BreakHis. Experimental results illustrate that, for features obtained from ResNet18-EO, the proposed approach achieves a 97.75% F-score by using the Support Vector Machine (SVM) with gaussian and radial-based functions (RBF).

1. Introduction

Cancer is a deadly disease caused by malignant tumors that occur with the abnormal division and proliferation of cells in the organs or tissues that make up our body. Breast cancer is a type of cancer that occurs when these malignant tumors arise on the breast tissue. According to World Health Organization (WHO),[1] in 2020, an estimated 19.3 million cancer diseases were detected in the world and approximately 10 million people died due to cancer. In this process, female breast cancer has surpassed other cancers as the most frequently diagnosed cancer among other female cancers [1,2]. With the increase in breast cancer cases, it is inevitable that medical experts who diagnose and treat cancer are faced with a limited time for diagnosis and treatment [3]. Further, analyses performed by medical experts are often done in an observational way by using visualization systems. However, factors such as the huge amount of medical images and the difficulty dealing with features have made manual classification by medical experts more complex and challenging [4,5]. In these cases, different analysis techniques have become an important need in this field.

In this direction, a significant degree of convenience is provided by medical image analysis for the diagnosis of many diseases [6]. Significant features are of the utmost relevance when using ML algorithms for medical images. Due to this, the majority of earlier algorithms used hand-crafted features by analyzing medical images. However, the hand-crafted features with task-specific limits are unable to provide much useful information and discriminative characteristics [7]. Analysis of medical images for breast cancer diagnosis has become more successful and faster with the use of deep learning techniques. Feature extraction with transfer learning networks provides features with higher accuracy than conventional feature extraction techniques (hand-crafted feature extraction) for image processing [8].

Meta-heuristic algorithms are algorithms that operate on probability but are used to obtain the optimum solution in high-level working environments by consciously performing a logical search. While not specific to any particular problem, they can be used to solve many problems such as complex scheduling problems [9,10], space allocation problems [11,12], clustering problems [13,14], cloud computing [15,16], parameter optimization [17,18] and so on. They usually give an approximate solution to the optimum solution. In order to obtain a good result over meta-heuristic algorithms, the method to be used must be correctly associated with the relevant problem.

As deep learning has advanced, features are now automatically acquired, and it has shown an increasing capacity to extract high-dimensional features [19,20]. The fact that data with very high dimensional features is the key issue here [21,22]. However, many features are redundant and irrelevant, which not only makes data analysis more

* Corresponding author.
E-mail address: furkanatban@subu.edu.tr (F. Atban).

[1] https://www.who.int/news-room/fact-sheets/detail/cancer

https://doi.org/10.1016/j.bspc.2022.104534
Received 17 May 2022; Received in revised form 4 October 2022; Accepted 18 December 2022
Available online 22 December 2022

time- and space-intensive but also reduces the effectiveness and generalization capacity of ML algorithms [23]. Consequently, feature selection has emerged as a key method for dealing with high-dimensional datasets.

In the literature, several meta-heuristic algorithms have been devised and applied to handle feature selection problems. These algorithms include Cuckoo Search (CS) [24], Flower Pollination Algorithm (FPA) [25], Dragonfly Algorithm (DA) [26,27], PSO [28], Genetic Algorithm (GA) [29], Artificial Bee Colony (ABC) [30], Ant Colony Optimization (ACO) [31], Grey Wolf Optimization (GWO) [32], WOA [33], EO [34], Bat Algorithm (BA) [35], Harris Hawks Optimisation (HHO) [36], and hybridization of them [37–39].

The recent success of pre-trained CNN architectures for feature extraction has prompted us to employ them for histopathology image classification, considering the need for correct and early diagnosis [40]. As a result, a framework based on ResNet18, a prominent CNN, is used in this study to generate an accurate and reliable feature set. Following that, the number of features in this set is reduced with different meta-heuristic algorithms PSO, ASO and EO to handle the curse of dimensionality. The model is then subjected to various classifiers in order to learn different patterns from these feature sets. For binary classification of cancer images, the proposed framework has presented an optimized deep feature classification with SVM, K Nearest Neighbour (KNN) and Decision Tree (DT). Experiments are conducted on the BreakHis dataset, and the best performance of 97.73% accuracy, 97.75% Precision, 97.75% Recall, and 97.75% F1-score is achieved on 200x magnification factor breast cancer histopathological images with ResNet18-EO and SVM. The main contributions of this paper are as follows:

- Proposing a model for early and accurate diagnosis of breast cancer
- Presenting a pipeline covering feature extraction, feature selection and classification
- Obtaining the minimal subset of features in all magnification factors using the proposed ResNet18-PSO, ResNET18-ASO, ResNet18-EO
- Comparing the performance of proposed feature selection algorithms with classification algorithms namely KNN, SVM, DT using histopathological images

The rest of the paper is organized as follows. Section 2 provides a brief review of the studies in which BreakHis classification is realized. In Section 3, we discuss the dataset and methods applied in the study. In Section 4, the experimental setup of the algorithms, feature selection and the experimental results are given in detail. Section 5 concludes the paper.

2. Literature review

In this section, the researches on the classification of BreakHis data have been presented [41]. We review briefly here the most recent research achievements and the advancements and applications in the biomedical field.

Hand-crafted features have a long history in computer vision tasks and are still in widespread use. Spanhol et al. used six hand-crafted feature descriptors namely Local Binary Patterns (LBP), Completed Local Binary Pattern (CLBP), Local Phase Quantization (LPQ), Gray Level Co-Occurrence Matrices (GLCM), Parameter-Free Threshold Adjacency Statistics (PFTAS), and Oriented FAST and Rotated BRIEF (ORB), in conjunction with four different classifiers: 1- NN, Quadratic Linear Analysis (QDA), SVM, and Random Forest (RF) [41]. The best result was achieved using PETAS and SVM together with a 0.851 accuracy rate in the 200X magnification factor images. Samah et al. evaluated into various hand-crafted features such as LBP, GLCM, pyramid-structured wavelet transforms (PWT) and tree-structured wavelet transform (TWT) with k-NN [42]. Experiments were conducted on only 400X images and the best overall accuracy rate was obtained with the PWT descriptor which was 85.62%. Kahya et al. attempted to realize adaptive sparse SVM (ASSVM) by combining L1 norm with SVM to select the most informative features from BreakHis images [43]. However, as a result of experiments, it was observed that the L1 norm performed inconsistent feature selection. Therefore, to improve feature selection Wilcoxon rank-sum based weighting was applied as a step to ASSVM. Compared with state-of-the-art algorithms, ASSVM performed better with fewer features for all magnification levels. Chattoraj and Vishwakarma extracted Zernaike moment, image entropy and fractal dimension – extracted from variational mode decomposition (VMD) components – as feature descriptors and applied the ReliefF algorithm to determine the most informative one [44]. Zernaike moment with Least-squares SVM (LS-SVM) was used as a classifier and it was seen that the best success was achieved with 88% accuracy for 200X images. In [45], KAZE features extracted from nonlinear scale spaces were fed into as input to Bag-of-Features (BOF) classifier. Then, the output of BOF was transformed into a feature histogram using the nearest neighbor algorithm and SVM was used for classifying images as benign or malignant. In the case of the 40X magnification factor, 91.53% was obtained as the best result.

CNN's feature extraction capability and representation power of extracted features are higher than hand-crafted ones, so it has been used instead of hand-crafted. Spanhol et al. adopted CNN for both feature extraction and classification tasks with different patch image generation strategies and fusion rules [46]. Among patch image generation strategies, randomly extracted 64 x 64 patch size was the best with 89.6% for 40X images; with max fusion rule, 90% for 40X images were achieved. Zerouaoui and Idri proposed 28 different hybrid approaches in which for feature extraction seven different deep learning models (DenseNet201, MobileNetV2, ResNet50, InceptionV3, Inception ResNetV2, VGG16, and VGG19) and for classification four different ML algorithms (MLP, SVM, DT, and KNN) are employed [8]. In the design of hybrid approaches, the usage of DenseNet 201 and MLP for classification achieved the best results.

CNN extracted features and hand-crafted features are also used together in the studies. In [47], the authors combined CNN based global features with hand-crafted features based local features namely The Contourlet Transform (CT), LBP and Histogram information and frequency-domain information namely Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT). The best performance was provided with CNN-CH for 200X images with 97.19% accuracy. Cascianelli et al. [48] applied Principal Component Analysis (PCA), Gaussian Random Projection (GRP) and Correlation-Based Feature Selection (CBFS) on CNN-based features extracted from images and concluded that while PCA, GRP and CBFS provided a significant reduction in the number of features, these methods also caused a moderate decrease in accuracy.

With the advent of pre-trained models, feature extraction is realized more simply. Spanhol et al. used pre-trained AlexNet and A Deep Convolutional Activation Feature (DeCAF) to classify breast cancer histopathological images [49]. The DeCAFs extracted from the 6th, 7th and 8th layers of AlexNet were used separately and in pairs with 1, 4 and 16 patches. For single-layer experiments, the 7th layer with 16 patches performed better for 200X images with 86.0% accuracy, for combined-layers experiments, the 7th and 8th layers with 16 patches performed better for 200X images with 86.3% accuracy. In [50], it was proposed a sequential framework with maximum voting. The proposed model realized deep multi-layered feature fusion by extracting high-level features from different layers. The best results were obtained for 200X images with 96.76% accuracy. Boumaraf et al. devised a new magnification-dependent and independent classification approach based on transfer learning for histopathology images [51]. They optimized the network by using ResNet18 as the backbone model and with a block-wise fine-tuning strategy. The accuracy rate of two-class images was 98.42%. In [52], it was built a classification model which

fused features and enhanced routing. The new fusion process combined semantic information extracted by BkNet with spatial features extracted by CapsNet. The proposed FE-BkCapsNet model was compared to BkNet, CapsNet, and CapsNet with enhanced routing but no capsule-level feature fusion (E-CapsNet), and it was found that FE-BkCapsNet improved classification accuracy for all magnification factors. Benhammou et al. compared CNN + InceptionV3 with classical methods used in [41], CNN+Pre-Processing (PP), CNN+PP+ Ensemble Learning (EL), CNN+PP+ Transfer Learning (TL) [53]. For 40X image, CNN+InceptionV3 achieved 93.0% F1-score. Wang et al. suggested a novel cross-domain transfer learning and multi-stage feature fusion technique (Cd-dtffNet) for histopathological image classification [54]. With cross-domain transfer learning, features were extracted from hierarchical layers, then adaptive feature fusion with L2 regularity on the fusion weights was realized and higher accuracy was obtained.

For more efficient feature extraction residual blocks are also preferred in the studies. Zou et al. proposed AHoNet which combined high-order statistics and attention mechanism modules in the residual network [55]. To produce local salient deep features, AHoNet used a channel attention module without dimensionality reduction and local cross-channel interaction. Then, second-order covariance statistics were computed to provide a more robust global feature presentation of salient deep features. Finally, AHoNet global and local salient features were combined. 99.292% accuracy was achieved for 200X images with the proposed architecture. Hu et al. developed myResNet-34 which was a residual learning-based CNN [56]. In this model, classification was realized in three steps namely stain normalization, data augmentation and deep learning-based classification. With 94.03% accuracy for 40X images, myResNet-34 dominated the compared models. Ibraheem et al. designed 3PCNNB-Net which had three stages network composed of three parallel CNN branches [57]. Three parallel CNN branches with deep residual blocks make up the first step. The three parallel branches were fused in the second stage to form a feature fusion path. In the third stage, the fused features were classified. Compared with state-of-the-art methods, 3PCNNB-Net obtained 97.04% accuracy on 200X images. Barzekar and Yu devised a new CNN model which was the concatenation of multiple networks outer, middle, and inner named C-Net [58]. While outer and middle were responsible for feature extract and inner was responsible for classification. With C-Net architecture 99.67% accuracy was reported for 200X as the best result.

In some of the works, pre-trained CNN's are used for fine-tuning. Zhi et al. devised a two-stage fine-tuning method. Firstly, only the fully connected (FC) layers were retrained then the entire network [59]. The authors used transfer learning, which transferred weights from ImageNet pretraining, on VGGNet and a custom six-layer CNN architecture to make a classification. The best results were obtained with the custom six-layer CNN architecture for 200X images with 94.8% accuracy. In another study, a deep manifold fusion of multilayer features named LPMF2Net were proposed to take into account low-level features [60]. To begin, a deep transfer model was created to extract convolutional features at various levels. Then, a deep manifold feature fusion layer was realized for feature reduction to obtain more significant features. Finally, to improve the classification accuracy, the objective function local preserving regularization term was used. The best experimental results were reported for 100X with a 96.12% accuracy rate.

Features extraction combined with fine-tuning is one of the approaches in this field. Zhang et al. proposed a classification model by using deep domain knowledge features [61]. The ensemble learning-based domain knowledge was injected into the features extracted from different layers of pre-trained CNN and then from the obtained output, supervised learning generated the final feature representation. And SVM was used to make a classification. The proposed method achieved 81.2% accuracy on the BreakHis dataset. Wang et al. devised a new hybrid architecture combining double-deep transfer learning by using fine-tuned ResNet50 (D^2TL) responsible for feature extraction and interactive cross-task extreme learning machine (ICELM) responsible for classification [?] dataset haste-of-the-art models, with D^2TL_ICELM the best results were obtained with 98.18% accuracy. Kumar et al. proposed a version of VGGNet-16 without FC layer and trained the model along with different classifiers namely, RF, SVM with linear, polynomial and RBF [62]. The best results were recorded with SVM with linear kernel for 200X magnification level with 95.77% accuracy.

Although the features extracted with CNN are stronger than the hand-crafted ones, there are also studies comparing these features in terms of performance. Badejo et al. compared hand-crafted features such as LBP, Local Ternary Pattern (LTP), CLBP, LPQ, Rotation Invariant Co-occurrence of adjacent LBPs (RICLBP) with pre-trained AlexNet features [63]. Thus, in the case of histopathologic image classification, LPQ hand-crafted features proved to be better than AlexNet features. Sharma and Kumar compared hand-crafted features with Xception model extracted features by training five traditional ML algorithms such as logistic regression (LR), Linear LDA, KNN, RF and linear and RBF [64]. For 40X and 100X images Xception with SVM (RBF) achieved the best accuracy rate with 96.25% accuracy.

Different from the above methods, our work focuses on overcoming the curse of dimensionality hence the computational complexity. For instance, when the curse of dimensionality arises, not only it causes irrelevant, redundant and noisy dimensions, but it also decreases the performance of the algorithm and significantly increases the time and memory requirements of algorithms. Moreover, when considered from the point of view of the patient, misclassification is a problem that will have fatal consequences. Therefore, feature selection is one of the most important steps to deal with the curse of dimensionality. In light of all these, we aim to extract features by CNN, and further, select features using meta-heuristic algorithms. Although there are a limited number of studies that make feature selection with metaheuristic algorithms in the BreakHis dataset [65,66], our difference is that feature selection is applied to the features obtained with CNN. The experimental results demonstrate that the approach improves the classification performance effectively.

3. Methods

The proposed method in the study is developed to show the effect of the optimized deep features, which will be obtained by using pre-trained transfer learning networks together with metaheuristic algorithms, on the classification performance. The system architecture used in the study is shown in Fig. 1.

3.1. Dataset

Within the scope of this study, 4 different datasets consisting of images obtained with 40x, 100x, 200x, 400x magnification factors from the open source BreakHis dataset, which includes 9109 breast histopathological images and obtained from 82 patients, are used. In each dataset, there are a total of 1174 microscopic images of $700 \times 460 \times 3$ size, representing 587 benign and malignant tumors for each. Images are in PNG format, with 3-channel RGB with 8-bit depth in each channel. Differentiating benign tumors from malignant tumors is the most critical challenge of cancer histopathology [67]. In order to make this distinction more successful, pathologists use different magnification factors. This process becomes very difficult due to the excessive workload of the pathologists and the analysis through different lenses at the microscope. Therefore, performing these processes automatically will make an important contribution to the field of medicine. In the literature, contributions have been made to automating this process with different approaches to the data set obtained with different magnification factors. The study we presented observed that training with 200x and 400x magnification coefficients gave better results than other magnification factors.

Images representing benign and malignant tumors can be divided into subheadings according to the way they are examined under the microscope. In the BreakHis dataset, the benign tumor class is formed with

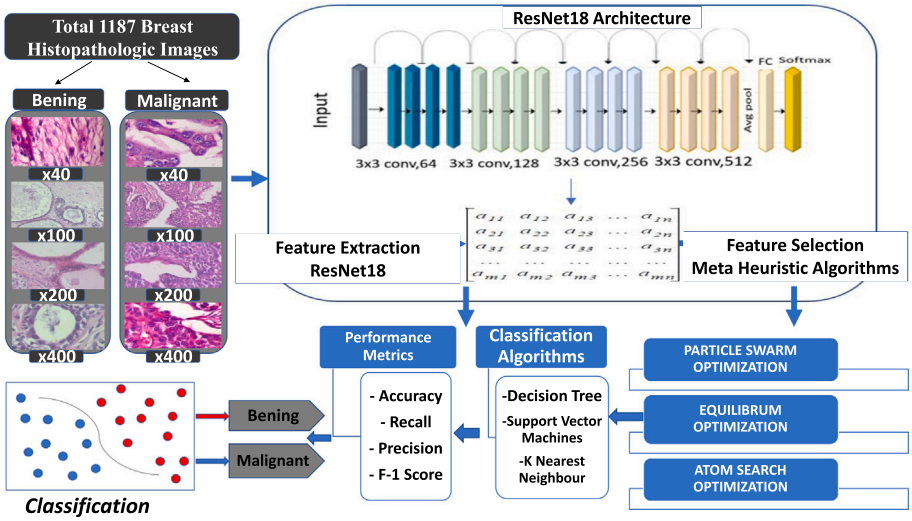


Fig. 1. A more comparable visualization of the parameter optimization effect in the training phase of two different CNN architectures.

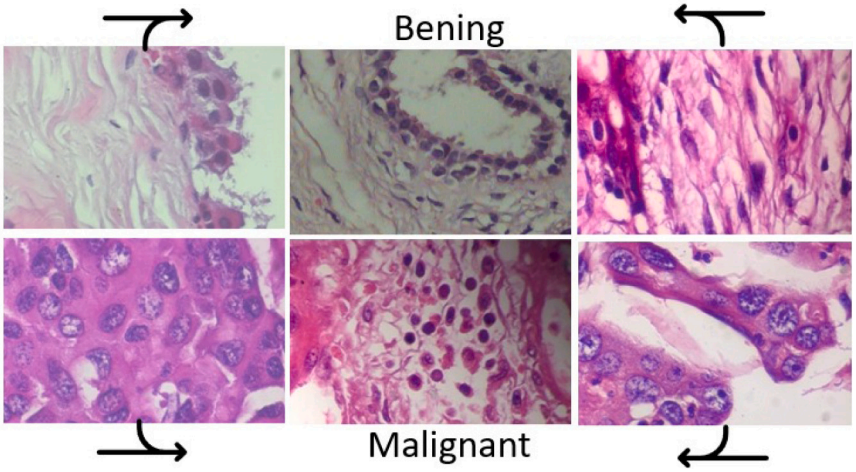


Fig. 2. Samples from dataset.

four different subclasses: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA). Images representing malignant tumor are generated with subclasses of ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). Images of the benign and malignant classes in the data set for 40X magnification factors are shown in Fig. 2.

3.2. Feature extraction with pre-trained network

Analysis with a large number of features in a data set requires a large amount of memory space and computational power in the same direction. Feature extraction aims to convert the raw data in the data set into numerical features that can be processed, thus reducing the amount of data [68]. In this way, the effect of an overfitting problem that a classification algorithm may create on the training data set, and the required computing resource will be reduced. The use of pre-trained transfer learning networks for feature extraction is one of the fastest methods. In the present study, feature extraction is performed from images using the ResNet18 architecture, which is a pre-trained transfer learning network, and a total of 512 features were obtained. In this way, features with higher accuracy are extracted than those obtained from classical feature extraction techniques using image processing. Transfer learning is one of the ML techniques in which the model trained to perform a particular task is reused for other related tasks. With transfer learning, the information obtained from the solution of a problem can be stored and used in solving different problems. It means improving the learning that will take place in a new task by transferring knowledge from a related task that has already been learned. However, given the big data resources required to train large and difficult datasets on which deep learning models are trained, it is highly advantageous that transfer learning can be stored and used in solving different problems. These advantages are because it is an optimization method that provides speed and high performance which are revealed with less training data in solving problems.

3.3. Feature selection with meta-heuristic algorithms

One of the pre-processing techniques to select a subset of the features that will best represent the dataset without information loss is feature selection. When the number of features rises, the search space for the problem will expand exponentially. As a result, the feature selection problem is categorized as NP-hard, and exact algorithms are unable to identify the ideal subset in a timely manner [69]. A model obtained from a dataset with many features is easier to understand thanks to feature selection. An over-fitting problem that may occur in the model is prevented. In this way, it is possible to increase the success rates expected from the model [70]. In order to tackle this issue, meta-heuristic algorithms are frequently used.

3.3.1. Particle swarm optimization

PSO algorithm was developed by Kennedy and Eberhart as a result of their inspiration from animal herd skills such as acclimating to new habitats, seeking food, and avoiding trappers [71–73]. This population-based approach keeps the swarm's information and state up to date. As a result, scanning the whole search space and finding the almost perfect solution becomes conceivable. The swarm's members have their own memories and are aware of the best position to take.

Each swarm member is referred to as a particle, and each particle has information about the swarm's position (p) and velocity (v). p_{t+1} and v_{t+1} are position and velocity at time $t+1$, respectively. In every iteration, this location and velocity information is updated based on the best position (p_{best}) of the particle in the iteration, the best position (g_{best}) in the swarm, and the velocity information available at the time. Speeds of individuals are updated with Eq. (1) at each iteration, and their positions are updated with Eq. (2).

$$v_{t+1} = vw_t + c_1 r_1 (p_{best} - p_t) + c_2 r_2 (g_{best} - g_t). \tag{1}$$

$$p_{t+1} = p_t + v_{t+1}. \tag{2}$$

3.3.2. Atom search optimization

ASO is a physics-based algorithm based on the analysis of molecular dynamics. ASO uses the basic molecular dynamics concept such as the properties of the potential function, interaction force and geometric restraint force and the principle of motion of atoms in order to converge to the global optimum point [74,75].

According to each atom's potential solution in the ASO algorithm, the solution with the best fitness value is obtained by the interaction of atoms among themselves. The motion of atoms is determined according to Eq. (3) and (4).

$$v_i^d(t+1) = rand_i^d v_i^d(t) + a_i^d(t). \tag{3}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1). \tag{4}$$

An atom in population i ($i = 1, 2, N$), d size, v_i^d and x_i^d values give ith velocity and position information. a_i^d represents the acceleration of each atom in the solution set.

The acceleration produced using Newton's law is expressed by Eq. (5).

$$\begin{aligned} a_i^d(t) &= \frac{F_i^d(t)}{m_i^d(t)} + \frac{G_i^d(t)}{m_i^d(t)} \\ a_i^d(t) &= -\alpha(1 - \frac{t-1}{T})^3 e^{-\frac{20t}{T}} x \sum_{j \in K_{best}} \frac{rand_j [2x(h_{ij})^{-13} - (h_{ij})^{-7}]}{m_i(t)} x \\ &\quad \frac{(x_j^d(t) - x_i^d(t))}{\|x_i(t), x_j(t)\|_2} + \beta e^{-\frac{20t}{T}} \frac{x_{best}^d(t) - x_i^d(t)}{m_i(t)} \end{aligned} \tag{5}$$

The force of interaction and geometric constraints between atoms are represented by F_i^d and G_i^d given in Eq. (6) and (7). Using Eq. (8) and (9), the mass of the ith atom is calculated with $m_i(t)$. K_{best} represents the best K atoms. These atoms exchange information with their neighbors. K_{best} is defined in Eq. (10). α is the depth weight and its value, β is the multiplier weight, T represents the maximum number of iterations. $X_{best}(t)$, $Fit_{best}(t)$ and $Fit_{worst}(t)$ represent the best position value, good and worst fitness values during the iteration, respectively.

$$F_i^d(t) = \sum_{j \in K_{best}} rand_j F_{ij}^d(t). \tag{6}$$

$$\begin{aligned} G_i^d(t) &= -\sigma(t)(x_i^d(t) - x_{best}^d(t)), \\ \sigma(t) &= \beta e^{-\frac{20t}{T}}. \end{aligned} \tag{7}$$

$$M_i^d(t+1) = e^{-} \frac{Fit_i(t) - Fit_{best}(t)}{Fit_{worst}(t) - Fit_{best}(t)}. \tag{8}$$

$$m_i(t) = \frac{M_i(t)}{\sum_{j=1}^{N} M_j(t)}. \tag{9}$$

$$K(t) = N - (N-2)x\sqrt{\frac{t}{T}}. \tag{10}$$

3.3.3. Equilibrium optimization

The EO algorithm calculates the solution of the problem in an efficient and effective manner with the exploration and exploitation search phases randomly. The equilibrium optimization method is a physics-based algorithm inspired by the modeling of dynamic mass balance behavior used to investigate the concentration of non-reactive components in control volume yields [76]. In the EO algorithm, the equilibrium state occurs when the particles reach the global optimal value [77]. The steps of the algorithm is given below.

(1) In the EO algorithm, the optimization process is started by formulating the individuals in the search space. During the process, by using Eq. (11) each particle updates its concentration according to the best solutions, which are called equilibrium candidates.

$$C_i^{initial} = C_{min} + rand_i(C_{max} - C_{min}), i = 1, 2, \ldots, n. \tag{11}$$

$C_i^{initial}$ is the initial concentration vector of the ith particle, C_{min} and C_{max} are the minimum and maximum values for the dimensions, n is the number of particles in the population, and rand are random numbers in the range $[0, 1]$. Particles are determined according to the fitness function value. They are then ranked according to their fitness values.

(2) The EO algorithm aims to find the balance of the system in the process. The EO algorithm aims to find the balance of the system in the process. The candidate solutions of the algorithm are chosen such that the four best particles in the process and their concentrations. The fifth equilibrium particle is obtained by taking the arithmetic average of the selected particles. Eq. (12) and (13) are used to create the vectors of the equilibrium pool.

$$C_{avg} = \frac{C_{eq1} + C_{eq2} + C_{eq3} + C_{eq4}}{4}. \tag{12}$$

$$C_{eq,pool} = \{C_{eq1}, C_{eq2}, C_{eq3}, C_{eq4}, C_{eq(avg)}\}. \tag{13}$$

(3) The EO algorithm consists of two stages, exploration and exploitation. In the exploration phase, the four best selected candidate solution values are used. The capacity of the exploitation phase is developed by the average value of the four calculated candidates. The balance between these two stages given in Eq. Eq. (14) is realized with an exponential term F.

$$\begin{aligned} F &= a_1 sign(r - 0.5)[e^{-\sigma t} - 1]. \\ t &= (1 - \frac{iter}{max\_iter})^{(a_2 \frac{iter}{max\_iter})}. \end{aligned} \tag{14}$$

It is a random vector between σ and r $[0, 1]$, the value of t decreases as the number of iterations increases. $iter$ and max_iter represent the current iteration count and the maximum iteration count, respectively, a_2 represents the discovery capability constant. Here, the higher a_2, the stronger the exploration ability, so the lower the exploitability. a_1 value is the opposite of the case of a_2 variable. The larger the value of a_2 the greater the impact on the exploit phase. On the contrary, the effect is less in the discovery phase.

(4) The effect of the production rate (G) has an important place in the exploitation phase of the EO algorithm. The production rate is modeled by Eq. (15).

$$\begin{aligned} G &= G_0 F. \\ G_0 &= GCP(C_{eq} - \sigma C). \\ GCP &= \begin{cases} 0.5 r_1 & r_2 \geq GP \\ 0 & r_2 \leq GP. \end{cases} \end{aligned} \tag{15}$$

The initial value G_0 represents a randomly varying number between r_1 and r_2 $[0, 1]$, GCP represents the control parameter of the production rate and GP represents the production rate used

in the status update. Finally, the general update rule of the states of the particles is modeled by Eq. (16). C and C_{eq} indicate the current particle and the equilibrium candidate solution.

$$C = C_{eq} + (C - C_{eq})F + \frac{G}{\sigma V}(1 - F). \tag{16}$$

3.4. Classification algorithms

3.4.1. Support vector machines

SVM, proposed by Vapnik in 1995, is a supervised ML algorithm to learn over nonlinear functions, on the basis of statistical learning theory and structural risk minimization [78]. SVM uses an appropriate kernel function namely linear, RBF, gaussian, and polynomial to turn the input space into a linearly separable high-dimensional space. The best linear hyperplane of the high-dimensional space is then found using a nonlinear transformation [79]. In this way, non-linear decision boundaries are allowed and the optimal solution is obtained.

3.4.2. K Nearest Neighbor

KNN is one of the conventional supervised classification algorithms that is both simple to use and effective. The main idea of this algorithm is to assign labels to every data in the test set by calculating distance between each unlabeled data and all labeled data in the training set. Assigning a label to the unlabeled data is performed by majority voting over the label of the K nearest neighbors. Different distance functions can be used when calculating distances over neighbors. The most commonly used distance functions can be specified as Minkowski, Euclidean, Manhattan, Chebyshev, Cosine, Jaccard, and Hamming distance [80].

3.4.3. Decision Tree

DT is an important classification method in ML. Its basic logic is to perform the classification process by recursively partitioning the data [81]. This classification process is carried out by means of a tree structure that proceeds through the internal nodes branching from the root node. At the lowest level of the created DT, there are leaf nodes showing the labels. When creating the model, first determine the most important input variable which is the root node, and then divide data at the root node and internal nodes to reach leaf nodes. To select an important node at each level, the DT uses entropy, Gini index, classification error, information gain, gain ratio, and twoing criteria.

4. Experimental results and analysis

4.1. Setup and performance metrics

In the presented study, the size of all the images in the dataset hen reduced to 224 × 224 × 3, which is accepted as input by the ResNet18 network used for feature extraction. Resnet18 architecture is used in many areas for feature extraction and classification. When the literature is examined, many successful performance outputs have been obtained with the ResNet18 architecture, especially in the medical field. Odusami et al. carried out a classification study on functional magnetic resonance images for the diagnosis of Alzheimer's disease [82]. Reported results show that the best performance output is given by fine-tuned Resnet18 architecture with 99.95% accuracy. Al-Falluji et al. In their study, they used X-ray images for the diagnosis of Covid 19 [83]. In the presented study, they reported an accuracy value of 96.73% with the proposed Resnet18-based ESA architecture. This indicates that the ResNet18 architecture will give successful outputs on the histopathological images in the study. For this reason, ResNet18 architecture is preferred as the starting point for the study.

For classification, the BreakHis dataset has been split into a training set and a test set with a ratio of 70% and 30% respectively. Binary classification has been realized on 40X, 100X, 200X, and 400X magnification levels. The models have been implemented in Matlab. The experimental environment is run on Windows 10 Home 64-bit, trained, and tested on an NVIDIA GeForce GTX 1050 GPU platform.

Due to their powerful capacities in higher-level feature extraction, the ResNet18 architecture, which is a pre-trained transfer learning network is used in this study. The deeper layers within the ResNet18 contain higher-level features that are derived from lower-level features which are created by previous layers. In order to get the characteristics of the training and test sets, global pooling layer of the ResNet18 architecture namely the 'pool5' layer is used. In this layer, features are extracted and these extracted features are collected across locations and images into a single pool. As a result, a total of 512 features are obtained for all datasets.

In this context, the performance of the K-nearest neighbor algorithm for k = 3 and k = 5 values was obtained with four different classification results using cosine similarity and euclidean distance. Then, four different classification performances were obtained by applying the SVM algorithm over linear, polynomial, gaussian and RBF kernel functions. Finally, classification performances of all data sets with 512 features obtained as a result of feature extraction were obtained by determining the classification performances through DTs.

In order to measure the performance of the model, precision, recall, accuracy and F1-score are used. Where true positive (TP) is the number of truly classified cancerous images. False positive (FP) shows the number of images predicted as cancerous but not cancerous. False negative (FN) is the incorrect classification of cancerous images as non-cancerous. True negative (TN) is the number of correctly classified non-cancerous images. The formulas of performance metrics in terms of TP, FP, FN and TN are given with Eq. (17), (18), (19), (20).

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{19}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

4.2. Feature selection

By choosing a subset of features from the original feature set, the dimension is lowered and classification accuracy can be improved. Considering this assumption, feature selection is performed with PSO, ASO and EO algorithms within the scope of this study.

A summary of all the parameters used for the meta-heuristic algorithms is shown in Table 1.

Table 2 displays the fitness values including the minimum (best), mean, maximum (worst), standard deviation and number of selected features with the best fitness value for all factors.

Table 1
Summary of all the parameters used in the meta-heuristic algorithms.

Algorithms	Parameters	Values
PSO	Population Size (N)	10
	# of Maximum Iteration	500
	Cognitive Factor (c1)	2
	Social Factor (c2)	2
	Inetria Weight (w)	1
ASO	Population Size (N)	10
	# of Maximum Iteration	500
	alpha	50
	beta	0.2
EO	Population Size (N)	10
	# of Maximum Iteration	500
	Constant (a1)	2
	Constant (a2)	1
	Generation Probability (gp)	0.5

Table 2
Results of meta-heuristic algorithms.

Magnification factor	Performance criteria	PSO	ASO	EO
40X	Minimum	0.018293	0.016253	0.006098
	Maximum	0.073171	0.065249	0.042683
	Standard deviation	0.01665774	0.013061319	0.011063411
	Mean	0.0359757	0.0328646	0.02347565
	# of Selected Features	265	214	131
100X	Minimum	0.006098	0.010216	0.012195
	Maximum	0.04878	0.05259	0.042683
	Standard deviation	0.009336768	0.01232098	0.009911305
	Mean	0.0265244	0.03007735	0.02865855
	# of Selected Features	275	214	173
200X	Minimum	0	0.004141	0
	Maximum	0.036585	0.028834	0.042683
	Standard deviation	0.009585013	0.007158917	0.00948758
	Mean	0.01554895	0.0124639	0.01219525
	# of Selected Features	268	212	173
400X	Minimum	0.006098	0.010509	0.006098
	Maximum	0.054878	0.052785	0.054878
	Standard deviation	0.013620065	0.010183338	0.014524018
	Mean	0.03414625	0.03467505	0.02987805
	# of Selected Features	269	229	173

Table 3
Classification results-feature extraction with ResNet18 architecture.

Magnification factor	Classifier	Recall	Precision	F1 score	Accuracy
40X	KNN K = 3, euclidean	91.05%	90.40%	90.72%	90.34%
	KNN K = 3, cosine	91.00%	90.30%	90.64%	90.34%
	KNN K = 5, euclidean	90.05%	90.40%	90.22%	90.06%
	KNN K = 5, cosine	91.85%	91.50%	91.67%	91.48%
	SVM, linear	90.10%	89.60%	89.84%	89.49%
	SVM, polynomial	93.45%	93.50%	93.47%	93.47%
	SVM, gaussian	94.05%	94.05%	94.05%	94.03%
	SVM, RBF	94.05%	94.05%	94.05%	94.03%
	DT	74.20%	74.15%	74.17%	74.15%
100X	KNN K = 3, euclidean	89.85%	89.50%	89.67%	89.49%
	KNN K = 3, cosine	90.75%	90.35%	90.54%	90.34%
	KNN K = 5, euclidean	89.95%	89.20%	89.57%	89.20%
	KNN K = 5, cosine	90.30%	89.80%	90.04%	89.77%
	SVM, linear	91.35%	91.20%	91.27%	91.19%
	SVM, polynomial	91.25%	91.20%	91.22%	91.19%
	SVM, gaussian	93.45%	93.45%	93.45%	93.47%
	SVM, RBF	93.80%	93.75%	93.77%	93.75%
	DT	76.45%	76.40%	76.42%	76.42%
200X	KNN K = 3, euclidean	90.55%	90.10%	90.32%	90.06%
	KNN K = 3, cosine	92.65%	92.35%	92.49%	92.33%
	KNN K = 5, euclidean	91.50%	91.20%	91.34%	91.19%
	KNN K = 5, cosine	93.20%	93.20%	93.20%	93.18%
	SVM, linear	91.80%	91.45%	91.62%	91.48%
	SVM, polynomial	93.45%	93.50%	93.47%	93.47%
	SVM, gaussian	94.90%	94.90%	94.90%	94.89%
	SVM, RBF	94.90%	94.90%	94.90%	94.89%
	DT	84.95%	84.95%	84.95%	84.94%
400X	KNN K = 3, euclidean	89.15%	88.65%	88.89%	88.64%
	KNN K = 3, cosine	89.00%	88.95%	88.97%	88.92%
	KNN K = 5, euclidean	89.30%	88.95%	89.12%	88.92%
	KNN K = 5, cosine	90.01%	90.05%	90.03%	90.06%
	SVM, linear	92.95%	92.90%	92.92%	92.90%
	SVM, polynomial	93.45%	93.45%	93.45%	93.47%
	SVM, gaussian	94.95%	94.90%	94.92%	94.89%
	SVM, RBF	94.95%	94.90%	94.92%	94.89%
	DT	77.20%	75.00%	76.08%	75.00%

According to Table 2, it can be seen that the meta-heuristic algorithms could effectively select fewer features. PSO algorithm selects 51.76%, ASO algorithm selects 41.8%, EO algorithm selects 25.59% of the initial feature set for 40X. For 100X, PSO algorithm selects 53.71%, ASO algorithm selects 41.8%, EO algorithm selects 33.79% of the initial feature set. For 200X, PSO algorithm selects 52.34%, ASO algorithm selects 41.41%, EO algorithm selects 33.79% of the initial feature set. For 400X, PSO algorithm selects 53.71%, ASO algorithm selects 44.73%, EO algorithm selects 33.79% of the initial feature set. When the algorithms are compared in terms of the number of features selected, it is observed that EO algorithm choose few features, while PSO algorithm choose more features than the others for each magnification factors. It can be seen that EO algorithm outperforms compared with PSO and ASO algorithms for all magnification factors.

Table 4
Classification results-feature selection with PSO algorithm.

Magnification factor	Classifier	Recall	Precision	F1 score	Accuracy
40X	KNN K = 3, euclidean	93.10%	92.90%	92.99%	92.90%
	KNN K = 3, cosine	93.20%	93.20%	93.20%	93.18%
	KNN K = 5, euclidean	91.70%	90.90%	91.29%	90.91%
	KNN K = 5, cosine	91.75%	91.75%	91.75%	91.76%
	SVM, linear	92.40%	92.35%	92.37%	92.33%
	SVM, polynomial	93.45%	93.50%	93.47%	93.47%
	SVM, gaussian	94.65%	94.65%	94.65%	94.60%
	SVM, RBF	94.65%	94.65%	94.65%	94.60%
	DT	82.65%	82.65%	82.65%	82.67%
100X	KNN K = 3, euclidean	90.30%	89.80%	90.04%	89.77%
	KNN K = 3, cosine	93.00%	92.65%	92.82%	92.61%
	KNN K = 5, euclidean	90.15%	89.50%	89.82%	89.49%
	KNN K = 5, cosine	92.15%	91.45%	91.79%	91.48%
	SVM, linear	92.95%	92.90%	92.92%	92.90%
	SVM, polynomial	94.90%	94.90%	94.90%	94.89%
	SVM, gaussian	94.90%	94.90%	94.90%	94.89%
	SVM, RBF	94.90%	94.90%	94.90%	94.89%
	DT	79.85%	79.80%	79.82%	79.83%
200X	KNN K = 3, euclidean	94.05%	93.50%	93.77%	93.47%
	KNN K = 3, cosine	94.20%	94.05%	94.12%	94.03%
	KNN K = 5, euclidean	93.30%	93.20%	93.25%	93.18%
	KNN K = 5, cosine	93.10%	92.65%	92.87%	92.61%
	SVM, linear	90.90%	90.65%	90.77%	90.62%
	SVM, polynomial	91.05%	96.60%	93.74%	96.59%
	SVM, gaussian	97.50%	97.20%	97.34%	97.16%
	SVM, RBF	97.50%	97.20%	97.34%	97.16%
	DT	91.00%	90.65%	90.82%	90.62%
400X	KNN K = 3, euclidean	91.80%	92.45%	92.12%	91.48%
	KNN K = 3, cosine	94.15%	94.10%	94.12%	94.83%
	KNN K = 5, euclidean	90.70%	90.35%	90.52%	90.34%
	KNN K = 5, cosine	92.25%	91.20%	91.72%	91.19%
	SVM, linear	91.80%	91.75%	91.77%	91.76%
	SVM, polynomial	93.10%	92.90%	92.99%	92.90%
	SVM, gaussian	95.75%	95.75%	95.75%	95.74%
	SVM, RBF	95.75%	95.75%	95.75%	95.74%
	DT	79.40%	79.25%	79.32%	79.26%

Smaller mean value says that EO algorithm is much more stable for feature selection problems.

4.3. Results

In this study, we aim to cope with the curse of the dimensionality problem by using meta-heuristic algorithms for feature selection purposes. The experimental results are composed of two parts for all magnification factors of BreakHis dataset. The performance for three traditional ML algorithms with pre-trained ResNet18 extracted features and selected features with PSO, ASO and EO algorithms for binary classification of histopathological images has been evaluated. Here, KNN with two different neighbor values (3 and 5) and two different distance functions (Euclidean and cosine), SVM with four different kernel functions (linear, polynomial, gaussian and RBF) and DT have been used to make the classification.

The classification performance for the ResNet18 extracted features is evaluated for all the classifiers at 40X, 100X, 200X and 400X magnification levels. The performance results at each magnification level are depicted in Table 3. While the best F-scores for 40X, 200X and 400X are achieved through SVM with gaussian and RBF kernels, for 100X the best score is achieved through SVM with RBF kernel. In fact, there is no significant difference between gaussian and RBF kernels for 100X. The best performance of SVM is due to the structural risk minimization principle which reduces the upper bound of expected risk. When the classification results are evaluated in terms of magnification factor, most of the algorithms have achieved the highest performance for 200X. For 200X images on average 91.91% F-score is obtained as the best result. So, we can say that for BreakHis the optimal magnification is 200X.

Different classification metrics are used to evaluate the performance of meta-heuristic algorithms. For, PSO, ASO and EO obtained results are depicted with Table 4, Table 5 and Table 6, respectively.

Regarding classification accuracy, feature selection algorithms provide improved accuracy with fewer features for each magnification factor. For 40X, while PSO algorithm improves accuracy to the extent of 8.5%, ASO algorithm improves a maximum of 4.8% and EO algorithm improves a maximum of 8.8%. Similarly, 3.41%, 2.84% and 3.13% maximum improvements have been achieved with PSO, ASO and EO algorithms, respectively, for 100X. Experiments for 200X show that all three algorithms improve the accuracy to the maximum of 5.7%. 5.9%, 7.1% and 3.69% with PSO, ASO and EO algorithms, respectively for 400X. Accordingly, the EO algorithm provides the best performance improvement for 40X images, increasing the success of DT from 74.15% to 82.95%. Maximum accuracy is also achieved with the EO algorithm. EO algorithm achieves 97.73% maximum accuracy with SVM (both gaussian and RBF) on 200X images.

In terms of precision, for 40X, all three feature selection algorithms improve the performance of the proportion of truly malignant among those predicted to be malignant maximum in the range of 4.47% to 8.88%. When 100X images are examined, it is seen those maximum improvements in precision are 1.65%, 2.85% and 3.15% for PSO, ASO and EO algorithms, respectively. On 200X images, 5.7% improvement has been achieved for all feature selection algorithms. And for 400X images, 5.15%, 7.1% and 3.8% maximum improvements are achieved. Recall shows the ratio of how many of the malignant ones are classified as malignant. While the maximum performance increase in 40X images for PSO, ASO and EO algorithms are 8.48%, 4.47% and 8.75%, increases for 100X images are 3.4%, 2.8% and 3.05% respectively. In 200X images, a 6.05% increase is achieved with both PSO and ASO algorithms, a 5.7% increase is achieved with the EO algorithm. For

Table 5
Classification results-feature selection with ASO algorithm.

Magnification factor	Classifier	Recall	Precision	F1 score	Accuracy
40X	KNN K = 3, euclidean	91.00%	90.65%	90.82%	90.62%
	KNN K = 3, cosine	91.60%	91.45%	91.52%	91.48%
	KNN K = 5, euclidean	91.40%	91.00%	91.19%	90.91%
	KNN K = 5, cosine	93.50%	93.45%	93.47%	93.47%
	SVM, linear	92.35%	92.35%	92.35%	92.33%
	SVM, polynomial	94.15%	93.80%	93.97%	93.75%
	SVM, gaussian	96.20%	96.00%	96.09%	96.02%
	SVM, RBF	96.20%	96.00%	96.09%	96.02%
	DT	78.30%	79.00%	78.64%	78.98%
100X	KNN K = 3, euclidean	90.60%	90.35%	90.47%	90.34%
	KNN K = 3, cosine	91.20%	90.65%	90.92%	90.62%
	KNN K = 5, euclidean	90.65%	90.05%	90.34%	90.06%
	KNN K = 5, cosine	90.85%	90.35%	90.59%	90.34%
	SVM, linear	92.75%	92.65%	92.70%	92.61%
	SVM, polynomial	94.05%	94.05%	94.05%	94.03%
	SVM, gaussian	94.30%	94.35%	94.32%	94.32%
	SVM, RBF	94.30%	94.35%	94.32%	94.32%
	DT	79.20%	78.70%	78.94%	78.69%
200X	KNN K = 3, euclidean	94.00%	93.50%	93.74%	93.47%
	KNN K = 3, cosine	94.20%	94.05%	94.12%	94.03%
	KNN K = 5, euclidean	93.20%	93.20%	93.20%	93.18%
	KNN K = 5, cosine	93.00%	92.65%	92.82%	92.61%
	SVM, linear	91.05%	90.65%	90.84%	90.62%
	SVM, polynomial	97.00%	96.60%	96.67%	96.59%
	SVM, gaussian	97.40%	97.20%	97.29%	97.16%
	SVM, RBF	97.40%	97.20%	97.29%	97.16%
	DT	91.00%	90.65%	90.82%	90.62%
400X	KNN K = 3, euclidean	91.90%	91.50%	91.69%	91.48%
	KNN K = 3, cosine	92.65%	92.60%	92.62%	92.61%
	KNN K = 5, euclidean	91.70%	91.20%	91.44%	91.19%
	KNN K = 5, cosine	92.10%	92.15%	92.12%	92.05%
	SVM, linear	93.45%	93.45%	93.45%	93.47%
	SVM, polynomial	94.30%	94.35%	94.32%	94.32%
	SVM, gaussian	96.30%	96.30%	96.30%	96.31%
	SVM, Rbf	96.30%	96.30%	96.30%	96.31%
	DT	82.15%	82.10%	82.12%	82.10%

400X images, a 5.15% increase is obtained with the PSO algorithm, a 4.95% increase is obtained with the ASO algorithm and a 2.6% increase is obtained with the EO algorithm.

The F-score represents the balance of precision and recall. The maximum increases in 40X images are 8.52%, 4.83% and 8.8% with PSO, ASO and EO algorithms respectively. When 100X images are examined, 3.67%, 2.52% and 1.6% improvements are achieved with PSO, ASO and EO algorithms, respectively. In 200X images, a 5.87% increase is achieved with both PSO and ASO algorithms, 5.72% increase is achieved with the EO algorithm. Finally in 400X images, with PSO algorithm 3.24%, with ASO algorithm 6.04% and with EO algorithm 3.2% the maximum performance increases are achieved.

When all results are evaluated according to Table 7, It has been found that the meta-heuristic feature selection algorithms provide notable performance increases in the classification of histopathological images for almost all magnification factors. Our conclusion here is that meta-heuristic algorithms select a fewer number of features, which best represents the dataset, with the highest classification accuracy for almost all classification algorithms.

With the features extracted with ResNet18, while the lowest classification is performed with DT, it is observed that feature selection increased the performance of DT the most. In addition, it is observed that the best classification success is obtained in 200X images by applying SVM (both gaussian and RBF kernels) to the features selected with the EO algorithm. Here, the reason why the gaussian kernel and the RBF kernel achieve the same or very close performance for most models is the resemblance to the gaussian distribution of RBF.

5. Conclusion

In this work, we propose a fused model of pre-trained transfer learning algorithm and metaheuristic algorithms with ML algorithms, for binary classification of histopathological images. The proposed framework is tested on the BreakHis dataset, using various magnification factors. Several performance metrics have been used to assess the reliability of the model. The model achieved the best accuracy of 97.73% and the best F1-score of 97.75%, for 200X magnification level with EO algorithm and SVM (both gaussian and RBF kernels). The experimental results indicate that the proposed model can be generalized for other binary classification tasks. It could also be considered for tasks that need the use of feature selection. In the future, we intend to apply this architecture to multi-class classification tasks.

CRediT authorship contribution statement

Furkan Atban: Data curation, Software, Writing – original draft, Visualization, Investigation, Validation. **Ekin Ekinci:** Conceptualization, Methodology, Writing – original draft, Investigation, Supervision. **Zeynep Garip:** Software, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have cite the dataset. The dataset is public.

Table 6
Classification results-feature selection with EO algorithm.

Magnification factor	Classifier	Recall	Precision	F1 score	Accuracy
40x	KNN K = 3, euclidean	91.15%	90.90%	91.02%	90.91%
	KNN K = 3, cosine	91.60%	91.50%	91.55%	91.48%
	KNN K = 5, euclidean	92.15%	91.45%	91.79%	91.48%
	KNN K = 5, cosine	92.40%	92.35%	92.37%	92.33%
	SVM, linear	92.65%	92.60%	92.62%	92.61%
	SVM, polynomial	93.80%	93.75%	93.77%	93.75%
	SVM, gaussian	94.30%	94.30%	94.30%	94.32%
	SVM, RBF	94.30%	94.30%	94.30%	94.32%
	DT	82.95%	82.95%	82.95%	82.95%
100X	KNN K = 3, euclidean	90.75%	89.80%	90.27%	89.77%
	KNN K = 3, cosine	92.25%	92.05%	92.14%	92.05%
	KNN K = 5, euclidean	89.60%	88.65%	89.12%	88.64%
	KNN K = 5, cosine	91.60%	91.20%	91.39%	91.19%
	SVM, linear	91.00%	90.90%	90.95%	90.91%
	SVM, polynomial	94.30%	94.35%	94.32%	94.32%
	SVM, gaussian	94.30%	94.30%	94.30%	94.32%
	SVM, RBF	94.30%	94.30%	94.30%	94.32%
	DT	77.10%	77.00%	77.05%	76.99%
200X	KNN K = 3, euclidean	90.70%	90.35%	90.52%	90.34%
	KNN K = 3, cosine	93.80%	93.75%	93.77%	93.75%
	KNN K = 5, euclidean	92.00%	91.50%	91.74%	91.48%
	KNN K = 5, cosine	93.45%	93.20%	93.32%	93.18%
	SVM, linear	93.80%	93.75%	93.77%	93.75%
	SVM, polynomial	97.20%	97.15%	97.17%	97.16%
	SVM, gaussian	97.75%	97.75%	97.75%	97.73%
	SVM, RBF	97.75%	97.75%	97.75%	97.73%
	DT	90.70%	90.65%	90.67%	90.62%
400X	KNN K = 3, euclidean	91.75%	92.00%	91.87%	91.48%
	KNN K = 3, cosine	91.30%	91.20%	91.25%	91.19%
	KNN K = 5, euclidean	91.19%	92.75%	92.32%	91.76%
	KNN K = 5, cosine	91.50%	91.50%	91.50%	91.48%
	SVM, linear	91.15%	90.95%	91.04%	90.91%
	SVM, polynomial	91.80%	91.75%	91.77%	91.76%
	SVM, gaussian	95.45%	95.45%	95.45%	95.45%
	SVM, RBF	95.20%	95.20%	95.20%	95.17%
	DT	78.75%	78.70%	78.72%	78.69%

Table 7
Comparison of conventional method with feature selection algorithms.

Magnification factor	Classifier	F1-Score			
		ResNwt18	ResNet18-PSO	ResNet18-ASO	ResNet18-EO
40x	KNN K = 3, euclidean	90.72%	92.99%	90.82%	91.02%
	KNN K = 3, cosine	90.64%	93.20%	91.52%	91.55%
	KNN K = 5, euclidean	90.22%	91.29%	91.19%	91.79%
	KNN K = 5, cosine	91.67%	91.75%	93.47%	92.37%
	SVM, linear	89.84%	92.37%	92.35%	92.62%
	SVM, polynomial	93.47%	93.47%	93.97%	93.77%
	SVM, gaussian	94.05%	94.65%	96.09%	94.30%
	SVM, RBF	94.05%	94.65%	96.09%	94.30%
	DT	74.17%	82.65%	78.64%	82.95%
100X	KNN K = 3, euclidean	89.67%	90.04%	90.47%	90.27%
	KNN K = 3, cosine	90.54%	92.82%	90.92%	92.14%
	KNN K = 5, euclidean	89.57%	89.82%	90.34%	89.12%
	KNN K = 5, cosine	90.04%	91.79%	90.59%	91.39%
	SVM,linear	91.27%	92.92%	92.70%	90.95%
	SVM, polynomial	91.22%	94.90%	94.05%	94.32%
	SVM, gaussian	93.45%	94.90%	94.32%	94.30%
	SVM, RBF	93.77%	94.90%	94.32%	94.30%
	DT	76.42%	79.82%	78.94%	77.05%
200X	KNN K = 3, euclidean	90.32%	93.77%	93.74%	90.52%
	KNN K = 3, cosine	92.49%	94.12%	94.12%	93.77%
	KNN K = 5, euclidean	91.34%	93.25%	93.20%	91.74%
	KNN K = 5, cosine	93.20%	92.87%	92.82%	93.32%
	SVM, linear	91.62%	90.77%	90.84%	93.77%
	SVM, polynomial	93.47%	93.74%	96.67%	97.17%

(*continued on next page*)

Table 7 (*continued*).

Magnification factor	Classifier	F1-Score			
		ResNwt18	ResNet18-PSO	ResNet18-ASO	ResNet18-EO
	SVM, gaussian	94.90%	97.34%	97.29%	97.75%
	SVM, RBF	94.90%	97.34%	97.29%	97.75%
	DT	84.95%	90.82%	90.82%	90.67%
400X	KNN K = 3, euclidean	88.89%	92.12%	91.69%	91.87%
	KNN K = 3, cosine	88.97%	94.12%	92.62%	91.25%
	KNN K = 5, euclidean	89.12%	90.52%	91.44%	92.32%
	KNN K = 5, cosine	90.03%	91.72%	92.12%	91.50%
	SVM, linear	92.92%	91.77%	93.45%	91.04%
	SVM, polynomial	93.45%	92.99%	94.32%	91.77%
	SVM, gaussian	94.92%	95.75%	96.30%	95.45%
	SVM, RBF	94.92%	95.75%	96.30%	95.20%
	DT	76.08%	79.32%	82.12%	78.72%

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 71 (3) (2021) 209–249.

[2] Y.-D. Zhang, C. Pan, X. Chen, F. Wang, Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling, J. Comput. Sci. 27 (2018) 57–68.

[3] G. Zhang, W. Wang, J. Moon, J.K. Pack, S.I. Jeon, A review of breast tissue classification in mammograms, in: Proceedings of the 2011 ACM Symposium on Research in Applied Computation, 2011, pp. 232–237.

[4] Y. Afriyie, B. A. Weyori, A. A. Opoku, Classification of blood cells using optimized capsule networks, Neural Process. Lett. (2022) 1–20.

[5] R.M. Prakash, R.S.S. Kumari, K. Valarmathi, K. Ramalakshmi, Classification of brain tumours from MR images with an enhanced deep learning approach using densely connected convolutional network, Comput. Methods Biomech. Biomed. Eng.: Imaging Vis. (2022) 1–12.

[6] Y.-D. Zhang, S.C. Satapathy, D.S. Guttery, J.M. Górriz, S.-H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, Inf. Process. Manage. 58 (2) (2021) 102439.

[7] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, A. Mittal, Pneumonia detection using CNN based feature extraction, in: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT, IEEE, 2019, pp. 1–7.

[8] H. Zerouaoui, A. Idri, Deep hybrid architectures for binary classification of medical breast cancer images, Biomed. Signal Process. Control 71 (2022) 103226.

[9] L. Zhang, Y. Hu, C. Wang, Q. Tang, X. Li, Effective dispatching rules mining based on near-optimal schedules in intelligent job shop environment, J. Manuf. Syst. 63 (2022) 424–438.

[10] N. Guo, B. Qian, J. Na, R. Hu, J.-L. Mao, A three-dimensional ant colony optimization algorithm for multi-compartment vehicle routing problem considering carbon emissions, Appl. Soft Comput. 127 (2022) 109326.

[11] G. Kim, I. Moon, Integrated planning for product selection, shelf-space allocation, and replenishment decision with elasticity and positioning effects, J. Retail. Consum. Serv. 58 (2021) 102274.

[12] H. Qin, X. Su, G. Li, X. Jin, M. Yu, A simulation based meta-heuristic approach for the inbound container housekeeping problem in the automated container terminals, Marit. Policy Manag. (2021) 1–23.

[13] S. Barshandeh, R. Dana, P. Eskandarian, A learning automata-based hybrid MPA and JS algorithm for numerical optimization problems and its application on data clustering, Knowl.-Based Syst. 236 (2022) 107682.

[14] A. Kaur, Y. Kumar, Neighborhood search based improved bat algorithm for data clustering, Appl. Intell. (2022) 1–35.

[15] R. Kumar, J. Bhagwan, A comparative study of meta-heuristic-based task scheduling in cloud computing, in: Artificial Intelligence and Sustainable Computing, Springer, 2022, pp. 129–141.

[16] N. Arora, R.K. Banyal, A particle grey wolf hybrid algorithm for workflow scheduling in cloud computing, Wirel. Pers. Commun. 122 (4) (2022) 3313–3345.

[17] C. Li, Y. He, D. Xiao, Z. Luo, J. Fan, P.X. Liu, A novel hybrid approach of ABC with SCA for the parameter optimization of SVR in blind image quality assessment, Neural Comput. Appl. 34 (6) (2022) 4165–4191.

[18] Q. Jin, Y. Zhang, Parameter optimization of active disturbance rejection controller using adaptive differential ant-lion optimizer, Algorithms 15 (1) (2022) 19.

[19] T. Liu, R. Su, C. Sun, X. Li, L. Wei, EOCSA: Predicting prognosis of epithelial ovarian cancer with whole slide histopathological images, Expert Syst. Appl. (2022) 117643.

[20] P. Yang, X. Yin, H. Lu, Z. Hu, X. Zhang, R. Jiang, H. Lv, CS-co: A hybrid self-supervised visual representation learning method for H&E-stained histopathological images, Med. Image Anal. 81 (2022) 102539.

[21] N. Singh, P. Singh, A hybrid ensemble-filter wrapper feature selection approach for medical data classification, Chemometr. Intell. Lab. Syst. 217 (2021) 104396.

[22] T. Li, Z.-H. Zhan, J.-C. Xu, Q. Yang, Y.-Y. Ma, A binary individual search strategy-based bi-objective evolutionary algorithm for high-dimensional feature selection, Inform. Sci. (2022).

[23] A.M. Khalid, H.M. Hamza, S. Mirjalili, K.M. Hosny, BCOVIDOA: A novel binary coronavirus disease optimization algorithm for feature selection, Knowl.-Based Syst. 248 (2022) 108789.

[24] M. Alzaqebah, K. Briki, N. Alrefai, S. Brini, S. Jawarneh, M.K. Alsmadi, R.M.A. Mohammad, I. ALmarashdeh, F.A. Alghamdi, N. Aldhafferi, et al., Memory based cuckoo search algorithm for feature selection of gene expression dataset, Inform. Med. Unlocked 24 (2021) 100572.

[25] S.A.-F. Sayed, E. Nabil, A. Badr, A binary clonal flower pollination algorithm for feature selection, Pattern Recognit. Lett. 77 (2016) 21–27.

[26] J. Too, S. Mirjalili, A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study, Knowl.-Based Syst. 212 (2021) 106553.

[27] S. Chatterjee, S. Biswas, A. Majee, S. Sen, D. Oliva, R. Sarkar, Breast cancer detection from thermal images using a grunwald-letnikov-aided dragonfly algorithm-based deep feature selection method, Comput. Biol. Med. 141 (2022) 105027.

[28] P. Hu, J.-S. Pan, S.-C. Chu, C. Sun, Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection, Appl. Soft Comput. 121 (2022) 108736.

[29] M. Wang, X. Li, L. Chen, An enhance multimodal multiobjective optimization genetic algorithm with special crowding distance for pulmonary hypertension feature selection, Comput. Biol. Med. 146 (2022) 105536.

[30] K. Hanbay, A new standard error based artificial bee colony algorithm and its applications in feature selection, J. King Saud Univ.-Comput. Inf. Sci. 34 (7) (2022) 4554–4567.

[31] M. Paniri, M.B. Dowlatshahi, H. Nezamabadi-pour, Ant-TD: Ant colony optimization plus temporal difference reinforcement learning for multi-label feature selection, Swarm Evol. Comput. 64 (2021) 100892.

[32] K. Deep, et al., A random walk grey wolf optimizer based on dispersion factor for feature selection on chronic disease prediction, Expert Syst. Appl. 206 (2022) 117864.

[33] M.H. Nadimi-Shahraki, H. Zamani, S. Mirjalili, Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study, Comput. Biol. Med. 148 (2022) 105858.

[34] Y. Gao, Y. Zhou, Q. Luo, An efficient binary equilibrium optimizer algorithm for feature selection, IEEE Access 8 (2020) 140936–140963.

[35] R. Chawla, S.M. Beram, C.R. Murthy, T. Thiruvenkadam, N. Bhavani, R. Saravanakumar, P. Sathishkumar, Brain tumor recognition using an integrated bat algorithm with a convolutional neural network approach, Meas.: Sens. (2022) 100426.

[36] J. Piri, P. Mohapatra, An analytical study of modified multi-objective Harris Hawk Optimizer towards medical data feature selection, Comput. Biol. Med. 135 (2021) 104558.

[37] R. Bandyopadhyay, A. Basu, E. Cuevas, R. Sarkar, Harris Hawks optimisation with Simulated Annealing as a deep feature selection method for screening of COVID-19 CT-scans, Appl. Soft Comput. 111 (2021) 107698.

[38] A. Adamu, M. Abdullahi, S.B. Junaidu, I.H. Hassan, An hybrid particle swarm optimization with crow search algorithm for feature selection, Mach. Learn. Appl. 6 (2021) 100108.

[39] R.K. Khadanga, A. Kumar, S. Panda, A modified Grey Wolf Optimization with Cuckoo Search Algorithm for load frequency controller design of hybrid power system, Appl. Soft Comput. (2022) 109011.

[40] X. Dong, M. Li, P. Zhou, X. Deng, S. Li, X. Zhao, Y. Wu, J. Qin, W. Guo, Fusing pre-trained convolutional neural networks features for multi-differentiated subtypes of liver cancer on histopathological images, BMC Med. Inform. Decis. Mak. 22 (1) (2022) 1–27.

[41] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (7) (2015) 1455–1462.

[42] A.A. Samah, M.F.A. Fauzi, S. Mansor, Classification of benign and malignant tumors in histopathology images, in: 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA, IEEE, 2017, pp. 102–106.
[43] M.A. Kahya, W. Al-Hayani, Z.Y. Algamal, Classification of breast cancer histopathology images based on adaptive sparse support vector machine, J. Appl. Math. Bioinf. 7 (1) (2017) 49.
[44] S. Chattoraj, K. Vishwakarma, Classification of histopathological breast cancer images using iterative VMD aided Zernike moments & textural signatures, 2018, arXiv preprint arXiv:1801.04880.
[45] D. Sanchez-Morillo, J. González, M. García-Rojo, J. Ortega, Classification of breast cancer histopathological images using KAZE features, in: International Conference on Bioinformatics and Biomedical Engineering, Springer, 2018, pp. 276–286.
[46] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: 2016 International Joint Conference on Neural Networks, IJCNN, IEEE, 2016, pp. 2560–2567.
[47] A.-A. Nahid, A. Mikaelian, Y. Kong, Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation, Biomed. Res. 29 (10) (2018) 2068–2077.
[48] S. Cascianelli, R. Bello-Cerezo, F. Bianconi, M.L. Fravolini, M. Belal, B. Palumbo, J.N. Kather, Dimensionality reduction strategies for cnn-based classification of histopathological images, in: International Conference on Intelligent Interactive Multimedia Systems and Services, Springer, 2018, pp. 21–30.
[49] F.A. Spanhol, L.S. Oliveira, P.R. Cavalin, C. Petitjean, L. Heutte, Deep features for breast cancer histopathological image classification, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE, 2017, pp. 1868–1873.
[50] V. Gupta, A. Bhavsar, Sequential modeling of deep features for breast cancer histopathological image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2254–2261.
[51] S. Boumaraf, X. Liu, Z. Zheng, X. Ma, C. Ferkous, A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images, Biomed. Signal Process. Control 63 (2021) 102192.
[52] P. Wang, J. Wang, Y. Li, P. Li, L. Li, M. Jiang, Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing, Biomed. Signal Process. Control 65 (2021) 102341.
[53] Y. Benhammou, S. Tabik, B. Achchab, F. Herrera, A first study exploring the performance of the state-of-the art cnn model in the problem of breast cancer, in: Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications, 2018, pp. 1–6.
[54] P. Wang, P. Li, Y. Li, J. Wang, J. Xu, Histopathological image classification based on cross-domain deep transferred feature fusion, Biomed. Signal Process. Control 68 (2021) 102705.
[55] Y. Zou, J. Zhang, S. Huang, B. Liu, Breast cancer histopathological image classification using attention high-order deep network, Int. J. Imaging Syst. Technol. 32 (1) (2022) 266–279.
[56] C. Hu, X. Sun, Z. Yuan, Y. Wu, Classification of breast cancer histopathological image with deep residual learning, Int. J. Imaging Syst. Technol. 31 (3) (2021) 1583–1594.
[57] A.M. Ibraheem, K.H. Rahouma, H.F. Hamed, 3PCNNB-net: Three parallel CNN branches for breast cancer classification through histopathological images, J. Med. Biol. Eng. 41 (4) (2021) 494–503.
[58] H. Barzekar, Z. Yu, C-net: A reliable convolutional neural network for biomedical image classification, Expert Syst. Appl. 187 (2022) 116003.
[59] W. Zhi, H.W.F. Yueng, Z. Chen, S.M. Zandavi, Z. Lu, Y.Y. Chung, Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images, in: International Conference on Neural Information Processing, Springer, 2017, pp. 669–676.
[60] P. Wang, P. Li, Y. Li, J. Xu, F. Yan, M. Jiang, Deep manifold feature fusion for classification of breast histopathology images, Digit. Signal Process. (2022) 103400.
[61] G. Zhang, M. Xiao, Y.-h. Huang, Histopathological image recognition with domain knowledge based deep features, in: International Conference on Intelligent Computing, Springer, 2018, pp. 349–359.
[62] A. Kumar, S.K. Singh, S. Saxena, K. Lakshmanan, A.K. Sangaiah, H. Chauhan, S. Shrivastava, R.K. Singh, Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer, Inform. Sci. 508 (2020) 405–421.
[63] J.A. Badejo, E. Adetiba, A. Akinrinmade, M.B. Akanle, Medical image classification with hand-designed or machine-designed texture descriptors: a performance evaluation, in: International Conference on Bioinformatics and Biomedical Engineering, Springer, 2018, pp. 266–275.
[64] S. Sharma, S. Kumar, The Xception model: A potential feature extractor in breast cancer histology images classification, ICT Express 8 (1) (2022) 101–108.
[65] M. Saraswat, R. Pal, R. Singh, H. Mittal, A. Pandey, J. Chand Bansal, An optimal feature selection approach using IBBO for histopathological image classification, in: Congress on Intelligent Systems, Springer, 2020, pp. 31–40.
[66] S. Vijh, S. Kumar, M. Saraswat, Efficient feature selection method for histopathological images using modified golden eagle optimization algorithm, in: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO, IEEE, 2021, pp. 1–5.
[67] A. Ashtaiwi, Optimal histopathological magnification factors for deep learning-based breast cancer prediction, Appl. Syst. Innov. 5 (5) (2022) 87.
[68] C. Iwendi, K. Mahboob, Z. Khalid, A.R. Javed, M. Rizwan, U. Ghosh, Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system, Multimedia Syst. 28 (4) (2022) 1223–1237.
[69] J. Zhou, Z. Hua, A correlation guided genetic algorithm and its application to feature selection, Appl. Soft Comput. 123 (2022) 108964.
[70] B. Zhang, Y. Li, Z. Chai, A novel random multi-subspace based ReliefF for feature selection, Knowl.-Based Syst. 252 (2022) 109400.
[71] R. Poli, J. Kennedy, T. Blackwell, Particle swarm optimization, Swarm Intell. 1 (1) (2007) 33–57.
[72] Z. Garip, M.E. Çimen, A.F. Boz, D. Karayel, Firefly algorithm and particle swarm optimization for optimal IIR system identification, in: 2018 6th International Conference on Control Engineering & Information Technology, CEIT, IEEE, 2018, pp. 1–6.
[73] Z. Garip, D. Karayel, M. Erhan Çimen, A study on path planning optimization of mobile robots based on hybrid algorithm, Concurr. Comput.: Pract. Exper. 34 (5) (2022) e6721.
[74] W. Zhao, L. Wang, Z. Zhang, Atom search optimization and its application to solve a hydrogeologic parameter estimation problem, Knowl.-Based Syst. 163 (2019) 283–304.
[75] M.A. Mossa, O.M. Kamel, H.M. Sultan, A.A.Z. Diab, Parameter estimation of PEMFC model based on Harris Hawks' optimization and atom search optimization algorithms, Neural Comput. Appl. 33 (11) (2021) 5555–5570.
[76] A. Faramarzi, M. Heidarinejad, B. Stephens, S. Mirjalili, Equilibrium optimizer: A novel optimization algorithm, Knowl.-Based Syst. 191 (2020) 105190.
[77] Y. Sun, J.-S. Pan, P. Hu, S.-C. Chu, Enhanced Equilibrium Optimizer algorithm applied in job shop scheduling problem, J. Intell. Manuf. (2022) 1–27.
[78] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[79] H.Y. git, H. Köylü, S. Eken, Estimation of road surface type from brake pressure pulses of ABS, Expert Syst. Appl. (ISSN: 0957-4174) (2022) 118726, http://dx.doi.org/10.1016/j.eswa.2022.118726, URL https://www.sciencedirect.com/science/article/pii/S095741742201747X.
[80] S. Nayak, M. Bhat, N.S. Reddy, B.A. Rao, Study of distance metrics on k-nearest neighbor algorithm for star categorization, in: Journal of Physics: Conference Series, 2161, (1) IOP Publishing, 2022, 012004.
[81] A. Kim, H. Kim, A new classification tree method with interaction detection capability, Comput. Statist. Data Anal. 165 (2022) 107324.
[82] M. Odusami, R. Maskeliūnas, R. Damaševičius, T. Krilavičius, Analysis of features of alzheimer's disease: detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network, Diagnostics 11 (6) (2021) 1071.
[83] R.A. Al-Falluji, Z.D. Katheeth, B. Alathari, Automatic detection of COVID-19 using chest X-ray images and modified ResNet18-based convolution neural networks, Comput. Mater. Continua (2021) 1301–1313.