Ozaltin, O., Coskun, O., Yeniay, O., & Subasi, A. (2022). Classification of brain hemorrhage computed tomography images using OzNet hybrid algorithm. *International Journal of Imaging Systems and Technology*.

Classification of Brain Hemorrhage Computed Tomography Images using OzNet Hybrid Algorithm

Oznur Ozaltin[1], Orhan Coskun[2], Ozgur Yeniay[1], Abdulhamit Subasi[3,4]

[1] Hacettepe University, Institute of Science, Department of Statistics, Ankara,Turkey
E-mail: oznurozaltin@hacettepe.edu.tr
E-mail: yeniay@hacettepe.edu.tr
[2]Health Sciences University Gaziosmanpaşa Training and Research Hospital, Pediatric Neurology
E-mail: dr.orhancoskun@hotmail.com
[3]Institute of Biomedicine, Faculty of Medicine, University of Turku, 20520, Turku, Finland.
E-mail: abdulhamit.subasi@utu.fi
[4]Department of Computer Science, College of Engineering, Effat University, Jeddah, 21478, Saudi Arabia.
E-mail: absubasi@effatuniversity.edu.sa

Abstract

Classification of brain hemorrhage Computed Tomography (CT) images provides a better diagnostic implementation for emergency patients. Attentively, each brain CT image must be examined by doctors. This situation is time-consuming, exhausting, and sometimes leads to making errors. Hence, we aim to find the best algorithm owing to a requirement for automatic classification of CT images to detect brain hemorrhage. In this study, we developed OzNet hybrid algorithm, which is a novel convolution neural networks (CNN) algorithm. Although OzNet achieves high classification performance, we combine it with Neighborhood Component Analysis (NCA) and many classifiers: Artificial Neural Networks (ANN), Adaboost, Bagging, Decision Tree, K-Nearest Neighbor (K-NN), Linear Discriminant Analysis (LDA), Naïve Bayes and Support Vector Machines (SVM). Additionally, Oznet is utilized for feature extraction, where 4096 features are extracted from the fully connected layer. These features are reduced to have significant and informative features with minimum loss by NCA. Eventually, we use these classifiers to classify these significant features. Finally, experimental results display that OzNet-NCA-ANN excellent classifier model and achieves 100% accuracy with created Dataset 2 from Brain Hemorrhage CT images.

Keywords: Classification, CNN, Feature extraction, Machine learning, OZNET, NCA.

1. Introduction

Intracerebral hemorrhage, or ICH, is a devastating disease. The overall incidence of spontaneous ICH worldwide is 24.6 per 100,000 person-years with approximately 40,000 to 67,000 cases per year in the United States [1]. Approximately half of this mortality occurs within the first 24 hours. Therefore, early diagnosis and treatment are very important [2]. The gold standard diagnostic approach is to initially perform a non-contrast Computed Tomography (CT) scan followed by a lumbar puncture (LP) and analysis of the cerebrospinal fluid (CSF) when the CT is negative [3]. Tomography findings are also important in the differentiation of chronic bleeding and acute bleeding. In the hyperacute phase, it can be difficult to delineate the hemorrhage. In the acute and early subacute phase, progressive blood clot retraction with concomitant extrusion of the serum component will rapidly increase the density of the

hemorrhage. In particular, the increasing concentration of the protein component of hemoglobin (globin) will appear hyperdense on CT. During the subsequent late subacute phase, progressive Red Blood Cell (RBC) lysis and proteolysis of the globin protein result in decreasing density of the hematoma. In the late subacute phase, the hematoma may become isodense to the brain parenchyma [4]. Therefore, we can express that CT absolutely helps doctors diagnose these brain problems [5, 6]. A very quick assessment is vital for patients. Additionally, it is also crucial for determining the region of the brain for surgery. Thus, the researchers aim to fast diagnoses from magnetic resonance imaging (MRI) or CT using artificial intelligence. Currently, deep learning algorithms are used in many studies in many fields [7-13]. Deep Learning algorithms eliminate feature extraction by handcrafting [10]. Therefore, these algorithms entirely automate feature extraction and classification.

In literature, many researchers study Brain Hemorrhage CT images with different applications. Shahangian and Pourghassem [14] have proposed a classification algorithm based on an automatic segmentation method that is modified Distance Regularized Level Set Evolution (DRLSE). They have utilized some machine learning algorithms: K-NN, MLP, SVM, and hierarchical classification. When they have classified three different types of hemorrhage classes, Epidural hemorrhage (EDH), Subdural hemorrhage (SDH), and Intracerebral hemorrhage (ICH), their study has been achieved an accuracy of 94.13 % with the SVM algorithm.

Kuo et al. [6] have examined 4396 head CT scans, which have been supplied from the University of California in San Francisco and affiliated hospitals, by using a fully connected convolutional network. Additionally, they have compared the performance of their algorithm with experts. As a result, the fully connected network performance has been exceeded that of experts. Moreover, their finding has been exhibited with an AUC of 0.991 for CT scans classification. Gautam et al. [15] have suggested the Robust Fuzzy C-Means algorithm and DRLSE for determining brain lesion regions from CT scan images. They have compared it standard Fuzzy C-Means (FCM) clustering, spatial FCM, robust kernel-based fuzzy clustering (RFCM), and DRLSE for accurate segment brain hemorrhagic lesion. Bhadauria and Dewal [16] have combined two popular segmentation methods: Fuzzy Clustering and Region-Based Active Contour Method to automatically detect ICH from Brain Hemorrhage CT images. Further, they have compared their proposed method with the Fuzzy Clustering and Region Growing method. Finally, they have achieved a sensitivity of 79.48% and a specificity of 99.42% via their proposed method.

Maas et al. [17] have investigated six different traumatic brain injury (TBI) morphological abnormalities on CT scans by using Marshall CT classification. When they improved it via logistic regression analysis, they achieved a 0.77 AUC result. Shahangian and Pourghassem [18] have automatically detected the hemorrhage regions from Brain CT images via Genetic Algorithm (GA). Besides, they have classified EDH, ICH, and SDH whose are hemorrhage classes. Eventually, they have obtained an accuracy of 93.33% with the multi-layer neural network. Solorio-Ramírez et al. [19] have proposed a new algorithm, they called as Minimalistic Machine Learning (MML) algorithm to classify Brain Hemorrhage CT images. They have used two types of classes: Normal (without hemorrhage) and Intraventricular (IVH) hemorrhage. Additionally, they have compared it with K-NN, MLP, Naive Bayes, SVM, Adaboost, and Hierarchical algorithms. Lastly, they have achieved an accuracy of 92.62% with

the Hierarchical classifier. Although it is provided supremacy to their proposed method, when they compared it in terms of time, the highest performance is theirs. Balasooriya [20] has used the watershed method to detect hemorrhage lesions from Brain CT images. This study has been aimed to develop surgeries and surgical simulations with artificial neural networks by using MATLAB. After all, this study has been expressed that it is taken an acceptable result. Alawad et al. [21] have developed a practical method, called Accurate Identification of Brain Hemorrhage (AIBH). Firstly, they have removed the skull region from Brain CT images via out the region of interest (ROI) using Otsu's method. Next, they have extracted features from images according to ROI distance. Moreover, they have utilized GA to choose related features. Then, they have applied stacking-based machine learning algorithms to estimate diverse types of hemorrhage. Eventually, they have obtained an accuracy of 99.5% by using 10-fold cross-validation with the proposed AIBH method.

Gautam and Raman [5] have proposed CNN architectures based and image fusion method for Brain CT images. In total, the dataset has included three classes which are hemorrhagic, ischemic, and normal. Besides, they have created another dataset from the general dataset. This dataset, namely dataset 1, has contained just two classes: hemorrhagic, ischemic. In fact, the general dataset has been named dataset 2. When they have divided into dataset 1 80% training set, and 20% testing set, they have achieved an accuracy of 98.33%. In addition, when they have applied 10-fold cross-validation on brain CT images, they have obtained an accuracy of 98.77%. When they have applied similar applications on dataset 2, they obtained an accuracy of 92.22% and 93.33%, respectively. Mansour and Aljehane [22] have proposed new deep learning (DL) model to diagnose ICH from brain CT images. Their study has included four-stage: preprocessing, segmentation, feature extraction, and classification. In the first stage, they have converted the format of the image to JPEG. In the second stage, they have used Kapur's thresholding with an elephant herd optimization (EHO) algorithm for image segmentation. Then, they have applied the Inceptionv4 network as a feature extractor for the feature extraction stage. Finally, they have classified with multi-layer perception networks obtained an accuracy of 95.06%.

Toğaçar et al. [23] have classified the brain hemorrhage CT images using AlexNet and they have aimed to evolve classification success by utilizing the autoencoder network model and heat maps of every image in the dataset. Additionally, they have also applied data augmentation for this goal. Consequently, they have achieved an accuracy of 98.57%. Anupama et al. [24] have utilized a deep learning algorithm for the classification of brain hemorrhage CT images. Before the classification phase, they have applied Gabor filtering for removing the noises. Moreover, they have developed GrabCut-based segmentation with synergic deep learning (SDL), named GC-SDL to identify disease regions. Hence, they have acquired an accuracy of 95.73%. Chawla and Kishore [25] have utilized contra-lateral symmetry to detect all types of strokes: acute, chronic infarcts, and hemorrhages from 347 brain CT slices of 15 patients. In addition, they have performed a two-level classification phase to detect abnormalities using features derived from the intensity and the wavelet domain. When they have applied their proposed method, they have achieved an accuracy of 90% for classification. Majumdar et al. [26] have applied a U-Net algorithm to automatically detect hemorrhage lesions from Brain CT images. In addition, they have utilized five types of hemorrhage classes from 134 cases whose are included 4300 CT images. Besides, they have performed data augmentation technique with random left-right flipping and ±10 degrees random rotation to

improve model performance. As a result, they achieved a sensitivity of 81% and a specificity of 98%.

Study	Database	State of art methods	Metrics
Uyar et al.[33]	2700 Brain CT images 3 classes	Faster R-CNN	Accuracy 0.9775
Ertuğrul and Akıl[34]	15979 Brain CT images 27203 bounding boxes 6 classes	YOLOv4*	Precision 0.94
Abramova et al.[35]	256 Brain CT images 2 classes	3D-UNet	Dice Accuracy 0.862
Santhoshkumar et al. [36]	2526 Brain CT images 6 classes	DenseNet-Extreme Learning	Accuracy 0.9634
Barin et al.[37]	752803 Brain CT images 6 classes	InceptionResNetv2-EfficientNetB3	Accuracy 0.9859
Burduja et al.[38]	25272 Brain CT images 6 classes	3D ResNeXt-BiLSTM	Accuracy 0.947 Sensitivity 0.756 Precision 0.972
Salehinejad et al.[39]	31277 Brain CT images 6 classes	ResNeXt-50-LightGBM CatBoost XGBoost	AUC 0.984 Sensitivity 0.988 Specificity 0.980
Pandimurugan et al. [40]	2450 Brain CT images 5 classes	CNN-GAN**	Accuracy 0.995
Iqbal et al.[41]	2501 Brain CT images 6 classes	VGG16-MLP	Accuracy 0.9724

Many other studies are also investigated Brain CT images with different approaches [27-32]. Table 1 shows the recent studies on Brain CT datasets.

Table 1. Results of performance on Brain CT images in recent studies

* YOLOv4: You Look Only Once ** GAN: Generative Adversarial Network

In this study, we targeted to determine the different types of brain hemorrhage from CT images, automatically. For this purpose, we designed a framework, which achieves the best classification performance. The contribution of this study is as follows.

We developed a novel convolution neural networks (CNN) algorithm, named as OzNet. This architecture is different from other pre-trained architectures in terms of layers, filter sizes, and the number of filters. In fact, we designed this architecture to classify biomedical datasets

with better accuracy. When we classified brain hemorrhage CT images, the performance of the architecture was excellent. However, in order to achieve better classification performance, we combine OzNet with Neighbor Component Analysis (NCA) and different classifiers: ANN, Adaboost, Bagging, Decision Tree, K-NN, LDA, Naïve Bayes, and SVM. This new framework is called as OzNet hybrid algorithms. Essentially, OzNet is utilized for feature extraction from brain hemorrhage CT images in this framework. While we have extracted 4096 features from the fully connected layer of OzNet, these features are reduced by using the NCA method. NCA is a non-parametric method and has no assumptions and is more practical than other feature selection methods. Thus, informative features are selected and utilized for classification. Also, in the classification phase, we employed several widely used classifiers. Another contribution of this study, we do not only perform classification and feature selection in just one dataset but also apply the same process in the datasets created according to age groups. Thus, the performance of these hybrid algorithms and OzNet are assessed on different datasets. Fig.1 demonstrates different types of brain hemorrhage CT images.

The novelties and significant contributions of this study are as follows:

- Brain Hemorrhage CT images are split according to age groups which are children, adults, and children+adults. Thus, all results are reported based on these datasets.
- A novel deep learning model called OzNet is developed in this study and trained using 10-fold cross-validation on created three different datasets.
- OzNet is compared with ResNet-18, MobileNetv2, ShuffleNet, and other CNNs models.
- Hybrid machine learning algorithms are designed to improve the accuracy of the proposed model.
- NCA, which is a feature selection method is applied to achieve the best performance in the detection of a brain hemorrhage, then built new hybrid algorithms.
- Though the best algorithm is determined, other non-parametric methods (minimum Redundancy Maximum Relevance, and Chi-Square) are utilized to be sure which one is the best.

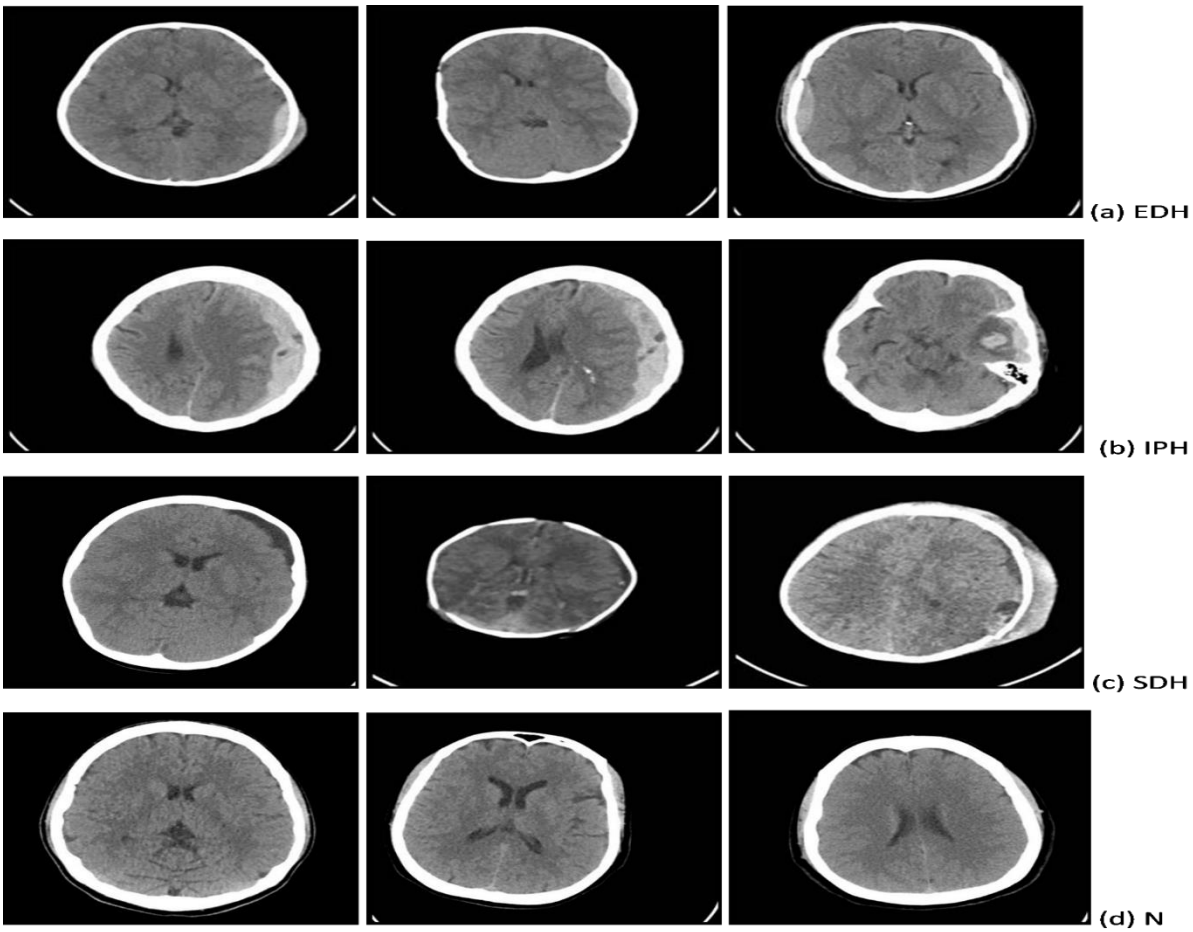


Fig.1 Type of Brain Hemorrhage CT images used in the study

The rest of the paper is organized as follows: In Section 2, we briefly present the utilized methods and datasets in the study. Next, in section 3, we explain the performance metrics result for each experiment and discuss the results, in section 4. In last, we conclude the study, in section 5.

2. Materials and Methods

2.1 Data set:

The dataset was gathered from Al Hilla Teaching Hospital Iraq in 2018. It was confirmed by the research and ethics board in the Iraqi ministry of health Babil Office. The dataset, which is shared by Hssayeni is publicly available online at Pysiyonet.org [42-44] and Kaggle [45]. The dataset is included both brain and skull CT images in 650x650 dimension jpg format. Totally, the dataset has consisted of 82 patients. When we investigate it in terms of patients' demographics, the age range is between 1 day and 72 years old. There are 46 male and 36 female patients. Besides, five different types of brain hemorrhage are included IVH, IPH, SAH, EDH, and SDH. Hssayeni et al. [44, 46] performed a U-Net model trained on 160x160

crops via 5-fold cross-validation. They also tested the model with a Jaccard index of 0.218 and a Dice coefficient of 0.315 on average.

In this study, we check each segmented region in the brain CT dataset and we can say that these are mostly correct. Additionally, we utilize these automatically segmented images to classify with developed OzNet architecture. Fig.2 shows the segmented brain CT image of Intraparenchymal Hemorrhage. In this study, we create three datasets which are included just brain CT images. Actually, we would like to examine whether the classification performance differs according to age distribution (children and adults). Besides this consideration, the datasets classes are not the same for datasets of children and adults. Thus, we also investigate all datasets regarding Oznet performance. First, we have created brain CT images for 18 ages and under, called Dataset 1. This dataset contains 3 classes of 300 images Epidural Hemorrhage (EDH), 300 images Subdural (SDH)+Intraparenchymal (IPH) Hemorrhage, and 250 images Normal (N). The reasons why we take Subdural and Intraparenchymal in the same class: children patients diagnosed with Subdural are also diagnosed as Intraparenchymal and the number of children patients diagnosed with Intraparenchymal is not enough in the whole dataset. Next, we have created adults' brain CT images, called Dataset 2. This dataset also included 3 classes of 265 images Epidural Hemorrhage (EDH), 150 images Intraparenchymal (IPH)+Subdural (SDH) Hemorrhage, and 330 images Normal (N). Contrary to the situation in children, adult patients diagnosed with Intraparenchymal are generally diagnosed as Subdural and the number of adult patients diagnosed with subdural is not enough in the whole dataset. In last, we have combined both children and adult datasets owing to see the Intraparenchymal (IPH) and Subdural (SDH) classes separately, called Dataset 3. Although the dataset includes five different types of hemorrhage: IVH, IPH, SAH, EDH, and SDH, we must eliminate IVH and SAH regarding not having enough images for classification. Hence, we have also measured OzNet's performance with four classes. Fig.3 demonstrates the flowchart of the study. Additionally, Algorithm 1 gives the pseudocode of the proposed hybrid algorithm.

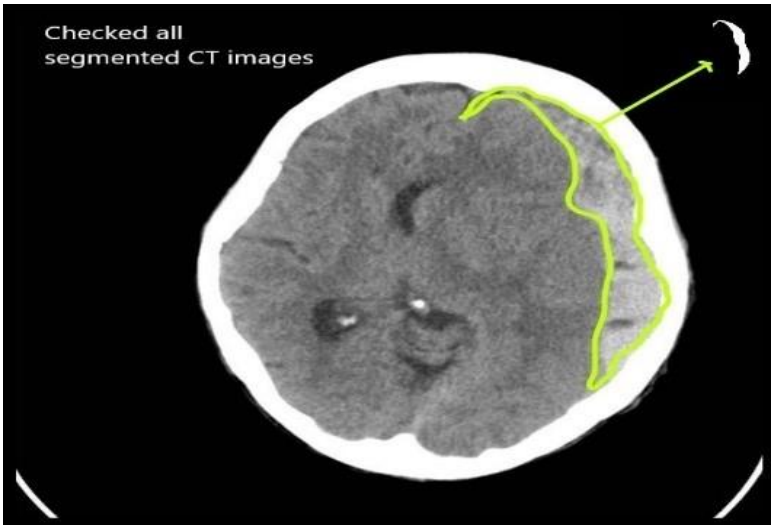


Fig.2. IPH class of the segmented CT used in the study.

Ozaltin, O., Coskun, O., Yeniay, O., & Subasi, A. (2022). Classification of brain hemorrhage computed tomography images using OzNet hybrid algorithm. *International Journal of Imaging Systems and Technology*.

Fig.3 Flowchart of this study

Algorithm 1: Pseudocode of the proposed hybrid algorithm

Input: Brain Hemorrhage CT images with size of 650x650

Result: ***pl***=predict labels

% Pre-processing phase

% Create three different datasets (Dataset1, Dataset2, Dataset3)

1. Load CT images ;
2. *I*=number of images;
3. **for** I=1: *I*
   imds= resized each image from 650x650 to 227x227 .jpg
4. **end**

% Classification and automatic feature extraction phase using OzNet

1. Load imds
2. Num_imgs=images' number of imds
3. K_folds=10
4. **for** id=1:num_imgs
   Create validation set;
   Create training set;
   Input size=[227 227];
   Number of classes;
   layers= $[Input \rightarrow Conv1 \rightarrow MaxPool1 \rightarrow Conv2 \rightarrow MaxPool2 \rightarrow \cdots \rightarrow FC-8 \rightarrow Drop \rightarrow FC-9 \rightarrow SoftMax \rightarrow Output]$*;%detailed in Table3*
   options=[sgdm, MaxEpochs=20, MiniBatchSize=8, Learning rate=0.0001];
   net=train(training set, layers, options);
   pl=classify(net, validation set);
   end
5. obtain performance metrics;
6. save net;

% Improve Classification reults

7. split imds 70% for training 30% for testing;
8. featurelayer=FC-8; %4096 features are obtained;
   %activations feature layers
9. Train_features=(net,training, featurelayer, *MiniBatchSize=8);*
10. Test_features=(net,testing, featurelayer, *MiniBatchSize=8);*

% Feature Selection using NCA method

11. nca=(train_features, fitmethod, exact, sgd, lambda=0.005,iteration=30, gradienttolerance=0.0001);
12. Draw plot graph; %for determining tolerance value
13. t=tolerance value
14. obtain featureweigth from nca
15. id=find(featureweight>t*max(1,max(featureweigth)));
    %Dimension reduction of features with NCA
16. trainfeatures=Train_features(id,:);
17. testfeatures=Test_features(id,:);
    %Classify by using ML algorithms
18. Pl=classify(ML algortihms);
19. obtain performance metrics;

2.2 OzNet: New CNN Architecture

With the increasing needs of the age, artificial intelligence is also advancing rapidly. Deep learning algorithms are one of them. Many researchers benefit from deep learning algorithms on different topics. Here, we present CNN architecture, which is one of the deep learning algorithms, named as OzNet. The reason why we named OzNet is that the first two letters of the names of the creators is Oz, like LeNet [47]. We aim to provide an architecture that is not overfitting and gives robust test results on biomedical data.

OzNet is a new CNN architecture which is consisted of 34 layers. There are 7 convolutional layers, 7 maximum pooling layers, 2 fully connected layers, a dropout layer, a SoftMax layer, and a classification layer. Further, each convolutional layer is merged with the batch normalization layer and activation layer. In this study, we prefer ReLU (Rectified Linear Unit) activation function owing to its faster than other activation functions. In fact, although OzNet's architectural structure seems similar to VGG, but it is different, since we used different convolutional layers, number of filters, filter sizes, and fully connected layers in OZNet. While a minimum VGG (VGG-11) has 11-convolutional layers, OzNet has 7-convolutional layers. In VGG, the filter size is taken as 3x3 for each layer but, OzNet uses not only 3x3 but also utilizes different filter sizes, as detailed in Table 3. Additionally, the first convolutional layer has a filter size of 64, then the filter sizes in the next layers are doubled till reached 512 in VGG [48, 49]. However, in OzNet, the filter size in the first layer 64, then the filter size in the next two layers is 128. Then, it reaches 256. The following three layers has a filter size of 128 again. Besides, fully connected layers are also different, while OzNet includes two fully connected layers with a dropout layer, VGG has three fully connected layers with different filter sizes. Hence, OzNet has a different structure from others.

Here, we utilize OzNet as a classification algorithm with a cross-entropy approach. In order to enhance the classification performance, we also benefit from OzNet to extract features from brain hemorrhage datasets. OzNet effectively extracts features from the dataset, thanks to its large number of convolution layers. Also, we create OzNet with different filter sizes, filter numbers, padding, and stride, detailed in Table 2. In addition, we select the optimization method as the stochastic gradient descent, 0.95 as the momentum parameter, and 0.0001 as the learning rate. In this study, we also compare with having different layers CNNs such as ResNet-18[50], MobileNetv2[51], and ShuffleNet[52]. These CNNs properties are shown in Table 2.

Table 2. Properties of CNN architectures compared with OzNet

Pre-trained architectures	Parameters (millions)	Input Image Size	Layers	Year
ResNet-18[50]	11.7	224x224	71	2016
MobileNetv2[51]	3.5	224x224	154	2017
ShuffleNet[52]	1.4	224x224	172	2018

Table 3. Details of OzNet parameters.

Name of Layer	***Type of Layer***	***Filter Size***	***Filters***	***Stride***	***Padding***	***Output Size***
Input	Image Input					227x227x3
Conv-1	Convolution 2D	64	5x5	1	1	225x225x64
MaxPool-1	Max Pooling		3x3	2	0	112x112x64
Conv-2	Convolution 2D	128	3x3	1	1	112x112x128
MaxPool-2	Max Pooling		3x3	2	0	55x55x128
Conv-3	Convolution 2D	128	13x13	1	0	55x55x128
MaxPool-3	Max Pooling		3x3	2	0	27x27x128
Conv-4	Convolution 2D	256	7x7	1	1	27x27x256
MaxPool-4	Max Pooling		2x2	2	0	13x13x256
Conv-5	Convolution 2D	128	3x3	1	1	13x13x128
MaxPool-5	Max Pooling		3x3	2	0	6x6x128
Conv-6	Convolution 2D	128	3x3	1	1	6x6x128
MaxPool-6	Max Pooling		3x3	2	0	3x3x128
Conv-7	Convolution 2D	128	3x3	1	1	3x3x128
MaxPool-7	Max Pooling	-	2x2	2	0	1x1x128
FC-8	Fully Connected	4096				1x1x4096
Drop-8	Dropout	50%				
FC-9	Fully Connected	number of classes				1x1x (number of classes)
Softmax	Softmax					1x1x (number of classes)
Output	Classification	Cross entropy				

2.3 Feature Selection with Neighborhood Component Analysis (NCA)

In high-dimensional data, an appropriate feature selection is prominent [53]. Many researchers focus on dimension reduction. However, significant features may be lost during dimension reduction [53, 54]. Principle component analysis (PCA) is a classical linear feature selection method. Essentially, this method is used for dimension reduction. Also, PCA may be caused to the loss of important features when the process has a lower dimension [55]. Linear Discriminant Analysis (LDA) is also a classic and parametric method for feature selection. LDA is a method that acquires knowledge of a linear combination of features to protect the discriminative information of the class labels [55]. LDA presumes that all classes come from the same Gaussian probability density and have the same covariance matrix. However, LDA encounters a petty sample size cause when tackled with high-dimensional data when the inside class scatters matrix is almost singular [56]. Moreover, Sequential Feature Selection (SFS) is also a popular method. But it is not able to take out features that become insignificant after attaching other features [53, 54]. On the contrary, NCA is not issued any assumptions. While the dimension reduction process, no information will be lost. This property takes advantage that it does not lose any significant features[53].

NCA is a method that uses a Mahalanobis distance metric in the supervised K-Nearest Neighbour algorithm with leave-one-out (LOO) [56]. The method immediately maximizes a stochastic variant of the LOO K-nearest neighbor score on the training set [56]. NCA learns a feature weighting vector via maximizing the anticipated LOO classification accuracy by a regularization term [53]. Moreover, we suggest Goldberger et. al [56], Yang et. al [57], and Shang et. al [53] for a detailed explanation of NCA.

Let $T = \{(x_1, y_1),...,(x_i, y_i),...,(x_N, y_N)\}$ be a set of training data. Here, x_i is shown as a feature vector that is d- dimensional, $y_i \in \{1,...,C\}$ is the corresponding class label. Additionally, N is shown as a number of samples [57]. Here, the aim is to ascertain a weighting vector w by using the Mahalanobis distance metric. w is demonstrated as follows [53, 57]:

$$D_w(x_i, x_j) = \sum_{l=1}^{d} w_l^2 \left| x_{il} - x_{jl} \right| \tag{1}$$

Here, l^{th} feature weight is shown as w_l. LOO maximizes classification accuracy on where choose a random point from a set of training data T to be labeled correspondingly [53, 57]. The reference point is approximately identified via a probability function where the probability of x_i chooses x_j as its determined point is represented as[57]:

$$p_{ij} = \begin{cases} \dfrac{K\left(D_w(x_i, x_j)\right)}{\sum_{k \neq i} K\left(D_w(x_i, x_k)\right)}, & if \quad i \neq j \\ 0, & if \quad i = j \end{cases} \tag{2}$$

here, $K(.)$ is a kernel function, $K(z)=\exp(-z/\sigma)$ and σ is a kernel width. The reference point is chosen only the nearest neighbor of the determined sample, while $\sigma \to 0$ is. Otherwise, each point has a similar advantage being chosen other than the determined point, while $\sigma \to \infty$ is [57]. The probability p_i can be expressed that the average probability of LOO accurate classification. It is identified as follows:

$$p_i = \sum_{j=1,\, j\neq i}^{n} p_{ij}\, y_{ij} \tag{3}$$

here,

$$y_{ij} = \begin{cases} 1, & if \quad y_i = y_j \\ 0, & otherwise \end{cases} \tag{4}$$

Thus, the approximate LOO classification accuracy is able to be computed as follows [57]:

$$\varepsilon(w) = \frac{1}{N}\sum_i p_i = \frac{1}{N}\sum_i \sum_j y_{ij}\, p_{ij} \tag{5}$$

NCA aims to perform feature selection and reduce overfitting [57]. Thus, it is added a regularization parameter to achieve the objective function as follows:

$$\varepsilon(w) = \sum_i p_i = \sum_i \sum_j y_{ij}\, p_{ij} - \lambda \sum_{l=1}^{d} w_l^{\,2} \tag{6}$$

here, equation (5), $1/N$ is an unconsidered coefficient because it only expresses that the parameter of λ performs a corresponding alter and the last solution vector is identical [57]. In addition, $\varepsilon(w)$ is a differentiable function, a derivative of $\varepsilon(w)$ with regard as w_l can be calculated as follows [57]:

$$\begin{aligned} \frac{\partial \varepsilon(w)}{\partial w_l} &= \sum_i \sum_j y_{ij} \left[\frac{2}{\sigma} p_{ij} \left(\sum_{k\neq i} p_{ik} \left| x_{il} - x_{kl} \right| - \left| x_{il} - x_{jl} \right| \right) w_l \right] - 2\lambda w_l \\ &= \frac{2}{\sigma} \sum_i \left(p_i \sum_{k\neq i} p_{ik} \left| x_{il} - x_{kl} \right| - \sum_j y_{ij}\, p_{ij} \left| x_{il} - x_{jl} \right| \right) w_l - 2\lambda w_l \\ &= 2 \left(\frac{1}{\sigma} \sum_i \left(p_i \sum_{j\neq i} p_{ij} \left| x_{il} - x_{jl} \right| - \sum_j y_{ij}\, p_{ij} \left| x_{il} - x_{jl} \right| \right) - \lambda \right) w_l \end{aligned} \tag{7}$$

In accordance with the above, it can be obtained the suitable gradient ascent modernized equation. Also, the gradient descent method is used to maximize the objective function [53].

NCA method is a powerful discriminator that can select significant features. In this study, we benefit from this non-parametric feature selection method to improve a type of brain

hemorrhage classification performance. Further, we utilize this method in our created three different datasets. As a result, we can emphasize that NCA increases classification performance.

2.4 Classifiers

In this section, we briefly present different classifiers that are ANNs, Decision Tree, K-NN, LDA, Naïve Bayes, SVM, Adaboost, and Bagging. Further, we also give tuning parameters which are determined by trial and error for the best classification results.

2.4.1 Artificial Neural Networks: ANNs

Artificial Neural Networks (ANNs) are a wealthy family of nonlinear models [58]. Recently, it is successfully implemented in machine learning, especially supervised learning. Firstly, ANNs were revealed from the neuron networks of animals and humans, but powerful mathematical reasons were also discovered to promote them in theory [58]. Further, ANNs have flexible structures that can be applied to fit diverse real-world data [58]. For detailed information, we suggest a book prepared by H. Jiang [58]. In this study, we utilize 10 hidden layers with a ReLU activation function and softmax layer. Besides, we use the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm as a training solver. Moreover, we determine the number of maximum iteration 1000, the learning rate 0.01, minimum gradient tolerance 1e-6, and loss tolerance 1e-6 for each experiment. In this study, selected parameters for ANNs give the best classification performance.

2.4.2 Decision Tree

The decision tree is used as a non-parametric machine learning algorithm for regression or classification [58, 59]. The algorithm was designed with the program C4.5 by Ross Quinlan over 20 years starting in the late 1970s [60]. When the decision tree is applied to the classification problem, it will have a hierarchical structure. Input tree nodes correspond to splits that separate the domain into areas, and terminal nodes appoint class labels to areas that rely on being sufficiently small or sufficiently uniform. Splits are identified by some related conditions based on selected features that might have two or more outcomes [61]. In the decision tree, when considered a two-dimension feature vector $\boldsymbol{x} = [\boldsymbol{x}_1\ \boldsymbol{x}_2]^T$, each node except the terminal node is related with a binary question regarding a feature element x_i and a threshold t_j receiving the form $x_i \le t_j$ [58]. Here, each leaf node indicates an area $\boldsymbol{R}_j$ in the input space. When taking any input feature such as x_1 the algorithm begins from the stem node and inquires about the question related with the node. While the answer is $True$, it drops into the left descendant node. Otherwise, it drops into the right descendant node. This transaction proceeds up to arrive at a leaf node [58]. In this study, we utilize standard CART for the selection of predictors and select Gini's diversity index as a split criterion. The maximum category in splits for decision trees is 10.

2.4.3 K-Nearest Neighbor (KNN)

The k-nearest-neighbor algorithm is labor demanding and did not obtain popularity until more computing power became available. The expression of nearest is calculated by a distance metric like Euclidean distance [60]. For classification, the fundamental of the nearest neighbor algorithm appoints an input sample vector y, that is of unknown classification, to its nearest neighbor class [62, 63]. This opinion might be expanded to the K-nearest neighbors with the vector y becoming appointed to the class that is symbolized by a larger number between the K-nearest neighbors [62]. Considering more than one neighbor, there is a probability of a link between the classes with the largest number of neighbors in the K-nearest neighbor group. A simple way to solve this problem is to limit the possible K values [62]. Here, the k value that generates the minimum error rate is chosen. In general, rising the value of k boosts the number of training [64]. In this study, we select the distance metric as Euclidean distance. Furthermore, we determine equal distance weight and 10-nearest neighbors according to the best classification performance.

2.4.4 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) was suggested by Ronald Fisher in 1936 [65]. It occurs in the detection of the projection hyperplane that reduces minimum into class variance and increases maximum the distance among the projected classes' means [65]. These two targets can be unraveled by handling an eigenvalue issue by correlating with eigenvectors identifying the hyperplane of regard [65]. Additionally, this case is similar to PCA. In order to explain the significance of the granted features, to classify and reduce dimension, this hyperplane can be applied [65]. In this study, the discriminant type is chosen as linear. As known, when the type is linear, each class has the same covariance matrix. The parameter of the delta is taken log scaled in the range [1e-6, 1e3] in this study. Additionally, the parameter of gamma is real values in [0,1].

2.4.5 Naïve Bayes (NBayes)

The Naive Bayes algorithm is very resilient in tackling a variety of features. According to the distribution of a feature x_i, it can be selected conditional distribution $p\left(x_i \middle| y\right)$ [58]. Namely, while a feature is binary, nonbinary discrete, and continuous, it can be chosen Bernoulli, multinomial and Gaussian distribution, respectively. In the algorithm, the number of features is totally linear with the number of parameters [58]. The algorithm's learning and inference can be performed with some enclosed-form remedies, that is also linear in the number of diverse features [58]. Consequently, this algorithm is extremely scalable to major issues that include an immense number of diverse features [58]. In this study, we chose multinomial distribution because of having the multinomial classes of our datasets.

2.4.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a well-known machine learning algorithm used for regression, classification, and outlier analysis. It was presented by Cortes and Vapnik [66] in 1995. Especially, SVM classifies with a high accuracy rate to many high dimensional data for classification [67]. Besides, SVM has many distinguishing properties. One of these properties is efficient separation with a kernel-based method to the datasets for classification [68].

Moreover, SVM can be worked with many algorithms thanks to its flexible and adaptable structure. Additionally, it can overcome overfitting. In this study, we utilize the Gaussian kernel function because this kernel can map the nonlinear features effectively.

2.4.7 Adaboost

Adaptive Boosting algorithm is designed for classification. It can be implemented in any classification method such as bagging. This algorithm starts by appointing equal weight to whole samples in the training data. Next, it invokes the learning algorithm to create a classifier for the data and each sample based on the output of the classifier. The weight of accurately classified samples is reduced, and that of misclassified ones is raised. This creates a group of "easy" samples with low weight and a group of "hard" ones with high weight. In the following stage, and whole upcoming ones, a classifier is set for reweighted data, which lastly adapts to classifying the hard samples accurately. And then, the weights of samples are changed considering this new classifier's output. Consequently, in practice, many possibilities can take place. Such as, some hard samples might be easier and easier ones harder. After all these, the weights show how often samples have been misclassified by the classifiers generated thus far. This implementation provides a graceful way to form a complementary set of experts while maintaining a measurement of "hardness" in each sample[60, 69]. In this study, we select Gini's diversity index as a split criterion and the maximum category in split samples as 10. Moreover, the decision trees are taken learner and 10 trees are grown in this study. Particularly, we utilize standard CART for the selection of predictors.

2.4.8 Bagging

Merging diverse models' decisions signifies integrating the various outputs into a single prediction. The uncomplicated way to perform this in the status of classification is to utilize voting. Although bagging and boosting both utilize this procedure, implementations are different. While bagging has equal weight, boosting uses efficient weight that it is determined the situation of success. The trees are established for the novel explored training sets, and their predictions also attend in the vote [60]. Especially, the ensemble classifier will generally be more correct than one decision tree built from just one of the datasets [60]. To deactivate the indecision of learning methods bagging alters the actual training data by removing some samples and multiplying others, rather than sampling a novel, independent training dataset every time [60]. Here, to create new samples of the same size as the actual dataset, these are randomly sampled by replacing. The generated different datasets are generally dependent since they are based on a single actual dataset. But, when bagging compared with another single classifier, it is often seen that its performance is better than the others. Hence, the primary advantage of bagging is that it significantly reduces the model variance [70]. When complicated classifiers are utilized, it can alleviate overfitting with this property [58]. In this study, we choose the decision tree as a learner. Also, we identified the number of trees as 10 and the number of learning cycles as 50.

Results and Discussion

In this study, we have combined the machine learning algorithms with OzNet. Firstly, we classify all created Datasets with OzNet to measure performance. Further, we use OzNet for the feature extractor as well. Next, we obtain reduced features by the fully connected layer of OzNet, then we classify them with these algorithms. After all, we perform NCA to select the significant feature and to eliminate others. In last, we also utilize different classification algorithms to classify selected features. Therefore, we achieve an excellent structure with high accuracy on brain hemorrhage CT images. Moreover, we performed 10- fold cross-validation because of a very confident method when the data is restricted. The dataset is divided into identified number folds randomly [68]. One sub-fold is considered the test fold, it trains the framework with the remaining folds. This process is repeated up to other folds and is tested in the framework [71].

3.1. Performance Metrics:

In this study, we assess OzNet and hybrid algorithms concerning performance metrics which are accuracy, sensitivity, specificity, precision, F1-Score, and G-mean as follows [72].

$$Accuracy = \left(TP + TN\right) / \left(TP + TN + FP + FN\right) \quad (8)$$

$$Sensitivity = TP / \left(TP + FN\right) \quad (9)$$

$$Specificity = TN / \left(TN + FP\right) \quad (10)$$

$$\mathrm{Pr}ecision = TP / \left(TP + FP\right) \quad (11)$$

$$F1 - Score = \left(2 \times TP\right) / \left(2 \times TP + FP + FN\right) \quad (12)$$

$$G - Mean = \sqrt{Sensitivity \times Specificity} \quad (13)$$

where TP: True Positive, FP: False Positive, TN : True Negative, and FN: False Negative are introduced in Equation (8-13).

3.2. Experimental Results

In this study, we developed a new CNN architecture named as OzNet in a MATLAB environment with Intel Core i7-7500U CPU, NVIDIA GeForce GTX 950M,16 GB RAM, 64-bit Operating System. Further, we examine three different created brain hemorrhage CT images, presented in the materials and methods section. Firstly, we reduce all CT image dimensions from 650x650 to 227x227. Then, we create two different data sets by age and combine them all for the third dataset. Essentially, we utilize no pre-processing method due to measuring just OzNet performance on raw datasets. Table 4 shows OzNet performance for each dataset. Besides, Fig. 4 exhibits a training and validation accuracy and loss graph using OzNet.

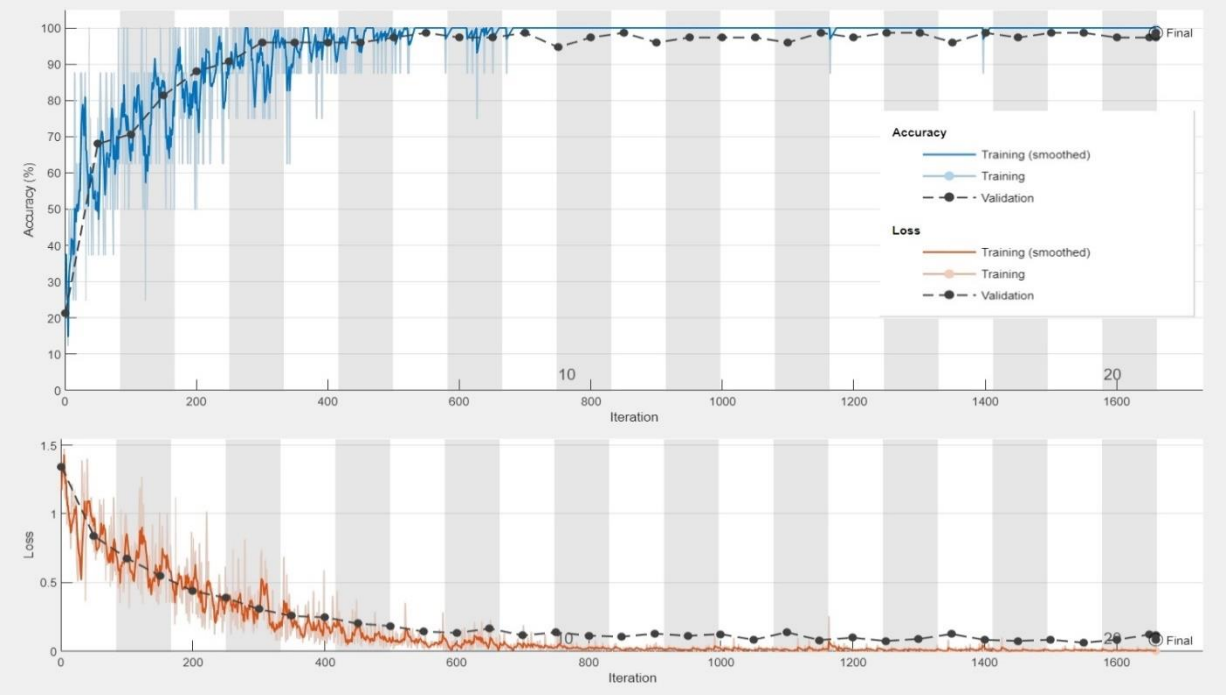


Fig. 4 A training and validation accuracy and loss graph using OzNet in this study.

While Dataset 1 is created for 18 ages and under or children of brain CT images to classify three different type classes: EDH, SDH+IPH, and N, Dataset 2 is generated over 18 ages or adults' brain CT images to classify three different types of classes: EDH, N, and IPH+SDH. Additionally, Dataset 3 is merged all brain CT images to classify four different type classes: EDH, IPH, N, and SDH.

Table 4. Performance of OzNet with three different datasets.

	AUC	Accuracy	Sensitivity	Specificity	Precision	F-measure	G-Means
Dataset 1	0.9935	0.9388	0.9391	0.9709	0.9382	0.9384	0.9549
Dataset 2	0.9923	0.9611	0.9556	0.9814	0.9608	0.9581	0.9684
Dataset 3	0.9838	0.9285	0.9195	0.9493	0.9317	0.9253	0.9343

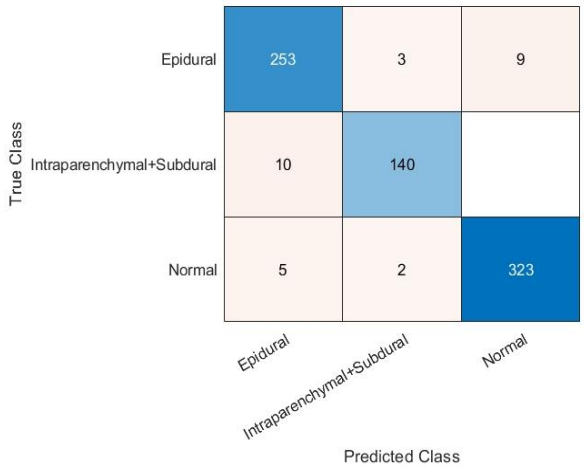
When Table 4 is investigated for Dataset 1, the overall accuracy is obtained 93.88%. However, SDH+IPH, EDH, and N classes accuracy are achieved 97.6%, 92.3%, and 91.1%, respectively. Here, OzNet distinguishes SDH+IPH class better than others.

When Table 4 is examined for Dataset 2, the overall accuracy is obtained 96.11%. However, IPH +SDH, EDH, and N classes accuracy are achieved 96.6%, 94.4%, and 97.3%,

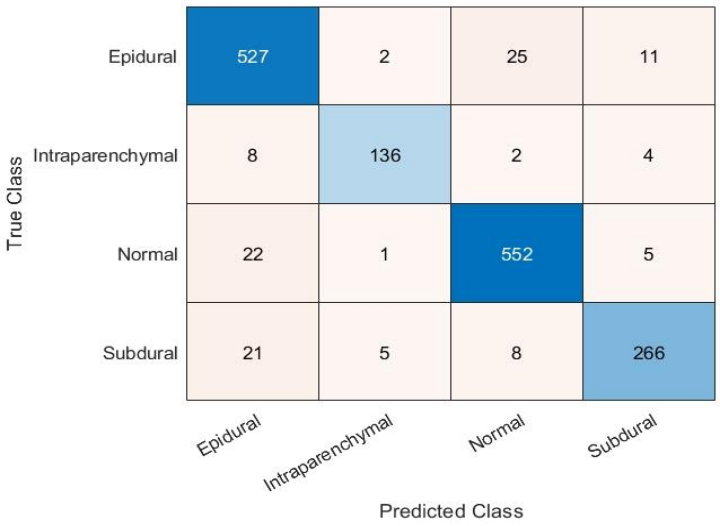
respectively. Here, OzNet distinguishes N class better than others. Essentially, we can express that OzNet classifies to Dataset 2 much better than others.

When Table 4 is analyzed for Dataset 3, the overall accuracy is obtained 92.85%. However, SDH, IPH, EDH, and N classes accuracy are achieved 93%, 94.4%, 91.2%, and 94%, respectively. Here, OzNet distinguishes IPH class better than others. Note that, Dataset 3 has four classes, and its sample size is larger than the others. Moreover, datasets are imbalanced datasets. Despite all these negative situations, Oznet classifies very effectively on brain hemorrhage CT datasets. While Fig.5 shows the confusion matrix of OzNet, Fig.6 demonstrates the ROC-Curve of OzNet on these datasets.

(a)Dataset 1



(b) Dataset 2



(c) Dataset 3

Fig.5 Confusion matrix of OzNet on different brain hemorrhage datasets

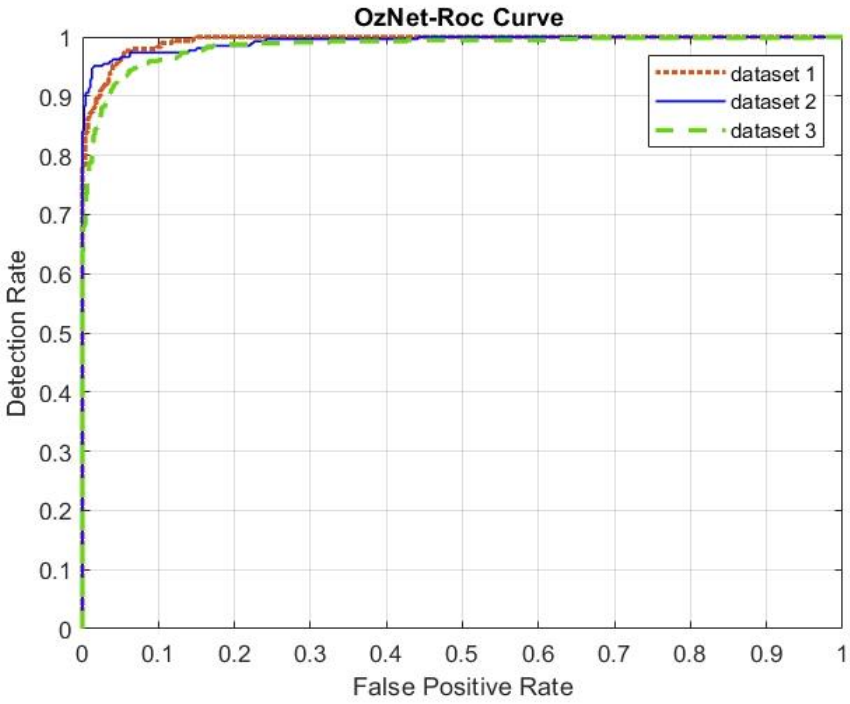


Fig.6 Roc-Curve graph of OzNet on different brain hemorrhage datasets.

In this study, we also compare OzNet performance with ResNet-18[50], MobileNetv2[51], and ShuffleNet[52] having 71, 154, and 172 layers, respectively, under the same conditions. Further, we create 2 different CNN architectures having fewer layers than OzNet and compare them with OzNet. First CNN architecture (CNN-1) has 16 layers which are 3 convolutional, 3 batch normalization, 3 Relu, 3 maximum pooling, 1 fully connected, and 1 SoftMax. Second CNN architecture (CNN-2) has 20 layers which are 4 convolutional, 4 batch normalization, 4 Relu, 4 maximum pooling, 1 fully connected, and 1 SoftMax. CNN-1 and CNN-2 have 3x3 filter sizes for each layer. Besides, the first convolutional filter numbers are beginning with 64, then the filter numbers in the next layers are doubled till reached 256 in CNN-1 and CNN-2. However, in CNN-2, the last layer's filter numbers have 256 again. All filter sizes and filter numbers are designed with trial and error.

According to Table 5, the overall accuracy is obtained 84% for Dataset 1. Besides, the sensitivity, specificity, precision, F-measure, and G-means are acquired 84.67%, 83.64%, 73.84%, 84.67%, and 84.15% respectively. While OzNet achieved all metrics over 93% for Dataset 1, ResNet-18 achieved lower performance. According to Table 6, the overall accuracy is acquired 81.76 % for Dataset 1. Similarly, OzNet's performance is superior to MobileNet. According to Table 7, the overall accuracy is obtained 70 % for Dataset 1. Identically, Oznet's performance is superior to ShuffleNet. According to Table 8 and Table 9, the overall accuracies are acquired 73.53 % and 81.76 % for Dataset 1, respectively. Eventually, we can express that OzNet was highly successful for Dataset 1.

According to Table 5, the overall accuracy is obtained 90.34% for Dataset 2. In addition, the sensitivity, specificity, precision, F-measure, and G-means are acquired 85.66%, 92.92%, 86.97%, 86.31%, and 89.22% respectively. Performance results of Oznet for Dataset 2 were given above in Table 4. While OzNet achieved all metrics over 95.8% for Dataset 2, ResNet-18 achieved lower performance. According to Table 6, the overall accuracy is obtained 85.5% for Dataset 2. In addition, the sensitivity, specificity, precision, F-measure, and G-means are acquired 79.25%, 88.96%, 79.85%, 79.25%, and 79.55% respectively. As seen, OzNet is more successful than MobileNet. According to Table 7, this situation does not differ from others. ShuffleNet achieves an accuracy of 74.5% and it is also lower performance than OzNet for Dataset 2. According to Table 8 and Table 9, the overall accuracies are acquired 87.92 % and 87.25% for Dataset 2, respectively.

According to Table 5, the overall accuracy is obtained 80.88% for Dataset 3. Further, the sensitivity, specificity, precision, F-measure, and G-means are acquired 72.57%, 85.44%, 73.22%, 72.89%, and 78.74% respectively. While OzNet achieved an accuracy of 92.85% for Dataset 3, ResNet-18 achieved an accuracy of 80.88%. According to Table 6, the overall accuracy is obtained 76.93% for Dataset 3. As seen, MobileNet performs lower than OzNet for Dataset 3. According to Table 7, ShuffleNet acquires an accuracy of 65.96% for Dataset 3. According to Table 8 and Table 9, the overall accuracies are acquired 86.21 % and 71.47% for Dataset 3, respectively.

In general, OzNet is a more successful performance than other CNN architectures for Brain Hemorrhage CT images.

Table 5. Performance of ResNet-18 with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	0.9370	0.8400	0.8467	0.8364	0.7384	0.8467	0.8415
Dataset 2	0.9663	0.9034	0.8566	0.9292	0.8697	0.8631	0.8922
Dataset 3	0.9231	0.8088	0.7257	0.8544	0.7322	0.7289	0.7874

Table 6. Performance of MobileNet with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	0.9148	0.8176	0.7700	0.8436	0.7287	0.7700	0.7488
Dataset 2	0.9460	0.8550	0.7925	0.8896	0.7985	0.7925	0.7955
Dataset 3	0.8936	0.7693	0.7027	0.8058	0.6650	0.7027	0.6833

Table 7. Performance of Shufflenet with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**

Dataset 1	0.8255	0.7000	0.6700	0.7164	0.5630	0.6700	0.6119
Dataset 2	0.8816	0.7450	0.7434	0.7458	0.6176	0.7434	0.6747
Dataset 3	0.8166	0.6596	0.6885	0.6437	0.5146	0.6885	0.5889

Table 8. Performance of CNN-1 with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	0,9314	0,7353	0,7444	0,8722	0,7796	0,7369	0,8058
Dataset 2	0,9515	0,8792	0,8423	0,9332	0,9050	0,8615	0,8866
Dataset 3	0,9805	0,8621	0,8397	0,9503	0,8565	0,8445	0,8933

Table 9. Performance of CNN-2 with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	0,9662	0,8176	0,8256	0,9116	0,8341	0,8122	0,8675
Dataset 2	0,9737	0,8725	0,8625	0,9295	0,8826	0,8676	0,8954
Dataset 3	0,9258	0,7147	0,6852	0,8968	0,7394	0,6855	0,7839

Although OzNet's performance is very high on raw datasets, we merged OzNet with Adaboost, ANN, Bagging, Decision Tree (DTree), K-NN, LDA, Naïve Bayes, and SVM. Because we would like to achieve the best performance for the classification of brain CT images. Thus, we investigate the performance of the machine learning algorithms, given in Table 10. Firstly, we obtain 4096 features for each image from the fully connected layer of OzNet. Next, we split the features as 70% training set and 30% testing set and activate them taking minibatch size 20 for the next step. Then, we classify with the presented algorithms. This process is implemented for each dataset: Dataset 1, Dataset 2, and Dataset 3.

When we investigate Table 10 in terms of performance metrics, performances of the hybrid algorithms give us very well results, except OzNet-Bagging, DTree, and LDA for Dataset 2. However, OzNet-SVM obtains the highest accuracy and sensitivity rate on both Dataset 1and Dataset 3. Its accuracy and sensitivity rates for Dataset 1 are 98.83% and 98.74%, respectively. Additionally, its accuracy and sensitivity rates for Dataset 3 are 98.95% and 98.32%, respectively. However, OzNet-Adaboost and OzNet-ANN achieve the same accuracy of 98.66% for Dataset 2. Other performance metrics of the classifiers are almost similar. As a result, we can express that OzNet-SVM is a good classifier for Dataset1 and Dataset 3, also OzNet-Adaboost and OzNet-ANN are good classifiers for Dataset 2.

Table 10. Performance of hybrid algorithms with three different datasets.

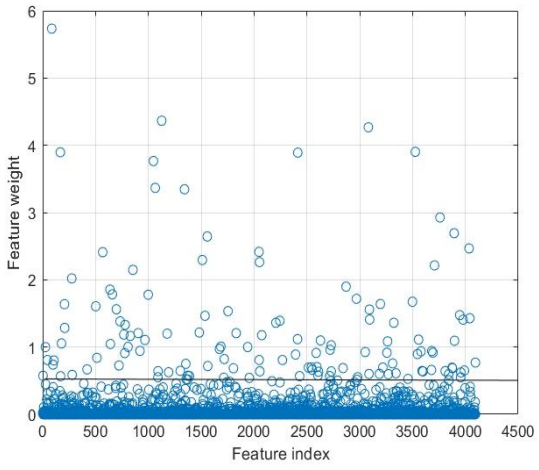
Classifier	Dataset	AUC	Accuracy	Sensitivity	Specificity	Precision	F-measure	G-Means
Oznet-AdaBoost	Dataset 1	**1**	0.9686	0.9689	0.9842	0.9689	0.9689	0.9765
Oznet -ANN	Dataset 1	0.9996	0.9843	0.9852	0.9919	0.9853	0.9852	0.9886
Oznet -Bagging	Dataset 1	0.9999	0.9843	0.9837	0.9921	0.9845	0.9841	0.9879
Oznet -DTree	Dataset 1	0.9993	0.9686	0.9681	0.9843	0.9691	0.9681	0.9762
Oznet -KNN	Dataset 1	0.9591	0.9529	0.9533	0.9761	0.9552	0.9538	0.9646
Oznet -LDA	Dataset 1	**1**	0.9843	0.9829	0.9921	0.9845	0.9836	0.9875
Oznet -NBayes	Dataset 1	0.9944	0.9843	0.9837	0.9921	0.9845	0.9841	0.9879
Oznet-SVM	Dataset 1	0.9999	**0.9883**	**0.9874**	**0.9941**	**0.9881**	**0.9877**	**0.9908**
Oznet-AdaBoost	Dataset 2	0.9988	**0.9866**	**0.9882**	**0.9936**	0.9825	0.9851	**0.9909**
Oznet -ANN	Dataset 2	**0.9996**	**0.9866**	0.9850	0.9927	0.9883	**0.9870**	0.9888
Oznet -Bagging	Dataset 2	0.9918	0.9641	0.9520	0.9807	0.9695	0.9593	0.9663
Oznet -DTree	Dataset 2	0.9799	0.9552	0.9404	0.9761	0.9619	0.9491	0.9581
Oznet -KNN	Dataset 2	0.9712	0.9776	0.9797	0.9886	0.9748	0.9769	0.9842
Oznet -LDA	Dataset 2	0.9967	0.9641	0.9575	0.9812	0.9621	0.9598	0.9693
Oznet -NBayes	Dataset 2	0.9936	0.9866	0.9850	0.9927	**0.9883**	0.9867	0.9888
Oznet-SVM	Dataset 2	0.9993	0.9821	0.9767	0.9899	0.9850	0.9807	0.9834
Oznet-AdaBoost	Dataset 3	**1**	0.9875	0.9831	0.9954	**0.9887**	**0.9857**	0.9892
Oznet -ANN	Dataset 3	0.9998	0.9875	0.9804	0.9955	0.9886	0.9844	0.9879
Oznet -Bagging	Dataset 3	0.9993	0.9833	0.9802	0.9938	0.9856	0.9829	0.9870
Oznet -DTree	Dataset 3	0.9883	0.9603	0.9426	0.9848	0.9691	0.9546	0.9635
Oznet -KNN	Dataset 3	0.9631	0.9498	0.9311	0.9815	0.9491	0.9396	0.9559
Oznet -LDA	Dataset 3	0.9949	0.9623	0.9383	0.9868	0.9570	0.9463	0.9623
Oznet -NBayes	Dataset 3	0.9985	0.9749	0.9624	0.9907	0.9748	0.9678	0.9765
Oznet-SVM	Dataset 3	0.9999	**0.9895**	**0.9832**	**0.9966**	0.9832	0.9832	**0.9899**

Even though these hybrid algorithms' performances are very high on these created datasets, we utilize NCA to achieve excellent classification performance on datasets. Besides, we apply OzNet to extract features as well as classification. As soon as trained OzNet for each dataset, we select features that have a dimension of 4096 x (number of images) from a fully connected layer (FC-8). The stage up to here is the same as the previous application. Then, a determined tolerance value and parameter of λ to get minimum loss is used for non-negative feature weights. NCA learns feature weights to minimize objective function with LOO on training data. Lastly, significant features are chosen with minimum loss. Eventually, these features were utilized for the SVM classifier. Table 11 and Fig. 7 demonstrate to tolerance value according to finding minimum loss for each dataset. In this study, the parameter of λ is determined as 0.005 to minimize loss for each experiment. Also, the optimization method is used SGD (stochastic gradient descent) and it is limited to 30 iterations with 0.0001 gradient tolerance.

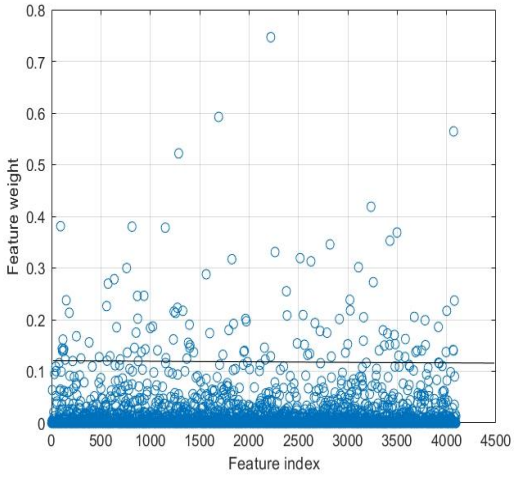
Table 11. Tolerance and loss values for each experiment.

	Dataset 1	***Dataset 2***	***Dataset 3***
Tolerance	0.05	0.12	0.2
Loss(MSE*)	0.0039	0.0045	0.0042
λ	0.005	0.005	0.005
Solver	Sgd	Sgd	Sgd
Iteration	30	30	30
Gradient Tolerance	0.0001	0.0001	0.0001

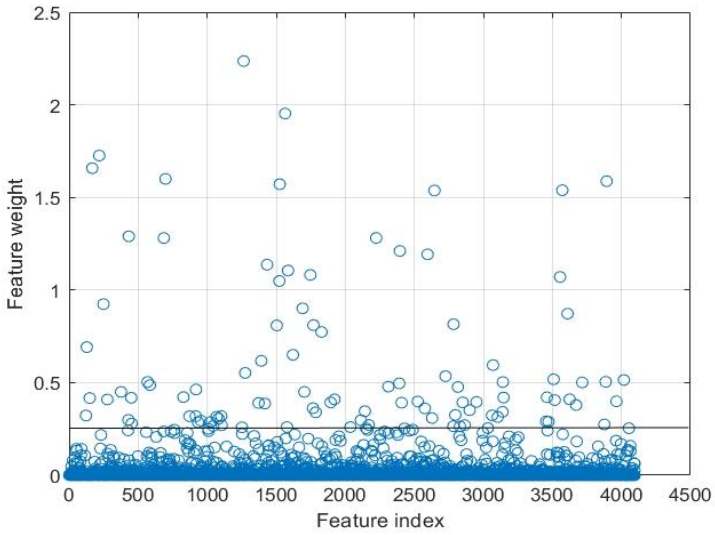
MSE*****: Mean Square Error



(a)Dataset 1



(b)Dataset 2



(c) Dataset 3

Fig.7 Graphs represent to select features by feature weight.

In these conditions, we select 229 features from Dataset 1, 112 features from Dataset 2, and 45 features from Dataset 3. As seen, 4096 features are reduced by NCA for each experiment. After all, we employed Adaboost, ANN, Bagging, Decision Tree (DTree), K-NN, LDA, Naïve Bayes, and SVM to classify these selected significant features, like the previous stage. Table 6 shows the performance metrics of the OzNet hybrid algorithms by using NCA for each dataset. Besides, Fig.8 demonstrates the confusion matrix for each experiment.

When Table 14 is examined for Dataset 1, OzNet-ANN, OzNet-KNN, and OzNet-SVM obtain the overall maximum accuracy of 99.61% by using NCA. Otherwise, it is maximum accuracy of 98.83%. Exactly, as seen this hybrid algorithm is almost 100% distinguishes each class. For example, OzNet-SVM achieves SDH+IPH, EDH, and N classes' accuracy of 98.9%, 100%, and 100%, respectively.

When Table 14 is investigated for Dataset 2, OzNet-ANN acquires the overall accuracy of 100% by using NCA. Moreover, the algorithm has an accuracy of 100% for each class. Particularly, each performance metric is also 100%. This is a perfect algorithm for Dataset 2 while the maximum accuracy is 98.66% without applying NCA.

When Table 14 is analyzed for Dataset 3, OzNet-ANN and OzNet-Bagging achieve an overall accuracy of 99.58% by using NCA. In the previous stage, the maximum accuracy was 98.95%.

In this study, when we utilized NCA, we generally obtained the best structure with Oznet-ANN. However, we show the results for two widely used feature selection methods: Chi-square and minimum Redundancy Maximum Relevance (mRMR) for the same structure. These methods are also non-parametric methods and features are ranked in both feature selection methods[73-75].

According to Table 12, the overall accuracies are 97.65 %, 98.21 %, and 99.16 for Dataset 1, Dataset 2, and Dataset 3 respectively. When we investigated Oznet-NCA-ANN, the overall accuracies are 99.61 %, 100 %, and 99.58% for Dataset 1, Dataset 2, and Dataset 3 respectively. Although the performance of OzNet-Chi-square-ANN is seen as successful, Oznet-NCA-ANN is superior.

According to Table 13, the overall accuracies are 99.21 %, 99.10 %, and 99.16% for Dataset 1, Dataset 2, and Dataset 3 respectively. As seen, OzNet-mRMR-ANN is also a powerful structure. However, OzNet-NCA-ANN possesses the highest performance to detect different types of hemorrhage.

Table 12. Performance of OzNet-Chi-square-ANN with three different datasets.

	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	0,9998	0,9765	0,9748	0,9882	0,9762	0,9754	0,9815
Dataset 2	0,9996	0,9821	0,9816	0,9916	0,9772	0,9793	0,9866
Dataset 3	0,9997	0,9916	0,9902	0,9974	0,9850	0,9875	0,9938

Table 13. Performance of OzNet-mRMR-ANN with three different datasets.

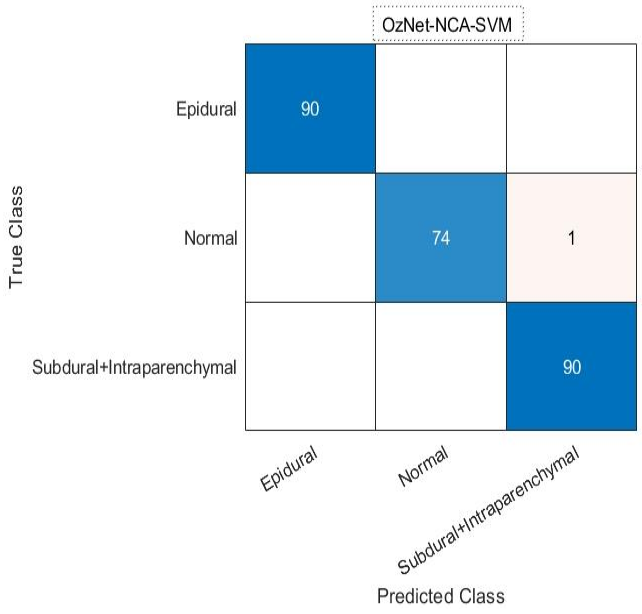
	AUC	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F-measure**	**G-Means**
Dataset 1	1	0,9921	0,9918	0,9961	0,9919	0,9918	0,9939
Dataset 2	0,9965	0,9910	0,9892	0,9954	0,9917	0,9922	0,9922
Dataset 3	0,9997	0,9916	0,9928	0,9971	0,9916	0,9922	0,9949

As seen, the success of these hybrid algorithms is increased with applying NCA. As mentioned before, dataset 3 has four classes, and datasets are imbalanced. In all these conditions, the Oznet-NCA-ANN algorithm achieved excellent performance on the brain hemorrhage CT dataset: Dataset 3. Additionally, Fig.8 shows the confusion matrix of the best algorithm for each dataset.

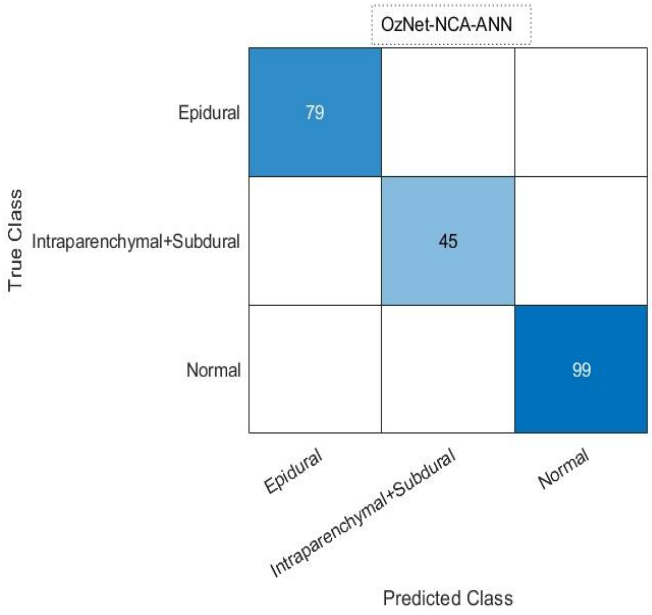
Ozaltin, O., Coskun, O., Yeniay, O., & Subasi, A. (2022). Classification of brain hemorrhage computed tomography images using OzNet hybrid algorithm. *International Journal of Imaging Systems and Technology*.

Table 14. Performance of hybrid algorithms with three different datasets by using NCA.

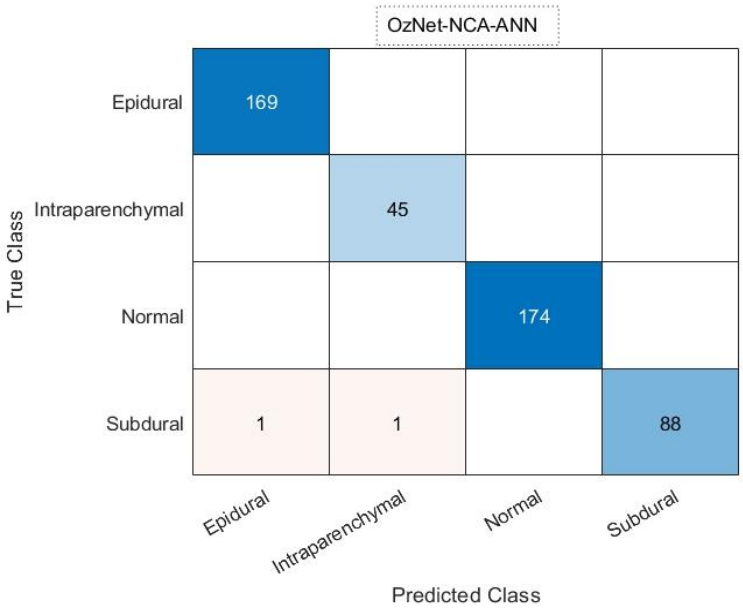
Classifier	Dataset	AUC	Accuracy	Sensitivity	Specificity	Precision	F-measure	G-Means
Oznet-AdaBoost	Dataset 1	**1**	0.9922	0.9926	0.9963	0.9913	0.9919	0.9944
Oznet -ANN	Dataset 1	0.9999	**0.9961**	0.9956	0.9979	**0.9963**	**0.9959**	0.9968
Oznet -Bagging	Dataset 1	0.9999	0.9922	0.9911	0.9959	0.9927	0.9918	0.9935
Oznet -DTree	Dataset 1	0.9996	0.9765	0.9755	0.9881	0.9774	0.9763	0.9817
Oznet -KNN	Dataset 1	0.9944	**0.9961**	0.9963	**0.9981**	0.9956	**0.9959**	**0.9972**
Oznet -LDA	Dataset 1	0.9999	0.9922	0.9918	0.9961	0.9919	0.9919	0.9940
Oznet -NBayes	Dataset 1	**1**	0.9882	0.9874	0.9939	0.9892	0.9882	0.9907
Oznet-SVM	Dataset 1	**1**	**0.9961**	**1**	0.9939	0.9890	0.9945	**0.9972**
Oznet-AdaBoost	Dataset 2	**1**	0.9955	0.9966	0.9977	0.9958	0.9962	0.9972
Oznet -ANN	Dataset 2	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Oznet -Bagging	Dataset 2	0.9998	0.9866	0.9890	0.9927	0.9883	0.9886	0.9909
Oznet -DTree	Dataset 2	0.9829	0.9686	0.9635	0.9842	0.9684	0.9654	0.9738
Oznet -KNN	Dataset 2	0.9873	0.9910	0.9916	0.9954	0.9894	0.9904	0.9935
Oznet -LDA	Dataset 2	0.9996	0.9866	0.9882	0.9923	0.9890	0.9886	0.9903
Oznet -NBayes	Dataset 2	0.9999	0.9910	0.9924	0.9950	0.9924	0.9924	0.9937
Oznet-SVM	Dataset 2	**1**	0.9955	**1**	0.9931	0.9875	0.9937	0.9965
Oznet-AdaBoost	Dataset 3	0.9999	0.9937	0.9915	0.9979	0.9931	0.9922	0.9947
Oznet -ANN	Dataset 3	0.9999	**0.9958**	0.9944	0.9986	0.9931	0.9937	0.9965
Oznet -Bagging	Dataset 3	0.9996	**0.9958**	0.9929	0.9987	0.9946	0.9937	0.9958
Oznet -DTree	Dataset 3	0.9988	0.9644	0.9591	0.9875	0.9573	0.9575	0.9732
Oznet -KNN	Dataset 3	0.9892	0.9874	0.9832	0.9957	0.9844	0.9837	0.9894
Oznet -LDA	Dataset 3	0.9999	0.9895	0.9875	0.9964	0.9850	0.9859	0.9919
Oznet -NBayes	Dataset 3	0.9968	0.9875	0.9819	0.9952	0.9914	0.9864	0.9885
Oznet-SVM	Dataset 3	0.9999	0.9937	0.9943	0.9953	0.9879	0.9910	0.9949



(a) Dataset 1



(b) Dataset 2



(c)Dataset 3

Fig.8 Confusion matrixes of OzNet hybrid algorithms on different brain hemorrhage datasets by using NCA.

3.3 Limitations of the study

In this section, we explain to limitations of the proposed hybrid algorithm. Primarily, the datasets include limited CT images and do not contain each type of hemorrhage. Additionally, more efficient results may obtain from MRI images, when CNN algorithms are used. Next, significant features are determined with a tolerance value on the graph of features. Essentially, this implementation does not different from trial and error. In this stage, there is a possibility of making a wrong decision. At this stage, there is a possibility of making a wrong decision. As a suggestion, we can say that the tolerance value should be determined more robustly.

3.4. Discussion

As we compare both performance metrics and confusion matrices, it can be seen clearly that NCA is an effective feature selection method. Before this implementation, OzNet could effectively classify despite the imbalanced datasets. However, we always aim to achieve better classification performance on biomedical images. Therefore, while we look for effective features selection method in the literature, we encounter the NCA non-parametric method. This method has practical properties that are no assumption and is non-parametric. Thus, the features can be easily selected since no conditions are sought. As known, Adaboost, ANN, Bagging, Decision Tree (DTree), K-NN, LDA, Naïve Bayes, and SVM are very powerful classification algorithms so preferred to classify these significant features. As a result, we can say that OzNet-NCA-SVM is a strong hybrid classification algorithm for Dataset1, namely child Brain CT Hemorrhage images. In addition, we can express that OzNet-NCA-ANN is an excellent hybrid classification algorithm for Dataset 2, namely adult Brain CT Hemorrhage images. Besides, we can represent that OzNet-NCA-ANN and OzNet-NCA-Bagging are powerful hybrid algorithms for Dataset 3 which is essentially included all Brain CT Hemorrhage images. In this study to eliminate the overfitting, while training, we use a dropout layer with 0.5 probability. Then, we utilize 10-fold cross-validation as well. Thus, we avoid overfitting and use NCA and machine learning algorithms. Therefore, excellent results are obtained using OzNet hybrid algorithms.

The most important advantage of this study is that it provides high accuracy classification results for different Brain CT Hemorrhage datasets. Of course, CNN architectures are powerful classifiers. However, it should be further strengthened with the hybrid structure. For this purpose, we further strengthened the architecture of OzNet via NCA and machine learning algorithms. We are aware that the performance of the OzNet architecture for other biomedical images needs to be measured, as well.

Therefore, we assess OzNet on other biomedical signals and images, for details we suggest examining the study of Ozaltin and Yeniay [76]. Further, Table 15 shows results in comparison with previous studies. As it can be seen from the table, the proposed method achieved better performance than the previous studies.

Table 15. Performance comparison with previous studies.

Study	Preprocessing	Classifier	Accuracy (%)
Dawud et al.[77]	AlexNet	SVM	93.48
Grewal et al.[78]	DenseNet	LSTM	81.82
Phong et al.[79]	-	LeNet	99.7
Singh et al.[28]	Shallow 3DCNN	Baseline 3D CNN	97
This study	OzNet+NCA	ANN	100

3. Conclusion

In this study, we developed a framework by combining OzNet with the Neighborhood Component Analysis (NCA) and classification algorithms such as Adaboost, ANN, Bagging, Decision Tree (DTree), K-NN, LDA, Naïve Bayes, and SVM. The proposed novel hybrid algorithms achieved excellent performance on real brain hemorrhage datasets. Firstly, we generate two different datasets from child and adult brain CT images named Dataset 1 and Dataset 2. These datasets have 3 classes. Then, we create the mixed dataset from these two datasets named Dataset 3. This third dataset has 4 classes. Secondly, we classify Dataset 1, Dataset 2, and Dataset 3 with OzNet and obtained an accuracy of 93.88%, 96.11%, and 92.85%, respectively. While the results are very well, our goal is to achieve better performance since the brain is a vital organ. Therefore, we utilized the NCA feature selection method. This method is famous to separate significant features from others. Additionally, we employ OzNet as a feature extractor and obtain features from a fully connected layer (FC-8). Next, we determine the tolerance value for minimum loss to select significant features via NCA. For each experiment, identified three different tolerance values. Thus, reducing feature dimensions for the next phase. Lastly, we benefit from these classification algorithms to classify these reduced features due to powerful machine learning algorithms. Also, we compare all results without NCA and with NCA, detailed in Table 4 and Table 6. As a result, we achieve maximum accuracy of 99.61%, 100%, and 99.58% for Dataset 1 by using OzNet-NCA-SVM, Dataset 2 by using OzNet-NCA-ANN, and Dataset 3 by using OzNet-NCA-ANN, respectively. Hence, the OzNet hybrid algorithm achieved better performance for each experiment by using NCA. Eventually, we are glad to say our developed framework is an excellent performance.

This study implements deep learning algorithms for Brain Hemorrhage (or ICH) CT image classification as a clinical decision support system. Therefore, doctors and experts will not consume much more time detecting brain hemorrhages, and the proposed framework help clinicians identify brain hemorrhage in a clinical environment in a better way. In future work, we are planning to apply the proposed framework to different biomedical datasets to check the validity algorithm. In fact, we want to collect real datasets such as MRI and CT images and detect different types of disorders by using OzNet model and the proposed hybrid algorithm. Besides, a new framework will be constructed with OzNet utilizing other feature selection methods to enhance the performance.

Acknowledgment

This study is a part of Oznur Ozaltin's Ph.D. thesis and supervised by Ozgur Yeniay.

Author contributions

Oznur Ozaltin idealized this study and analyzed the datasets. Dr. Orhan Coskun checked each image, created datasets classes, and informed about hemorrhage details for the study. Ozgur Yeniay supervised the study and approved the final manuscript. Abdulhamit Subasi also checked all experimental results, supervised the manuscript, and revised the manuscript.

Conflict of Interest

The authors announce no conflicts of interest.

Data Availability Statement

Dataset can be available from https://www.kaggle.com/vbookshelf/computed-tomography-ct-images

ORCID

Oznur Ozaltin https://orcid.org/0000-0001-9841-1702

References

[1] J. Broderick *et al.*, "Guidelines for the Management of Spontaneous Intracerebral Hemorrhage in Adults: 2007 Update: A Guideline from the American Heart Association/American Stroke Association Stroke Council, High Blood Pressure Research Council, and the Quality of Care and Outcomes in Research Interdisciplinary Working Group: The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists," *Stroke,* vol. 38, no. 6, pp. 2001-2023, 2007.

[2] J. Elliott and M. Smith, "The acute management of intracerebral hemorrhage: a clinical review," *Anesthesia & Analgesia,* vol. 110, no. 5, pp. 1419-1427, 2010.

[3] R. F. McCormack and A. Hutson, "Can computed tomography angiography of the brain replace lumbar puncture in the evaluation of acute-onset headache after a negative noncontrast cranial computed tomography scan?," *Academic Emergency Medicine,* vol. 17, no. 4, pp. 444-451, 2010.

[4] P. Parizel, S. Makkat, E. Van Miert, J. Van Goethem, L. Van den Hauwe, and A. De Schepper, "Intracranial hemorrhage: principles of CT and MRI interpretation," *European radiology,* vol. 11, no. 9, pp. 1770-1783, 2001.

[5] A. Gautam and B. Raman, "Towards effective classification of brain hemorrhagic and ischemic stroke using CNN," *Biomedical Signal Processing and Control,* vol. 63, p. 102178, 2021, doi: https://doi.org/10.1016/j.bspc.2020.102178.

[6] W. Kuo, C. Häne, P. Mukherjee, J. Malik, and E. L. Yuh, "Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning," *Proceedings of the National Academy of Sciences,* vol. 116, no. 45, pp. 22737-22745, 2019, doi: https://doi.org/10.1073/pnas.1908021116.

[7] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," *Computer methods and programs in biomedicine,* vol. 138, pp. 49-56, 2017, doi: https://doi.org/10.1016/j.cmpb.2016.10.007.

[8] L. Li *et al.*, "Deep learning for hemorrhagic lesion detection and segmentation on brain ct images," *IEEE Journal of Biomedical and Health Informatics,* vol. 25, no. 5, pp. 1646-1659, 2020, doi: 10.1109/JBHI.2020.3028243.

[9] M. Srikrishna *et al.*, "Deep learning from MRI-derived labels enables automatic brain tissue classification on human brain CT," *Neuroimage,* vol. 244, p. 118606, 2021, doi: https://doi.org/10.1016/j.neuroimage.2021.118606.

[10] A. Shoeibi *et al.*, "Applications of epileptic seizures detection in neuroimaging modalities using deep learning techniques: methods, challenges, and future works," *arXiv preprint arXiv:2105.14278,* 2021.

[11] A. Shoeibi *et al.*, "Detection of epileptic seizures on EEG signals using ANFIS classifier, autoencoders and fuzzy entropies," *Biomedical Signal Processing and Control,* vol. 73, p. 103417, 2022.

[12] A. Shoeibi *et al.*, "Automatic Diagnosis of Schizophrenia in EEG Signals Using CNN-LSTM Models," *Frontiers in Neuroinformatics,* vol. 15, 2021.

[13] D. Sharifrazi *et al.*, "CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering," 2020.

[14] B. Shahangian and H. Pourghassem, "Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure," *Biocybernetics and Biomedical Engineering,* vol. 36, no. 1, pp. 217-232, 2016, doi: https://doi.org/10.1016/j.bbe.2015.12.001.

[15] A. Gautam, B. Raman, and S. Raghuvanshi, "A hybrid approach for the delineation of brain lesion from CT images," *Biocybernetics and Biomedical Engineering,* vol. 38, no. 3, pp. 504-518, 2018, doi: https://doi.org/10.1016/j.bbe.2018.04.003.

[16] H. Bhadauria and M. Dewal, "Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain CT imaging," *Signal, Image and Video Processing,* vol. 8, no. 2, pp. 357-364, 2014, doi: https://doi.org/10.1007/s11760-012-0298-0.

[17] A. I. Maas, C. W. Hukkelhoven, L. F. Marshall, and E. W. Steyerberg, "Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors," *Neurosurgery,* vol. 57, no. 6, pp. 1173-1182, 2005, doi: https://doi.org/10.1227/01.NEU.0000186013.63046.6B.

[18] B. Shahangian and H. Pourghassem, "Automatic brain hemorrhage segmentation and classification in CT scan images," in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2013: IEEE, pp. 467-471, doi: doi: 10.1109/IranianMVIP.2013.6780031.

[19] J.-L. Solorio-Ramírez, M. Saldana-Perez, M. D. Lytras, M.-A. Moreno-Ibarra, and C. Yáñez-Márquez, "Brain Hemorrhage Classification in CT Scan Images Using Minimalist Machine Learning," *Diagnostics,* vol. 11, no. 8, p. 1449, 2021, doi: https://doi.org/10.3390/diagnostics11081449.

[20] U. Balasooriya and M. S. Perera, "Intelligent brain hemorrhage diagnosis using artificial neural networks," in *2012 IEEE Business, Engineering & Industrial Applications Colloquium (BEIAC)*, 2012: IEEE, pp. 128-133, doi: doi: 10.1109/BEIAC.2012.6226036.

[21] D. M. Alawad, A. Mishra, and M. T. Hoque, "AIBH: accurate identification of brain hemorrhage using genetic algorithm based feature selection and stacking," *Machine*

Learning and Knowledge Extraction, vol. 2, no. 2, pp. 56-77, 2020, doi: https://doi.org/10.3390/make2020005.

[22] R. F. Mansour and N. O. Aljehane, "An optimal segmentation with deep learning based inception network model for intracranial hemorrhage diagnosis," *Neural Computing and Applications*, vol. 33, no. 20, pp. 13831-13843, 2021, doi: https://doi.org/10.1007/s00521-021-06020-8.

[23] M. Toğaçar, Z. Cömert, B. Ergen, and Ü. Budak, "Brain Hemorrhage Detection based on Heat Maps, Autoencoder and CNN Architecture," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 2019: IEEE, pp. 1-5, doi: 10.1109/UBMYK48245.2019.8965576.

[24] C. Anupama, M. Sivaram, E. L. Lydia, D. Gupta, and K. Shankar, "Synergic deep learning model–based automated detection and classification of brain intracranial hemorrhage images in wearable networks," *Personal and Ubiquitous Computing*, pp. 1-10, 2020, doi: https://doi.org/10.1007/s00779-020-01492-2.

[25] M. Chawla, S. Sharma, J. Sivaswamy, and L. Kishore, "A method for automatic detection and classification of stroke from brain CT images," in *2009 Annual international conference of the IEEE engineering in medicine and biology society*, 2009: IEEE, pp. 3581-3584, doi: 10.1109/IEMBS.2009.5335289.

[26] A. Majumdar, L. Brattain, B. Telfer, C. Farris, and J. Scalera, "Detecting intracranial hemorrhage with deep learning," in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2018: IEEE, pp. 583-587, doi: doi: 10.1109/EMBC.2018.8512336.

[27] A. Muhammad and W. Guojun, "Segmentation of calcification and brain hemorrhage with midline detection," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 2017: IEEE, pp. 1082-1090, doi: doi: 10.1109/ISPA/IUCC.2017.00164.

[28] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, and P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14290-14299, 2020, doi: doi: 10.1109/JSEN.2020.3023471.

[29] A. Rovlias, S. Theodoropoulos, and D. Papoutsakis, "Chronic subdural hematoma: surgical management and outcome in 986 cases: a classification and regression tree approach," *Surgical neurology international*, vol. 6, 2015, doi: doi: 10.4103/2152-7806.161788.

[30] S.-J. Yeh, S.-C. Tang, L.-K. Tsai, and J.-S. Jeng, "Pathogenetical subtypes of recurrent intracerebral hemorrhage: designations by SMASH-U classification system," *Stroke*, vol. 45, no. 9, pp. 2636-2642, 2014, doi: https://doi.org/10.1161/STROKEAHA.114.005598.

[31] M. S. DİN *et al.*, "Exemplar deep and hand-modeled features based automate and accurate cerebral hemorrhage classification method," *Medical Engineering & Physics*, p. 103819, 2022.

[32] K. Sharada, V. Prashanthi, and S. Kanakala, "Detection and Classification of Intracranial Brain Hemorrhage," in *IoT and Analytics for Sensor Networks*: Springer, 2022, pp. 455-464.

[33] K. Uyar, Ş. Taşdemir, E. Ülker, M. Öztürk, and H. Kasap, "Multi-Class brain normality and abnormality diagnosis using modified Faster R-CNN," *International journal of medical informatics*, vol. 155, p. 104576, 2021.

[34] Ö. F. Ertuğrul and M. F. Akıl, "Detecting hemorrhage types and bounding box of hemorrhage by deep learning," *Biomedical Signal Processing and Control,* vol. 71, p. 103085, 2022.

[35] V. Abramova *et al.*, "Hemorrhagic stroke lesion segmentation using a 3D U-Net with squeeze-and-excitation blocks," *Computerized Medical Imaging and Graphics,* vol. 90, p. 101908, 2021.

[36] S. Santhoshkumar, V. Varadarajan, S. Gavaskar, J. J. Amalraj, and A. Sumathi, "Machine learning model for intracranial hemorrhage diagnosis and classification," *Electronics,* vol. 10, no. 21, p. 2574, 2021.

[37] S. Barin, M. Saribaş, B. G. Çiltaş, G. E. Güraksin, and K. Utku, "Hybrid Convolutional Neural Network-Based Diagnosis System for Intracranial Hemorrhage," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience,* vol. 12, no. 4, pp. 01-27, 2021.

[38] M. Burduja, R. T. Ionescu, and N. Verga, "Accurate and efficient intracranial hemorrhage detection and subtype classification in 3D CT scans with convolutional and long short-term memory neural networks," *Sensors,* vol. 20, no. 19, p. 5611, 2020.

[39] H. Salehinejad *et al.*, "A real-world demonstration of machine learning generalizability in the detection of intracranial hemorrhage on head computerized tomography," *Scientific reports,* vol. 11, no. 1, pp. 1-11, 2021.

[40] V. Pandimurugan *et al.*, "Detecting and Extracting Brain Hemorrhages from CT Images Using Generative Convolutional Imaging Scheme," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.

[41] K. N. Iqbal, I. Azad, M. Emon, I. Haque, N. S. Amlan, and A. A. Aporna, "Brain hemorrhage detection using hybrid machine learning algorithm," Brac University, 2022.

[42] M. Hssayeni. Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation (version 1.3.0). [Online] Available: https://physionet.org/content/ct-ich/1.2.0/

[43] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online].

[44] M. D. Hssayeni, Croock, M. S., Al-Ani, A., Al-khafaji, H. F., Yahya, Z. A., & Ghoraani, B. , "Intracranial Hemorrhage Segmentation Using Deep Convolutional Model. ," *arXiv preprint* vol. arXiv:1910.08643., 2019. [Online]. Available: https://arxiv.org/abs/1910.08643.

[45] M. Hssayeni. https://www.kaggle.com/vbookshelf/computed-tomography-ct-images (accessed 19.11.2021, 18.55

[46] M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial hemorrhage segmentation using a deep convolutional model," *Data,* vol. 5, no. 1, p. 14, 2020, doi: https://doi.org/10.3390/data5010014.

[47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[49] Y. Ioannou, D. Robertson, J. Shotton, R. Cipolla, and A. Criminisi, "Training cnns with low-rank filters for efficient image classification," *arXiv preprint arXiv:1511.06744,* 2015.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[51] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[52] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848-6856.
[53] Q. Shang, D. Tan, S. Gao, and L. Feng, "A hybrid method for traffic incident duration prediction using BOA-optimized random forest combined with neighborhood components analysis," *Journal of Advanced Transportation*, vol. 2019, 2019, doi: https://doi.org/10.1155/2019/4202735.
[54] M. Jin and W. Deng, "Predication of different stages of Alzheimer's disease using neighborhood component analysis and ensemble decision tree," *Journal of neuroscience methods*, vol. 302, pp. 35-41, 2018, doi: https://doi.org/10.1016/j.jneumeth.2018.02.014.
[55] S. Raghu and N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms," *Expert Systems with Applications*, vol. 113, pp. 18-32, 2018, doi: https://doi.org/10.1016/j.eswa.2018.06.031.
[56] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," *Advances in neural information processing systems*, vol. 17, 2004.
[57] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161-168, 2012, doi: 10.4304/jcp.7.1.161-168.
[58] H. Jiang, *Machine Learning Fundamentals: A Concise Introduction*. Cambridge University Press, 2021.
[59] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.
[60] A. Subasi, M. Balfaqih, Z. Balfagih, and K. Alfawwaz, "A Comparative Evaluation of Ensemble Classifiers for Malicious Webpage Detection," *Procedia Computer Science*, vol. 194, pp. 272-279, 2021, doi: https://doi.org/10.1016/j.procs.2021.10.082.
[61] P. Cichosz, *Data mining algorithms: explained using R*. John Wiley & Sons, 2014.
[62] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580-585, 1985.
[63] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21-27, 1967.
[64] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
[65] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," in *Robust data mining*: Springer, 2013, pp. 27-33.
[66] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273-297, 1995, doi: https://doi.org/10.1007/BF00994018.
[67] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (covid-19) classification using ct images by machine learning methods," *arXiv preprint arXiv:2003.09424*, 2020.
[68] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Computers and Electronics in Agriculture*, vol. 174, p. 105507, 2020, doi: https://doi.org/10.1016/j.compag.2020.105507.

[69] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *DATA MINING: Practical machine learning tools and techniques*, Fourth ed. (DATA MINING). Morgan Kaufmann, 2017, p. 601.
[70] C. C. Aggarwal, *Data mining: the textbook*. Springer, 2015.
[71] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics surveys*, vol. 4, pp. 40-79, 2010.
[72] D. Sharifrazi *et al.*, "Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images," *Biomedical Signal Processing and Control*, vol. 68, p. 102622, 2021.
[73] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185-205, 2005.
[74] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1315-1321, 1999.
[75] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University-Computer and Information Sciences*, vol. 29, no. 4, pp. 462-472, 2017.
[76] Ö. Özaltın and Ö. Yeniay, "Ecg classification performing feature extraction automatically using a hybrid cnn-svm algorithm," in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2021: IEEE, pp. 1-5, doi: 10.1109/HORA52670.2021.9461295.
[77] A. M. Dawud, K. Yurtkan, and H. Oztoprak, "Application of deep learning in neuroradiology: brain haemorrhage classification using transfer learning," *Computational Intelligence and Neuroscience*, vol. 2019, 2019.
[78] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018: IEEE, pp. 281-284.
[79] T. D. Phong *et al.*, "Brain hemorrhage diagnosis by using deep learning," in *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*, 2017, pp. 34-39.