Diagnosis of MS by processing brain MRI images

Author: Zainab Ejazi, Ph.D., Islamic Azad University, South Tehran Branch
Mehdi Eslami, Ph.D., Islamic Azad University, Science and Research Unit, Tehran

Abstract
MS is a disease that affects the central nervous system. As a result of this disease, multiple areas of scar tissue called sclerosis may appear. MRI imaging is a complementary initial examination for the monitoring and diagnosis of MS. Timely detection of the disease can slow down or stop its progress. Today, various methods have been identified to predict the disease. In this study, a machine learning method is proposed to identify the disease with high accuracy. For this purpose, image processing tools such as wavelet transform and GLCM matrix have been used. Then, various characteristics such as energy, contrast, correlation coefficient, entropy and homogeneity of the image have been extracted from the output of these tools. These features were classified by three classifiers: SVM, KNN and decision tree. The method of the reference article was a hybrid method that examined both machine learning methods and deep learning methods. The results of the proposed method showed a high accuracy of 97%, which was an improvement of about 2% compared to the reference article.

Keywords: MS, Violet transform, GLCM matrix, SVM classifier, KNN classifier, decision tree classifier

.١Introduction:

MS or multiple sclerosis, which means multiple sclerosis or multiple wounds, is a chronic and progressive autoimmune disease that affects the central nervous system [1]. MS occurs when the immune system attacks myelin. Myelin protects nerve fibers in the brain and spinal cord. This event is known as demyelination and causes communication problems between the nerves and the brain [2]; It means that the myelin covering the nerve is destroyed. Finally, it can cause nerve damage. This disease can cause different symptoms for the patient. In most cases, at the beginning, the complications of MS occur and disappear [3,4]. Over time, some of these symptoms persist and can lead to disability. Although there is no specific treatment for this disease, drugs and treatment of MS disease in different ways can reduce the number of recurrences of the disease and related symptoms and disabilities [5]. There is no cure for MS, but timely diagnosis and control of its symptoms reduce the speed of disease progression [6]. Therefore, timely diagnosis of this disease is very important. Due to the lack of specialists in deprived areas, many patients may not be aware of the disease in the early and controllable stages. This issue makes it important to have a system for timely disease diagnosis.[٧]

MS diagnosis by MRI imaging is the best method to help diagnose this disease [8]. This method can be used repeatedly to track the progress of MS disease, determine disease status and drug performance, although the doctor may not use this method widely. In MRI, a powerful magnetic field and computer-generated radio waves are used to measure the amount of water in areas of the body such as tissues, nerves, organs, and bones, and it is the most non-invasive and sensitive imaging method available. It is from the brain, spinal cord and other parts of the body.[٩]

Various works have been done to identify this disease. Tadayon et al. [10] in 2016 introduced a fuzzy KNN algorithm (F-KNN) to classify MS lesions into three types. This classifier combines features obtained from contrast-weighted FLAIR, 1T-w, 2T-w, and 1T-w gadolinium-enhanced images (1Gad-E-T-w) and additional features extracted from DT-MR images (Publication of indices) including fractional anisotropy of FA and average publication of MD has been done.

Fuladi et al. [11] in 2018 proposed a three-class classification process based on ANN from different samples with relapsing-remitting MS. The input features of ANN algorithm, which included three versions of RBF, MLP, and ENN, depended on Akaike's Information Criterion (ENN-AIC) as average ideals of TI and 1QMT were extracted from parametric maps. The results showed that the technique based on ENN-AIC achieved 86% accuracy, 92% sensitivity and 90% accuracy, which was better than other different ANN techniques they investigated.

In 2020, Jain and co-workers [12] proposed a classification technique based on ensemble learning to identify multiple sclerosis diseases from a database of healthy and unhealthy brain magnetic resonance (MR) images. Feature extraction from brain MR images has been performed using eighteen different features based on gray level co-occurrence matrix (GLCoM). Then, decision tree-based ensemble learning is performed on these features using three different augmentation techniques for MR image classification. The results showed that the group learning technique has a high accuracy of 94.91%.Acar et al. [13] in 2022, developed a CNN model for MS diagnosis through lesion detection in brain FLAIR MRI. The dataset used consists of brain MRI, brain mask and ground truth data of 30 MS patients obtained from the Laboratory of Imaging Technologies (LabIT). Features of MS lesions in MRI are extracted with a small set of trainable parameters. Results were generated by dividing the data at the slice level as well as at the patient level. By using division in the cutting surface, the proposed model reached 98.0% ± 0.02% accuracy, 97.9% ± 0.03% sensitivity, 98.3% ± 0.03% specificity, and 98.2% ± 0.03% accuracy. By using division at the patient level, the proposed model reached accuracy, sensitivity, specificity and accuracy of 90.3 ± 0.05, 90.5 ± 0.05, 90.1 ± 0.09 percent and 91.1 ± 0.09 percent, respectively.

In this project, our goal is to diagnose MS with brain MRI images. For this purpose, wavelet transform is first applied to the image. Then the co-occurrence matrix (GLCM) is calculated from the images. In the next step, homogeneity, contrast, entropy and energy features are extracted from the image wavelet transform. Then these features are given to the neural network and the images with lesions are detected.

In the reference article [13], the images are first placed in a pre-processing process and the skull is removed from the images. Then, with the help of histogram of directional gradients, the feature is extracted from the image. Then the extracted features are reduced with the help of PCA method. Finally, classification is done with the help of 3 decision tree classifiers, support vector machine and K nearest neighbor. In addition, a deep learning-based approach has also been implemented with CNN networks. Figure 1 shows the block diagram of the proposed method:

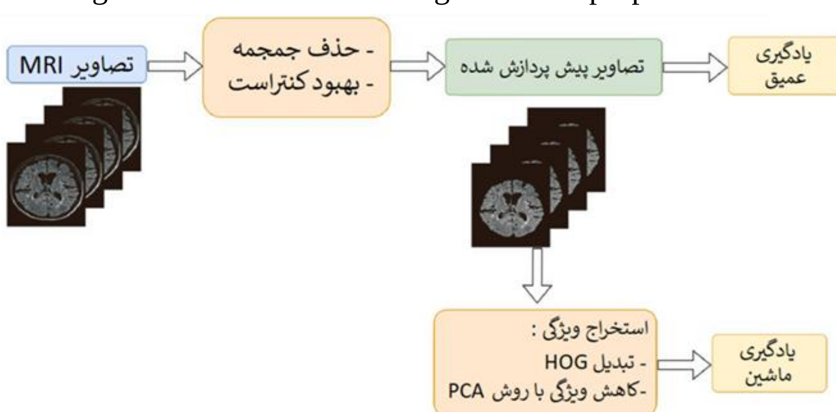


Figure 1: General block diagram of the reference article method

۲- Materials and methods

۲-۱- Pre-processing

In the pre-processing stage, skull removal and histogram stretching were first applied to all MRI data to improve contrast. Removing the skull with the help of opening morphological operations, the skull is removed with the help of a circular mask with a radius of 5 pixels. The result of this process is shown in Figure 2. After removing the skull, it is time to correct the histogram to improve the contrast. Since the histogram distribution of the images is different from each other, the process of histogram conjugation is shown in Figure 3 as a result of contrast improvement.

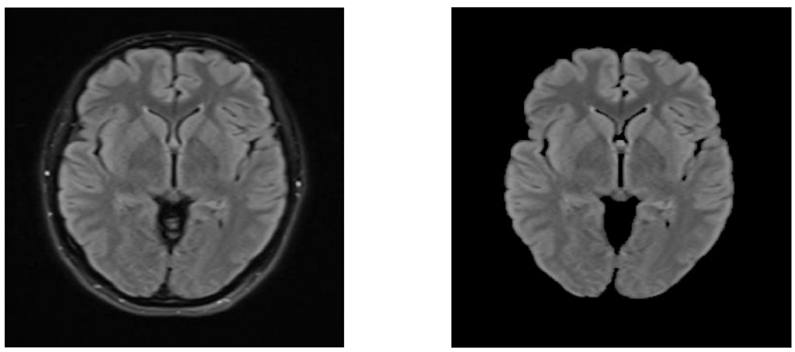


Figure 2: Result of skull removal from brain MRI images. The left side of the original image and the right side of the image after removing the skull

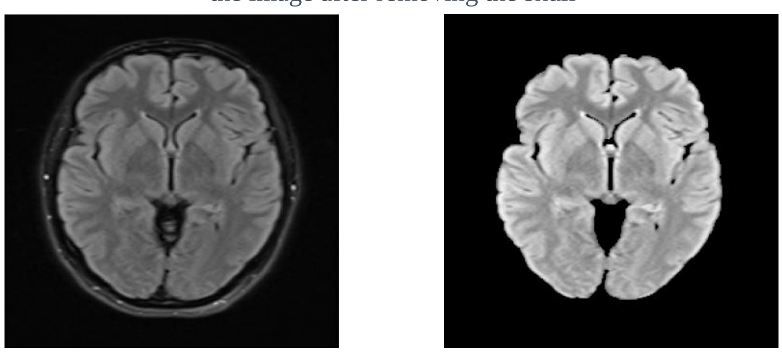


Figure 3: Contrast enhancement result of brain MRI image. The left side of the original image and the right side of the image after contrast improvement

۲-۲- Image processing and feature extraction

As stated at the beginning of the chapter, this study consists of two parts. In the first part, after pre-processing, the images are entered into pre-trained convolutional networks to extract features from the images using the transfer learning technique. In the second part, the images are extracted from the feature images with the help of the directional gradient histogram algorithm.

۲-۲-۱- feature extraction in the proposed method

Wavelet transformation is one of the important mathematical transformations that is used in various fields of science. The main idea of wavelet transform is to overcome the weaknesses and limitations of Fourier transform. Unlike the Fourier transform, this transformation can be used for non-stationary signals and dynamic systems [14]. The signal passes through two filters, high-pass filters and low-pass filters. Then the image is decomposed into high frequency (detail) and low frequency (approximation) components. At each level, we receive 4 sub-signals. Approximation shows the overall trend of pixel values and details as horizontal, vertical and diagonal components. If these details are insignificant, they can be evaluated as zero without significantly affecting the image, thus achieving filtering and compression [15]. The result of applying wavelet transformation on one of the database images can be seen in Figure 3:

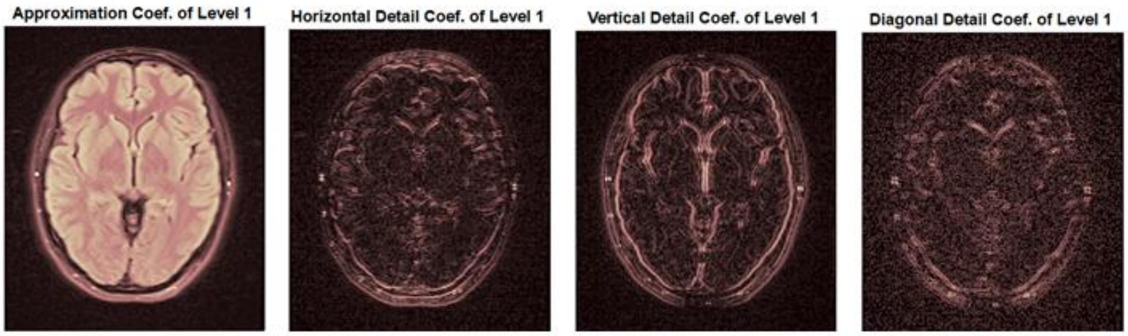


Figure 4: Sample images after applying wavelet transformation

After applying the wavelet transformation, it is time to use the co-occurrence matrices of the gray level. The theory of using texture feature with the help of gray level co-occurrence matrix or abbreviated GLCM was presented by Haralick in 1973 [16]. Gray level per-occurrence matrix is a statistical method for examining texture features that considers the spatial relationship of pixels. GLCM is also known as gray spatial dependence matrix. GLCM functions create the texture of an image by calculating the displacement of pixel pairs with specific values and in a specific spatial relationship in an image. Other filters cannot give us information about the placement of different pixels in an image. Calculating the GLCM of the image and then extracting statistical criteria from this matrix can determine features for image classification or histology that are well distinguishable [17]. To create GLCM, graycomatrix function is used. This function creates a gray level co-occurrence matrix (GLCM) by calculating how many pixels with luminance value (gray level) i are in a specific spatial relationship with a pixel with luminance value (gray level) j . By default, spatial relationships are defined as the pixel of interest and the pixel to its right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i, j) in the glcm matrix is simply the sum of the number of times that the pixel with value i occurs in the specified spatial relationship to the pixel with value j in the input image. After calculating these two transformations, it is time to extract features. Feature extraction is a process in which, by performing an operation on the data, its salient and defining features are identified. The purpose of feature extraction is to transform the raw data into a more usable form for further statistical processing. The statistical features obtained from GLCM are computable as a structure with fields specified by the features. In this project, we have 4 categories of statistical features that can be deduced from this matrix.

Contrast is the difference in color brightness or contrast in the pixels of the image, which makes them (or their image) different from each other. The French word contrast means difference and separation and is one of the basic principles of various fields of art. To calculate the contrast, we first need to calculate the GLCM matrix. After calculating this matrix, the contrast is calculated from equation 1:[۱۸]

$$Contrast = \sum_{i,j} |i-j|^2 \, C(i,j)$$ Relationship 1

where variable C is the GLCM matrix of the image. Considering that we used various offsets to calculate the GLCM matrix, a value for contrast will be obtained for each matrix.

Correlation coefficient is a statistical tool to determine the type and degree of relationship of one quantitative variable with another quantitative variable. The correlation coefficient is one of the criteria used to determine the correlation between two variables. The correlation coefficient shows the intensity of the relationship as well as the type of relationship (direct or inverse). This coefficient is between 1 and -1, and if there is no relationship between two variables, it is equal to zero [18]. The correlation between two random variables X and Y is defined as the following relationship:

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$ Relationship 2

where E is the mathematical expectation operator, cov is the covariance, corr is the usual symbol for Pearson correlation, and sigma is the standard deviation. Energy and power have different definitions than the definition of energy and power in physics. In physics, energy is equal to work and power is defined as work per unit of time. But in the science of image processing, the energy and power of the image are considered without physical units; Because image pixels may represent different physical quantities. It can be said that they obtain the energy and power of the signal based on the size of the signal [18]. To calculate the energy of the image from the GLCM matrix, it is enough to raise all the values of this matrix to the second power and then add them together:

$$Energy = \sum_{i,j} |C(i,j)|^2$$ Relationship 3

The distribution of different brightness values in an image can be determined by a homogeneity factor. In general, an image is homogeneous if every point (pixel) in the image has the same color. If there are strong contrasts in an image, it is heterogeneous. Statistically, in the simplest case, the standard deviation of each pixel from the average gray value can be calculated. If the variance of this standard deviation is high, the homogeneity is low [18]. To calculate this parameter, we use the following relationship:

$$Homogeneity = \sum_{i,j} \frac{C(i,j)}{1+|i-j|}$$ Relationship 4

۲-۲-۲- Feature extraction in the reference article

Histogram of Oriented Gradients (HOG) is a feature descriptor used in computer vision and image processing for object recognition. This technique counts gradient orientation in local parts of an image. This method is similar to edge orientation histograms, scale-invariant feature transform descriptors, and shape fields, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization to improve accuracy. .[۱۹]

Dalal and Triggs investigated four different methods for block normalization. If v is the non-normalized vector containing all the histograms in a given block, its k-norm‖v‖_k is k=1,2 and e is a small constant value. Then the normalization coefficient can be one of the following:[۲۰]

$$L2-norm: \quad f=\frac{v}{\sqrt{\|v\|_2^2+e^2}}$$

Relationship 2

$$L1-norm: \quad f=\frac{v}{(\|v\|_1+e)}$$

$$L1-sqrt: \quad f=\sqrt{\frac{v}{(\|v\|_1+e)}}$$

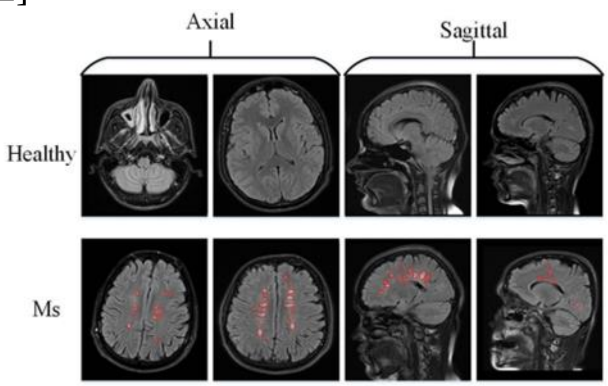
In their experiments, Dalal and Triggs found that L2-norm and L1-sqrt schemes provide similar performance, while L1-norm provides slightly less reliable performance. However, all three methods showed a very significant improvement over non-normal data. In this study, L2-norm is used for feature extraction.

After extracting the feature in this method, it is time to reduce the feature with the PCA method. This method is the main linear method to reduce dimensions; This method performs a linear mapping of the data to a space with a lower dimension, so that the variance explanation of the original data is maximized in the transferred data (to lower dimensions). In practice, the variance (and sometimes covariance) matrix of the data is constructed and the eigenvector of this matrix is calculated. The eigenvectors corresponding to the largest eigenvalues contain the most information from the original data and can now be used to reconstruct a large portion of the variance of the original data. Roughly, the first several vectors can be interpreted as representing the behavior of the big data. The data is constructed in lower dimensions using these principal vectors with some loss of information (hopefully retaining variance explanatory power). After reducing the features, we increase the number of features to 10.

The next step is classification. Classification is the problem of identifying the belonging of a new observation to which of the set of categories (sub-populations), based on a set of data used for training, including observations whose category membership is known [21]. In machine learning terms, classification is a form of supervised learning, where a set of data is available for training. In this study, 3 classes of support vector machine, k nearest neighbor and decision tree are used. The working basis of the SVM classifier is the linear classification of the data, and in the linear division of the data, we try to choose a hyperplane that has a higher confidence margin. K Nearest Neighbor or KNN algorithm is one of the simplest machine learning algorithms. This algorithm finds the k nearest neighbor for classification problems and predicts the nearest neighbors of the class with the majority of votes. Decision trees belong to the group of supervised learning algorithms and most of them are built based on quantitative minimization called entropy.

3-Total data

The data set of the study included axial and sagittal FLAIR MRI images of the brain obtained from 72 patients with MS and 59 healthy male and female nondiseased individuals who attended Ozal University Medical School in 2021. Institutional ethics committee approved this study. Medical experts read the sections of the FLAIR image. From 72 MS patients, 1441 axial and sagittal brain image sections containing identifiable MS lesions were assigned to MS class. From 59 healthy people, 2016 axial and sagittal image slices with normal appearance, i.e. without, were assigned to the healthy class. A sample of images in the database is shown in Figure 5 [22]



[۲۲] Figure 5: Examples of used database images

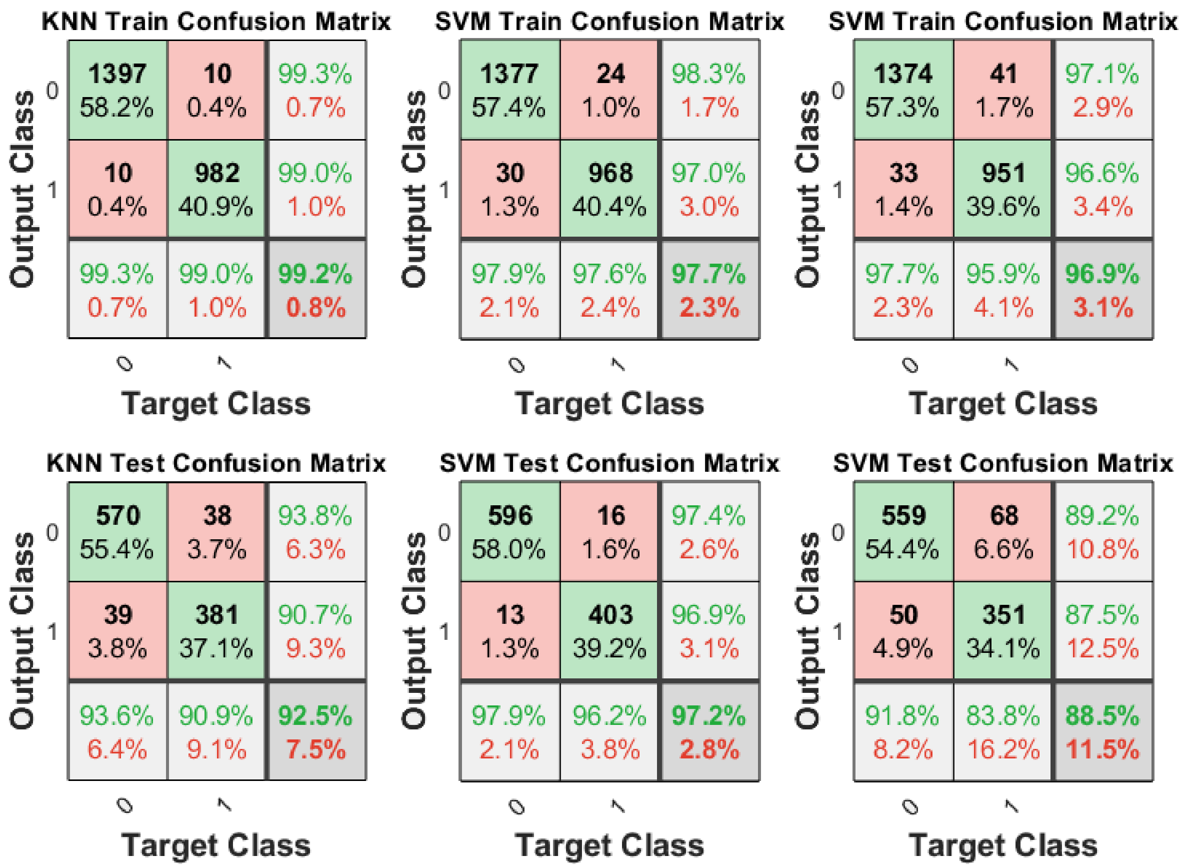
4-Results

In the discussion of machine learning, it is customary to study and build algorithms that can learn and prevent based on data sets. These algorithms work through prediction or data-driven decision-making by building a mathematical model based on input data. The data used to build the final model is usually prepared from multiple datasets. In particular, 2 datasets are usually used in different stages of model building. The model is initially built on a training dataset, to provide the ability to fit the model parameters using a set of examples. A test dataset is a dataset used to provide an unbiased evaluation of the final model fit to the training dataset. In this project, the data were divided into 2 categories: training data and test data, and the stated evaluation criteria were calculated for each of the above two modes, as well as the total data to be evaluated. Table 1 shows the classification results. Using the proposed method, it is reported:

Table 1: Classification results with the proposed method

	Train			Test		
	KNN	SVM	DT	KNN	SVM	DT
Accuracy	99.166	97.749	96.915	92.51	97.179	88.521
Sensitivity	99.289	97.868	97.655	93.596	97.865	91.79
Specificity	98.992	97.581	95.867	90.931	96.181	83.771
Percision	99.289	98.287	97.102	93.75	97.386	89.155
F1-Score	99.289	98.077	97.378	93.673	97.625	90.453

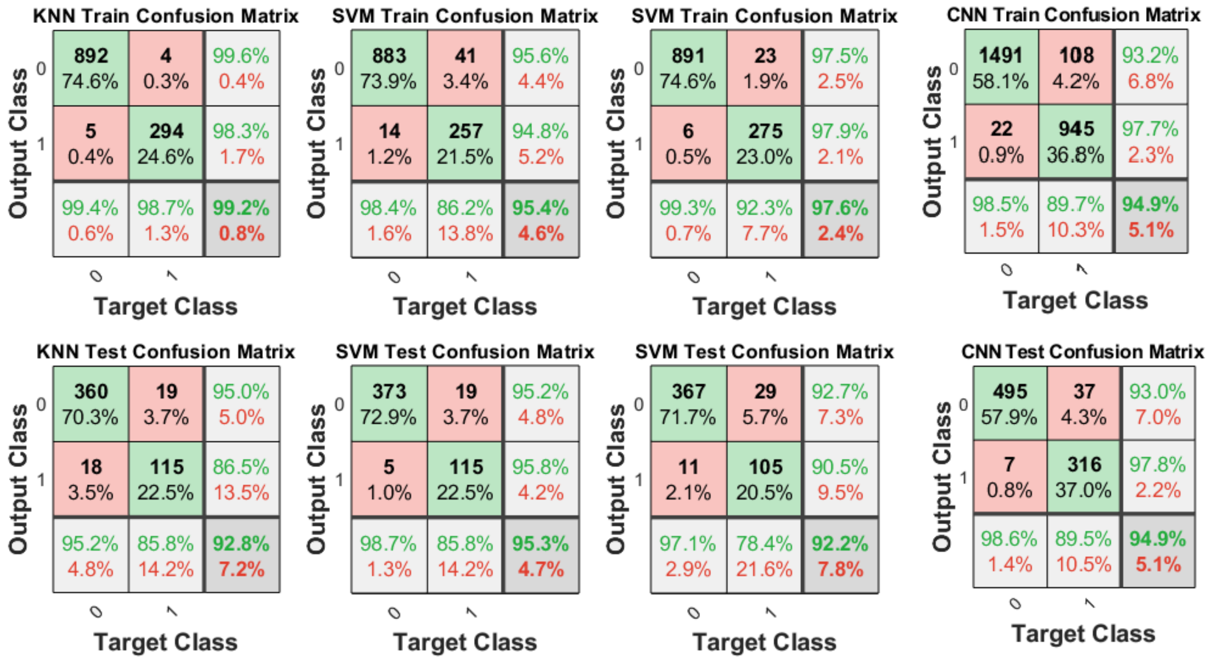
Figure 6: Clutter matrix of the proposed method



[13] Table 2: Classification results with the second method for comparison with the reference article

	Train				Test			
	KNN	SVM	DT	CNN	KNN	SVM	DT	CNN
Accuracy	99.247	95.397	97.573	94.934	92.773	95.313	92.188	94.854
Sensitivity	99.443	98.439	99.331	98.546	95.238	98.677	97.09	98.606
Specificity	98.658	86.242	92.282	89.744	85.821	85.821	78.358	89.518
Percision	99.554	95.563	97.484	93.246	94.987	95.153	92.677	93.045
F1-Score	99.498	96.98	98.399	95.823	95.112	96.883	94.832	95.745

Figure 7: Clutter matrix of the reference article method



5 -Summary

In this study, we tried to find an accurate classification based on these images and the labels assigned by the expert doctor. We used brain MRI images based on data from the Kaggle collection. In the continuation of this project, first a pre-processing stage is applied to remove the skull and improve the contrast. Then, with the help of Violet transformation and GLCM matrix, we extracted features from the images. 4 characteristics of energy, contrast, homogeneity and correlation coefficient were extracted for this purpose. After feature extraction, these features are classified using SVM, KNN and neighborhood tree classifiers. To compare with the reference article, the second method was also applied. In this method, after pre-processing, with the help of directional gradient histogram transformation, features are extracted from the image and then optimal features are selected with the help of PCA method.

The results show that the proposed method has achieved the highest performance with 97.2% accuracy in the test phase with SVM network. On the other hand, using the 10-fold validation algorithm in the proposed algorithm shows the lowest error compared to using networks independently.

Reference

[1] Klineova, Sylvia, and Fred D. Lublin. "Clinical course of multiple sclerosis." Cold Spring Harbor perspectives in medicine 8.9 (2018): a028928.

[2] Filippi, Massimo, et al. "Pediatric Multiple Sclerosis." White Matter Diseases: An Update for Neurologists (2020): 37-66.

[3] Thrue, C., et al. "Time matters: early-phase multiple sclerosis is accompanied by considerable impairments across multiple domains." Multiple Sclerosis Journal 27.10 (2021): 1477-1485.

[4] Baroncini, Damiano, et al. "Risk of persistent disability in patients with pediatric-onset multiple sclerosis." JAMA neurology 78.6 (2021): 726-735.

[5] Simonsen, Cecilia Smith, et al. "Early high efficacy treatment in multiple sclerosis is the best predictor of future disease activity over 1 and 2 years in a Norwegian population-based registry." Frontiers in Neurology 12 (2021): 693017.

[6] Jain S, Rajpal N, Yadav J. Multiple Sclerosis Identification Based on Ensemble Machine Learning Technique. InProceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020) 2020 Nov 21.

[7] Macin G, Tasci B, Tasci I, Faust O, Barua PD, Dogan S, Tuncer T, Tan RS, Acharya UR. An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: ExMPLPQ. Applied Sciences. 2022 May 12;12(10):4920.

[8] Caverly, Robert H. "MRI fundamentals: RF aspects of magnetic resonance imaging (MRI)." IEEE Microwave Magazine 16.6 (2015): 20-33.

[9] Khashami, Fatemeh. Fundamentals of NMR and MRI: From Quantum Principles to Medical Applications. Springer Nature, 2023.

[10] Tadayon E, Khayati RM, Karami V, Nabavi SM. A novel method for automatic classification of multiple sclerosis lesion subtypes using diffusion tensor MR images. Biomedical Engineering: Applications, Basis and Communications. 2016 Oct 31;28(05):1650038.

[11] Fooladi M, Sharini H, Masjoodi S, Khodamoradi E. A novel classification method using effective neural network and quantitative magnetization transfer imaging of brain white matter in relapsing remitting multiple sclerosis. Journal of biomedical physics & engineering. 2018 Dec;8(4):409.

[12] Jain S, Rajpal N, Yadav J. Multiple Sclerosis Identification Based on Ensemble Machine Learning Technique. InProceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020) 2020 Nov 21.

[13] Acar ZY, Başçiftçi F, Ekmekci AH. A Convolutional Neural Network model for identifying Multiple Sclerosis on brain FLAIR MRI. Sustainable Computing: Informatics and Systems. 2022 Sep 1;35:100706.

[14] Mourad, Talbi. "Wavelets and wavelet transforms." ECG Denoising Based on Total Variation Denoising and Wavelets. Cham: Springer International Publishing, 2023. 1-18.

[15] Wang, Jizeng, Xiaojing Liu, and Youhe Zhou. "Application of wavelet methods in computational physics." Annalen der Physik 536.5 (2024): 2300461.

[16] Zubair, Abdul Rasak, and Oluwaseun Adewunmi Alo. "Grey level co-occurrence matrix (GLCM) based second order statistics for image texture analysis." arXiv preprint arXiv:2403.04038 (2024).

[17] Pantic, Igor, et al. "Gray level Co-occurrence matrix, fractal and wavelet analyses of discrete changes in cell nuclear structure following osmotic stress: Focus on machine learning methods." Fractal and Fractional 7.3 (2023): 272.

[18] Yunianto, Mohtar, et al. "Gray Level Co-Occurrence Matrices and Support Vector Machine for Improved Lung Cancer Detection." International Journal of Online & Biomedical Engineering 19.5 (2023).

[19] Patel, Chirag I., et al. "Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences." Sensors 20.24 (2020): 7299.

[20] Zhou, Wei, et al. "Histogram of oriented gradients feature extraction from raw bayer pattern images." IEEE Transactions on Circuits and Systems II: Express Briefs 67.5 (2020): 946-950.

[21] Uddin, Md Palash, Md Al Mamun, and Md Ali Hossain. "PCA-based feature reduction for hyperspectral remote sensing image classification." IETE Technical Review 38.4 (2021): 377-396.

[22] Macin, Gulay, et al. "An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: ExMPLPQ." Applied Sciences 12.10 (2022): 4920.