ORIGINAL ARTICLE

Wilson disease tissue classification and characterization using seven artificial intelligence models embedded with 3D optimization paradigm on a weak training brain magnetic resonance imaging datasets: a supercomputer application

Mohit Agarwal[1] · Luca Saba[2] · Suneet K. Gupta[1] · Amer M. Johri[3] · Narendra N. Khanna[4] · Sophie Mavrogeni[5] · John R. Laird[6] · Gyan Pareek[7] · Martin Miner[8] · Petros P. Sfikakis[9] · Athanasios Protogerou[10] · Aditya M. Sharma[11] · Vijay Viswanathan[12] · George D. Kitas[13] · Andrew Nicolaides[14] · Jasjit S. Suri[15]

Abstract

Wilson's disease (WD) is caused by copper accumulation in the brain and liver, and if not treated early, can lead to severe disability and death. WD has shown white matter hyperintensity (WMH) in the brain magnetic resonance scans (MRI) scans, but the diagnosis is challenging due to (i) subtle intensity changes and (ii) weak training MRI when using artificial intelligence (AI). Design and validate seven types of high-performing AI-based computer-aided design (CADx) systems consisting of 3D optimized classification, and characterization of WD against controls. We propose a "conventional deep convolution neural network" (cDCNN) and an "improved DCNN" (iDCNN) where rectified linear unit (ReLU) activation function was modified ensuring "differentiable at zero." Three-dimensional optimization was achieved by recording accuracy while changing the CNN layers and augmentation by several folds. WD was characterized using (i) CNN-based feature map strength and (ii) Bispectrum strengths of pixels having higher probabilities of WD. We further computed the (a) area under the curve (AUC), (b) diagnostic odds ratio (DOR), (c) reliability, and (d) stability and (e) benchmarking. Optimal results were achieved using 9 layers of CNN, with 4-fold augmentation. iDCNN yields superior performance compared to cDCNN with accuracy and AUC of 98.28 ± 1.55, 0.99 ($p < 0.0001$), and $97.19 \pm 2.53\%$, 0.984 ($p < 0.0001$), respectively. DOR of iDCNN outperformed cDCNN fourfold. iDCNN also outperformed (a) transfer learning–based "Inception V3" paradigm by 11.92% and (b) four types of "conventional machine learning–based systems": k-NN, decision tree, support vector machine, and random forest by 55.13%, 28.36%, 15.35%, and 14.11%, respectively. The AI-based systems can potentially be useful in the early WD diagnosis.

✉ Jasjit S. Suri
Jasjit.Suri@AtheroPoint.com

[1] CSE Department, Bennett University, Greater Noida, UP, India

[2] Department of Radiology, Azienda Ospedaliero Universitaria (A.O.U.), Cagliari, Italy

[3] Department of Medicine, Division of Cardiology, Queen's University, Ontario, Kingston, Canada

[4] Department of Cardiology, Indraprastha APOLLO Hospitals, New Delhi, India

[5] Cardiology Clinic, Onassis Cardiac Surgery Center, Athens, Greece

[6] Heart and Vascular Institute, Adventist Health St. Helena, St. Helena, CA, USA

[7] Minimally Invasive Urology Institute, Brown University, Providence, RI, USA

[8] Men's Health Center, Miriam Hospital Providence, Providence, RI, USA

[9] Rheumatology Unit, National Kapodistrian University of Athens, Athens, Greece

[10] Department of Cardiovascular Prevention, National and Kapodistrian Univ. of Athens, Athens, Greece

[11] Division of Cardiovascular Medicine, University of Virginia, Charlottesville, VA, USA

[12] MV Hospital for Diabetes & Professor M Viswanathan Diabetes Research Centre, Chennai, India

[13] R & D Academic Affairs, Dudley Group NHS Foundation Trust, Dudley, UK

[14] Vascular Screening and Diagnostic Centre, University of Nicosia, Nicosia, Cyprus

[15] Stroke Monitoring and Diagnostic Division, AtheroPoint™, Roseville, CA 95661, USA

1 Introduction

Wilson's disease (WD) is due to excessive copper accumulation in the liver and brain [1]. The National Organizations for Rare Disorders (NORD) reported that 1 in every 30,000 to 40,000 people in the world are affected by WD [1]. It is estimated that there were nearly 600 cases of WD in the USA in the year 2007[1] and there will be 9000 people affected by WD in the USA by the end of 2020.

WD causes severe disability and death, if not treated early. The present diagnosis of WD uses anatomical tests, but they are not reliable [2, 3]. MRI has shown promising signs for diagnosing WD since it shows white matter hyperintensity (WMH) in the brain [4–7]. However, due to the volumetric nature of the MRI and subtle nature of the hyperintensity between WD and controls, human bias and interobserver variability may complicate diagnosis. To overcome these challenges, computer-aided diagnosis (CADx) methods can play a vital role in improving the classification and characterization of WD [8, 9].

Artificial intelligence (AI) is a branch of computer science that can handle classification effectively as it can map nonlinearity between input variations and disease severity [10]. AI methods can be broadly divided into two supervised learning categories namely machine learning (ML) and deep learning (DL). Machine learning methods [11–16] like decision tree (DT), k-nearest neighbor (k-NN), support vector machine (SVM), and random forest (RF) can be applied for classification, but they use manually identified features and can yield low performance. In contrast, deep learning (DL) [17–20] methods are more reliable because they can automatically learn features using hidden layers within a dataset. One such example is a deep convolution neural network (DCNN) that has been well-adopted by industry for image classification [11, 19].

Furthermore, DL systems can augment the size of the input data to ensure a balance between classes leading to stronger learning protocols.

The DL systems are having several parameters, in particular the number of layers in the DCNN architecture. These layers are typically (i) convolution layers, (ii) max-pooling layers, and the (iii) combination of dense layers that constitutes the neural network, along with the softmax layer. Since the low-level and high-level MRI features of the WD are extracted by the convolution and max-pooling set of layers, it is, therefore, important to define as to how many such sets are needed for the optimization of the DCNN system [21]. Furthermore, since DCNN is sensitive to the training data size,

it is customary to understand what should be the total size of the input training data. This training data size can be altered by the "augmentation procedure" well-developed in the AI industry [22]. For the best combination of "augmentation folds" and the "number of layers" of the DCNN, we, therefore, need to optimize the DL system for best classification and characterization of the WD. Thus, there is a direct bearing between the WD classification and (a) number of DCNN layers and (b) augmentation folds of the training data size. Such an optimization paradigm is being attempted the first time for the WD application. Furthermore, because conventional DCNN (cDCNN) uses the rectified linear unit (ReLU) as an activation function that is not continuously differentiable at the origin, we have designed an improved DCNN (iDCNN) which is smooth near the origin thereby improving its performance. Finally, we, validate the hypothesis that there is WMH in WD MRI images. The system uses two novel characterization strategies by taking advantage of the CNN layers and the Bispectrum signal processing framework [12, 19, 23]. This is another unique feature of our paradigm. Furthermore, we benchmark our two DCNN systems against the transfer learning–based "Inception V3" [24] framework and four types of machine learning systems. Overall, we designed, applied, and compared seven kinds of AI approaches for the diagnosis of WD.

As part of performance evaluation, we conducted several new experiments: (i) computing diagnostics odds ratio (DOR) and correlating this to classification accuracy were conducted, which were never attempted previously. (ii) Furthermore, a power study was also conducted to estimate the optimum dataset size. (iii) Since all implementation were conducted on a supercomputer having 8 GPUs, timing analysis was performed to demonstrate the horsepower of our design. (iv) For best operating point characteristics of the DL system in terms of training data size, we optimize the CNN models by computing classification accuracy on varying the percentage of training data.

This study is the first study of its kind having the following novel approaches:

(i) Design of 3D optimization of two deep learning systems by classification accuracy with (a) changing DL layers and (b) folds of augmentation. The DL layers were varied between 5, 7, 9, and 11, and each design was tested with different dataset sizes created using the augmentation protocol. Augmentation was utilized to increase the dataset size by 2×, 3×, 4×, and 5× folds to create 5 sets of data including the original size. Three-dimensional

optimization was carried amongst these 4 DL designs and 5 augmented dataset sizes creating 20 combinations, which could then be used for choosing the best DCNN design for optimized classification. The final estimation would then be the number of layers and augmentation fold for the highest accuracy.

(ii) Design of inter-comparison benchmarking between the three kinds of DL systems and four types of ML classification systems.

(iii) Hypothesis validation by characterization of WD vs. controls using a combination of AI-based feature map strength (FMS), and Bispectrum signal processing approach.

(iv) Performance evaluation by computing (a) DOR, (b) generalization of DL paradigms, (c) reliability, (d) stability, and (e) time analysis of supercomputing.

This study is subdivided into the following sections: Section 1 details the ongoing research in the field of classification in MR neuroimaging. Section 2 explains the methods and materials used in this study. Section 3 has the details of all classification results, and Section 4 discusses the characterization of WD. Section 5 shows the performance evaluation, and Section 6 presents the discussion on the novel techniques used in this paper. The paper concludes with Section 7.

2 Background literature

The role of AI in WD classification has not taken the front stage yet. Few studies are using AI, on this topic compared with other diseases such as Alzheimer's, Parkinson's, or cancer and neuroimaging in general. One can therefore not ignore studies in the WD area which are not AI-based. Our recent work is the culmination of two decades of research, where biomarkers like serum or urine were used for diagnosis or segregation of patients having WD, primarily based on the threshold ranges of these biomarkers [1, 2, 25]. These methods consisted of 24-h urinary and serum laboratory tests for the identification of WD. Another class of method for diagnosis of WD consisted of eye examination for Kayser-Fischer rings and gene mutations [25]. A more recent method used all four types of biomarkers such as serum, eye, urine, and brain imaging to confirm WD [1].

Other frameworks for diagnosis of WD have been studied such as laboratory-based (blood) tests or genetic mechanisms for WD classification. Vrabelova et al. [26] described the utilization of blood tests involving DNA analysis for the ATP7B gene mutation study. Rosencrantz and Schilsky [3] used mutation of ATP7B analysis along with Kayser-Fischer ring in eyes and elevated copper level in urine. These tests were better at diagnosing disease compared to the serum/urine biomarker tests.

With the advent of MRI, several studies diverted towards neuroimaging-based classification approaches; however, they remained manual in nature. WMH has recently been explored recently in many diseases [27–29]. Kim et al. [30] analyzed hyperintensity in T1-weighted (T1W) and T2-weighted (T2W) MRI scans of suspected patients and found WMH in different parts of the brain such as the globus pallidus, thalamus, midbrain, and pons.

Recently, AI-community has started implementing this technology for characterization and classification of WD. In 2011, imaging took a leap towards fMRI for WD. Hu et al. [31] studied changes in the amplitude of low-frequency fluctuations (ALFF) while conducting fMRI on WD patients. Resting-state functional magnetic resonance (fMRI) images have shown promise and were employed to measure ALFF in different parts of the brain [32, 33]. Furthermore, the evolution of ML started to penetrate the imaging domain for WD classification [34, 35]. Kaden et al. [34] have demonstrated the use of support vector machine (SVM) and parameterized

Fig. 1 Control (left) and WD (right). Both images show the skull and background removed
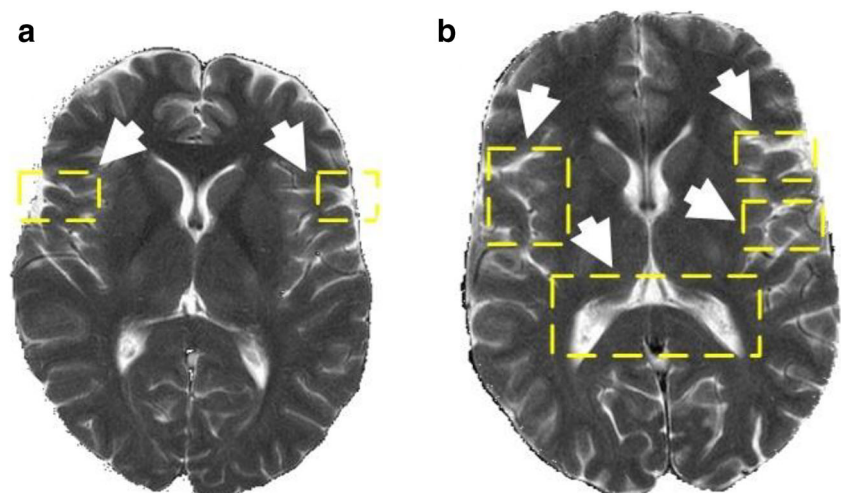
Table 1 Five types of cDCNN models consisting of different CNN layers

	C1 DCNN type	C2 Convolution 2D layers	C3 Max-pool layers	C4 Dense layers
R1	DCNN5	1	1	3
R2	DCNN7	2	2	3
R3	DCNN9	3	3	3
R4	DCNN11	4	4	3
R5	DCNN13	5	5	3

generalized learning vector quantization (PGLVQ) for WD classification with an accuracy of 87.5% and 90.1%, respectively. Jing et al. [35] used independent component analysis (ICA) with functional networks and SVM for WD classification and obtained the AUC of 0.94 and accuracy of 89.4% (specificity: 90.0%, sensitivity: 89.3%) with aberrant functional networks (FN). None of the above studies demonstrated automated approaches for WD classification and characterization. Our study uses a 3D optimized deep learning–based paradigm for classification of WD against controls and further extends the DL models combined with signal processing for tissue characterization. The CADx system shows three kinds of DL and four kinds of ML for WD classification and offers a novel approach to the diagnosis of WD.

3 Methodology

### 3.1 Patient demographics, acquisition, and data augmentation

A cohort of 46 patients T2W-TSE MRI scans (average age: 40.73 ± 11.3 years, equal M/F ratio) between the years 2011 and 2015 was analyzed (approval was obtained from the Institutional Ethics Committee, Azienda Ospedaliero Universitaria (A.O.U.), Cagliari, Italy).

Imaging examinations were performed using a 1.5-T superconducting magnet (Philips, Best, The Netherlands) with a head coil according to a standardized protocol. In each subject, the conventional diffusion-weighted imaging (DWI) was performed with single-shot spin-echo with 2 diffusion-sensitivity values of 0 and 1000 s/mm^2 along the transverse axis. As part of our general brain protocol, axial and sagittal 2D FLAIR images (10000/140/2200 ms for TR/TE/TI; matrix: 512×512; FOV: 240×240 mm^2; section thickness: 5 mm) were acquired. In addition to FLAIR and DWI sequences, axial spin-echo T1-weighted images (500–600/15/2 for TR/TE/excitations) and fast spin-echo T2-weighted images (2200–3200/80–120/1,2 for TR/TE/excitations; turbo factor, 2) were also obtained with the same section thickness.

3.1.1 Data augmentation

The initial MRI data were manually classified by our radiological team which was then prepared for further processing. Because the cohort consisted of 37 controls and 9 WD patients, we had an unequal number of images in both classes. As each patient MRI study had 12–13 slices, this resulted in 458 control images and 115 WD images. For optimal

Fig. 2 A 3D view of DCNN used for training and testing of the WD/control dataset
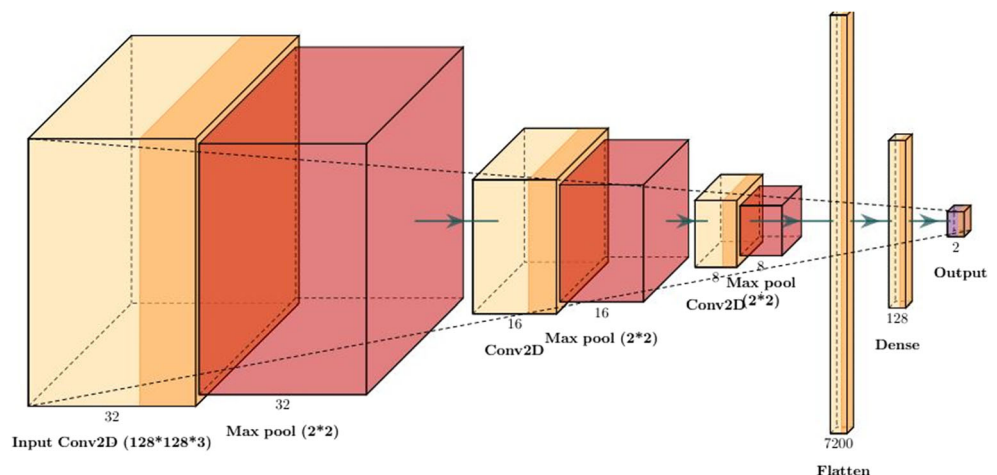
performance with an unbalanced dataset, the augmentation protocol using a python "Augmentor" API was applied in WD class resulting in 343 more WD images. Because deep CNN (DCNN) needs a large number of images for proper training and performance, we increased the number of images from 458 by 2×, 3×, 4×, and 5× folds in both classes, and the system was then trained and tested to find which augmented set yield optimal results. To avoid the unrealistic brain MRI scans during the augmentation protocol, we followed the acceptable protocol of rotating the image by − 10 to 10° randomly. This would prevent methods like flipping horizontal or vertical or rotating by larger angles.

3.1.2 Preprocessing: skull and background removal

Preprocessing is an essential component of the classification process. It helps extract the region-of-interest (ROI) from the MRI images. There are two important steps: (i) removal of the skull region and (ii) removal of the black background to prepare for the segmented ROI. As there are standard packages available which are well-accepted and published, we used BrainSuite [36] combined with volBrain [37] to segment and remove the background images. BrainSuite was used to read DICOM images that converted to nii files (nii file type is primarily associated with NIfTI-1 Data Format by

Fig. 3 Block diagram of DL-based classification and characterization system
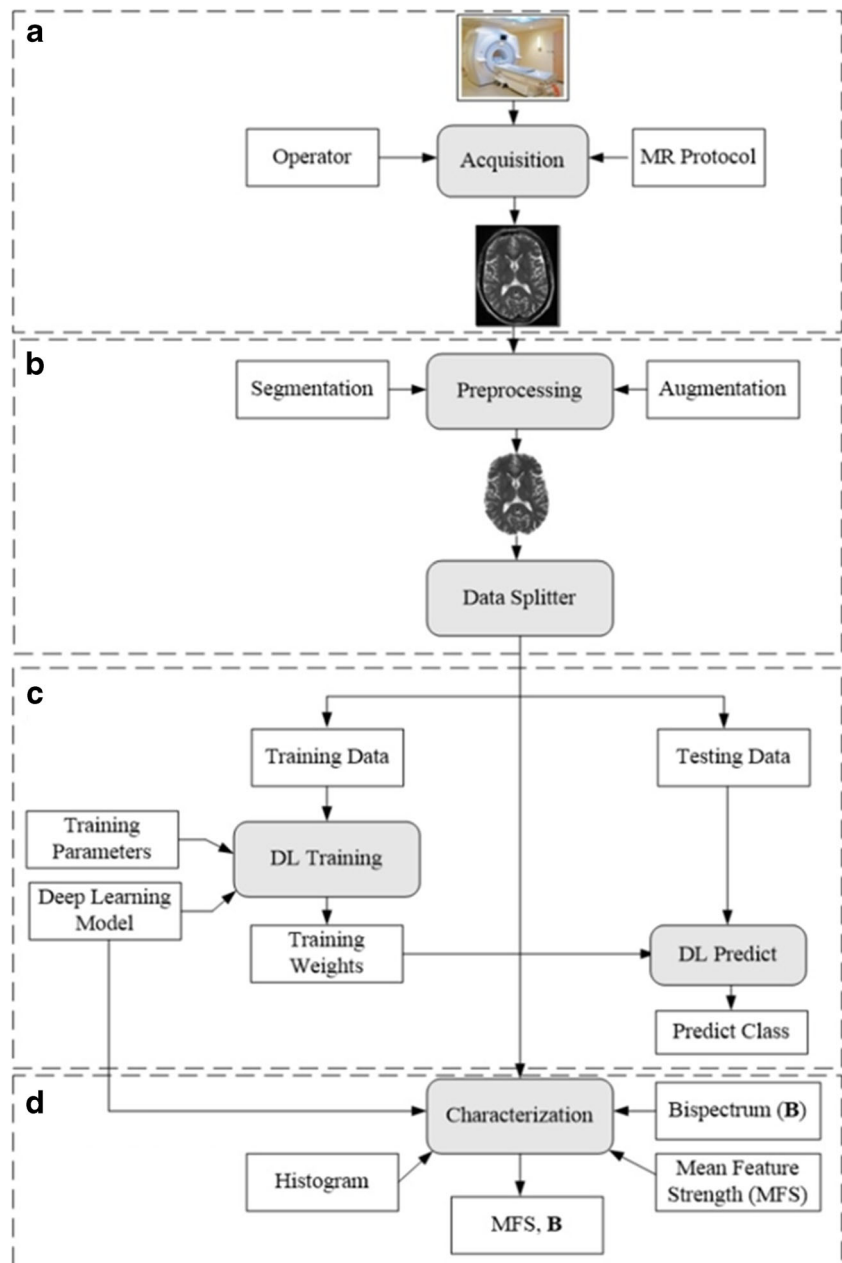
Table 2 Combination of feature types for four types of ML systems (bold cell indicates maximum accuracy obtained with FC3 feature combination and random forest ML technique)

Feature combinations (FC)		Four types of ML systems			
		ML1 k-NN	ML2 DT	ML3 SVM	ML4 RF
FC1	Haralick	63.813 ± 1.54	75.43 ± 1.58	84.08 ± 2.21	85.55 ± 1.67
FC2	Haralick + LBP	63.348 ± 2.41	76.56 ± 1..86	85.19 ± 1.71	86.12 ± 2.00
FC3	Haralick + Hu moments	63.156 ± 2.41	75.34 ± 1.65	84.62 ± 2.58	**87.43 ± 1.54**
FC4	LBP + Hu moments	68.276 ± 1.44	76.03 ± 1.35	47.40 ± 2.46	85.93 ± 2.46

Neuroimaging Informatics Technology Initiative) and obtain the grayscale images of the brain with the skull. volBrain helps to create a mask of the brain which can be used to remove the skull from the original MRI grayscale images. These were then further segmented and morphologically cleaned to remove the background. A sample pair of images from a patient with WD and control is shown in Fig. 1. The WD segmented brain images had brighter regions (higher WMH) in the convoluted zones of the brain (as shown inside yellow dotted rectangles Fig. 1) compared to control images.

3.2 Local architecture: deep CNN configurations

Our group had developed several CNN architectures covering a wide number of applications, namely, radiological imaging [38], stroke [19, 21, 39–41], liver [42], and cancer [43]. We have extended this to Wilson disease first time, and this is the first study of its kind which uses deep learning. The DCNN architecture used is shown in Fig. 2. It is composed of three convolution layers, each followed by a max-pooling layer, thus a total of six layers. A flattening layer that follows after these six layers converts the 2D signal to 1D. The final layer is a hidden dense layer consisting of 128 nodes. As usual, the final output is a softmax layer that has two outputs

corresponding to WD or control. This design of lesser layers was chosen as the number of classes was only two, and the DCNN was able to work with the desired accuracy. Thus, the current configuration would need less storage space and inference time as compared to pre-trained CNN models. We adapted the ReLU function for the convolution and dense layers since that helps with fast convergence to the solution as compared to sigmoid or tanh activations functions [44]. Because the DCNN had augmentation implemented, we, therefore, consider several layered options corresponding to different DCNN configurations, shown in Table 1. It shows 5 types of DCNN combination consisting of different convolution layers, max-pool layers, and dense layers. Thus, for the adaption of all experiments, it is, therefore, necessary to undergo 3D optimization between the accuracy, CNN layers, and the folds of augmentations. The block diagram of the DL-based classification and characterization system is shown in Fig. 3. As seen in Fig. 3, the MRI scans are preprocessed and split into training and testing. Training images are used to train the deep learning model along with gold-standard labels, generating the training weights. These weights transformed the test patients to predict their class labels. Bispectrum, DL model's mean feature strength, and histogram were used for the characterization process to yield mean feature strength (MFS) and Bispectrum (B) values.
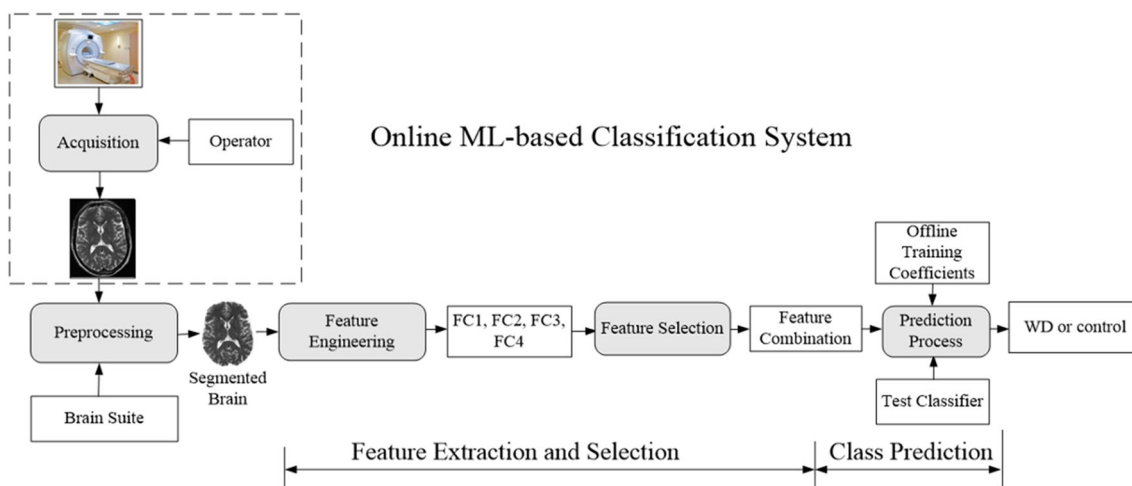
Fig. 4 Block diagram of the ML-based classification system

Table 3 Accuracy of different cDCNN layers vs. augmentation (bold cell indicates that maximum accuracy obtained with 4× fold augmentation using 9 layered cDCNN)

		C1	C2	C3	C4
	Classification accuracy (%) vs. CNN layers vs. augmentation				
	cDCNN layers	Augmentation types			
		2× times	3× times	4× times*	5× times
R1	5	91.56 ± 4.72	84.059 ± 12.14	86.715 ± 5.62	94.997 ± 2.18
R2	7	95.465 ± 2.49	97.025 ± 2.00	97.028 ± 2.6	95.465 ± 1.55
R3	9*	96.303 ± 2.66	96.558 ± 2.30	**97.185 ± 2.52**	96.402 ± 2.33
R4	11	95.621 ± 2.18	96.559 ± 1.91	96.558 ± 2.53	95.621 ± 3.10
R5	13	93.121 ± 4.30	93.278 ± 3.29	94.215 ± 2.85	93.745 ± 2.55

The definition of conventional ReLU is as given as $\sigma = \max(0, x)$. Here, σ is the activation value and x is the input to ReLU function. This equation was modified to $\sigma = (\max(0, x))^{\wedge} 1.00001$. Note that the differential of this equation is 0 at point $x = 0$, whereas the conventional ReLU is not differentiable at the origin. Since loss minimization in DCNN used gradient descent process which needs differential of various variables, it is, therefore necessary to have ReLU made continuously differentiable at $x = 0$. The equation for loss is given in formula 13 in Appendix. In the improved deep CNN (iDCNN) for better performance, this change of activation function was implemented in for all the convolution layers and the dense layer.

3.2.1 Transfer learning

As part of the overall CADx system, we benchmark our DL systems against the "transfer learning–based Inception V3" [24] pre-trained CNN and four machine learning paradigms such as k-NN, DT, SVM, and RF. InceptionV3 is a 42-layered deep model consisting of 11 inception modules (each comprising of multiple convolution layers and max-pooling filters), followed by three fully connected layers and a softmax activation layer. It was originally designed for a 1000 class

ImageNet dataset for the famous ImageNet Large-Scale Visual Recognition Competition (ILSVRC) and the model is customized to two class problems for this study and trained further after loading the pre-trained weights of the ImageNet dataset. Inception V3 was designed to reduce the overall number of parameters to reduce network size and inference time. The reduction in parameters is done with help of factorizing convolutions. For example, a 5×5 filter convolution can be done by two 3×3 filter convolutions. The parameters in this process reduce from $5 \times 5 = 25$ to $3 \times 3 + 3 \times 3 = 18$. Thus, it brings a 28% reduction in the number of parameters. With a smaller number of parameters, the model will less overfit and thus also increases the accuracy.

3.2.2 Machine learning

Our group has been very active in machine learning (ML) in several tissue characterization and classification applications, namely, diabetes [45], plaque [46–50], thyroid cancer [51, 52], ovarian cancer [53, 54], prostate cancer [23], liver cancer [42, 55, 56], lung cancer [57], skin cancer [58], bladder cancer [59], heart [60], cardiovascular disease risk [61–64], coronary artery disease [65], stroke [50, 66, 67], arrhythmia [68], and gene expression characterization [69]. We adapted similar
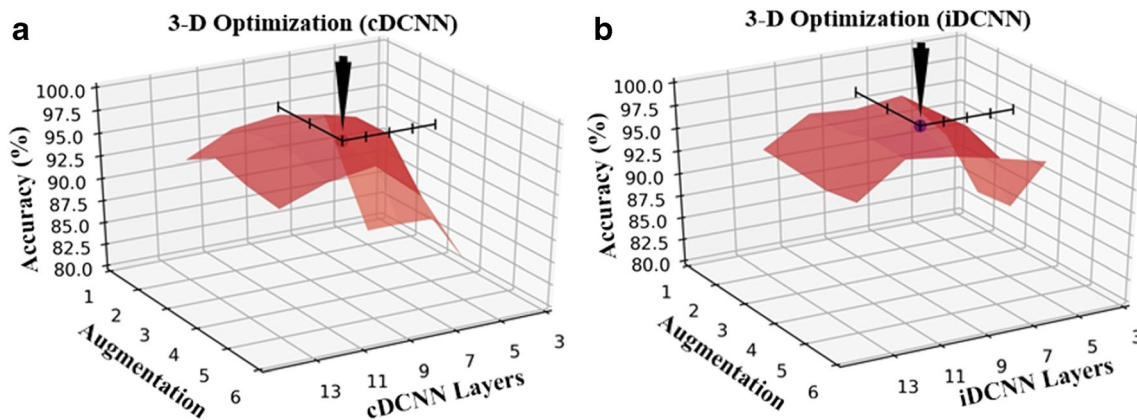
Fig. 5 Three-dimensional optimization for best DCNN layers. **a** cDCNN showing optimization point at cDCNN = 9 layers and **b** iDCNN showing optimization point at iDCNN = 9 layers

Table 4 Results on the "effect of training data size" using DCNN9*-Augm4* (DL9A4) combination

	K-fold	C1 Train-test ratio	C2 TC&	C3 cDCNN accuracy (%)	C4 iDCNN accuracy (%)	C5 tDCNN accuracy (%)
R1	K2	50:50	2	94.53 ± 1.10	95.31 ± 2.20	78.12 ± 8.83
R2	K3	67:33	2	96.04 ± 2.10	96.87 ± 1.56	83.85 ± 6.76
R3	K4	75:25	4	96.48 ± 1.49	97.26 ± 0.78	86.45 ± 2.38
R4	K5	80:20	5	96.79 ± 1.12	97.49 ± 1.78	86.71 ± 2.70
R5	K10	90:10	10	97.19 ± 2.52	98.28 ± 1.55	87.81 ± 2.42
R6	TT	100:100	1	98.43	100	90.62

$^{\&}$ *TC*, total combinations

paradigm in our current setting for Wilson disease application. Different feature selection methods were used consisting of Haralick, Hu moments, and LBP feature extraction frameworks. Table 2 ML4 (consisting of random forest) shows the highest accuracy corresponding to selected feature combination FC3 that consisted of Haralick's and Hu's moments. A brief description of these features is as given below:

- Haralick features: These are based on the texture of the image and generated using Gray Level Co-occurrence Matrix (GLCM) using one of energy, entropy, or homogeneity of these matrix element values.
- Hu moments: These are features of the object in the image and generated using centralized moments.
- LBP features: LBP (Local Binary Pattern) is also a powerful texture-based feature calculated by comparing a pixel with 8 neighboring pixels.

The equations for energy, entropy, and homogeneity used in Haralick features are given in Eqs. 1–3.

$$Energy = \sum_i \sum_j P_d(i,j)^2 \tag{1}$$

$$Entropy = -\sum_i \sum_j P_d(i,i) \log(P_d(i,j)) \tag{2}$$

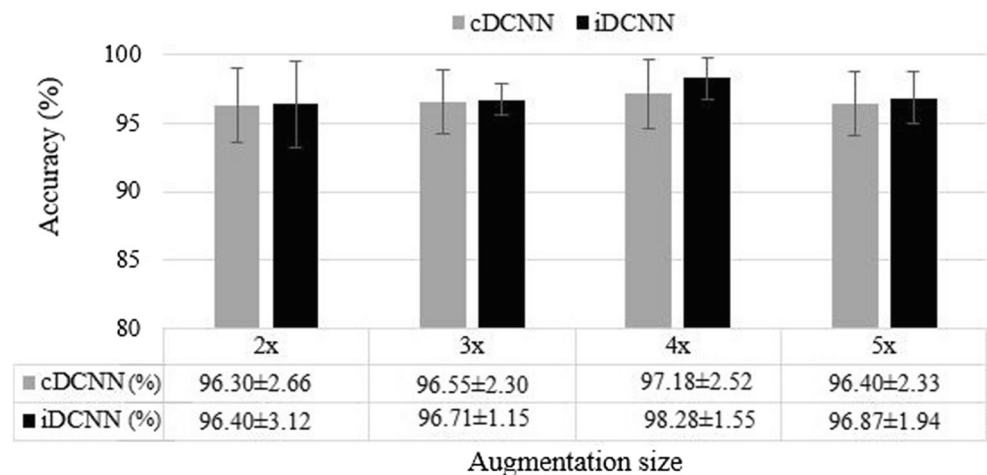$$Homogeneity = \sum_i \sum_j \frac{1}{1+(i+j)^2} P_d(i,j) \tag{3}$$

The equation for Hu moments is given by Eq. 4.

$$\mu_{pq} = \sum_x \sum_y (x-x)^p (y-y)^q f(x,y) \tag{4}$$

where μ_{pq} are the centralized moments; x, y are pixel coordinates and $f(x,y)$ are pixel intensities at these coordinates. Here $p = 0, 1, 2, 3$ and $q = 0, 1, 2, 3$.

The block diagram of the ML-based classification system is shown in Fig. 4. As seen in Fig. 4, the preprocessing block processes the acquired MRI scans to yield the segmented brain region. This was implemented using BrainSuite and volBrain software which gives a very clean mask used to segment grayscale images very clearly. The engineering features were extracted using a combination of Haralick, Hu moments, and LBP feature–based methods. ML-based methods (k-NN, DT, SVM, or RF) already trained on labeled segmented images are used as input to the prediction process for classification of the test MRI input. The final output consisted of a binary output class consisting of either WD or control class.

Fig. 6 Comparison of cDCNN9* and iDCNN9* for different augmentation size

	2x	3x	4x	5x
cDCNN (%)	96.30±2.66	96.55±2.30	97.18±2.52	96.40±2.33
iDCNN (%)	96.40±3.12	96.71±1.15	98.28±1.55	96.87±1.94

Augmentation size

Fig. 7 Generalization of the cDCNN9* using K10 protocol

% of Training data	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Accuracy	82.1	89.18	91.27	93.9	95.93	96.71	95.62	95.93	96.71	97.19
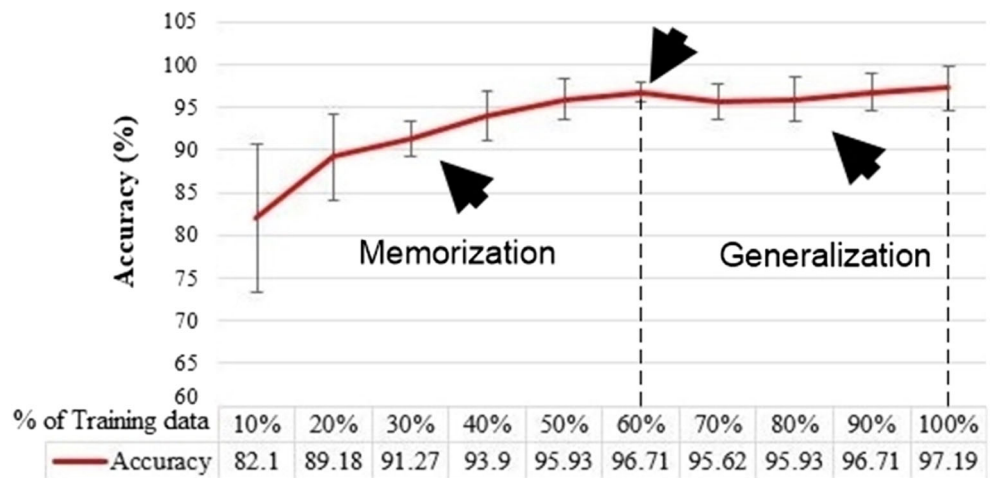
3.3 Performance evaluation protocol

To evaluate the model performance of the DCNN systems, different tests were adopted such as (a) Wilson Disease Segregation Index (WDSI) to estimate the feature strength of different AI techniques between two classes; (b) diagnostics odds ratio (DOR) for DCNN and ML methods; (c) power analysis to find the optimum dataset size; (d) timing analysis of supercomputer vs. local computer; (e) for best operating point characteristics, optimization of the CNN model with a percentage of the training datasets; and (f) finally, the validation of DL models against the well-accepted and published biometric facial dataset.

3.3.1 Wilson Disease Segregation Index of WD against control

We compute the WDSI, which is an indicator for the class separation between the controls and Wilson disease, expressed as percentage, and is mathematically given by Eq. 5:
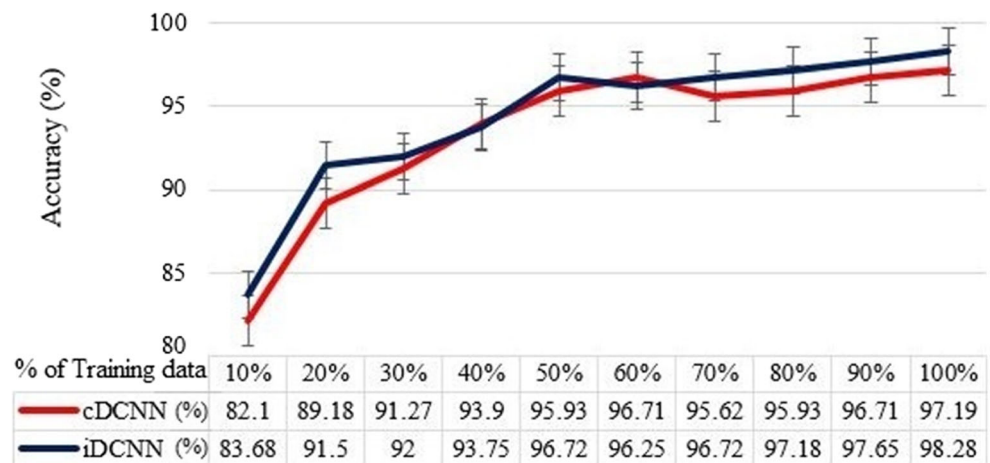
$$\text{WDSI} = \left(\frac{|\mu_{WD} - \mu_C|}{\mu_C} \right) \times 100 \qquad (5)$$

where μ_C is the mean feature strength of control class and μ_{WD} is the mean feature strength of WD class.

4 Results

This section primarily demonstrates three optimization experiments on DCNN, and four comparative experiments between DCNN and ML systems. The first three experiments show 3D optimization of DCNN layers during the augmentation process (DCNN9*), the effect of training on DCNN performance, and the optimal sample size selection for DCNN to be generalized. The batch is focused on benchmarking of the DL system against four ML systems, AUCs of DL vs. ML systems, and the segregation index of WD vs. controls.

Fig. 8 Comparison of cDCNN9* and iDCNN9* for different percent of training data

% of Training data	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
cDCNN (%)	82.1	89.18	91.27	93.9	95.93	96.71	95.62	95.93	96.71	97.19
iDCNN (%)	83.68	91.5	92	93.75	96.72	96.25	96.72	97.18	97.65	98.28

4.1 Three-dimensional optimization of DCNN layers during augmentation process (DCNN9*)

The objective of this protocol is to find out the best CNN layer and augmentation combination. Since there are 20 "DCNN + Augm" combinations (5 types of CNN layers and 4 types of augmentations) and each combination is a K10 protocol (10 combinations in each of the K10 protocols), then a total of 200 different runs (or jobs), we, therefore, take advantage of supercomputer power to run five types of DCNN designs over four kinds of augmentations. Since DCNN accuracy varies depending upon the number of hidden convolution layers, it is, therefore, vital to undergo the optimization run (see Table 1). The results of this can be seen in the 3D surface plot in Fig. 5 and 3D bar graphs in Fig. 14 in Appendix, and the corresponding values for cDCNN are shown in Table 3. As seen in the 3D surface plot, with an increase in the CNN layers (down the rows R1 to R5), the accuracy increases and then gradually falls. Similarly, with an increase in augmentation, the accuracy increases initially (from C1 to C3) and then falls. The best CNN layer-augmentation combination was DCNN9*-Augm4*. All the subsequent experiments were then conducted at this combination point of DCNN9*-Augm4* or a short form as "DL9A4". The equation of accuracy is given as formula 8 in Appendix. The equation for standard deviation is given as formula 14 of Appendix.

4.2 Effect of training on DCNN performance using "DL9A4"

K-fold cross-validation protocols were executed on DCNN9*-Augm4* combination dataset. For this different train and test split (K2, K3, K4, K5, K10, and TT), we used different combinations, as required.

Table 5 Comparison of 7 AI systems for WD classification (in the increasing order of AUC) (bold cell indicates the maximum AUC value obtained with iDCNN)

	Model	C1 Accuracy (%)	C2 AUC	C3 p value	C4 Confidence interval
4 machine learning systems
R1	k-NN	63.35 ± 2.42	0.781	< 0.0001	0.683339 to 0.776441
R2	DT	76.56 ± 2.14	0.812	< 0.0001	0.717490 to 0.806813
R3	SVM	85.20 ± 1.72	0.851	< 0.0001	0.714949 to 0.793839
R4	RF	86.13 ± 2.00	0.918	<0.0001	0.732858 to 0.820271
3 deep learning systems
R5	tDCNN	87.81 ± 2.42	0.927	< 0.0001	0.809110 to 0.965826
R6	cDCNN	97.19 ± 2.53	0.984	< 0.0001	0.894384 to 0.996754
R7	iDCNN	98.28 ± 1.55	**0.990**	< 0.0001	0.943991 to 1.000000
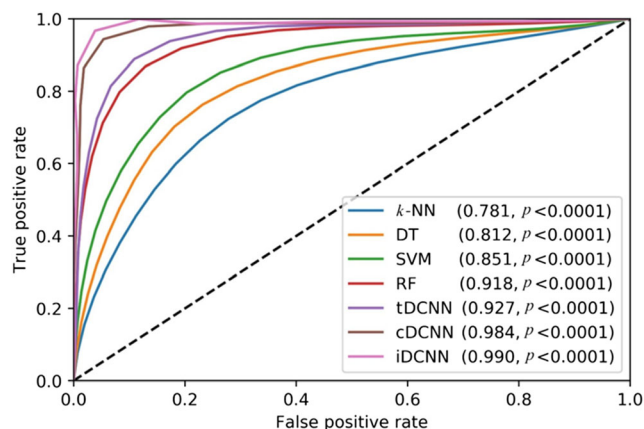
Fig. 9 ROC curves for 7 AI systems: 3 DCNN systems (cDCNN, iDCNN, tDCNN) and 4 ML (k-NN, DT, DVM, RF). The AUC values are shown from lowest to highest, where iDCNN does the best

Table 4 shows the effect of training data size (with increasing K-fold) on the three types DCNN systems. For convenience, we have added the transfer learning system (tDCNN) here as well. The comparisons of cDCNN and iDCNN results are given in Fig. 15 in Appendix and Fig. 6.

As seen, accuracy slowly rises from K2 to K10 and is best for TT (training is the same as testing) protocol, which was used for validation. Note the order of performance was iDCNN > cDCNN > tDCNN.

4.3 Optimal sample selection for generalization of the DL system

The DL9A4 was tested with a different percent of training data, and K10 accuracy was found for each data size as shown in Fig. 7. The curve shows accuracy increases until the point of inflection, which is 60% of the training dataset. This shows the data size has the capacity to generalize after 60% of the dataset.

The comparison of cDCNN and iDCNN performance for different percent of training data is shown in Fig. 8.

4.4 Benchmarking of three DL systems against four ML systems

The benchmarking was conducted for DCNN9*-Augm4* (DL9A4) combination against the transfer learning–based

Table 6 WDSI between DCNN and ML systems

Method	Control	WD	WDSI
tDCNN	189.92 ± 54.35	239.69 ± 62.57	26.20
cDCNN	148.30 ± 44.93	247.74 ± 69.52	67.05
iDCNN	146.53 ± 43.56	251.65 ± 45.39	71.73
ML	954.99 ± 276.43	1037.66 ± 292.01	8.65

Fig. 10 **a** Comparison of FMS values for control (green color) and WD class (red color). **b** Comparison of Bispectrum (B) values for control (green color) and WD class (red color)

	Input	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7	Layer 8	Mean
Control	84.6	654.57	742.73	1123.33	1346.67	116.43	253	31.62	148.3	500.14
Wilson Disease	88.45	687.1	773.33	1169.03	1394.75	117.36	257.24	32.15	247.74	529.68

	0	15	30	45	60	75	90	105	120	135	150	165	180	Mean
Control	27.87	16.38	11.23	9.84	8.7	9.27	11.27	8.64	8.56	9.68	10.78	15	27.91	13.472
Wilson Disease	39.58	29.07	20.29	15.88	14.07	14.36	17.19	13.37	12.72	14.12	17.19	23.6	39.9	20.872

Inception V3 pre-trained model (tDCNN) and four types of ML systems (k-NN, DT, SVM, and RF) using K10 protocol as discussed in Section 3. The comparative results of benchmarking can be seen in Table 5.

The best performance was achieved for iDCNN9* using a modified ReLU function.

4.5 Receiver operating characteristics curves (DL vs. ML)

Receiver operating characteristics curve shows the relationship between the false-positive rate (FPR) and the true-positive rate (TPR). The AUC validates our hypothesis. The

ROC curve for 4 ML classifiers and three DCNNs are given in Fig. 9, while the AUC values are shown in Table 5 (column C2). ROC curve is a plot between TPR (y-axis) and FPR (x-axis). The equation for TPR and FPR is given as formula 11 and 12 in Appendix.

4.6 Wilson Disease Segregation Index

The Wilson Disease Segregation Index (WDSI) was calculated using the mean feature strengths of DCNN and ML features. The large value of WDSI shows larger segregation between WD and controls which justifies the ability of the AI
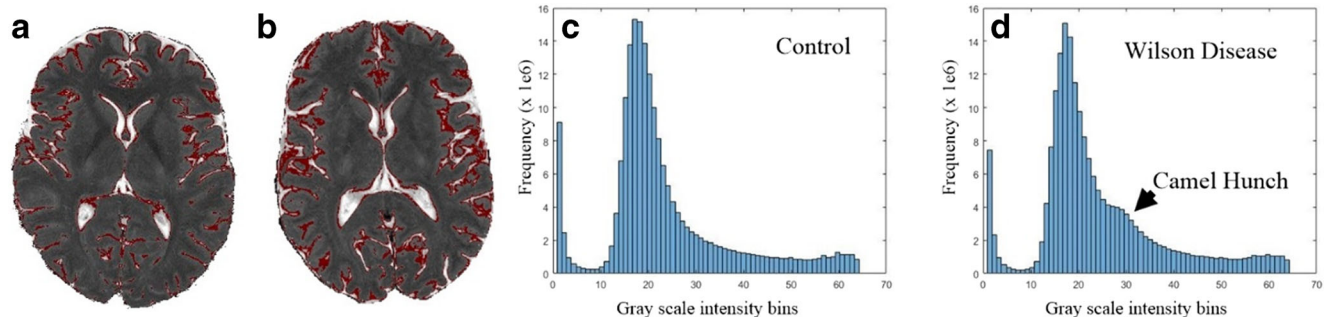
Fig. 11 Pixels in range 100–140 marked red in **a** control and **b** WD. **c** and **d** show the comparison of the histogram for control and WD class. The camel hunch is seen in the WD class representing the WMH
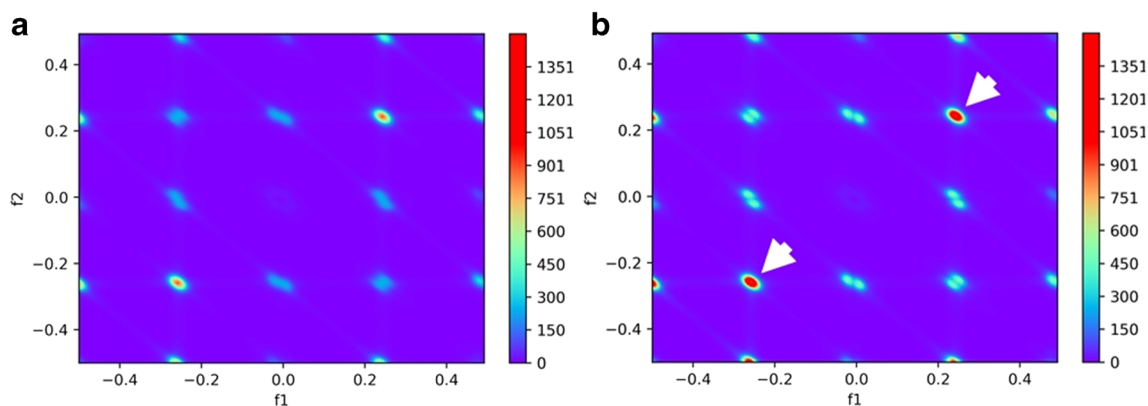
Fig. 12 Bispectrum 2D plot of **a** control and **b** WD. (White arrows show higher strengths of the B-values in the WD class in 2D Bispectrum plots)

method as seen in Table 6. The order of the WDSI is iDCNN > cDCNN > tDCNN > ML.

5 WD characterization

Characterization [12, 19, 23, 70] is vital for validation of our hypothesis that WD has a higher WMH compared to controls and this accounts for the increase in feature strength in the layers of the hidden layers of the DCNN. We thus evaluate the FMS for the layers and see the value of the FMS between the WD against controls.

5.1 Hypothesis validation 1: mean feature map strength using DNN9*-Augm4* (DL9A4)

Feature map strength (FMS) is the mean of activation values over all images in a class. FMS for a trained DNN9*-Augm4* model at 8 hidden layers is shown in Fig. 10a.

As seen from Fig. 10a, the FMS values are consistently higher for WD class of the output layers. The mean FMS values for control and WD class are 500.14 ± 46.09 and 529.68 ± 47.23, respectively, showing an increase of 5.77% (C.I. 4.5463 to 54.5426, p value < 0.0001). This supports our

1st hypothesis that WMH of WD class is higher than the controls.

5.2 Camel hunch phenomenon

The WMH could be better visualized by understanding the histogram distribution of the brain region. The histogram is computed by considering the bin size of 4, leading to 64 bins (256 values, divided by 4). This is repeated for both classes. As seen in Fig. 11c, d, the histograms show a camel hunch-like shape in WD class from 25th to 35th bin corresponding to intensity range 100–140. This phenomenon occurs due to regions of WMH [5–7] in WD MRI scans, mainly at the convoluted edge of the folds of the brain.

5.3 Hypothesis validation 2: bispectrum strength computation

Bispectrum (B) falls in the category of higher order spectra (HOS) [23]. To calculate HOS, the Radon transform of images was calculated at various angles from 0 to 180° in an interval of 15°. Here, the Radon transform was applied to the images where pixels of MRI scans in range 100–140 are segregated (shown in red: Fig. 11a, b). On calculating mean
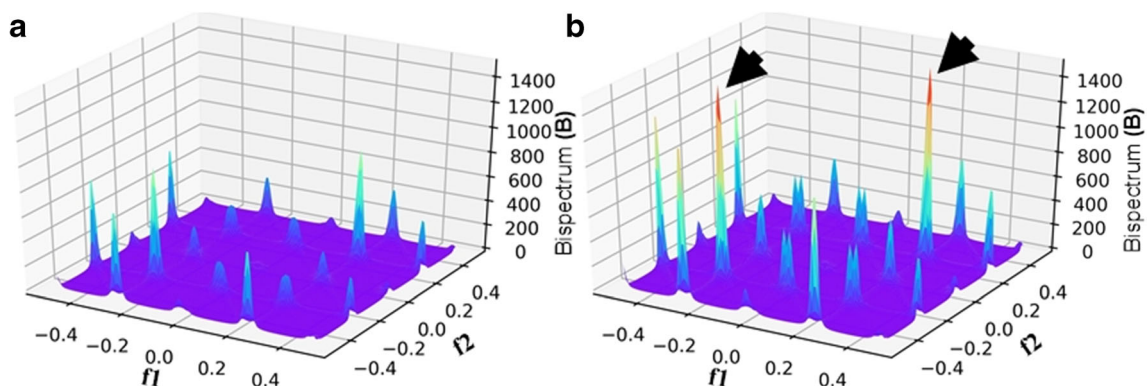
Fig. 13 Bispectrum 3D plot of **a** control class and **b** WD class. (Black arrows show more B-value for the WD class indicated by higher peaks in 3D Bispectrum plots)

B-values on all images of two classes, the B-value for WD is found consistently higher than control (see Fig. 10b), with mean of 20.87 for WD and 13.47 for control class showing a rise by 54.71% (C.I. 5.5490 to 9.2526, p value < 0.0001). Figure 12 shows the 2D representation of the Bispectrum strength for WD against the control, and Fig. 13 shows more Bispectrum strength for WD in its 3D plot. The equation for Bispectrum is given as $Bispectrum(B(f1, f2)) = E[\mathcal{F}(f1) \times \mathcal{F}(f2) \times \mathcal{F}(f1 + f2)]$, where, B is the Bispectrum value, $\mathcal{F}$ is the Fourier transforms and E is the expectation operator. The region Ω of computation of bispectrum and bispectral features of a real signal is uniquely given by a triangle $0 < = f2 < = f1 < = f1 + f2 < = 1$.

6 Performance evaluation

6.1 Diagnostics odds ratio

Diagnostic odds ratio (DOR) is used to discriminate subjects with a target disorder from subjects without it. DOR is calculated according to Eq. 6 [71]. DOR can take any value from 0 to infinity. A test with a more positive value means better test performance. A test with a value of 1 means it gives no information about the disease and with a value less than 1 means it is in the wrong direction and predicts opposite outcomes.

$$DOR = \frac{TP/FN}{FP/TN} = \frac{sens/(1-sens)}{(1-spec)/spec} \quad (6)$$

where TP, FP, TN, and FN represent true positive, false positive, true negative, false negative. Sens and spec stand for the sensitivity and specificity, respectively, and equation is given in Formula 9 and 10 of Appendix. The DOR values for all ML and DL methods are shown in Table 7.

Table 8 Timing analysis and gains in time for the supercomputer against the local computer

Comparison between LC# vs. SC+ (1-epoch in s)			
	LC#	SC+	Gain (folds)
2× Augm	170	19	8.94
3× Augm	179	25	7.16
4× Augm	215	27	7.96
5× Augm	234	29	8.06

+ SC, supercomputer

LC, local computer

6.2 Power analysis

The sample size was calculated according to Eq. 7 [72] using the mean difference between K10 mean accuracy of DCNN9* and DCNN11 while keeping the augmentation 4× (see Table 3, cell number (C3, R3), and cell number (C3, R4)).

$$\text{Sample Size} = \frac{2 \times (Z_\alpha + Z_{1-\beta})^2 \times \sigma^2}{\Delta^2} \quad (7)$$

Here, the value of $Z_\alpha = 3.2905$ for type 1 error having a value of 1%, and $Z_{1-\beta} = 1.6449$ for type II error having a value of 1%. Here, σ (standard deviation) = 2.53 and Δ (mean difference) = 0.627. Substituting these values in Eq. 2, the sample size returns a value of 793. This is the required sample size. The database we adapted has 1832 samples for WD or controls using 4× augmentations. Thus, our database is 2.31 times the required limit and we are above the limit by 1039 samples.

6.3 Timing analysis

The supercomputer was adapted during the training of DCNN for the optimal performance of the CADx system. We,

Table 7 Sensitivity and specificity with increasing order of DOR for the 7 AI systems (the bold rows indicate the maximum values using cDCNN and iDCNN)

	Model	C1 Sensitivity (%)	C2 Specificity (%)	C3 DOR
R1	k-NN	62.50	50.75	1.71
R2	DT	80.95	71.91	10.88
R3	RF	92.43	82.41	57.25
R4	tDCNN	93.40	82.16	65.25
R5	SVM	80.40	98.29	235.86
R6	**cDCNN**	**98.11**	**95.34**	**1066**
R7	**iDCNN**	**99.06**	**97.64**	**4399**

Table 9 Reliability analysis for different percent of training data

	C1 Data percent	C2 Accuracy (%)	C3 Reliability (%)	C4 Standard deviation
R1	10	82.10	89.45	8.66
R2	20	89.18	94.28	5.09
R3	30%	91.26	97.74	2.06
R4	40	93.90	96.92	2.89
R5	50	95.93	97.55	2.35
R6	60	96.71	98.81	1.15
R7	70	95.62	97.85	2.05
R8	80	95.93	97.32	2.57
R8	90	96.71	97.79	2.14
R10	100	97.18	97.40	2.52

Table 10 Benchmarking of our proposed DCNN strategy against the previously published literature (the bold rows indicate that best accuracy/AUC values obtained using cDCNN and iDCNN)

SN	C1 Author	C2 Appl.	C3 Technique	C4 Imaging modality	C5 AI[$]	C6 Accuracy (%)	C7 AUC (p value)
R1	Kaden et al. [35] (2015)	WD	SVM and PGLVQ	NG	ML	90.1	NG
R2	Suk *et al.* [74] (2017)	ALZ[a]	DBN+NN	MRI (ADNI)	DL	85.91	0.9082
R3	Abiwinanda *et al.* [11] (2018)	BT[&]	CNN	MRI	DL	84.19	NG[#]
R4	Zhang *et al.* [75] (2019)	MCI[b]	SSGSR	fMRI	ML	88.50	0.965
R5	Abrol *et al.* [76] (2018)	MCI[b] vs. ALZ[a]	ResNet	MRI	DL	82.7	0.89
R6	Jing *et al.* [35] (2019)	WD	SVM and ICA	fMRI	ML	89.4	0.94
R7	Richhariya *et al.* [77] (2020)	MCI[b]	SVM	MRI	ML	90	NG[#]
R9	Liu *et al.* [78] (2020)	MCI[b]	SVM	MRI	ML	88.5	0.897
R10	**Proposed study (cDCNN)**	**WD**	**CNN**	**MRI**	**DL**	**97.19**	**0.984 (*p* < 0.0001)**
R11	**Proposed study (iDCNN)**	**WD**	**CNN**	**MRI**	**DL**	**98.28**	**0.990 (*p* < 0.0001)**

therefore, calculated the gain as the ratio of time taken by local computer (LC) (which was HP Desktop 2010) to the time taken by the supercomputer (SC) (which was NVIDIA). The gain values are shown in Table 8.

As seen, the time taken by a local computer using CPU is around 7–9 times more than that of a supercomputer. Thus, it will take 72 h or 3 days for a job to run on a local computer which will take only 8 h on a supercomputer. The C.I. and p value of the timings are 113.2824 to 215.3842, p value = 0.0052.

6.4 Reliability analysis

Reliability is calculated using the formula: Reliability $(\%) = \left(1 - \frac{\mu_N}{\sigma_N}\right) \times 100$, where μ_N and σ_N are the mean and standard deviation of the classification accuracy. The variation of reliability according to data size is given in Table 9. For the system to be reliable and stable, we must meet three criteria [70, 73]: (i) If the reliability of the DCNN > 95%, then the system is reliable; (ii) If the SD < 5%, then the DCNN is considered as stable; (iii) Furthermore, if the variation in accuracy is not more than 5%, then the system is considered stable. In our case, we meet all the above 3 criteria. For data size above 20% (row R3), reliability (column C3) is above 95%, SD < 5% (row R3, column C4), and variation in accuracy < 5% for rows R4 to rows R10. This concludes the compliance of our DCNN system to be stable and reliable.

7 Discussion

The objective of this study was to classify images from patients with WD against control in unbalanced and weak brain MRI training datasets. The system design consisted of 3D optimization of the best DCNN model under best-

augmented conditions. Our optimization uses the best combination of DCNN9*-Augm4*. The design of iDCNN was comparable but superior to cDCNN. We also showed that iDCNN outperformed tDCNN by 11.92% and four types of "conventional machine learning–based systems" such as k-NN, decision tree, support vector machine, and random forest by 55.13%, 28.36%, 15.35%, and 14.11%, respectively. The performance evaluation of the DCNN system was evaluated using DOR and WDSI and all showed consistent results. We also showed the effect of training data on the system accuracy for this optimal point. The hypothesis was validated using two novel strategies for WD characterization using FMS and Bispectrum analysis.

7.1 Benchmarking

We benchmarked our DCNN systems against existing systems as shown in Table 10. As found from existing research, no classification work has been done in WD. Benchmarking table also shows a comparison between the previous studies and current proposed study. Overall, this was the first paper using state-of-the-art technology to optimize several AI methods in the diagnosis of WD. The benchmarking is done with recent papers referred by author and year and 1st column C1; it is followed by the type of brain diseases such as Alzheimer (ALZ), brain tumor (BT), and mild cognitive impairment (MCI) in column C2. In column C3, the techniques used are mentioned such as SVM, CNN, and DBM. Column C4 tells the imaging modality such as MRI and fMRI. In column C5, we have mentioned where authors have used ML- or DL-based AI technique. Columns C6 and C7 describe the accuracy and AUC (*p* value) comparison with these referred papers. The column C4 shows the neurological applications such as spatial MRI imaging or functional MRI imaging. Majority of the studies use ML or DL models. The

accuracy of the systems has an average value of 87%, while our system had about 10% improvement compared to the average value. R10 and R11 showed the proposed cDCNN and iDCNN methods which had an accuracy of ~ 97.2% and ~ 98.3%, respectively. The AUC for our proposed methods were 0.98 ($p < 0.0001$) and 0.99 ($p < 0.0001$).

7.2 A short note on WD characterization

Even though we are able to characterize the WD by computing the Bispectrum values and AI models, it is important to note that the Bispectrum values are computed in the pixel zone corresponding to the camel hunch region. This was valid for our datasets, but definitely need a wider set of clinical validations. Furthermore, the spatial slices considered were on an average of 12 per patient. To begin with the number of patients on control was four times the WD, which shows a slight imbalance. Thus, a strong pool of data is required to more validate this paradigm.

7.3 A short note on the role of the gold standard for the design of the Wilson disease system

The role of the gold standard is very crucial in the design of deep learning systems. They act like the binary event, "high risk and low risk," or "benign cancer vs. malignant cancer," or "cardiovascular event vs. no-cardiovascular event" or "cerebrovascular event vs no-cerebrovascular event." Such binary events can be well-detected diagnostically if they are trained using the deep learning model. Recently, our group developed a method for classification of symptomatic risk likely to have stroke vs. asymptomatic patients not likely to have the stroke [21]. The deep learning solution was very successful by training the deep CNN model using the gold standard based design by the neurologist. Such a process is also called as characterization of the disease since the deep learning system is able to use the features of the disease to classify into binary events such as Wilson disease vs. normal. Examples of several kinds of well-defined characterization systems can be seen in the machine learning section. Even though typically, the characterization can be used for binary classification, but the multiclass scenarios can also be developed when it comes to characterization [61, 79]. It just provides several levels of risk rather than two types of risk.

7.4 Strength, weakness, and extensions

This is the first study of its kind that considered DCNN architecture for classification and characterization of the WD. The architecture was optimized by changing the number of layers of the DCNN architecture and augmentation protocol. The AI system showed high performance and the results were validated. The WD characterization was conducted using two different models, first using the AI framework, and second using signal processing framework using higher order spectra by computing the Bispectrum values. The system showed consistent results and the hypothesis was validated.

Even though the pilot study showed powerful results, one can automate the manual segmentation step by automated methods [9, 80, 81]. Furthermore, more ML alternative and more features can be used in future [59]. More validations need to be conducted in the future, such as cross-modality fusion using registration methods [82, 83]. More neurological model-based techniques can be designed [84, 85].

The system can be extended to transfer learning–based approaches to avoid the heavy supercomputer processing time during training, therefore using the pre-trained weights [79]. In spite of our successful pilot study showing a set of seven successful AI models, this can be taken as a launching pad for multicenter data collection for bigger trials. The scanners used for imaging also can play a role while acquiring the MRI data, just like other modalities [86].

8 Conclusion

This is the first study of its kind to use an advanced CADx system based on seven kinds of AI combinations to classify images from patients with WD vs. controls to achieve the best possible architecture of DCNN and to attain the best accuracy of 98.28 ± 1.55%. The three DCNN methods were compared with four ML methods showing the benefit of deep learning. The study also used the characterization of WD using two hypotheses showing feature map strength and Bispectrum strength of WD higher due to regions of WMH in MRI scans. A detailed performance evaluation was also implemented using diagnostics odds ratio, power analysis, supercomputer timing analysis, and generalization analysis of DCNN performance.

Appendix

If TP, FP, TN, FN, TPR, and FPR represent true positive, false positive, true negative, false negative, true-positive rate, and false-positive rate, respectively, then the performance parameters can be computed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

$$\text{TPR} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \qquad (12)$$

$$\pounds(\Theta) = -[y_i \times \log p_i + (1-y_i) \times \log(1-p_i)] \qquad (13)$$

where y_i is the class label for input and p_i is the predicted probability of class being y_i.

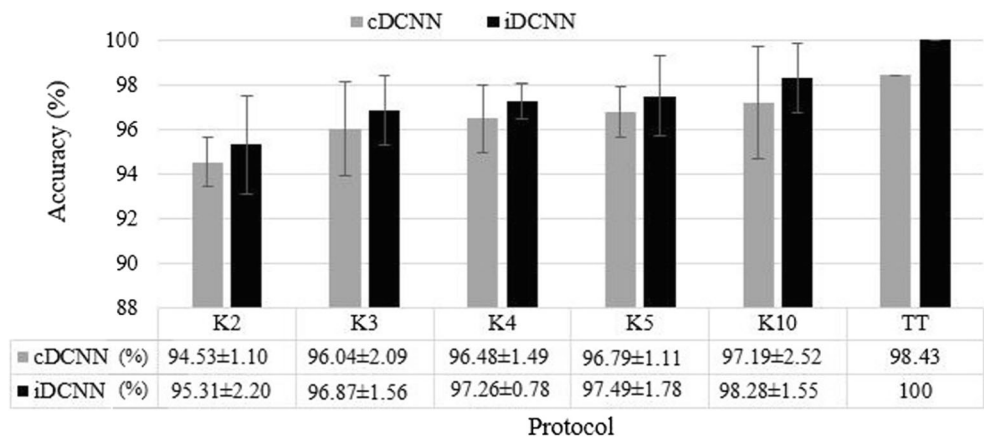$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}} \qquad (14)$$

where σ is the standard deviation and x_i is the accuracy at ith combination of K-fold cross-validation, μ is the mean accuracy, and N is the number of combinations, equal to 10 for K10.
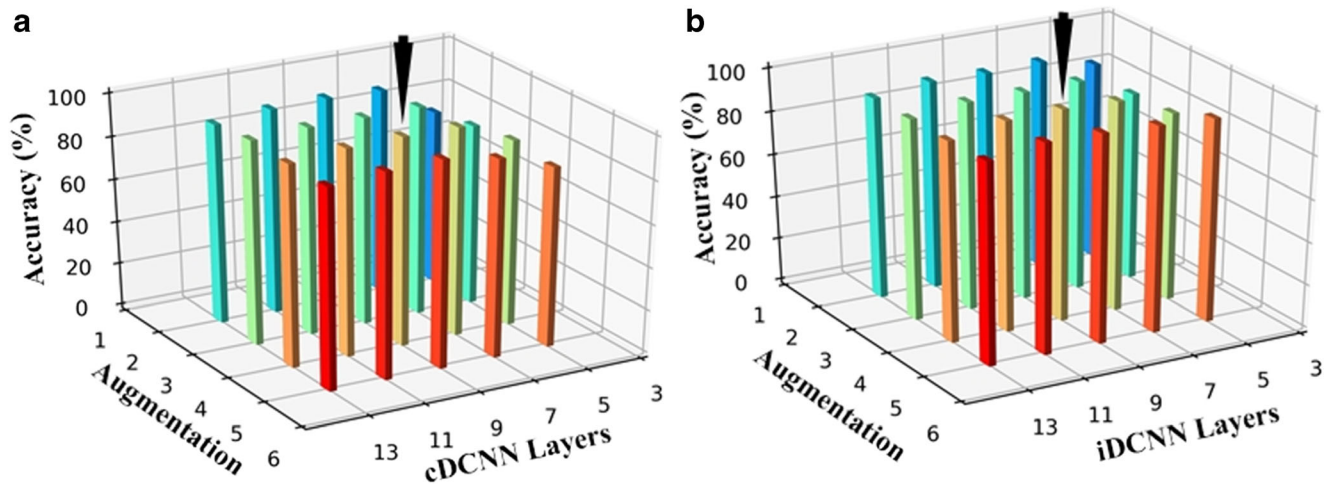
Fig. 14. **a** Three-dimensional optimization for best cDCNN layer and augmentation combination (cDCNN9*), where * shows the optimized CNN layer. The black arrow represents optimized value at cDL9A4. **b** Three-dimensional optimization for best iDCNN layer and augmentation combination (iDCNN9*), where * shows the optimized CNN layer. The black arrow represents the optimized value at iDL9A4.

Fig. 15 Comparison of cDCNN9* and iDCNN9* for different training protocols

	K2	K3	K4	K5	K10	TT
cDCNN (%)	94.53±1.10	96.04±2.09	96.48±1.49	96.79±1.11	97.19±2.52	98.43
iDCNN (%)	95.31±2.20	96.87±1.56	97.26±0.78	97.49±1.78	98.28±1.55	100

Protocol

Scientific validation of DCNN systems

The DCNN9*-Augm4* system was validated using well-accepted and well-published facial biometric data. The facial dataset consisted of 72 subjects, and each subject represented a class. Each subject had 20 different face images totaling to 1440 images in the dataset. Using K10 protocol, the accuracy obtained with cDCNN, tDCNN, and iDCNN was $96.72 \pm 2.01\%$, $97.18 \pm 1.23\%$, and $98.27 \pm 1.55\%$, respectively. These numbers are comparable with the accuracy obtained in WD. The order of performance was iDCNN > tDCNN > cDCNN. The results demonstrated the proposed DCNN methods as they promised encouraging high accuracy on an already published dataset. Table 11 shows the 10 K10 combinations of cDCNN, tDCNN, and iDCNN for the facial biometric dataset in sorted form.

Table 11 K10 performance of three DCNN models on the facial biometric dataset

		C1 Accuracy (%)	C2	C3
		cDCNN	tDCNN	iDCNN
R1	Comb 1	93.75	95.31	95.31
R2	Comb 2	95.31	95.31	96.87
R3	Comb 3	95.31	96.87	96.87
R4	Comb 4	95.31	96.87	98.43
R5	Comb 5	95.31	96.87	98.43
R6	Comb 6	96.87	96.87	98.43
R7	Comb 7	98.43	98.43	98.43
R8	Comb 8	98.43	98.43	100
R9	Comb 9	98.43	98.43	100
R10	Comb 10	100	98.43	100
R11	Mean	96.71 ± 2.01	97.18 ± 1.23	98.27 ± 1.55

The ML methods used in this study were also validated against well-published and well-accepted Cleveland clinic heart data in "csv" format. Using the K10 protocol, the accuracies corresponding to k-NN, DT, SVM, and RF were $84.52 \pm 0.04\%$, $0.8032 \pm 0.08\%$, $83.87 \pm 0.04\%$, and $84.19 \pm 0.07\%$, respectively.

Table 12 Symbol table

SN	Symbol	Description of the symbol
1	$£(\Theta)$	Loss of a network
2	p	Probability of image belonging to a class i
3	B	Bispectrum value
4	μ_C	Mean feature strength for control class
5	μ_{WD}	Mean feature strength of Wilson disease class
6	$\mathcal{F}$	Fourier transform
7	E	Expectation operator
8	DCNN	Deep convolution neural network
9	cDCNN	Conventional DCNN
10	iDCNN	Improved DCNN
11	tDCNN	Transfer learning–based DCNN
12	k-NN	k-nearest neighbor
13	DT	Decision tree
14	SVM	Support vector machine
15	RF	Random forest
16	K2	2-fold cross-validation
17	K3	3-fold cross-validation
18	K4	4-fold cross-validation
19	K5	5-fold cross-validation
20	K10	10-fold cross-validation
21	TT	Training and testing on same data
22	ML1	Machine learning 1
23	ML2	Machine learning 2
24	ML3	Machine learning 3
25	ML4	Machine learning 4
26	FC1	Feature combination 1
27	FC2	Feature combination 2
28	FC3	Feature combination 3
29	FC4	Feature combination 4

References

1. Dusek P, Litwin T, Czlonkowska A (2015) Wilson disease and other neurodegenerations with metal accumulations. Neurol Clin 33(1):175–204
2. Medici V, Rossaro L, Sturniolo G (2007) Wilson disease—a practical approach to diagnosis, treatment and follow-up. Dig Liver Dis 39(7):601–609
3. Rosencrantz R and Schilsky M (2011) "Wilson disease: pathogenesis and clinical considerations in diagnosis and treatment," in Seminars in liver disease 31(03):245–259. © Thieme Medical Publishers, pp.
4. Roberts EA, Schilsky ML (2003) A practice guideline on Wilson disease. Hepatology 37(6):1475–1492
5. Singh P, Ahluwalia A, Saggar K, Grewal CS (2011) Wilson's disease: Mri features. J Pediatr Neurosci 6(1):27
6. Parekh JR, Agrawal PR (2014) Wilson's disease: 'face of giant panda' and 'trident' signs together. Oxford Medical Case Reports 2014(1):16–17

7. Yousaf M, Kumar M, Ramakrishnaiah R, Vanhemert R, Angtuaco E (2009) Atypical MRI features involving the brain in Wilson's disease. Radiol Case Rep 4(3):312

8. El-Baz A, Suri JS (2011) Lung imaging and computer aided diagnosis. CRC Press

9. El-Baz A, Jiang X, Suri JS (2016) Biomedical image segmentation: advances and trends. CRC Press

10. Khanna NN, Jamthikar AD, Gupta D, Araki T, Piga M, Saba L, Carcassi C, Nicolaides A, Laird JR, Suri HS, Gupta A, Mavrogeni S, Protogerou A, Sfikakis P, Kitas GD, Suri JS (2019) Effect of carotid image-based phenotypes on cardiovascular risk calculator: Aecrs1. 0. Med Biol Eng Comput 57(7):1553–1566

11. Abiwinanda N, Hanif M, Hesaputra ST, Handayani A, and Mengko TR (2019) "Brain tumor classification using convolutional neural network," in World Congress on Medical Physics and Biomedical Engineering 2018. Springer, pp. 183–189.

12. Sharma AM, Gupta A, Kumar PK, Rajan J, Saba L, Nobutaka I, Laird JR, Nicolades A, Suri JS (2015) A review on carotid ultrasound atherosclerotic tissue characterization and stroke risk stratification in machine learning framework. Curr Atheroscler Rep 17(9): 55

13. Saba L, Tiwari A, Biswas M, Gupta SK, Godia-Cuadrado E, Chaturvedi A, Turk M, Suri HS, Orru S, Sanches JM et al (2019) Wilson's disease: a new perspective review on its genetics, diagnosis and treatment. Front Biosci (Elite edition) 11:166–185

14. Wee C-Y, Yap P-T, Zhang D, Wang L, Shen D (2014) Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Struct Funct 219(2):641–656

15. Prasad G, Joshi SH, Nir TM, Toga AW, Thompson PM, A. D. N. I. ADNI et al (2015) Brain connectivity and novel network measures for Alzheimer's disease classification. Neurobiol Aging 36:S121–S131

16. Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F, Initiative ADN et al (2015) Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Neurocomputing 151: 139–150

17. Affonso C, Rossi ALD, Vieira FHA, de Leon Ferreira ACP et al (2017) Deep learning for biological image classification. Expert Syst Appl 85:114–122

18. Baloglu UB, Talo M, Yildirim O, San Tan R, Acharya UR (2019) Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recogn Lett 122:23–30

19. Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Sharma A, Cuadrado-Godia E, Laird JR, Nicolaides A, Suri JS (2019) Deep learning fully convolution network for lumen characterization in diabetic patients using carotid ultrasound: a tool for stroke risk. Med Biol Eng Comput 57(2):543–564

20. Druzhkov P, Kustikova V (2016) A survey of deep learning methods and software tools for image classification and object detection. Pattern Recognit Image Anal 26(1):9–15

21. Skandha SS, Gupta SK, Saba L, Koppula VK, Johri AM, Khanna NN, Mavrogeni S, Laird JR, Pareek G, Miner M, Sfikakis PP, Protogerou A, Misra DP, Agarwal V, Sharma AM, Viswanathan V, Rathore VS, Turk M, Kolluri R, Viskovic K, Cuadrado-Godia E, Kitas GD, Nicolaides A, Suri JS (2020) 3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic™ 2.0. Comput Biol Med 125:103958

22. Saba L, Agarwal M, Sanagala S, Gupta S, Sinha G, Johri A, Khanna N, Mavrogeni S, Laird J, Pareek G et al (2020) Brain MRI-based Wilson disease tissue classification: an optimised deep transfer learning approach. Electronics Letters

23. Pareek G, Acharya UR, Sree SV, Swapna G, Yantri R, Martis RJ, Saba L, Krishnamurthi G, Mallarini G, El-Baz A et al (2013) Prostate tissue characterization/classification in 144 patient population using wavelet and higher order spectra features from transrectal ultrasound images. Technol Cancer Res Treat 12(6):545–557

24. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, and Wojna Z (2016) "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

25. Ferenci P (2006) Regional distribution of mutations of the atp7b gene in patients with Wilson disease: impact on genetic testing. Hum Genet 120(2):151–159

26. Vrabelova S, Letocha O, Borsky M, Kozak L (2005) Mutation analysis of the atp7b gene and genotype/phenotype correlation in 227 patients with Wilson disease. Mol Genet Metab 86(1-2):277–285

27. Saba L, Lucatelli P, Anzidei M, di Martino M, Suri JS, Montisci R (2018) Volumetric distribution of the white matter hyper-intensities in subject with mild to severe carotid artery stenosis: does the side play a role? J Stroke Cerebrovasc Dis 27(8):2059–2066

28. Saba L, Sanfilippo R, Porcu M, Lucatelli P, Montisci R, Zaccagna F, Suri JS, Anzidei M, Wintermark M (2017) Relationship between white matter hyperintensities volume and the circle of Willis configurations in patients with carotid artery pathology. Eur J Radiol 89:111–116

29. Porcu M, Balestrieri A, Siotto P, Lucatelli P, Anzidei M, Suri JS, Zaccagna F, Argiolas GM, Saba L (2018) Clinical neuroimaging markers of response to treatment in mood disorders. Neurosci Lett 669:43–54

30. Kim T, Kim IO, Kim WS, Cheon J-E, Moon S, Kwon J, Seo J, Yeon K (2006) MR imaging of the brain in Wilson disease of childhood: findings before and after treatment with clinical correlation. Am J Neuroradiol 27(6):1373–1378

31. Hu X, Chen S, Huang C-B, Qian Y, Yu Y (2017) Frequency-dependent changes in the amplitude of low-frequency fluctuations in patients with Wilson's disease: a resting-state fMRI study. Metab Brain Dis 32(3):685–692

32. Corgiolu S, Barberini L, Suri JS, Mandas A, Costaggiu D, Piano P, Zaccagna F, Lucatelli P, Balestrieri A, Saba L (2018) Resting-state functional connectivity MRI analysis in human immunodeficiency virus and hepatitis c virus co-infected subjects. a pilot study. Eur J Radiol 102:220–227

33. Porcu M, Wintermark M, Suri JS, Saba L (2020) The influence of the volumetric composition of the intracranial space on neural activity in healthy subjects: a resting-state functional magnetic resonance study. Eur J Neurosci 51(9):1944–1961

34. Kaden M, Riedel M, Hermann W, Villmann T (2015) Border-sensitive learning in generalized learning vector quantization: an alternative to support vector machines. Soft Comput 19(9):2423–2434

35. Jing R, Han Y, Cheng H, Han Y, Wang K, Weintraub D, Fan Y (2019) Altered large-scale functional brain networks in neurological Wilson's disease. Brain Imaging Behav:1–11

36. Shattuck DW, Leahy RM (2002) Brainsuite: an automated cortical surface identification tool. Med Image Anal 6(2):129–142

37. Manjón JV, Coupé P (2016) volBrain: an online MRI brain volumetry system. Front Neuroinform 10:30

38. Saba L, Biswas M, Kuppili V, Godia EC, Suri HS, Edla DR, Omerzu T, Laird JR, Khanna NN, Mavrogeni S et al (2019) The present and future of deep learning in radiology. Eur J Radiol 114: 14–24

39. Saba L, Biswas M, Suri HS, Viskovic K, Laird JR, Cuadrado-Godia E, Nicolaides A, Khanna N, Viswanathan V, Suri JS (2019) Ultrasound-based carotid stenosis measurement and risk stratification in diabetic cohort: a deep learning paradigm. Cardiovasc Diagn Ther 9(5):439

40. Biswas M, Kuppili V, Araki T, Edla DR, Godia EC, Saba L, Suri HS, Omerzu T, Laird JR, Khanna NN, Nicolaides A, Suri JS (2018)

Deep learning strategy for accurate carotid intima-media thickness measurement: an ultrasound study on Japanese diabetic cohort. Comput Biol Med 98:100–117

41. Biswas M, Saba L, Chakrabartty S, Khanna NN, Song H, Suri HS, Sfikakis PP, Mavrogeni S, Viskovic K, Laird JR, Cuadrado-Godia E, Nicolaides A, Sharma A, Viswanathan V, Protogerou A, Kitas G, Pareek G, Miner M, Suri JS (2020) Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: a screening tool for cardiovascular/stroke risk assessment. Comput Biol Med 123: 103847

42. Biswas M, Kuppili V, Edla DR, Suri HS, Saba L, Marinhoe RT, Sanches JM, Suri JS (2018) Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. Comput Methods Prog Biomed 155:165–177

43. Tandel GS, Biswas M, Kakde OG, Tiwari A, Suri HS, Turk M, Laird JR, Asare CK, Ankrah AA, Khanna N et al (2019) A review on a deep learning perspective in brain cancer classification. Cancers 11(1):111

44. Li Y and Yuan Y (2017) "Convergence analysis of two-layer neural networks with relu activation," in Adv Neural Inf Proces Syst, pp. 597–607.

45. Maniruzzaman M, Kumar N, Abedin MM, Islam MS, Suri HS, El-Baz AS, Suri JS (2017) Comparative approaches for classification of diabetes mellitus data: machine learning paradigm. Comput Methods Prog Biomed 152:23–34

46. Acharya UR, Faust O, Sree SV, Molinari F, Saba L, Nicolaides A, Suri JS (2011) An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans. IEEE Trans Instrum Meas 61(4):1045–1053

47. Suri JS, Kathuria C, and Molinari F (2010) Atherosclerosis disease management. Springer Science & Business Media.

48. Saba L, Jain PK, Suri HS, Ikeda N, Araki T, Singh BK, Nicolaides A, Shafique S, Gupta A, Laird JR et al (2017) Plaque tissue morphology-based stroke risk stratification using carotid ultrasound: a polling-based pca learning paradigm. J Med Syst 41(6):98

49. Acharya UR, Faust O, Alvin A, Krishnamurthi G, Seabra JC, Sanches J, Suri JS et al (2013) Understanding symptomatology of atherosclerotic plaque by image-based tissue characterization. Comput Methods Prog Biomed 110(1):66–75

50. Acharya UR, Mookiah MRK, Sree SV, Afonso D, Sanches J, Shafique S, Nicolaides A, Pedro LM, Fernandes JFE, Suri JS (2013) Atherosclerotic plaque tissue characterization in 2d ultrasound longitudinal carotid scans for automated classification: a paradigm for stroke risk assessment. Med Biol Eng Comput 51(5): 513–523

51. Molinari F, Mantovani A, Deandrea M, Limone P, Garberoglio R, Suri JS (2010) Characterization of single thyroid nodules by contrast-enhanced 3-d ultrasound. Ultrasound Med Biol 36(10): 1616–1625

52. Acharya UR, Swapna G, Sree SV, Molinari F, Gupta S, Bardales RH, Witkowska A, Suri JS (2014) A review on ultrasound-based thyroid cancer tissue characterization and automated classification. Technol Cancer Res Treat 13(4):289–301

53. Acharya UR, Molinari F, Sree SV, Swapna G, Saba L, Guerriero S, Suri JS (2015) Ovarian tissue characterization in ultrasound: a review. Technol Cancer Res Treat 14(3):251–261

54. Acharya UR, Sree SV, Kulshreshtha S, Molinari F, Koh JEW, Saba L, Suri JS (2014) Gynescan: an improved online paradigm for screening of ovarian cancer via tissue characterization. Technol Cancer Res Treat 13(6):529–539

55. Acharya UR, Sree SV, Ribeiro R, Krishnamurthi G, Marinho RT, Sanches J, Suri JS (2012) Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm. Med Phys 39(7Part1):4255–4264

56. Saba L, Dey N, Ashour AS, Samanta S, Nath SS, Chakraborty S, Sanches J, Kumar D, Marinho R, Suri JS (2016) Automated stratification of liver disease in ultrasound: an online accurate feature classification paradigm. Comput Methods Prog Biomed 130:118–134

57. Than JC, Saba L, Noor NM, Rijal OM, Kassim RM, Yunus A, Suri HS, Porcu M, Suri JS (2017) Lung disease stratification using amalgamation of Riesz and Gabor transforms in machine learning framework. Comput Biol Med 89:197–211

58. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS (2016) Computer-aided diagnosis of psoriasis skin images with hos, texture and color features: a first comparative study of its kind. Comput Methods Prog Biomed 126:98–109

59. Wu DH, Chen Z, North JC, Biswas M, Vo J, Suri JS (2020) Machine learning paradigm for dynamic contrast-enhanced MRI evaluation of expanding bladder. Front Biosci (Landmark Edition) 25:1746–1764

60. Acharya UR, Sree SV, Krishnan MMR, Krishnananda N, Ranjan S, Umesh P, Suri JS (2013) Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images. Comput Methods Prog Biomed 112(3): 624–632

61. Jamthikar AD, Gupta D, Mantella LE, Saba L, Laird JR, Johri AM, Suri JS (2020) Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. Int J Cardiovasc Imaging:1–17

62. Jamthikar A, Gupta D, Khanna NN, Saba L, Laird JR, Suri JS (2020) Cardiovascular/stroke risk prevention: a new machine learning framework integrating carotid ultrasound image-based phenotypes and its harmonics with conventional risk factors. Indian Heart J 72(4):258–264

63. Jamthikar A, Gupta D, Khanna NN, Saba L, Araki T, Viskovic K, Suri HS, Gupta A, Mavrogeni S, Turk M et al (2019) A low-cost machine learning-based cardiovascular/stroke risk assessment system: integration of conventional factors with image phenotypes. Cardiovasc Diagn Ther 9(5):420

64. Khanna NN, Jamthikar AD, Gupta D, Piga M, Saba L, Carcassi C, Giannopoulos AA, Nicolaides A, Laird JR, Suri HS et al (2019) Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning–based tissue characterization. Curr Atheroscler Rep 21(2):7

65. Banchhor SK, Londhe ND, Araki T, Saba L, Radeva P, Laird JR, Suri JS (2017) Wall-based measurement features provides an improved IVUS coronary artery risk assessment when fused with plaque texture-based features during machine learning paradigm. Comput Biol Med 91:198–212

66. Suri JS, "Imaging based symptomatic classification and cardiovascular stroke risk score estimation," Oct. 20 2011, uS Patent App. 13/053,971.

67. Cuadrado-Godia E, Dwivedi P, Sharma S, Santiago AO, Gonzalez JR, Balcells M, Laird J, Turk M, Suri HS, Nicolaides A et al (2018) Cerebral small vessel disease: a review focusing on pathophysiology, biomarkers, and machine learning strategies. Journal of Stroke 20(3):302

68. Martis RJ, Acharya UR, Prasad H, Chua CK, Lim CM, Suri JS (2013) Application of higher order statistics for atrial arrhythmia classification. Biomed Signal Proces Control 8(6):888–900

69. Maniruzzaman M, Rahman MJ, Ahammed B, Abedin MM, Suri HS, Biswas M, El-Baz A, Bangeas P, Tsoulfas G, Suri JS (2019) Statistical characterization and classification of colon microarray gene expression data using multiple machine learning paradigms. Comput Methods Prog Biomed 176:173–193

70. Kuppili V, Biswas M, Sreekumar A, Suri HS, Saba L, Edla DR, Marinhoe RT, Sanches JM, Suri JS (2017) Extreme learning

machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization. J Med Syst 41(10):152

71. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56(11):1129–1135

72. Kadam P, Bhalerao S (2010) Sample size calculation. Int J Ayurveda Res 1(1):55

73. Araki T, Ikeda N, Shukla D, Jain PK, Londhe ND, Shrivastava VK, Banchhor SK, Saba L, Nicolaides A, Shafique S, Laird JR, Suri JS (2016) PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology. Comput Methods Prog Biomed 128:137–158

74. Suk H-I, Lee S-W, Shen D, Initiative ADN et al (2017) Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal 37:101–113

75. Zhang Y, Zhang H, Chen X, Liu M, Zhu X, Lee S-W, Shen D (2019) Strength and similarity guided group-level brain functional network construction for mci diagnosis. Pattern Recogn 88:421–430

76. Abrol A, Bhattarai M, Fedorov A, Du Y, Plis S, Calhoun V, Initiative ADN et al. (2020) "Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease, " J Neurosci Methods, p. 108701.

77. Richhariya B, Tanveer M, Rashid A, Initiative ADN et al (2020) Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomed Signal Proces Control 59:101903

78. Liu J, Pan Y, Wu F.-X, and Wang J (2020) "Enhancing the feature representation of multi-modal MRI data by combining multi-view information for mci classification," Neurocomputing

79. Tandel GS, Balestrieri A, Jujaray T, Khanna NN, Saba L, and Suri JS (2020) "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," Computers in Biology and Medicine, p. 103804.

80. El-Baz A, Gimel'farb G, Suri JS (2015) Stochastic modeling for medical image analysis. CRC Press

81. Suri JS, Wilson DL, and Laxminarayan S (2005) Handbook of Biomedical Image Analysis: Segmentation models part B. Kluwer Academic/Plenum Publishers

82. Suri JS, Wilson D, Laxminarayan S (2005) Handbook of biomedical image analysis. Springer Science & Business Media 2

83. Narayanan R, Kurhanewicz J, K. Shinohara, E. D. Crawford, A. Simoneau, and J. S. Suri (2009) "MRI-ultrasound registration for targeted prostate biopsy," in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Mael2019neurologicalcro. IEEE, pp. 991–994.

84. El-Baz A and Suri JS (2019) "Neurological disorders and imaging physics, volume 2; engineering and clinical perspectives of multiple sclerosis," ndi2

85. Acharya R, Ng YE, and J. S. Suri (2008) Image modeling of the human eye. Artech House.

86. Molinari F, Liboni W, Giustetto P, Badalamenti S, Suri JS (2009) Automatic computer-based tracings (act) in longitudinal 2-d ultrasound images using different scanners. J Mech Med Biol 9(04): 481–505

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mohit Agarwal received his B.Tech from IIT Delhi and MTech from ABES, Ghaziabad, in 1995 and 2016, respectively. He is veteran in AI programming by gathering massive experience from MNC's in India and USA with concentration in Python and C#.NET. Currently, he is pursuing his PhD from Bennett University, India. His research interests include Artificial Intelligence and healthcare

Luca Saba, MD, is currently Professor and Head of Radiology at A.O.U., Cagliari, Italy. His research interests are in computed tomography, magnetic resonance, ultrasound, neuroradiology, diagnostic in vascular sciences, and lung imaging. Dr. Saba is authored 12 books and has H-index of 40. He has authored more than 370 peer-reviewed papers and is a frequently invited as guest speaker in radiological and cardiovascular conferences.

Suneet K. Gupta, PhD, has been working as a Professor in the Department of Computer Science Engineering at Bennett University, Gr. Noida since 2017. His current research interests are wireless sensor network, Internet of things, natural language processing, and optimization of neural network. Recently, He has completed a project funded by Council of Science and Technology, UP, India.

Amer M. Johri, MD, working as an assistant professor in Division of Cardiology, Undergraduate Medical Education, Cardiovascular Imaging Network at Queen's, Queen's University, Kingston, Ontario.

John R. Laird, MD, FACC, is with St. Helena Hospital, CA, USA. Professor Laird is an internationally renowned interventional cardiologist, and his expertise is innovative procedures for carotid artery disease.

Narendra N. Khanna, MD, DM, FACC, is Advisor to Apollo Group of Hospitals in India and is working as Senior Consultant in Cardiology & Coordinator of Vascular Services at Indraprastha Apollo Hospital, New Delhi.

Gyan Pareek, MD, is a director of the Minimally Invasive Urology Institute and director of the Kidney Stone Center at The Miriam Hospital. He is also a professor of surgery (urology) at The Warren Alpert Medical School of Brown University. His areas of expertise include kidney stones, prostate cancer, and benign prostatic hyperplasia. Dr. Pareek is board certified in urology and is a fellow of the American College of Surgeons. He is advisor to AtheroPoint, CA, and Global Biomedical Technologies, CA, USA.

Sophie Mavrogeni, MD, PhD, currently working at Cardiology Clinic, Onassis, Athens, Greece. Her research is focused on nonischemic cardiomyopathy, dystrophinopathies, myocarditis, and rheumatic diseases.

Martin M. Miner, MD, has practiced preventive and primary care medicine for more than 28 years and is currently an internist at the Miriam Hospital. He is a fellow of the American Urological Association and the author of more than 75 publications in the areas of erectile dysfunction and cardiovascular disease, benign prostatic hyperplasia, and lower urinary tract symptoms in reference to male sexuality, and hormonal replacement therapy in men. Miner is president-elect of the American Society of Men's Health, associate editor of the *Journal of Men's Health*, and serves on multiple journal boards and reviews for several publications.

Petros P. Sfikakis, MD, is the Dean of the School of Medicine at National and Kapodistrian University of Athens, Greece, and a Professor of Internal Medicine and Rheumatology. One of his main research interests focus in the cardiovascular outcomes of patients with immune-mediated diseases.

Vijay Viswanathan, MD, PhD, FRCP, is currently working as a chairman and director of MV Hospital for Diabetes and M Viswanathan Diabetes Research Centre, Chennai, India. His Research interests are endocrinology diabetes and metabolism, diabetes, and foot research.

Athanasios Protogerou, MD, PhD, currently working at the Department of Cardiovascular Prevention & Research Unit Clinic, National and Kapodistrian Univ. of Athens, Greece. His research is focused on cardiovascular prevention.

George D. Kitas, MD, PhD, FRCP, is director of Research & Development-Academic Affairs, Dudley Group NHS Foundation Trust, Dudley, UK. He is an Honorary Professor of Rheumatology at the Arthritis Research UK Epidemiology Unit. He has an H-index of 81.

Aditya Sharma, MD, is a cardiologist at University of Virginia, where he directs the anticoagulation clinic and the medical optimization clinic, which helps patients with vascular disorders and manages their risk factors with medication. His research interests are peripheral arterial disease, venous thromboembolism, and fibromuscular dysplasia.

Andrew Nicolaides, MS, FRCS, PhD (Hon), is currently the professor emeritus at Imperial College, London. He is the co-author of more than 500 original papers and editor of 14 books and has an H-index of 104.

Jasjit S. Suri, PhD, MBA, is an innovator, visionary, scientist, and an internationally known world leader in field of Biomedical Engineering and its Management. Dr. Suri received the (i) *Director General's Gold medal* in 1980 and is a Fellow of: (ii) *American Institute of Medical and Biological Engineering* (FAIMBE), awarded by National Academy of Sciences, Washington DC, (iii) *American Institute of Ultrasound in Medicine* (FAIUM), (iv) *Asia Pacific Vascular Society* (FAPVS), and (v) *the Institute of Electrical and Electronics Engineers* (FIEEE). He is also a recipient of *Life Time Achievement Award* from Marquis. He is currently Chairman of AtheroPoint, Roseville, CA, USA, dedicated to imaging technologies for cardiovascular and stroke. He has nearly ~ 19,000 citations, co-authored 50 books, over 50 patents, and has an H-index of 66.