





Deep Learning Approach to Classify Parkinson's Disease from MRI Samples

Nanziba Basnin¹ , Nazmun Nahar² , Fahmida Ahmed Anika³ ,
Mohammad Shahadat Hossain⁴ , and Karl Andersson⁵ 

¹ International Islamic University Chittagong, Chittagong, Bangladesh

² BGC Trust University Bangladesh Bidyanagar, Chandanaish, Bangladesh

³ Independent University Bangladesh, Dhaka, Bangladesh

⁴ University of Chittagong, Chittagong, Bangladesh

⁵ University of Technology, Skelleftea, Sweden

Abstract. Parkinson's disease is a progressive degenerative disorder that comes from a recognized clinical parkinsonian syndrome. The manifestations of Parkinson's disease include both motor and nonmotor symptoms identified as tremor, bradykinesia (slowed movements), rigidity, and postural instability. PD is marked as one of the most prevalent disorders from various researches and surveys because it has been observed in 90% of people out of 100. It is imperative to design CAD to develop an advanced model for the determination of this disease with accuracy since up to date there is no accurate clinical intervention for the diagnosis of PD. In contrast to conventional methods. Deep learning convolutional neural network tools are implied for the faster and accurate identification of PD through MRI. The purpose of this research is to contribute to the development of an accurate PD detection method. To conduct the research a public dataset NTU (National Technical University of Athens) is used. The data samples are categorized into three sets (Training, Test, and Validation). A DenseNet integrated with LSTM is applied to the MRI data samples. DenseNet is used to strengthen the feature selection ability, as each layer selects features depending on the temporal closeness of the image. The output is then fed into the LSTM layer, for discovering the significant dependencies in temporal features. The performance of the proposed DenseNet-LSTM is compared to other CNN state-of-the-art models. The proposed model outputs a training accuracy of 93.75%, testing accuracy of 90%, and validation accuracy of 93.8% respectively.

Keywords: Parkinson's Disease · CNN state-of-the-art-model · DenseNet-LSTM · MRI data samples

1 Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder that is the primary stage of parkinsonism, (an idiopathic chronic disorder) usually predominant in the elderly. Approximately 1% of the world population encounters PD, From

which the majority of cases manifested complicated motor and cognitive issues. As the disease starts to develop, a prodromal phase of cognitive and behavioral symptoms inclusive of various personality changes, depressive disorders, memory dysfunction, and emotion dysregulation may become apparent. To date, clinicians still follow conventional methods to diagnose PD based on the prodromal symptoms mentioned above including slowness, stiffness, tremor, and movement difficulties. However, these symptoms tend to vary according to each individual being affected. Hence, at present, there is no particular blood test or biomarker to diagnose PD meticulously or detect prime changes as the condition rises. Researchers have perceived with the provision that changes in disease progression can be detected by brain MRI under a distinctive protocol.

PD in the primary stage can be misdiagnosed as other syndromes. PD is a most remarkable disease to be misinterpreted with abundant atypical parkinsonian disorders (APDs). For instance, progressive supranuclear palsy (PSP), multiple system atrophy (MSA), specifically the Parkinson variant of multiple system atrophy (MSA-P), and corticobasal degeneration (CBD). In addition to this, categorizing clinical entities of PD is difficult. Furthermore, classification of APDs and PD is imperative for accurate diagnosis during the prodromal phase as well as for choosing a specific intervention. MRI, a revolutionary and widespread method used in modern diagnosis, has proven to be significant for detecting PD. In general, MRI images undergo distributional differences in image resolution, contrast, signal to noise ratio due to hardware acquisitions, or techniques used in scanning. All these factors impact the composition of such images, in turn influencing the performance of training models.

In this research, we are adapting Deep Learning methods because it authorizes a higher quality learning representation of MRI image data [18, 21] enabling multiple levels of abstraction, achieved through multiple processing of layers. Thus providing an optimal rate of model performance in comparison to other learning approaches. The dataset selected for this study is a public dataset NTUA (National Technical University of Athens), which is split into training, testing, and validation sets. A DenseNet Model is integrated with LSTM to train the MRI data samples. The use of learning parameters is significantly reduced in DenseNet, as the input in each layer is provided as output in the next layer. Enabling reuse, and smooth flow of prominent features throughout the layer without any loss. Further, within each layer, some mechanism exists that select features according to their temporal relatedness apt to predicting class. These selected features are passed through the LSTM layer where their temporal dependencies are discovered. This exposes the model to more complex temporal dynamics of imaging, enhancing the classification and learning of the model [11]. In addition to this LSTM handles back prop and vanishing gradient problems [3] to enable effective training of datasets. The proposed model is further compared to other CNN state-of-the-art models, namely, DenseNet, VGG19, InceptionV3, ResNet, MobileNet for further evaluation.

Thus, through this section, the motive behind this research is understood. In the following Sect. 2, other works related to this research are assessed. Section 3

illustrated the methodology. Section 4 investigates and evaluates the proposed model. Section 5 outlines the conclusion and future prospects. While Sect. 6 provides acknowledgment.

2 Literature Review

Ref. No.	Methodology	Accuracy	Limitation
[26]	A CAD-based CNN model for auto-diagnosis of PD is proposed. A T2-weighted Magnetic Resonance Imaging (MRI) data samples are used in training the model. The model is compared to other ML models	96%	The complex manifestations of MRI data make it difficult to select the appropriate structure of CNN
[14]	T1-weighted sMRI scans from 416 patients are included to build the OASIS 3D dataset. A 2D architecture of 3 layers of distinct configuration of CNN is proposed to build the classification system for Alzheimer's. It is compared with other CNN state-of-the-art models	93.18%	The model may stop working if the vanishing gradient is relatively small, preventing the weights from getting updated
[28]	This is a review study consisting of some experiments with different layers of CNN, CNN-RNN, DNN. It was conducted to survey different types of datasets (image, speech handwriting, sensor, DATscan, MRI) on ML and DL models. Neurodegenerative disorders such as Alzheimer's and PD were investigated	96% for CNN-LSTM	Retraining the DNN model can lead to unforgettable prior knowledge
[19]	A classification model is generated to dissimilate PD-induced olfactory dysfunction from other types of non-Parkinsonian olfactory dysfunction (NPOD) from T1-weighted axial MRI samples of 30 patients. A 4 layer of CNN architecture along with other ML models are applied to train the data samples	96.6%	The size of MRI samples are small in order to reduce bias. The proposed model suffers from the problem of overfitting
[23]	A classification model is generated to dissimilate PD-induced olfactory dysfunction from other types of non-Parkinsonian olfactory dysfunction (NPOD) from T1-weighted axial MRI samples of 30 patients. A 4 layer of CNN architecture along with other ML models are applied to train the data samples	96.6%	The size of MRI samples are small in order to reduce bias

3 Methodology

Figure 1 demonstrated the methodology which is undertaken to carry out this research. In the beginning, the dataset is preprocessed by applying normalization and other perimeters found in data augmentation, followed by segmentation using K-means algorithm. Afterwards, the dataset is split into three parts. This dataset is then fed into the training model. The training model is labeled and used to predict PD from a sample of test datasets.

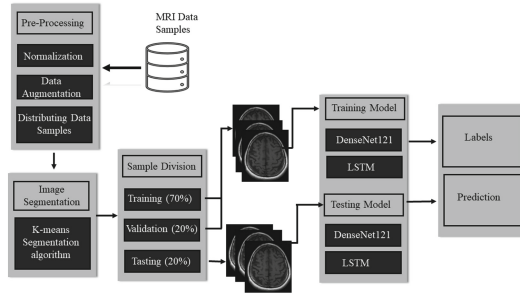


Fig. 1. Methodology

3.1 MRI Data Samples

This dataset comprises MRI examination report samples from a total of 78 individuals, 55 of which are suffering from Parkinson’s Disease, 23 are healthy and serve as control subjects. This dataset is made accessible for the public, which includes the epidemiological, clinical and paraclinical data sample of patients, and is named as the ‘National Technical University of Athens (NTUA) Parkinson’s Dataset’. This dataset contains T1, T2, and Flair MRI image samples where the frames per sequence and the resolution differ for individual images. Although the images were derived in DICOM format, which is standard in medical imaging, the images were published in PNG format for efficient storage (Fig. 2).

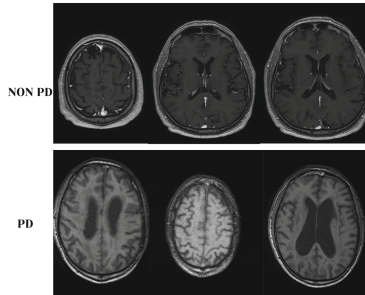


Fig. 2. Sample of NTUA MRI

The total number of data samples is 1387, of which 472 are NON PD patients and 915 PD patients. The samples were divided into three sets, Test, Training, and Validation samples respectively [28]. The Table 1 represents how the samples are distributed into sets.

Table 1. Distribution of MRI data samples

Test		Train		Validation	
Non PD	PD	Non PD	PD	Non PD	PD
35	51	369	605	68	259

3.2 Pre-processing of MRI Data Samples

In the beginning the MRI image data samples are preprocessed with an ImageDataGenerator class available in keras. The ImageDataGenerator class provides access to a wide range of pixel scaling methods and data augmentation techniques. The ImageDataGenerator class comprises three prominent methods of pixel scaling procedures. Here, the Normalization method is implied to re-scale the pixel values from a range of 0–255 to a range of 0–1. The rescale argument is invoked in order to carry out normalization of the image dataset before feeding it to the neural network for training [24]. The data augmentation increases the data quality and its robustness. Parameters such as shifting, rotation, shear and zoom, and flip are instigated. Moreover, augmented datasets aid in increasing data points by reducing the length between the testing and training dataset. Hence preventing the overfitting in the training dataset [27].

3.3 MRI Sample Segmantation

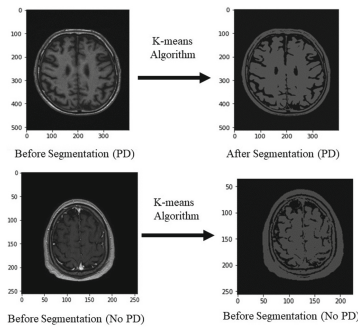


Fig. 3. Sample of MRI segmented

The MRI samples are converted to LAB color space. The adaptive K-means algorithm is employed to segment each image by a value of K which falls between a cycle two to ten. Secondly, the morphological operation is applied to transform the image into two values. Lastly, the maximum threshold is registered by iterating the process, to obtain the segmentation outcome. If the number of the outcome matches the value of K in that instant, a stop iteration command is invoked and the outcome of division becomes the final outcome.

3.4 Proposed Model Architecture

Table 2. System architecture description.

Layers	Output size
Convolution	112×112
Pooling	56×56
Dense Block (1)	56×56
Transitional Block (1)	56×56
Dense Block (2)	28×28
Transitional Block (2)	28×28
Dense Block (3)	14×14
Transitional Block (3)	14×14
Dense Block	7×7
ReLu	-
LSTM	(None, 148, 70)
Global Average Pooling	(None, 1024)
LSTM	(None, 70)
Concatenate	(None, 1024)
Dense	(None, 1)

DenseNet is a state-of-the-art which is the most suggested convolutional neural networks, because it can join both the previous and updraught layers. The structure offers advantages superior to existing structures like mitigating the vanishing gradient drawback, reinforcing propagation of feature, enables reuse of features, and turn down the number of parameters. In general, deep DenseNets are sets of dense blocks which is associated consecutively, with subsidiary convolutional and operations of pooling allying consecutive dense blocks. This development allows to construct a deep neural network able to constitute difficult transformations [13]. Thus, the task consists of two major obstructions: (i) Invoking CNN to deal with image sequences is inapt as CNN formerly established for static data. And leads to obtaining features from the image. These features are provided to the LSTM as it is capable of dealing with sequence data. Hence it is used to acquire image-sequence differentiation. The real image sequences are noisy and high dimensional, so it provides poor results when fed onto the LSTM Model (Table 2).

3.5 System Implementation

Google Colab was employed to generate the training module [7]. This platform expedites the implementation of programs in run time and supports deep learning libraries by providing access to robust GPU (Graphics Processing Unit) and TPU (Tensor Processing Unit). Python is compatible with the Google Colab

environment. To build the programs, libraries obligatory for implementation include Tensor flow, Keras, OpenCV, PIL, sklearn, Matplotlib, Pandas, and Numpy. In this study, Tensor was implied for the backend of the system and Keras was employed to build the pre-trained models of CNN [1] because it can aid in built-in functions for the purpose of layering, optimization, and activation. OpenCV is vital for image processing [25]. Conversely, Sklearn provides access to various supervised and unsupervised algorithms [12]. A confusion matrix was implemented to build Matplotlib [12]. Image processing was carried out by the utilization of an integration tool PIL, whereas Numpy was employed to aid in the operations of arrays [10]. Callbacks are implemented to train the model. The advantage of using Callbacks includes not only overfitting which is the result of the occurrence of numerous epochs as well as circumvents under models. Check points, early stopping and lowering the learning rate on a plateau are utilized in callbacks. Check-points aid in the preservation of the best models by inspection of the loss within validation. Early stopping halts the training epochs, once the model exhibits no significant change in model performance. When the validation loss declines for any further enhancement, a Reducing learning rate on the plateau is invoked.

4 Experimental Result

This section evaluates the performance of the models when trained using the MRI data samples for the classification of PD.

4.1 Comparison of Results

After the MRI data samples are pre-processed, they are fed into state-of-the-art CNN models. These models are, namely, DenseNet, VGG19, ResNet, MobileNet, and Inception V3. It is observed that VGG19, ResNet, MobileNet have a training accuracy of 1.0 with very insignificant loss, in contrast, their testing accuracy is 0.81, 0.93, and 0.84. Whereas, these models have a validation accuracy of 0.82, 0.96, and 0.94 respectively. Since the training accuracy is 100%, the model is likely to be an overfitted model. In order to justify this cross-validation needs to be carried out for VGG-19, ResNet, and MobileNet.

Table 3. Comparison of result

Transfer learning model	Training accuracy	Training loss	Testing accuracy	Testing loss	Validation accuracy	Validation loss
DenseNet	0.93	0.061	0.87	0.10	0.94	0.023
VGG19	1.0	0.095	0.8125	0.07	0.828	0.114
ResNet	1.0	0.00028	0.93	0.07	0.96	0.038
MobileNet	1.0	0.0018	0.84	0.114	0.94	0.067
InceptionV3	0.95	0.038	0.843	0.112	0.90	0.0431

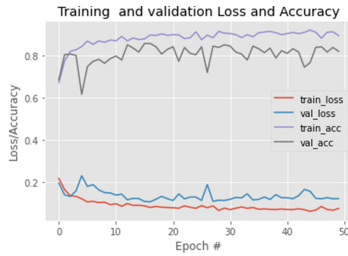


Fig. 4. Performance curve of InceptionV3

On the other hand, InceptionV3 has a training accuracy of 0.95, a testing accuracy of 0.84, and a validation accuracy of 0.94. Since there is an 11% difference between the training and testing. Furthermore, it can be said that the model suffers from overfitting. In the case of DenseNet, the model has a relatively close training and testing accuracy as well as shows little or no significant difference between the validation and training accuracy. Thus, this model can be taken into account for further experimenting.

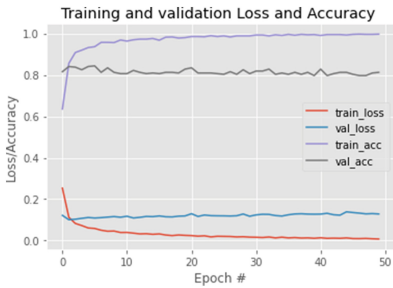


Fig. 5. Performance curve of DenseNet

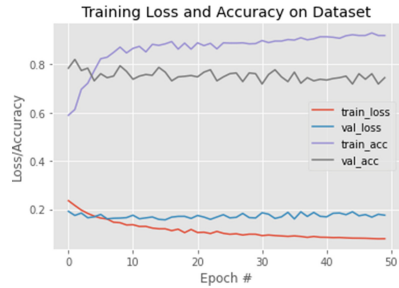


Fig. 6. Performance curve of VGG19

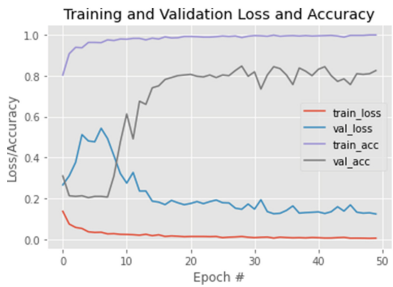


Fig. 7. Performance curve of ResNet

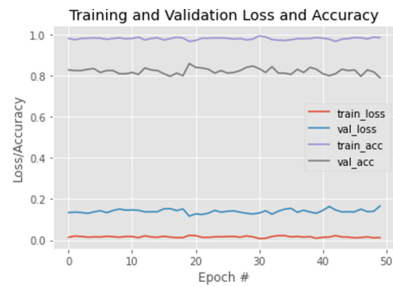


Fig. 8. Performance curve of MobileNet

Figures 3, 4, 5, and, 6 graphically represents the performance curve of all the state of the art model applied to train the MRI samples for the classification of PD (Fig. 7).

4.2 Proposed Model Result

Table 3, demonstrates the training accuracy, testing accuracy as well as validation accuracy of the model. It is observed that the model is a well-fit model since there is not any significant differences between, testing, training and validation. Figures 8 and 9 represents the Model accuracy and that Model loss. In Fig. 8 the red line depicts the training accuracy which is observed to increase after 20 epochs while the blue line depicts the validation accuracy. Both the red line and the blue line is observed to meet at a particular point after 30 epochs. On the other hand, in Fig. 9 the blue line represents the validation loss which is slightly less than the training loss represented by the red line .

4.3 Performance Matrix of Various Deep Learning Models

To evaluate each models used in this research, the performance matrices such as precision, recall, f1-score, AUC are used (Tables 4 and 5).

Table 4. DenseNet-LSTM result

Training accuracy	Testing accuracy	Validation accuracy	Training loss	Testing loss	Validation loss
93.75%	90.0%	93.8%	0.06	0.11	0.06

Table 5. Comparison of performance matrices

Model	Precision	Recall	f1-score	AUC
DenseNet121	0.865	0.902	0.798	0.907
VGG16	0.793	0.886	0.798	0.829
ResNet	0.881	0.769	0.798	0.882
MobileNet	0.860	0.902	0.802	0.902
InceptionV3	0.817	0.844	0.798	0.827

5 Conclusion and Future Work

This research is aimed at the development of a method that will help to determine and differentiate PD with precision as well as without the interference of traditional approaches. To detect PD early, it is crucial to use an application that can segment the data sample image for better visualization and image resolution. With the emergence of various applications, in the modern era, MRI

plays an important role as it can aid in the segmentation of images for a vivid understanding of the target affected side, leading to the recognition of PD while in the prodromal stage. To build an optimal learning model, DenseNet integrated with the LSTM model is trained with the publicly available NTUA dataset. This model is deployed for this research because it is better at learning and storing significant features within layers. The LSTM layers benefit by sorting the features according to their dependencies to be more relatable to predictable targets. The performance is further evaluated against other CNN state-of-the-art models in order to find which model is more appropriate for the data sample. However, the downside of Deep learning approaches adhered to in this research is that it is inadequate for in-depth comprehension as these models used are similar to a “Black-box” [17]. So to remove prior intuitive interpretations, explainable Artificial Intelligence can be introduced in near future. Further, this model can be enhanced to classify other syndromes like PD and we use other model to predict PD[?] [2, 4–6, 8, 9, 14, 15, 20, 22, 29].

References

1. Abadi, M., et al.: TensorFlow: a system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283 (2016)
2. Abedin, M.Z., Akther, S., Hossain, M.S.: An artificial neural network model for epilepsy seizure detection. In: 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), pp. 860–865. IEEE (2019)
3. Ahmed, T.U., Hossain, M.S., Alam, M.J., Andersson, K.: An integrated CNN-RNN framework to assess road crack. In: 2019 22nd International Conference on Computer and Information Technology (ICCIT). pp. 1–6. IEEE (2019)
4. Ahmed, T.U., Jamil, M.N., Hossain, M.S., Andersson, K., Hossain, M.S.: An integrated real-time deep learning and belief rule base intelligent system to assess facial expression under uncertainty. In: 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 1–6. IEEE (2020)
5. Basnin, N., Nahar, L., Hossain, M.S.: An integrated CNN-LSTM model for micro hand gesture recognition. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) ICO 2020. AISC, vol. 1324, pp. 379–392. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68154-8_35
6. Basnin, N., Nahar, L., Hossain, M.S.: An integrated CNN-LSTM model for Bangla lexical sign language recognition. Proceedings of International Conference on Trends in Computational and Cognitive Engineering ,pp. 695–707 (2021). https://doi.org/10.1007/978-981-33-4673-4_57
7. Carneiro, T., Da Nóbrega, R.V.M., Nepomuceno, T., Bian, G.B., De Albuquerque, V.H.C., Reboucas Filho, P.P.: Performance analysis of google colab as a tool for accelerating deep learning applications. IEEE Access **6**, 61677–61685 (2018)
8. Chowdhury, R.R., Hossain, M.S., Hossain, S., Andersson, K.: Analyzing sentiment of movie reviews in Bangla by applying machine learning techniques. In: 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), pp. 1–6. IEEE (2019)

9. Chowdhury, R.R., Hossain, M.S., ul Islam, R., Andersson, K., Hossain, S.: Bangla handwritten character recognition using convolutional neural network with data augmentation. In: 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 318–323. IEEE (2019)
10. Diez, P.: *Smart Wheelchairs and Brain-Computer Interfaces: Mobile Assistive Technologies*. Academic Press, London (2018)
11. Donahue, J., et al.: Long-term recurrent convolutional networks for visual recognition and description. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2625–2634 (2015)
12. Greenfield, P., Miller, J.T., Hsu, J., White, R.L.: *numarray: a new scientific array package for python*. PyCon DC (2003)
13. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)
14. Islam, J., Zhang, Y.: Brain MRI analysis for Alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inf.* **5**(2), 1–14 (2018)
15. Islam, R.U., Hossain, M.S., Andersson, K.: A deep learning inspired belief rule-based expert system. *IEEE Access* **8**, 190637–190651 (2020)
16. Islam, R.U., Ruci, X., Hossain, M.S., Andersson, K., Kor, A.L.: Capacity management of hyperscale data centers using predictive modelling. *Energies* **12**(18), 3438 (2019)
17. Mahmud, M., Kaiser, M.S., McGinnity, T.M., Hussain, A.: Deep learning in mining biological data. *Cogn. Comput.* **13**(1), 1–33 (2021)
18. Mahmud, M., Kaiser, M.S., Hussain, A., Vassanelli, S.: Applications of deep learning and reinforcement learning to biological data. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(6), 2063–2079 (2018)
19. Mei, J., Tremblay, C., Stikov, N., Desrosiers, C., Frasnelli, J.: Differentiation of Parkinson’s disease and non-parkinsonian olfactory dysfunction with structural MRI data. In: *Medical Imaging 2021: Computer-Aided Diagnosis*. vol. 11597, p. 115971E. International Society for Optics and Photonics (2021)
20. Nahar, N., Hossain, M.S., Andersson, K.: A machine learning based fall detection for elderly people with neurodegenerative disorders. In: Mahmud, M., Vassanelli, S., Kaiser, M.S., Zhong, N. (eds.) *BI 2020. LNCS (LNAI)*, vol. 12241, pp. 194–203. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59277-6_18
21. Noor, M.B.T., Zenia, N.Z., Kaiser, M.S., Al Mamun, S., Mahmud, M.: Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer’s disease, Parkinson’s disease and schizophrenia. *Brain Inform.* **7**(1), 1–21 (2020)
22. Pathan, R.K.: Gender classification from inertial sensor-based gait dataset. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) *ICO 2020. AISC*, vol. 1324, pp. 583–596. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68154-8_51
23. Pugalenth, R., Rajakumar, M., Ramya, J., Rajinikanth, V.: Evaluation and classification of the brain tumor MRI using machine learning technique. *J. Control Eng. Appl. Inform.* **21**(4), 12–21 (2019)
24. Reisfeld, D., Yeshurun, Y.: Preprocessing of face images: detection of features and pose normalization. *Comput. Vis. Image Understand.* **71**(3), 413–430 (1998)
25. Samal, B., Behera, A.K., Panda, M.: Performance analysis of supervised machine learning techniques for sentiment analysis. In: *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)*, pp. 128–133. IEEE (2017)

26. Shah, P.M., Zeb, A., Shafi, U., Zaidi, S.F.A., Shah, M.A.: Detection of Parkinson disease in brain MRI using convolutional neural network. In: 2018 24th International Conference on Automation and Computing (ICAC), pp. 1–6. IEEE (2018)
27. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019)
28. Tagaris, A., Kollias, D., Stafylopatis, A., Tagaris, G., Kollias, S.: Machine learning for neurodegenerative disorder diagnosis-survey of practices and launch of benchmark dataset. *Int. J. Artif. Intell. Tools* **27**(03), 1850011 (2018)
29. Zisad, S.N., Hossain, M.S., Andersson, K.: Speech emotion recognition in neurological disorders using convolutional neural network. In: Mahmud, M., Vassanelli, S., Kaiser, M.S., Zhong, N. (eds.) *BI 2020. LNCS (LNAI)*, vol. 12241, pp. 287–296. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59277-6_26