Received: 7 April 2024 | Revised: 13 June 2024 | Accepted: 17 July 2024 | Published online: 1 August 2024

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2024, Vol. 00(00) 1–10
DOI: 10.47852/bonviewJCCE42023040

BON VIEW PUBLISHING

Comparison Among Deep Learning Approaches and Biomarkers in Early Detection of Parkinson's Disease

Gunjan Goswami[1] and Bhanu Prasad[2,*]

[1]*Independent Researcher, Canada*

[2]*Department of Computer and Information Sciences, Florida A&M University, USA*

Abstract: Parkinson's disease (PD) is a chronic neurodegenerative disorder that is affecting millions of people worldwide. There is a lot of research reported on using different biomarkers and different deep learning (DL) approaches for the early detection of PD. However, there is no research reported on which biomarker is better suited when a specific DL model is used and vice versa. The current research is aimed at addressing this issue. We use two widely used DL-based methods, Autoencoder (AE)-based and Convolution Neural Network (CNN)-based, to find out which one is better suited for PD classification when imaging, voice, and cerebrospinal fluid (CSF) biomarkers are used for evaluation. Further, we explore which biomarker, among imaging, voice, and CSF, is better suited for PD classification when the AE-based approach and CNN-based approach are used for evaluation. Two variants of AE-based approaches, Sparse AE (SAE) and Stacked Sparse AE (SSAE), are implemented in this research. The biomarkers' values are used from publicly available databases. Various experiments are conducted by using several performance parameters, and it is observed that (i) SSAE is better than SAE as well as CNN for the considered magnetic resonance imaging (MRI), CSF, and voice features, based on the average of all the performance parameters' values for training as well as for testing datasets, and (ii) MRI is better than CSF as well as voice features for the considered SAE, SSAE, and CSF frameworks, based on the average of all the performance parameters' values, for training as well as for testing databases: for training databases, these values are 91.83% for MRI features, 84.55% for voice features, and 74.88% for CSF, and for testing databases, these values are 96.53 % for MRI features, 92.08 % for voice features, and 85.51% for CSF.

Keywords: Parkinson's disease (PD), artificial intelligence (AI), deep learning (DL), magnetic resonance imaging (MRI), cerebrospinal fluid (CSF) biomarkers, voice recordings

1. Introduction

Parkinson's disease (PD) is affecting millions of people all over the globe. According to Parkinson's Foundation [1], in the USA, approximately one million people are affected with PD. This count is likely to be 1.2 million by the year 2030. Another country in which PD is more prevalent is China. According to Li et al. [2], by the year 2030, China will be having approximately half of the PD population in the world. Because of the increasing population and aging, PD cases are anticipated to reach 4.94 million by 2030 in China. A similar situation is with India too. In 2016, there were approximately 0.58 million PD-affected people in India [3], with a major increase expected in the coming years. The number of PD cases is growing very rapidly. The risk of PD development is twice as high in males than females, but faster progression of PD is observed in females [4]. As the number of PD cases is increasing, it is burdening the countries' economies. Besides the economic interruption, this neurodegenerative disease is also causing social hindrance to the community. In 2017, in the USA, the collective economic burden was $51.9 billion including the direct medical cost of $25.4 billion and indirect and nonmedical cost of $26.5 billion [5]. Since PD is impacting a lot of people economically, socially, and personally, it is very important to optimize the ways to control this disease, and it can be done by diagnosing PD while it is in the early stages. If PD is not detected in the early stages, it could lead to the death of the affected person. It is important to find out the ways to detect PD in the early stages so that the treatment can be started on time, the symptoms will be less severe, and the quality or quantity of life of the patient will be better.

It is difficult to diagnose PD in its initial stages because other neurodegenerative diseases such as Alzheimer's disease (AD) have similar symptoms. Though, in advanced stages, it is easy to differentiate PD from AD, it is difficult to find an effective treatment for PD in those stages. Levodopa is a common medicine that can suppress the motor symptoms of PD for a short duration, but a permanent cure is still unavailable [6, 7].

Neuroimaging methods (NIMs), rating scale methods (RSMs), and other biological/biochemical biomarkers (BBs) can be utilized to discriminate between various neurological disorders and the stages of these disorders. Some NIMs used for diagnosing neurodegenerative diseases are CT, MRI, fMRI, PET, and SPECT [6]. Some RSMs used to evaluate the progression of the PD symptoms, functional changes, and treatment-related observations are the Unified

***Corresponding author:** Bhanu Prasad, Department of Computer and Information Sciences, Florida A&M University, USA. Email: bhanu.prasad@famu.edu

Parkinson's Disease Rating Scale [8], Hoehn and Yahr (H&Y) [7], and Schwab and England Activities of Daily Living Scale.

Biomarkers play a very vital role in diagnosing PD at different stages, hence in curing it to some extent. Biomarkers are categorized as clinical, genetic, biochemical/biological, and imaging biomarkers [9]. Though each type of biomarker is important to assess the progression of PD, in this study, we are evaluating voice, imaging (MRI), and biological biomarkers (cerebrospinal fluid (CSF)) individually to see the effectiveness of these biomarkers, in classifying PD subjects from healthy (HY) subjects, in a given collection of subjects. In this research, the term "PD classification" also includes determining or diagnosing if a given subject is a PD subject or not. In addition, the terms "feature" and "biomarker" are used interchangeably in this research. The same is true for "approach" and "framework".

The contribution of this study is provided next.

The present research is the earliest one that is aimed at finding out (a) the deep learning (DL)-based framework, among Sparse Autoencoder (SAE), Stacked Sparse Autoencoder (SSAE), and Convolution Neural Network (CNN), which is better suited for PD classification when imaging, voice, and CSF biomarkers are used, and (b) the biomarker, among imaging, voice, and CSF, which is better suited for PD classification when DL-based frameworks are used.

The rest of the research is composed as follows: Section 2 discusses the research done so far about PD classification using artificial intelligence (AI) methods, including machine learning (ML) and DL, on the datasets consisting of voice recordings, imaging, and biological biomarkers. Section 3 represents the data collection methods, the statistical summary of CSF, and the voice recording datasets used in this research. Further, AI methods used for PD classification based on these datasets are discussed in the same section. The experimental details are provided in Section 4, followed by a conclusion in Section 5.

2. Existing Research

The basal ganglia are the core brain region compromised in PD [10]. These researchers also mentioned that PD is caused by dysfunction of the whole basal ganglia-cortex-cerebellum system rather than just by basal ganglia. There are other region of interests (ROIs) such as the visual cortex, cortex, and temporal gyrus that are affected by PD [11].

Along with the changes in the brain, there are certain other BBs that are also affected as PD progresses. Parnetti et al. [12] discussed the importance of CSF and blood biomarkers in PD diagnosis. Pahuja and Prasad [13] discussed various CSF biomarkers, like Aβ42, α-Syn, Total-tau, and P-tau, which are affected by PD. Despite a lot of research and scientific advancements, there is no single biomarker that can be used for the early detection of PD. Thus, in this study, various biomarkers are being compared to find out the categories of biomarkers that are more prevalent during PD and can be used for PD classification.

Most of the work done so far for PD classification is based on voice recordings, MRI, and BBs, using AI methods. Some of the important findings based on these biomarkers are provided next. Das [14] found that "voice" plays a very important role in PD classification. Das [14] used four different classification methods, namely, regression, DMNeural, decision tree, and Artificial Neural Network (ANN) on the voice dataset. An overall classification score of 92.9% was achieved by that author by using ANN. Karaman et al. [15] developed a CNN model for PD identification based on biomarkers-derived voice characteristics. SqueezeNet1_1, DenseNet161, and ResNet101 were assessed to find out which architecture identifies the PD subjects more accurately. These authors identified DenseNet-161 architecture as the most suitable one with the highest testing accuracy at 89.75% among the other considered architectures.

Pahuja and Nagabhushan [16] observed an accuracy of 95.89% on the same voice dataset as used by Das [14], by using ANN with the Levenberg–Marquardt algorithm. Suppa et al. [17] demonstrated that voice disorders are obvious in the early stages of PD and as the disease progresses speech gets worse. These authors classified the PD patients from HY subjects by considering the disease's stage, severity of PD, and effect of levodopa treatment. Cavallieri et al. [18] noticed that the acoustic speech variables and bradykinesia are correlated with each other in advanced PD. Thus, it is evident that voice plays a vital role in PD classification.

Various NIMs such as MRI, SPECT, fMRI, and CT have been used for PD classification. MRI is noninvasive [19] and does not use radiation, thus making MRI safe as compared to other imaging methods. Hence, we are considering MRI in the present study as well. Some of the findings for PD classification based on the MRI features are discussed next. Chakraborty et al. [20] employed 3D CNN architecture on 3T T1-weighted MRI scans of 406 (203 PD and 203 HY) subjects collected from Parkinson's Progression Markers Initiative (PPMI). These authors observed the following results for both classes: an overall accuracy of 95.29% with an average precision of 0.927, average recall of 0.943, f1-score of 0.936, average specificity of 0.9430, and receiver operating characteristic-area under curve (ROC-AUC) of 0.98. Sivaranjini and Sujatha [21] used deep learning neural networks (CNN AlexNet) for PD classification based on MRI and observed an accuracy of 88.9%. Solana-Lavalle and Rosas-Romero [11] performed MRI-based experiments for PD classification separately in men and women. T1-weighted MRI images of 312 male subjects and 168 female subjects are collected from PPMI. These authors used voxel-based morphometry (VBM) for feature extraction and observed that different brain regions are affected in men and women; specifically, the basal ganglia, fourth ventricle, brainstem, cerebellum, and lateral ventricle are the affected brain regions in men, while the affected brain regions in women are the thalamus, basal ganglia, the frontal lobe, and small part of the cerebellum. The following performance values are obtained by these authors: in men, an accuracy of 99.01%, 99.35% of sensitivity, 100% of specificity, and 100% of precision, while in women, an accuracy of 96.97%, 100% of sensitivity, 96.15% of specificity, and 97.22% precision. Vyas et al. [22] employed 318 MRI scans in the axial plane for PD classification and observed an accuracy of 72.22% with an AUC of 0.50 and an accuracy of 88.9% with an AUC of 0.86 by using 2D and 3D CNN models, respectively. Pahuja and Prasad [13] used DL architectures on MRI scans for PD classification. These authors observed an accuracy of 61%, 76.5%, and 91.43% by using Softmax Classifier (SC) with one SAE, stacked SAE, and CNN, respectively. Thus, it is evident that MRI plays a vital role in PD classification. Not only MRI but other imaging methods such as fMRI, SPECT [6], and so on are also being used for PD diagnosis.

CSF proteins play a very important role in diagnosing early stages of PD clinically [23]. These authors collected 15 CSF protein markers for 80 PD and 80 controls. After performing univariate analysis, these authors observed that there are six potential biomarkers for PD diagnosis, and these are α-syn, S100β, DJ-1, p-Tau, Aβ42, and t-Tau. After that, based on the ML approach, these authors found that the important biomarkers for PD diagnosis are α-syn, S100β, and UCHL1. Katayama et al. [24] presented a review of CSF biomarkers for PD diagnosis based on the latest literature and meta-analysis data. Although recently available PD detection methods depend on imaging and clinical features, CSF is a very important

biomarker that reflects early variation occurring in the brains of PD subjects [25]. Kwon et al. [26] discussed that CSF is a promising candidate biomarker for PD diagnosis as it is near to the brain structures. Hence, it could indicate the changes in pathologic procedures. Thus, it is observed that imaging, voice, and CSF biomarkers play a vital role in PD classification.

3. Materials and Methods

3.1. Details of the biomarkers employed in this study

Voice recordings, MRI images/scans, and CSF are the three biomarkers used in this study. Details are provided in subsections 3.1.1, 3.1.2, and 3.1.3.

3.1.1. MRI scans

From the literature, it is evident that PD leads to a change in the brain volumes of gray matter (GM)/white matter (WM)/CSF of the patients as this disease progresses [27]. ROI and VBM approaches are the two well-known approaches that are employed to relate brain volume changes with neurodegenerative disease's progression [28]. VBM is being used here in this to evaluate the GM variations between PD and HY groups. Before using the VBM approach, it is required to preprocess the MRI scans, and this is done by Statistical Parametric Mapping Version 8 (SPM8). Preprocessing is required to ensure the proper alignment of MRI images. The following steps are involved in the preprocessing of MRI images: conversion of DICOM (Digital Imaging and Communications in Medicine) to 3D NIFTI (Neuroimaging Informatics Technology Initiative) format conversion, spatial normalization, unified segmentation, and smoothing [29]. After this preprocessing, either of the smoothed GM/WM/CSF TPMs (tissue probability maps) can be used for extracting the MRI features. In this study, we used GM volumes for extracting the features.

In this research, to perform PD classification, T1-weighted MRI scans are collected, for 200 subjects, from PPMI. Because of the failure of the segmentation process on approximately 40 MRI scans, MRI features from the remaining 160 (200 minus 40) subjects are used for PD classification in this study. Details of these MRI scans are provided in Table 1. By using SPM8/VBM, a total of 2038 features are extracted from MRI scans of these 160 subjects. There could be a possibility that, by using the same set of MRI scans, different researchers got different numbers of features from MRI scans [27] as there are many settings that can be varied in SPM8/VBM, thus affecting the accuracy of the classification results.

3.1.2. CSF biomarkers' values

In this research, to perform PD classification, four CSF biomarkers' (i.e., α-synuclein, Aβ42, total-tau, and P-tau181P) values for 474 subjects are collected from PPMI. Out of these 474 subjects, 350 are early PD because they are in Stages 1 and 2 of the H&Y rating scale, and the remaining 124 are HY. The statistical significance of each of these CSF biomarkers is done at a significance level (p-value) of 0.10. It is observed that, for each of these four CSF biomarkers, the p-value is below 0.10, and this shows that the CSF biomarkers' values that are considered are significant. Hence, these biomarkers' values should be considered for PD classification. Table 2 and Figure 1 show the details and the boxplots, respectively, of the CSF biomarkers' values used in this study.

From Figure 1, it is evident that the notches for CSF biomarkers' values for early PD and HY subjects are separated from each other. Therefore, the above-mentioned CSF markers have an important part in distinguishing the HY and early PD subjects.

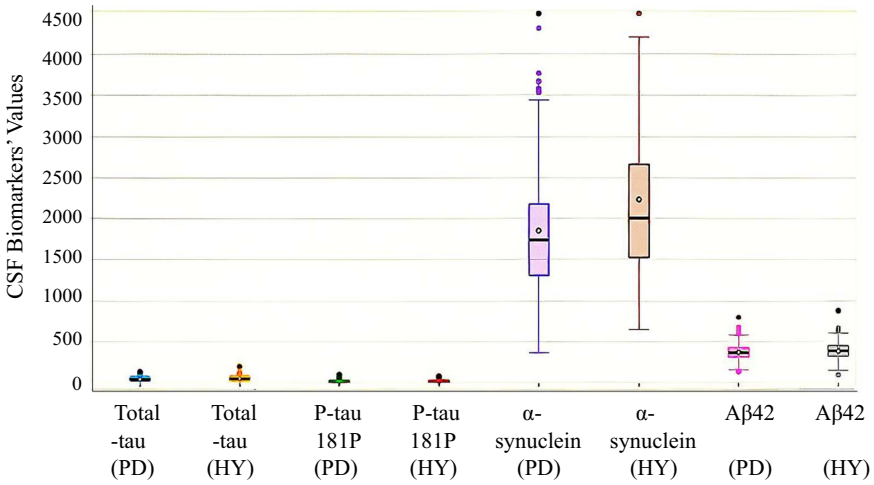
Table 1
Details of MRI scans of 80 early PD and 80 HY subjects collected from PPMI

Variable	Early PD subjects (80)	HY subjects (80)
Gender (F/M)	24/58	22/60
Age (range in years)	31–82	34–77
Slice thickness	1–1.5 mm	1–1.5 mm
Voxel dimensions	1_1_1.20 mm	1_1_1.20 mm

Table 2
Details of the CSF biomarkers' values for the PD (350) and HY (124) subjects

		Total-tau	P-tau181P	α-synuclein	Aβ42
Min (minimum value)	PD	14.4	5.00	363.12	129.2
	HY	18.23	5.10	641.90	94.00
Q1 (first quartile)	PD	32.03	10.01	1309.49	311.43
	HY	37.24	12.40	1526.32	324.25
Median (median)	PD	40.12	13.54	1736.09	362.40
	HY	46.15	14.72	2002.73	385.15
Q3 (third quartile)	PD	50.39	19.06	2172.53	419.78
	HY	57.14	21.41	2661.74	445.08
Max (maximum value)	PD	128.2	94.10	4822.22	796.50
	HY	194.97	73.30	5621.51	879.50
Mean (average value)	PD	43.94	15.77	1850.77	368.43
	HY	51.92	18.11	2228.94	382.91
SD (standard deviation)	PD	17.96	8.91	741.47	96.31
	HY	24.73	10.70	989.77	109.87

Figure 1
Boxplots of CSF biomarkers' values



3.1.3. Voice dataset

In this research, to perform PD classification, a voice recording dataset consisting of 195 voice measurements from 31 (23 PD and 8 HY) subjects is obtained from the UCI machine learning repository [30]. Table 3 contains the details of the voice dataset used in this study.

3.2. Details of the DL-based approaches used in this study

Autoencoder (AE)-based (i.e., SAE and SSAE) and CNN-based are the two DL-based approaches used in this study, and they are explained in subsections 3.2.1 and 3.2.2.

3.2.1. AE-based approach

Analogous to "multilayer perceptron", AE also consists of an input layer (InL), hidden layer (HiL), and output layer (OuL). The count of nodes in OuL and InL is the same. The AE-based approach compresses the data; thus, it is a dimensionality reduction approach [31]. Also, AEs are very data specific; that is, they can only compress the input data that is similar to what these AEs have been trained for. The number of HiL nodes is a hyper-parameter that is set before the training of AE starts, and in our case, we assumed that number to be 10. The other hyper-parameters, which are important to initialize before the training of AE starts, are explained later in this section. SAEs are variations of the AEs, and the only difference is that in SAEs, a sparsity penalty is involved in the training phase (Figure 2).

Table 3
Details of the voice dataset

	Min		Q1		Median		Q3		Max		Mean		SD	
Class	PD(23)	HY(8)	PD(23)	HY(8)	PD(23)	HY(8)	PD(23)	HY(8)	PD(23)	HY(8)	PD(23)	HY(8)	PD(23)	HY(8)
MDVP:Fo(Hz)	88.33	110.74	117.96	117.06	148.14	198.17	173.92	234.50	223.36	260.11	147.34	179.32	33.47	55.95
MDVP:Fhi(Hz)	102.15	113.60	134.66	135.81	163.74	238.24	209.51	252.82	588.52	592.03	189.32	225.46	86.69	103.43
MDVP:Flo(Hz)	65.48	74.29	81.74	96.93	100.76	109.17	133.75	216.21	199.02	239.17	110.32	138.18	35.93	59.61
MDVP:Jitter(%)	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.03	0.01	0.01	0.00	0.01	0.00
MDVP:Jitter(Abs)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MDVP:RAP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00
MDVP:PPQ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00
Jitter:DDP	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.06	0.02	0.01	0.01	0.01	0.00
MDVP:Shimmer	0.01	0.01	0.02	0.02	0.03	0.02	0.04	0.02	0.12	0.04	0.03	0.02	0.02	0.01
MDVP:Shimmer(dB)	0.09	0.10	0.16	0.14	0.26	0.15	0.38	0.20	1.30	0.41	0.31	0.17	0.21	0.06
Shimmer:APQ3	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.06	0.02	0.02	0.01	0.01	0.00
Shimmer:APQ5	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.08	0.02	0.02	0.01	0.01	0.00
MDVP:APQ	0.01	0.01	0.01	0.01	0.02	0.01	0.03	0.02	0.14	0.03	0.03	0.01	0.02	0.00
Shimmer:DDA	0.01	0.02	0.03	0.02	0.04	0.03	0.07	0.04	0.17	0.07	0.05	0.03	0.03	0.01
NHR	0.00	0.00	0.01	0.00	0.01	0.01	0.03	0.01	0.31	0.11	0.03	0.01	0.04	0.02
HNR	8.44	17.88	18.80	22.25	21.53	24.69	24.89	25.53	33.05	26.81	21.37	23.78	4.70	2.50
RPDE	0.26	0.26	0.43	0.38	0.51	0.44	0.60	0.52	0.69	0.66	0.51	0.45	0.10	0.10
D2	0.57	0.63	0.69	0.65	0.73	0.67	0.76	0.75	0.83	0.79	0.73	0.69	0.05	0.05
DFA	–7.96	–7.52	–6.14	–7.07	–5.48	–6.77	–4.71	–6.23	–2.43	–5.20	–5.42	–6.77	1.05	0.59
spread1	0.06	0.01	0.19	0.11	0.24	0.16	0.30	0.20	0.45	0.29	0.25	0.16	0.08	0.07
spread2	1.42	1.84	2.14	2.01	2.43	2.16	2.66	2.36	3.67	2.88	2.43	2.16	0.40	0.26
PPE	0.04	0.07	0.16	0.10	0.22	0.12	0.27	0.16	0.53	0.25	0.23	0.12	0.09	0.04

Figure 2
SAE architecture

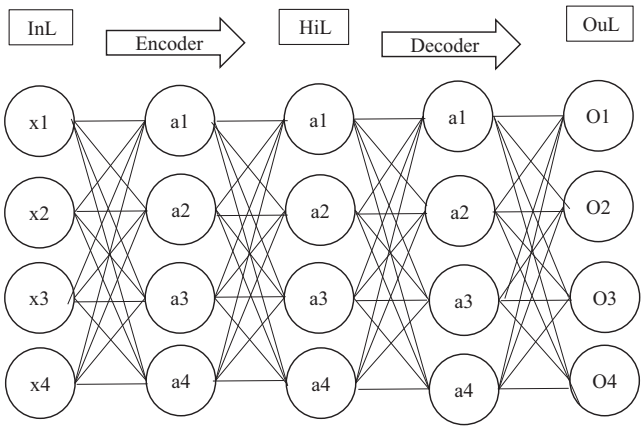
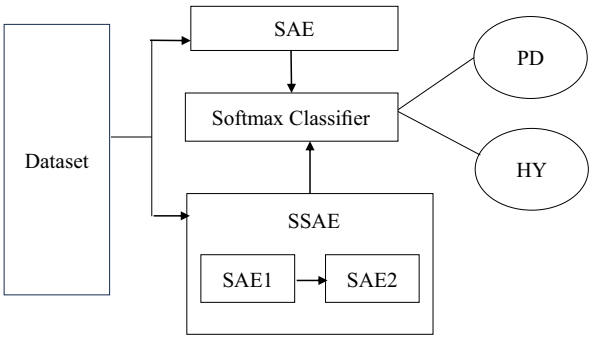


Figure 3
AE (i.e., SAE and SSAE) architecture with Softmax Classifier

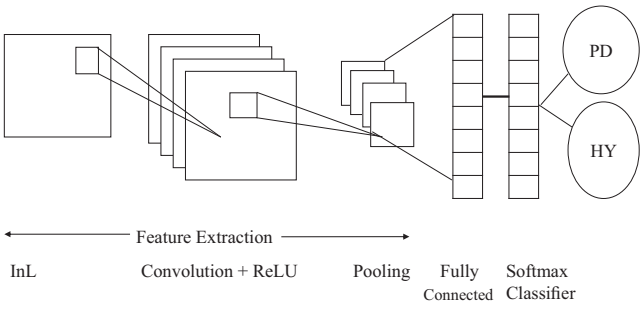


In SAEs, not all, but only a fraction of HiL nodes is activated, and the other HiL nodes are penalized. This strategy is implemented by "regularization," and by using this strategy, it is certain that SAEs are learning only the latent representations, not the redundant information, from the input data. This approach works fine if the HiL size is even larger as only a few nodes are active at any time and AEs will be learning the useful features only. L2 regularization is used in this study as it adds the squared magnitude of the coefficients as the penalty. Also, L2 regularization moves the weights "w" toward 0, but because of small step sizes, "w" never reaches 0. The mean squared error and the cross-entropy are the types of loss functions that are used with AEs.

Since we are dealing with the PD classification problem here, cross-entropy is being used. When the output of one SAE (i.e., SAE1) relates to the input of the next SAE (i.e., SAE2), it is called SSAE [31]. Since the dataset (voice, imaging, and CSF) we are using here for PD classification is a labeled dataset, SC is used to classify the PD and HY subjects as shown in Figure 3.

More details about PD classification by using SC with SAE and SSAE can be found in Pahuja and Prasad [13]. All the experiments are performed by using the online Matlab free version (https://matlab.mathworks.com/). For AE-based approach, the hyper-parameters' values used for sparsity regularization, L2 weight regularization, and sparsity proportion are 4, 0.01, and 0.05, respectively. Purelin and cross-entropy are used as the decoder transfer function and loss function, respectively.

Figure 4
CNN-based approach



3.2.2. CNN-based approach

CNN is a kind of deep neural network that learns from the data directly, and the need for manual feature extraction is eliminated. Like ANN, CNN consists of an InL, an OuL, and many HiLs in between the InL and OuL (Figure 4).

Convolution layer, pooling layer, and rectified linear unit (ReLU) are the three common layers that can be heaped in accordance with the functionality [32]. After feature extraction is performed, the next phase is classification. As observed from Figure 4, in this research, SC is a layer that is used as OuL. SC gives the classification output to the user. Additional details about CNN architecture, hyper-parameters, and layers can be found in O'Shea and Nash [32]. For the CNN-based approach, adaptive moment estimation optimizer (Adam), ReLU, and cross-entropy are used as an optimizer, activation function, and loss function, respectively.

3.3. Biomarker datasets and DL-based approaches in action

In this study, DL-based (i.e., AE-based and CNN-based) frameworks are implemented to perform PD classification based on collected voice, imaging, and CSF biomarkers' values.

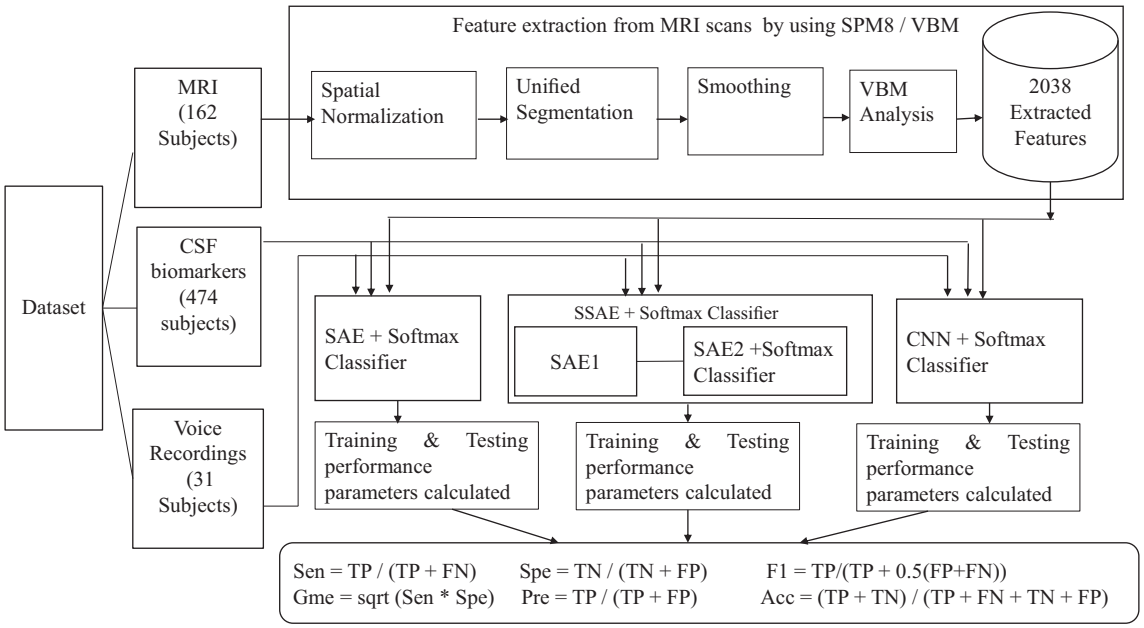
Figure 5 represents the schematic diagram of the approach used in this research. The datasets used in this work are not balanced; that is, the number of HY subjects is unequal to the number of PD subjects. Hence, the performance parameters sensitivity (Sen), specificity (Spe), precision (Pre), F1-score (F1), and geometric mean (Gme) are used along with accuracy to determine the performance of DL-based frameworks [33].

In both the DL-based approaches, the three features (voice, features extracted from MRI scans, and CSF) are used to find out the most promising feature for PD classification, as shown in Figure 5. In addition, for each of these three features, the two DL-based frameworks are used to find out the most promising DL-based framework (among these two DL-based frameworks) for PD classification, as shown in Figure 5. Since the subjects are different for each of the three features considered, and due to the nonavailability of the same set of subjects with these three features, experiments are conducted as follows for different combinations of these features and these two DL-based frameworks:

1) Voice + SAE, Voice + SSAE, and Voice + CNN
2) MRI + SAE, MRI + SSAE, and MRI + CNN
3) CSF + SAE, CSF + SSAE, and CSF + CNN.

The results of these experiments are provided in the following section.

Figure 5
A schematic diagram of the approach used in this study



Note: TP is true positive, FN is false negative, TN is true negative, and FP is false positive

4. Experimental Results and Discussion

The six performance parameters used in all the experiments in this research are Sen, Spe, Acc, F1, Pre, and Gme. As explained earlier, SC is used with SAE, SSAE, and CNN in all these experiments. Therefore, in the rest of this work, SAE, SSAE, and CNN mean SAE + SC, SSAE + SC, and CNN + SC, respectively.

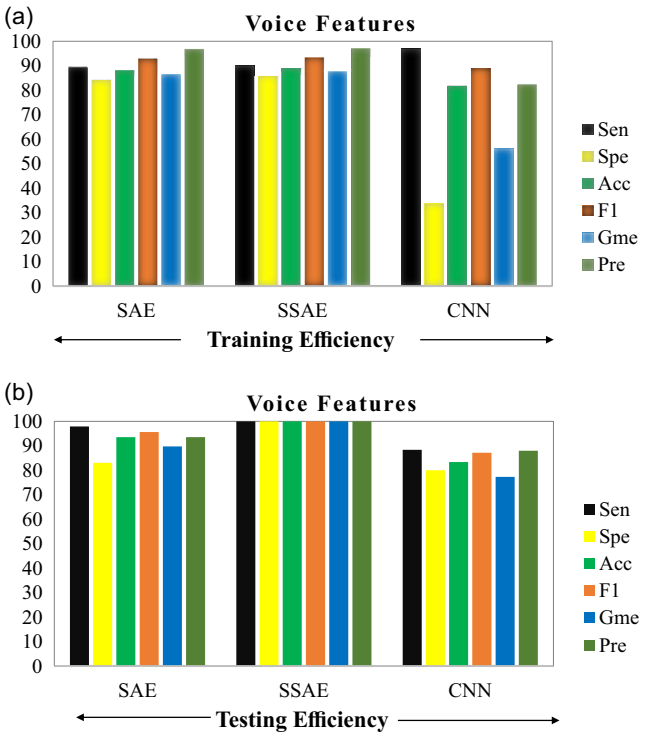
The mean values of each of the six performance parameters, received from five experiments conducted on the voice features (five experiments on the training voice dataset to determine the training efficiency in terms of voice and another separate five experiments on the testing dataset to determine the testing efficiency in terms of voice), are presented graphically in Figure 6. Similar calculations are done for MRI and CSF features and are represented in Figures 7 and 8, respectively.

4.1. Comparison between the performance of DL-based frameworks, by considering voice features

From Figure 6(a), it is evident that SSAE (90.01% Sen, 85.75% Spe, 89.03% accuracy, 93.33% F1, 87.73% Gme, and 97.05% Pre) is giving better results, on the training voice dataset, as compared to SAE (89.29% Sen, 84.14% Spe, 88.13% accuracy, 92.78% F1, 86.49% Gme, 96.72% Pre) in terms of the value of each of these performance parameters. In addition, from Figure 6(a), it is evident that, on the training voice dataset, SSAE is better than CNN (96.89% Sen, 35.57% Spe, 81.67% accuracy, 88.91% F1, 56.11% Gme, and 82.32% Pre) in terms of the value of each of these performance parameters except for Sen.

Furthermore, for the testing voice features, as shown in Figure 6(b), SSAE (100% Sen, 100% Spe, 100% accuracy, 100% F1, 100% Gme, and 100% Pre) is outperforming SAE (97.88% Sen, 83.03% Spe, 93.50% accuracy, 95.60% F1, 89.76% Gme, and 93.55% Pre) as well as CNN (88.33% Sen, 80.00% Spe, 83.33% accuracy, 87.14% F1, 77.32% Gme, and 88.00% Pre) in terms of the value of each of these performance parameters.

Figure 6
Performance parameter values obtained on training and testing the voice features by using SAE, SSAE, and CNN



4.2. Comparison between the performance of DL-based frameworks, by considering CSF features

For the training CSF dataset (Figure 7(a)), the performance parameters' values obtained by SSAE are 80.65% Sen, 79.02% Spe, 80.58% accuracy, 88.07% F1, 79.81% Gme, and 97.07%

Figure 7
Performance parameter values obtained on training and testing CSF dataset by using SAE, SSAE, and CNN

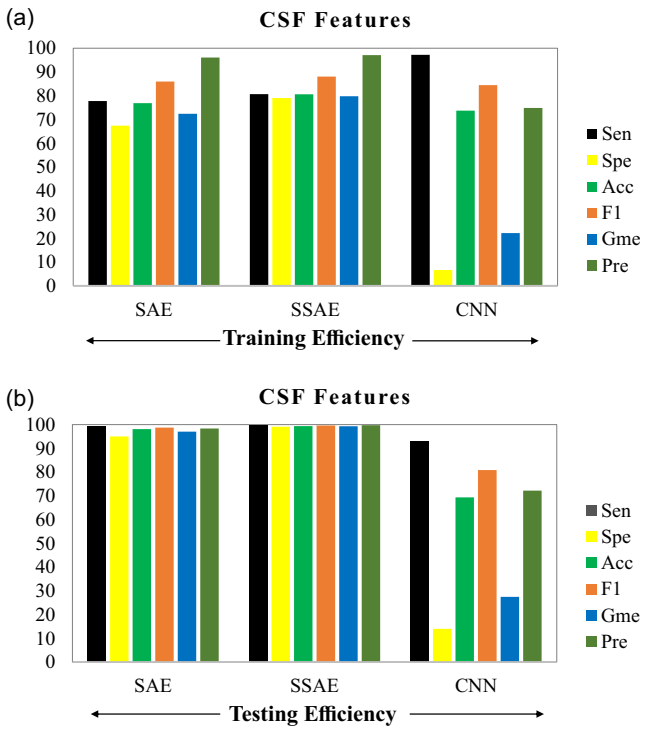
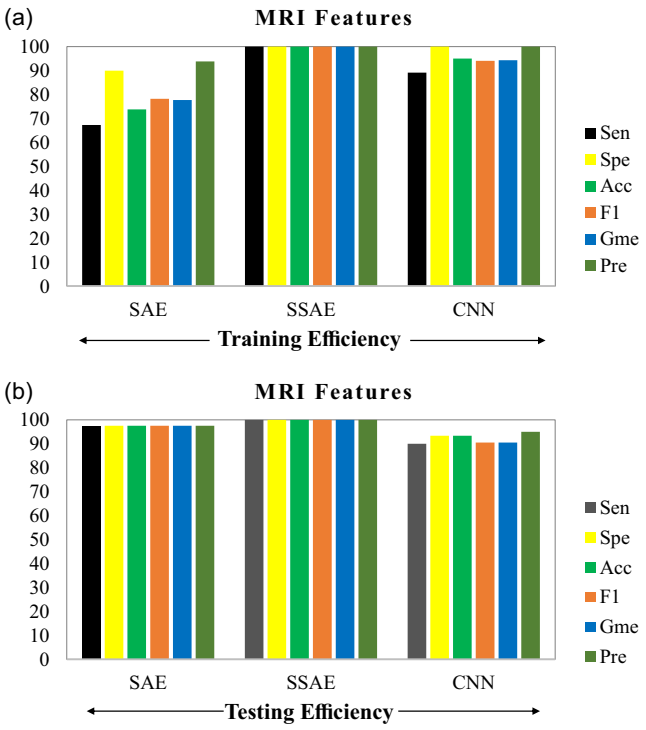


Figure 8
Performance parameter values obtained on training and testing MRI dataset by using SAE, SSAE, and CNN



Pre. These values are greater than the values achieved by SAE (77.76% Sen, 67.40% Spe, 76.84%, accuracy, 85.95% F1, 72.37% Gme, and 96.07% Pre). Furthermore, for the training CSF dataset, SSAE is also outperforming CNN (97.19% Sen, 6.65% Spe, 73.68% accuracy, 84.43% F1, 22.21% Gme, and 74.83% Pre) in terms of the value of each of these performance parameters except for Sen.

Furthermore, for the testing CSF dataset (Figure 7(b)), SSAE (99.44% Sen, 99.09% Spe, 99.36% accuracy, 99.58% F1, 99.27% Gme, and 99.72% Pre) is outperforming SAE (99.16% Sen, 94.95% Spe, 98.09% accuracy, 98.73% F1, 97.03% Gme, and 98.31% Pre) as well as CNN (92.73% Sen, 14.00% Spe, 69.33% accuracy, 80.81% F1, 27.47% Gme, and 72.16% Pre) in terms of the value of each of these performance parameters.

4.3. Comparison between the performance of DL-based frameworks, by considering MRI features

For the training MRI features, as shown in Figure 8(a), all the performance parameters' values obtained by using SSAE (100% Sen, 100% Spe, 100% accuracy, 100% F1, 100% Gme, and 100% pre) are better than or equal to the corresponding performance parameters' values obtained by using SAE (67.23% Sen, 89.90% Spe, 73.75% accuracy, 78.20% F1, 77.69% Gme, and 93.75% Pre) as well as CNN (89.11% Sen, 100.00% Spe, 95.00% accuracy, 94.04% F1, 94.30% Gme, and 100% Pre).

Furthermore, for the testing MRI features, as shown in Figure 8(b), SSAE (100% Sen, 100% Spe, 100% accuracy, 100% F1, 100% Gme, and 100% Pre) is outperforming SAE (97.50% Sen, 97.50% Spe, 97.50% accuracy, 97.50% F1, 97.50% Gme, and 97.50% Pre) as well as CNN (90.00% Sen, 93.33% Spe, 93.33% accuracy, 90.48% F1, 90.47% Gme, and 95.00% Pre) in terms of the value of each of the performance parameters.

4.4. Comparison between the performance of DL-based frameworks, by considering all the performance parameters

To determine which DL-based approach provides better training performance for PD classification, when voice, imaging, and CSF are used, we have performed the following experiment on the considered MRI, CSF, and voice features:

1) Calculated the average value of all the six performance parameters obtained, for training, for the voice dataset (i.e., we added the values of all these six performance parameters, shown in Figure 6(a), then divided that sum by 6).
2) Added the values of all those six performance parameters shown in Figure 7(a) then divided by 6.
3) Added the values of all those six performance parameters shown in Figure 8(a) then divided by 6.

Similarly, to determine which DL-based approach provides better testing performance for PD classification, when voice, imaging, and CSF are used, we have repeated the above experiment by using Figures 6(b), 7(b), and 8(b), in place of Figures 6(a), 7(a), and 8(a), respectively.

The results obtained from the above two experiments are depicted in Figure 9(a) and (b), respectively.

From Figure 9(a), it can be observed that, for the training dataset, SSAE (100% for MRI, 90.48% for voice, and 84.20% for CSF) is better than SAE (80.09% for MRI, 89.59% for voice, and 79.40% for CSF) as well as CNN (95.41% for MRI, 73.58% for voice, and 59.83% for CSF) for the MRI, CSF, and voice datasets that are used.

From Figure 9(b), it can be noticed that, for the testing dataset, SSAE (100% for MRI, 100% for voice, and 99.41% for CSF) is better than SAE (97.50% for MRI, 92.22% for voice, and 97.71%

Figure 9
Comparison between the performance of DL-based frameworks when all the six performance parameters are considered together

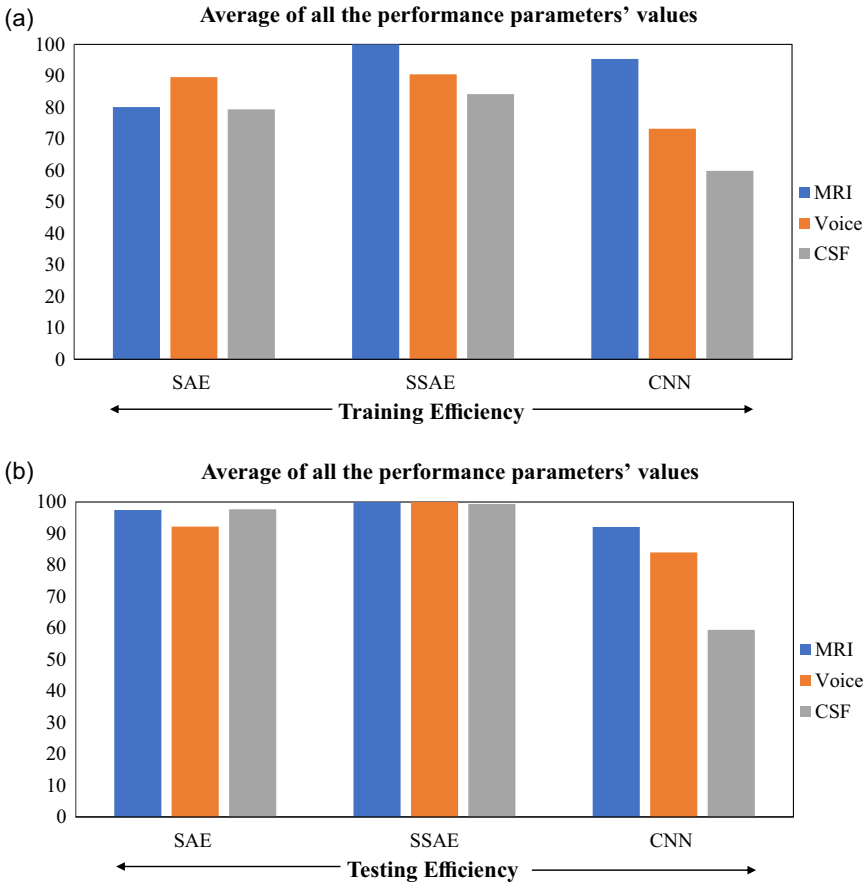
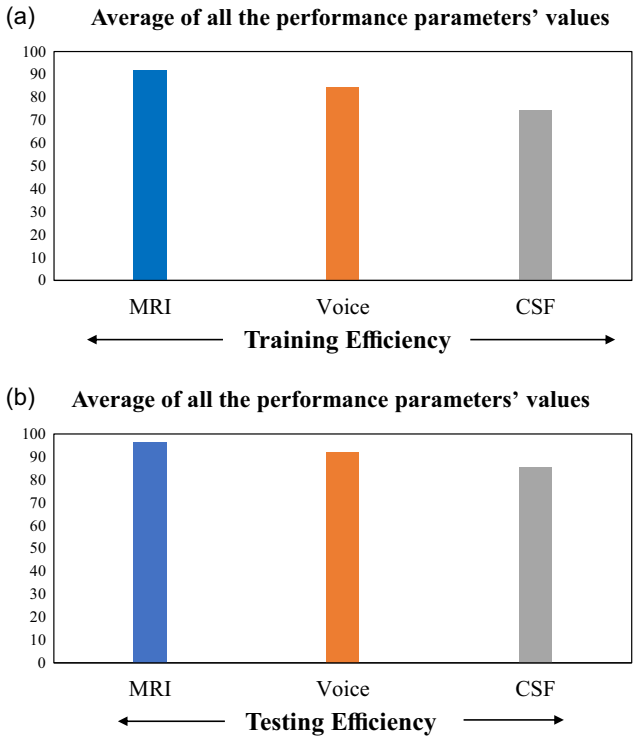


Figure 10
Comparison between the performance of biomarkers when all the six performance parameters are considered together



for CSF) and CNN (92.10% for MRI, 84.02% for voice, and 59.42% for CSF) for the MRI, CSF, and voice datasets that are used.

4.5. Comparison between the performance of biomarkers, by considering all the performance parameters

To determine which biomarker provides better training performance for PD classification when DL-based frameworks are used, we calculated the average value of all the six performance parameters obtained by SAE, SSAE, and CNN, for training the voice dataset (i.e., we added the values of all the six performance parameters received by DL-based frameworks, i.e., Figures 6(a), 7(a), and 8(a), then divided that sum by 18).

To determine which biomarker (among voice, imaging, and CSF) provides better testing performance for PD classification, when DL-based frameworks are used, we have repeated the experiment explained in the above paragraph by using Figures 6(b), 7(b), and 8(b), in place of Figures 6(a), 7(a), and 8(a), respectively.

The results obtained from the above two experiments are depicted in Figure 10(a) and (b), respectively.

From Figure 10(a), it can be noticed that, for the training dataset, the performance parameters' value (91.83%) for MRI features is better than that for the voice features (84.55%) as well as CSF (74.48%).

From Figure 10(b), it can be noticed that, for the testing dataset, the performance parameters' value (96.53%) obtained for MRI features is better than that for the voice features (92.08%) as well as CSF (85.51%).

5. Conclusion

This research is conducted by using AE-based and CNN-based frameworks and by using MRI, CSF, and voice biomarkers. Based on the experiments, the following observations are made:

For the testing voice dataset, SSAE is better than SAE as well as CNN, in terms of the value of each of the performance parameters. For the training voice dataset, SSAE is better than SAE in terms of the value of each of the performance parameters. In addition, for the training voice dataset, SSAE is better than CNN in terms of the value of each of the performance parameters except for Sen.

For the testing CSF dataset, SSAE is better than SAE as well as CNN, in terms of the value of each of the performance parameters. For the training CSF dataset, SSAE is better than SAE in terms of the value of each of the performance parameters. However, for the training CSF dataset, SSAE is better than CNN in terms of the value of each of the performance parameters except for Sen.

For the testing MRI dataset, SSAE is better than SAE Classifier as well as CNN, in terms of the value of each of the performance parameters. The same is true for the training MRI dataset.

Based on the average of all the performance parameters' values, for training as well as for testing, SSAE is better than SAE as well as CNN for the considered MRI, CSF, and voice features.

Based on the average of all the performance parameters' values, for training as well as for testing, efficiencies of each of the SAE, SSAE, and CSF are calculated, it is observed that MRI is better than CSF as well as voice features for the considered SAE, SSAE, and CSF frameworks.

Recommendations

We believe that improved results may be achieved with these DL-based frameworks if more features are considered (i.e., larger dataset is used). Though the results achieved in this study can be noticed as a significant step for PD classification, these DL-based frameworks may be evaluated by changing the SAE, SSAE, and CNN hyper-parameter values and by considering other DL-based frameworks. Further, we have compared only three types of features here. Therefore, future studies must also consider other features from motor, non-motor, and imaging categories.

Acknowledgments

The authors are thankful to PPMI. The investigators within PPMI contributed to the design and implementation of PPMI and/or provided data and collected biospecimen but did not participate in the analysis or writing of this report. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Amathus Therapeutics, Avid Radiopharmaceuticals, Bial Biotech, Biogen, BioLegend, Bristol Myers Squibb, Calico, Celgene, Denali Therapeutics, 4D Pharma PLC, GE Healthcare, Genentech, GlaxoSmithKline, Golub Capital, Handl Therapeutics, Insitro, Janssen, Lilly, Lundbeck, Merck, Meso Scale Discovery, Neurocrine, Pfizer, Piramal Imaging, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB, Verily, and Voyager Therapeutics.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The voice datasets that support the findings of this study are openly available at https://archive.ics.uci.edu/ml/datasets/parkinsons. The MRI and CSF data that support the findings of this study are openly available in PPMI at https://www.ppmi-info.org.

Author Contribution Statement

Gunjan Goswami: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Project administration. **Bhanu Prasad:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Parkinson's Foundation. (2024). *Who has Parkinson's*. Retrieved from: https://www.parkinson.org/understanding-parkinsons/statistics

[2] Li, G., Ma, J., Cui, S., He, Y., Xiao, Q., Liu, J., & Chen, S. (2019). Parkinson's disease in China: A forty-year growing track of bedside work. *Translational Neurodegeneration, 8,* 22. https://doi.org/10.1186/s40035-019-0162-z

[3] Rajan, R., Divya, K. P., Kandadai, R. M., Yadav, R., Satagopam, V. P., Madhusoodanan, U. K., ..., & Sharma, M. (2020). Genetic architecture of Parkinson's disease in the Indian population: Harnessing genetic diversity to address critical gaps in Parkinson's disease research. *Frontiers in Neurology, 11,* 524. https://doi.org/10.3389/fneur.2020.00524

[4] Cerri, S., Mus, L., & Blandini, F. (2019). Parkinson's disease in women and men: What's the difference? *Journal of Parkinson's Disease, 9*(3), 501–515. https://doi.org/10.3233%2FJPD-191683

[5] Yang, W., Hamilton, J. L., Kopil, C., Beck, J. C., Tanner, C. M., Albin, R. L., ..., & Thompson, T. (2020). Current and projected future economic burden of Parkinson's disease in the US. *npj Parkinson's Disease, 6*(1), 15. https://doi.org/10.1038/s41531-020-0117-1

[6] Bidesi, N. S. R., Andersen, I. V., Windhorst, A. D., Shalgunov, V., & Herth, M. M. (2021). The role of neuroimaging in Parkinson's disease. *Journal of Neurochemistry, 159*(4), 660–689. https://doi.org/10.1111/jnc.15516

[7] Jackson, H., Anzures-Cabrera, J., Taylor, K. I., Pagano, G., PASADENA Investigators, & Prasinezumab Study Group. (2021). Hoehn and Yahr stage and striatal Dat-SPECT uptake are predictors of Parkinson's disease motor progression. *Frontiers in Neuroscience, 15,* 765765. https://doi.org/10.3389/fnins.2021.765765

[8] Ramsay, N., Macleod, A. D., Alves, G., Camacho, M., Forsgren, L., Lawson, R. A., . . . , & Parkinson's Incidence Cohorts Collaboration. (2020). Validation of a UPDRS-/MDS-UPDRS-based definition of functional dependency for Parkinson's disease. *Parkinsonism & Related Disorders, 76,* 49–53. https://doi.org/10.1016/j.parkreldis.2020.05.034

[9] Surguchov, A. (2022). Biomarkers in Parkinson's disease. In P. V. Peplow, B. Martinez & T. A. Gennarelli (Eds.),

Neurodegenerative diseases biomarkers: Towards translating research to clinical practice (pp. 155–180). Springer. https://doi.org/10.1007/978-1-0716-1712-0_7

[10] Caligiore, D., Helmich, R. C., Hallett, M., Moustafa, A. A., Timmermann, L., Toni, I., & Baldassarre, G. (2016). Parkinson's disease as a system-level disorder. *npj Parkinson's Disease, 2,* 16025. https://doi.org/10.1038/npjparkd.2016.25

[11] Solana-Lavalle, G., & Rosas-Romero, R. (2021). Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine, 198,* 105793. https://doi.org/10.1016/j.cmpb.2020.105793

[12] Parnetti, L., Gaetani, L., Eusebi, P., Paciotti, S., Hansson, O., El-Agnaf, O., ..., & Calabresi, P. (2019). CSF and blood biomarkers for Parkinson's disease. *The Lancet Neurology, 18*(6), 573–586. https://doi.org/10.1016/s1474-4422(19)30024-9

[13] Pahuja, G., & Prasad, B. (2022). Deep learning architectures for Parkinson's disease detection by using multi-modal features. *Computers in Biology and Medicine, 146,* 105610. https://doi.org/10.1016/j.compbiomed.2022.105610

[14] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications, 37*(2), 1568–1572. https://doi.org/10.1016/j.eswa.2009.06.040

[15] Karaman, O., Çakın, H., Alhudhaif, A., & Polat, K. (2021). Robust automated Parkinson disease detection based on voice signals with transfer learning. *Expert Systems with Applications, 178,* 115013. https://doi.org/10.1016/j.eswa.2021.115013

[16] Pahuja, G., & Nagabhushan, T. N. (2021). A comparative study of existing machine learning approaches for Parkinson's disease detection. *IETE Journal of Research, 67*(1), 4–14. https://doi.org/10.1080/03772063.2018.1531730

[17] Suppa, A., Costantini, G., Asci, F., Di Leo, P., Al-Wardat, M. S., Di Lazzaro, G., ..., & Saggio, G. (2022). Voice in Parkinson's disease: A machine learning study. *Frontiers in Neurology, 13,* 831428. https://doi.org/10.3389/fneur.2022.831428

[18] Cavallieri, F., Di Rauso, G., Gessani, A., Budriesi, C., Fioravanti, V., Contardi, S., ..., & Valzania, F. (2023). A study on the correlations between acoustic speech variables and bradykinesia in advanced Parkinson's disease. *Frontiers in Neurology, 14,* 1213772. https://doi.org/10.3389%2Ffneur.2023.1213772

[19] Ryman, S. G., & Poston, K. L. (2020). MRI biomarkers of motor and non-motor symptoms in Parkinson's disease. *Parkinsonism & Related Disorders, 73,* 85–93. https://doi.org/10.1016%2Fj.parkreldis.2019.10.002

[20] Chakraborty, S., Aich, S., & Kim, H. C. (2020). Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network. *Diagnostics, 10*(6), 402. https://doi.org/10.3390%2Fdiagnostics10060402

[21] Sivaranjini, S., & Sujatha, C. M. (2020). Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications, 79,* 15467–15479. https://doi.org/10.1007/s11042-019-7469-8

[22] Vyas, T., Yadav, R., Solanki, C., Darji, R., Desai, S., & Tanwar, S. (2022). Deep learning-based scheme to diagnose Parkinson's disease. *Expert Systems, 39*(3), e12739. https://doi.org/10.1111/exsy.12739

[23] dos Santos, M. C. T., Scheller, D., Schulte, C., Mesa, I. R., Colman, P., Bujac, S. R., ..., & da Costa, A. N. (2018). Evaluation of cerebrospinal fluid proteins as potential biomarkers for early stage Parkinson's disease diagnosis. *PLOS ONE, 13*(11), e0206536. https://doi.org/10.1371/journal.pone.0206536

[24] Katayama, T., Sawada, J., Takahashi, K., & Yahara, O. (2020). Cerebrospinal fluid biomarkers in Parkinson's disease: A critical overview of the literature and meta-analyses. *Brain Sciences, 10*(7), 466. https://doi.org/10.3390/brainsci10070466

[25] Paolini Paoletti, F., Gaetani, L., & Parnetti, L. (2020). The challenge of disease-modifying therapies in Parkinson's disease: Role of CSF biomarkers. *Biomolecules, 10*(2), 335. https://doi.org/10.3390%2Fbiom10020335

[26] Kwon, E. H., Tennagels, S., Gold, R., Gerwert, K., Beyer, L., & Tönges, L. (2022). Update on CSF biomarkers in Parkinson's disease. *Biomolecules, 12*(2), 329. https://doi.org/10.3390%2Fbiom12020329

[27] Whitwell, J. L. (2009). Voxel-based morphometry: An automated technique for assessing structural changes in the brain. *Journal of Neuroscience, 29*(31), 9661–9664. https://doi.org/10.1523%2FJNEUROSCI.2160-09.2009

[28] Seyedi, S., Jafari, R., Talaei, A., Naseri, S., Momennezhad, M., Moghaddam, M. D., & Akbari-Lalimi, H. (2020). Comparing VBM and ROI analyses for detection of gray matter abnormalities in patients with bipolar disorder using MRI. *Middle East Current Psychiatry, 27,* 69. https://doi.org/10.1186/s43045-020-00076-3

[29] Kurth, F., Gaser, C., & Luders, E. (2015). A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (SPM). *Nature Protocols, 10*(2), 293–304. https://doi.org/10.1038/nprot.2015.014

[30] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 56*(4), 1015–1022. https://doi.org/10.1109/TBME.2008.2005954

[31] Kadam, V. J., & Jadhav, S. M. (2019). Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease. In *Computing, Communication and Signal Processing: Proceedings of ICCASP 2018,* 567–581. https://doi.org/10.1007/978-981-13-1513-8_58

[32] O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv Preprint:1511.08458.* https://doi.org/10.48550/arXiv.1511.08458

[33] Nguyen, G. H., Bouzerdoum, A., & Phung, S. L. (2009). Learning pattern classification tasks with imbalanced data sets. In P. Y. Yin (Ed.), *Pattern recognition* (pp. 193–208). Intech Open.

How to Cite: Goswami, G., & Prasad, B. (2024). Comparison Among Deep Learning Approaches and Biomarkers in Early Detection of Parkinson's Disease. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE42023040