

Research Article

Data Clustering Improves Siamese Neural Networks Classification of Parkinson's Disease

Mohamed Shalaby ¹, Nahla A. Belal ², and Yasser Omar ²

¹Egyptian Armed Forces, Cairo, Egypt

²College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Smart Village, Cairo, Egypt

Correspondence should be addressed to Nahla A. Belal; nahlabelal@aast.edu

Received 17 April 2021; Accepted 27 May 2021; Published 3 June 2021

Academic Editor: Danilo Comminiello

Copyright © 2021 Mohamed Shalaby et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Parkinson's disease (PD) is a clinical neurodegenerative disease having symptoms like tremor, rigidity, and postural disability. According to Harvard, about 60,000 of American citizens are diagnosed with PD yearly, with more than 10 million people infected worldwide. An estimate of 4% of the people have PD before they reach the age 50; however, the incident increases with age. Diagnosis of PD relies on the expertise of the physician and depends on several established clinical criteria. This makes the diagnosis subjective and inefficient. Hence, continuous efforts are being made to enhance the diagnosis of PD using deep learning approaches that rely on experienced neurologists. Siamese neural networks mainly work on two different input vectors and are used in comparison of output vectors. Moreover, clustering a dataset before applying classification enhances the distribution of similar samples among groups. In addition, applying the Siamese network can overcome the limitation of samples per class in the dataset by guiding the network to learn differences between samples rather than focusing on learning specific classes. In this paper, a Siamese neural network is applied to diagnose PD. Siamese networks predict the sample class by estimating how similar a sample is to other samples. The idea behind this work is clustering the dataset before training the network, as different pairs that belong to the same cluster are candidates to be mistaken by the network and assumed to be matched pairs. To overcome this problem, the dataset is first clustered, and then the architecture feeds the network to pairs of the same cluster. The proposed framework is concerned with comparing the performance when using clustered against unclustered data. The proposed framework outperforms the conventional framework without clustering. The accuracy achieved for classifying unclustered PD patients reached 76.75%, while it reached 84.02% for clustered data, outperforming the same technique on unclustered data. The significance of this study is in the enhanced performance achieved due to the clustering of data, which shows a promising framework to enhance the diagnostic capability of computer-aided disease diagnostic tools.

1. Introduction

Machine and deep learning are increasingly used in numerous fields. Medical and health applications are among those fields where machine learning and deep learning are used to diagnose, detect, and early predict diseases like Alzheimer's [1], cardiovascular disease [2], cancer [3], and Parkinson's disease [4]. The models used for diagnosing, detecting, and predicting diseases include algorithms such as decision trees [5], clustering [6], support vector machine [5], naïve bias [6], logistic regression [7], and neural network

[3, 5, 6]. Neural networks, especially deep neural networks, have high classification accuracy; however, these models fail when the number of samples used for training is small. Siamese neural network [1, 4] is one type of neural network model that works well under this limitation. Siamese neural network was first presented by [4] for signature verification, and this work was later extended for text similarity [8], face recognition [9, 10], video object tracking [11], and other image classification work [1, 12].

This work is motivated by the importance of Parkinson's disease (PD), which is a neurodegenerative disease (NDD)

that has motor-related symptoms, like tremor and instability [13]. Parkinson's disease patients are expected to reach 1% and 4% in people of age 60 and 80, respectively [14]. The key concern of computer-aided PD diagnosis is the early diagnosis of the first signs of PD for better quality of patients' life [15]. PD is diagnosed using many techniques and types of data. Some techniques rely on biomarkers [16]. Other methods use handwritten dynamics and speech assessment for PD diagnosis. Various machine learning techniques are used to diagnose PD using a variety of data types [17].

Siamese neural networks mainly work on two different input vectors and is used in comparison of output vectors. Moreover, clustering a dataset before applying classification enhances the distribution of similar samples among groups. In addition, applying the Siamese network can overcome the limitation of samples per class in the dataset by guiding the network to learn differences between samples rather than focusing on learning specific classes. A few number of researches addressed the concept of clustering prior to classification. The concept was applied to a number of datasets in [18], including lung cancer and Coli2000, and the clustering of data showed improved classification accuracy. Moreover, in [19], a combination of clustering with classification was proven to give an increase of up to 10% in accuracy. However, the choice of classification and clustering techniques is critical in achieving the increased performance. In this paper, a framework that clusters the Parkinson's dataset [20] is adopted, which then applies Siamese neural network to classify the patients of this dataset into three classes, namely, Parkinson's disease (PD), REM sleep behavior disorder (RBD), and healthy controls (HC). The proposed framework shows enhanced performance when the data is clustered, which is a promising enhancement that can be applied in computer-aided diagnostic systems. The accuracy achieved for classifying PD patients reached 76.75% for unclustered data, while it reached 84.02% for clustered data, outperforming the same technique on unclustered data. The significance of this study lies in the increased accuracy that is achieved due to the clustering of data, which shows promising performance enhancement when adding clustering before classification in computer-aided diagnostic systems.

The main contribution of this work can be highlighted in the points below:

- (i) Proposing a framework for clustering data before classification.
- (ii) Employing the k-means clustering in conjunction with Siamese neural networks.
- (iii) Applying the proposed model on Parkinson's disease patients.
- (iv) Showing that data clustering improves the diagnostic ability of computer-aided diagnostic systems.

The remaining sections of this paper are summarized as follows. Section 2 gives the work relevant to clustering, Siamese neural network, and PD diagnosis. In Section 3, the Methodology section, the proposed framework is presented. The framework is followed by the Results and Discussion section, Section 4. The results focus on a comparison of

applying Siamese neural network to both clustered Parkinson's dataset and Parkinson's dataset without clustering. Finally, the paper concludes and highlights some future suggestions in Section 5.

2. Related Work

In this section, the state-of-the-art methods of clustering and a review of Siamese neural networks are presented, in addition to an overview of Parkinson's disease diagnostic systems.

2.1. Clustering. Clustering is an unsupervised learning technique that groups instances together based on feature similarities, with the objective to increase similarity within the same class and decrease similarity between classes. The effectiveness of clustering is measured by its ability to identify unknown patterns. This is achieved by using distance measures such as Euclidean distance, Manhattan distance, and Minkowski distance [21]. Data clustering is used in many different applications, among which are machine learning, pattern recognition, and disease prediction. According to [21], data clustering can be categorized into linear and nonlinear clustering. Linear clustering algorithms include k-means clustering, quality threshold clustering, hierarchical clustering, fuzzy c-mean clustering, and Gaussian clustering, while nonlinear clustering algorithms include density-based clustering techniques such as minimum spanning tree-based clustering and kernel k-mean clustering. Clustering algorithms can be further categorized into partition-based, hierarchical, fuzzy, density-based, distribution-based, graph theory-based, model-based, and grid-based clustering [22]. Partition-based clustering is based on grouping instances according to the center of data points [22]. Algorithms based on partitioning include but are not limited to k-means [23], k-medoids [24], and CLARANS [25]. K-means is considered the most popular clustering technique; it is implemented by iteratively updating the center of a cluster (center of data) until a convergence criterion is met. K-medoids is an improved version of k-means that deals with discrete data. Hierarchical clustering algorithms are based on constructing hierarchical relationships among data [26]. Among hierarchical clustering algorithms are BIRCH [27] and CURE [28]. Fuzzy clustering algorithms label instances by changing the discrete value of belonging to a certain cluster into a continuous interval. FCM [29], FCS [30], and MM [31] are among the well-known fuzzy clustering algorithms. Density-based algorithms use the density of data for clustering, where the data belonging to high density regions are considered in the same cluster [32]. Distribution-based clustering, like DBCLASD [33] and GMM [34], generates clusters of data from the same distribution. In graph theory-based clustering, data is represented as a graph, where the nodes are used to represent the data points and the edges represent relationships among data. Examples of algorithms in this category of clustering are minimum spanning tree-based clustering [35]. Model-based clustering selects a specific model to represent each cluster and then selects the data that

best fits each cluster according to that model, using statistical learning or neural network learning. Examples of algorithms in this category are GMM [34]. Grid clustering represents the original data on a grid and uses this structure to cluster data points. Algorithms for grid clustering include STING [36] and CLIQUE [37]. More recent clustering algorithms include user kernel ensemble, quantum theory, swarm intelligence, spectral graph theory, affinity propagation, clustering algorithms for large-scale data, and spatial data [22].

2.2. Siamese Neural Networks. A Siamese neural network is a type of neural network that is used to solve the problem of one-shot learning [38], where the class must be correctly predicted even if only a few examples are available for each new class. A Siamese neural network is structured into two similar neural networks (sometimes called twins); Figure 1 shows the architecture of a Siamese neural network. The input to Siamese neural network, used for training, is a couple of samples, one sample for the top twin and the other for the bottom one, in addition to a label that shows whether the two samples belong to the same class or not. The output of each twin network is a feature vector; these two feature vectors are combined through a cost function; the output of this function is a scalar energy. The output of the cost function is then combined with the label through a loss function. In the training phase, the network parameters are updated using the backpropagation method such that the loss function value is minimum for the pairs that belong to the same class and the maximum for the pairs that belong to different classes. In [6, 39], the authors used a contrastive energy function as a loss function. Koch et al. [40] used a different approach, as they used L1-distance function followed by a sigmoid activation function. The concept of Siamese neural network was extended to what is called “triplet loss,” where three samples are used as input to the neural network, namely, the anchor, the positive sample (which belongs to the same class of the anchor), and the negative sample (which belongs to a class different from the anchor class) [41, 42]. The idea behind the triplet loss is the minimization of the distance between the anchor and positive samples and the maximization of the distance between the anchor and the negative samples. In 2018, Utkin et al. [43] introduced an alternative approach to Siamese neural network which is called Siamese deep forest. It is based on gcForest [44], a structure consisting of multiple layers, where each layer consists of groups of decision trees. The Siamese deep forest prevents the overfitting from happening in the conventional neural networks due to the limitation of the available training data.

2.3. Parkinson’s Disease Diagnosis. Different sources of data are employed to help in the diagnosis of PD. This shows the significance of the disease. Among those data types are images, speech, sensor data, and handwriting motor data. The study presented in [45] on the diagnosis of Parkinson’s disease (PD) discusses the features that can be used in diagnosis. They employed a Med-Line search to measure the clinical characteristics of PD. The study highlighted the

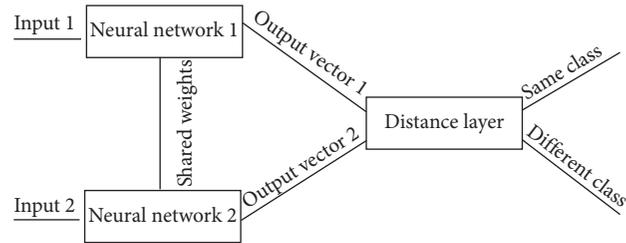


FIGURE 1: The Siamese neural network architecture.

importance of clinical characteristics, such as tremor, rigidity, and loss of postural reflexes, to differentiate PD from other diseases. Based on this study, the dataset used in this paper was selected. Moreover, it was concluded that genetic variations and neuroimaging tests may also be used to improve diagnosis.

Pereira et al. [46] used various CNN architectures to classify PD from non-PD patients and proposed a model with an accuracy of up to 95% using handwritten images. In addition, Moetesum et al. [47] used CNN to extract visual features from different samples of handwritten images and a support vector machine (SVM) classifier was used for classification with an accuracy of 83%. Singh and Nasoz [48] also worked on handwritten images to decrease the loss function of the images from the validation set. Classification accuracy reached 83.11% and 90.38%, for meander and spiral tests, respectively, using CNNs and SVM. An accuracy of 88% was achieved by Khatamino et al. [49] using a CNN system that employs dynamic features of spirals, as well as visual attributes, to detect PD. A fusion of a fuzzy system with neural networks was also proposed in [50] with an improved performance over other classification methods. An expert system using a genetic algorithm and wavelet kernel extreme learning was proposed in [51] with an accuracy of 96.81%.

Other methods rely on speech data. For example, Al-Fatlawi et al. [52] used DBN with two stacked RBMs. The proposed system reached an accuracy of 94%. Furthermore, in [53], speech impairments were used to diagnose PD using the deep neural network (DNN) classifier. The obtained accuracy was up to 93.79%. Additionally, Gunduz [54] proposed two CNN-based frameworks to classify PD using vocal features with an accuracy of 86.9%. The work in [55] used a speech dataset with many machine learning techniques. The highest performance reported was for the Light Gradient Boosting model, with an area under curve of 0.951.

Other techniques use image analysis on different scans. In this context, Segovia et al. [56] used brain images for classification using SVM and the Partial Least Squares were used. Accuracy reached up to 94.7%. In [57], a deep CNN model was used on SPECT images. The performance of this approach reached an accuracy of 90.7%. Finally, Sivaranjini and Sujatha [58] employed a CNN architecture to classify the MRI of healthy controls and PD patients with an accuracy of 88.9%.

Techniques that use sensor data are also employed in some studies for the diagnosis of PD. Aich et al. [59] showed that decision trees outperformed other methods, obtaining an accuracy of 88.46% to study the effect of medicine on the

gait characteristics of PD patients. Wearable sensors attached to the wrist are used to detect FoG. Moreover, Xu et al. [60] used EEG signals to feed a deep neural network, pooling-based deep recurrent neural network (DPRNN). The proposed DPRNN model achieved sensitivity, specificity, and precision of 84.84%, 91.81%, and 88.31%, respectively. Recently, [61] presented an automated gait differentiation procedure for the diagnosis of PD through a holistic, nonintrusive method that uses Vertical Ground Reaction Force (VGRF). Gait features are extracted from the VGRF for training of a neural network that achieves an accuracy of 97.4%.

3. Methodology

This section explains the methods used in the proposed framework to classify Parkinson disease patients.

In this context, k-means clustering is first applied to the dataset; then the Siamese neural network is employed to classify the clustered dataset. The k-means method is used because it does not need complex computation. However, prior knowledge of the number of clusters is a disadvantage to this method. This is overcome by choosing the optimal number of clusters which minimizes the distance between each sample vector in the dataset and the centroid of its corresponding cluster.

3.1. Parkinson Disease Dataset. The original Parkinson disease dataset includes 130 patients prepared by professional neurologists, so that the proposed framework is not affected by the subjectivity of nonprofessional neurologists and clinicians. Patients are classified into three classes, namely, early untreated Parkinson’s disease (PD), REM sleep behavior disorder (RBD), and healthy controls (HC). The dataset contains 30 patients with early untreated Parkinson’s disease (PD), 50 patients with REM sleep behavior disorder (RBD), which have a high risk of developing the disease or other synucleinopathies, and 50 healthy controls (HC). Each sample in the dataset is described by 65 features collected from patients [20]. These features include age, gender, positive history of Parkinson disease in the family, medications, dosage, motor examination, speech, facial expression, measures of tremor, measures of rigidity, finger taps, hand movements, leg agility, posture, and respiration. The features in the dataset represent clinical data, as well as data obtained from hand movements (as in handwriting images), and data to measure speech features. Figure 2 shows the distribution of the dataset on age, gender, age of disease onset, and duration of disease from first symptoms. For gender, 0 represents male and 1 represents female. Year 0 in the graph indicates a healthy person. The dataset consists of 103 males and 27 females, their ages range from 34 to 83 years, the age where they first experience the symptoms ranges from 30 to 81 years, and the time duration of disease from first symptoms ranges from 0.5 to 17 years.

3.2. Data Preprocessing. The data preprocessing prepares the data to be suitable for the proposed framework.

The steps done in the data preprocessing step are as follows:

- (i) The “patient code” field is eliminated as it has no effect on the learning process.
- (ii) The string values of the data field “Gender” are enumerated, where 0 represents male and 1 represents female.
- (iii) The fields “Positive,” “Antidepressant therapy,” “Antiparkinsonian medication,” “Antipsychotic medication,” and “Benzodiazepine medication” are enumerated.
- (iv) The healthy control class has no applicable value for the “Age of disease onset” and “Duration of disease from the first symptoms,” and therefore, in the preprocessing stage those values are replaced by zeros.
- (v) The missing values from the dataset are replaced by the value -1 . This approach is selected to avoid deleting samples and losing information and to avoid using calculated values, like median or mean, which may lead to variance in results. The 130 patients’ dataset can generate $130 \times 129 = 16,770$ pairs, of which 747 and 296 pairs are used for training and testing, respectively. The pairs are created randomly from the same clusters such that the percentage of positive pairs to the negative pairs is 1 : 2 because we have 3 classes.

3.3. K-Means Clustering. Clustering is carried out via the k-means clustering algorithm. The algorithm is used to cluster the Parkinson disease dataset. The input to the k-means algorithm is the number of clusters, M , and the size of the dataset, N . Choosing the most optimum number of clusters depends on minimizing the cluster inertia, I , which measures the sum of the Euclidean distances between each sample vector in the dataset and the centroid of its corresponding cluster according to

$$I = \sum_j \sum_i w_{ij} \sqrt{(X_i - C_j)^2}, \quad (1)$$

where $i \in \{1, 2, \dots, N\}$, $j \in \{1, 2, \dots, M\}$, X_i is the vector that represents sample i , in the dataset, C_j is the centroid of cluster j , and w_{ij} equals 1 when $X_i \in$ cluster j and 0, otherwise. There is a trade-off between the number of clusters and the inertia. To solve this problem, the elbow method [62] is used. The elbow method examines the relation between the number of clusters and inertia and finds the point after which the decrease in inertia is not significant and then considers the number of clusters corresponding to this point as the optimal number of clusters. Figure 3 shows how to get the optimal number of clusters graphically.

To get the elbow point mathematically, the perpendicular distance between each point of the curve and the line that connects the first point of the curve with the last point,

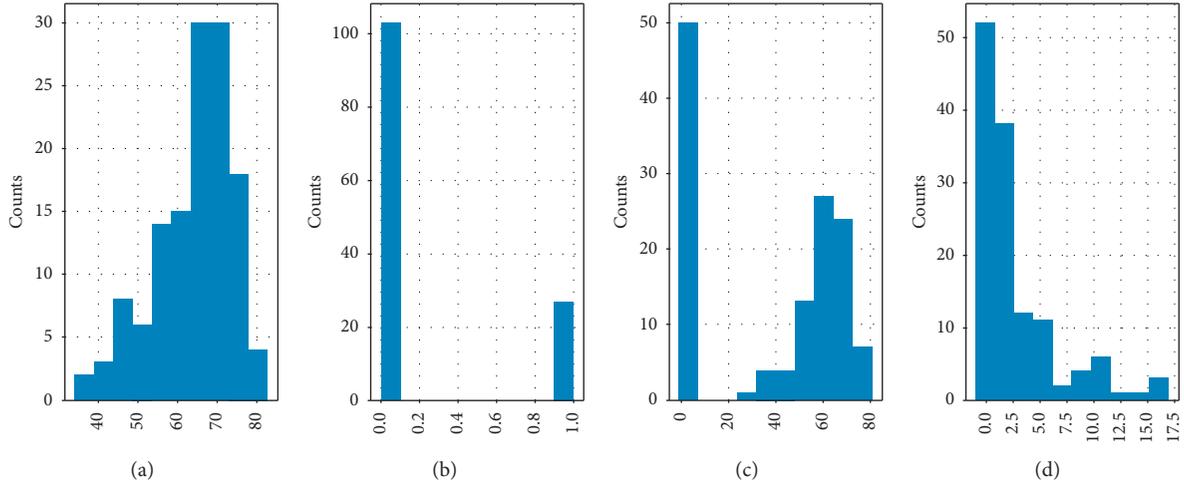


FIGURE 2: The distribution of Parkinson's disease dataset against (a) age, (b) gender, (c) age of disease onset (years), and (d) duration of disease from first symptoms (years).

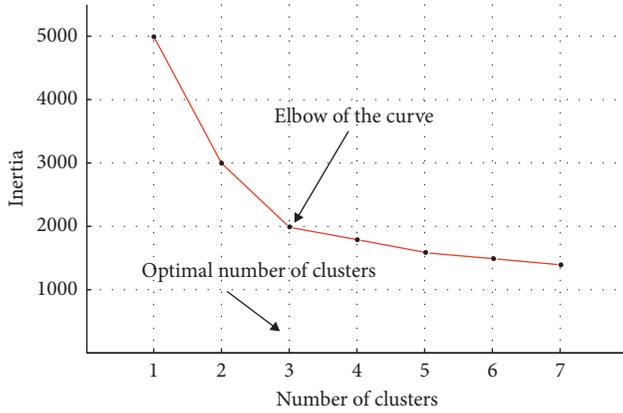


FIGURE 3: The relation between the number of clusters and inertia. The optimal number of clusters corresponds to the elbow of the curve.

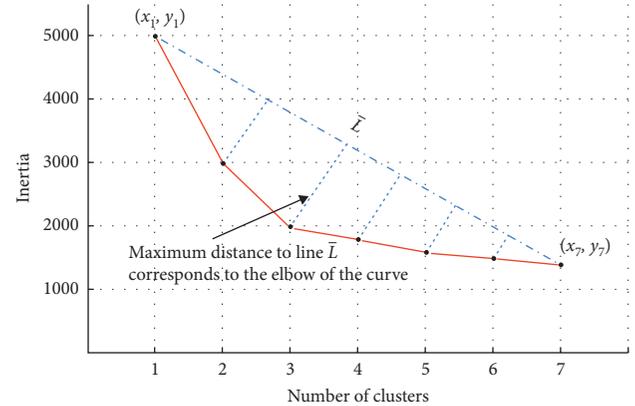


FIGURE 4: Obtaining the elbow point of the curve mathematically by calculating the maximum of perpendicular distances between curve points and line \bar{L} .

\bar{L} , is calculated. The elbow point corresponds to the maximum distance, d_{\max} , as shown in Figure 4.

The distance d_{\max} is calculated using

$$d_{\max} = \max \left\{ \frac{|x_k(y_M - y_1) - y_k(x_M - x_1) + x_M y_1 - x_1 y_M|}{\sqrt{(y_M - y_1)^2 + (x_M - x_1)^2}} \right\}, \quad (2)$$

where (x_1, y_1) is the first point of the curve, (x_M, y_M) is the last point of the curve, and (x_k, y_k) is the point from which the distance to the line, \bar{L} , is calculated.

After the optimal number of clusters is determined, the k-means algorithm is implemented according to the following steps:

- (1) Obtain the centroid of each cluster randomly.
- (2) Assign each sample to its nearest centroid.
- (3) Update the centroid of each cluster.
- (4) If the centroids change, repeat steps 2 and 3.

3.4. Siamese Neural Network Framework. The Siamese neural network is a feedforward network with error back-propagation. The network comprises of two indistinguishable feedforward neural networks joined at their yield, as shown in Figure 1. Amid training, each network peruses a profile made of genuine values and forms its values at each layer. The network enacts a few of the neurons based on these values and upgrades its weights through error back-propagation, and at the end it creates an output profile that is compared with the output of the other network. The Siamese neural network compares the output of the upper and the lower networks by calculating a distance metric. Through this distance, the network states that the two outputs are distinctive or comparative. The algorithm at that point names occurrence as positive or negative, depending on the distance metric. The final output can at last be compared with its corresponding ground truth value.

The proposed Siamese neural network framework consists of twin neural networks (top and bottom networks), where each of them consists of three dense layers (other than the input layer). The input layer consists of 64 neurons, and

each other layer consists of 128 neurons. The activation function applied for the first three layers is “tanh.” To avoid overfitting, dropout is applied between layer 1 and layer 2, and layer 2 and layer 3 with fraction 10%. Table 1 shows a summary of both top and bottom neural networks. The two twin neural network outputs are joined together to form the output layer with one neuron, where the Euclidean distance is applied between the outputs of the top and bottom networks to measure the similarity between the two outputs.

The dataset is split into two partitions, such that two-thirds of the dataset are dedicated for training and the rest is dedicated for testing. Random pairs of records of the dataset are created; the pair is positive if the two records belong to the same class; otherwise the pair is negative. The data is fit to the model in batches. The batch size is 128. The network is trained through 200 epochs. The objective of the network is minimizing positive pairs distance, while maximizing negative pairs distance. This is done using the contrastive loss function:

$$L = \begin{cases} D(X, Y_p), & \text{for positive pairs,} \\ \max(0, m - D(X, Y_n)), & \text{for negative pairs,} \end{cases} \quad (3)$$

where L is the loss function, (X, Y_p) is a positive pair, (X, Y_n) is a negative pair, D is the distance between two records of the same pair, and m is the margin value which shows that the two records of the negative pair are distant enough. To get the values of network weights at which the loss is optimum, the RMSprop optimization algorithm is applied.

4. Results and Discussion

This section presents the results of the experiment on Parkinson’s disease dataset. The proposed model is analyzed with the performance metrics accuracy, precision, recall, specificity, and f_1 -score [20]. In a binary classification problem (where we have only positive and negative classes), the performance metrics are calculated using True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), where TP and TN are the number of positive samples and negative samples that are correctly predicted, and FP and FN are the number of positive and negative samples that are falsely predicted. The total number of test samples is given by N . The performance metrics are calculated as shown in equations.

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{N}, \\ \text{precision} &= \frac{TP}{TP + FP}, \\ \text{recall} &= \frac{TP}{TP + FN}, \\ \text{specificity} &= \frac{TN}{TN + FP}, \\ f_1\text{-score} &= \frac{2 \times \text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}. \end{aligned} \quad (4)$$

TABLE 1: A summary of both top and bottom twin networks of the proposed Siamese neural network model.

Layer	Type	Output shape
Input layer	Input layer	64
Layer 1	Dense	128
Layer 2	Dense	128
Layer 3	Dense	128

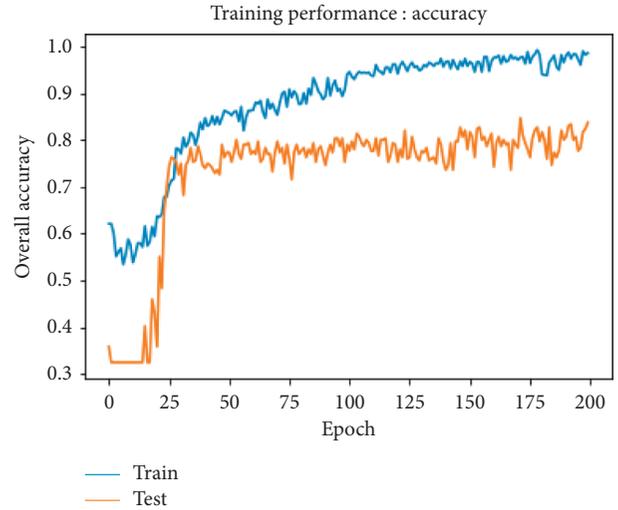


FIGURE 5: The overall accuracy on the training set and the accuracy on the validation set for the proposed model training.

However, our classification problem is multiclass (where there are more than two classes); in this case the performance metrics are calculated with respect to each class, and then the weighted average is calculated for each metric. The weighted average for metric p is $\sum_{i=1}^C N_i p_i / N$, where C is the number of classes, N_i is the number of test samples in the class, and p_i is the value of this metric with respect to class i .

4.1. Parkinson’s Disease Experiment. Applying k-means algorithm to cluster Parkinson disease dataset results in partitioning the data into eight clusters. Two-thirds of the dataset are used for creating 747 pairs to train the proposed Siamese neural network framework, and the rest is used to create 296 pairs to test it. Figure 5 shows the overall accuracy on both the training set and the validation set, for the proposed framework during training. Figure 6 shows the loss on both the training set and the validation set, for the proposed framework during training. The performance of the proposed framework changes insignificantly at epoch 175; hence, the iteration can be terminated as there is no significant change after that.

To study the effect of clustering dataset, the Siamese neural network model is applied on the dataset before clustering, and then the model is applied on the clustered dataset. Table 2 shows the performance metrics in the two cases, clustered and unclustered Parkinson’s dataset. The weighted average Precision/Recall/ F_1 (PRF) of the unclustered model is 71.55%, 70.83%, and 71.19%, and the accuracy



FIGURE 6: The loss on the training set and the loss on the validation set for the proposed model training.

TABLE 2: A comparison between applying the proposed model on Parkinson dataset without clustering and clustered PD dataset with the performance metrics accuracy, precision, recall, specificity, and f_1 -score.

Class	Parkinson dataset without clustering					Clustered Parkinson dataset				
	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F_1 -score (%)	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F_1 -score (%)
PD	71.88	56.76	65.63	75	60.87	77.05	50	42.86	87.23	46.15
RBD	72.92	57.89	68.75	75	62.86	75.36	60.87	63.64	80.85	62.22
HC	92.71	100	78.13	100	87.72	100	100	100	100	100
Weighted average	79.17	71.55	70.83	83.33	71.19	85.4	73.06	72	90.06	72.53

TABLE 3: The overall accuracy of applying the model on the dataset with/without clustering on the PD dataset.

Iteration number	Parkinson dataset without clustering (%)	Clustered Parkinson dataset (%)
1	78.03	84.9
2	82.85	87.7
3	75.25	84.7
4	76.52	87.4
5	74.75	85.7
6	78.54	81.5
7	78.28	84.2
8	76.01	80.2
9	75.76	83.2
10	71.46	80.7
Average	76.75	84.02

is 79.17%. The PRF of the clustering model is 73.06%, 72.00%, and 72.53%, and the accuracy is 85.40%. This increase in accuracy comes from better predicting the true-negative and true-positive samples.

The system iterates the above experiment 10 times. Table 3 shows the overall accuracy for each iteration. Applying the model on the clustered dataset outperforms applying it on the dataset without clustering. On average, the overall accuracy is 84.02% and 76.75%, correspondingly.

5. Conclusions

In this work, we adopt a new method of training Siamese neural networks on Parkinson’s disease dataset. The new method depends on clustering the dataset prior to the training phase and concentrates on training the network on pairs of the same clusters. We have compared the proposed framework to the conventional framework. We run both models (the conventional and proposed framework) 10 times. On average, we get an overall accuracy of 76.75% when applying the conventional model, while we get an accuracy of 84.02% when applying the proposed framework. Moreover, both frameworks were analyzed with the performance metrics, accuracy, precision, recall, specificity, and f_1 -score. On average, while the conventional model achieves 79.17% accuracy, 71.55% precision, 70.83% recall, 83.33% specificity, and 71.19% f_1 -score, the proposed framework achieves 85.4% accuracy, 73.06% precision, 72% recall, 90.06% specificity, and 72.53% f_1 -score. The above experimental results prove that the proposed framework outperforms the conventional framework. The proposed framework shows a promising improvement in performance, and hence, the model is expected to be further tested on multiple datasets for Parkinson’s disease, in addition to other classification problems.

Data Availability

The data used in this manuscript were taken from [20] and are available at <https://archive.ics.uci.edu/ml/machine-learning-databases/00392/>. Any other data required are available upon request from the author.

Conflicts of Interest

The authors have no conflicts of interest.

References

- [1] C.-F. Liu, S. Padhy, S. Ramachandran et al., "Using deep siamese neural networks for detection of brain asymmetries associated with Alzheimer's disease and mild cognitive impairment," *Magnetic Resonance Imaging*, vol. 64, pp. 190–199, 2019.
- [2] R. Sivaranjani and D. N. Yuvaraj, "Artificial intelligence model for earlier prediction of cardiac functionalities using multilayer perceptron," *Journal of Physics: Conference Series*, vol. 1362, 2019.
- [3] S. Agrawal and J. Agrawal, "Neural network techniques for cancer prediction: a survey," *Procedia Computer Science*, vol. 60, pp. 769–774, 2015.
- [4] J. Bromley, I. Guyon, Y. Lecun, E. Sackinger, and R. Shah, "Signature verification using a siamese time delay neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 669–688, 1993.
- [5] K. Kouroua, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.
- [6] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, 2006.
- [7] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, Wiley, Hoboken, NJ, USA, 2013.
- [8] K. Amin, G. Lancaster, S. Kapetanakis, K. Althoff, A. Dengel, and M. Petridis, "Advanced similarity measures using word embeddings and siamese networks in CBR," in *Intelligent Systems and Applications, IntelliSys*, Springer, Cham, Switzerland, 2019.
- [9] N. Chakraborty, A. Dan, A. Chakraborty, and S. Neogy, "Effect of dropout and batch normalization in siamese network for face recognition," *Advances in Intelligent Systems and Computing*, vol. 1059, pp. 21–37, 2020.
- [10] W. Wang, J. Yang, J. Xiao, S. Li, and D. Zhou, "Face recognition based on deep learning," in *Human Centered Computing, HCC 2014, Lecture Notes in Computer Science*, Springer, Cham, Switzerland, 2015.
- [11] A. D. Rhodes and M. Goel, "Deep siamese networks with Bayesian non-parametrics for video object tracking," in *Proceedings of the Future Technologies Conference (FTC)*, The Park Central Hotel, San Francisco, CA, USA, 2019.
- [12] M. Zinzuvadiya, V. Dhameliya, S. Vaghela, S. Patki, N. Nanavati, and A. Bhavsar, "Co-detection in images using saliency and siamese networks," *Advances in Intelligent Systems and Computing*, vol. 1024, pp. 351–362, 2019.
- [13] S. Chakraborty, S. Aich, E. Han, J. Park, and H. Kim, "Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: a multistage classifier approach," in *Proceedings of the 22nd International Conference on Advanced Communication Technology (ICACT)*, pp. 298–303, Phoenix Park, PyeongChang, South Korea, 2020.
- [14] O. B. Tysnes and A. Storstein, "Epidemiology of Parkinson's disease," *Journal of Neural Transmission*, vol. 124, pp. 901–905, 2017.
- [15] C. R. Pereira, D. R. Pereira, J. Papa, G. H. Rosa, and X. S. Yang, "Convolutional neural networks applied for Parkinson's disease identification," *Machine Learning for Health Informatics*, vol. 9605, pp. 377–390, 2016.
- [16] M. C. T. Dos Santos, R. Bell, and A. N. Da Costa, "Recent developments in circulating biomarkers in Parkinson's disease: the potential use of miRNAs in a clinical setting," *Bioanalysis*, vol. 8, no. 23, pp. 2497–2518, 2016.
- [17] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.
- [18] Y. K. Alapati and K. Sindhu, "Combining clustering with classification: a technique to improve classification accuracy," *International Journal of Computer Science Engineering*, vol. 5, no. 6, pp. 336–338, 2016.
- [19] T. Chakraborty, "EC3: combining clustering and classification for ensemble learning," in *Proceedings of the IEEE International Conference on Data Mining*, pp. 781–786, New Orleans, LA, USA, 2017.
- [20] J. Hlavnicka, R. Cmejla, T. Tykalova, K. Sonka, E. Ruzicka, and J. Ruzs, "Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder," *Nature*, vol. 7, no. 1, 2017.
- [21] S. Bano and M. N. Khan, "A survey of data clustering methods," *International Journal of Advanced Science and Technology*, vol. 113, pp. 133–142, 2018.
- [22] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.
- [23] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, University of California, Berkeley, Ca, USA, 1967.
- [24] H. S. Park and C. H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336–3341, 2009.
- [25] R. T. Ng and J. Han, "Clarans: a method for clustering objects for spatial data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 5, pp. 1003–1016, 2002.
- [26] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
- [27] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," *ACM SIGMOD Record*, vol. 25, no. 2, pp. 103–114, 1996.
- [28] S. Guha, R. Rastogi, and K. Shim, "CURE: an efficient clustering algorithm for large databases," *ACM SIGMOD Record*, vol. 27, no. 2, pp. 73–84, 1998.
- [29] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984.
- [30] R. N. Dave and K. Bhaswan, "Adaptive fuzzy c-shells clustering and detection of ellipses," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 643–662, 1992.

- [31] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 8, pp. 1279–1284, 1994.
- [32] H. P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 231–240, 2011.
- [33] X. Xu, M. Ester, H. P. Kriegel, and J. Sander, "A distribution-based clustering algorithm for mining in large spatial databases," in *Proceedings 14th International Conference on Data Engineering*, pp. 324–331, Orlando, FL, USA, February 1998.
- [34] C. Rasmussen, "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems*, vol. 12, pp. 554–560, 1999.
- [35] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [36] W. Wang, J. Yang, and R. Muntz, "STING: a statistical information grid approach to spatial data mining," *VLDB*, vol. 97, pp. 186–195, 1997.
- [37] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, pp. 94–105, Seattle, WA, USA, 1998.
- [38] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 594–611, 2006.
- [39] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 539–546, San Diego, CA, USA, June 2005.
- [40] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proceedings of the ICML Deep Learning Workshop*, Lille, France, 2015.
- [41] F. Schroff, D. Kalenichenko, J. Philbin, and "Facenet, "A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, Boston, MA, USA, June 2015.
- [42] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. 2, 2009.
- [43] L. V. Utkin and M. A. Ryabinin, "A siamese deep forest," *Knowledge-Based Systems*, vol. 139, pp. 13–22, 2018.
- [44] Z. H. Zhou and J. Feng, "Deep forest: towards an alternative to deep neural networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pp. 3553–3559, Melbourne, Australia, August 2017.
- [45] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.
- [46] C. R. Pereira, D. R. Pereira, G. H. Rosa et al., "Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification," *Artificial Intelligence in Medicine*, vol. 87, pp. 67–77, 2018.
- [47] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders—a case study on Parkinson's disease," *Pattern Recognition Letter*, vol. 121, pp. 19–27, 2019.
- [48] S. Singh and F. Nasoz, "Facial expression recognition with convolutional neural networks," in *Proceedings of the 2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 324–328, Las Vegas, NV, USA, January 2020.
- [49] P. Khatamino, I. Cantürk, and L. Özyılmaz, "A deep learning-CNN based system for medical diagnosis: an application on Parkinson's disease handwriting drawings," in *Proceedings of the 6th International Conference on Control Engineering & Information Technology (CEIT)*, pp. 1–6, Istanbul, Turkey, October 2018.
- [50] R. H. Abiyev and S. Abizade, "Diagnosing Parkinson's diseases using fuzzy neural system," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 1267919, 2016.
- [51] D. Avci and A. Dogantekin, "An expert diagnosis system for Parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine," *Parkinson's Disease*, vol. 2016, Article ID 5264743, 2016.
- [52] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, "Efficient diagnosis system for Parkinson's disease using deep belief network," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1324–1330, Vancouver, BC, Canada, July 2016.
- [53] A. Caliskan, H. Badem, A. Basturk, and M. E. Yuksel, "Diagnosis of the Parkinson disease by using deep neural network classifier," *Istanbul University-Journal of Electrical and Electronics Engineering*, vol. 17, pp. 3311–3318, 2017.
- [54] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019.
- [55] I. Karabayir, S. M. Goldman, S. Pappu et al., "Gradient boosting for Parkinson's disease diagnosis from voice recordings," *BMC Medical Informatics Decision Making*, vol. 20, no. 228, 2020.
- [56] F. Segovia, J. Gorriiz, J. Ramirez, I. Alvarez, J. Jiménezhoyuela, and S. Ortega, "Improved parkinsonism diagnosis using a partial least squares based approach," *Medical Physics*, vol. 39, no. 7, pp. 4395–4403, 2012.
- [57] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.
- [58] S. Sivaranjini and C. M. Sujatha, "Deep learning based diagnosis of Parkinson's disease using convolutional neural network," *Multimed Tools and Applications*, vol. 79, pp. 15467–15479, 2020.
- [59] S. Aich, P. Pradhan, S. Chakraborty et al., "Design of a machine learning-assisted wearable accelerometer-based automated system for studying the effect of dopaminergic medicine of gait characteristics of Parkinson's patients," *Journal of Healthcare Engineering*, vol. 2020, Article ID 1823268, 2020.
- [60] S. Xu, Z. Wang, J. Sun et al., "Using a deep recurrent neural network with EEG signal to detect Parkinson's disease," *Annals of Translational Medicine*, vol. 8, no. 14, pp. 1–13, 2020.
- [61] S. Veeraragavan, A. Gopalia, D. Gouwanda, and S. Ahmad, "Parkinson's disease diagnosis and severity assessment using ground reaction forces and neural networks," *Frontiers in Physiology*, vol. 11, 2020.
- [62] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.