IET Image Processing

Special Section: Multidisciplinary advancement of imaging technologies: from medical diagnostics and genomics to cognitive machine vision, and artificial intelligence

IET Journals
The Institution of Engineering and Technology

Unified deep learning approach for prediction of Parkinson's disease

ISSN 1751-9659
Received on 18th November 2019
Revised 11th March 2020
Accepted on 29th April 2020
E-First on 22nd June 2020
doi: 10.1049/iet-ipr.2019.1526
www.ietdl.org

James Wingate[1], Ilianna Kollia[2] ✉, Luc Bidaut[1], Stefanos Kollias[1,2]

[1]*School of Computer Science, University of Lincoln, Brayford Pool, LN6 7TS, Lincoln, UK*
[2]*School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou street, Zografou, 15780, Athens, Greece*
✉ *E-mail: ilianna2@mail.ntua.gr*

Abstract: The study presents a novel approach, based on deep learning, for diagnosis of Parkinson's disease through medical imaging. The approach includes analysis and use of the knowledge extracted by deep convolutional and recurrent neural networks when trained with medical images, such as magnetic resonance images and dopamine transporters scans. Internal representations of the trained DNNs constitute the extracted knowledge which is used in a transfer learning and domain adaptation manner, so as to create a unified framework for prediction of Parkinson's across different medical environments. A large experimental study is presented illustrating the ability of the proposed approach to effectively predict Parkinson's, using different medical image sets from real environments.

1 Introduction

Current biomedical signal analysis, including medical imaging, has been used for long based on feature extraction combined with quantitative and qualitative processing. Recent advances in machine learning (ML) and deep neural networks (DNNs) have provided state-of-the-art performance in major signal processing tasks, such as computer vision, speech recognition, human computer interaction and natural language processing. DNNs can be trained as end-to-end architectures which include different network types and provide numerical or symbolic outputs [1]. Medical diagnosis is an area in which ML and DNNs can be effectively used. This is due to their ability to analyse big amounts of data, signals, images and image sequences, to find patterns in them and use those for effective classification, regression and prediction purposes. Various promising results have been obtained in a variety of problems [2–4].

Parkinson's is one of the most common neurodegenerative disorders among people from 50 to 70 years old, especially in countries with elderly population, such as the United States and the European Union. Early detection and prognosis are crucial for assisting patients to retain a good quality of life. Therefore, developing techniques that are able to provide accurate and trustworthy prediction of Parkinson's in subjects is of major significance for a society that cares about people's well-being.

Prediction of Parkinson's [5, 6] can be based on analysis of medical images, in particular magnetic resonance images (MRIs) and dopamine transporters scans (DaTscans). MRI analysis principally focuses on the detection of morphological variations in brain areas, especially examining the volume of the surface of substantia nigra, the lenticular nucleus and the head of the caudate nucleus. DaTscans are produced by single photon emission computer tomography, with 123-I-Ioflupane being provided to the patients. DaTscans are used for detecting whether there is degeneration of dopaminergic neurons in the substantia nigra. For the diagnosis of Parkinson's, doctors focus on the images and scans that are considered most representative, select the areas around the caudate nucleus head, make comparison with the cerebellum, calculate and use ratios of defined volumes for making their prediction.

ML and classification methods [7] have been used for diagnosis of Parkinson's based on MRIs [8], or DaTscans [9] in the last decade. Recent developments in deep learning have provided further progress in this direction. Deep convolutional neural network and recurrent neural networks (CNN-RNNs) have been developed and used for prediction of Parkinson's [10], achieving high prediction accuracy based on a new Parkinson's database including MRI and DaTscan image data [11].

However, although DNNs are capable of analysing complex data, they lack transparency in their decision-making, in the sense that it is not straightforward to justify their prediction, or to visualise the features on which the decision was based. Moreover, they generally require large amounts of data in order to learn and become able to adapt to different medical environments, or different patient cases. This makes their use difficult in health care, where trust and personalisation are key issues.

In this study we adopt the DNN architecture developed in [10] as a model that can potentially be applied to other medical environments, or respective data sets. However, the latter generally include medical images with different characteristics, e.g. scans can be colour or gray-scale, they may have different sizes, or there can be different numbers of images per subject. As a consequence, direct application of trained DNN to other data sets is not generally successful.

Various methods can be used to address this problem. Training the DNN model from scratch with each new data set is a possibility, but this would result in creating many different DNNs solving the same problem albeit for different data cases and with no interoperability among them. Merging all possible data sets, so that a single DNN is trained on all of them, would be another possibility, but this is rather unfeasible due to issues with both implementation and privacy.

Transfer learning is another approach usually adopted in deep learning methodologies [12, 13], according to which the DNN model trained with the original data set is used to initialise DNN re-training with the new data set. However, a serious problem then arises: as the refined DNN learns to predict from the new data set, it tends to forget the old data that are not used in the retraining procedure; this is known as ‘catastrophic forgetting’. As in learning from scratch, local rather than global prediction models would be generated through such an approach.

Recent research has focused on extracting trained DNN representations and using them for classification purposes [3, 14], either by an auto-encoder methodology, or by monitoring neuron

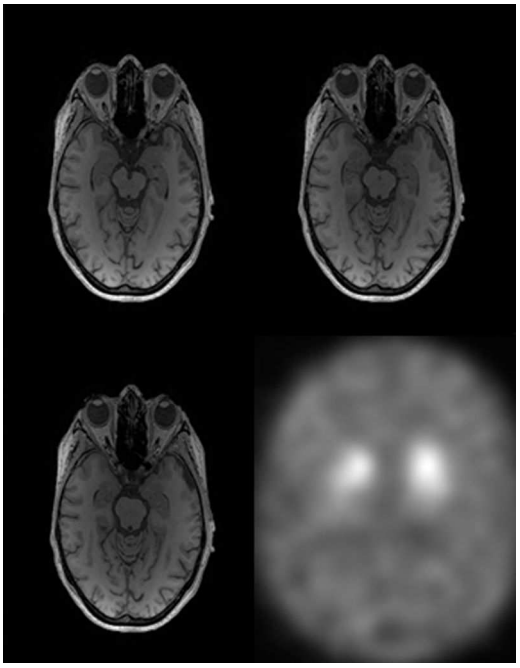

Fig. 1 *A DNN input including a DaTscan (gray scale, bottom right hand side) and three consecutive MRIs (top, and bottom left hand side) from the PPMI data set [15]*

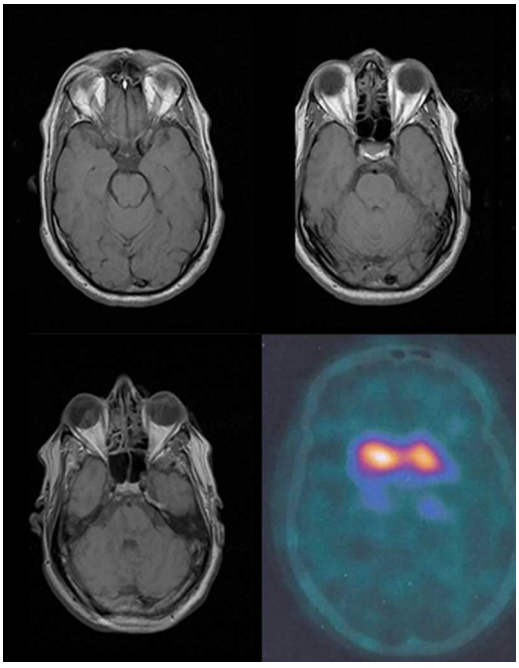

Fig. 2 *A DNN input including a DaTscan (colour, bottom right hand side) and three consecutive MRIs (top, and bottom left hand side) from the data set [11]*

outputs in the convolutional or/and fully connected (FC) network layers. Such developments are what is exploited in this study, proposing a novel approach that is able to overcome the above mentioned shortcomings and problems, while generating a unified prediction model for Parkinson's based on DaTscans and/or MRI data.

At first, we extract appropriate internal features, say features $\boldsymbol{v}$, from the DNN model trained with the data set developed in [11]. Using a clustering methodology, we generate concise representations, say $\boldsymbol{c}$, of these features, which are then annotated by medical experts to denote patient or non-patient categories. Using these representations and the nearest neighbour criterion, we can then predict, in an efficient and transparent way, whether new subjects' data indicate Parkinson's status or not.

We then present a new transfer-learning methodology that alleviates the 'catastrophic forgetting' problem by generating a unified model over different data sets. According to this methodology, we apply the originally trained DNN to a new data set deriving a corresponding set of representations, through which we train a new DNN. From the latter DNN, we extract a new set of features, say $\boldsymbol{v}'$ and a concise representation $\boldsymbol{c}'$. The unified Parkinson's prediction model is then produced by merging the $\boldsymbol{c}$ and $\boldsymbol{c}'$ representation sets. Having achieved high precision and recall metrics in the derivation of each one of these representations ensures that the generated unified model provides high prediction accuracy in the derived representation space.

We also show that the proposed approach can improve Parkinson's prediction in cases and environments where some input data types, e.g. DaTscans, are not available and prediction is made only through MRI analysis. A domain adaptation methodology for new DNN training is presented, which uses a novel error criterion based on the above-described $\boldsymbol{c}$ representations.

Finally, we provide an extensive experimental study, in which we develop, adapt and evaluate DNNs, using two different data sets: the database described in [11] and the Parkinson's progression markers initiative (PPMI) database [15], which include MRI and DaTscan data as well as textual information from patients and controls.

In particular, Section 2 presents the theoretical background, describing the two databases and presenting related work that mainly focuses on methods that have been recently applied to these databases. Section 3 presents the proposed methodology. At first, it describes the derivation of the above-mentioned $\boldsymbol{v}$ and $\boldsymbol{c}$ representations from a DNN trained with the first data set [11]. It then uses these representations to train a new DNN with the second (PPMI) data set, thus deriving the unified prediction model. It finally uses the obtained knowledge, through domain adaptation, to improve Parkinson's prediction in environments that do not possess DaTscan equipment and therefore lack the corresponding information. Section 4 presents the experimental study, evaluating all aspects of the proposed methodology in both data sets. Section 5 presents discussion and analysis of the obtained results. Section 6 provides conclusions and the foreseen directions of our future work.

2 Theoretical background

2.1 The Parkinson's image databases

The PPMI data set has been created through collaboration of researchers, funders and study participants so as to improve Parkinson's disease (PD) therapeutics via the identification of progression biomarkers. The PPMI study includes a cohort of 423 patients with PD, who have been diagnosed for two years or less and do not take PD medications; 196 control subjects, with no PD (NPD). Other categories, such as subjects who have been consented as PD, but whose DaTscans do not reveal dopaminergic deficit (SWEDD), prodromal ones, or subjects with genetic mutations are also followed in the study. There is at least one DaTscan, in the form of gray scale image, as well as MRIs for each subject.

In order to be able to extract and deal with volumetric information, the MRIs are considered in consecutive triplets. As a consequence, the medical image inputs to the DNNs consist of a DaTscan and/or three consecutive MRIs. Such an input sample from the PPMI data set, including a DaTscan (bottom right hand side) and an MRI triplet (top and bottom left hand side) is shown in Fig. 1.

Another Parkinson's database has been recently developed [11], based on anonymised data from 75 subjects, 50 subjects with PD and 25 controls, of the Georgios Gennimatas Hospital in Athens, Greece. It includes at least one DaTscan, in the form of colour image, and many MRIs per subject. In total, it includes 925 DaTscans, 595 of which come from subjects with PD and 330 from controls; and 41528 MRIs, 31147 of which represent PD and 10381 NPD.

For comparison purposes, Fig. 2 shows a respective input from this data set, including a DaTscan and a triplet of MRIs. It can be

seen that the DaTscans in this database are colour images, in contrast to the gray scale scans of the PPMI database.

2.2 Related work

A variety of techniques have been applied to the PPMI data set. During the last three years, ML techniques, such as support vector machines (SVMs), logistic regression, random forests (RFs), and decision trees have been used for PD diagnosis. Such methods have been applied based on patient questionnaires [16], reporting an accuracy over 95%. They were also used to analyse extracted features (related to uptake ratios on the striatum, volume and length of the striatal area) from 652 DaTscans [17], reporting an accuracy of 97.9%.

ML techniques, such as SVMs and RFs, were also applied to features extracted from MRI data [18], reporting an accuracy ranging from 88 to 93%, in which clinical features were also considered, apart from network features.

Techniques based on self-organising maps combined with SVMs have been used to understand the pathology and provide PD diagnosis [19], reporting an accuracy of about 95.4%. Techniques using Fisher's linear discriminant analysis and locality preserving projection for feature selection, as well as a multitask framework, have been applied to discriminate among PD, control and SWEDD subjects [20, 21], reporting accuracy of about 84%.

Use of Tensorflow as an interface for PD diagnosis based on medical imaging has been proposed [22], using a neural network model and providing an accuracy of 97.34%.

DNNs, including CNNs, CNN-RNNs have been developed in [10, 23] for PD prediction using the DaTscan and MRI data included in the above-mentioned database [11].

In contrast to most of the techniques which were applied to the PPMI data set, DNNs do not require a feature selection step, since features are automatically detected and extracted during DNN training. The DaTscans and/or the MRI triplets were directly presented at the input of the DNN. Moreover, to tackle imbalanced data between the two categories, a data augmentation strategy has been used [1, 10], rising the number of combined, i.e. DaTscan and MRI inputs to a balanced number of 150.000 inputs.

DNN training was implemented by using the pre-trained ResNet-50 structure [24], transfer learning and adaptation [14, 25] of its convolutional layers' weights, followed by training the FC layers and the recurrent part of the architecture; the latter was composed of gated recurrent units (GRUs) [26].

Experiments have been presented [10, 11] comparing the obtained accuracy, when feeding the DNNs with only DaTscan inputs, or with only MRI inputs, or with both DaTscans and MRI inputs. By training CNN and CNN-RNN architectures with the resulting data set, a highest accuracy of 98% was achieved when using both types of data as inputs. An accuracy of 94% was achieved when using only DaTscan inputs, while a much lower accuracy of 70% was obtained when using only MRI inputs.

In the following, we extend this DNN architecture, as well as some early results on extraction of latent information from it which we recently presented in [27], to derive a unified prediction model, which can be effectively and efficiently applied for PD diagnosis across both the database [11] and the PPMI data set, overcoming the DNN shortcomings described in the previous Section.

3 The proposed methodology

3.1 The extracted features from deep neural networks

Our approach starts by training a deep neural architecture, such as a convolutional, or convolutional-recurrent network to predict the status (PD, or NPD) of subjects. This is based on analysis of medical images, i.e. DaTscans and/or MRI images, collected in a specific medical centre, or hospital (in particular, that in [11]).

As in [10] we consider a CNN part that has a well-known structure, such as ResNet-50, generally composed of convolutional and pooling layers, followed by one, or two fully-connected layers. ReLU neuron models are used in this part. In the case of convolutional and recurrent network, which we adopt in the following, two hidden layers with long short-term memory neuron models, or GRUs are used on top of the CNN part, providing the final classification, or prediction, outputs.

In our approach we select to extract and further analyse the, say M, outputs of the last FC layer, or last hidden layer of the trained CNN, or CNN-RNN, respectively. This is due to the fact that these outputs constitute high level, semantic extracts, based on which the trained DNN provides its final predictions. Other choices can also be used, involving features extracted, not only from high level, but also from mid and lower level layers. From our experiments, such choices have not proven capable of significantly improving the achieved performance.

In the following we present the extraction of concise semantic information from these representations, using unsupervised analysis.

Let us assume that the data set S, including DaTscans and MRI inputs has been collected and used for training the DNN to predict the PD or NPD status of subjects. Let also T denote the respective test set used to evaluate the performance of the trained network:

$$S = \{(\boldsymbol{x}_s(k), y_s(k));\ k = 1, \ldots, N_s\} \tag{1}$$

$$T = \{(\boldsymbol{x}_t(k), y_t(k));\ k = 1, \ldots, N_t\} \tag{2}$$

In (1), (2), $\boldsymbol{x}_s(k)$ and $y_s(k)$ denote the N_s training inputs and the category to which each one of them belongs. We use a 1 to denote a patient category, and a 0 to denote a control/non-patient one. Similarly, $\boldsymbol{x}_t(k)$ and $y_t(k)$ denote the N_t inputs and the corresponding category over the test set.

Let us assume that we train the DNN using the data in S and, for each input k, we collect the M values of the outputs of neurons in the selected DNN FC or hidden layer, generating a vector $\boldsymbol{v}_s(k)$. A similar vector $\boldsymbol{v}_t(k)$ is generated when applying the trained DNN to each input k of the test set

$$V_s = \{(\boldsymbol{v}_s(k),\ k = 1, \ldots, N_s\} \tag{3}$$

and

$$V_t = \{(\boldsymbol{v}_t(k),\ k = 1, \ldots, N_t\} \tag{4}$$

In the following we derive a concise representation of these $\boldsymbol{v}$ vectors, by using an unsupervised, clustering procedure. In particular, we use the k-means++ algorithm [28] to generate, say, L clusters $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_L\}$ through minimisation of the following function:

$$\widehat{Q}_{k\text{-means}} = \arg\min_{Q} \sum_{i=1}^{L} \sum_{\mathbf{v}_s \in V_s} ||\boldsymbol{v}_s - \boldsymbol{\mu}_i||^2 \tag{5}$$

in which $\boldsymbol{\mu}_i$ denotes the mean of $\boldsymbol{v}$ values belonging to cluster i.

For each cluster i, we then compute the corresponding cluster centre $\boldsymbol{c}(i)$, thus defining the set of cluster centres C, which forms a concise representation and prediction model for Parkinson's diagnosis.

$$C = \{(\boldsymbol{c}(i),\ i = 1, \ldots, L\} \tag{6}$$

This procedure, of using data set S to generate the set of cluster centres C is illustrated in Fig. 3.

Since the derived representation consists of a small number of cluster centres, medical experts can examine and annotate the respective DaTscans and MRI images with relevant textual information. This information can include the subject's status (i.e. PD, or non-PD), the stage of Parkinson's for patients, as well as other metrics.

Let us now focus on using the set C for diagnosis of Parkinson's in new subject cases, e.g. those included in the test data set T. For each input in T, we compute the $\boldsymbol{v}_s$ value. We then calculate the euclidean distance of this value from each cluster centre in C and classify it to the category of the closest cluster centre. As a result,

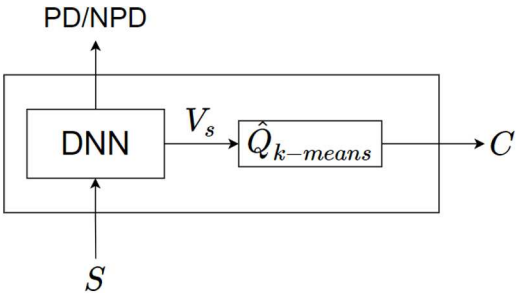

Fig. 3 *Input set S is used to train the DNN; clustering of the extracted V_s vector generates representation set C*

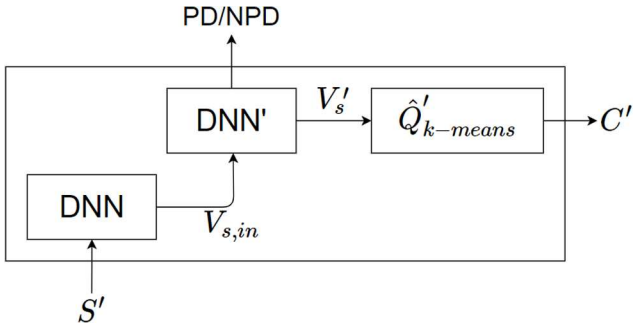

Fig. 4 *Set S′ is fed to DNN, with the extracted V_s vector being used as input for training DNN′; clustering of the extracted V'_s vector generates representation set C′*

we classify each test input to a respective category, thus predicting the subject's status.

It should be mentioned that, using this approach, we can predict a new subject's status in a rather efficient and transparent way. At first, only L distances between M-dimensional vectors have to be computed and the minimum of them be selected. Then, the subject can be informed of why the specific diagnosis was made, through visualisation of the medical images and presentation of the medical annotations corresponding to the selected cluster centre.

3.2 The unified prediction model

Following the above described approach: (i) we design a DNN (as shown in Fig. 3) and extensively train it for predicting PD, based on image data provided by a specific hospital, medical centre, or available database, (ii) we generate a concise representation (set C) composed of the derived cluster centre representation that can be used to predict Parkinson's in an efficient way. This information, i.e. the DNN weights and the set C, represent, in the proposed unified approach, the knowledge obtained through the analysis of the respective database S.

Let us now consider another medical environment, where another database related to Parkinson's has been generated. Let us assume that it can be, similarly, described through the following training and test sets

$$S' = \{(\boldsymbol{x}'_s(k), y'_s(k));\ k = 1, \ldots, N'_s\} \tag{7}$$

$$T' = \{(\boldsymbol{x}'_t(k), y'_t(k));\ k = 1, \ldots, N'_t\} \tag{8}$$

In (7), (8), $\boldsymbol{x}'_s(k)$ and $y'_s(k)$ denote the N'_s training inputs and the corresponding category, whilst $\boldsymbol{x}'_t(k)$ and $y'_t(k)$ denote the N'_t inputs and the corresponding category over the test set.

In the deep-learning field it is known that when applying a network, trained on a specific data set, to another data set with different characteristics, the performance is expected to be poor. Transfer learning, along with network retraining is the usual technique for obtaining a good performance over the new data set. However, the 'catastrophic forgetting' problem that was mentioned in the Introduction appears, obstructing the derivation of a unified prediction model over all data sets.

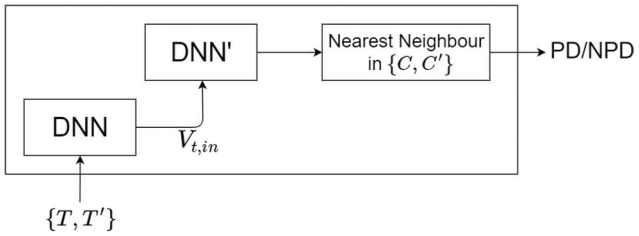

Fig. 5 *Any subject's data from either set T, or T′, is fed to the DNN-DNN′ architecture, with the extracted V'_s vector being classified to the category of the nearest cluster centre in C and C′; thus predicting subject's status*

In the following we show how the proposed approach can alleviate this problem.

Fig. 4 shows the procedure we follow to achieve such a model. According to it, we present all inputs of the new training data set S' to the available DNN that we have already trained with the original data set S; we compute the V_s representations, similarly to (3), named as $V_{s,in}$ in Fig. 4. These representations, which were generated using the knowledge obtained from the original data set, form the input to a new DNN, named DNN′ in Fig. 4; this network is trained to use these inputs so as to predict the PD/NPD status of the subjects whose data are in set S'.

In a similar way, as in (3)–(5), we compute the new set of representations, named V'_s and through clustering the new set of cluster centres C':

$$C' = \{(\boldsymbol{c}'(i),\ i = 1, \ldots, L'\} \tag{9}$$

The next step is to merge the sets C and C', creating the unified prediction model. Using the two network structures (DNN and DNN′ in Fig. 4), in a testing formulation, and the nearest neighbour criterion with respect to the union of C and C', we can predict the PD/NPD status of all subjects in both test sets T and T', as shown in Fig. 5.

The resulting representation, consisting of the C and C' sets, is, therefore, able to predict a new subject's status, using the knowledge acquired by the DNN and DNN′ networks trained on both data sets, in an efficient and transparent way.

3.3 Domain adaptation in Parkinson's prediction

In the former sections it was assumed that the inputs to the DNN consisted of both DaTscans and MRI data, so that the networks learn to detect and use correlations between both types of inputs. From [10, 11] we know that DaTscan inputs provide DNNs with more discriminating ability than MRI inputs. However, DaTscan facilities are generally available in big medical centres and hospitals. As a consequence, in many medical environments, prediction should be achieved using only MRI information.

In the following, we present a novel domain adaptation extension of the proposed approach for improving the prediction provided by a DNN when using only MRI inputs, based on the concise C representations derived from a DNN trained with both types of inputs.

To achieve this, we introduce a novel error criterion for training the new DNN with MRI inputs, which is expressed in terms of the internal $\boldsymbol{v}_s$ representations generated by this network, as well as by the representation set C obtained during training of the original network.

In particular, let us consider that the training and test data sets in (1) and (2) consist of only MRI data.

By training a DNN with data set S, we can obtain, similarly to (3), a vector V''_s, defined as follows:

$$V''_s = \{(\boldsymbol{v}''_s(k),\ k = 1, \ldots, N''_s\} \tag{10}$$

where each $\boldsymbol{v}''(s)$ vector is of M dimensionality, equal to the size of the last layer in the CNN or CNN-RNN architecture, for the N''_s training data.

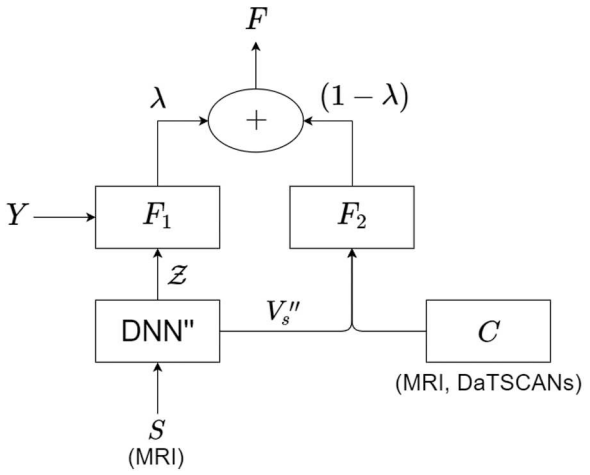

Fig. 6 *MRI data in S are fed to DNN′′ which is trained through minimisation of error function F; this is computed based*
(a) On the difference between output Z and desired output Y (F_1 component), ***(b)*** On comparison of the extracted V''_s vector to the representation set C computed with both MRI and DaTscan data (F_2 component)

Table 1 The accuracy obtained by CNN and CNN-RNN architectures

Structure	No FC layers	No hidden Layers	No units in FC Layer(s)	No units in hidden layers	Accuracy, %
CNN	2	—	2622–1500	—	94
CNN-RNN	1	2	1500	128–128	98

Our target is to train the new network to produce $\boldsymbol{v}''_s$ values that are close to one of the cluster centres in C extracted from the original network, which had been trained with both types of inputs. If this is possible, then the achieved prediction will be closer to the one provided by the original network. As a consequence, a higher prediction accuracy will be obtained by the new network.

In mathematical terms, we compare the $\boldsymbol{v}''_s$ values with the L cluster centres $\boldsymbol{c}_s$ defined in (6). By computing the minimum euclidean distance, we select a particular cluster centre, to form the desired target value for each one of the $\boldsymbol{v}''_s$. As a result, the following U vector of desired values $u(m,n)$is generated:

$$U_s = \{(u(m,n), m = 1, \ldots, L; n = 1, \ldots, N''_s\} \tag{11}$$

in which $u(m,n)$ equals 1, if the respective cluster centre is the selected one among the L dimensional set C, or equals 0, if the cluster centre is not selected.

The U_s values are used in the following to define the new error function. Minimisation of this error function would provide the new network with the ability to make decisions that are similar to those of the original network; thus, providing improved predictions of the subjects' status.

The proposed error function is composed of two distinct terms. The first term is the normal mean squared error criterion computed at the network output level and defined as follows:

$$F_1 = \frac{1}{N''_s}\sum_{k=1}^{N''_s}(y(k) - z(k))^2 \tag{12}$$

in which $z(k)$ represents the category of the input and $y(k)$ represents the respective category prediction provided at the network output.

The following variables are introduced to define the second term in the error function:

$$\mathbf{e}(m,n) = \boldsymbol{v}''_s(n) - \boldsymbol{c}(m),\ m = 1, \ldots, L;\ n = 1, \ldots, N''_s \tag{13}$$

$$E(m,n) = \mathbf{e}(m,n) * (\mathbf{e}(m,n))^{\mathrm{T}} \tag{14}$$

where T denotes transposition.

To achieve the targeted goal, we perform minimisation of all $E(m,n)$ values, when $u(m,n)$ equals unity, with simultaneous maximisation of the $E(m,n)$ values, when $u(m,n)$ equals zero. Thus, we feed $E(m,n)$ to a non-linear activation function, of the softmax type and reverse the result, by subtracting it from unity.

The second error term is computed as the mean squared error between the resulting values and the respective U_s ones:

$$F_2 = \frac{1}{LN''_s}\sum_{m=1}^{L}\sum_{n=1}^{N''_s}(u(m,n) - [1 - f(E(m,n)])^2 \tag{15}$$

where f is the used softmax function.

Using (12) and (15), the resulting total Error Criterion is computed as follows:

$$F_{new} = \lambda\mathcal{F}_1 + (1-\lambda)\mathcal{F}_2 \tag{16}$$

in which λ is a positive number less than unity. When the value of λ is close to zero, the significance of the proposed approach is more evident in the obtained results. In general, a value of 0.5 is used in our approach.

Fig. 6 presents the proposed training procedure, in which: the set of original cluster centres C is compared to the extracted V''_s representations, as defined in the error criterion F_2; the DNN′′ outputs Z, composed of $z(k)$ values, are compared to the desired network predictions Y, composed of respective $y(k)$ values, as defined in the error criterion F_1; F_1 and F_2 are used to compute the proposed error function minimised during DNN′′ network training.

4 Experimental study

As already described, our experimental study is performed on two databases; the first is the database generated in Greece [11] and the second is the PPMI database [15]. Both of them include DaTscans and MRI information for all their subjects. For training and evaluation purposes the respective data sets have been separated to training, validation and test data. The specific settings can be provided, upon request, from mlearn.lincoln.ac.uk. All experiments have been based on ten-fold cross validation.

Creation of the main deep neural architecture and generation of the respective cluster centre representation set for predicting Parkinson's is made using the database in [11]. Based on this database we also evaluate the domain adaptation approach for predicting Parkinson's using only MRI information. The unified approach for predicting Parkinson's is based on the data of both databases. All DNN training implementations were made with Python and Tensorflow.

4.1 Extracting DNN concise representations

In [10], DNNs were trained with an augmented data set based on database [11], achieving very good performance on this database. The convolutional part of the network was applied to each image component, i.e. to the RGB DaTscan image and to the three (gray-scale) MRIs, using the same pretrained ResNet-50 structure. The outputs of these two ResNet structures were concatenated and fed to the FC layer of the CNN part of the network.

This structure has been able to analyse the spatial characteristics of the DaTscans and MRIs, achieving a high accuracy in the database test set, of 94%, as shown in Table 1.

The complete CNN-RNN architecture included two hidden layers on top of the CNN part, each containing 128 GRU neurons, as shown in the Table. This has been able to also analyse the temporal evolution of the MRI data, achieving an improved performance of 98% over the test data.

We trained this CNN-RNN network so as to classify the DaTscans and MRIs to the correct PD/NPD category, using a batch

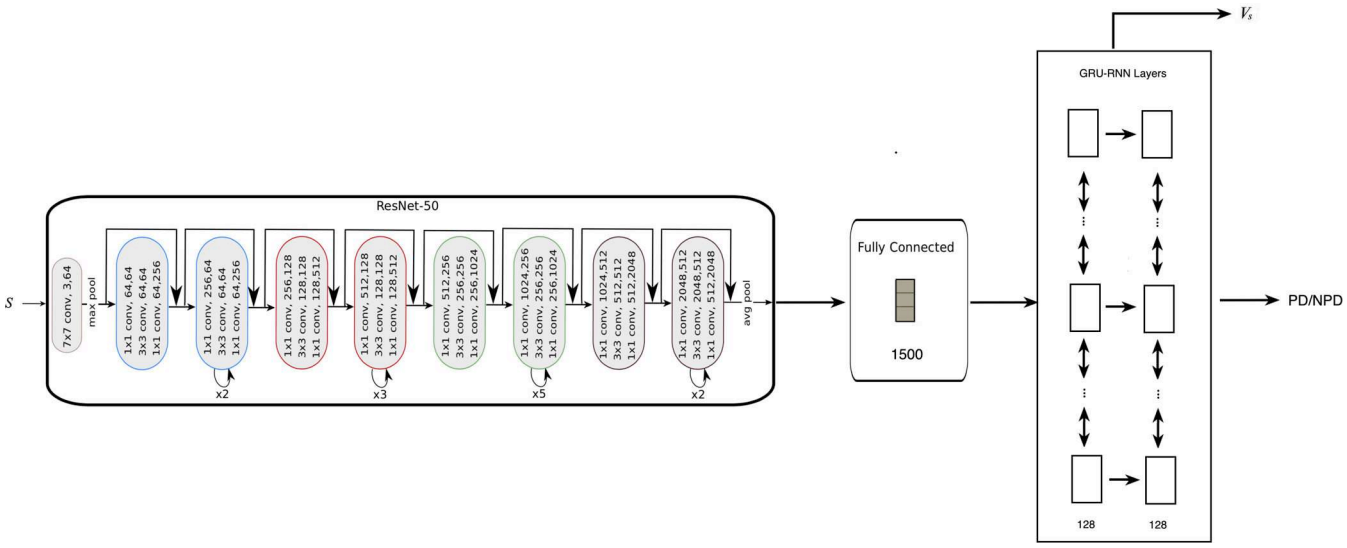

Fig. 7 *A CNN-RNN network, based on the ResNet-50 CNN structure, with one FC layer (1500 units) and two GRU RNN layers (with 128 units each); the V_s are extracted from the second GRU layer*

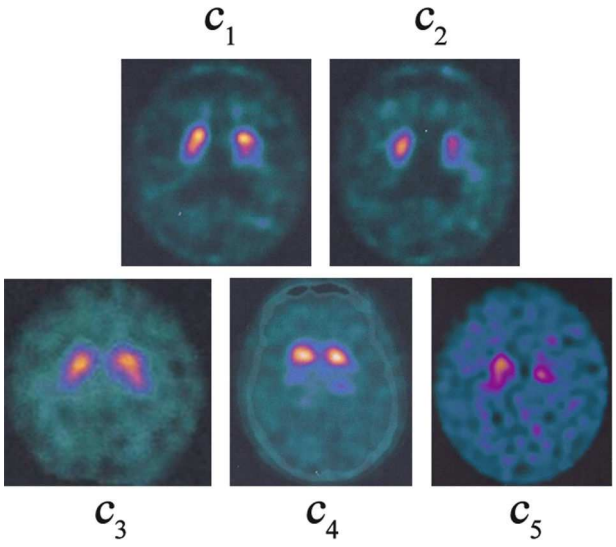

Fig. 8 *The DaTscans of the 5 selected cluster centres: $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ correspond to NPD cases, whilst $\boldsymbol{c}_3$ - $\boldsymbol{c}_5$ to progressing stages of Parkinson's*

size of 10, a fixed learning rate of 0.001 and a dropout probability of 50%.

Fig. 7 shows the extraction of the V_s vector representations from the second GRU layer of this CNN-RNN network.

These vectors include 128 elements, as is also shown in Table 1.

The clustering process, using k-means was then applied to the V_s vectors, as shown in Fig. 3. We extracted five clusters, two of which correspond to control subjects, i.e. NPD ones, with three clusters corresponding to patients, as in the original study [10]. Since the k-means algorithm depends on the initial conditions, the cluster centres are not identical, but very similar to the ones in [10]. These constitute the extracted concise representation C set; consequently, C is composed of five 128-dimensional vectors.

The DaTscans corresponding to the extracted cluster centres are shown in Fig. 8. Through the assistance of medical experts we were able to verify that the three DaTscans corresponding to patient cases represent different stages of PD. In particular: the first of them ($\boldsymbol{c}_3$) represents an early occurence, between stage 1 and stage 2; the second ($\boldsymbol{c}_4$) shows a pathological case, at stage 2; the third ($\boldsymbol{c}_5$) represents a case that has reached stage 3 of Parkinson's. In the case of controls, there are differences between the first ($\boldsymbol{c}_1$), which is a clear NPD case and the second ($\boldsymbol{c}_2$), which is a more obscure case.

Following the above annotations, it can be said that the derived representations convey more information about the subjects' status

Table 2 Percentage of inputs in the five different clusters

Cluster	No of data, %
$\boldsymbol{c}_1$	4,3
$\boldsymbol{c}_2$	38,4
$\boldsymbol{c}_3$	27,6
$\boldsymbol{c}_4$	2,3
$\boldsymbol{c}_5$	27,4

Table 3 Test data in each generated cluster and PD/NPD accuracy

Test case	$\boldsymbol{c}_1$	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	$\boldsymbol{c}_4$	$\boldsymbol{c}_5$	PD/NPD, %
non patient 1	44	**398**	0	0	0	100
non patient 2	10	**90**	0	0	0	100
patient 1	3	7	**94**	8	8	91.6
patient 2	1	7	**139**	17	20	95.6
patient 3	3	0	**145**	18	38	98.5
patient 4	0	0	0	8	**72**	100

than trained DNN outputs. This information can be used by medical experts to evaluate the predictions made by the original DNN when new subjects' data have to be analysed. The computed V_s representations in the new cases can be efficiently classified to the category of the nearest cluster centre of C; the cluster centre's DaTscan, MRIs and annotations will then be used to justify, in a transparent way, the provided prediction. In Table 2 we present the percentage of training inputs included in every cluster category. Since a large number of cases belong to an early stage of PD, it is of high significance to develop tools, such as the proposed one, which have the ability to provide highly accurate predictions over different data sets and different medical environments.

Let us consider six new subjects, with their data (many combinations of DaTscans and MRIs) having to be analysed by the clustered representation extracted from the trained DNN. There are two NPD and four PD subjects.

We applied the procedure shown in Fig. 5 to classify these test data. Table 3 presents the classification of these data to the five generated clusters and consequently to the PD or NPD category. It can be seen that the proposed approach was able to discriminate all cases, including the early stage Parkinson's cases (cluster $\boldsymbol{c}_3$), with a very high accuracy. This illustrates its ability to provide accurate predictions of PD based on DaTscan and MRI data.

Moreover, let us assume that a new case appears, for which (i) the DNN outputs are of low confidence, for example providing output values around 0.5, when a value near to 0 or 1 is required

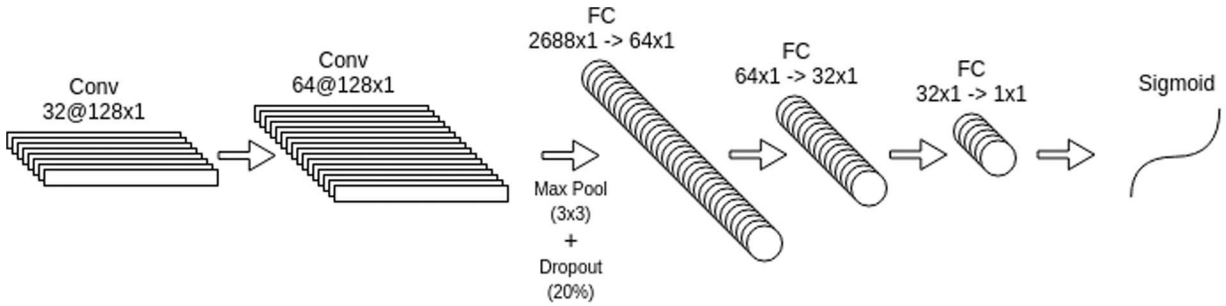

Fig. 9 *A CNN structure, with three FC layers; the V_s are extracted from the last FC layer*

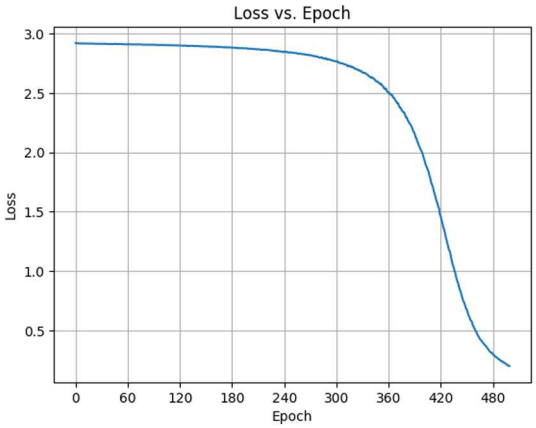

Fig. 10 *Minimisation of the CNN Loss Function in terms of the number of training epochs*

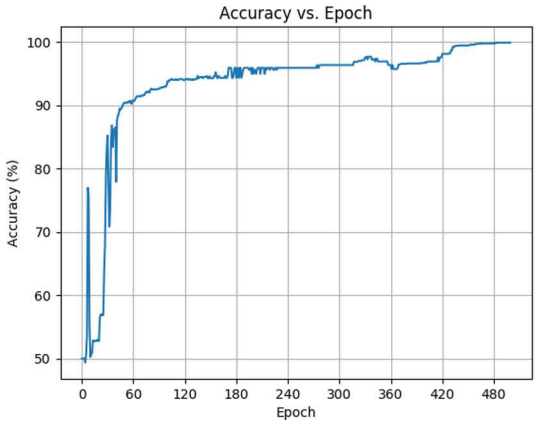

Fig. 11 *Accuracy (%) of the CNN when applied to the test data set in terms of the number of training epochs*

Table 4 PD/NPD Accuracy (%) on PPMI Data set

S'_{PD}	S'_{NPD}	T'_{PD}	T'_{NPD}	Total
99.80	99.69	99.61	99.9	99.76%

for good prediction; (ii) the V_s values are quite faraway from all existing cluster centres in C. This means that this is a case that the DNN cannot generalise its learning. As a consequence, a medical expert should annotate these data.

Following the annotation of the new data by the expert, we would need to insert the new data in our prediction system. It should be mentioned that retraining of the DNN would be required, so as to retain the old knowledge and include the new one; this would be computationally intensive and possibly unfeasible. On the contrary, the proposed approach would only require extension of the C set with one, or more, cluster centres, corresponding to the new information; as a consequence, this would be done in a very efficient way.

4.2 The unified prediction model

In the following we examine the ability of the proposed approach to generate a unified prediction model for Parkinson's. In particular, we examine the ability of the procedure shown in Fig. 4, using the trained DNN (CNN-RNN) architecture, to be successfully applied to the PPMI database [15], for PD/NPD prediction.

Since the DaTscans were the basic source of the DNN's discriminating ability, we focus our new developments on the DaTscans included in the PPMI database. For this reason, we have retained 609 subjects from the PPMI database, excluding some patients for which we were not able to extract DaTscans of good quality. In total we selected 1481 DaTscans, which we combined with MRI triplets from the respective subjects, generating a data set of 7700 inputs; each input was composed, of one (gray-scale) DaTscan and a triplet of MRI images.

We split the data into training, validation and test sets, each representing about 65, 15 and 20% of the data, respectively. During separation, care was taken to ensure the split was subject independent. No subject's data were included in more than one set, ensuring that the model learns to solve the problem and not the specific data. Since the two categories were unbalanced, we performed data augmentation of the NPD category, through addition of small amount of noise, so as to generate a balanced set of 10,240 inputs.

At first, for comparison purposes, we trained CNN and CNN-RNN networks, similar to the ones presented in the previous section, from scratch, on the selected PPMI training set (6656 inputs). We used the validation set (1584 inputs) to test the obtained accuracy in the end of each training epoch. We then tested the performance of the networks on the test set (2000 inputs). The obtained accuracy was in the range of 96–97%, similar to the accuracy achieved by other techniques, as reported in Section 2.2 of the study. We also used transfer learning of the networks generated in the first section of our experimental study, to initialise the re-training of the new networks. Similar results were obtained in this case as well.

We then applied the procedure shown in Fig. 4, to train DNN' with the V_s vectors extracted from the last hidden layer of the DNN that had been trained on the [11] data set.

We used a CNN model, in place of DNN' in Fig. 4. The CNN was fed with the 128-dimensional V_s vectors, and its structure included two Convolution layers, a Max Pooling layer, a Dropout layer with 20% probability and three Fully-Connected layers, containing 2688-64-32 neurons, respectively, as shown in Fig. 9.

The performance of the network was very high, classifying in the correct PD/NPD category 99.76% of the inputs. The minimisation of the Loss function over 500 epochs and the respective accuracy over the test data are shown in Figs. 10 and 11, respectively, while the obtained per class accuracy in the training S' and test T' sets, for the PD and NPD categories, is shown in Table 4.

By then implementing the clustering procedure shown in Fig. 4, we were able to extract five new clusters, three of which represent NPD subjects' cases and two of which represent PD cases. Table 5 presents the split of PPMI data to these five clusters. These cluster centres are 32-dimensional vectors, since they were extracted from the last FC layer of DNN', which includes 32 neurons.

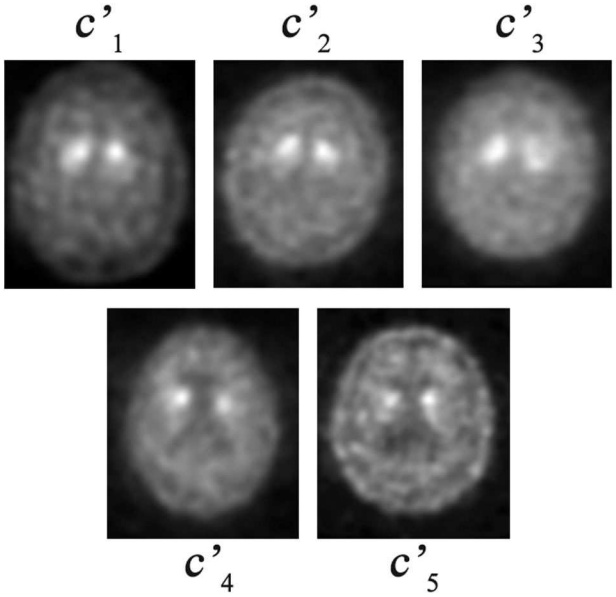

Fig. 12 *The DaTscans of the five cluster centres in C′; the three on top represent NPD cases, whilst the two at bottom represent PD cases*

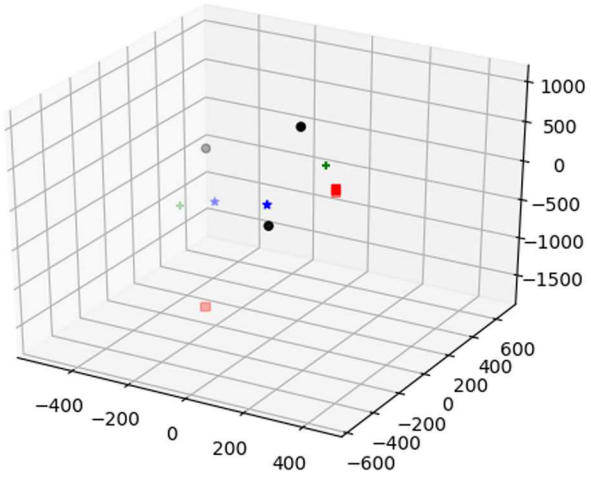

Fig. 13 *The obtained ten cluster centres in 3-D: 5 of them (squares with red/rose colour, & plus (+) symbols with green colour) depict patients; five of them (stars with blue colour & circles with black/grey colour) depict non-patients*

Fig. 12 shows the DaTscans corresponding to the cluster centres $\boldsymbol{c}_1'$–$\boldsymbol{c}_5'$. Since the patients in the PPMI Database generally belong to early stages of Parkinson's (stage 1 to stage 2), it can be seen that two cluster centres, i.e. $\boldsymbol{c}_4'$ and $\boldsymbol{c}_5'$ were enough to represent these cases. Variations in the appearance of the non-Parkinson's cases can be seen in $\boldsymbol{c}_1'$–$\boldsymbol{c}_3'$ DaTscans.

We then applied the merging of sets C and C'. It should be mentioned that the five centres in set C were 128-dimensional, whilst the five centres on set C' were 32-dimensional. To produce a unified representation, we made an ablation study, through PCA analysis, on the classification performance achieved in data set [11], if we represented the five cluster centres in C through only 32 principal components. We were able to achieve a classification performance of 97.92%, which is very close to the 98% performance in Table 1.

Consequently, we were able to generate a unified model consisting of 10 32-dimensional cluster centres. Fig. 13 shows a three-dimensional (3D) projection of the ten cluster centres. The three (red/rose) squares denote the patient cases in the data set [11] and the two (green) plus (+) symbols represent the patient cases in the PPMI data set. The two (blue) stars represent the normal cases in data set [11] and the three (black/grey) circles represent the normal cases in the PPMI data set. It can be seen that the PD cenres are distinguishable from the NPD ones.

Table 5 Percentage of inputs in the new five clusters

Cluster	No of data, %
$\boldsymbol{c}_1'$	14
$\boldsymbol{c}_2'$	13
$\boldsymbol{c}_3'$	23
$\boldsymbol{c}_4'$	27
$\boldsymbol{c}_5'$	23

Table 6 Test data in each cluster and PD/NPD accuracy (no adaptation)

Test case	$\boldsymbol{c}_1$	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	$\boldsymbol{c}_4$	$\boldsymbol{c}_5$	PD/NPD, %
non patient 1	**181**	74	179	8	0	57.7
non patient 2	14	4	**44**	33	5	25.5
patient 1	16	0	**53**	49	2	86.7
patient 2	6	0	**83**	80	15	96.7
patient 3	26	3	**130**	35	10	85.8
patient 4	12	0	**51**	11	6	85

Table 7 Test data in each cluster and PD/NPD accuracy (with domain adaptation)

Test case	$\boldsymbol{c}_1$	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	$\boldsymbol{c}_4$	$\boldsymbol{c}_5$	PD/NPD, %
non patient 1	**176**	147	114	5	0	73
non patient 2	13	**41**	25	18	3	54
patient 1	13	0	**70**	35	2	89.2
patient 2	5	0	**116**	54	9	97.3
patient 3	20	2	**140**	34	8	89.2
patient 4	9	0	31	5	**35**	88.8

This has been verified by testing the ability of the unified prediction model to correctly classify all input data in test sets T and T', i.e. the data from both data sets. There was no effect on the performance of the prediction achieved by each prediction model, i.e. C and C' when applied, separately, to their respective data sets, as shown in Tables 1 and 4.

This illustrates that the unified representation set, composed of the union of C and C', has been able to provide exactly the same prediction results, as the original representation sets.

4.3 Domain adaptation

If we train a DNN with only MRI data over the data set [11], then the obtained accuracy is just over 70%. In particular, if we apply the trained DNN to classify the six new subject cases mentioned in Section 4.1 and split the data in the five cluster centres C, the obtained results are shown in Table 6. It can be seen that the prediction, especially in the NPD cases, is low, with one subject being wrongly classified as PD.

In the following we examine the application of the domain adaptation approach, so as to improve the DNN prediction accuracy when using only MRI inputs.

To do this, we implemented the procedure shown in Fig. 6, training the CNN-RNN with only the MRI training data of the database [11]. We used the five cluster centres in C to compute the F_2 error criterion, combining it with the normal mean-squared error criterion F_1, thus calculating and minimising the total F error criterion. A value of $\lambda = 0.5$ was selected, compensating the contribution of both error components.

After training, we tested the performance of the adapted DNN over the same test set, obtaining a prediction accuracy of 81.1%. Table 7 illustrates the improvement that was obtained, by using the cluster centres in C as desired values for the extracted $\mathscr{V}_s$ values, when compared to the respective results of Table 6.

It can be seen that all subjects have been correctly classified to the correct PD/NPD category.

5 Discussion

The results obtained in the experimental study illustrate the ability of the proposed approach to generate information rich, concise representations of latent features extracted from trained DNNs and use them for developing a unified prediction model for Parkinson's.

Such a concise representation was first developed and validated over the Greek Parkinson's database. Based on clustering of the representations extracted from the database training data, and simultaneous testing on the validation data, best precision accuracy was obtained when generating five clusters. Two of them represent NPD cases, whilst three of them represent PD cases. The corresponding class centres' DaTscans are shown in Fig. 8.

As was verified by the medical team of the Department of Neurology of the Hellenic Georgios Gennimatas Hospital, these DaTscans correspond, on the one hand to different NPD cases and on the other hand, to three different stages of Parkinson's.

Table 2 shows the percentage of inputs belonging in each of these clusters. It can be seen that a large part of the subjects (about two thirds) belong either to an, obscure, NPD case (second cluster), or to an early stage of Parkinson's (third cluster). This indicates the significance of the very high prediction accuracy provided by the proposed system. Table 3 further illustrates the very good prediction achieved when the developed system is applied to data from six new subjects, especially when focusing on those who are in an early stage of Parkinson's (cluster $\boldsymbol{c}_3$).

Through the proposed approach, we were able to produce a respective concise representation of the PPMI database. The clustering procedure resulted in five clusters as well; however, three represented the NPD cases and two the PD ones. Table 5 shows the percentage of inputs in each cluster, while Table 4 shows that an excellent prediction accuracy was achieved in this data set. The cluster centres' respective DaTscans are shown in Fig. 12.

It was further verified that a unified highly accurate prediction model has been generated, through merging of the above ten cluster centres; the nearest neighbour criterion was successfully used to classify all data inputs from both Greek and PPMI databases in the respective categories.

Finally Tables 6 and 7 illustrate the improved prediction accuracy obtained when using the proposed domain adaptation procedure, in comparison to the prediction obtained when simply using MRI data as inputs.

Let us further discuss the significance of the derived cluster centres, for generating trustworthy DNN decision making in health care. Whenever a PD/NPD prediction is provided to the medical expert for a specific subject, it will also show the subject's DaTscan, together with the DaTscan of the centre of the selected cluster. The latter will indicate what type of data were used by the system to generate its prediction. In this way, the medical expert, and the subject, could decide by themselves whether to trust, or not, the suggested decision.

It should be finally stressed that the existing literature on PD/NPD DaTscan image classification (e.g. https://www.accessdata.fda.gov), mentions that: (i) normal images are characterised by two symmetric comma- or crescent-shaped focal regions of activity mirrored about the median plane, while the striatal activity is distinct, relative to surrounding brain tissue, (ii) abnormal images fall into at least one of the following three categories:

- assymetric activity, e.g. when activity in the region of the putamen of one hemisphere is absent or greatly reduced with respect to the other; activity is still visible in the caudate nuclei of both hemispheres, resulting in a comma or crescent shape in one and a circular or oval focus in the other; there may be reduced activity between at least one striatum and surrounding tissues
- activity that is absent in the putamen of both hemispheres and confined to the caudate nuclei; activity is relatively symmetric and forms two roughly circular or oval foci; activity of one or both is generally reduced
- activity that is absent in the putamen of both hemispheres and greatly reduced in one or both caudate nuclei; activity of the striata with respect to the background is reduced.

Let us now examine the DaTscans of the ten extracted cluster centres. Although most of them are consistent with the above rules, there are some DaTscans that differ, thus providing more specific information on each subject's status prediction. Such a case occurs in the DaTscan of cluster $\boldsymbol{c}_2$ centre shown in Fig. 8. Remarkably, this cluster contains >38% of the input data of the respective database.

6 Conclusions and future work

In this study we have developed a new approach for deriving a unified prediction model for PD.

We first extracted concise representations from DNNs after training them with DaTscans and MRI data. A set of vectors corresponding to the centres of clusters of these representations, together with the respective DNN structure/weights, constitute the information used to model the knowledge extracted from the PPMI database [15] and the Greek database [11].

It has been then shown that the unified model generated over these different data sets can provide efficient and transparent prediction of PD. Predictions of very high accuracy, which extend the state-of-the-art, have been obtained in both databases.

A domain adaptation methodology, based on the proposed approach was also developed; this introduces a novel error criterion and uses the representations extracted from the DNN that was trained with DaTscans and MRIs, for effectively training respective DNNs in environments that only possess MRI information for their subjects.

Our future work will follow three directions.

The first will be to extend the derived unified prediction model for Parkinson's to cover more data cases and be used in real medical environments. We have been collaborating with medical experts and hospitals in Greece and UK for achieving this goal.

The second will be to extend our former and current research for derivation of a transparent and trustworthy prediction-making process; this will include combining the data driven deep neural architectures with knowledge-based methods and ontological representation of knowledge [29], as well as considering the use of fuzzy descriptors in them [30, 31]. We have been working on extending the early models developed in these works in the current framework of explainable deep-learning methodologies.

The third direction will be to apply the proposed approach to other neurodegenerative diseases, including Alzheimer's disease. Deep-learning methodologies have been recently applied to Alzheimer's data [11, 32, 33]. The proposed approach can be applied to these frameworks for unified prediction and for making the deep-learning procedure more efficient and transparent.

7 Acknowledgments

We thank Dr Georgios Tagaris and the Department of Neurology of the Georgios Gennimatas General Hospital, Athens, Greece, for providing the data set with Parkinson's data and for annotating the DaTscans corresponding to the extracted cluster centre representations.

The PPMI data used in the preparation of this article were obtained from the PPMI database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. 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, Biogen, BioLegend, Bristol-Myers Squibb, Celgene, Denali, GE Health care, Genentech, GSK, Eli Lilly and Company, Lundbeck, Merck, MSD, Pfizer, Piramal Imaging, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, .Takeda, Teva, USB, Verily and Voyager Therapeutics.

We thank them for providing us with the PPMI data set used in our experiments to illustrate the performance of the proposed unified prediction model for Parkinson's disease.

8 References

[1] Goodfellow, I., Bengio, Y., Courville, A.: '*Deep learning*' (MIT Press, USA., 2016)

[2] Sadja, P.: 'Machine learning for detection and diagnosis of disease', *Annual Rev. Biomed. Eng.*, 2006, **8**, pp. 537–565

[3] Azizi, S., Bayat, S., Yan, P., *et al.*: 'Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations', *Int. J. Comput. Assist. Radiol. Surg.*, 2017, **12**, pp. 1293–1305

[4] Li, R., Zhang, W., Suk, H., *et al.*: 'Deep learning based imaging data completion for improved brain disease diagnosis'. Int. Conf. on Medical Image Computing and Computer-assisted Intervention, Boston, MA, USA., 2014, pp. 305–312

[5] Goetz, C.G., Tilley, B.C., Shaftman, S.R., *et al.*: 'Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results', *Movement Disorders*, 2008, **23**, (15), pp. 2129–2170

[6] Hoehn, M.M., Yahr, M.D.: 'Parkinsonism: onset, progression, and mortality', *Neurology*, 1998, **50**, p. 318

[7] Das, R.: 'A comparison of multiple classification methods for diagnosis of parkinson disease', *Expert Syst. Appl.*, 2010, **37**, (2), pp. 1568–1572

[8] Salvatore, C., Cerasa, A., Castiglioni, I., *et al.*: 'Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy', *J. Neurosci. Methods*, 2014, **222**, pp. 230–237

[9] Rojas, A., Górriz, J.M., Ramírez, J., *et al.*: 'Application of empirical mode decomposition on datscan spect images to explore parkinson disease', *Expert Syst. Appl.*, 2013, **40**, (7), pp. 2756–2766

[10] Kollias, D., Tagaris, A., Stafylopatis, S., *et al.*: 'Deep neural architectures for prediction in healthcare', *Complex Intell. Syst.*, 2018, **4**, (2), pp. 119–131

[11] Tagaris, A., Kollias, D., Stafylopatis, A., *et al.*: 'Machine learning for neurodegenerative disorder diagnosis - survey of practices and launch of benchmark dataset', *Int. J. Artif. Intell. Tools*, 2018, **27**, (3), pp. 1850011:1–1850011:28

[12] Tan, C., Sun, F., Kong, T., *et al.*: 'A survey on deep transfer learning'. 27th Int. Conf. on Artificial Neural Networks, Rhodes, Greece, 2018, pp. 270–279

[13] Kollias, D., Zafeiriou, S.P.: 'Training deep neural networks with different datasets in-the-wild: the emotion recognition paradigm'. Int. Joint Conf. on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1–8

[14] Kollias, D., Yu, M., Tagaris, A., *et al.*: 'Adaptation and contextualization of deep neural network models'. 2017 IEEE Symp. Series on Computational Intelligence (SSCI), Honolulu, HI, USA., 2017, pp. 1–8

[15] Marek, K., Jennings, D., Lasch, S., *et al.*: 'The Parkinson Progression Marker Initiative (PPMI)', *Prog. Neurobiol.*, 2011, **95**, (4), pp. 629–635

[16] Prashanth, R., Dutta-Roy, S.: 'Early detection of parkinson's disease through patient questionnaire and predictive modelling', *Int. J. Med. Inform.*, 2018, **119**, pp. 75–87

[17] Oliveira, F.P.M., Faria, D.B., Costa, D.C., *et al.*: 'Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of parkinson's disease based on [123i]fp-cit spect images', *Eur. J. Nucl. Med. Mol. Imaging*, 2018, **45**, (6), pp. 1052–1062

[18] Amoroso, N., La-Rocca, M., Monaco, A., *et al.*: 'Complex networks reveal early MRI markers of parkinson's disease', *Med. Image Anal.*, 2018, **48**, pp. 12–24

[19] Singh, G., Samavedham, L., Lim, E.C.: 'Determination of imaging biomarkers to decipher disease trajectories and differential diagnosis of neurodegenerative diseases (DIsease TreND)', *J. Neurosci. Methods*, 2018, **305**, pp. 105–116

[20] Lei, H., Huang, Z., Han, T., *et al.*: 'Joint regression and classification via relational regularization for Parkinson's disease diagnosis', *Technology and Healthcare*, 2018, **26**, pp. 19–30

[21] Lei, H., Zhao, Y., Wen, Y., *et al.*: 'Sparse feature learning for multi-class parkinson's disease classifcation', *Technol. Health Care*, 2018, **26**, (S1), pp. 193–203

[22] Zhang, Y., Kagen, A.: 'Machine learning interface for medical image analysis', *J. Digit. Imaging*, 2017, **30**, (5), pp. 615–621

[23] Tagaris, A., Kollias, D., Stafylopatis, A.: 'Assessment of parkinson's disease based on deep neural networks'. Proc. 17th Int. Conf. on Engineering Applications of Neural Networks, Athens, Greece, 2017, pp. 391–403

[24] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA., 2016

[25] Ng, H., Nguyen, V.D., Vonikakis, V., *et al.*: 'Deep learning for emotional recognition on small datasets using transfer learning'. Proc. 2015 Int. Conf. on Multimodal Interaction, Seattle, WA, USA., 2015, pp. 443–449

[26] Chung, J., Gulcehre, C., Cho, K., *et al.*: 'Empirical evaluation of gated recurrent neural networks on sequence modeling'. NIPS 2014 Workshop on Deep Learning, Montreal, Canada, 2014, pp. 1–9

[27] Kollia, I., Stafylopatis, A., Kollias, S.: 'Predicting parkinson's disease using latent information extracted from deep neural networks'. 2019 IEEE Int. Joint Conf. on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1–8

[28] Arthur, D., Vassilvitskii, S.: 'K-means++: the advantages of careful seeding'. Proc. 18th Annual ACM-SIAM Symp. on Discrete Algorithms, New Orleans, LA, USA., 2007, pp. 1027–1035

[29] Kollia, I., Glimm, B., Horrocks, I.: 2011 'Answering queries over owl ontologies with sparql'. OWL ED

[30] Simou, N., Athanasiadis, T., Stoilos, G., *et al.*: 'Image indexing and retrieval using expressive fuzzy description logics', *Signal Image Video Process. J.*, 2008, **2**, pp. 321–335

[31] Simou, N., Kollias, S.: 'FiRE: a fuzzy reasoning engine for imprecise knowledge', 2007

[32] Ortiz, A., Munilla, J., Górriz, J.M., *et al.*: 'Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease', *Int. J. Neural Syst.*, 2016, **26**, (7), p. 1650025

[33] Jo, T., Nho, K., Saykin, A.J.: 'Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data', *Front. Aging Neurosci.*, 2019, **11**, p. 220