Visualization of Decision Tree State for the Classification of Parkinson's Disease

[1]David P. Williams, [1]Deborah Mudali , [2]Hugo Buddelmeijer , [2,5]Parisa Noorishad, [3]Sanne Meles , [3]Remco J. Renken , [4]Klaus L. Leenders, [2]Edwin A. Valentijn, [1]Jos B.T.M. Roerdink

[1]Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands;
[2]Kapteyn Astronomical Institute, University of Groningen, The Netherlands;
[3]Neuroimaging Center, University Medical Center Groningen, The Netherlands;
[4]Department of Neurology, University Medical Center Groningen, The Netherlands;
[5]Netherlands eScience Center, Amsterdam, The Netherlands;
david@david-williams.info

ABSTRACT

Decision trees have been shown to be effective at classifying subjects with Parkinson's disease when provided with features (subject scores) derived from FDG-PET data. Such subject scores have strong discriminative power but are not intuitive to understand. We therefore augment each decision node with thumbnails of the principal component (PC) images from which the subject scores are computed, and also provide labeled scatter plots of the distribution of scores. These plots allow the progress of individual subjects to be traced through the tree and enable the user to focus on complex or unexpected classifications. In addition, we present a visual representation of a typical brain activity pattern arriving at each leaf node, and show how this can be compared to a known reference to validate the behaviour of the tree.

Keywords: Decision tree; Visualization; Medical imaging; Parkinson's Disease; Computer aided diagnosis.

1   Introduction

Decision trees provide a flexible approach to classification which has been applied to a wide range of problem domains. Although the construction of decision trees can be a complex process [1], the resulting trees are typically simple and can be understood by users with little background in machine learning. The structure of the tree also encodes knowledge which has been learned, and this is easy to interpret provided that the features which are used have some intuitive meaning. This is in contrast to more advanced classification algorithms which operate as a 'black box' and for which the internal representation is difficult to comprehend.

Previous work [2] has applied decision trees to the problem of separating healthy subjects from those suffering from Parkinsonian syndromes [3, 4, 5], based on FDG-PET images obtained before a diagnosis could be made by traditional means. Principal component analysis (PCA) is applied to the images and each subject is assigned a set of *subject scores* which are the features used by the decision tree for

classification. This yielded a classification accuracy of up to 80% (dependent on the precise Parkinsonian syndrome being diagnosed), but the subject scores are not intuitive nor easy to reason about. Hence it is difficult to extract meaningful information directly from the resulting tree (an example is shown in Figure 1).

In this paper we extend this tree visualization with the aim of understanding how a given decision tree

Figure 1: This decision tree (adapted from [2]) separates subjects with multiple-system atrophy (MSA) from healthy controls. Although we can see exactly which conditions are applied to classify a patient, this is not enough to give an intuitive understanding of what the classifier has learned due to the abstract nature of the subject scores.
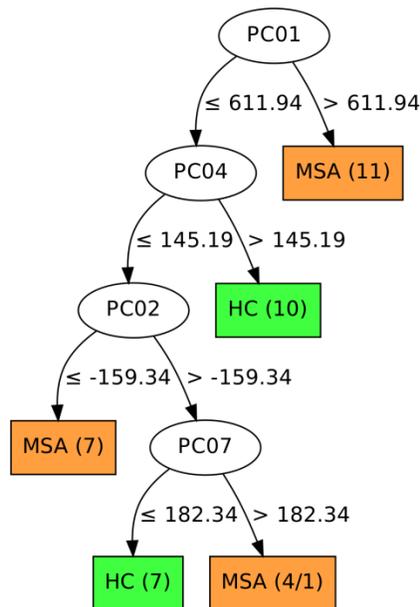
was constructed, determining why specific parameters were chosen and, most importantly, revealing the knowledge which is captured by the resulting tree. To do this we augment internal nodes with thumbnails of the corresponding principal component images, and show how these can be combined to yield typical patterns of relative brain activity for the leaf nodes. Note that our goal is therefore to improve the visualization of the tree and not to show that the decision tree is necessarily the best solution to the classification problem.

The remainder of this paper is structured as follows. Section 2 introduces the literature related to our image analysis approach and the visualization of decision trees, and Section 3 then elaborates further on how the analysis is performed. Section 4 presents the enhanced decision tree and a discussion of its salient features, and Section 5 expands on these ideas to create an interactive system for decision tree exploration. Section 6 then presents a quantitative analysis and informal feedback from neuroscientists before Section 7 gives our closing remarks and ideas for future work.

2 Related Work

Parkinsonian syndromes are a family of neurodegenerative conditions primarily affecting the elderly (over 60 years of age), and are typically characterized by progressive complaints of muscle stiffness,

slowness of movement and sometimes tremor. The family includes idiopathic Parkinson's disease (PD) [3] as well as related disorders such as multiple-system atrophy (MSA) [5] and progressive supranuclear palsy (PSP) [4]. These conditions exhibit similar early symptoms resulting in frequent misdiagnosis in daily clinical practice [6].

Modern brain imaging techniques allow the onset of such neurological conditions to be detected by identifying the associated changes in brain activity[7]. This is most widely done with univariate methods such as Statistical Parametric Mapping (SPM) [8], but recent work has shown that the multivariate Scaled Subprofile Modelling (SSM) approach has a number of advantages and is effective at differentiating between different Parkinsonian syndromes [9, 10]. The SSM technique involves performing principal component analysis on two or more groups (e.g. diseased subjects and healthy controls) and then projecting the brain images onto the principal component images to obtain a set of scores for each patient. We discuss this process in more detail in Section 3.

Decision trees [11] are a widely used approach to classification which have a number of desirable properties. They can be applied to data sets containing a mixture of numerical and categorical variables, are robust in the presence of missing attributes, and perform well with large datasets. Although easy to understand, the process of constructing the tree is non-trivial with numerous algorithms and metrics being proposed in the literature [12, 13, 1]. In this work we construct decision trees using the C4.5 algorithm [1], which uses the concept of information gain [14] to recursively split the dataset.

The structure of a decision tree is most often visualized as a simple hierarchy of nodes (see Figure 1), although more elaborate schemes such as mosaic plots [15], icicle plots [16], and the sunburst layout [17] have also been used. Comprehension can be aided by displaying the distribution of instances at each node or by showing an indication of the confidence at each stage of the tree [18]. More recently, interactive versions of these visualization have been developed, allowing the user to actively participate in the tree construction process [19, 20].

3   Data acquisition

An appreciation of this work is dependent on a high-level understanding of the processing pipeline which transforms FDG-PET scans into decision trees. Full details of the process can be found in [9] (preprocessing, PCA analysis, derivation of subject scores) and [2] (decision tree construction). Here we restrict ourselves to providing a brief overview.

The inputs to the system are FDG-PET scans selected from a previous study [21] describing 18 Healthy Controls (HC), 20 PD, 21 MSA and 17 PSP patients, with clear retrospective diagnoses according to established clinical research criteria. PD patients were 9 males (M), 11 females (F), 6 right body-side affected, 14 left-side affected, with mean age of 63±9 years and disease duration (DD) at scanning of 3±2 years. Also 14 probable MSA and 7 possible MSA patients (10 M, 11 F, age 64±10; DD 4±2), and 13 probable and 4 possible PSP patients (9 M, 8 F, age 68±8; DD 2±1) were included.

To illustrate the SSM process we refer to the images shown in Figure 2(a). Note that the scans are actually three dimensional, and that all such images in this paper are simply depicting a single slice of the full data. Also, absolute intensity values of FDG-PET data are generally not meaningful for diagnostic purposes as they naturally vary according to the patient, the scanner, and the amount of FDG injected. The SSM procedure therefore performs a log transformation and double-centering of the data to

eliminate such scaling effects, and applies a mask to remove voxels outside of the brain. Principal component analysis is then applied to generate a number of 3D principal component images which is equal to the number of input images. Slices through the first four of these principal component images are shown in Figure 2(b).
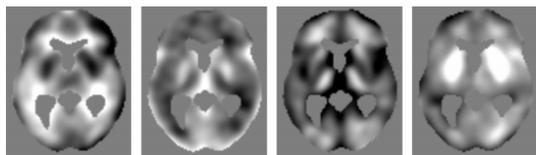
The subject scores are computed by projecting each input image onto each principal component image, and the resulting scores indicate the extent to which each principal component is expressed in each subject.

That is, each subject receives a number of scores which matches the number of principal components and in turn the number of subjects. Examples of these scores are shown in Figure 2(c) and constitute the

Figure 2: Input images are hard to differentiate by eye, but principal component analysis leads to subject scores which can be used to separate the classes.
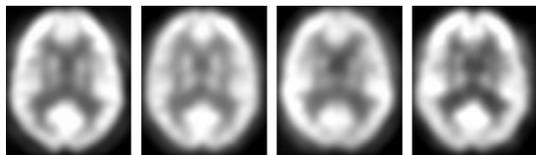
(a) Preprocessed FDG-PET scans.

(b) The first four component images.

	PC1	PC2	PC3	PC4	...	Class
Scan 1	-625	826	-1164	149	...	HC
Scan 2	186	1395	135	207	...	HC
Scan 3	1273	-1420	-1070	947	...	MSA
Scan 4	-1331	-159	887	-1501	...	MSA
Scan 5	...	...	...	...	...	...
Scan 6	...	...	...	...	...	...
...	...	...	...	...	...	...

(c) Subject scores are computed as the projection of each scan onto each principal component image.

feature (or *attribute*) vectors which are used to train the decision tree. It should be clear at this point that the subject scores and the principal component images are intrinsically related, and it is hard to interpret the former without the latter being available for context.

Lastly, we use the C4.5 algorithm to learn from the subject scores. This part of the process is covered in detail in [2], and when separating healthy subjects from those with MSA it results in a decision tree such as the one shown previously in Figure 1.

4   Static visualization of the decision tree

Although Figure 1 clearly shows us the steps involved in the decision-making process, it does not provide much further insight into what has been learned. The tree is appropriate for use as a classifier on new data, but it is desirable to also understand why certain parameter values are chosen, what the implications would be of changing these parameters, and why certain instances are misclassified. Perhaps most importantly, we wish to extract examples of what a typical instance of each class looks like. Such a typical example can be compared to a clinician's own expectations to further increase confidence that the tree is behaving as expected.

The principles which we will describe are applicable to all disease groups, but we choose to explain them in the context of separating subjects with MSA from healthy controls (HC). This classification task was moderately successful as presented by [2], but the reported classification rate of 72% means we can still hope to gain insight into why the classifier does not perform better.

The primary reason why Figure 1 provides only limited insight into the problem domain is that the subject scores are not intuitive to reason about. It therefore makes sense to consider the chosen subject score and threshold for a given node in the context of its corresponding principal component image.

Figure 3: Scatterplots and thumbnails provide additional information on each principal component (PC). Positive values are shown in red with negative values are shown in blue.
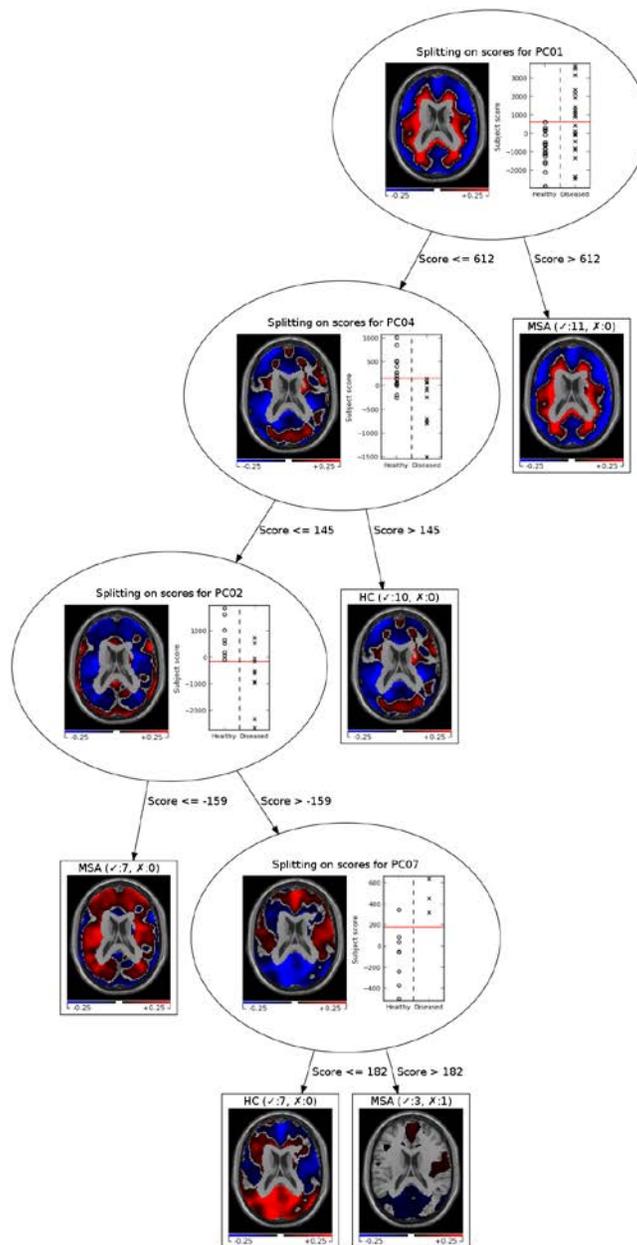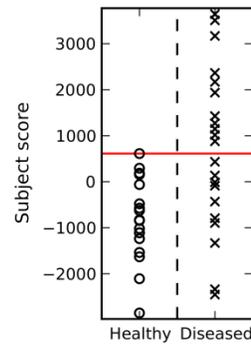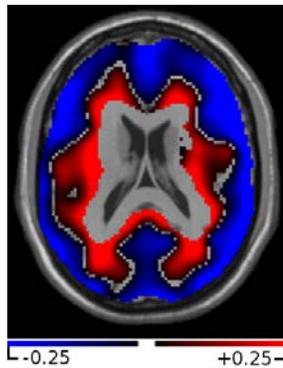

Figure 4: A number of diseased subjects can immediately be identified by their strong expression of the first

principal component.

(a) The first principal component image.

(b) Distribution of subjects around the threshold.

We enhance the decision tree shown in Figure 1 with the following additional attributes/properties:

o   For each non-leaf node, we add a scatterplot showing the distribution of subject scores for the chosen principal component, along with a visual indication of where the threshold is positioned.

o   For each non-leaf node, we add a thumbnail of the principal component image. This is overlaid on a template magnetic resonance (MR) image for anatomical context.

o   For each leaf node, we add a thumbnail of the typical pattern of incoming instances. This is a linear combination of the principal component images encountered when traversing the tree to that node.

The resulting visualization is shown in Figure 3. However, decision trees can grow relatively large and both the image thumbnails and scatter plots include fine detail which cannot be easily resolved given the limited physical space available in this paper. We therefore provide enlarged views of the relevant elements as we discuss their functionality, and use Figure 3 to simply provide a structural overview.

We begin our analysis by looking at the root node of the tree, the visual components of which are enlarged in Figure 4. The C4.5 algorithm has first chosen to split the data set based on the subject scores derived from the first principal component image. This is a common choice, as the first principal component image explains the greatest amount of variance in the entire dataset and we can expect that this will lead to a high information gain.

Although the input data contains only positive values, all image thumbnails contain both positive and negative components as a result of the double-centering operation which is performed prior to the PCA. We apply a red-blue colourmap to highlight these regions, and scans which yield a high subject score when projected onto this PC image are exhibiting a significant difference in brain activity between patients and healthy controls in the marked areas.

Looking at the scatterplot in Figure 4(b), we can also see why the given threshold was chosen. The subject scores do not cleanly separate the healthy and diseased groups into two distinct clusters as we may have hoped, but instead define two overlapping clusters where the upper bound of the diseased group extends significantly beyond the upper bound of the healthy group. The threshold has been selected to separate instances in this extended range and classify them as diseased. From a clinical

Figure 5: Enlarged version of two scatterplots from Figure 3. Markers can be replaced with numerical identifiers so that a given instance can be found in different plots. Horizontal offsets are applied to reduce marker overlap. The blue arrow indicates the subject which was eventually misclassified.

(a) Distribution for PC07   (b) Distribution for PC02
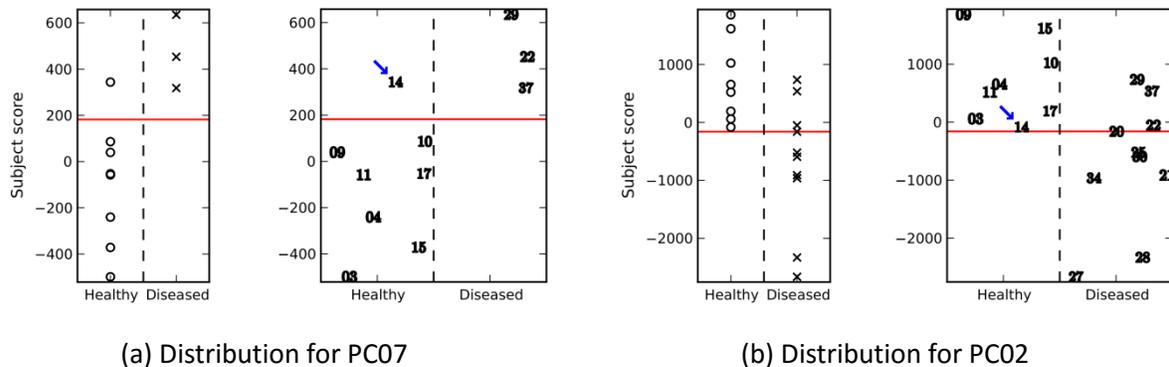
perspective, this corresponds to subjects having such high activity in the areas identified on the corresponding thumbnail that they can immediately be classified as diseased.

Further insight is provided by subsequent scatter plots as we follow the left branches of Figure 3 to examine the subject scores corresponding to the fourth and then second principal components (PC04 and PC02). The scatter plots reveal that the subject scores have a distribution in which the two classes are largely distinct but overlap. The C4.5 algorithm always uses a single threshold to perform a binary split of the data, but we can see in these cases that a tertiary split may also have been appropriate to separate the classes into 'healthy', 'diseased', and 'unknown'. Of course, if desired the algorithm could have emulated such tertiary splitting by using the same split attribute for two successive nodes.

The final test is performed according to the subject scores derived from PC07, and it is here that we see our first (and only) misclassification occur as a healthy subject is incorrectly classified as MSA. The scatter plot in Figure 5(a) shows that this happens as a result of a slight overlap between the ranges of the healthy and diseased scores, and if the groups were slightly more separated we can expect that the threshold would have been chosen to separate them correctly. Accepting the distribution as it is, we can also see why the C4.5 algorithm stopped the decision tree construction process at this point. Attempting to further separate would result in leaf nodes with only a single instance, meaning the tree would almost certainly be over-fitted to the dataset.

4.1 Tracing individual instances

Our misclassified subject also serves as an illustration of why it can be useful to trace the route of an individual subject through the decision tree. As discussed, this particular subject was only narrowly misclassified by the last internal node, and we may be interested in knowing where this particular subject fell relative to the other attributes and split points which were used earlier in the decision-making process. We do not get this information from Figure 3 because we cannot differentiate any given subject from another of the same class, as we have only high-level statistical information about the distribution at each node.

To solve this we wish to make each instance visually distinct in the scatter plots. We considered the use of different colours and shapes for the markers, but eventually settled on using a unique numerical

identifier which is drawn at the appropriate position on the scatter plots (see Figure 5). An advantage of this approach is that we can use the identifier to refer back to the appropriate row of the input data table or to the original image. We also apply a random horizontal offset to the markers to reduce problems with markers overlapping. Note that the horizontal position of the markers is therefore *not meaningful* (except with respect to the dashed line separating diseased from healthy subjects) and is used only to make the markers more legible.

Looking to Figure 5(a), our misclassified subject is healthy and so appears on the left of the vertical dashed line, but lies above the threshold indicated by the red horizontal line (hence the misclassification). The version of the plot with numerical identifiers shows that this is subject number 14, which is information we did not have previously. We can now track this subject back up the tree. For example, the previous node (shown with identifiers in Figure 5(b)) made use of PC02 and subject 14 is again very close to the threshold. At this point the subject has been narrowly misclassified by two separate nodes, and this may warrant a closer examination of the image data by clinicians.

4.2 Reconstruction of typical disease patterns

The techniques described so far allow the user to better understand the behaviour of the decision tree with respect to the individual subjects which are classified by it. Ideally, we would also like to understand its behaviour at a more global level, by extracting the knowledge which has been learned during its construction. More specifically, we would like to extract a representation of what a typical subject image at each node actually looks like. In neuroimaging this is usually presented in the form of a pattern showing relative differences in neural activity between healthy and diseased subjects. Such a pattern can be used directly to classify new subjects and also serves as validation of the decision tree's behaviour because it can be compared to reference patterns obtained using other methods [22].

We can form a typical pattern for a leaf node by taking a linear combination of the principal component images which are used to reach it. The weights which are used should reflect the importance of a given image in the classification process, but they may also change depending on which class we are working with. For example, PC01 is instrumental in identifying the MSA class because a *high* score immediately gives a correct classification, but the reverse is not true in that a *low* score does not imply the subject is healthy. Figure 4(b) instead shows a very mixed distribution below the threshold.

We can measure the purity of a group of instances using the concept of information [14]. This same principle (actually information *gain*) is used by the C4.5 algorithm to determine which attribute should be used to split the dataset, and so it is a natural extension to use this as a weighting factor. We compute the amount of information in each of the two groups leaving a node, and weight the principal component image by this value when following the path to a given leaf.

After weighting in this way we observe that the largest contributor to the typical pattern in each leaf node is usually its parent node. Intuitively this makes sense, as instances arriving at a given leaf were not able to be classified by decision nodes prior to the leaf node's immediate parent. Typically we see that approximately 90% of a leaf node's pattern is obtained from its parent's PC image with the remainder of the pattern being provided by ancestors higher up in the tree.

Figure 6: The derived pattern is a linear combination of components 1, 2, 4 and 7 (components and weights as chosen by the tree) while the reference pattern is a different linear combination of components 1, 2 and 4 (see [22]). A similar structure can be seen in each of these two patterns.

(a) Derived pattern   (b) Reference pattern
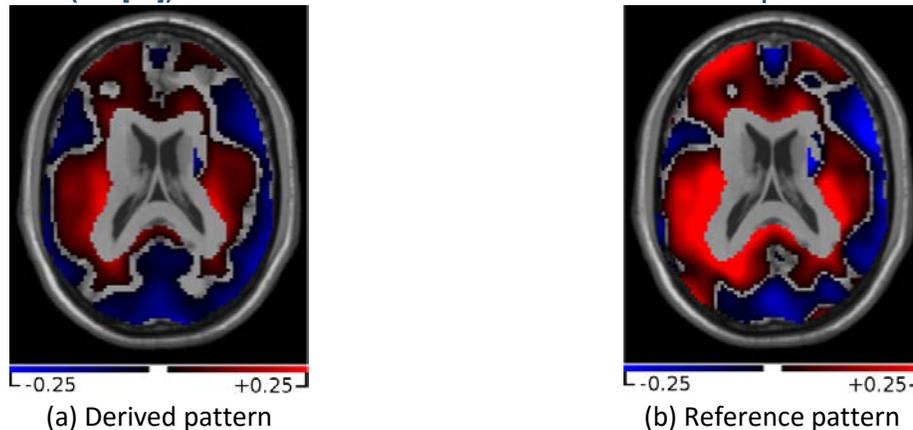
Additional consideration must be given to subjects with negative scores. Looking at Figure 3 and considering node PC02, we can see that subjects are classified as diseased if they have a score *below* -159, whereas earlier nodes assigned a class if the score was *above* a given threshold. This means these diseased subjects have an *inverse* expression of this principal component image. This is a valid outcome as patterns are expressed as *relative* differences in brain activity, and so healthy and diseased patterns are expected to be the inverse of each other. Hence the thumbnail for leaf node 'MSA (✓:7, ✗:0)' is composed mostly of the inverse of its parent's principal component (PC02) with small contributions from other ancestors.

The significant variation between the leaf node patterns indicates that successive stages of the classification process are making use of different regions of the brain. However, it is useful to also derive a single overall pattern as a linear combination of all the diseased leaf nodes and the inverses of all the healthy leaf nodes. We therefore compute such an overall pattern where the weighting factor of each node is simply the number of subjects it classifies. The resulting pattern is presented in Figure 6(a) and discussed further in Section 6.

5   Interactive visualization of the decision tree

We have now presented the key components of our decision tree visualization and explained how they improve user understanding. The tree has thus far been computed as an offline process which generates a static image (Figure 3) which can be examined. Henceforth, this image will be referred to as the *static visualization*, because in this section we demonstrate how similar concepts can be applied in an *interactive system* to bring increased flexibility and a better workflow to the user.

Our interactive system is implemented as an extension to the Orange data analysis software [23] which provides a generic framework for machine learning and visualization. Orange allows an analysis pipeline to be defined as an ordered collection of nodes, with data flowing from input nodes, through processing nodes, and finally onto visualization nodes. Within the context of our problem, this means creating a pipeline (Figure 7(a)) which reads subject scores from a file on disk, feeds them into a decision tree learner, and then allows the result to be examined with the 'decision tree viewer' node. Users can click

on any pipeline node to open a window containing configuration and visualizations as shown in Figure 7(b–e).

Figure 7: Images of our interactive decision tree exploration tool based on Orange [23]. A sample data flow graph (a) demonstrates the use of widgets for decision tree construction and image visualization. The user can click on nodes in the tree explorer (b) to see the images updated in real time. The visualization can display principal component images (c), group reconstructions (d) or the original subject scans (e).
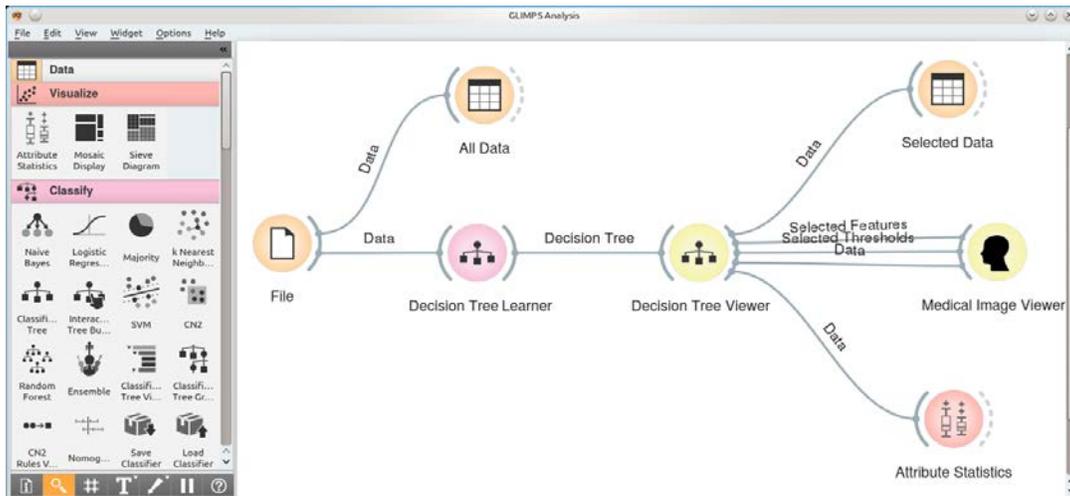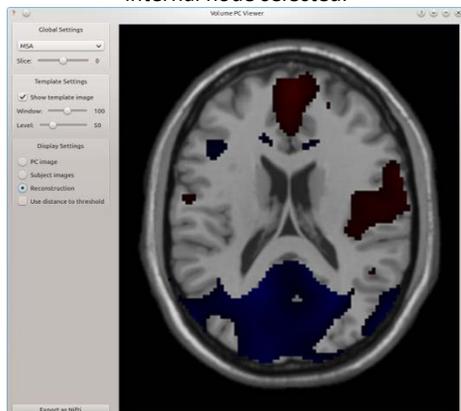

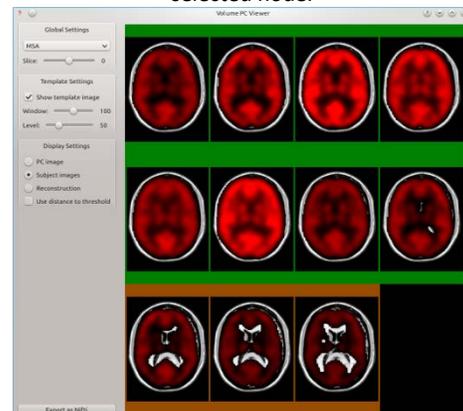(a) Data flow pipeline defined in the Orange software.
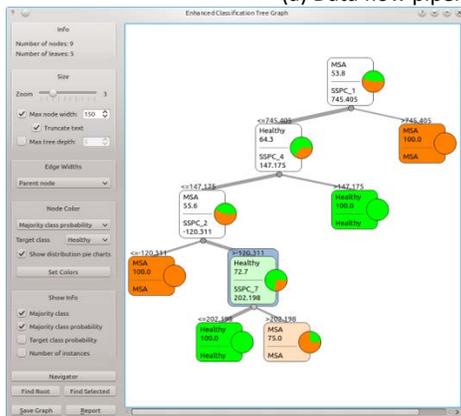
(b) The existing 'decision tree viewer' with the lowest internal node selected.
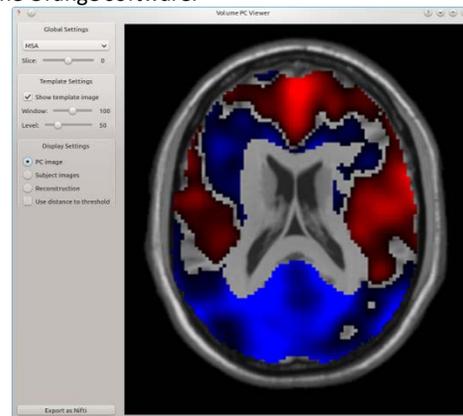
(c) The principal component image corresponding to the selected node.

(d) Reconstruction of the typical pattern arriving at the selected node.

(e) Subject scans being classified by the selected node.

Existing functionality in Orange already enables a more flexible approach to classification than the static tree shown previously, because the 'decision tree learner' node allows the modification of key learning parameters such as the attribute selection criterion, the maximum number of instances per leaf, and the maximum tree depth, with changes being shown in real time. Additionally, the 'decision tree viewer' is interactive and shows additional information such as the distribution of classes at each node via a small pie chart (Figure 7(b)). But our key contribution to this system is the addition of a new type of visualization node which embeds medical image data directly inside Orange. This 'medical image viewer' node can be seen on the right-hand side of Figure 7(a) and the expanded form is shown in Figure 7(c-e).

The image viewer features several modes of operation. At a basic level it allows the user to visualize single 3D images such as preprocessed subject scans or principal component images. As such it serves as a direct replacement for the thumbnails which form part of the static visualization, with the additional benefit that the user is able to scroll through the individual slices of an image and perform common operations such as adjusting the window-level settings. This clearly allows for more careful examination of the data than is possible with a single thumbnail image. The only drawback compared to the static visualization is that our interactive system currently only displays image data for a single node at a time, and so in this sense the static visualization provides a better overview. Note that Orange also provides a built-in widget for displaying scatterplots of the subject scores but this is not shown in Figure 7.

The interactive system also exposes additional functionality which would be difficult to implement in the static version. Most notably, the image viewer can present a 'small multiples' view in order to display several 3D images at the same time. The images which are displayed are synchronized to the currently selected node and this allows the individual subject scans of patients to be examined as they pass through the tree. Furthermore, we color the background (dark green and dark orange in Figure 7(e)) of these thumbnails to indicate their true class so that the subject scans of misclassified patients can be easily identified. Adding such additional images to the static visualization would quickly overload the user with information.

The workflow of the interactive system improves significantly over the static visualization. Further analysis of the subjects in Figure 3 requires separate medical imaging software and burdens the user with the task of finding and loading the appropriate images based on the numerical identifiers which are read from the scatterplots. By integrating the image viewer and tree visualization into a single package the user is relieved of this task, and is able to quickly switch between different images as desired. If further analysis in an external application is desired then the user is able to export the reconstructed images in standard NIfTI format.

6 Evaluation

We believe that the methods presented in this paper can aid a user in understanding and visualizing the classification process, but we do not claim that a decision tree is necessarily the most accurate method of classification. Promising results have also been obtained using other methods [9, 24]. However, a decision tree is certainly a *viable* method for separating MSA from healthy controls as shown by the results of a set of leave-one-out cross-validations which were performed in previous work [2]. In this

section we do not focus any further on the performance of the tree, but instead provide validation of the derived pattern and present clinical feedback which was obtained from neuroscientists.

Figure 8: Individual subjects and the pattern derived from the tree are both projected onto the reference pattern. Derived patterns for both MSA and PSP score very highly. The PD pattern is still representative of the group but does not serve as such a distinguished example. Horizontal jittering of markers serves only to reduce clutter.

(a) MSA   (b) PSP   (c) PD
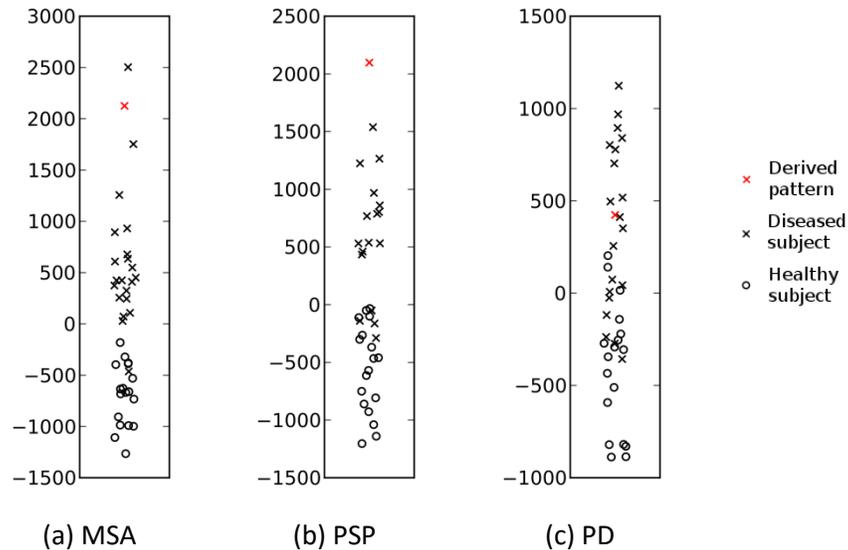
6.1 Validation of the derived pattern

In Section 4.2 we described how a typical pattern for a diseased subject can be extracted from the tree as a linear combination of the principal component images (Figure 6(a)). For comparison, Figure 6(b) shows the reference pattern created using the method of Teune *et al.* [22]. This reference pattern is again a linear combination of the principal component images, with the weights being computed according to the Akaike information criterion. The structural similarities between the two images are immediately apparent, and show that the decision tree has indeed captured some knowledge regarding the typical pattern of brain activity in MSA subjects.

We can perform a more quantitative validation of this claim by treating the derived pattern as a new subject and projecting it onto the reference image along with the rest of the subjects. This is shown in Figure 8(a) with the derived pattern marked in red. It can be seen that the derived MSA pattern does indeed have a strong expression of the reference pattern (actually stronger than most of the individual subjects) and in this sense it can be considered representative of what a typical MSA subject would look like. We can also apply the same principle to the PSP and PD trees from [2] and observe that the PSP derived pattern is stronger than any of the individual PSP subjects whereas the PD result is less striking. This corresponds with the relative performance of the trees given by Mudali *et al.* [2].

6.2 Clinical feedback

Although the presented results serve to validate the behavior of the tree in a quantitative fashion, for the purpose of this work we are actually more interested in understanding and maximizing the benefits

of the visualization. Our key claim is that the depiction of the decision tree given in Figure 3 is more useful to neuroscientists than the depiction given in Figure 1. To verify this claim we presented the

Figure 9: Enlarged image and scatterplot for the first principal component (the root node of the tree). This is the same data as Figure 4 but with the markers replaced by subject numbers as described in Section 4.1.

(a)   (b)
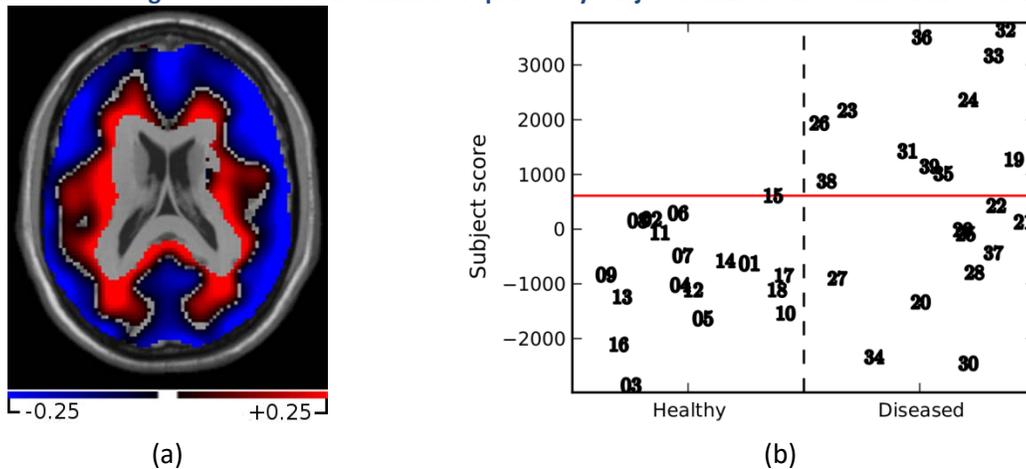
visualization to neuroscientists at the University Medical Center Groningen, The Netherlands.

The evaluation was relatively informal and involved identifying interesting subjects in the tree (e.g. those with extreme values, those near thresholds, and the misclassified subject) and then analyzing the corresponding subject scans or clinical diagnosis. Information about the distribution and classification of

individual subjects was not provided by the original decision tree in Figure 1, so any insight gained in this process is the result of the additional visualizations added by Figure 3. Note that we are not trying to prove any clinical hypothesis at this point as we have insufficient data for that purpose, but instead we simply wish to demonstrate that the tree can guide the user to additional insight.

During the evaluation the following key observations were made:

o The principal components images do indeed show activity patterns which are meaningful to neuroscientists. For example, the first principal component image shows relatively high activity in the cerebellum and low activity in the basal ganglia. However, these patterns can be difficult to identify in the thumbnail images (Figure 9(a)) as it is often necessary to scroll through all the slices of the volume.

o The scatterplots show an interesting distribution of values, particularly for the first principal component (shown enlarged in Figure 9(b)). It can be clearly seen that a highly positive expression of this principal component is indicative of a diseased subject, while the inverse is not true. Indeed, a highly negative expression can be associated with both healthy and diseased groups. This characteristic is already known to neuroscientists though its explanation remains unclear. There is speculation that it is related to different subgroups of MSA (MSA-P vs MSA-C) but we do not have enough data to support this.

o The same scatterplot also makes it easy to identify the extreme cases and look for common traits. For example, the two highest-scoring subjects (32 and 36) are both males of the same age

and exhibit similar clinical symptoms. Conversely, the two lowest-scoring subjects (30 and 34) are both younger females and have symptoms not present in the high scoring subjects (criteria autonomic dysfunction, orthostatic fall in blood pressure). Such traits are relatively easy to

o identify when using the decision tree visualization in conjunction with the existing patient diagnosis.

o A new examination of subject 14 has still found no clinical signs of MSA, despite being misclassified as MSA by the decision tree (see Figures 3 and 5). However, it was observed that this subject was originally scanned because it was the sibling of a sufferer of Huntington's disease (a neurodegenerative genetic disorder). Unfortunately no further information is available on their present situation.

o Subjects 9, 10, 14, and 37 were the most difficult to classify with the decision tree, as they passed through to the final node and still ended up near the threshold (as shown previously in Figure 5(a)). Cross referencing the medical records revealed that subject 37 was a particularly difficult case due to a hypometabolic defect laterally to the right side of the striatum. The assessment of the nuclear medicine physician was that this was not a clear case of MSA and could instead be a result of corticobasal degeneration or a local defect such as stroke or trauma.

o There were some concerns over the quality of the input data as the healthy controls are actually taken from two separate studies and have some unknown reconstruction parameters. This can have an impact on the resulting images and, in turn, the principal components. There is on-going work to obtain a larger and cleaner dataset for a more thorough evaluation.

Overall, feedback from neuroscientists was positive and they agreed that the tree led to interesting observations about the input data. This validates our claim that the visualization is useful for gaining insight, but we cannot draw strong clinical conclusions at this point.

7   Conclusion

In this paper we have shown how additional visualizations can improve a user's understanding of the classification process, and provide insight into the knowledge which has been learned by a decision tree. We have focused on understanding the diagnosis of Parkinsonian syndromes, but it is likely that similar ideas can be applied to other problem domains (particularly those involving the classification of images). Informal feedback has been very positive and the system shows strong potential for aiding the diagnostic process.

While using the static decision tree we have encountered a few areas of potential improvement. Most importantly, the layout of the subject numbers in the scatter plots should be improved to prevent them from overlapping with each other, as this makes some numbers difficult to read. We currently apply a random horizontal offset to each subject number at the moment it is drawn to the tree, but a better algorithm could wait until the set of all subject numbers is known before computing an optimal layout.

Furthermore, the decision tree gives a binary classification at each node but it may be desirable to also compute a confidence measure for each classification. The exact approach will be subject to further consideration, but it would seem appropriate to use the distance from the threshold(s) or the amount of information gain as an initial metric. This level of confidence should then also be shown visually on the tree.

Future improvements to the interactive visualization are likely to focus on the workflow, and particularly on allowing data and selections to be passed between the different nodes of the Orange pipeline. For example, selecting a node in the decision tree will currently display the corresponding data in the image viewer, but the reverse is not possible as data only flows left-to-right. The interactive system is currently at the prototype stage so there are many opportunities to improve this kind of workflow issue.

ACKNOWLEDGMENTS

REFERENCES

[1].   J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[2].   D. Mudali, L. K. Teune, R. J. Renken, K. L. Leenders, J. B. T. M. Roerdink, Classification of Parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features, Computational and Mathematical Methods in Medicine Article ID 136921 (2015) 1–10. doi:10.1155/2015/136921.

[3].   J. M. Shulman, P. L. De Jager, M. B. Feany, Parkinson's disease: Genetics and pathogenesis, Annual Review of Pathology: Mechanisms of Disease 6 (1) (2011) 193–222. doi:10.1146/annurev-pathol-011110-130242.

[4].   J. Steele, J. Richardson, J. Olszewski, Progressive supranuclear palsy: A heterogeneous degeneration involving the brain stem, basal ganglia and cerebellum with vertical gaze and pseudobulbar palsy, nuchal dystonia and dementia, Archives of Neurology 10 (4) (1964) 333–359. doi:10.1055/s-0034-1377058.

[5].   L. Swan, J. Dupont, Multiple system atrophy, Physical Therapy 79 (5) (1999) 488–494. URL http://www.ncbi.nlm.nih.gov/pubmed/10331752

[6].   A. J. Hughes, S. E. Daniel, Y. Ben-Shlomo, A. J. Lees, The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service, Brain 125 (4) (2002) 861–870. URL http://www.ncbi.nlm.nih.gov/pubmed/11912118

[7].   L. K. Teune, A. L. Bartels, K. L. Leenders, FDG-PET imaging in neurodegenerative brain diseases, in: F. Signorelli (Ed.), Functional Brain Mapping and the Endeavor to Understand the Working Brain, InTech, 2013.

[8].   T. Eckert, A. Barnes, V. Dhawan, S. Frucht, M. F. Gordon, A. S. Feigin, D. Eidelberg, FDG PET in the differential diagnosis of parkinsonian disorders, Neuroimage 26 (2005) 912–921. URL http://www.ncbi.nlm.nih.gov/pubmed/15955501

[9].   P. G. Spetsieris, D. Eidelberg, Scaled subprofile modeling of resting state imaging data in Parkinson's disease: Methodological issues, NeuroImage 54 (4) (2011) 2899 – 2914. doi:10.1016/j.neuroimage.2010.10.025.

[10]. C. C. Tang, K. L. Poston, T. Eckert, A. Feigin, S. Frucht, M. Gudesblatt, V. Dhawan, M. Lesser, J.-P. Vonsattel, S. Fahn, D. Eidelberg, Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis, The Lancet Neurology 9 (2) (2010) 149–158.

[11]. L. Rokach, O. Maimon, Data Mining with Decision Trees, Vol. 69, World Scientific, 2008.

[12]. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, Chapman and Hall, 1984.

[13]. J. Quinlan, Induction of decision trees, Machine Learning 1 (1) (1986) 81– 106. doi:10.1023/A:1022643204877.

[14]. C. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (3) (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

[15]. H. Hofmann, A. P. J. M. Siebes, A. F. X. Wilhelm, Visualizing association rules with interactive mosaic plots, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '00, ACM, 2000, pp. 227–235. doi:10.1145/347090.347133.

[16]. M. Ankerst, M. Ester, H.-P. Kriegel, Towards an effective cooperation of the user and the computer for classification, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '00, ACM, 2000, pp. 179–188. doi:10.1145/347090.347124.

[17]. J. Stasko, An evaluation of space-filling information visualizations for depicting hierarchical structures, Int. J. Hum.-Comput. Stud. 53 (5) (2000) 663–694. doi:10.1006/ijhc.2000.0420.

[18]. M. de Oliveira, H. Levkowitz, From visual data exploration to visual data mining: a survey, IEEE Transactions on Visualization and Computer Graphics 9 (3) (2003) 378–394. doi:10.1109/TVCG.2003.1207445.

[19]. S. van den Elzen, J. van Wijk, Baobabview: Interactive construction and analysis of decision trees, in: Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on, 2011, pp. 151–160. doi:10.1109/VAST.2011.6102453.

[20]. S. T. Teoh, K.-L. Ma, Paintingclass: Interactive construction, visualization and exploration of decision trees, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, ACM, 2003, pp. 667–672. doi:10.1145/956750.956837.

[21]. L. K. Teune, A. L. Bartels, B. M. de Jong, A. T. Willemsen, S. A. Eshuis, J. J. de Vries, J. C. van Oostrom, K. L. Leenders, Typical cerebral metabolic patterns in neurodegenerative brain diseases, Mov Disord 25 (2010) 2395–404. doi:10.1002/mds.23291.

[22]. L. K. Teune, R. J. Renken, D. Mudali, B. M. D. Jong, R. A. Dierckx, J. B. T. M. Roerdink, K. L. Leenders, Validation of parkinsonian diseaserelated metabolic brain patterns, Movement Disorders 28 (4) (2013) 547–551. doi:10.1002/mds.25361.

[23]. J. Demar, B. Zupan, Orange: Data mining fruitful and fun - a historical perspective, Informatica 37 (1) (2013) 55–60.

[24]. G. Garraux, C. Phillips, J. Schrouff, A. Kreisler, C. Lemaire, C. Degueldre, C. Delcour, R. Hustinx, A. Luxen, A. Deste, E. Salmon, Multiclass classification of FDG PET scans for the distinction between parkinson's disease and atypical parkinsonian syndromes, NeuroImage: Clinical 14 (2) (2013) 883–893. doi:10.1016/j.nicl.2013.06.004.