Using deep-learning to predict outcome of patients with Parkinson’s disease


Abstract— There are currently no established disease modifying therapies for PD, and prediction of outcome in PD to power clinical studies is a very important area of research. Assessment of PD is informed by imaging the dopamine system with dopamine transporter (DAT) single-photon emission computed tomography (SPECT) imaging and by the presence of key symptoms. Recently, deep-learning based methods have shown promise for medical image analysis tasks and disease detection. The purpose of this study was to develop a deep-learning based approach to predict outcome of patients with PD using longitudinal clinical data containing imaging and non-imaging information. Features were first extracted from the clinical data by the proposed deep-learning based approach and then combined to predict motor performance (MDS-UPDRS-III) in year 4. The performance of the proposed approach was evaluated via a 10-fold cross-validation. We evaluated the performance of the network on the basis of mean absolute error (MAE) between the predicted and true MDS-UPDRS part III scores in year 4. The proposed approach yielded a MAE of 4.33±3.36 when given only imaging features, 3.71±2.91 when given only non-imaging features, and 3.22±2.71 when given all input data. While the approach given only non-imaging input data outperformed the approach given only imaging data, we found that the performance of the proposed approach substantially improved when given both imaging and non-imaging information. Our results indicate that the addition of imaging data to non-imaging clinical data is helpful for the prediction of outcome in patients with PD. The proposed approach that incorporated both imaging and non-imaging clinical data shows significant promise for prediction of outcome in patients with PD.

I. INTRODUCTION

Parkinson’s disease (PD) is the second most common neurodegenerative disorder which is characterized by neuronal loss of dopaminergic neurons in the substantia nigra [1]. PD is a progressive movement disorder where the loss of dopamine levels can cause progressive motor and non-motor symptoms. Patients with PD may exhibit motor symptoms, such as resting tremor, bradykinesia, muscle stiffness and postural instability, as well as non-motor symptoms including cognitive problems and autonomic nervous system dysfunction usually occurring in the later stages of the disease [1]. The diagnosis of PD is informed by the presence of such key symptoms and by imaging the dopamine system with 123I-isoflupane-dopamine transporter (DAT) single-photon emission computed tomography (SPECT) [1]. There are currently no established disease modifying therapies for PD, and prediction of outcome in PD to power clinical studies is an active area of research [1]-[5]. Due to the need for identifying biomarkers of PD progression, the Parkinson’s Progression Markers Initiative (PPMI) has made available longitudinal clinical data of patients with PD that included a database of non-imaging clinical measures of PD and DAT-SPECT images [6].

Recently, deep-learning based methods have shown promise for medical image analysis tasks and disease detection [7]. Therefore, in this project, our aim is to develop a deep-learning approach to predict motor outcome of patients with PD by incorporating both imaging and non-imaging information. We aim to develop this deep-learning approach as a prognostic tool that may further characterize patients into different groups. This could lead to determining different treatments or therapy regimens for each patient to ultimately reduce symptoms and to delay the disease progression.

II. MATERIALS AND METHODS

A. Patient data

The longitudinal clinical data, including DAT-SPECT images and clinical measures of patients with PD, were extracted from 198 patients (144 males and 54 females, mean age 67.6±9.98 years, range [39,91]) in the PPMI database. DAT-SPECT images and clinical measures from year 0 (baseline) and year 1 were used as predictors. The non-imaging clinical measures included movement disorder society unified Parkinson’s disease rating scale (MDS-UPDRS) – part III from both year 0 and year 1 as well as age, gender, and diagnosis duration with respect to time of diagnosis and time of appearance of symptoms. For the prediction task, we define the composite MDS-UPDRS-III score in year 4 as outcome.

The DAT-SPECT images were preprocessed by selecting a continuous segment of 21 image slices of each image where the center slice had the highest relative intensity in the transaxial direction. The images were then zero padded resulting in 128 x 128 x 21 sized images for both year 0 and year 1.

B. Varying the input to the proposed approach

Given the availability of a heterogeneous longitudinal dataset, we developed several deep-learning based approaches that used different input data. The first method uses only information from DAT-SPECT images. The second method...
uses data only from the non-imaging clinical features. And the last method is given both imaging and non-imaging features. These methods are shown in Table 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>Input Data (Years 0 &amp; 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Only DAT-SPECT Images</td>
</tr>
<tr>
<td>2</td>
<td>Only non-imaging clinical features</td>
</tr>
<tr>
<td>3</td>
<td>Both imaging and non-imaging features</td>
</tr>
</tbody>
</table>

C. Proposed deep-learning based approach

Typically, deep-learning methods require very large training set sizes, on the order of thousands of images to adequately train a deep neural network on various image analysis tasks [8]. Due to the availability of our limited dataset of only 198 patients with DAT-SPECT images in years 0 and 1, we first extract features from the DAT-SPECT images from years 0 and 1 with the Google Inception network (Fig. 1), a convolutional neural network (CNN)-based architecture that has been pretrained with the ImageNet dataset, which consists of hundreds of thousands of natural images [9]. The Google Inceptionv3 network has been previously used in transfer learning for a variety of image analysis tasks and has been very successfully applied in several deep-learning based medical applications [10]-[12].

The features learned via the LSTM-based network are then combined with other non-imaging clinical measures such as age, gender, diagnostic duration, as well as the imaging features learned from Google Inception network. These combined features are input into a final fully-connected layer that outputs the prediction of MDS-UPDRS-III scores in year 4. The complete deep-learning based approach is shown in Fig. 3.

D. Evaluation of the proposed approach

To evaluate the performance of each the method, we perform a 10-fold cross validation on the clinical dataset of 198 patients. We first randomly split the training data into 10 separate folds and repeat training 10 times for each method using a different fold as a test set. As previously mentioned, we compare three different methods which take different subsets of the clinical data as input (Table 1). The deep-learning based approach is trained by optimizing a mean squared error loss function that quantifies the error between the true and predicted MDS-UPDRS-III scores in year 4. The proposed approach is evaluated on the remaining test fold for each iteration of the cross validation. The generalization error is estimated by averaging the error over all examples in each test fold of the cross validation. The method that minimizes this generalization error is considered the best performing method.

The performance of the proposed approach was evaluated on the basis of mean absolute error where a lower value is better. Statistical significance was determined by a paired sample t-test where p-value < 0.05 is used to infer a statistical difference.
III. RESULTS

Results using the procedure Section II.D. are shown in Figs. 4 and 5. Methods 1, 2, and 3 yielded a MAE of 4.33±3.36, 3.71±2.91, and 3.22±2.71, respectively. Method 3, which used both imaging and non-imaging clinical features, outperformed other methods (p-value < 0.05) that used only either DAT-SPECT imaging features or non-imaging features as input to predict MDS-UPDRS-III at year 4. Method 1, which used only imaging information from DAT-SPECT images, had worse performance than method 2, which only used the non-imaging clinical information. However, when the imaging features learned from the Inception network are added to the input along with the non-imaging clinical features, the performance of the method substantially improves and results in the lowest MAE for predicting MDS-UPDRS-III at year 4 (Fig. 4).

![Generalization Error](image)

The predicted versus observed outcome measures are shown in Fig. 5. The solid line represents perfect outcome prediction, and the datapoints represent the predicted outcome measures for MDS-UPDRS-III in year 4. Again, method 1, which used only imaging features (Fig. 5a), results in the worse prediction when compared to method 2, which used only non-imaging features (Fig. 5b). Addition of imaging to non-imaging features allows the method to further improve performance by reducing both the MAE and the variance of the prediction (Fig. 5c).

IV. DISCUSSION AND CONCLUSION

While we see that method 3, which used both imaging and non-imaging clinical measures, outperformed methods 1 and 2 for prediction of UPDRS-III scores in year 4, it is important to note that previous methods that attempted to predict UPDRS-III scores in year 4 directly from the DAT-SPECT images were not effective. This may be due to our limited clinical dataset with only 198 patients. In contrast, using the Google Inception network to first extract the DAT-SPECT image features enabled the proposed approach to incorporate information from DAT-SPECT images for improved outcome prediction.

Other strategies to address the lack of clinical data include various data augmentation techniques, such as applying random translations and rotations to images [14]. However, our attempts to train the proposed deep-learning approach with such data augmentation techniques were not effective. This highlights the need for larger clinical datasets for such analysis. Thus, exploring alternative deep-learning based approaches for extracting imaging features that are important for outcome prediction from a large clinical dataset of DAT-SPECT images is an important area of future research.

In conclusion, a deep-learning based approach that incorporated both imaging and non-imaging clinical features was developed and showed significant promise for prediction of outcome in patients with PD.

ACKNOWLEDGMENT

Financial support for this project was provided, in part, by the National Institutes of Health under grant numbers P41-EB024495 and R01-NS094227. The project was also supported by the Michael J. Fox Foundation, including use of data available from the PPMI—a public-private partnership—funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners (listed at www.ppmi-info.org/fundingpartners).

REFERENCES


