Prediction of Parkinson's Disease Based on Medical Imaging Using Multi-Task Learning

Hemalatha[1], Payal A Makhija[2], S Aishwarya[3], Saarah Asad[4]
[1, 2, 3, 4] Dept of Computer Engineering
[1, 2, 3, 4] Atria Institute of Technology, Bangalore, India

Abstract- *Parkinson's disease (PD) is a long-term degenerative disorder of the central nervous system, with symptoms generally appearing slowly over time. Predicting the PD disease is critical as motor and non-motor manifestations occur many years after the onset of neurodegeneration, hence its early management of disease is a significant challenge in the field of PD therapeutics. While part of previous studies with respect to the prediction of Parkinson's Disease has been based mainly on brain images, dependencies between additional patients' information have not been taken into account. This observation suggests that prediction of Parkinson's Disease along with additional patients' data with a unified framework should outperform Machine Learning (ML) algorithms that treat different sources of patients' information separately. Our presented framework relies on Multi-Task Learning (MTL) implemented with Deep Neural Networks (DNNs) with shared hidden layers. Our preliminary experimental results confirm the benefits of MTL over Single-Task Learning (STL), underlying the capability of our proposed system to achieve an increased Area Under the Curve (AUC) as high as 92% and helping at the same time to reduce human error.*

Keywords- Deep Neural Networks, Parkinson's dis-ease, Medical Data Analysis, Multi-Task learning, Computer-Aided Diagnosis

I. INTRODUCTION

Parkinson's Disease (PD) is a neurological disorder expressed through a progressive decline in motor precision and sensor motor integration stemming presumably from a disorder of the basal ganglia [2]. Due to the disease's nature, both motor and non-motor manifestations of PD significantly influence the patients' life in terms of shaking, slowness of movement and postural instability. Up to now, there is no objective med-ical test to make a certain diagnosis of PD. Instead, doctors perform routine basis neurological tests such as MMSE [3],

UPDRS [4] and brain scans. Early prediction of the disease is critical, as it can greatly reduce the medical treatment's economic burden, as well as improve the patient's overall quality of life.. Early prediction is therefore a main challenge in the field of PD therapeutics and the absence of a validated computational unified framework, which could jointly train the PD or non-PD condition along with the patient's epidemiological data, is one of the major obstacles in understanding PD progression.

These observations suggest that an appealing solution to address the aforementioned issues is to jointly learn the patients' complementary indicators. Instead of considering patients' characteristics as not interrelated, we should formulate ML algorithms that leverage the dependencies between them. Motivated by these observations, we attempt to build a unified framework to predict Parkinson's Disease by leveraging patients' complementary characteristics' interdependencies.

Our study formulates the prediction of Parkinson's Dis-ease (corresponding normal condition (non-PD) and PD) as a multi-task learning (MTL) problem. Thus, in our frame-work, presented in Figure 2, learning the attribute of interest, namely the prediction of the PD through imaging data, is treated as the primary task, while the epidemiological data serve as auxiliary tasks. We consider PD, non-PD, age and sex tasks and based on DNN architectures we train systems that learn to jointly model the above tasks. According to our preliminary evaluations, we demonstrate that jointly learning patients' data leads to significant improvements over Single-Task Learning (STL) approaches [5]. This is in line with the suggestions given in [6]. Our goal is to assist doctors to predict accurate and faster the risk of the PD and prevent it in time. Thus, our framework aims to enhance the abilities of doctors and researchers to understand how to analyze the generic variations that will lead to disease

II. RELATED WORK

Previous Computer-Aided Diagnosis methodologies have shown that dependencies between patient's attributes ex-tracted from images through ML approaches, can represent information effectively and efficiently. Up to know, research work in the field PD diagnosis and prognosis has focused on treating different sources of patients' information (patient's information, patient's history, scans etc.) separately [16, 17].

To the best of our knowledge, there are only two recent studies related to our work in terms of assessing Parkinson's Disease based on DNNs [7] and in terms of joint learning multiple sources of patients' clinical information [8]. The former work presents a system based on DNN that performs analysis of medical imaging data. Particularly, the system's inputs, which consisted of consecutive MRI triplets and a single frame from a DaT Scan examination, were combined to perform the final prediction. According to the latter work, even though a MTL approach was followed for the early pre-diction of Parkinson's Disease, the prediction of the PD progression is based on the UPDRS clinical rating scale and the whole approach is treated as a regression problem. Based on this work, each task refers to the prediction of PD rating scales at one future time point. The contributions of our work with respect to previous studies are based on the idea that when the tasks are related, the information learned from each task can be used to enhance the learning of other tasks. It is therefore beneficial to learn relevant tasks together simultaneously, as opposed to learning each task independently [9]. Particularly:

We are the first to apply a neural MTL model to predict PD from Dat Scan images and patients' epidemiological data (i.e. age and sex).

We enrich the number of the Dat Scan images we are using, by applying a number of augmentation techniques (Section 3.1).

We increase our MTL system's Specificity (Spec.) up to 83%, while maintaining a Sensitivity (Sens.) of 100% leading to a AUC of 92%. This result has a direct impact for Healthcare applications as presented in Table 1.

III. METHODS

3.1. Parkinson's Disease Database

This study uses a new Medical Dataset[1] related to Parkin-son's Disease that is currently under development as new data are constantly added to it. This Dataset is being constructed based on collaboration of the Intelligent Systems Research Group, National Technical University of Athens, with the Department of Neurology, Georgios Gennimatas General

Hospital, Athens, Greece [7]. To the best of our knowledge this is the first publicly available dataset of this type. Consequently, researchers will have the possibility of using it for the development of systems which will learn to make predictions and assist medical doctors in early detection of Parkinson's Disease. The dataset is currently populated and includes:

a) Magnetic Resonance Images (MRI) of the brain,
b) Images obtained through scintigraphy with 123-ioflupane (DaT Scans),
c) Epidemiological data such as patient's current age, sex and disease duration,
d) Treatment data in terms of duration of dopaminergic treatment, dose
e) Clinical data relating to patient status to several scales (Unified Parkinsons Disease Rating Scale (UPDRS) [4], PDQ-39, etc.) that reflect the patient's mobility, everyday activities, therapy complications and quality of life.

At the time being, the dataset is composed of 55 patients with Parkinson's and 23 subjects with Parkinson's related syndromes, including subjects' MRI, DaT Scans and clinical data. Due to the fact that at the moment the size of the dataset is not sufficient to train complex DNN architectures, we applied several augmentations techniques to improve the generalization capability of the system [10, 11]. The python library "imgaug" was used to augment the brain images[2].

3.2. Multi-Task Learning with Neural Networks

Single Task Learning (STL) NN: This task involves the use of a single type of data as an input for the system. In the current framework the main input is the DaT Scan images, as they have been found to carry the most information, out of all the available data. The images are fed to a DNN follow-ing the 50-layer ResNet architecture [13]. Transfer learning is performed, as the model's weights are initialized from a sim-ilar architecture, previously trained on the ImageNet dataset

- This network has proven to be capable of high-level feature extraction from the input images and can be applied to multiple fields. Two Fully Connected (FC) layers, of 128 neurons each, were added on top of the ResNet followed by the output layer. The second FC layer is used for the MTL in the case of auxiliary input data. Our STL architecture is depicted in Figure 1.

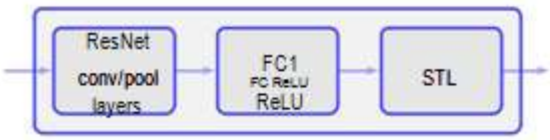

Fig. 1: Our Single-Task Learning (STL) model architecture.

The input consists exclusively of DaT Scan images, whichare fed to a CNN for feature extraction (in our case a 50-layer ResNet). The output of the model is, in turn, passed into a STL module that outputs the final prediction.

Multi-Task Learning (MTL) NN: The MTL architecture is fairly similar to the STL one. Two auxiliary inputs were also added to the system corresponding to the age and sex of each Subject. These two inputs were directly fed to the MTL layer of the network, where we concatenated them with the DaT Scans' extracted representations. The layers are the same size with the ones in the STL model. The architecture of the proposed system is illustrated in Figure 2.

To quantify and compare the performance of our four architectures on the PD database, commonly used performance measures such as Accuracy (Acc.), Precision (Prec.), Sensitivity/Recall (Sens.), Specificity (Spec.) and Area Under the Receiver Operating Characteristic Curve (AUC) [12] were estimated and are presented in Table 1 and Figure 4.

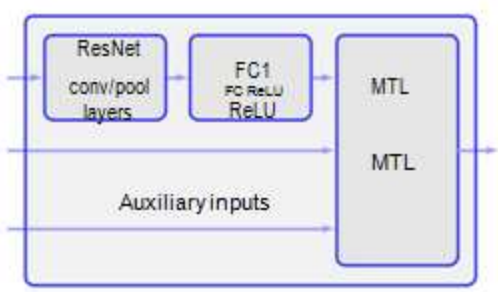

Fig. 2: A simplified version of our Multi-Task Learning (MTL) sys-tem.

The main input (in our case the DaT Scan images) is presented in the same way as in the STL system. The auxiliary data used in this experiment are all categorical, however in theory relevant data of any type can be fed to the system as well. Any preprocessing required for the auxiliary data (e.g. feature extraction) should be performed prior to them entering the MTL system. The architectures of the two systems were selected to be as similar as possible for a better comparison of the results. Essentially, the only thing that differs in the two models are the auxiliary inputs.

For the training procedure the Adam [1] optimizer was used with a base learning rate of 0.001. The experiments were run on an NVIDIA Titan Xp GPU and lasted about one day each. All our models were implemented in Keras[3] and TensorFlow [4].

IV. RESULTS AND DISCUSSION

To assess the performance of our models (both STL and MTLs) and to select the most appropriate architecture for the early prediction of the Parkinson's Disease, we carried out four separate sets of experiments on the PD database. The first set solely relied on the DaT Scan images to make the predictions (i.e. STL) and serves as the system's baseline. The remaining employed MTL with different examining conditions of the auxiliary inputs (age, sex or both). Testing was

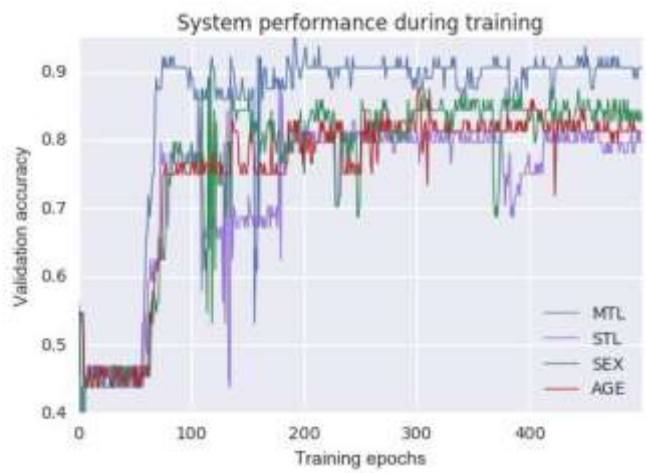

Fig. 3: Comparison of the validation accuracy achieved by the different systems.

Our results, illustrated in Figure 3, show that the system that employs all available information outperforms the rest, leading to a 91% accuracy, while the worst system seems to be the STL architecture reaching an accuracy percentage of 77%. This result is in concordance to previous studies such as [6]. The wide difference between the MTL and STL models can be explained in part by the increased feature set size; MTL training may, in this case, provided a form of regularization that STL cannot exploit.

Additionally, an important observation, is that all four systems register a sensitivity (recall) score of 100%, as a result of the lack of False Negative predictions on the test set. This is paramount for systems aiming at Computer-Aided Diagnosis, since these types of errors could have severe consequences [14].

Table 1: Top MTL systems' performance metrics.

Model	Acc.	Prec.	Sens.	Spec.	AUC
MTL	0.91	0.83	1.0	0.83	0.92
SEX	0.83	0.70	1.0	0.64	0.85
AGE	0.80	0.70	1.0	0.64	0.82
DAT	0.77	0.67	1.0	0.58	0.80

As far as the the new Medical Image Database with re-spect to the Parkinson's disease, even though it is in its con-struction phase, nevertheless, our preliminary experimental results of 91%, 83%, 80% and 77% accuracy for the four

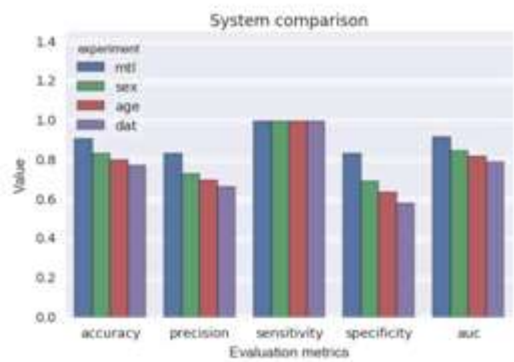

Fig. 4: Comparison of the performance achieved by our four pro-posed model systems. The model that benefits from all available inputs is denoted as 'mtl'. The ones that use only the age/sex vari-ables as auxiliary inputs are denoted as 'age'/'sex' respectively. The STL model that relies solely on DaT scan images is denoted as 'dat'.model architectures, one MTL with two auxiliary tasks of age and sex, two MTL systems with having the subject's sex and age as auxiliary tasks respectively and one STL model that re-lies solely on DaT scan images, confirm the potential of using our proposed approach for predicting PD.

V. CONCLUSIONS AND FUTURE WORK

We are the first to demonstrate how to use the patients' data (DaT Scans, sex and age) jointly to improve the prediction of mental health. We model the different conditions as tasks in a Multi-Task Learning framework. The main task of our model is the PD and non-PD prediction, while the patients' age and sex serve as auxiliary tasks. Finally, all three combinations of MTL frameworks we proposed significantly outperform the Single-Task model (baseline) for predicting Parkinson's dis-ease. Apart from that, within the three MTL models, the 'mtl' outperforms both the 'age' and 'sex' models, reaching per-centage accuracies of 91% over 83% and 80% respectively.

Considering that the experimental setup is flexible enough, future work will explore the extension of this model to further factors than the ones shown here, such as disease duration and treatment data (i.e. treatment duration and dose). At last, our intention is also to explore the effect of auxiliary-task se-lection on model performance for a given prediction task. Similar to [18], we expect to find that choosing auxiliary tasks, which are prerequisites or related to the prediction of PD disease task, is critical for learning a strong prediction model when developing treatments that can delay, prevent or reverse disease progress.

Our research is ongoing and will be extended to cover all cases met in Parkinson's Dataset. It is expected that the gen-erated systems will provide significant support to clinicians and medical doctors for early detection of PD.

REFERENCES

[1] Kingma, Diederik P and Ba, Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980,2014.

[2] Mahlon R DeLong, "Primate models of movement dis-orders of basal ganglia origin," Trends in neurosciences, vol. 13, no. 7, pp. 281–285, 1990.

[3] Eugenia Mamikonyan, Paul J Moberg, Andrew Siderowf, John E Duda, Tom Ten Have, Howard I Hur-tig, Matthew B Stern, and Daniel Weintraub, "Mild cognitive impairment is common in parkinson's dis-ease patients with normal mini-mental state examination (mmse) scores," Parkinsonism & related disorders, vol. 15, no. 3, pp. 226–231, 2009.

[4] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease et al., "The unified parkinson's disease rating scale (updrs): status and rec-ommendations.," Movement disorders: official journal of the Movement Disorder Society, vol. 18, no. 7, pp. 738, 2003.

[5] Rich Caruana, Multitask Learning, Ph.D. thesis, Carnegie Mellon University Pittsburgh, PA, 1997.

[6] Jonathan Baxter et al., "A model of inductive bias learn-ing," Journal of Artificial Intelligence Resources(JAIR), vol. 12, no. 149-198, pp. 3, 2000.

[7] Athanasios Tagaris, Dimitrios Kollias, and Andreas Stafylopatis, "Assessment of parkinsons disease based on deep neural networks," in International Confer-ence on Engineering Applications of Neural Networks (EANN). Springer, 2017, pp. 391–403.

[8] Saba Emrani, Anya McGuirk, and Wei Xiao, "Prog-nosis and diagnosis of parkinson's disease using multi-task learning," in Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining (SIGKDD). ACM, 2017, pp. 1457–1466.

[9] R Caruna, “Multitask learning: A knowledge-based source of inductive bias,” in Proceedings of the Tenth International in Machine Learning Conference, 1993, pp. 41–48.

[10] Alexey Dosovitskiy, Jost Tobias Springenberg, and Thomas Brox, “Unsupervised feature learning by augmenting single images,”arXiv preprint arXiv:1312.5242, 2013.

[11] Ren Wu, Shengen Yan, Yi Shan, Qingqing Dang, and Gang Sun, “Deep image: Scaling up image recogni-tion,” arXiv preprint arXiv:1501.02876, vol. 7, no. 8, 2015.

[12] Marina Sokolova and Guy Lapalme, “A systematic anal-ysis of performance measures for classification tasks,” Information Processing & Management, vol. 45, no. 4,427–437, 2009.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vi-sion and Pattern Recognition (CVPR), 2016, pp. 770–778.

[14] Kaplan, Robert M and Bush, James W, “Health-related quality of life measurement for evaluation research and policy analysis,” in Health Psychology, vol. 1, no. 1, 1982.

[15] Deng, Jia and Dong, Wei and Socher, “Imagenet: A large-scale hierarchical image database,” IEEE Con-ference on Computer Vision and Pattern Recognition, (CVPR), 248–255, 2009, IEEE.

[16] Er, Orhan and Cetin, Onursal and Bascil, M Serdar and Temurtas, Feyzullah, “A Comparative study on parkin-son’s disease diagnosis using neural networks and artifi-cial immune system,” Journal of Medical Imaging and Health Informatics, vol. 6, no. 1, 264–268, 2016, Amer-ican Scientific Publishers.

[17] Gil, David and Manuel, Devadoss Johnson, “Diagnos-ing parkinson by using artificial neural networks and support vector machines,” Global Journal of Computer Science and Technology, vol.9, no. 4,2009.

[18] Rich Caruana, Shumeet Baluja, and Tom Mitchell, “Us-ing the future to “sort out” the present: Rankprop and multitask learning for medical risk evaluation,” in Ad-vances in neural information processing systems, 1996,959–965.