Early Detection & Stage Classification of Parkinson's Disease using Deep Learning

Author

Muhammad Muzzamil Zeeshan

Reg No. 359590

Supervised by:

Dr. Kashif Javed

DEPARTMENT BMES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

SEPTEMBER 2023

Early Detection & Stage Classification of Parkinson's Disease using Deep Learning

Author

MUHAMMAD MUZZAMIL ZEESHAN

Regn Number

359590

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Biomedical Engineering

Thesis Supervisor:

DR. KASHIF JAVED

Thesis Supervisor's Signature: _____ _____

DEPARTMENT BMES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

SEPTEMBER, 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Regn No. 00000359590 Muhammad Muzzamil Zeeshan** of **School of Mechanical & Manufacturing Engineering (SMME) (SMME)** has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis titled. **Early detection and stage classification of Parkinson's disease using deep learning.**

Signature: 

Name (Supervisor): Kashif Javed

Date: 01 - Sep - 2023

Signature (HOD): 

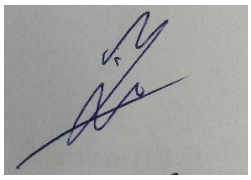
Date: 01 - Sep - 2023

Signature (DEAN): 

Date: 01 - Sep - 2023

Declaration

I certify that this research work titled "*Early Detection & Stage Classification of Parkinson's Disease using Deep Learning"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.



Signature of Student

Muhammad Muzzamil Zeeshan

Regn No. 359590

MS Biomedical Engineering

Proposed Certificate for Plagiarism

It is certified that MS Thesis Titled **Early detection and stage classification of Parkinson's disease using deep learning** by Muhammad Muzzamil Zeeshan has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b. The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c. There is no fabrication of data or results which have been compiled/analyzed.
d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

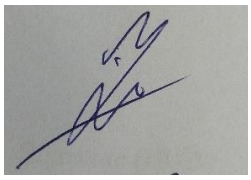
Name & Signature of Supervisor

Dr. Kashif Javed

Signature :

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.



Signature of Student

Muhammad Muzzamil Zeeshan

Registration Number

359590

Signature of Supervisor

Copyright Statement

- Copyright in text of this thesis rests with the student author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of NUST School of Mechanical & Manufacturing Engineering (SMME). Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights which may be described in this thesis is vested in NUST School of Mechanical & Manufacturing Engineering, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the SMME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST School of Mechanical & Manufacturing Engineering, Islamabad.

Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which You setup in my mind to improve it. Indeed, I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual, was due to Your will, so indeed none be worthy of praise but You.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor Dr. Kashif Javed for his help throughout my thesis. For guiding me, despite having occupied with lots of responsibilities.

I would also like to pay special thanks to my brother, Muhammad Osama Zeeshan for his tremendous support and cooperation. Each time I got stuck on something, he came up with the solution. Without his help, I would not have been able to complete my thesis. I appreciate his honesty, patience, and guidance throughout the whole thesis.

I would also like to thank Dr. Asim Waris, Dr. Amir Kashif and Dr. Adeeb Shehzad for being on my thesis guidance and evaluation committee and express my special thanks to Dr. Syed Omer Gilani for his help. I am also thankful to Dr. Mahira Zeeshan, Dr. Aneeqa Noor and Dr. Nosheen Fatima Rana for their support and cooperation.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Parkinson's disease (PD) is caused by a lack of dopamine production by the substantia nigra in the brain. It is an enduring disorder without any cure, making it a burden on the patient and the society. PD is a complex disorder marked by many physical and non-physical manifestations, which differ for everyone. Clinicians might misdiagnose, waste time and resources to get a patient's diagnosis or do not have enough expertise to diagnose a patient. Deep learning models tend to overfit with new data; thus, to prevent variance in the model, merging outputs has been proven effective. This study proposes an ensemble deep learning model, to automate PD detection and stage classification, which can handle different data by combining rules. The ensemble model (TransConvNet) links two state-of-the art deep learning models in decision level ensembling. The outputs are fused using averaging voting. Both neural networks utilize gait data provided by Physionet. The validation accuracy for PD detection reached 82%, while for PD stage classification, it reached 73%. This model delivers competitive and top-notch performance for severity and detection prediction for PD using gait. This can be used as a tool for PD detection or monitoring its development. Future work might include addition of models for better performance and reducing the training time of the models.

Key Words: *Parkinson's disease, deep learning, ensemble deep learning, transformer, 1D-ConvNet*

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disease, described as 'shaking palsy' by Dr. James Parkinson in 1817. It is characterized by both motor and non-motor function loss, manifested by variable degrees. Motor dysfunction is attributed to damage to nigral dopaminergic neurons and other regions of neurons [1] shown in **Figure 1-1**. The damage resulted in lesser production of dopamine (DA) leading to impaired movement issues. The main motor biomarkers of PD include tremor, gait freezing, rigidity, akinesia (bradykinesia), flexed and instable posture [2]. The patients may suffer from mental and behavioral changes, memory issues, mood swings, depression, and fatigue. PD patients have developed Parkinsonian gait attributed to a lean posture, and they take small quick steps. In this paper gait data is investigated to distinguish healthy from PD patients and their stages of severity.

People with PD often develop a parkinsonian gait that includes a tendency to lean forward; take small, quick steps; and reduce swinging their arms. They also may have trouble initiating or continuing movement.

Symptoms often begin on one side of the body or on one limb on one side of the body. As the disease progresses, it eventually affects both sides. However, the symptoms may still be more severe on one side than on the other.

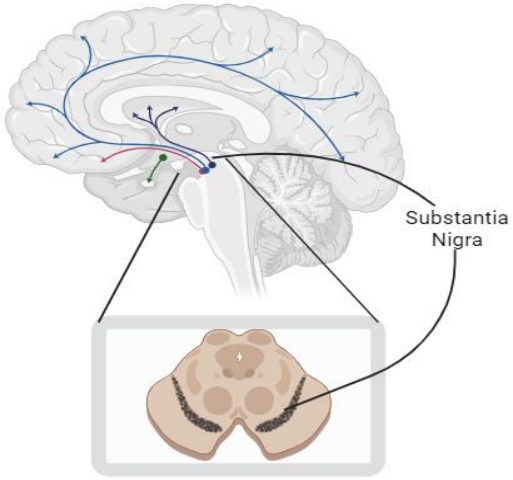


Figure 1-1: Nigrostriatal pathway (black) and cross-section of midbrain showing substantia nigra

1.1 Diagnosis

The diagnosis of PD remains highly subjective according to clinicians. The diagnosis of PD is either based on genetic testing or postmortem detection of neuropathological changes in the brain which means it is evaluated on stages of family history, motor function, atypical parkinsonism symptoms and levodopa response [3]. It requires a lot of time and effort for doctors to diagnose PD because it might overlap with other disease symptoms e.g., Alzheimer's etc. Several rating scales are used for PD evaluation, two of them include Hoehn & Yahr (H&Y) & Unified Parkinson's disease Rating Scale (UPDRS). H&Y scale ranges from 0 (healthy) to 5 (advanced PD) with later additions of 1.5 and 2.5 for intermediate stages.

1.2 Rationale

Since being a progressive disease with no confirmation even if the symptoms worsen or improve like resting tremors [4], there needs to be a way to determine whether a patient has PD and its severity. Therefore, one of the symptoms that shows the most fluctuations in early to high disease progression is gait disturbance [5]. Gait evaluation is trivial for physicians to indicate for accurate PD diagnosis because it is affected by many factors. Quantifying gait and evaluating its parameters can be useful inputs to machine learning and/or deep learning models to evaluate PD. Moreover, ensemble learning is applied, to see effects of different deep learning models, on gait dataset. For a model, this would Enhance accuracy and prevent overfitting.

1.2.1 Problem Statement

Currently, there is no reported technique that identifies PD in gait through ensemble deep learning, as per our knowledge. The deep learning methods employed in PD detection are more likely to overfit even if their given accuracies are high because they are sensitive to minute changes in input. The model may lose the original knowledge on trained data when adapting to a new dataset. Hence, retraining of deep models from the same dataset is required and an ensemble learning technique would be useful for reducing variance [6]. Thus, can be used to detect PD gait and its stages.

1.2.2 Aims & Objectives

Aim: To develop an advanced automation technique for detection and severity prediction of PD.

Objectives:

1. To develop up-to-date deep learning model for detection & monitoring.
2. To ensemble multiple deep learning networks to improve performance [6].
3. To build reliable and accurate ensemble deep learning model.

CHAPTER 2: LITERATURE REVIEW

2.1 Background

PD is a debilitating progressive disorder without any cure. After Alzheimer's disease, it is the second most common neurological disorder affecting around 10 million people [7]. Elder people are most likely to suffer from its effects to their Quality of life (QOL), making life burdensome to themselves, to their family and to the society [5]. Besides that, it needs highly specialized care for each patient as it varies differently for them [6, 7]. An early diagnosis is therefore necessary. For severity prediction, UPDRS scale is used containing four parts for comprehensive diagnosis – non-motor experiences of daily living (nM-EDL), motor experiences of daily living (M-EDL), motor examination and motor complication [9].

2.1.1 Structure

People with PD have a particularly altered Gait and posture. The parkinsonian gait is marked by smaller stride length, smaller steps, longer stance, freezing of gait and flat foot, which is usually due to faulty subcortical control mechanism. In this paper, gait data taken from PD in gait [10] dataset is applied. Thus, the studies analyzed, compare methods to obtain good performance of PD diagnosis and severity prediction. The algorithms in the studies are either based on machine learning, deep neural networks, or ensemble learning. Machine learning (ML) methods use handcrafted features i.e., spatiotemporal, kinematic, or kinetic whereas Deep learning (DL) networks learn features automatically through various filtering and layering techniques. Sections 2.2 and 2.3 show papers with deep learning i.e., utilizing 1D-CNN and transformers, section 2.4 explains ensemble and its related papers, section 2.5 shows articles using ML in related work and section 2.6 concludes.

2.2 1D-Convnet

Basic principle of 1D convolutional neural network (1D-Convnet) involves input sequence which is modified by a kernel to give modified output. Its main components include convolution, dropout, max pooling, flatten, fully connected, and softmax layer. A fully connected layer is combination of an affine function and non-linear function. 1D-Covnet is mainly used for one-dimensional time series data.

El Maachi et al. [7], employs specific 1D-CNN to process 18 one-dimensional signals from each of the foot sensors from [10] dataset. The algorithm measures PD detection and severity (using UPDRS). The first part of the network processes individual inputs in 1D-ConvNets while the second part processes concatenated vectors in fully connected layers to give prediction. It achieved 98.7% accuracy for PD detection and 85.3% accuracy for UPDRS severity prediction which is state of the art, considering its application on PD in gait dataset.

Chenhui et al. [11] proposed static-dynamic neural networks using gait analysis [10] for PD diagnosis and severity. The static pathway uses 1D-CNN for each sensor's one-dimensional vGRF and the dynamic pathway uses 2D-CNN to capture motion transfer as a 2D image between 16 sensors through time. The accuracies are 96.7% for PD diagnosis and 92.3% for PD severity prediction.

2.3 Transformers

Transformers are introduced in the paper “Attention is all you need” to replace recurrent neural networks [12]. These are the neural networks which contain an encoder and a decoder each with 6 identical layers. Each of these layers contains two sublayers for encoder and three sublayers for decoder. The encoder sublayers consist of Multi-Head Attention and feed-forward network. The decoder contains the sublayers Masked Multi-Head Attention, Multi-Head Attention (receives inputs from encoder) and feed-forward network, each with their own residual connection – input adds to output – and normalization. The inputs to encoder and decoder are tokenized in the embeddings and then they are encoded by position encodings. The transformers are used to analyze sequence of words for language translation. The decoder shifts input 1 position to the right to process words from encoder, and processes sequences from the left to predict output.

Nguyen et al. [8] analyzed 1D signals from [10] for transformer encoders. The data preprocessing for signals is the same as the one used in [7] in the code available, using the same dataset. The time-series data is run through temporal transformers for each 18 one-dimensional signals, followed by a fully connected layer (FC-0) and then concatenated to spatial transformer layer. For classification, the classifier has two fully connected layers and an output layer. This paper only focused on PD detection for which the accuracy reached 95.2%. This is a state-of-the-art result for transformer model in PD in gait dataset.

2.4 Ensemble Learning

Ensemble learning model (**Figure 2-1**) fuses multiple models to obtain a powerful model. They combine many different baseline models in multiple techniques i.e., averaging, boosting, bagging, stacking and random forest. This nature of model reduces overfitting and exceeding in performance from many single models; hence it is applicable in many fields [6]. For deep learning as baseline models, only average voting is tested in papers, which concludes bad performance from weak learners. Its basic structure [6] consists of dataset which baseline classifiers use for predictions followed by aggregation G function on which baseline classifiers $(c_1, c_2 \ldots c_k)$ combine to give a single output. Shown in equation 1:

$$y = \Phi(x_i) = G(c_1, c_2 \ldots c_k) \tag{2.1}$$

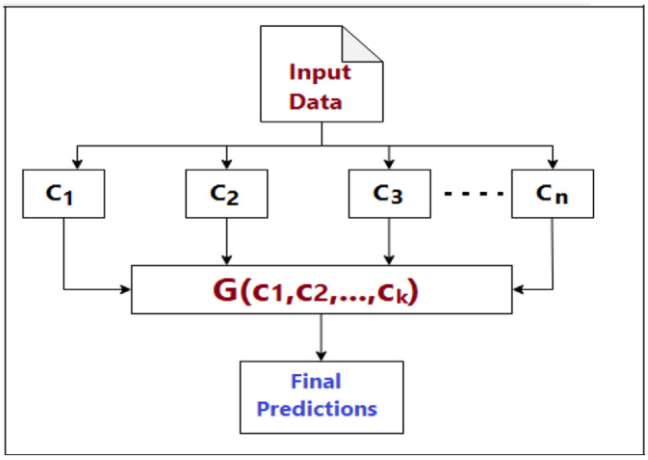


Figure 2-1: Ensemble learning architecture

Three factors can affect its performance: 1) the arrangement of baseline classifiers can be either parallel (above) or sequential. 2) processes for combining output in aggregation function include weight voting or meta-learning methods. 3) heterogeneity or homogeneity of baseline classifiers i.e., difference or sameness.

For ensemble deep learning, to reduce variance 4 strategies are proposed [6]:

A. Applying many different basic models with a single data.
B. Applying many different same basic models with a single data.
C. Applying many different basic models with different data samples.
D. Applying many different same basic models with different data samples.

Furthermore, ensemble deep learning fusion methods can be either feature level or decision level ensembling. Feature level ensembling involves concatenation of features from various models for final prediction whereas decision level ensembling involves normalization of predictions from different classifiers for final prediction.

Decision level ensembling shows promising results in [13] for skin cancer lesion detection combining baseline deep learners VGG, CapsNet and ResNet having 93.5% accuracy. It is very simple to implement and reduces overfitting.

2.5 Related Work

2.5.1 Articles with PD in Gait Dataset

Other papers utilizing PD in gait dataset with confident results include:

Nicholas et al. [14] used vertical ground reaction force (vGRF) data [10] for diagnosing PD, the features are selected through wrapper approach with RF. These were inputs to supervised algorithms i.e., K-nearest neighbour (KNN), classification and regression tree (CART), random forest (RF), support vector machines (SVM) and Naïve Bayes (NB), and unsupervised algorithms i.e., K-means and Gaussian mixture model (GMM). Additional dataset with patients having ALS and HD is used for validation. They reached an accuracy of 86.05% for Yogev et al. [15] sub-dataset, 90.91% for Hausdorff et al. [16] sub-dataset and 82.81% for Frenkel-Toledo et al. [17] sub-dataset.

Rami et al. [18] used the database [10] to classify PD patients on load distribution during gait. They classified them having balanced or unbalanced gaits using centre of pressure (COP). Then they further distinguish them into balanced-normal and balanced-diseased subjects using linear decision boundary. 95% accuracy is achieved.

2.5.2 Articles with Gait Data Investigation

These papers perform physical experiments in their study to subjects to obtain gait data:

Milica et al. [19] utilized data from 40 mild PD patients – H&Y 1 to 1.5 - and healthy subjects from three different conditions: simple walking, motor task and mental task. For feature selection, affinity propagation clustering was applied. The features are further reduced using random forest feature importance and transferred to SVM classifier. Final parameters were stride length with its coefficient of variation (CV), stride time, stride time (CV), swing time, swing

time (CV), heel-to-heel base support CV and step time asymmetry. The accuracy reached 85% for PD detection.

Carlo et al. [20] purpose was to detect PD patients with or without Mild Cognitive Impairment (MCI). The PD patients went through three different conditions of gait as in Milica et al. study, features were compared through statistical analysis and given to DT, RF and KNN classifiers. DT involving motor dual-task conditions achieved highest accuracy 86.8%.

2.6 Conclusion

The papers using ML models do not display good accuracies, that is from 85-95% for every type of classification involving PD patients. They require additional work to extract handcrafted features [18, 20], while they require more algorithms to filter out features to reach better to average accuracies [14, 19]. In contrast, deep learning models obtained better accuracies for PD detection, ranging from 95.2 to 98.7% and almost the same for PD severity ranging from 85.3 to 92.3%. These models do not need to specify which features to extract, because they only need specific inputs to form connections between them i.e., temporal transformers getting inputs from 18 sensors to get temporal features [8], making handcrafted features limiting and unnecessary. Since, ensembling learner gives good accuracy combining several deep learning models with decision level ensembling [13], and for proof of concept, this simple ensembling would be used to combine bigger deep learning models. These models are [6, 7], 1D-Covnet and transformer.

2.6.1 Gaps

1D-ConvNets identifies patterns in data regardless of their position, unlike the need of positional encoding in transformers. Transformers mostly captures relation between the tokens (elements) and they can capture long-range dependencies by attention head symmetry which does not happen in 1D-ConvNets. Both the models can be combined to leverage both of their benefits using decision level ensembling – which would prevent overfitting and reduce load time compared to feature level ensembling – this would provide robust classification from the same dataset.

CHAPTER 3: METHODOLOGY

From literature [6], we used ensemble deep learning technique in our methodology, involving two neural networks. In this technique, strategy A is selected out of four strategies, which involves training different baseline models with single dataset. This is due to higher performance, and different state-of-the-art deep learning models are combined in ensemble learning, in contrast to weak ML learners.

3.1 Method

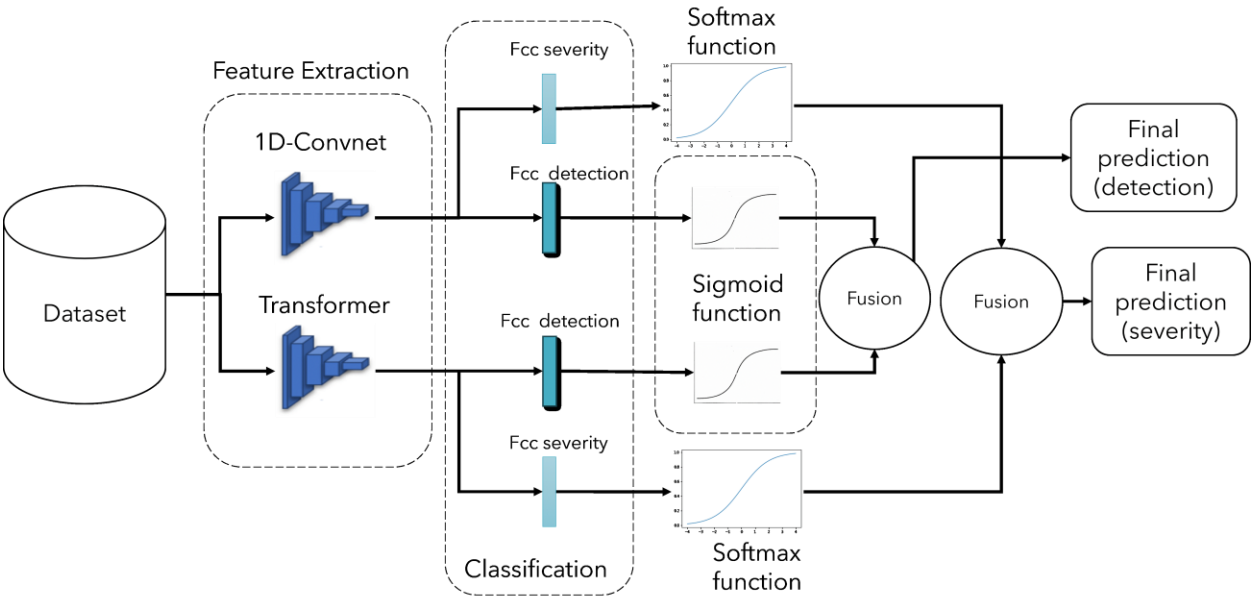


Figure 3-1: Method Diagram for TransConvNet model

The dataset containing time-series data, sends force signal as function of time. These are 16 vGRF signals with two total vGRF, for right and left foot. These are preprocessed into vectors with 100 elements having 50% overlap and sent to both feature extractors: transformer and 1D-Convnet.

In 1D-ConvNet feature extractor, for a single sensor, the vector passes through 4 convolutional layers, and two max-pooling layers before flattening towards output layer. From each of 18 1D-ConvNet feature maps are concatenated and sent to two parallel fully connected layers as shown in **Figure 3-1** above. They give detection probabilities with one neuron and severity probabilities with five neurons (according to PD rating scales) after softmax and sigmoid function.

In transformer feature extractor, temporal transformer encoder extracts temporal features, shortens - through FC-0 layer - and concatenates vectors from 18 encoders. It sends them to the spatial transformer encoder, which extracts spatial features – to capture relations between the sensors i.e., motion flow – and then passes the feature vector to fully connected layers as shown in **Figure 3-1** above. They give probabilities for detection with one neuron and for severity with five neurons, from softmax and sigmoid function.

The arithmetic mean of probability of each class is taken through average voting, and gives output ŷ, to the probability with highest value. It is shown by following equation:

$$\hat{y} = \text{argmax}_i \frac{1}{m} \sum_{j=1}^{m} P_{ij} \tag{3.1}$$

Where ŷ is predicted value, P_{ij} is probability of i^{th} class of j^{th} classifier and m is total number of classifiers.

For PD detection, with i=2 classes, it would give:

$$\hat{y} = \text{argmax}_i(\hat{y}_0, \hat{y}_1) \tag{3.2}$$

For PD severity prediction, with i=5 classes, it would give:

$$\hat{y} = \text{argmax}_i(\hat{y}_0, \hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4) \tag{3.3}$$

To reduce loss in backpropagation, cross entropy loss is applied after each iteration, the equation is:

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^{N} y_i \times \log f_\theta(x_i) \tag{3.4}$$

Where L_{ce} is cross entropy loss, f_θ is feature extractor and N is total number of examples.

3.1.1 Performance Metrics

Training and validation accuracy is used to check the performance of models. The equation for accuracy is:

$$Accuracy = \frac{correct\ number\ of\ predictions}{total\ number\ of\ predictions} \tag{3.5}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.6}$$

CHAPTER 4: EXPERIMENTS & RESULTS

4.1 Database

PD in gait [10] is an open access database from Physionet repository. It consists of gait data of 166 individuals, combined from the studies of Yogev et al., Hausdorff et al., and Frenkel-Toledo et al. For gait analysis, forces sensors are placed inside their shoes (Computer Dyno Graphy (CDG) ® system Infotronic, Netherlands) for both feet. The sensors are placed according to cartesian axis with the origin at center of both feet shown in **Figure 4-1** the below:

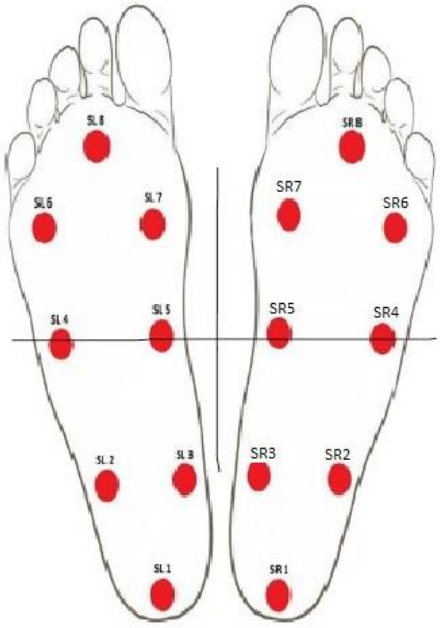


Figure 4-1: Sensor position, SR represents right foot sensors and SL represents left foot sensors.

The database contains gait data from 93 PD patients and 73 healthy controls with their mean age 66.3 yrs. Database contains vGRF (in Newtons) as function of time, as patients walk with their normal pace from 25 to 77 m for 2-5 minutes. The 16 sensors output was converted into digital values by sampling at frequency of 100 Hz. The sum of the first and the last 8 sensors' signals

were also included. Yogev et al. studied dual-task conditions during gait for PD and healthy controls. Hausdorff et al. studied stimulation of walking with or without rhythmic auditory stimulation (RAS) for PD and healthy controls. Frenkel-Toledo et al. studied walking on a treadmill with or without wheeled walker for PD and healthy controls.

For each subject, vGRF data is stored inside a text file. Title of each file is displayed as:

SiPt20_01, GaCo15_10, JuPt09_03 …

a. The first two texts in the title of the file indicate type of study i.e., Ga for Galit Yogev et al. study, Ju for Hausdorff et al. study and Si for Silvi Frenkel-Toledo et al. study.
b. The third and fourth text indicate type of subject i.e., Co for healthy control and Pt for PD patient.
c. The fifth and sixth numbers indicate the subject number in that study.
d. The type and number of walks for a subject is coded by the last two numbers where 10 indicates dual task walking – serial-7 subtraction – and numbers excluding 10 indicate normal walk.

The content of each file contains 19 columns in which:

a. 1st column indicates time (in seconds) correct to 4 decimal places.
b. 2nd to 9th columns indicate vGRF (in newtons) for left foot sensors correct to 2 decimal places.
c. 10th to 17th columns indicate vGRF (in newtons) for right foot sensors correct to 2 decimal places.
d. 18th column is sum of vGRFs from left foot.
e. 19th column is sum of vGRFs from right foot.

4.2 Initial Experimentation on Handcrafted Features

An attempt was made to extract handcrafted features from vGRF data using [21] as a guide. In this paper, statistical and spatiotemporal features were extracted for PD detection and stage classification and compared performance of ML models using confusion matrix and region of convergence (ROC).

In an attempt to extract features, different sections of algorithms were made in MATLAB; a section to extract mean vGRF of a single sensor for healthy subject and PD patient, a section to trim, organize and store vGRF time series data for each patient, and other sections to store spatiotemporal features processed from vGRF time series data.

These include temporal features such as stance time, swing time, step time, stride time, cadence, speed; spatial features calculated from subjects' speed such as stance, swing, step and stride length and ratios such as swing/stance ratio. The **Figure 4-2** and **Figure 4-3** shows a healthy subject's vGRF and spatiotemporal features:

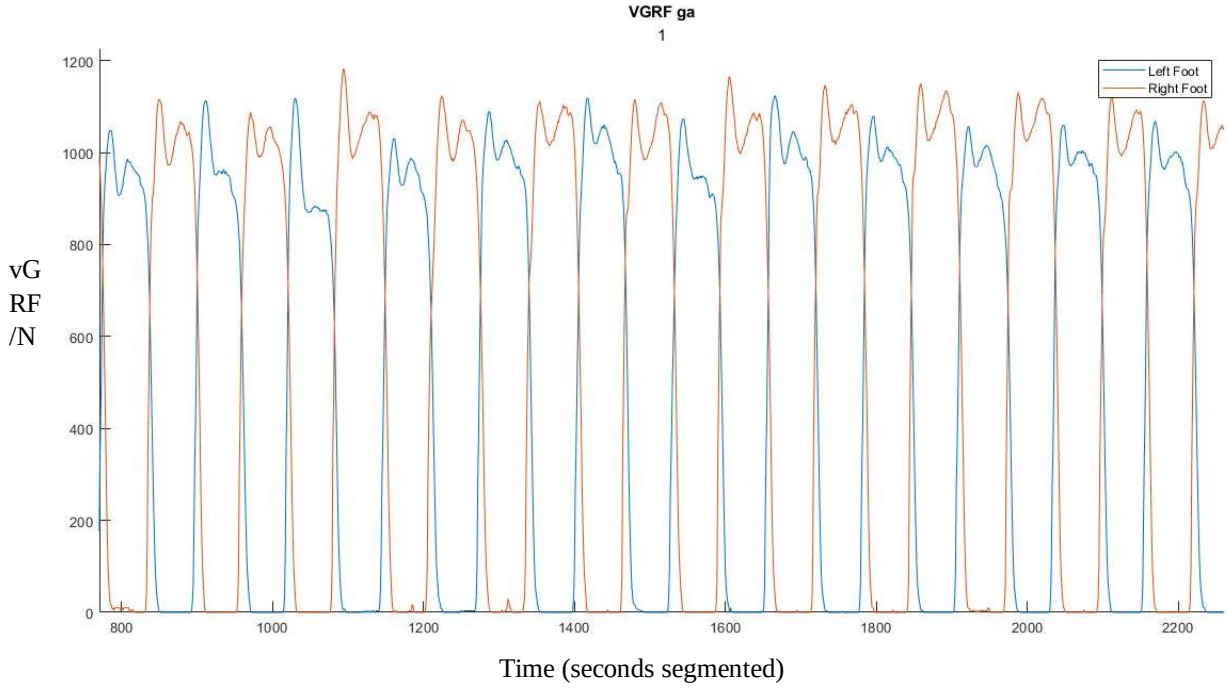


Figure 4-2: total vGRF of right and left foot plotted as function of time (segmented into numbers) for a single walk of a healthy subject

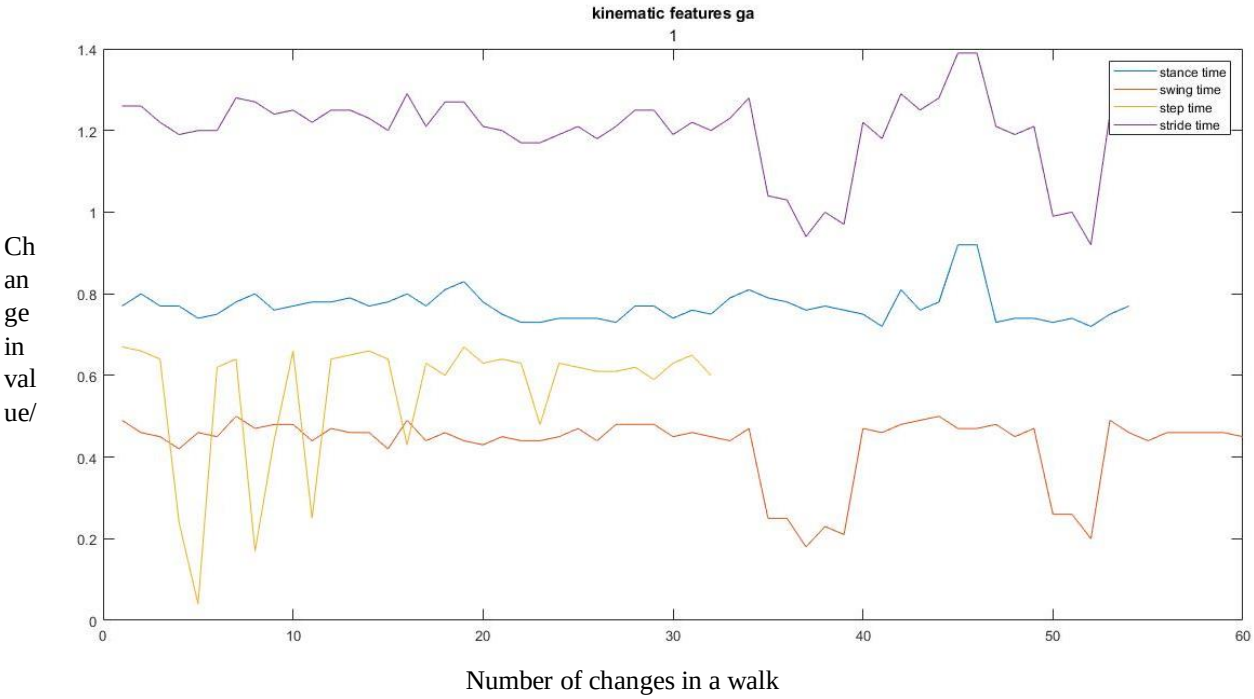


Figure 4-3: change in temporal features per number of change in a single walk of a healthy subject

4.3 Experiments on proposed ensemble based technique

4.3.1 Implementation Details

All algorithms are run on colab pro with NVIDIA T4 Tensor Core GPU on cloud. Data is divided into 64468 segments and split according to 5-fold cross validation. Each model is assigned 100 epochs initially but runs until good training accuracy is achieved, which is usually around 20 epochs for transformer and 10 epochs for 1D-Convnet, thus adjusted to 30 epochs.

Transformer model hyperparameters for each fully connected layer contains scaled exponential linear unit (Selu) activation function while for the output layer contains sigmoid activation function. The batch size is 110 and the learning is started at 0.001.

1D-Convnet model hyperparameters for their hidden layers use Selu activation function while for the output sigmoid activation is used. Nesterov Adam optimizer is used with the batch size 800 and learning rate 0.001 – which is decreased by factor of 2. Dropout layer is also applied.

4.3.2 Deep Learning

For each epoch, accuracy, validation accuracy, loss and validation loss are calculated. The transformer paper [8], did not calculate severity prediction, but it is calculated here as to compare severity scores. The validation accuracies for transformer are shown in **Table 4-1**:

Table 4-1: Validation accuracies of transformer for binary and severity prediction in 5 folds

Folds / classification	**1**	**2**	**3**	**4**	**5**	**Avg**	**Max**
Binary	0.869	0.876	0.808	0.833	0.846	**0.846**	0.876
Severity	0.71	0.71	0.74	0.64	0.79	**0.71**	0.79

The validation accuracies of 1D-Convnet are shown in **Table 4-2**:

Table 4-2: Validation accuracies of 1D-Convnet for binary and severity prediction in 5 folds

Folds / classification	**1**	**2**	**3**	**4**	**5**	**Avg**	**Max**
Binary	0.7	0.68	0.67	0.68	0.71	0.68	0.71
Severity	0.30	0.26	0.63	0.46	0.52	0.43	0.63

The **Table 4-1** and **Table 4-2**, illustrates the comparison between binary and severity prediction, by their validation accuracies, average and maximum values for a transformer and 1D-ConvNet. Each cell in the table shows the average of validation accuracies from 30 epochs in a single fold. In general, binary rows shows better classification validation accuracies in both of the tables, with the transformers leading the mark.

4.3.3 Ensemble Learning

The validation accuracies of ensemble TransCovNet are shown in **Table 4-3**:

Table 4-3: Validation accuracies of TransCovNet for binary and severity prediction in 5 folds

Folds/ classification	1	2	3	4	5	Avg	Max
Binary	0.85	0.82	0.83	0.82	0.81	**0.826**	0.85
Severity	0.82	0.74	0.68	0.70	0.71	**0.732**	0.82

Table 4-3, shows improved performance in general for both binary and severity classifications but validation accuracy for binary classification of transformer remains the best.

4.3.4 Comparison

To see trends in PD detection, training accuracy and loss for deep learning models, graphs are plotted, shown in **Figure 4-4** & **Figure 4-5**. The fold which reached the highest accuracy or loss is plotted in classification and severity graphs.

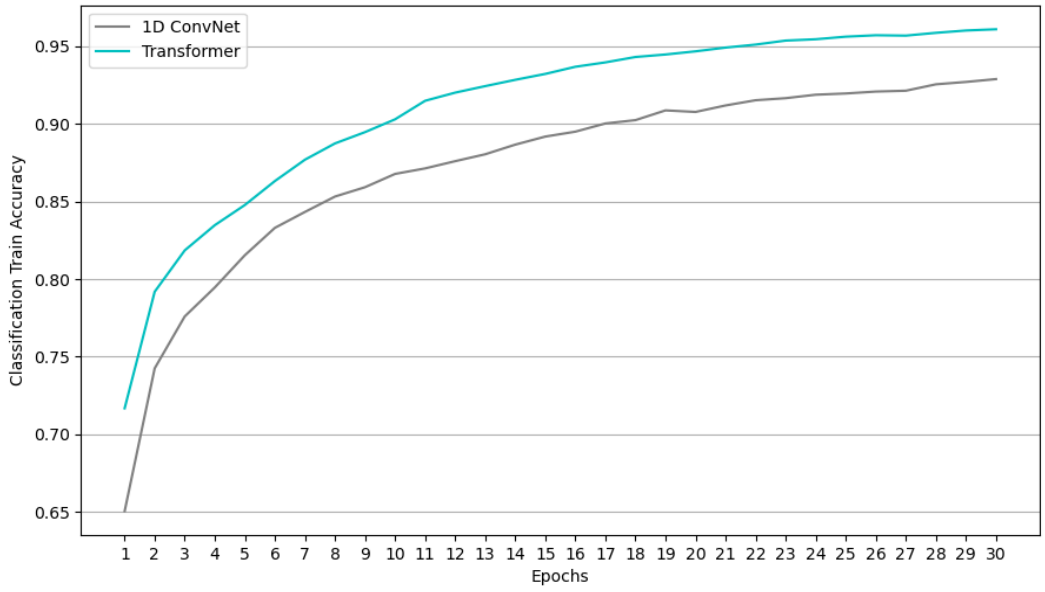


Figure 4-4: classification training accuracy per epoch for a single fold of 1D-ConvNet (grey) and transformer (blue)

A non-linear curve in **Figure 4-4**, shows training accuracy value rapidly increasing during the first four epochs till there is no increase in their gradient close to 30 epochs.

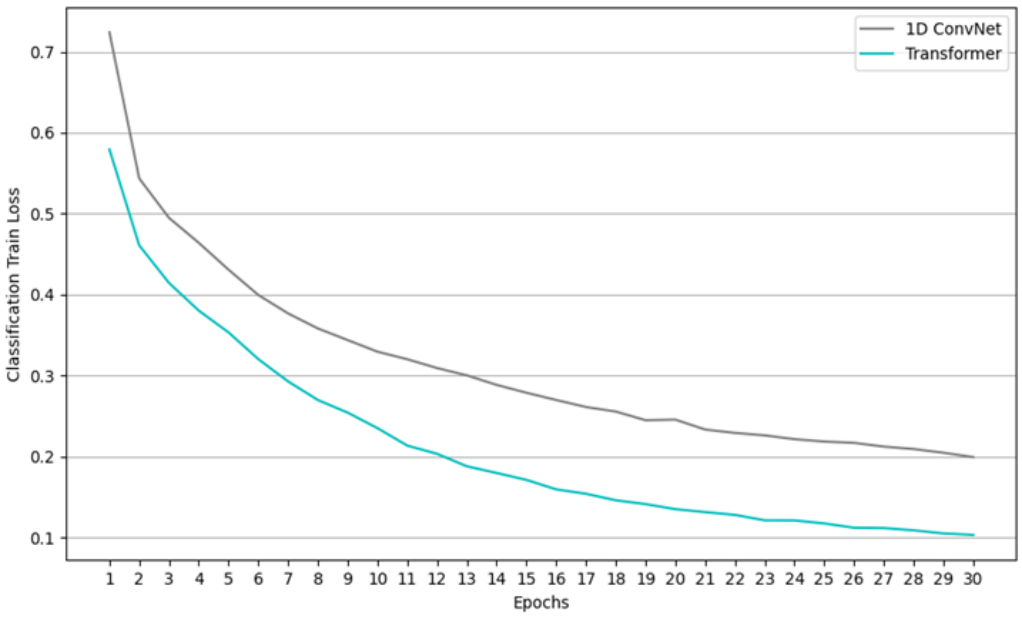


Figure 4-5: Classification training loss per epoch for a single fold of 1D-ConvNet (grey) and transformer (blue)

In contrast, with non-linear curve, training loss (**Figure 4-5**) decreases rapidly during the first few epochs till there is little to no difference in their gradient close to 30 epochs. They both display almost the same trend regarding changes in their gradients over time.

For comparison, the result of fusing 1D-ConvNet and transformer is shown in **Figure 4-6** to investigate trend of validation accuracies for TransConvNet:

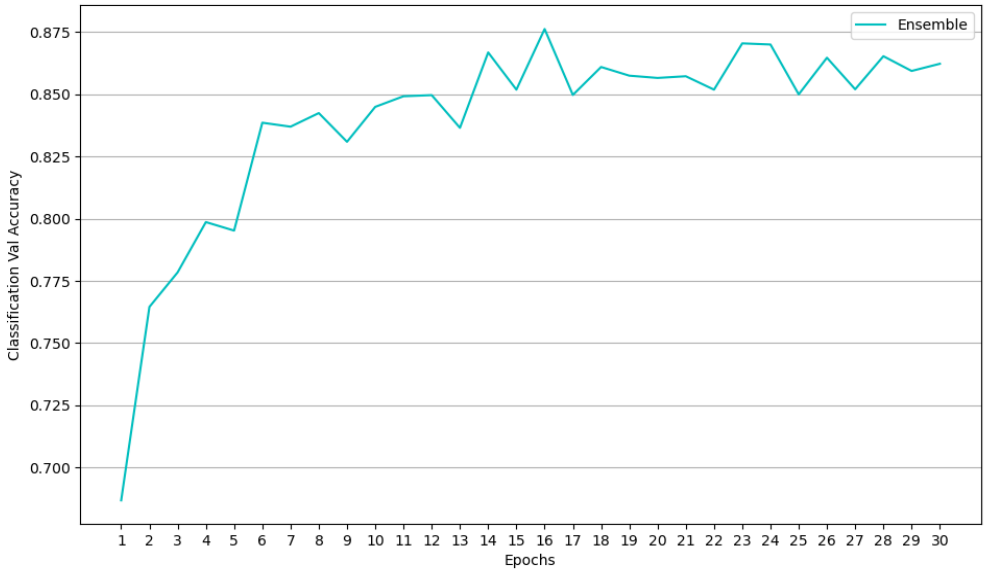


Figure 4-6: Classification validation accuracy per epoch for a single fold of TransCovNet

The main difference observed between the training accuracy and loss, is that the gap between transformer and 1D-ConvNet increases with the increase in epochs. This would explain the fluctuations in classification validation accuracy in **Figure 4-6** at later epochs.

To see trends in PD severity prediction, training accuracy and loss for deep learning models is shown in **Figure 4-7** & **Figure 4-8**:

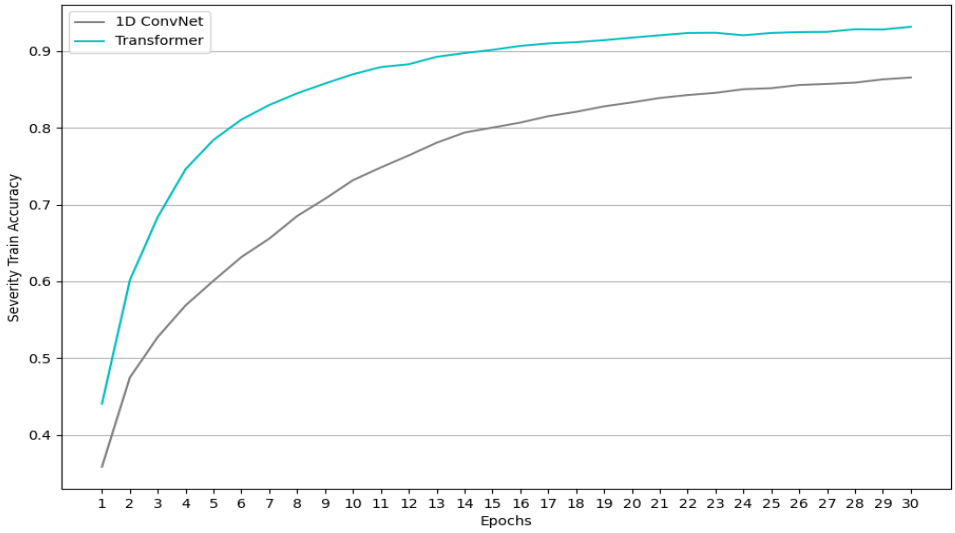


Figure 4-7: severity training accuracy per epoch for a single fold of 1D-ConvNet and transformer

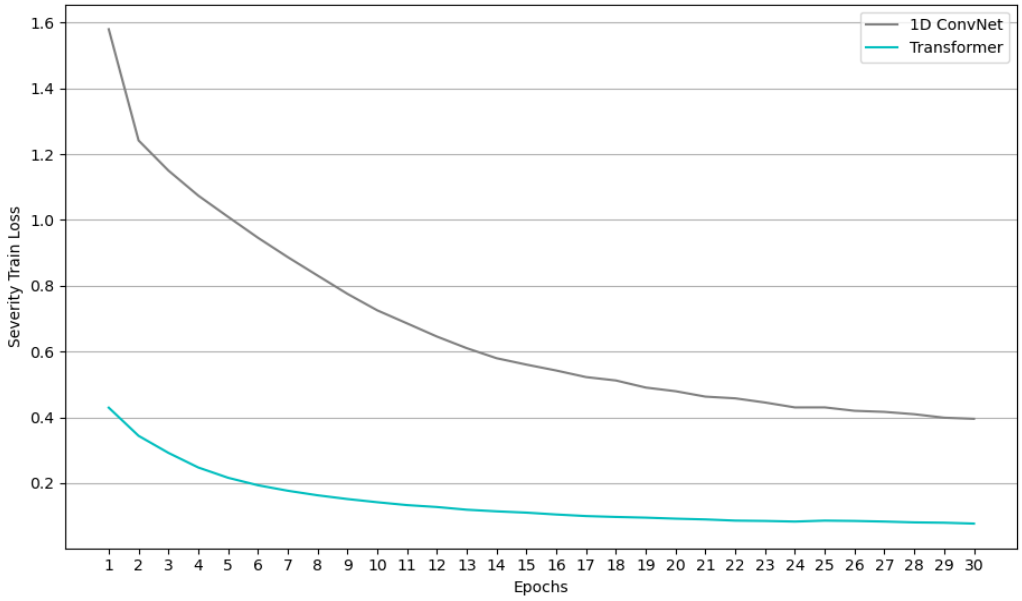


Figure 4-8: Severity training loss per epoch for a single fold of 1D-ConvNet and transformer

For comparison, the result of fusing 1D-ConvNet and transformer is shown in **Figure 4-9** for severity:

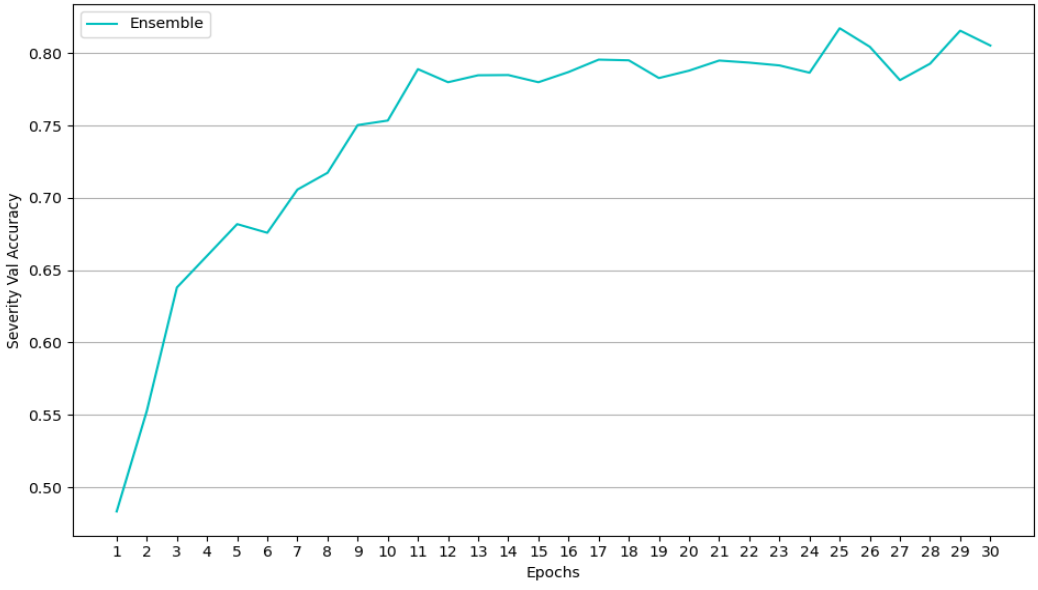


Figure 4-9: severity validation accuracy per epoch for a single fold of TransCovNet

In comparison to previous graphs, there is a significant gap between transformer's and 1D-ConvNet's curve in severity training accuracy and loss. Which reduces and stabilizes, when the

gap between them diminishes, after several epochs. That might be the reason why TransConvNet validation accuracy is more stable albeit still fluctuating.

4.3.5 Summary

From **Table 4-4**, we can see validation accuracies from three models and their comparison for detection (binary) and severity (multiclass) prediction:

Table 4-4: average validation accuracies of three models

Classifier / Models	Transformer	1D-ConvNet	Ensembling (TransConvNet)
Binary	**0.84**	0.68	0.82
Severity	0.71	0.43	**0.73**

In binary and severity classification, TransConvNet performs better than 1D-ConvNet. TransConvNet gives competitive performance as close to the transformer values as they swing very slightly from each other. TransConvNet gives promising validation accuracy for severity prediction compared to transformers and 1D-ConvNet models.

CONCLUSIONS & FUTURE WORK

To date, no reported work has used ensemble-based learning on PD in gait dataset for Parkinsonian classification. We achieved state-of-the art results for severity (with accuracy of 73%) and competitive performance for detection (with accuracy of 82%) in PD. Fusion by average voting was performed successfully by combining two deep learning models proving reliability of decision level ensembling. Thus, in the future, we will consider incorporating more ML or DL models to further improve the model performance, recall and precision. We would explore other ensemble methods e.g., feature level ensembling to reduce complexity and training time.

REFERENCES

[1] M. Romanato, F. Spolaor, D. Volpe, S. Fioretti, Z. Sawacha, and F. Di Nardo, "Recruitment of Gastrocnemius Lateralis during walking in Parkinson's Disease assessed by EMG analysis," in *2022 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2022 - Conference Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/MeMeA54994.2022.9856506.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, p. 368, Apr. 2008, doi: 10.1136/jnnp.2007.131045.

[3] M. J. Armstrong and M. S. Okun, "Diagnosis and Treatment of Parkinson Disease: A Review," *JAMA*, vol. 323, no. 6, pp. 548–560, Feb. 2020, doi: 10.1001/jama.2019.22360.

[4] B. R. Bloem, M. S. Okun, and C. Klein, "Parkinson's disease," *The Lancet*, vol. 397, no. 10291, pp. 2284–2303, 2021.

[5] A. Balakrishnan, J. Medikonda, and P. K. Namboothiri, "Role of wearable sensors with machine learning approaches in gait analysis for Parkinson's disease assessment: a review," *Engineered Science*, vol. 19, pp. 5–19, 2022.

[6] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, 2023, doi: https://doi.org/10.1016/j.jksuci.2023.01.014.

[7] I. El Maachi, G. A. Bilodeau, and W. Bouachir, "Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait," *Expert Syst Appl*, vol. 143, Apr. 2020, doi: 10.1016/j.eswa.2019.113075.

[8] D. M. D. Nguyen, M. Miah, G.-A. Bilodeau, and W. Bouachir, "Transformers for 1D Signals in Parkinson's Disease Detection from Gait," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2204.00423

[9] C. G. Goetz *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Mov Disord*, vol. 23, no. 15, pp. 2129–2170, 2008.

[10] J. M. Hausdorff and S. Frenkel-Toledo, "Gait in parkinson's disease," *MIT Laboratory for Computational Physiology*. 2008.

[11] C. Dong, Y. Chen, Z. Huan, Z. Li, B. Zhou, and Y. Liu, "Static-Dynamic Temporal Networks for Parkinson's Disease Detection and Severity Prediction," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2205–2213, 2023, doi: 10.1109/TNSRE.2023.3269569.

[12] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[13] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, and A. Almuhaimeed, “Skin cancer detection using combined decision of deep learners,” *IEEE Access*, vol. 10, pp. 118198–118212, 2022.

[14] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou, and S. Mohammed, “Data-driven based approach to aid Parkinson’s disease diagnosis,” *Sensors (Switzerland)*, vol. 19, no. 2, Jan. 2019, doi: 10.3390/s19020242.

[15] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff, “Dual tasking, gait rhythmicity, and Parkinson’s disease: which aspects of gait are attention demanding?,” *European journal of neuroscience*, vol. 22, no. 5, pp. 1248–1256, 2005.

[16] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, “Rhythmic auditory stimulation modulates gait variability in Parkinson’s disease,” *European Journal of Neuroscience*, vol. 26, no. 8, pp. 2369–2375, 2007.

[17] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, “Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson’s disease,” *Mov Disord*, vol. 20, no. 9, pp. 1109–1114, 2005.

[18] R. Alkhatib, M. O. DIab, C. Corbier, and M. El Badaoui, “Machine Learning Algorithm for Gait Analysis and Classification on Early Detection of Parkinson,” *IEEE Sens Lett*, vol. 4, no. 6, Jun. 2020, doi: 10.1109/LSENS.2020.2994938.

[19] M. Djurić-Jovičić, M. Belić, I. Stanković, S. Radovanović, and V. S. Kostić, “Selection of gait parameters for differential diagnostics of patients with de novo Parkinson’s disease,” *Neurol Res*, vol. 39, no. 10, pp. 853–861, Oct. 2017, doi: 10.1080/01616412.2017.1348690.

[20] C. Ricciardi *et al.*, “Machine learning can detect the presence of Mild cognitive impairment in patients affected by Parkinson’s Disease,” in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2020, pp. 1–6.

[21] B. E., B. D., and B. R., “Supervised machine learning based gait classification system for early detection and stage classification of Parkinson’s disease,” *Applied Soft Computing Journal*, vol. 94, Sep. 2020, doi: 10.1016/j.asoc.2020.106494.