



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Parkinson's disease detection using voice recordings
and Deep Learning**

Φωτεινή Α. Βασιλοπούλου
A.M. : 03114854

Επιβλέπουσα: Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Parkinson's disease detection using voice recordings and Deep Learning

Φωτεινή Α. Βασιλοπούλου
Α.Μ. : 03114854

Επιβλέπουσα: Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή τηνⁿ Νοεμβρίου, 2021.

.....
Κωνσταντίνα Νικήτα
Καθηγήτρια
Ε.Μ.Π.

.....
Α.-Γ. Σταφυλοπάτης
Καθηγητής
Ε.Μ.Π.

.....
Δ.-Δ. Κουτσούρης
Καθηγητής
Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021

Copyright © Φωτεινή Α. Βασιλοπούλου, 2021.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

.....

Φωτεινή Α. Βασιλοπούλου

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

©2021 - All rights reserved.

Περίληψη

Η νόσος του Πάρκινσον είναι μια προοδευτική νευροεκφυλιστική διαταραχή με σημαντικές κοινωνικές και οικονομικές επιπτώσεις, η οποία χαρακτηρίζεται από κινητικά και μη-κινητικά συμπτώματα. Διαταραχές σχετικές με την ομιλία όπως δυσαρθρία, δυσφωνία και υποφωνία εμφανίζονται στο 90% των ασθενών με Πάρκινσον στα πρώτα στάδια της νόσου. Η διάγνωση της ασθένειας σε αρχικό στάδιο θεωρείται κρίσιμη για την εξέλιξη και την διαχείρισή της. Το TabNet είναι ένα μοντέλο βαθιάς μάθησης που χρησιμοποιεί μηχανισμούς προσοχής και δομικά στοιχεία βαθιών νευρωνικών δικτύων για να δημιουργήσει μια αρχιτεκτονική που ακολουθεί τη λογική των δέντρων απόφασης. Η επιλογή των κυρίαρχων χαρακτηριστικών για την ταξινόμηση παρέχει στο μοντέλο τοπική και συνολική ερμηνευσιμότητα για κάθε μεμονωμένο δείγμα εισόδου. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε σετ δεδομένων διαθέσιμο από το UCI Machine Learning Repository, το οποίο περιλαμβάνει χαρακτηριστικά από ηχητικές καταγραφές 252 ατόμων (188 ασθενών με Πάρκινσον και 64 υγείων ατόμων). Για κάθε συμμετέχοντα το φωνήεν /a/ ηχογραφήθηκε τρεις φορές. Λόγω των πολλαπλών δειγμάτων ανά υποκείμενο και της ανισορροπίας μεταξύ ασθενών και υγείων, για την εκπαίδευση χρησιμοποιήθηκε η μέθοδος ομαδοποιημένων ταξινομητών και η τεχνική Leave-One-Subject-Out cross-validation. Το TabNet υπερσχύει των ταξινομητών που περιλαμβάνονται στη βιβλιογραφία, με μέση και μέγιστη ακρίβεια για την ταξινόμηση κάθε δείγματος να φθάνει το 94.5% και 95.2%, αντίστοιχα. Οι μέσες και μέγιστες τιμές για τις μετρικές F1-score και MCC προέκυψαν 96.3%, 96.8% και 86%, 87.8%, αντίστοιχα. Στην περίπτωση προβλέψεων για κάθε συμμετέχοντα, με βάση τα τρία δείγματά του, ο μέσος όρος και η μέγιστη τιμή που προέκυψε ήταν 95.9% και 97.2% accuracy, 97.3% και 98.1% F1-score, 89.4% και 92.8% MCC.

Λέξεις Κλειδιά - Πάρκινσον, ηχητικά χαρακτηριστικά, pyTorch-TabNet, ομαδοποιημένοι ταξινομητές, ερμηνευσιμότητα

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disease with significant social and economic impact. PD is characterized by motor and non-motor symptoms, with vocal disorders such as dysarthria, dysphonia, and hypophonia preceding common motor symptoms in approximately 90% of PD patients at the early stages of the disease. PD's detection and diagnosis during the early stages are considered crucial for the progression and management of the disease. TabNet is a canonical DNN architecture that uses sequential attention and DNN building blocks to implement a Decision Tree-like output manifold with soft, instant-wise feature selection giving the model local and global interpretability. TabNet was trained with a vocal features dataset available from the UCI Machine Learning repository. The dataset includes extracted features from voice recordings gathered from 252 subjects (188 PD – 64 control), with each subject providing three samples of the sustained phonation of the vowel /a/. Due to the multiple samples per subject and the unbalanced number of healthy individuals and patients with PD in the dataset, ensemble learning and the Leave-One-Subject-Out cross-validation were used to train TabNet in order to keep it unbiased. TabNet outperforms the classifiers in the literature with average and maximum accuracy for every sample reaching 94.5% and 95.2%, respectively. The average and maximum values for F1-score and MCC were 96.3%, 96.8%, and 86%, 87.8%, respectively. When doing majority voting of the three samples to make a final prediction of every subject, the average and maximum observed results were 95.9% and 97.2% for accuracy, 97.3% and 98.1% for F1-score, 89.4% and 92.8% for MCC metric.

Key words - *Parkinson's disease, vocal features, pyTorch-TabNet, ensemble learning, interpretability*

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω την επιβλέπουσά μου, Καθηγήτρια Κωνσταντίνα Νικήτα ΕΜΠ, η οποία με εμπιστεύθηκε και μου έδωσε την ευκαιρία να εκπονήσω τη διπλωματική μου εργασία στο Εργαστήριο Βιοϊατρικών Προσομοιώσεων και Απεικονιστικής Τεχνολογίας (BIOSIM) στο ΕΜΠ.

Επίσης, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στον υποψήφιο διδάκτορα Κωνσταντίνο Μήτση, την μεταδιδακτορική ερευνήτρια Καλλιόπη Δαλακλείδη και την Κωνσταντία Ζαρκογιάννη (ΕΔΙΠ, ΣΗΜΜΥ ΕΜΠ) για τη βοήθεια και την καθοδήγησή τους κατά τη διάρκεια εκπόνησης της διπλωματικής μου. Μέσω της συνεργασίας μας κατάφερα να αποκτήσω περισσότερες γνώσεις σχετικά με το αντικείμενο της εργασίας αυτής και να ανακαλύψω το ενδιαφέρον μου για την έρευνα.

Ακόμη θα ήθελα να ευχαριστήσω την οικογένειά μου για την υποστηρίξή τους όλα αυτά τα χρόνια. Τέλος, θα ήθελα να ευχαριστήσω τους ανθρώπους που γνώρισα μέσω της σχολής οι οποίοι ήταν δίπλα μου και τόσο στις καλές όσο και στις κακές στιγμές και με βοήθησαν να γίνω πιο όμορφος άνθρωπος.

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
List of figures	13
List of tables	17
1 Εκτεταμένη ελληνική περίληψη	19
1.1 Η νόσος του Πάρκινσον	19
1.1.1 Εισαγωγή	19
1.1.2 Επιδημιολογικά δεδομένα	19
1.1.3 Παθοφυσιολογία της νόσου	20
1.1.4 Διάγνωση	21
1.1.5 Θεραπεία και παρακολούθηση	22
1.2 Βαθιά Μάθηση	23
1.2.1 Εισαγωγή	23
1.2.2 TabNet	24
1.2.2.1 Εισαγωγή	24
1.2.2.2 Αρχιτεκτονική του δικτύου	24
1.2.2.3 Επεξεργασία των χαρακτηριστικών	26
1.2.2.4 Επιλογή των κυρίαρχων χαρακτηριστικών	27
1.2.3 Ερμηνευσιμότητα του μοντέλου	28
1.2.4 Αρχιτεκτονική του αποκωδικοποιητή	29
1.3 Υλικά και μεθοδοι	29
1.3.1 Δεδομένα	29
1.3.2 Διαδικασία εκπαίδευσης	30
1.3.3 Εξαγωγή των προβλέψεων του μοντέλου	31
1.4 Αποτελέσματα	31
1.4.1 Ταξινόμηση με χρήση του TabNet	31
1.4.1.1 Βέλτιστα αποτελέσματα ταξινόμησης	32
1.4.1.2 Επιλογή υπερ-παραμέτρων	33
1.4.1.3 Μέθοδος αρχικοποίησης του TabNet	36
1.4.1.4 Feature reduction	37
1.4.2 Ερμηνευσιμότητα του TabNet	39
1.5 Συζήτηση	42

1.6	Σύνοψη και μελλοντικές επεκτάσεις	45
1.6.1	Σύνοψη	45
1.6.2	Μελλοντικές επεκτάσεις	45
2	Parkinson's disease	47
2.1	Introduction	47
2.2	Epidemiology	47
2.3	Pathophysiology	48
2.4	Diagnosis	49
2.4.1	Vocal disorder based diagnosis	49
2.5	Management	50
3	Deep Learning	51
3.1	Introduction	51
3.2	Artificial Neural Networks	51
3.3	Deep Neural Networks	52
3.4	TabNet	53
3.4.1	TabNet Encoder architecture	53
3.4.1.1	Feature processing	55
3.4.1.2	Feature selection	55
3.4.2	Interpretability of the model	57
3.4.3	TabNet Decoder architecture	57
4	Materials and methods	59
4.1	Data	59
4.2	Training method	60
4.3	Model	60
4.4	Hyperparameter tuning	61
4.5	Classification predictions	62
4.6	Evaluation metrics	62
5	Experimental Results	65
5.1	Classification performance of TabNet	65
5.1.1	Best results	65
5.1.2	Hyperparameter tuning	66
5.1.3	TabNet's initialization method	69
5.1.4	Feature reduction	70
5.2	Interpretability of the model	72
6	Discussion	77
7	Conclusion and future work	81
7.1	Summary	81
7.2	Future work	81
	Bibliography	86

List of Figures

1.1	Αρχιτεκτονική του κωδικοποιητή του TabNet.	25
1.2	Εσωτερική δομή του μετασχηματιστή χαρακτηριστικών του TabNet. .	26
1.3	Εσωτερική δομή του μετασχηματιστή προσοχής του TabNet.	28
1.4	Αρχιτεκτονική του αποκωδικοποιητή του TabNet.	29
1.5	Αποτελέσματα ταξινόμησης του TabNet για κάθε συμμετέχοντα στην έρευνα, συνυπολογίζοντας τα τρία δείγματα ηχητικών καταγραφών φωνής που παρείχε.	32
1.6	Πίνακας σύγχυσης (confusion matrix) των αποτελεσμάτων ταξινόμησης για κάθε δείγμα.	32
1.7	Πίνακας σύγχυσης (confusion matrix) των αποτελεσμάτων ταξινόμησης για κάθε συμμετέχοντα συνυπολογίζοντας τα τρία του δείγματα. . . .	33
1.8	Αποτελέσματα της ταξινόμησης του TabNet με ένα decision step και διαφορετικά μεγέθη validation set.	34
1.9	Αποτελέσματα της ταξινόμησης του TabNet με τρία decision steps και διαφορετικά μεγέθη validation set.	34
1.10	Αποτελέσματα της ταξινόμησης του TabNet σε μορφή box plots με ένα decision step και διαφορετικά μεγέθη validation set.	35
1.11	Αποτελέσματα της ταξινόμησης του TabNet σε μορφή box plots με τρία decision steps και διαφορετικά μεγέθη validation set.	36
1.12	Box plots των προβλέψεων για τα δείγματα με σταθερή και τυχαία επιλεγμένη παράμετρο seed για την αρχικοποίηση του TabNet. Το μέγεθος του validation set ορίστηκε στο 10% των δειγμάτων κάθε ensemble ταξινομητή. Τα δείγματα που περιλαμβάνονταν σε κάθε validation set επιλέχθηκαν τυχαία, με την προϋπόθεση ότι υπήρχε πάντα ίσος αριθμός από υγιή άτομα και ασθενείς με Πάρκινσον. . . .	37
1.13	Μέση τιμή των μετρικών αξιολόγησης της ικανότητας γενίκευσης του TabNet που προέκυψαν από τις προβλέψεις για τα δείγματα κατά τη μείωση των αρχικά διαθέσιμων χαρακτηριστικών του dataset. . . .	38
1.14	Μέση τιμή των μετρικών αξιολόγησης της ικανότητας γενίκευσης του TabNet που προέκυψαν από τις προβλέψεις για κάθε συμμετέχοντα κατά τη μείωση των αρχικά διαθέσιμων χαρακτηριστικών του dataset.	38
1.15	Μάσκες επιλογής χαρακτηριστικών για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).	40
1.16	Μήτρα επεξηγησιμότητας για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).	40
1.17	Μάσκες επιλογής χαρακτηριστικών για τον υγιή συμμετέχοντα (ID:251).	41
1.18	Μήτρα επεξηγησιμότητας για τον υγιή συμμετέχοντα (ID:251).	41
1.19	Συντελεστές σημαντικότητας/συμμετοχής των χαρακτηριστικών στην ταξινόμηση για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).	42

1.20	Συντελεστές σημαντικότητας/συμμετοχής των χαρακτηριστικών στην ταξινόμηση για τον υγιή συμμετέχοντα (ID:251).	42
3.1	Artificial neural network architecture.[43]	52
3.2	TabNet Encoder architecture.	54
3.3	Topological structure of the Feature transformer block.	55
3.4	Topological structure of the Attentive transformer block.	57
3.5	TabNet Decoder architecture.	58
5.1	Best results obtained with TabNet in the cases of Parkinson’s disease classification for every individual sample and each subject using its three provided voice recordings.	65
5.2	Confusion matrix that presents the correct and incorrect classification results for every sample from the iteration that achieved the highest accuracy score.	66
5.3	Confusion matrix that presents the correct and incorrect classification results for every subject from the iteration that achieved the highest accuracy score.	66
5.4	TabNet’s classification results with single-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.	67
5.5	TabNet’s classification results with three-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.	67
5.6	TabNet’s classification box plots with single-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.	68
5.7	TabNet’s classification box plots with three-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.	69
5.8	Box plots of TabNet’s predictions for the test samples with constant and randomly selected seeds for every model’s initialization. The validation set’s size was set to 10% of the samples of every ensemble classifier. The samples of every validation set were randomly chosen, ensuring that it was always balanced.	70
5.9	Average TabNet’s evaluation metrics for classification of individual samples during feature reduction.	71
5.10	Average TabNet’s evaluation metrics for classification of subjects during feature reduction.	71
5.11	Masks for the samples of a subject with Parkinson’s disease (ID:100).	72
5.12	Explain matrices for the samples of a subject with Parkinson’s disease (ID:100).	73
5.13	Masks for the samples of a subject from the control group (ID:251).	73
5.14	Explain matrices for the samples of a subject from the control group (ID:251).	74

5.15	Feature importances for the samples of a subject with Parkinson's disease (ID:100).	74
5.16	Feature importances for the samples of a subject from the control group (ID:251).	75

List of Tables

1.1	Επιλεγμένες τιμές για τις υπερ-παραμέτρους του TabNet.	31
1.2	Αποτελέσματα που παρουσιάζονται στη βιβλιογραφία για την ταξινόμηση ασθενών με Πάρκινσον και υγιών ατόμων με χρήση του dataset Sakar et al. (2019) [3] για την εκπαίδευση και την αξιολόγηση του επιλεγμένου μοντέλου.	44
4.1	Extracted features from spectrograms of speech signals.[3, 32].	59
4.2	Selected hyperparameters and training parameters for TabNet.	62
4.3	Evaluation metrics.[44]	63
6.1	Parkinson’s disease classification results in the literature. All the included studies use the Sakar et al. (2019) [3] dataset to train and inference the proposed models.	79

Κεφάλαιο 1

Εκτεταμένη ελληνική περίληψη

1.1 Η νόσος του Πάρκινσον

1.1.1 Εισαγωγή

Στον τομέα της νευρολογίας, οι νευροεκφυλιστικές διαταραχές καθορίζονται από την βαθμιαία και μη αναστρέψιμη απώλεια νευρώνων από συγκεκριμένες περιοχές του εγκεφάλου. Σύμφωνα με έρευνες, η γήρανση είναι η κύρια αιτία τέτοιων ασθενειών μέσω τριών επιστημονικά αποδεδειγμένων παραγόντων στους οποίους περιλαμβάνονται οι βλάβες στο πυρηνικό DNA των νευρώνων, το αυξημένο οξειδωτικό στρες και η χρόνια φλεγμονή των νευρώνων.[1] Η νόσος του Πάρκινσον, που περιγράφηκε πρώτη φορά το 1817 από τον Δρ. James Parkinson, είναι μια προοδευτική νευροεκφυλιστική διαταραχή η οποία προκαλείται από τη νέκρωση ντοπαμινεργικών νευρώνων που βρίσκονται στη μέλαινα ουσία (substantia nigra), μία περιοχή του εγκεφάλου που είναι υπεύθυνη για τη σύνθεση της νευροδιαβιβαστικής ντοπαμίνης.[1, 2] Η νόσος του Πάρκινσον αποτελεί τη δεύτερη πιο συχνή νευροεκφυλιστική διαταραχή σε ηλικιωμένους, μετά από το Alzheimer, ανεξαρτήτου φυλής ή κοινωνικής τάξης.[1, 3]

1.1.2 Επιδημιολογικά δεδομένα

Σύμφωνα με τα παγκόσμια δεδομένα, η επίπτωση της νόσου του Πάρκινσον εκτιμάται ότι κυμαίνεται από 5 έως και πάνω από 35 περιστατικά ανά 100.000 άτομα, γεγονός που πιθανώς αντικατοπτρίζει τις διαφορές ανάμεσα στα δημογραφικά στοιχεία των πληθυσμών που μελετήθηκαν, καθώς και στον τρόπο διεξαγωγής των ερευνών.[2] Η νόσος του Πάρκινσον θεωρείται σπάνια για άτομα ηλικίας κάτω των 50 ετών, ενώ η επίπτωσή της αυξάνεται 5 έως 10 φορές από την έκτη έως την ένατη δεκαετία της ζωής ενός ατόμου. Η νόσος του Πάρκινσον σε νεαρά άτομα ορίζεται συνήθως από ηλικία έναρξης μικρότερη των 45 ετών, με το 10% των περιπτώσεων αυτών να έχουν γενετική βάση, ενώ το ποσοστό αυξάνεται σε >40% όταν η εμφάνιση γίνεται σε ηλικία κάτω των 30 ετών.[2] Ο συνολικός επιπολασμός υπολογίζεται στο 0.3% και αυξάνεται ραγδαία καθώς μεγαλώνει η ηλικία, φτάνοντας σε ποσοστά μεγαλύτερα από 3% στον πληθυσμό άνω των 80 ετών.[2]

Όσον αφορά τη θνησιμότητα, δεν παρατηρείται σημαντική μεταβολή κατά την πρώτη δεκαετία μετά από την έναρξη της νόσου. Αντίθετα, μετά το πέρας του διαστήματος αυτού σημειώνεται αύξηση του ποσοστού θανάτων, τα οποία σταδιακά γίνονται έως και διπλάσια σε σύγκριση με τον γενικό πληθυσμό.[2] Η βελτίωση της υγειονομικής περιθαλψης τα τελευταία χρόνια έχει οδηγήσει σε επέκταση του

προσδόκιμου ζωής, γεγονός το οποίο σχετίζεται με την αύξηση του επιπολασμού της νόσου σύμφωνα με μία μελέτη 20 ετών, με τον αριθμό των ατόμων με νόσο του Πάρκινσον να αναμένεται να διπλασιαστεί μεταξύ 2005 και 2030.[2] Η επέκταση του χρόνου επιβίωσης των ατόμων που διαθέτουν τη νόσο, η οποία παρατηρήθηκε μεταξύ 1990 και 2010, αναμένεται να οδηγήσει σε σταδιακή αύξηση της προσωπικής, κοινωνικής και οικονομικής επιβάρυνσης που σχετίζεται με τη νόσο καθώς ο παγκόσμιος πληθυσμός γερνά.[2]

Η νόσος του Πάρκινσον είναι δύο φορές πιο συχνή σε άνδρες από ό,τι σε γυναίκες στη γενική περίπτωση, αν και σε ορισμένους πληθυσμούς, συμπεριλαμβανομένης και μίας μελέτης από την Ιαπωνία, δεν παρατηρήθηκε διαφορά ή προέκυψαν και μεγαλύτερα ποσοστά στις γυναίκες.[2] Η προστατευτική επίδραση των γυναικείων αναπαραγωγικών ορμονών, ορισμένες γενετικές λειτουργίες που σχετίζονται με το φύλο ή η διαφορετική αντιμετώπιση των οργανισμών των δύο φύλων όταν εκτίθενται σε περιβαλλοντικούς παράγοντες κινδύνου ίσως είναι ικανά να εξηγήσουν την υπερίσχυση της νόσου του Πάρκινσον στον ανδρικό πληθυσμό.[2]

Η επίπτωση της νόσου φαίνεται να ποικίλλει εντός υποομάδων που καθορίζονται από τη φυλή, την εθνικότητα, τον γονότυπο ή το περιβάλλον. Ο τρόπος ζωής, η έκθεση σε οργανικούς ρύπους μέσω της διατροφής και οι περιβαλλοντικοί παράγοντες είναι πιθανό να εξηγούν τις αποκλίσεις που παρατηρούνται στον κίνδυνο νόσησης μεταξύ διαφορετικών πληθυσμών.[2] Για παράδειγμα, η συχνότητα εμφάνισης της νόσου του Πάρκινσον είναι σημαντικά μεγαλύτερη σε άτομα που εκτίθενται σε ουσίες όπως τα φυτοφάρμακα ή που έχουν υποστεί εγκεφαλικές βλάβες λόγω τραυμάτων και χαμηλότερα σε καπνιστές ή όσους καταναλώνουν καφέινη.[2, 4, 5] Η αυξημένη κατανάλωση γαλακτοκομικών προϊόντων έχει επίσης συνδεθεί με αυξημένο κίνδυνο εμφάνισης της νόσου. [5] Αντιθέτως, η σωματική δραστηριότητα και η υγιεινή διατροφή αποτελούμενη από σημαντικές ποσότητες φρούτων, λαχανικών και δημητριακών φαίνεται να μειώνει τον κίνδυνο νόσησης.[5]

1.1.3 Παθοφυσιολογία της νόσου

Στα χαρακτηριστικά συμπτώματα της νόσου του Πάρκινσον περιλαμβάνονται η νέκρωση νευρώνων κυρίως στην περιοχή της συμπαγούς μοίρας της μέλαινας ουσίας (substantia nigra pars compacta) του μέσου εγκεφάλου και η παρουσία συσσωματωμάτων Lewy (Lewy bodies) και νευριτών Lewy (Lewy neurites) στους εναπομείναντες νευρώνες του εγκεφάλου.[2, 4] Τα συσσωματώματα Lewy και οι νευρίτες Lewy είναι έγκλειστα συσσωματώματα πρωτεϊνών, κυρίως α-συνουκλείνης, στο κυτταρόπλασμα συγκεκριμένων νευρώνων σε διάφορες περιοχές του εγκεφάλου.[2, 4] Τα συσσωματώματα Lewy φαίνεται να είναι πιο σημαντικά στην εξέλιξη της νόσου, αφού παρατηρούνται σε πιο προχωρημένα στάδια της.[4] Παρόλο που ούτε η απώλεια ντοπαμινεργικών νευρώνων που περιέχουν νευρομελανίνη από την μέλαινα ουσία, ούτε η συγκέντρωση α-συνουκλείνης στους νευρώνες παρατηρούνται αποκλειστικά στη νόσο του Πάρκινσον, αποτελούν τις δύο κύριες νευροπαθολογίες που καθορίζουν την διάγνωση της νόσου όταν εμφανίζονται μαζί.[2] Η απώλεια των ντοπαμινεργικών νευρώνων που εντοπίζεται ακόμα και από την έναρξη της νόσου, υποδεικνύει ότι ο εκφυλισμός σε αυτήν την περιοχή του εγκεφάλου ξεκινά πριν εμφανιστούν τα πρώτα κινητικά συμπτώματα στον ασθενή, γεγονός το οποίο υποστηρίζεται από αρκετές πρόσφατες κλινικοπαθολογικές μελέτες.[2]

Τα αποτελέσματα ερευνών έχουν υποδείξει την δυσλειτουργία των μιτοχονδρίων ως έναν από τους κύριους παράγοντες στην παθογένεση της νόσου του Πάρκιν-

σον. Συγκεκριμένα, φαίνεται να δημιουργείται ένας φαύλος κύκλος στον οποίο η συσσώρευση α-συνουκλεΐνης και η μιτοχονδριακή δυσλειτουργία επιδεινώνουν η μία την άλλη, γεγονός που θα μπορούσε να εξηγήσει γιατί αυτές οι κυτταρικές αλλαγές παρατηρούνται μαζί σε εκφυλισμένους νευρώνες στη νόσο του Πάρκινσον.[2]

Οι ενδείξεις ότι το οξειδωτικό στρες, ως συνέπεια της μιτοχονδριακής δυσλειτουργίας, αυξάνεται στον εγκεφαλικό ιστό των ασθενών με νόσο του Πάρκινσον θεωρούνται αδιάσειστες, αλλά είναι αμφισβητήσιμο το αν εμφανίζεται νωρίς ή σε μεταγενέστερο χρόνο κατά τη διάρκεια του νέκρωσης των νευρώνων.[2] Παράλληλα μεγάλος αριθμός απεικονίσεων του εγκεφάλου και αναλύσεων βιοδεικτών που λήφθηκαν μετά από το θάνατο των ασθενών, δείχνουν ότι η φλεγμονή του νευρικού ιστού αποτελεί επίσης σημαντικό χαρακτηριστικό της νόσου του Πάρκινσον, το οποίο αν και δεν φαίνεται να αποτελεί το αρχικό έναυσμα για την εκδήλωση της νόσου, είναι πιθανώς βασικός παράγοντας στην παθογένεση.[2]

1.1.4 Διάγνωση

Η νόσος ορίζεται κλινικά από την παρουσία βραδυκινησίας και τουλάχιστον ενός ακόμα χαρακτηριστικού κινητικού χαρακτήρα (δυσκαμψία ή τρόμο ανάπαυσης), καθώς και πρόσθετων κριτηρίων που βοηθούν στην επιβεβαίωση ή τον αποκλεισμό της πιθανότητας νόσησης.[2] Στα κινητικά συμπτώματα της νόσου του Πάρκινσον περιλαμβάνονται η δυσκαμψία (κυρίως των κάτω άκρων) και το μοτίβο βάδισης που χαρακτηρίζεται από μικρά, ασταθή βήματα και συνεπώς μειωμένη ταχύτητα κίνησης. Καθώς η ασθένεια εξελίσσεται, το βάδισμα διακόπτεται συχνά λόγω της αίσθησης αβεβαιότητας κατά την έναρξη της κίνησης και "παγωματος", ειδικά όταν ο ασθενής συναντά πόρτες ή άλλου είδους εμπόδια.[6] Κατά την έναρξη των κινητικών συμπτωμάτων, αυτά εντοπίζονται συνήθως στην μία πλευρά του σώματος και η ασυμμετρία αυτή παραμένει σε όλη τη διάρκεια της ασθένειας.[2]

Εκτός από τα βασικά κινητικά συμπτώματα, η πλειονότητα των ασθενών με νόσο του Πάρκινσον εμφανίζει επίσης μη-κινητικά συμπτώματα.[2] Σε αυτά περιλαμβάνονται διαταραχές του ύπνου, εξασθένηση των γνωσιακών ικανοτήτων (γνωστικές, συναισθηματικές και συμπεριφορικές δυσκολίες, προβλήματα ανάκτησης μνήμης, παραισθησία, άνοια), διαταραχές της διάθεσης, δυσλειτουργία του αυτόνομου κεντρικού συστήματος (ορθοστατική υπόταση, προβλήματα στο ουροποιητικό σύστημα, δυσκοιλιότητα και υπεριδρωσία) καθώς και αισθητικά συμπτώματα (κυρίως μειωμένη αίσθηση γεύσης) και πόνος.[1, 2] Στα κλινικά χαρακτηριστικά της νόσου περιλαμβάνονται επίσης δευτερογενή κινητικά συμπτώματα όπως απώλεια έκφρασης του προσώπου, διαταραχή της άρθρωσης του λόγου, δυσφαγία και σιελορροια.[1] Η έλλειψη ντοπαμίνης επηρεάζει επίσης τις κινητικές και γνωστικές λειτουργίες που σχετίζονται με την ομιλία με τους ασθενείς σε προχωρημένο στάδιο να χάνουν πλήρως την ικανότητα ομιλίας.[1] Ορισμένα από τα παραπάνω μη-κινητικά συμπτώματα της νόσου μπορεί να προηγηθούν της εμφάνισης κλασικών κινητικών συμπτωμάτων για χρόνια ή και δεκαετίες.[2]

Σε μία μακροχρόνια μελέτη, η άνοια εντοπίστηκε στο 83%, η παραισθησία στο 74%, η συμπτωματική ορθοστατική υπόταση στο 48%, η δυσκοιλιότητα στο 40% και η ακράτεια ούρων στο 71% των ατόμων με νόσο του Πάρκινσον που επιβίωσαν περισσότερα από 20 χρόνια με την ασθένεια. Σε προχωρημένα στάδια, περιλαμβάνονταν κινητικά συμπτώματα τα οποία δεν ανταποκρίνονταν πλέον στη θεραπεία. Μερικά από ήταν η παροδική, ακούσια αδυναμία κίνησης (freezing of gate) (81%), η αστάθεια και η πτώση (87%, με κάταγμα το 35%) και ο πνιγμός

(48%).[2]

Παρόλο που η παρακολούθηση της πορείας των παραπάνω συμπτωμάτων έχει καθοριστικό ρόλο στην μακροπρόθεσμη εξέλιξη της νόσου του Πάρκινσον, οι κλινικές δοκιμές και οι μελέτες έχουν μέχρι στιγμής επικεντρωθεί στην εξέλιξη των κινητικών προβλημάτων, όπως καταγράφονται από την Unified Parkinson Disease Rating Scale (UPDRS). Η UPDRS είναι η πιο κοινώς χρησιμοποιούμενη κλίμακα για την παρακολούθηση των κινητικών δυσλειτουργιών που σχετίζονται με τη νόσο του Πάρκινσον.[2]

Παρά το γεγονός ότι σε περιπτώσεις που παρουσιάζουν πλήρως ανεπτυγμένα κλασικά χαρακτηριστικά της νόσου του Πάρκινσον η διάγνωση είναι εύκολη, σε αρχικά στάδια της παρατηρούνται ποσοστά σφάλματος στις κλινικές διαγνώσεις που προσεγγίζουν το 20% ακόμα και σε εξειδικευμένα κέντρα.[1, 2] Αυτή τη στιγμή η πιο συχνά χρησιμοποιούμενη κλινική διάγνωση που αντικαθιστά το UKPDS-BB είναι τα Movement Disorder Society Clinical Diagnostic Criteria, που δημοσιεύθηκαν το 2015 και στόχο έχουν την βελτιστοποίηση της ακρίβειας και της ευαισθησίας του αποτελέσματος της διάγνωσης.[1, 7] Η εγκυρότητα της κλινικής διάγνωσης μπορεί να βελτιωθεί με την αυστηρή χρήση τυπικών κλινικών κριτηρίων όπως το United Kingdom Parkinson's Disease Society Brain Bank (UKPDS-BB).[1, 2] Στα διαθέσιμα διαγνωστικά τεστ για την διάγνωση της νόσου του Πάρκινσον περιλαμβάνονται απεικονιστικές τεχνικές (PET, SPECT και MRI), γενετικές εξετάσεις, και εξετάσεις εγκεφαλονωτιαίου υγρού και αίματος.[2]

1.1.5 Θεραπεία και παρακολούθηση

Όπως αναφέρθηκε προηγουμένως, η απώλεια ντοπαμινεργικών νευρώνων από την συμπαγή μοίρα της μέλαινας ουσίας του εγκεφάλου που οδηγεί σε μείωση της ντοπαμίνης είναι ο βασικός μηχανισμός που προκαλεί τα κύρια κινητικά χαρακτηριστικά της ασθένειας του Πάρκινσον.[8] Αυτή τη στιγμή δεν υπάρχει κάποιου είδους θεραπεία για τη νόσο. Ως εκ τούτου, η κύρια μέθοδος που ακολουθείται στοχεύει στην υποκατάσταση της απώλειας ντοπαμίνης με συστηματική χορήγηση του προδρόμου αμινοξέος της ντοπαμίνης L-DOPA ή αγωνιστών ντοπαμίνης.[2, 9] Από τα διαθέσιμα στοιχεία για την ασθένεια φαίνεται ότι αν η θεραπεία ξεκινήσει όσο το δυνατόν νωρίτερα, η χορήγηση φαρμακευτικής αγωγής που στοχεύει στην επιβράδυνση ή τη διακοπή του νευροεκφυλισμού έχει μεγαλύτερη αποτελεσματικότητα.[6] Ορισμένα επιπλέον φάρμακα συνταγογραφούνται επίσης στους ασθενείς με Πάρκινσον, με σκοπό να αντιστρέψουν τις παρενέργειες του L-DOPA, όπως η δυσκινησία.[2]

Σε ορισμένους ασθενείς με Πάρκινσον, παρατηρείται μία ποικιλία μη-κινητικών συμπτωμάτων (πόνος, άγχος, πανικός, κατάθλιψη, νευρικότητα) που φαίνεται να παρουσιάζουν διακυμάνσεις εξαιτίας της χορηγούμενης φαρμακευτικής αγωγής, επιβαρύνοντας τους ίσως και περισσότερο σε σύγκριση με τα κλασικά κινητικά συμπτώματα.[2] Μη-κινητικά συμπτώματα όπως η γνωστική δυσλειτουργία, η κατάθλιψη και η έλλειψη αυτονομίας αντιμετωπίζονται με αντιψυχωσικά και αντικαταθλιπτικά φάρμακα, όπως η κλοζαπίνη. Από το 1993, η τεχνική Deep Brain Stimulation (DBS) χρησιμοποιείται επίσης ως θεραπεία για την νόσο του Πάρκινσον. Η ηλεκτρική διέγερση με υψηλές συχνότητες χρησιμοποιείται σε συγκεκριμένες περιοχές-στόχους του εγκεφάλου. Η θεραπεία περιλαμβάνει επίσης την εμφύτευση ενός ηλεκτροδίου στον ιστό του εγκεφάλου.[2]

Οι περισσότεροι ασθενείς με νόσο του Πάρκινσον καλούνται παράλληλα να αντιμετωπίσουν υπολειπόμενες κινητικές αναπηρίες που επηρεάζουν την βάδιση,

την ευκινησία, τον έλεγχο της στάσης του σώματος, την ισορροπία, την ομιλία και την κατάποση, οι οποίες συνήθως δεν ανταποκρίνονται επαρκώς στα φάρμακα και στην DBS.[2] Αυτά τα συμπτώματα φαίνεται να βελτιώνονται με ορισμένες μορφές σωματικής άσκησης.[2, 6]

1.2 Βαθιά Μάθηση

1.2.1 Εισαγωγή

Η τεχνητή νοημοσύνη αποτελεί έναν από τους παλαιότερους τομείς στην επιστήμη των υπολογιστών.[10] Η τεχνητή νοημοσύνη (AI) ορίζεται ως «ένα πεδίο επιστήμης και μηχανικής που ασχολείται με την υπολογιστική κατανόηση της ευφυούς συμπεριφοράς και τη δημιουργία τεχνουργημάτων που επιδεικνύουν τέτοια συμπεριφορά».[11, 12] Ο Alan Turing θεωρείται ως ένας από τους θεμελιωτές της σύγχρονης επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης, καθώς καθόρισε την «έξυπνη συμπεριφορά» στον υπολογιστή την ικανότητά του να επιδεικνύει ευφυή συμπεριφορά ισοδύναμη με, ή μη διακριτή από αυτήν του ανθρώπου, που αργότερα αναφέρθηκε ως «Turing test».[12]

Τα Τεχνητά Νευρωνικά Δίκτυα (ANN), τα οποία εντάσσονται στον τομέα της τεχνητής νοημοσύνης, είναι υπολογιστικά εργαλεία ανάλυσης και βασίστηκαν στο ανθρώπινο βιολογικό νευρικό σύστημα και τη δομή του εγκεφάλου.[12, 13] Τα νευρωνικά δίκτυα χρησιμοποιούνται σήμερα σε μια μεγάλη ποικιλία προβλημάτων όπως αναγνώριση προτύπων, ταξινόμηση, παλινδρόμηση, ομαδοποίηση, μείωση διαστάσεων, όραση υπολογιστών και επεξεργασία φυσικής γλώσσας.[13] Τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks - DNN) είναι τα νευρωνικά δίκτυα που διαθέτουν μεγάλο αριθμό κρυφών στρωμάτων.[13] Η εμφάνιση των DNNs κατέστησε δυνατή τη απεικόνιση πολύπλοκων, μη γραμμικών συναρτήσεων, η οποία είναι αδύνατη με τα απλά νευρωνικά δίκτυα.[13]

Η διάγνωση, η θεραπεία και η πρόβλεψη διαφόρων ασθενειών εξαρτώνται από έναν συνδυασμό πολλών κλινικών, βιολογικών και παθολογικών μεταβλητών. Αυτό το γεγονός οδήγησε σε μια αυξανόμενη ανάγκη για χρήση αναλυτικών εργαλείων στην ιατρική.[12] Η τεχνητή νοημοσύνη χρησιμοποιείται στον τομέα της ιατρικής με σκοπό την ανάπτυξη αλγορίθμων που θα βοηθήσουν τους επαγγελματίες υγείας να πραγματοποιήσουν μια διάγνωση και να λάβουν θεραπευτικές αποφάσεις.[12, 14, 15, 16] Τα τελευταία χρόνια, η βαθιά μηχανική μάθηση έχει ενταχθεί στον χώρο της ιατρικής εφόσον παρέχει την ικανότητα βελτίωσης της ανθρώπινης ζωής, την παροχή ακριβέστερης διάγνωσης σε ασθένειες όπως ο καρκίνος και την ανακάλυψη νέων φαρμάκων.[13] Στην βιβλιογραφία αναφέρεται ότι ένα βαθύ νευρωνικό δίκτυο είχε τη δυνατότητα να εκπαιδευτεί χρησιμοποιώντας 129.450 εικόνες από 2.032 διαφορετικές ασθένειες και να δώσει διαγνώσεις στο ίδιο επίπεδο με 21 πιστοποιημένους δερματολόγους.[17] Επιπλέον, η Google AI κατάφερε να υπερβεί τη μέση ακρίβεια διάγνωσης του καρκίνου του προστάτη, επιτυγχάνοντας ποσοστό 9% μεγαλύτερο σε σύγκριση με εκείνο γενικών παθολόγων πιστοποιημένων από την αμερικανική επιτροπή.[18]

1.2.2 TabNet

1.2.2.1 Εισαγωγή

Το 2019, οι S. Arik και T. Pfister από το Google Cloud AI πρότειναν μια νέα αρχιτεκτονική βαθιών νευρωνικών δικτύων για δεδομένα σε μορφή πίνακα, που μπορεί να χρησιμοποιηθεί για ταξινόμηση ή παλινδρόμηση, που ονομάστηκε TabNet. Το TabNet δέχεται δεδομένα σε μορφή πίνακα χωρίς να είναι απαραίτητη οποιαδήποτε προηγούμενη επεξεργασία και εκπαιδεύεται χρησιμοποιώντας τεχνικές gradient descent και εκμάθηση end-to-end. Το μοντέλο χρησιμοποιεί δομικά στοιχεία που έχουν μορφή βαθιών νευρωνικών δικτύων για να δημιουργήσει μια αρχιτεκτονική που ακολουθεί τη λογική των δέντρων απόφασης.[19]

Το TabNet χρησιμοποιεί sequential attention και υπολογίζει τον βαθμό συμμετοχής των χαρακτηριστικών στην ταξινόμηση (soft feature selection). Πιο συγκεκριμένα, η έξοδος του μοντέλου αποτελεί γραμμικό συνδυασμό των χαρακτηριστικών του πίνακα εισόδου με συντελεστές που καθορίζουν την συμμετοχή κάθε χαρακτηριστικού στο τελικό αποτέλεσμα. Το TabNet χρησιμοποιεί αραιή επιλογή χαρακτηριστικών (sparse feature selection) και κατασκευάζει μια αρχιτεκτονική με ένα ή περισσότερα διαδοχικά βήματα που καθορίζουν από κοινού την τελική απόφαση δημιουργώντας έναν ενιαίο, συλλογικό ταξινομητή (ensemble classifier). Με αυτόν τον τρόπο, το μοντέλο επιλέγει τα καλύτερα χαρακτηριστικά για κάθε μεμονωμένη είσοδο, σε αντίθεση με πολλές άλλες μεθόδους.[19]

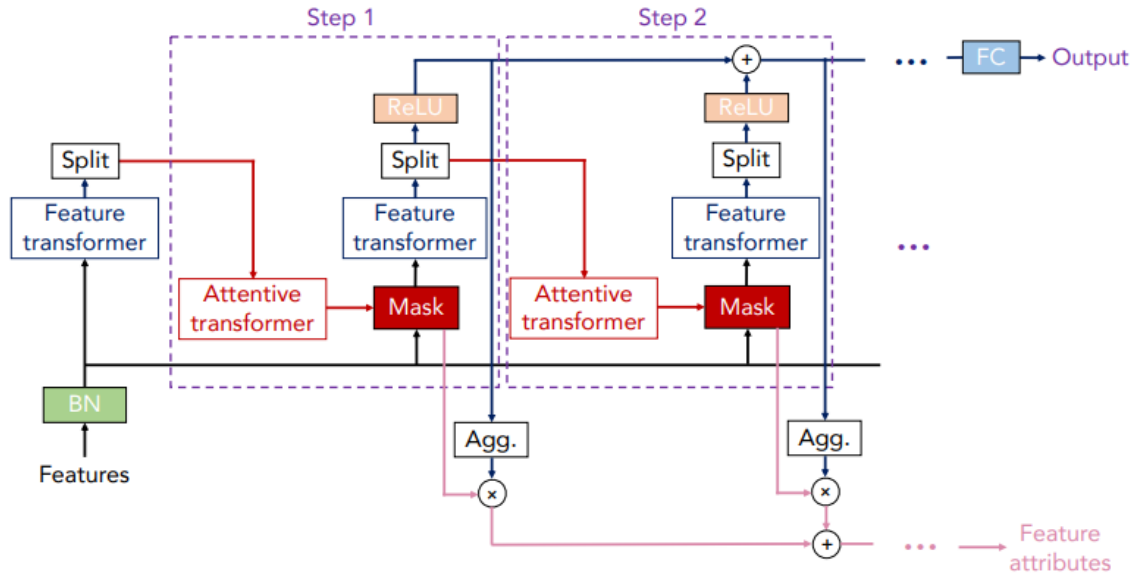
Με την επιλογή χαρακτηριστικών για κάθε δείγμα ανεξάρτητα, το μοντέλο επιτρέπει την ερμηνεία των αποτελεσμάτων ενώ οι ικανότητες εκμάθησής του αυξάνονται λόγω της μη γραμμικής επεξεργασίας των επιλεγμένων χαρακτηριστικών. Το μοντέλο χρησιμοποιεί τους πίνακες εισόδου χωρίς προηγούμενη επεξεργασία και δημιουργεί απεικονίσεις για τα κατηγορηματικά χαρακτηριστικά (categorical features). Εκτελεί επίσης μόνο κανονικοποίηση κατά παρτίδες (batch normalization) και όχι καθολική κανονικοποίηση όλων των χαρακτηριστικών εισόδου.[19]

Το TabNet χρησιμοποιεί τη μέθοδο Xavier Initialization για να αρχικοποιήσει όλα τα βάρη των νευρώνων που περιλαμβάνει, έτσι ώστε η διακύμανση να είναι η ίδια στις συναρτήσεις ενεργοποίησης κάθε επιπέδου, χρησιμοποιώντας την ομοιόμορφη κατανομή στο διάστημα $\left[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right]$, όπου n είναι το μέγεθος του προηγούμενου στρώματος. Αυτή η σταθερή διακύμανση βοηθά στην αποφυγή του απειρισμού ή εξαφάνισης της του διανύσματος κλίσης στον αλγόριθμο gradient descent.[20]

1.2.2.2 Αρχιτεκτονική του δικτύου

Ο κωδικοποιητής του TabNet αποτελείται από N_{steps} διαδοχικά επίπεδα απόφασης (decision steps). Το $i^{στο}$ επίπεδο δέχεται πληροφορία από το αμέσως προηγούμενό του ($i - 1$) ώστε να επιλέξει τα χαρακτηριστικά που θα χρησιμοποιηθούν για την δημιουργία της $i^{στης}$ Μάσκας καθώς και του διανύσματος εξόδου του επιπέδου αυτού. Αθροίζοντας τις εξόδους των επιμέρους επιπέδων προκύπτει μετά από επεξεργασία το τελικό αποτέλεσμα του TabNet. Η ιδέα της επιλογής κυρίαρχων χαρακτηριστικών με την τεχνική «από πάνω προς τα κάτω» (top-down attention) στην ακολουθιακή της μορφή είναι εμπνευσμένη από την επεξεργασία οπτικών δεδομένων και δεδομένων κειμένου και την ενισχυτική μάθηση (reinforcement learning) που εφαρμόζεται κατά τη διάρκεια αναζήτησης μικρών υποσυνόλων σχετικής πληροφορίας μέσα σε εισόδους μεγάλων διαστάσεων.[19]

Τα βασικά δομικά στοιχεία κάθε decision step του TabNet είναι ο μετασχηματιστής χαρακτηριστικών (Feature transformer), ο μετασχηματιστής προσοχής (Attentive transformer) και η Μάσκα (Mask). Κάθε στάδιο δέχεται ως είσοδο τον πίνακα δεδομένων που δίνεται για την εκπαίδευση του μοντέλου, διαστάσεων $(B \times D)$, όπου B είναι το μέγεθος του batch και D ο αριθμός των χαρακτηριστικών κάθε γραμμής του πίνακα.[19]



Εικόνα 1.1: Αρχιτεκτονική του κωδικοποιητή του TabNet.

Κάθε παρτίδα δεδομένων (batch) $X \in \mathbb{R}^{B \times D}$ του πίνακα χαρακτηριστικών εισόδου, που δίνεται ως είσοδος στο TabNet, πολλαπλασιάζεται με τη μάσκα $M[i]$ του τρέχοντος βήματος απόφασης και στη συνέχεια περνάει στον Feature transformer για την απαραίτητη επεξεργασία, η οποία συμβολίζεται με f_i . Ο τένσορας εξόδου του Feature transformer αποστέλλεται στη μονάδα Split, όπου χωρίζεται σε δύο ξεχωριστούς τένσορες που περιέχουν διαφορετικές απεικονίσεις των αρχικών χαρακτηριστικών του batch. Το πρώτο τμήμα, $d[i] \in \mathbb{R}^{B \times n_d}$, αφού περάσει από μία ReLU (ή άλλη συνάρτηση ενεργοποίησης) αποτελεί την έξοδο του τρέχοντος επιπέδου απόφασης. Το δεύτερο σκέλος, $a[i] \in \mathbb{R}^{B \times n_a}$, χρησιμοποιείται ως είσοδος στον Attentive transformer του αμέσως επόμενου βήματος απόφασης ($i + 1$) για την δημιουργία της μάσκας του, $M[i + 1]$. [19]

$$[d_i, a_i] = f_i(M_i \times X) \quad (1.1)$$

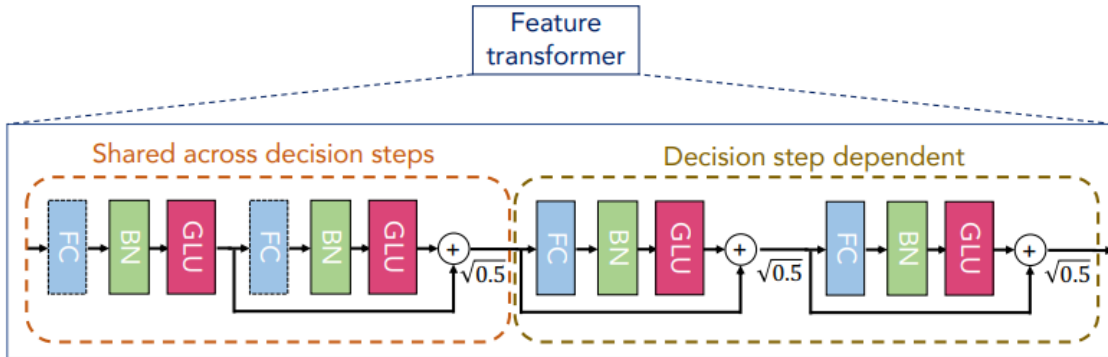
Ο τένσορας εξόδου του κωδικοποιητή, ο οποίος προκύπτει από το γραμμικό μετασχηματισμό των εξόδων όλων των decision steps, περιέχει τα επεξεργασμένα χαρακτηριστικά που θα χρησιμοποιηθούν ώστε να δημιουργηθεί η τελική απόφαση του μοντέλου όσον αφορά την ταξινόμηση ή την παλινδρόμηση, ακολουθώντας τη λογική των δέντρων απόφασης.[19]

$$d_{out} = \sum_{i=1}^{N_{steps}} ReLU(d_i) \quad (1.2)$$

Τέλος, εφαρμόζεται ένας γραμμικός μετασχηματισμός $W_{final}d_{out}$ χρησιμοποιώντας ένα πλήρως συνδεδεμένο επίπεδο νευρώνων, ώστε να λάβουμε την τελική έξοδο του TabNet. Στην έξοδο του κωδικοποιητή μπορούν να προστεθούν επιπλέον στρώματα, δημιουργώντας ένα νευρωνικό δίκτυο με οποιονδήποτε αριθμό κρυφών επιπέδων με βάση τις απαιτήσεις του προβλήματος που ζητείται να επιλυθεί, αυξάνοντας την πολυπλοκότητα του δικτύου.[19]

1.2.2.3 Επεξεργασία των χαρακτηριστικών

Όπως φαίνεται στην Εικόνα 3.3, ο Feature transformer αποτελείται από επιμέρους επίπεδα. Προκειμένου να επιτευχθεί εύρωστη μάθηση (robust learning), ο Feature transformer περιλαμβάνει ένα ή περισσότερα επίπεδα που είναι κοινά σε όλα τα decision steps (εφόσον κάθε επίπεδο δέχεται ως είσοδο τον ίδιο πίνακα χαρακτηριστικών) και άλλα τα οποία περιορίζονται αποκλειστικά στα πλαίσια του κάθε επιπέδου. Κάθε επίπεδο του Feature transformer αποτελείται από ένα πλήρως συνδεδεμένο στρώμα νευρώνων (fully-connected layer - FC), ένα επίπεδο batch normalization (BN) και μία Gated Linear Unit (GLU). Κάθε επίπεδο συνδέεται επίσης με μια υπολειμματική σύνδεση με την οποία πραγματοποιείται κανονικοποίηση με $\sqrt{0.5}$. Η λειτουργία αυτή βοηθά στη σταθεροποίηση της διαδικασίας εκμάθησης και διασφαλίζει ότι η διακύμανση δεν θα αλλάξει σημαντικά κατά μήκος του δικτύου. Για ταχύτερη εκπαίδευση, χρησιμοποιούνται batches μεγάλου μεγέθους. Επιπλέον, εκτός από το batch normalization, το TabNet χρησιμοποιεί και τη μέθοδο ghost batch normalization, χρησιμοποιώντας εικονικά batches μεγέθους B_V και momentum ίσο με m_B . [19]



Εικόνα 1.2: Εσωτερική δομή του μετασχηματιστή χαρακτηριστικών του TabNet.

Στο πρώτο κοινό επίπεδο του Feature transformer, το FC στρώμα απεικονίζει, μέσω γραμμικού μετασχηματισμού, τον τανυστή εισόδου $X \in \mathbb{R}^{BD}$ σε έναν νέο τανυστή $X \in \mathbb{R}^{B \times 2(n_d+n_a)}$ που περνά στο BN για κανονικοποίηση και στη συνέχεια στην GLU. Μέσα στην GLU (εξίσωση 3.3), το δεύτερο μισό του τένσορα λειτουργεί ως πύλη που ελέγχει (με πολλαπλασιασμό στοιχείου με στοιχείο) τις πληροφορίες που μεταφέρονται από το πρώτο του μισό στην έξοδο, $X \in \mathbb{R}^{B \times (n_d+n_a)}$, η οποία στη συνέχεια δίνεται ως είσοδος στα ακόλουθα επίπεδα που λειτουργούν με τον ίδιο τρόπο. Όλα τα fully-connected στρώματα του Feature transformer εκτός από το πρώτο, όπως είναι αναμενόμενο, έχουν $(n_d + n_a)$ εισόδους και $2(n_d + n_a)$ εξόδους.[19]

$$GLU(x) = x[:, : (n_d + n_a)] \otimes \sigma(x[:, (n_d + n_a) :]) \quad (1.3)$$

Οι Gated Linear Units μειώνουν το πρόβλημα της εξαφάνισης της κλίσης στον

αλγόριθμο gradient descent κατά την εφαρμογή βαθιάς αρχιτεκτονικής με άλλα μοντέλα, όπως τα LSTM. Σε αντίθεση με άλλα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs), οι GLU δεν έχουν χρονικές εξαρτήσεις, οπότε απλοποιούν την παραλληλοποίηση του αλγορίθμου. Η κλίση της συνάρτησης ενεργοποίησης της πύλης (εξίσωση 3.4) διαθέτει έναν επιπλέον παράγοντα $\nabla X \otimes \sigma(X)$, ο οποίος βοηθά τις κλίσεις να περνούν μέσα από τα διαδοχικά επίπεδα χωρίς να υφίστανται παραμορφώσεις.[21]

$$\nabla[X \otimes \sigma(X)] = \nabla X \otimes \sigma(X) + X \otimes \sigma'(X)\nabla X \quad (1.4)$$

1.2.2.4 Επιλογή των κυρίαρχων χαρακτηριστικών

Η επιλογή των σημαντικότερων χαρακτηριστικών σε κάθε decision step πραγματοποιείται μέσω της μάσκα του, $M[i]$. Ο Attentive transformer του $i^{\text{στου}}$ επιπέδου επιλέγει τα βέλτιστα χαρακτηριστικά εκπαιδεύοντας μια μάσκα. Πιο συγκεκριμένα, ο τανυστής $a[i-1]$ από τη μονάδα Split του επιπέδου $(i-1)$ που δίνεται ως είσοδος στον Attentive transformer, περνάει από το επίπεδο h_i , που αντιπροσωπεύει ένα πλήρως συνδεδεμένο στρώμα νευρώνων και ένα στρώμα BN. Το στρώμα h_i απεικονίζει τον τένσορα $a[i-1] \in \mathbb{R}^{B \times n_a}$ σε έναν νέο, πιο αραιό τένσορα μεγαλύτερης διάστασης, με μέγεθος $B \times D$. Η έξοδος του επιπέδου h_i στη συνέχεια πολλαπλασιάζεται με την κλίμακα Prior, $P[i-1]$, του προηγούμενου επιπέδου απόφασης. Η κλίμακα Prior ένας πίνακας τα στοιχεία του οποίου εκφράζουν τη χρήση κάθε χαρακτηριστικού από τα decision steps που έχουν προηγηθεί. Η επιθυμητή μάσκα, $M[i]$, δημιουργείται μέσω της συνάρτησης Sparsemax().[19] Η εξίσωση 3.5 περιγράφει την διαδικασία που αναφέρθηκε.

$$M[i] = \text{Sparsemax}(P[i-1] \times h_i(a[i-1])) \quad (1.5)$$

Η συνάρτηση Sparsemax απεικονίζει την Ευκλείδεια προβολή στον χώρο πιθανοτήτων, ενώ παράλληλα είναι εύκολα υπολογίσιμη και διαφορίσιμη. Η συνάρτηση επιστρέφει αραιές εκ των υστέρων κατανομές και επομένως έχει τη δυνατότητα να δώσει ακριβώς μηδενική πιθανότητα σε μερικές από τις μεταβλητές εξόδου της. Αυτή η ιδιότητα καθιστά τη Sparsemax κατάλληλη για χρήση ως φίλτρο για προβλήματα μη δυαδικής ταξινόμησης ή ως δομικό στοιχείο για τον προσδιορισμό μιας ομάδας μεταβλητών ή χαρακτηριστικών που μπορεί να σχετίζονται με μια απόφαση, με αποτέλεσμα να δημιουργείται ένα μοντέλο με μεγαλύτερη ερμηνευσιμότητα. Μπορεί επίσης να χρησιμοποιηθεί και ως συνάρτηση κόστους. [22] Η Sparsemax δημιουργεί βάρη για κάθε χαρακτηριστικό, j , κάθε δείγματος, b , με τα βάρη για κάθε δείγμα να έχουν άθροισμα ίσο με τη μονάδα, $\sum_{j=1}^D M_{b,j}[i] = 1$. Με αυτό τον τρόπο πραγματοποιείται η επιλογή των χαρακτηριστικών που θα αξιοποιηθούν από το τρέχον decision step, η οποία γίνεται για κάθε γραμμή του πίνακα εισόδου ανεξάρτητα από τις υπόλοιπες (instance-wise feature selection). Το TabNet χρησιμοποιεί την παράμετρο L_{sparse} για τον έλεγχο του ποσοστού των μη μηδενικών χαρακτηριστικών, η οποία δίνεται από την εξίσωση:

$$L_{\text{sparse}} = \sum_{i=1}^{N_{\text{steps}}} \sum_{b=1}^B \sum_{j=1}^D \frac{-M_{b,j}[i] \log(M_{b,j}[i] + \epsilon)}{N_{\text{steps}} \cdot B} \quad (1.6)$$

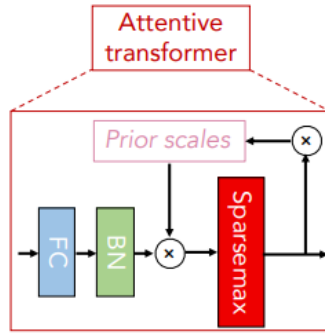
όπου ϵ είναι ένας πολύ μικρός αριθμός που προστίθεται για αριθμητική σταθερότητα.[19]

Η κλίμακα Prior $P[i]$ υποδεικνύει τη χρήση κάθε χαρακτηριστικού στα decision steps που έχουν προηγηθεί και ενημερώνεται σε κάθε decision step από την εξίσωση 3.7, όπου το γ είναι μια παράμετρος χαλάρωσης. Για $\gamma = 1$ επιβάλλεται ότι κάθε χαρακτηριστικό θα χρησιμοποιηθεί ακριβώς σε ένα στάδιο απόφασης, ενώ μεγαλύτερες τιμές παρέχουν αντίστοιχη ευελιξία στη διαδικασία επιλογής χαρακτηριστικών. Η κλίμακα Prior αρχικοποιείται στην μονάδα κατά την έναρξη της εκπαίδευσης.[19]

$$P[i] = \prod_{j=1}^i (\gamma - M[j]) \quad (1.7)$$

$$P[0] = 1^{B \times D}$$

Η εκπαιδευμένη μάσκα $M[i]$ πολλαπλασιάζεται με τον τανυστή εισόδου $X \in \mathbb{R}^{B \times D}$ προκειμένου να προκύψει ο αραιός τένσορας με τα επιλεγμένα χαρακτηριστικά του τρέχοντος βήματος απόφασης, ο οποίος στη συνέχεια εισάγεται στον Feature transformer.[19]



Εικόνα 1.3: Εσωτερική δομή του μετασχηματιστή προσοχής του TabNet.

1.2.3 Ερμηνευσιμότητα του μοντέλου

Η Μάσκα κάθε decision step επιλέγει τα πιο σημαντικά χαρακτηριστικά που στη συνέχεια συμμετέχουν στην ταξινόμηση. Το βάρος $M_{b,j}[i]$ εκφράζει τη συμμετοχή του $j^{\text{στου}}$ χαρακτηριστικού του $b^{\text{στου}}$ δείγματος, που συμβολίζεται με $f_{b,j}$, στην τελική απόφαση του μοντέλου. Η εκπαιδευμένη πλέον μάσκα κάθε σταδίου απόφασης είναι διαθέσιμη μετά το τέλος της εκπαίδευσης, προσδίδοντας στο μοντέλο τοπική ερμηνευσιμότητα (local interpretability). Λόγω της μη-γραμμικής επεξεργασίας των χαρακτηριστικών σε κάθε decision step, απαιτείται μια πρόσθετη διαδικασία προκειμένου να υπολογιστεί η σημαντικότητα των $f_{b,j}$ για ολόκληρο το μοντέλο.[19]

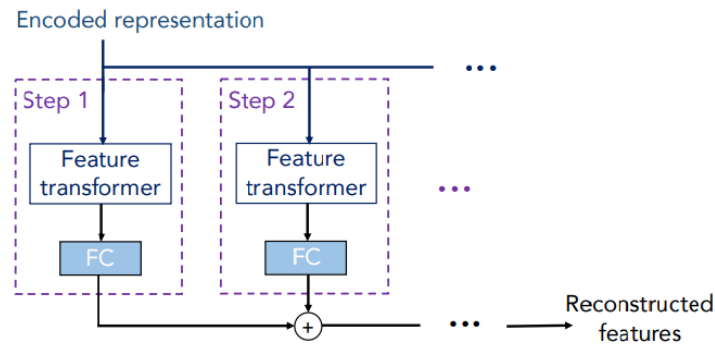
Η καθολική ερμηνευσιμότητα (global interpretability) του TabNet επιτυγχάνεται μέσω γραμμικού συνδυασμού των μασκών των επομέρους decision steps. Πιο συγκεκριμένα, το TabNet ποσοτικοποιεί την συνολική συμμετοχή κάθε χαρακτηριστικού σταθμίζοντας τον ρόλο κάθε decision step στην τελική πρόβλεψη, μέσω του συντελεστή $\eta_b[i] = \sum_{c=1}^{n_d} \text{ReLU}(d_{b,c}[i])$ για να διπλώσει τη συνολική συνεισφορά του δείγματος b^{th} στο βήμα απόφασης i^{th} . Όταν ο τένσορας $d_{b,c}[i]$ έχει μόνο αρνητικές τιμές, κανένα στοιχείο του $i^{\text{στου}}$ επιπέδου απόφασης δεν συμβάλλει στην ταξινόμηση. Η μάσκα που προκύπτει για την καθολική ερμηνευσιμότητα του TabNet

υπολογίζεται μέσω του πολλαπλασιασμού της μάσκας κάθε βήματος με τον αντίστοιχο συντελεστή βαρύτητας $\eta_b[i]$ (εξίσωση 3.8). Η κανονικοποίηση της $M_{agg-b,j}$ απαιτείται ώστε να διασφαλιστεί ότι $\sum_{j=1}^D M_{agg-b,j} = 1$. [19]

$$M_{agg-b,j} = \frac{\sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i]}{\sum_{j=1}^D \sum_{i=1}^{N_{steps}} \eta_b[i] M(b, j)[i]} \quad (1.8)$$

1.2.4 Αρχιτεκτονική του αποκωδικοποιητή

Η είσοδος του αποκωδικοποιητή του TabNet είναι η απεικόνιση του αρχικού πίνακα εισόδου, όπως προέκυψε από τον κωδικοποιητή, δηλαδή το άθροισμα των διανυσμάτων των εξόδων όλων των βημάτων απόφασης του κωδικοποιητή. Ο αποκωδικοποιητής χρησιμοποιεί τον Feature transformer κάθε βήματος απόφασης για να ανασυνθέσει τα χαρακτηριστικά και τελικά εξάγει τον ανακατασκευασμένο πίνακα εισόδου. [19]



Εικόνα 1.4: Αρχιτεκτονική του αποκωδικοποιητή του TabNet.

1.3 Υλικά και μεθοδοι

1.3.1 Δεδομένα

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε το σετ δεδομένων [3] το οποίο είναι διαθέσιμο από το UCI Machine Learning Repository από το Πανεπιστήμιο της Καλιφόρνια στο Ιρβάν. [23] Τα δεδομένα περιλαμβάνουν χαρακτηριστικά που έχουν εξαχθεί από ηχητικές καταγραφές φωνής 188 ασθενών που έχουν διαγνωσθεί με Πάρκινσον (107 ανδρών και 81 γυναικών) με ηλικίες μεταξύ 33 και 87 ετών (65.1 ± 10.9), τα οποία συλλέχθηκαν από τον Τομέα Νευρολογίας του Cerrahpaşa Faculty of Medicine του Πανεπιστημίου της Κωνσταντινούπολης. Στο dataset έχουν προστεθεί χαρακτηριστικά από 64 υγιή άτομα (23 άνδρες και 41 γυναίκες) με ηλικίες από 41 έως 82 έτη (61.1 ± 8.9).

Κάθε άτομο που συμμετείχε στη δημιουργία του dataset εξετάστηκε αρχικά από το ιατρικό προσωπικό και στη συνέχεια το φωνήεν /a/ ηχογραφήθηκε τρεις φορές, με τη συχνότητα του μικροφώνου να έχει οριστεί στα 44.1 kHz. Όλοι οι συμμετέχοντες ενημερώθηκαν σχετικά με την συλλογή των δεδομένων και συμμετείχαν εθελοντικά στην διαδικασία ακολουθώντας τους κανονισμούς της Επιτροπής Ηθικής Κλινικής Έρευνας του Πανεπιστημίου Bahcesehir. [3]

Στα χαρακτηριστικά περιλαμβάνονται βασικά χαρακτηριστικά ηχητικών σημάτων, χαρακτηριστικά συχνότητας (time frequency features), συντελεστές Mel-Freque-

ncy Cepstral (MFCCs), χαρακτηριστικά φωνητικών χορδών (vocal fold features), καθώς και επιπλέον χαρακτηριστικά που προέκυψαν με χρήση του μετασχηματισμού tunable Q-factor wavelet transform (TQWT) που σχετίζονται με την θεμελιώδη συχνότητα των φωνηέντων που καταγράφηκαν και προτάθηκε από τον Sakar και τους συνεργάτες του [3]. Τα χαρακτηριστικά που προέκυψαν για κάθε δείγμα κανονικοποιήθηκαν έτσι ώστε να έχουν μηδενική μέση τιμή και μοναδιαία διασπορά.[3]

1.3.2 Διαδικασία εκπαίδευσης

Για την εκπαίδευση του μοντέλου χρησιμοποιούμε τη μέθοδο leave-one-subject-out (LOSO) cross-validation. Συγκεκριμένα πραγματοποιούνται ανεξάρτητες εκπαιδεύσεις ίσες με τον αριθμό των ατόμων που περιλαμβάνονται στο dataset, όπου σε κάθε επανάληψη παραλείπεται και από ένα άτομο (τα τρία δείγματα που ανήκουν σε αυτό). Τα δείγματα αυτά αποτελούν το test σετ για το μοντέλο που εκπαιδεύεται εκείνη τη στιγμή. Προκειμένου να περιοριστεί το πρόβλημα του μη ισορροπημένου αριθμού υγιών-ασθενών στα δεδομένα, ακολουθούμε τη μέθοδο ομαδοποιημένων ταξινομητών (ensemble classification). Πιο συγκεκριμένα, μετά από την αφαίρεση των δειγμάτων που θα χρησιμοποιηθούν για τον έλεγχο γενίκευσης του μοντέλου, τα υπόλοιπα δείγματα χωρίζονται σε δύο επιμέρους datasets. Τα δείγματα των υγιών συμμετεχόντων περιλαμβάνονται και στα δύο datasets, ενώ τα δείγματα των ασθενών με Πάρκινσον μοιράζονται σε αυτά εξ' ίσου, με τυχαίο τρόπο, διασφαλίζοντας ότι τα τρία δείγματα κάθε συμμετέχοντα περιλαμβάνονται στο ίδιο υποσύνολο. Τα dataset που προκύπτουν συνεχίζουν να έχουν μεγαλύτερο αριθμό ασθενών, όμως χωρίζοντας το αρχικό dataset σε περισσότερα τμήματα παρατηρήθηκε ότι δεν υπήρχαν αρκετά δείγματα για την εκπαίδευση του μοντέλου.

Μετά τη διάσπαση του αρχικού dataset σε δύο υποσύνολα, καθένα από αυτά χωρίστηκε περαιτέρω σε train και validation σετ. Το validation σετ αποτελείται κάθε φορά από τυχαία επιλεγμένα δείγματα, διασφαλίζοντας και πάλι ότι δεν υπάρχουν δείγματα του ίδιου ατόμου στο train και το validation σετ ώστε το μοντέλο να παραμείνει αμερόληπτο. Επιπλέον, το validation σετ κάθε ταξινομητή είχε πάντα ίσο αριθμό υγιών-ασθενών και το μέγεθός του ήταν ένα συγκεκριμένο ποσοστό του συνόλου των δειγμάτων του ταξινομητή. Τα υπόλοιπα δείγματα χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Το validation σετ χρησιμοποιήθηκε επίσης για την ρύθμιση των υπερ-παραμέτρων και την διακοπή της εκπαίδευσης του TabNet πριν ξεκινήσει το overfitting. Αυτή η μέθοδος διαχωρισμού διασφαλίζει ότι οι ensemble ταξινομητές έχουν validation set που αποτελούνται από διαφορετικά δείγματα διαφορετικών ατόμων, που σε αντίθετη περίπτωση θα οδηγούσε σε μη αμερόληπτα μοντέλα.

Στην παρούσα διπλωματική εργασία χρησιμοποιείται η υλοποίηση του TabNet της Google AI σε PyTorch, η οποία είναι διαθέσιμη στη σελίδα [24]. Το Pytorch-TabNet αποτελεί λογισμικό ανοιχτού κώδικα που δημιουργήθηκε από την Dreamquark, με κύριους συντελεστές στο έργο τους Sebastien Fischman, Eduardo Carvahlo Pinto και Bazire Houssin.

Αρχικά, πραγματοποιήσαμε ρύθμιση των υπερ-παραμέτρων χρησιμοποιώντας τη μέθοδο αναζήτησης πλέγματος (grid search). Παρ'όλα αυτά παρατηρήθηκε ότι δεν υπήρχε σημαντική μεταβολή μεταξύ των τιμών που συμπεριλήφθηκαν στο grid search set, επομένως οι τιμές των υπερ-παραμέτρων καθορίστηκαν και παρέμειναν σταθερές κατά τη διάρκεια της εκπαίδευσης ώστε να περιοριστεί το υπολογιστικό

κόστος ανά εκπαίδευση για το σύνολο των συμμετεχόντων στο dataset (Πίνακας 1.1). Τέλος, λόγω της προσθήκης του ελέγχου της παραμέτρου *seed* για την εκπαίδευση του Pytorch-TabNet για λόγους επαναληψιμότητας των αποτελεσμάτων, πραγματοποιήθηκαν δοκιμές τόσο με σταθερή όσο τυχαία τιμή για την παράμετρο αυτή.

Πίνακας 1.1: Επιλεγμένες τιμές για τις υπερ-παραμέτρους του TabNet.

Υπερ-παράμετρος	Τιμή
n_d	8
n_a	8
n_{steps}	1
$batch_size$	64
$virtual_batch_size$	32
$gamma$	1.3 (default)
$n_{independent}$	2
n_{shared}	2
$lamda_sparse$	$1e - 3$ (default)
$learning\ rate$	0.01
$scheduler_params$	{ $step_size : 10, gamma : 0.9$ }
$patience$	40

1.3.3 Εξαγωγή των προβλέψεων του μοντέλου

Το TabNet έχει τη δυνατότητα να εξάγει αποτελέσματα για την ταξινόμηση τόσο σε δυαδική όσο και πιθανοτική μορφή. Προκειμένου να αυξηθεί τη σταθερότητα και την αξιοπιστία του μοντέλου, αποφασίσαμε να πραγματοποιήσουμε τεχνική soft voting χρησιμοποιώντας τις προβλέψεις πιθανοτήτων από τους ensemble ταξινομητές και στη συνέχεια να μετατρέψουμε το αποτέλεσμα σε δυαδική μορφή.

Τέλος, ακολουθώντας τη μέθοδο που χρησιμοποιείται συχνά στην έρευνα για την ταξινόμηση ατόμων σε υγιείς και νοσούντες με Πάρκινσον μέσα από ηχητικές καταγραφές φωνής, υπολογίσαμε μία πρόβλεψη ανά συμμετέχοντα με βάση την πλειοψηφία (majority voting) των τριών δειγμάτων που παρείχε.

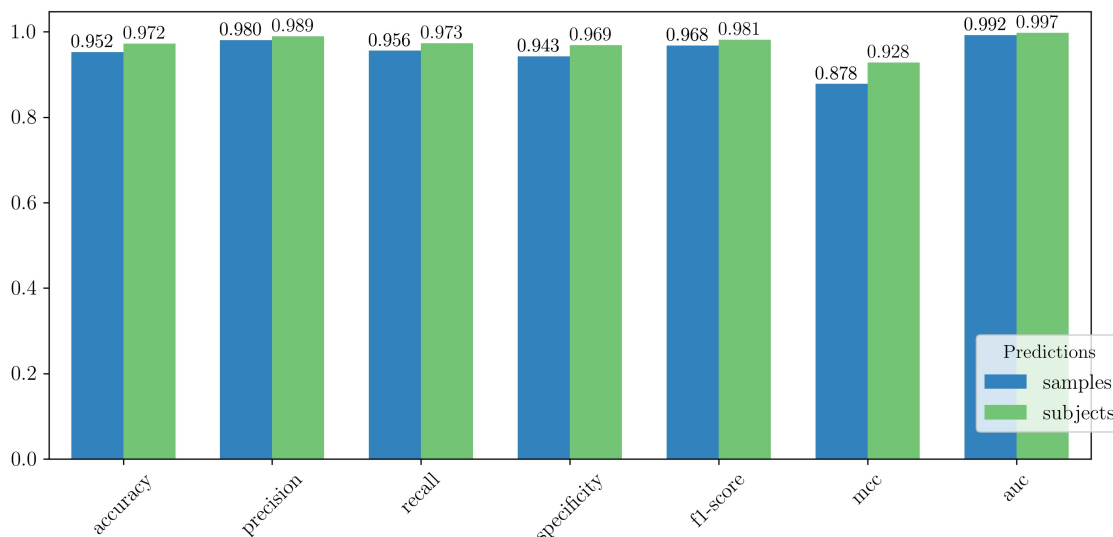
1.4 Αποτελέσματα

1.4.1 Ταξινόμηση με χρήση του TabNet

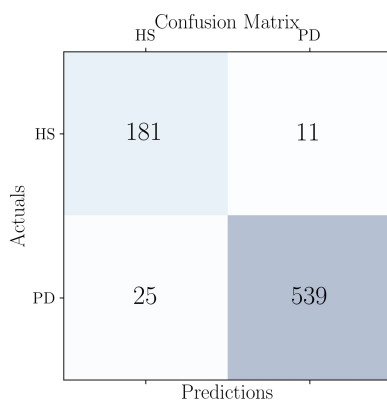
Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα του TabNet για την ταξινόμηση ασθενών με Πάρκινσον και υγιών δειγμάτων. Για κάθε πείραμα που πραγματοποιήθηκε, υπολογίσαμε τις επιλεγμένες μετρικές για κάθε δείγμα ως ανεξάρτητη δοκιμαστική περίπτωση. Ακολουθώντας τη μέθοδο που χρησιμοποιείται συχνά σε παρόμοιες έρευνες στη βιβλιογραφία, πραγματοποιήσαμε επίσης ψηφοφορία κατά πλειοψηφία των τριών διαθέσιμων δειγμάτων για να κάνουμε μια τελική πρόβλεψη για κάθε συμμετέχοντα αφού κάθε άτομο που πήρε μέρος στη μελέτη παρείχε τρεις διαφορετικές ηχητικές καταγραφές φωνής.

1.4.1.1 Βέλτιστα αποτελέσματα ταξινόμησης

Η Εικόνα 1.5 παρουσιάζει τα καλύτερα αποτελέσματα που προέκυψαν από την αξιολόγηση της ικανότητας γενίκευσης του TabNet με τη μέθοδο LOSO cross-validation. Πιο συγκεκριμένα, το μοντέλο ορίστηκε με βάση τις επιλεγμένες υπερπαραμέτρους (Πίνακας 1.1) με το μέγεθος του validation σετ ίσο με το 10% των συνολικών δειγμάτων εκπαίδευσης κάθε ensemble ταξινομητή.



Εικόνα 1.5: Αποτελέσματα ταξινόμησης του TabNet για κάθε συμμετέχοντα στην έρευνα, συνυπολογίζοντας τα τρία δείγματα ηχητικών καταγραφών φωνής που παρείχε.



Εικόνα 1.6: Πίνακας σύγχυσης (confusion matrix) των αποτελεσμάτων ταξινόμησης για κάθε δείγμα.

Το TabNet πραγματοποίησε ταξινόμηση με ακρίβεια 95.24% για τα δείγματα και 97.22% για τα άτομα συνυπολογίζοντας και τα τρία δείγματά τους. Όσον αφορά τις υπόλοιπες μετρικές του TabNet, τα αποτελέσματα των προβλέψεων ανά άτομο ήταν 98.92% precision, 97.34% recall, 96.88% specificity, 98.12% F1-score και 92.83% MCC, με τις αντίστοιχες τιμές να είναι ελαφρώς χαμηλότερες για την περίπτωση των μεμονωμένων δειγμάτων. Από τα 64 άτομα στην ομάδα ελέγχου, 2 ταξινομήθηκαν λανθασμένα ως ασθενείς με Πάρκινσον, ενώ 5 άτομα που διαγνώστηκαν με νόσο του Πάρκινσον από τους γιατρούς ταξινομήθηκαν λανθασμένα ως υγιή (Εικόνα 1.7).

Confusion Matrix

		HS	PD
		HS	PD
Actuals	HS	62	2
	PD	5	183
		Predictions	

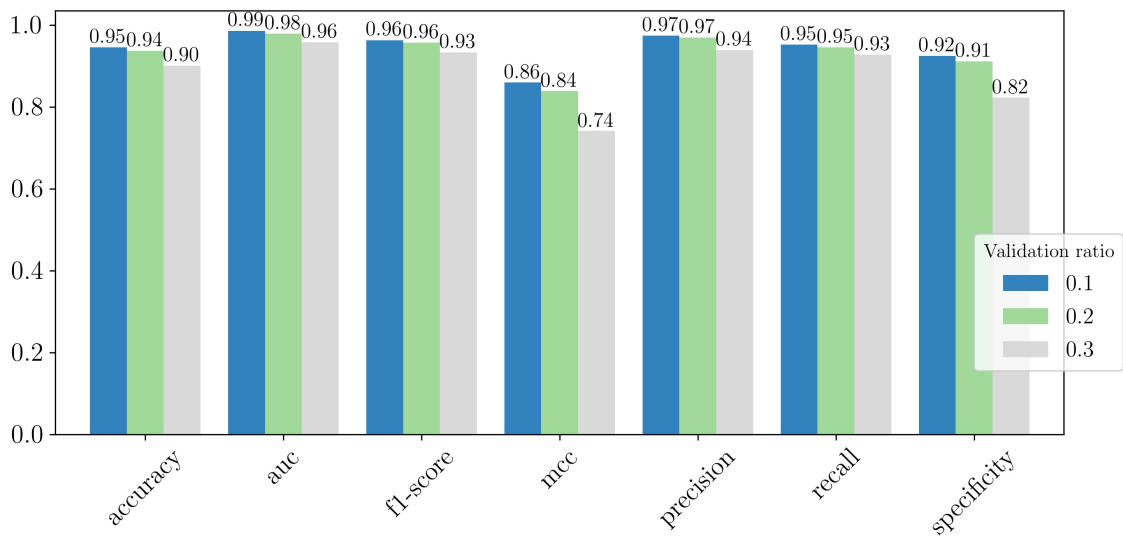
Εικόνα 1.7: Πίνακας σύγχυσης (confusion matrix) των αποτελεσμάτων ταξινόμησης για κάθε συμμετέχοντα συνυπολογίζοντας τα τρία του δείγματα.

1.4.1.2 Επιλογή υπερ-παραμέτρων

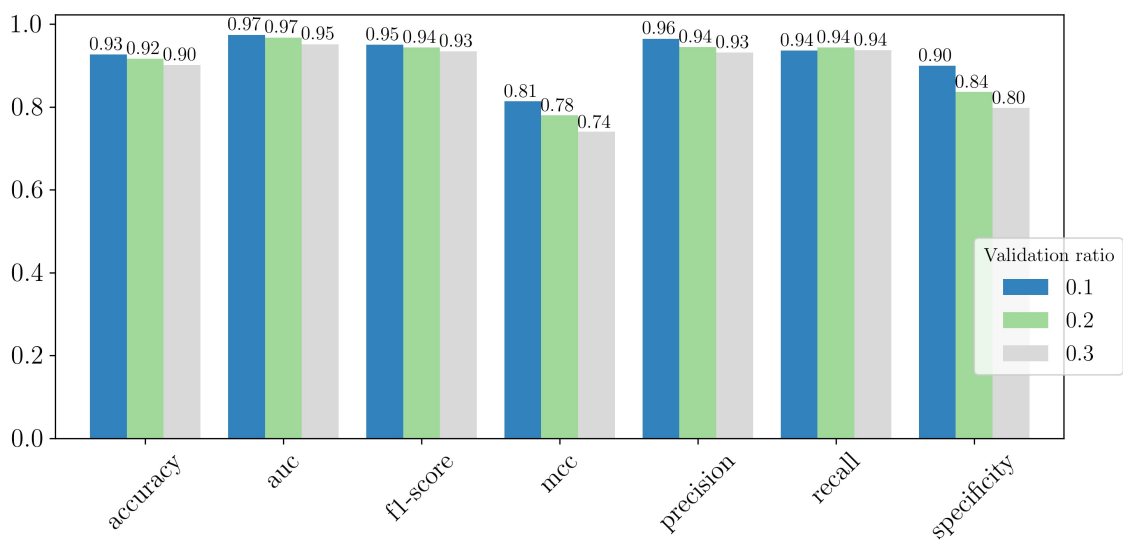
Στις Εικόνες 1.8 και 1.9 φαίνεται η μεταβολή στη μέση απόδοση του μοντέλου για τις ενδεικτικές περιπτώσεις όπου χρησιμοποιείται αρχιτεκτονική ενός και τριών decision steps. Σύμφωνα με τα αποτελέσματα, η χρήση μιας βαθύτερης αρχιτεκτονικής (με τρία decision steps) είχε ως αποτέλεσμα μια μικρή πτώση στις μετρικές του TabNet. Το γεγονός αυτό θα μπορούσε να είναι συνέπεια του εξαιρετικά μικρού συνόλου δεδομένων που διατίθενται για την εκπαίδευση του μοντέλου, καθώς τα μεγαλύτερα μοντέλα χρειάζονται μεγάλο όγκο δεδομένων για να εκπαιδευτούν επαρκώς. Το ίδιο αποτέλεσμα παρατηρήθηκε όταν αλλάξαμε τις υπερπαραμέτρους προκειμένου να δημιουργήσουμε ένα πιο σύνθετο μοντέλο. Για το λόγο αυτό διατηρήσαμε την αρχιτεκτονική των μοντέλων όσο το δυνατόν πιο απλή, με ένα decision step, δύο ανεξάρτητα και δύο κοινά επίπεδα απόφασης για τον Feature transformer και μικρά batch sizes.

Προκειμένου να διασφαλιστεί ότι το μέγεθος των σετ που χρησιμοποιήθηκαν για το validation δεν επηρέασε την επιλογή των υπερ-παραμέτρων, συμπεριλάβαμε σε αυτές το ποσοστό των δειγμάτων που χρησιμοποίησε το TabNet για το validation. Το μοντέλο δοκιμάστηκε με μεγέθη για το validation set ίσα με το 10%, 20% και 30% των συνολικών δειγμάτων εκπαίδευσης κάθε ensemble ταξινομητή. Μικρές αλλαγές στις παραμέτρους οδήγησαν σε ελαφρώς υψηλότερες ή χαμηλότερες μετρικές σχετικά με την ικανότητα γενίκευσης του TabNet. Ωστόσο, όπως αναφέρθηκε προηγουμένως, αποφασίσαμε να μην χρησιμοποιήσουμε grid search για την επιλογή των υπερ-παραμέτρων και ορίσαμε εξ'αρχής τις τιμές έτσι ώστε να μειώσουμε τις υπολογιστικές ανάγκες της εκπαίδευσης.

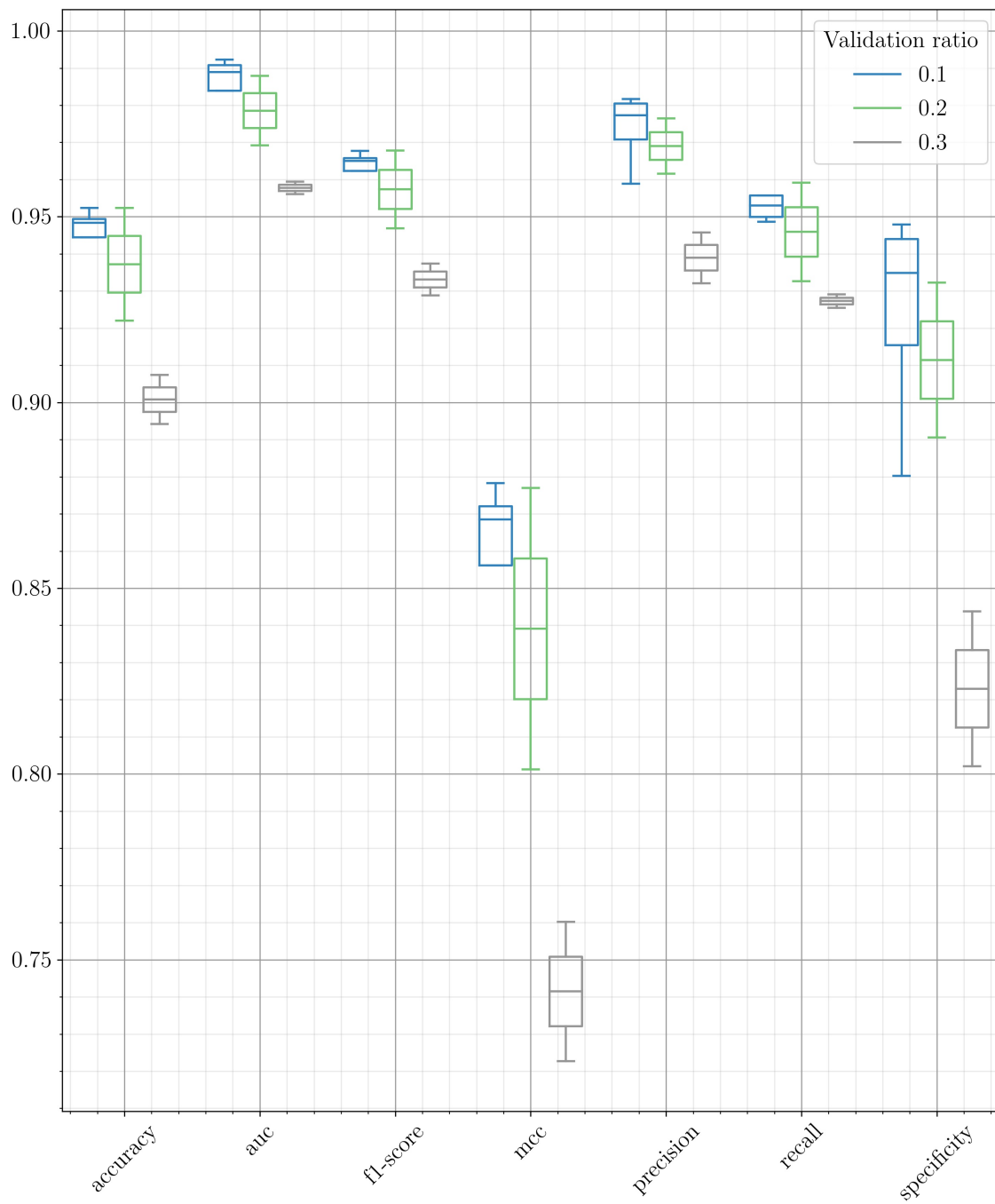
Για κάθε συνδυασμό των υπερπαραμέτρων στο grid search, επαναλάβαμε την LOSO cross-validation εκπαίδευση τέσσερις ή περισσότερες φορές, προκειμένου να έχουμε μια γενικότερη εκτίμηση της απόδοσης του TabNet. Στις Εικόνες 1.10, 1.11 παρουσιάζονται τα box plots των αποτελεσμάτων ταξινόμησης χρησιμοποιώντας αρχιτεκτονική ενός και τριών decision steps. Η αύξηση της πολυπλοκότητας του μοντέλου είχε ως αποτέλεσμα λιγότερο εύρωστους ταξινομητές με μικρότερες μέσες τιμές και μεγαλύτερες διακυμάνσεις στις μετρικές μεταξύ των μοντέλων που εκπαιδεύτηκαν με τις ίδιες παραμέτρους.



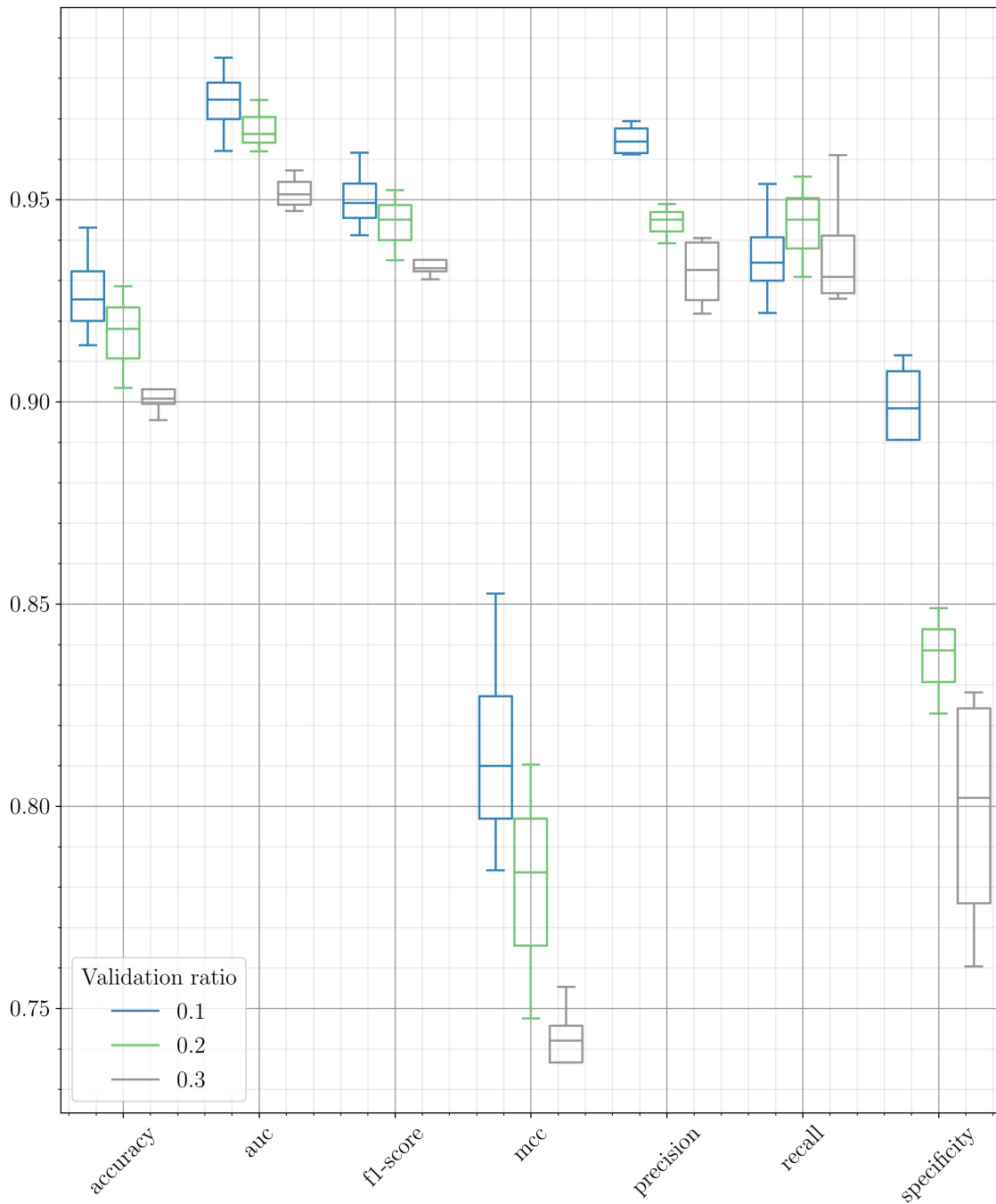
Εικόνα 1.8: Αποτελέσματα της ταξινόμησης του TabNet με ένα decision step και διαφορετικά μεγέθη validation set.



Εικόνα 1.9: Αποτελέσματα της ταξινόμησης του TabNet με τρία decision steps και διαφορετικά μεγέθη validation set.



Εικόνα 1.10: Αποτελέσματα της ταξινόμησης του TabNet σε μορφή box plots με ένα decision step και διαφορετικά μεγέθη validation set.



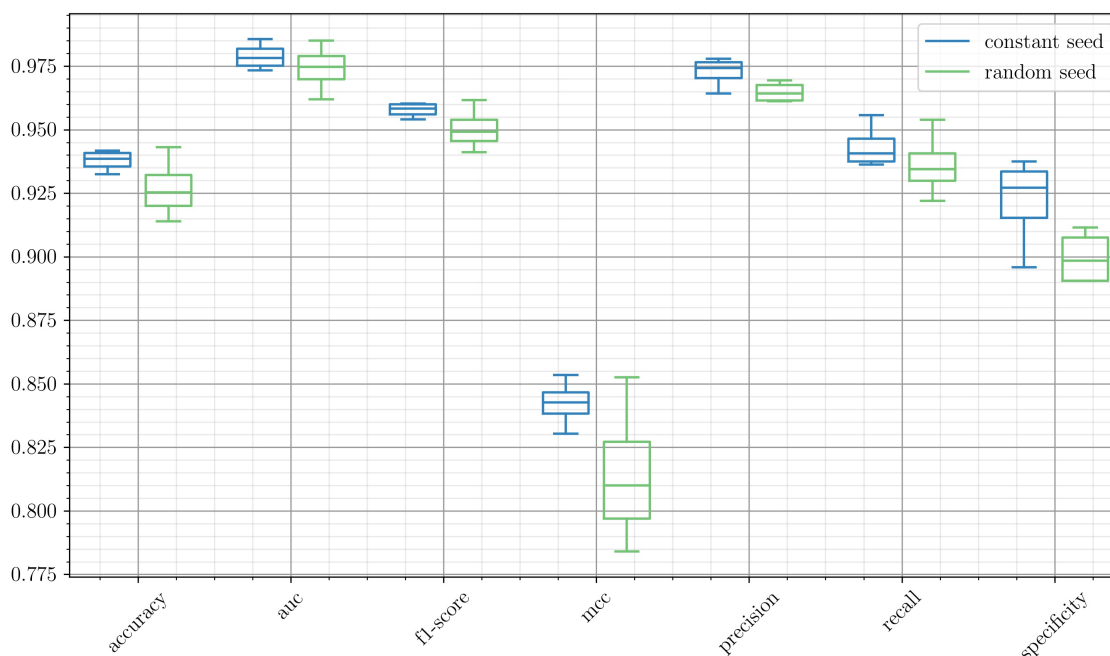
Εικόνα 1.11: Αποτελέσματα της ταξινόμησης του TabNet σε μορφή box plots με τρία decision steps και διαφορετικά μεγέθη validation set.

1.4.1.3 Μέθοδος αρχικοποίησης του TabNet

Όπως αναφέρθηκε προηγουμένως, η υλοποίηση του Pytorch-TabNet περιλαμβάνει τον χειροκίνητο καθορισμό της παραμέτρου seed για την αρχικοποίηση του μοντέλου, προκειμένου να διασφαλιστεί η επαναληψιμότητα των αποτελεσμάτων. Λαμβάνοντας υπόψη ότι το σύνολο των δεδομένων χωρίζεται τυχαία σε επιμέρους dataset για τους ensemble ταξινομητές και μετέπειτα σε validation και test σετ, πραγματοποιήσαμε ορισμένες επιπλέον δοκιμές ώστε να εκτιμήσουμε την επίδραση της επιλογής των δειγμάτων που απαρτίζουν το validation set στην απόδοση του

TabNet. Το μέγεθος του validation set ορίστηκε στο 10% των δειγμάτων κάθε ensemble ταξινομητή. Τα δείγματα κάθε validation set επιλέχθηκαν τυχαία, διασφαλίζοντας ότι υπήρχε πάντα ίσος αριθμός από υγιή άτομα και ασθενείς με Πάρκινσον. Η Εικόνα 1.12 περιλαμβάνει τα αποτελέσματα ταξινόμησης των προβλέψεων του TabNet με σταθερά και τυχαία επιλεγμένα seed για την αρχικοποίηση κάθε μοντέλου της LOSO cross-validation. Τα αποτελέσματα υποδεικνύουν ότι το validation set που επιλέγεται επηρεάζει την ικανότητα ταξινόμησης του μοντέλου, λαμβάνοντας υπόψη ότι το TabNet διαθέτει πλήρη επαναληψιμότητα για συγκεκριμένη τιμή της παραμέτρου seed.

Όταν η παράμετρος seed επιλέγεται τυχαία για κάθε μοντέλο κατά την LOSO cross-validation εκπαίδευση, η απόδοση του μοντέλου όσον αφορά την ευρωστία και την ακρίβεια των προβλέψεων ποικίλλει σημαντικά. Για παράδειγμα, η ακρίβεια (accuracy) κυμαίνεται μεταξύ 91,4% και 94,31%, ενώ οι τιμές άλλων μετρήσεων όπως το MCC διαφέρουν έως και 6,85%.



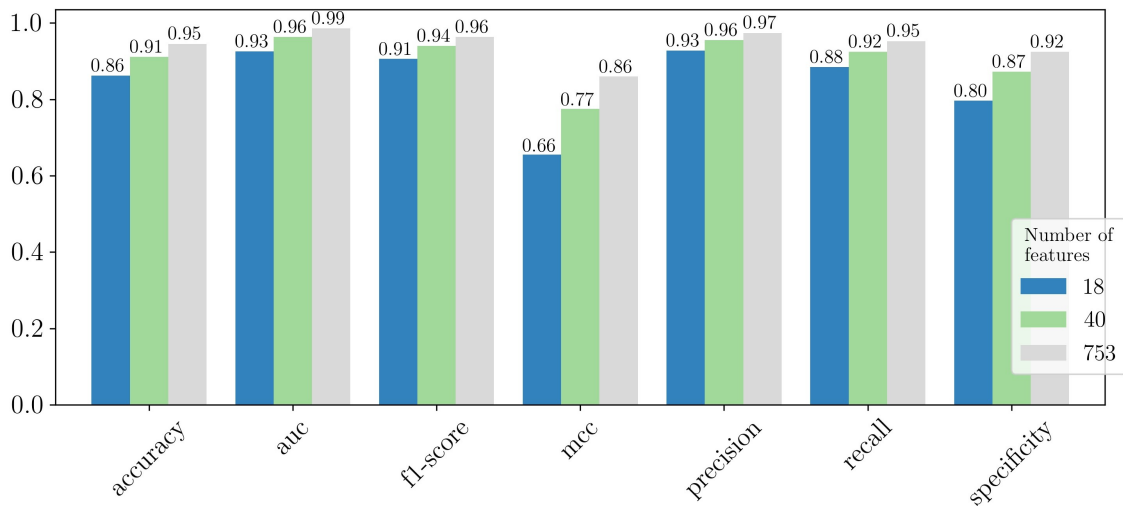
Εικόνα 1.12: Box plots των προβλέψεων για τα δείγματα με σταθερή και τυχαία επιλεγμένη παράμετρο seed για την αρχικοποίηση του TabNet. Το μέγεθος του validation set ορίστηκε στο 10% των δειγμάτων κάθε ensemble ταξινομητή. Τα δείγματα που περιλαμβάνονταν σε κάθε validation set επιλέχθηκαν τυχαία, με την προϋπόθεση ότι υπήρχε πάντα ίσος αριθμός από υγιή άτομα και ασθενείς με Πάρκινσον.

1.4.1.4 Feature reduction

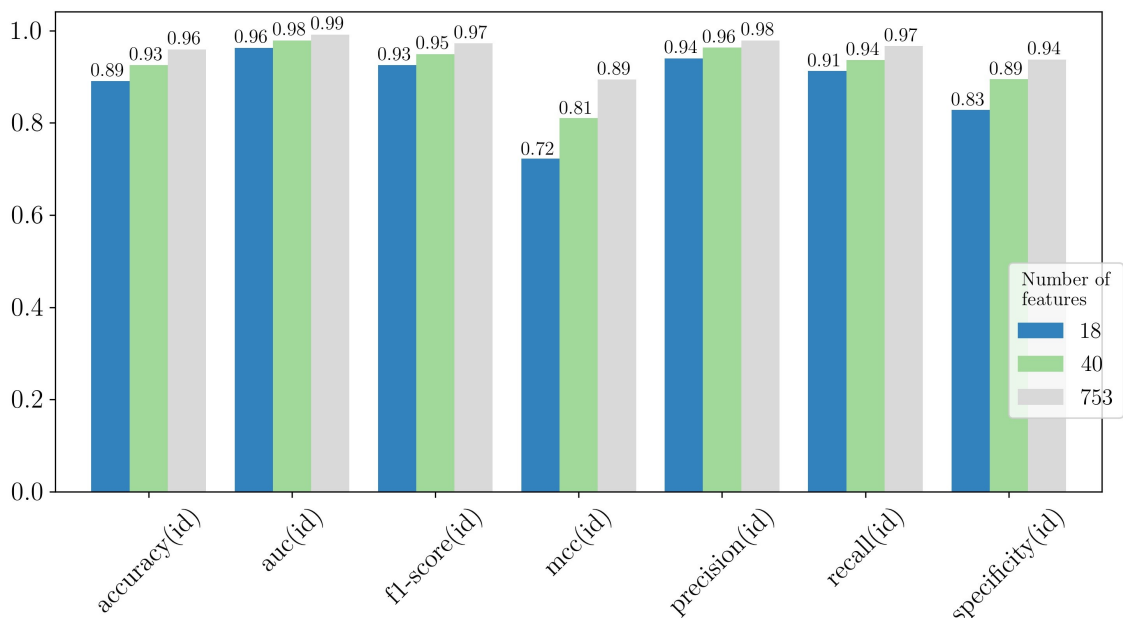
Ο Sakar και οι συνεργάτες του [3] δημιούργησαν το μεγαλύτερο έως αυτή τη στιγμή σύνολο δεδομένων για την ταξινόμηση του Parkinson από ηχητικές καταγραφές φωνής τόσο ως προς τον αριθμό των συμμετεχόντων όσο και των χαρακτηριστικών, με τα άλλα διαθέσιμα dataset να περιέχουν λιγότερα από 50 χαρακτηριστικά. Στο τελευταίο μέρος αυτής της παραγράφου, περιλαμβάνεται η διερεύνηση της επίδρασης του αρχικού αριθμού χαρακτηριστικών μεταξύ των οποίων το TabNet καλείται να επιλέξει τα πιο σημαντικά ώστε να τα χρησιμοποιήσει για την ταξινόμηση. Για το λόγο αυτό, μειώσαμε τον αριθμό των χαρακ-

τηριστικών από 753 σε 40 και 18. Και στις δύο περιπτώσεις, παραλείψαμε το μεγαλύτερο ποσοστό των χαρακτηριστικών, διατηρώντας μόνο εκείνα τα οποία ήταν κοινά μεταξύ του dataset [3] και δύο άλλων, συχνά χρησιμοποιούμενων στη βιβλιογραφία, dataset.

Το dataset [25] περιλαμβάνει συνολικά 195 δείγματα με 18 χαρακτηριστικά από 23 ασθενείς με Πάρκινσον και 8 υγιή άτομα και το dataset [26] αποτελείται από 40 χαρακτηριστικά από 40 ασθενείς με PD και 40 άτομα στην ομάδα ελέγχου.



Εικόνα 1.13: Μέση τιμή των μετρικών αξιολόγησης της ικανότητας γενίκευσης του TabNet που προέκυψαν από τις προβλέψεις για τα δείγματα κατά τη μείωση των αρχικά διαθέσιμων χαρακτηριστικών του dataset.



Εικόνα 1.14: Μέση τιμή των μετρικών αξιολόγησης της ικανότητας γενίκευσης του TabNet που προέκυψαν από τις προβλέψεις για κάθε συμμετέχοντα κατά τη μείωση των αρχικά διαθέσιμων χαρακτηριστικών του dataset.

Οι Εικόνες 1.13 και 1.14 παρουσιάζουν την σημαντική πτώση που παρατηρήθηκε σε όλες τις μετρικές αξιολόγησης της ικανότητας γενίκευσης του TabNet, κα-

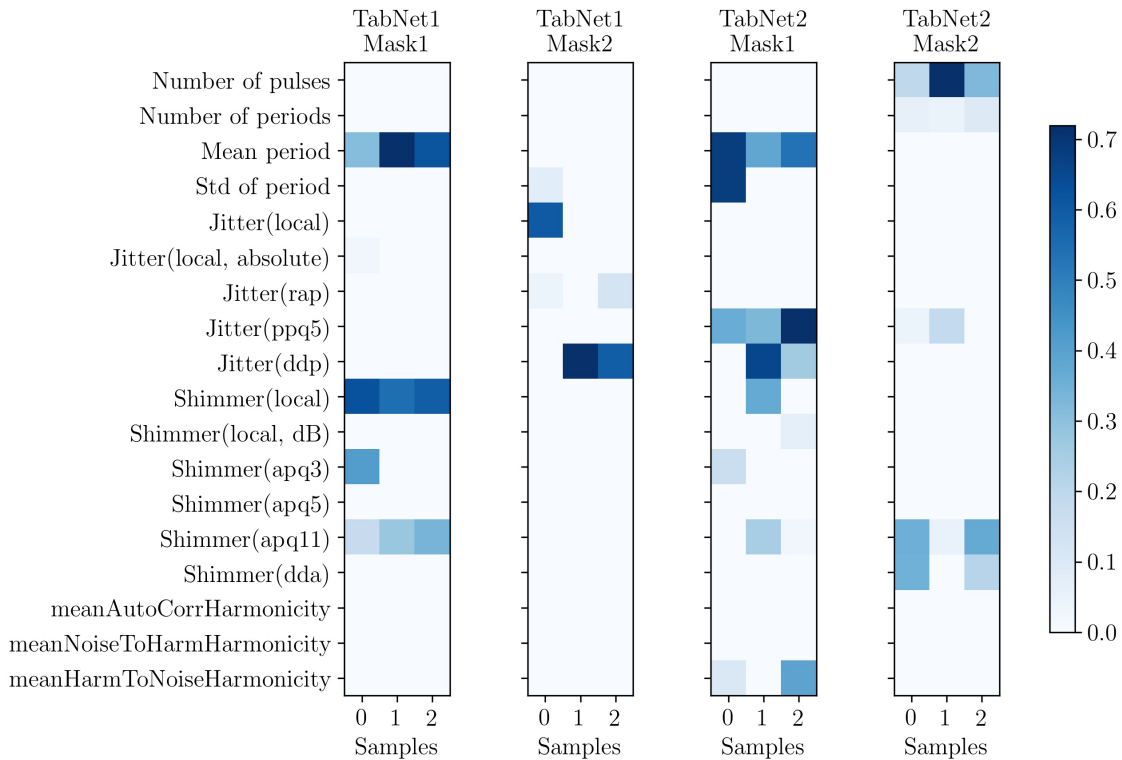
θώς μειώνεται ο αριθμός των αρχικά διαθέσιμων χαρακτηριστικών. Η ακρίβεια των προβλέψεων κάθε δείγματος μειώνεται από 96% σε 89% στην περίπτωση των 18 χαρακτηριστικών. Κατά τη διάρκεια αυτής της φάσης, χρησιμοποιήσαμε τυχαία αρχικοποίηση για το μοντέλο και validation set με μέγεθος ίσο με το 10% του συνόλου των δειγμάτων εκπαίδευσης κάθε ensemble ταξινομητή. Οι υπερ-παραμέτροι ορίστηκαν στις τιμές που αναφέρονται στον Πίνακα 1.1.

1.4.2 Ερμηνευσιμότητα του TabNet

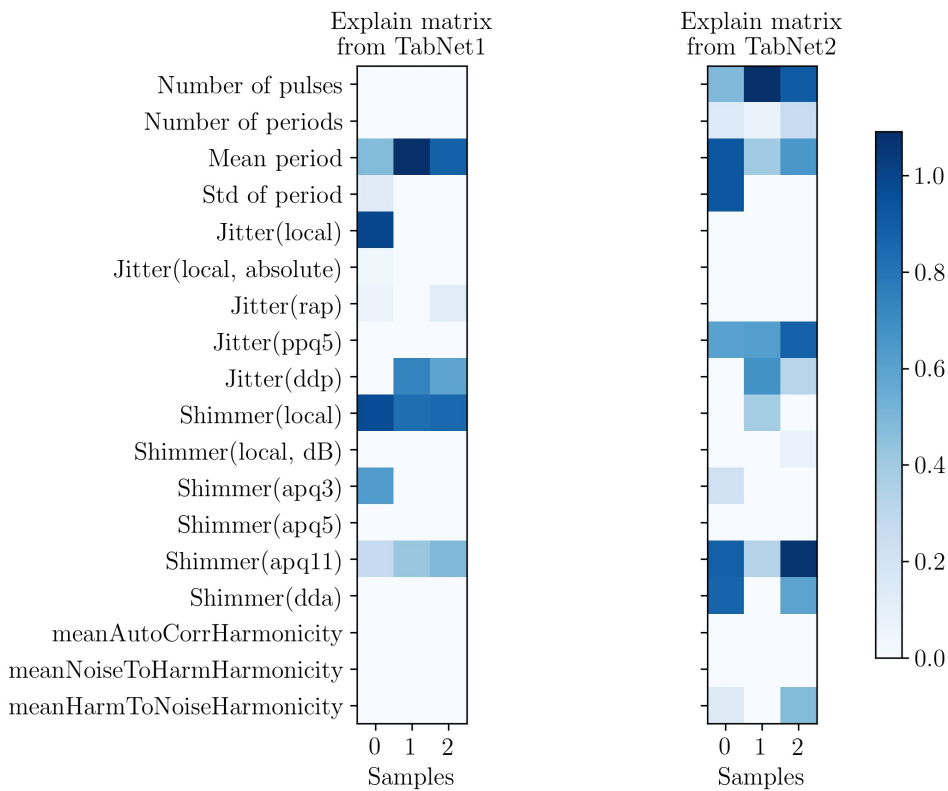
Το TabNet δημιουργεί κατά τη διάρκεια της εκπαίδευσής του μια μάσκα επιλογής χαρακτηριστικών για κάθε decision step, η οποία καθορίζει τα πιο σημαντικά χαρακτηριστικά τα οποία θα συμμετάσχουν στη λήψη αποφάσεων σχετικά με την ταξινόμηση. Το μοντέλο πραγματοποιεί επιλογή χαρακτηριστικών ανά δείγμα, πράγμα που σημαίνει ότι η μάσκα επιλέγει τον καλύτερο συνδυασμό χαρακτηριστικών για κάθε μεμονωμένο δείγμα σε αντίθεση με τις περισσότερες άλλες τεχνικές μείωσης διαστάσεων που λαμβάνουν αποφάσεις συνυπολογίζοντας τα χαρακτηριστικά όλων των διαθέσιμων δειγμάτων. Οι μάσκες που προκύπτουν σε κάθε decision step είναι διαθέσιμες στο τέλος της εκπαίδευσης, δίνοντας στο μοντέλο τοπική ερμηνευσιμότητα. Αθροίζοντας όλες τις μάσκες των decision steps, το TabNet δημιουργεί μια μήτρα που συνοψίζει τον βαθμό συμμετοχής κάθε χαρακτηριστικού στην πρόβλεψη του μοντέλου για κάθε ξεχωριστό δείγμα. Επιπλέον, η συνολική ερμηνευσιμότητα του μοντέλου επιτυγχάνεται μέσω γραμμικού συνδυασμού όλων των μασκών. Το TabNet ποσοτικοποιεί τη συνολική σημασία των χαρακτηριστικών στην ταξινόμηση λαμβάνοντας υπόψη τη σχετική σημασία κάθε βήματος στον τελική απόφαση.

Για να παρουσιάσουμε τα αποτελέσματα της ερμηνευσιμότητας του TabNet, επιλέξαμε μια αρχιτεκτονική δύο σταδίων απόφασης με τις τιμές των υπόλοιπων υπερ-παραμέτρων να είναι οι ίδιες με πριν (Πίνακας 1.1). Επιλέχθηκαν τυχαία δύο άτομα, ένα διαγνωσμένο με νόσο του Πάρκινσον (άτομο με ID 100) και ένα από την ομάδα ελέγχου (άτομο με ID 251). Επίσης, παρουσιάζουμε τα αποτελέσματα στην περίπτωση 18 αρχικών χαρακτηριστικών προκειμένου να διατηρήσουμε τα γραφήματα ευανάγνωστα. Οι Εικόνες 1.15, 1.16 και 1.19 αφορούν τον συμμετέχοντα No.100 και οι Εικόνες 1.17, 1.18 και 1.20 τον συμμετέχοντα No.251.

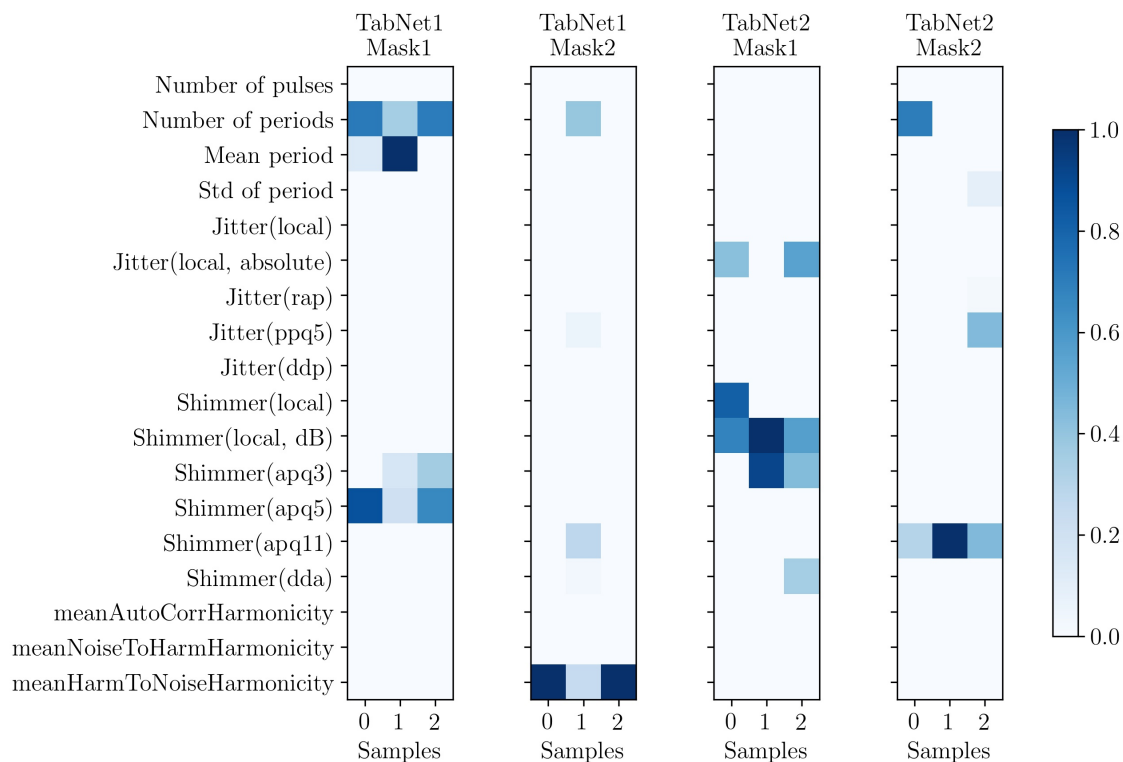
Είναι σημαντικό να σημειωθεί ότι το TabNet επιλέγει διαφορετικά χαρακτηριστικά με βάση τα validation sets που χρησιμοποιούνται για την εκπαίδευση των ensemble ταξινομητών. Επιπλέον, η συνεισφορά των χαρακτηριστικών στην τελική πρόβλεψη ποικίλλει σημαντικά, ανάλογα με το αν τα δείγματα εισόδου ανήκουν σε υγιές άτομο ή σε ασθενή με Πάρκινσον. Όλα τα δείγματα από αυτά τα δύο άτομα ταξινομήθηκαν σωστά από το TabNet.



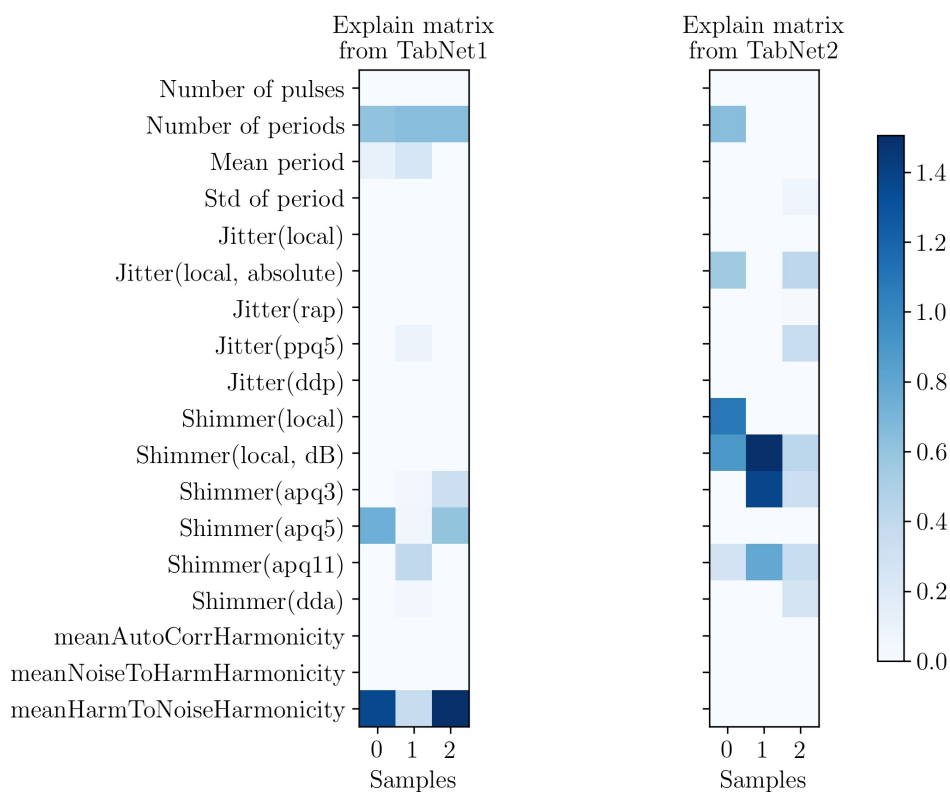
Εικόνα 1.15: Μάσκες επιλογής χαρακτηριστικών για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).



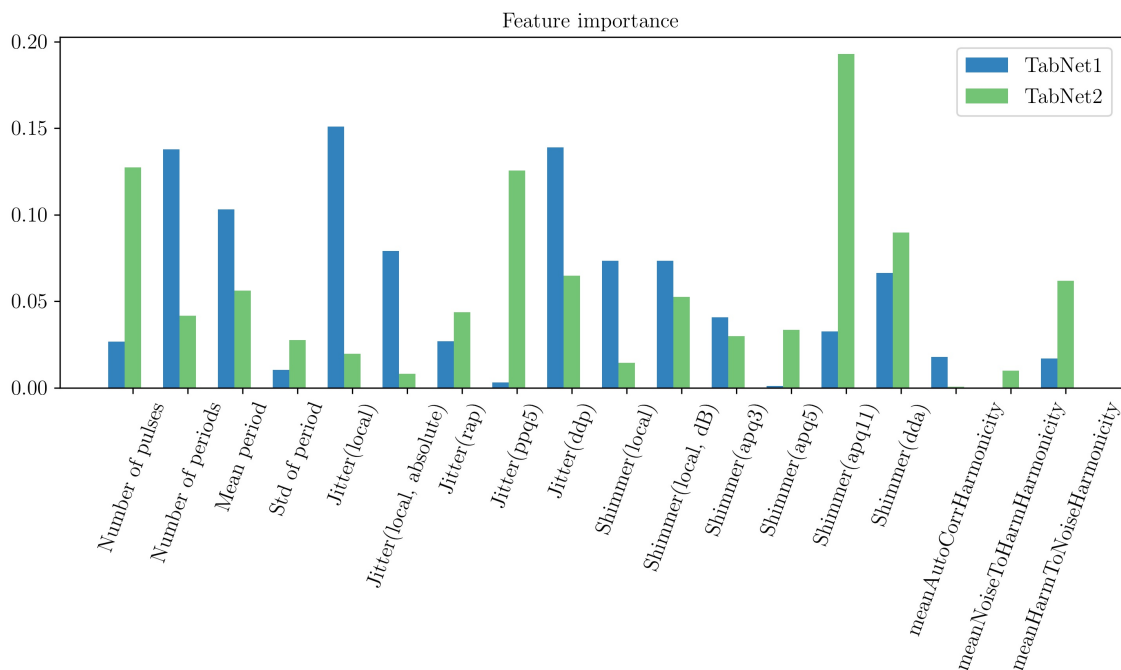
Εικόνα 1.16: Μήτρα επεξηγησιμότητας για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).



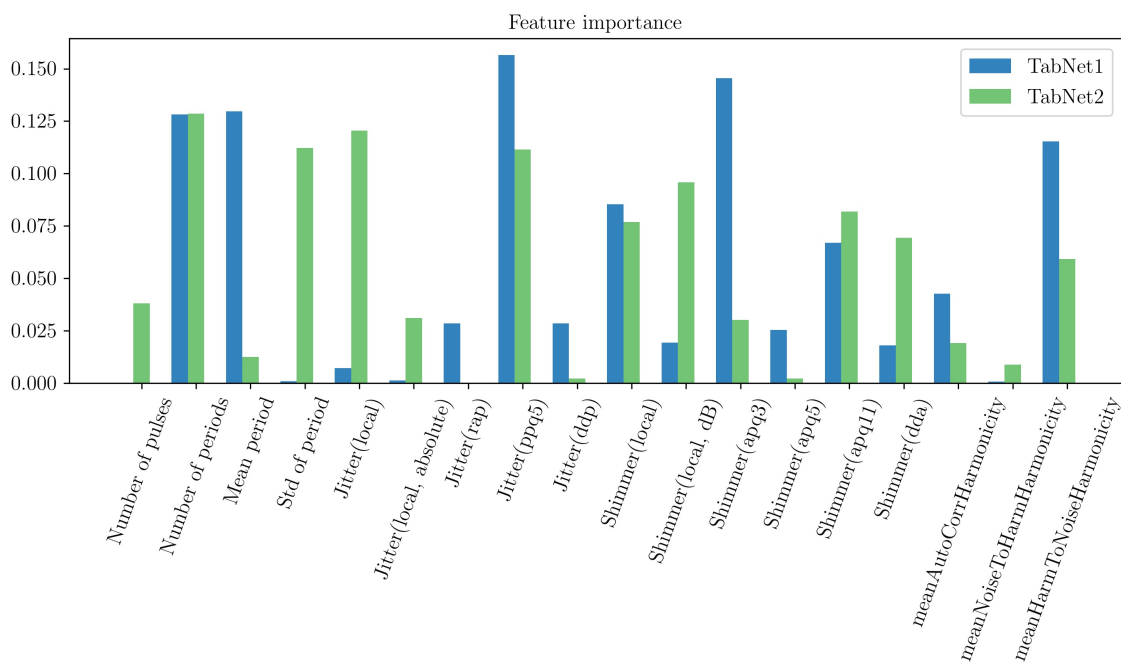
Εικόνα 1.17: Μάσκες επιλογής χαρακτηριστικών για τον υγί συμμετέχοντα (ID:251).



Εικόνα 1.18: Μήτρα επεξηγησιμότητας για τον υγί συμμετέχοντα (ID:251).



Εικόνα 1.19: Συντελεστές σημαντικότητας/συμμετοχής των χαρακτηριστικών στην ταξινόμηση για τον συμμετέχοντα που νοσεί από Πάρκινσον (ID:100).



Εικόνα 1.20: Συντελεστές σημαντικότητας/συμμετοχής των χαρακτηριστικών στην ταξινόμηση για τον υγιή συμμετέχοντα (ID:251).

1.5 Συζήτηση

Στην βιβλιογραφία περιλαμβάνονται πολυάριθμες μελέτες (Πίνακας 1.2) σχετικά με την ταξινόμηση ασθενών με Πάρκινσον και υγιών ατόμων με βάση τα χαρακτηριστικά που εξάγονται από τις ηχητικές καταγραφές φωνής. Σύμφωνα με τα

αποτελέσματα των μελετών, οι μέθοδοι που θα χρησιμοποιηθούν για την επιλογή των κυρίαρχων χαρακτηριστικών και την εκπαίδευση των μοντέλων σχετίζονται άμεσα με την ακρίβεια και την αξιοπιστία των προτεινόμενων συστημάτων. Το dataset [3], που χρησιμοποιήθηκε για τους σκοπούς της παρούσας διπλωματικής εργασίας, αποτελεί το μεγαλύτερο ελεύθερα προσβάσιμο σύνολο δεδομένων έως αυτή την στιγμή. Άλλα συχνά χρησιμοποιούμενα και διαθέσιμα σύνολα δεδομένων είναι το [25] που περιλαμβάνει συνολικά 195 δείγματα φωνής από 23 ασθενείς με Πάρκινσον και 8 υγιή άτομα, το [27] που αποτελείται από 20 ασθενείς με Πάρκινσον και 20 υγιή άτομα και το [26] με 40 ασθενείς και 40 υγιείς.[3]

Ορισμένες βιβλιογραφικές μελέτες έχουν παρουσιάσει μοντέλα ταξινόμησης που επιτυγχάνουν σωστή διάγνωση σε ποσοστά που προσεγγίζουν το 100%.[3]. Ο Tuncer και οι συνεργάτες του παρουσίασαν ένα νέο σύστημα εξαγωγής χαρακτηριστικών βασισμένο στη μέθοδο multiple pooling. Χρησιμοποίησαν πολλούς ταξινομητές για να αξιολογήσουν την ικανότητα διάκρισης των υγιών ατόμων από τους ασθενείς με Πάρκινσον, φτάνοντας ακρίβεια 97.62% χρησιμοποιώντας τον ταξινομητή ενός πλησιέστερου γείτονα (1NN).[28] Ο Olivares και οι συνεργάτες του χρησιμοποίησαν την μέθοδο Bat Algorithm in σε Extreme Learning Machine (BA-ELM), με αποτέλεσμα μέγιστη ακρίβεια προβλέψεων στο 96.7%.[29] Σε μια άλλη μελέτη χρησιμοποιείται η τεχνική SMOTE (Synthetic Minority Over-sampling Technique) και ο ταξινομητής Random Forest επιτυγχάνοντας 94.9% ακρίβεια στην ταξινόμηση.[30] Ο ταξινομητής AdaBoost χρησιμοποιήθηκε με grid search hyperparameter tuning και προ-επεξεργασία των χαρακτηριστικών με Standard Scaler, με την υψηλότερη ακρίβεια στο 94%.[31] Τέλος, στη βιβλιογραφία αναφέρεται επίσης ταξινόμηση με Wrappers feature subset selection που είχε ακρίβεια 94.7% με τον ταξινομητή SVM(RBF).[32].

Ωστόσο, παρόλο που αυτές οι μελέτες, μεταξύ άλλων, ανέφεραν πολύ υψηλά ποσοστά ταξινόμησης, οι τεχνικές cross-validation που χρησιμοποιήθηκαν από τους συγγραφείς οδηγούν σε αποτελέσματα που δεν είναι αμερόληπτα, εφόσον υπάρχουν δείγματα από το ίδιο άτομο ταυτόχρονα στο train, το validation και το test σετ. Η τήρηση του ενδεδειγμένου τρόπου χωρισμού των δειγμάτων για την εκπαίδευση (Leave-One-Subject-Out cross-validation) έχει αποδειχθεί ότι μειώνει δραματικά τις ικανότητες ταξινόμησης των μοντέλων.[33]

Ορισμένες δημοσιευμένες μελέτες ακολουθούν τις κατάλληλες μεθόδους, κατασκευάζοντας έτσι ένα αμερόληπτο μοντέλο τόσο για τα στάδια της εκπαίδευσης όσο και της αξιολόγησης του μοντέλου στο test σετ.[3] Η χρήση Gradient Boosted Tree (GBT) με βάρος ίσο ή μεγαλύτερο από 9, οδήγησε σε 92% accuracy, 94% F1-score και 78% MCC.[34] Ο Sakar και οι συνεργάτες του χρησιμοποίησαν επίσης LOSO cross-validation για την αξιολόγηση της ικανότητας γενίκευσης του μοντέλου τους, επιτυγχάνοντας 86% accuracy, 84% F1-score και 59% MCC με τον ταξινομητή SVM(RBF) και τον αλγόριθμο mRMR για την επιλογή των καλύτερων 50 χαρακτηριστικών.[3] Τέλος, ο H. Gunduz δημιούργησε ένα CNN 9 επιπέδων επιτυγχάνοντας 86.9% accuracy, 91.7% F1-score και 63.2% MCC.[35]

Είναι σημαντικό να σημειωθεί πως, δεδομένου ότι κάθε υποκείμενο της μελέτης παρείχε τρία δείγματα, οι περισσότερες μελέτες που χρησιμοποιούν LOSO cross-validation αναφέρουν αποτελέσματα βασισμένα στην πρόβλεψη του μοντέλου τους για κάθε συμμετέχοντα, η οποία υπολογίστηκε μέσω πλειοψηφίας των αποτελεσμάτων ταξινόμησης για τα τρία επιμέρους δείγματα κάθε ατόμου. Αυτό έχει ως αποτέλεσμα να αυξάνονται ελαφρώς οι τιμές των μετρικών που περιγράφουν τις ικανότητες γενίκευσης των μοντέλων σε άγνωστα δείγματα.

Πίνακας 1.2: Αποτελέσματα που παρουσιάζονται στη βιβλιογραφία για την ταξινόμηση ασθενών με Πάρκινσον και υγιών ατόμων με χρήση του dataset Sakar et al. (2019) [3] για την εκπαίδευση και την αξιολόγηση του επιλεγμένου μοντέλου.

Author, Year	Methods	Train/test split	Best results
Sakar et al.(2019) [3]	SVM(RBF)	LOSO cv	Accuracy = 0.86 F1 score = 0.84 MCC = 0.59
Solana-Lavalle et al. (2020) [32]	SVM(RBF) with Wrappers feature subset selection	10-fold cv	Accuracy = 0.95
Tuncer et al.(2019) [28]	INN classifier with octopus-based feature extraction network	10-fold cv	Accuracy = 0.98 F1 score = 0.98
Olivares et al.(2020) [29]	Bat Algorithm in the Extreme Learning Machine (BA-ELM)	80 – 20	Accuracy = 0.96
Polat(2019) [30]	SMOTE (Synthetic Minority Over-sampling Technique) and Random Forest	10-fold cv	Accuracy = 0.95 F1 score = 0.95
Anisha et al.(2020) [31]	AdaBoost with grid search hyperparameter tuning and preprocessing with Standard Scaler	10-fold cv	Accuracy = 0.94 F1 score = 0.94
Wijaya et al.(2019) [34]	Gradient Boosted Tree (GBT)	LOSO cv	Accuracy = 0.92 F1 score = 0.94 MCC = 0.78
Gunduz(2019) [35]	9-layered CNN	LOSO cv	Accuracy = 0.87 F1 score = 0.82 MCC = 0.63
Xiong et al.(2020) [36]	Adaptive Grey Wolf Optimization (GWO) with sparse autoencoder and Linear Discriminant Analysis(LDA)	10-fold cv	Accuracy = 0.95
Ashour et al.(2020) [37]	SVM(cubic), weighted-hybrid selected features that	70-30	Accuracy = 0.94
Yücelbaş(2020) [38]	Simple Logistic hybrid system(SLGS)	10-fold cv	Accuracy = 0.88 F1 score = 0.87 MCC = 0.63
Nissar et al.(2019) [39]	XGBoost with Minimum Redundancy - Maximum Relevance (mRMR) feature selection	not mentioned	Accuracy = 0.95 F1 score = 0.95
Bchir(2020) [40]	Gaussian mixture model (GMM) with mRMR feature selection	10-fold cv	Accuracy = 0.95 MCC = 0.70
Proposed approach	TabNet	LOSO cv	Accuracy = 0.97 F1 score = 0.98 MCC = 0.93

Σύμφωνα με τα αποτελέσματά της παρούσας διπλωματικής εργασίας, το TabNet υπερισχύει των μεθόδων που περιλαμβάνονται στη βιβλιογραφία, με μέση και μέγιστη ακρίβεια για την ταξινόμηση κάθε δείγματος να φθάνει το 94.54% και 95.24%, αντίστοιχα, όταν χρησιμοποιούνται όλα τα διαθέσιμα χαρακτηριστικά από στο σύνολο δεδομένων [3] και LOSO cross-validation. Οι μέσες και μέγιστες τιμές για τις μετρικές F1-score και MCC ήταν 96.31%, 96.77% και 85.97%, 87.83%, αντίστοιχα. Στην περίπτωση προβλέψεων για κάθε συμμετέχοντα, ο μέσος όρος και η μέγιστη τιμή που προέκυψε ήταν 95.93% και 97.22% accuracy, 97.26% και 98.12% F1-score, 89.43% και 92.83% MCC. Επιπλέον, η δυνατότητα του TabNet να επιλέγει εσωτερικά τα καλύτερα χαρακτηριστικά για την ταξινόμηση κάθε δείγματος προσθέτει ερμηνευσιμότητα στο μοντέλο, ιδιότητα η οποία είναι αναγκαία για τα μοντέλα τεχνητής νοημοσύνης ιδιαίτερα στον τομέα της υγείας και δεν περιλαμβάνεται στους ταξινομητές που αναφέρονται στην βιβλιογραφία.

1.6 Σύνοψη και μελλοντικές επεκτάσεις

1.6.1 Σύνοψη

Στα πλαίσια της παρούσας διπλωματικής εργασίας διερευνήθηκε η δυνατότητα χρήσης αρχιτεκτονικών βαθιάς μάθησης για την διάγνωση της ασθένειας του Πάρκινσον από χαρακτηριστικά που έχουν εξαχθεί από ηχητικές καταγραφές φωνής. Το Pytorch-TabNet, το οποίο εκπαιδεύτηκε με τα διαθέσιμα δεδομένα και την τεχνική LOSO cross-validation, σημείωσε καλύτερες επιδόσεις από όσους ταξινομητές αναφέρονται στην βιβλιογραφία και ακολούθησαν τις ενδεδειγμένες μεθόδους εκπαίδευσης.

1.6.2 Μελλοντικές επεκτάσεις

Ένα από τα σημαντικότερα ζητήματα που δυσχεραίνουν την ανάπτυξη συστημάτων μηχανικής μάθησης για την διάγνωση της ασθένειας του Πάρκινσον είναι η έλλειψη δεδομένων εκπαίδευσης, με το μεγαλύτερο σετ δεδομένων που είναι διαθέσιμο αυτή τη στιγμή να περιλαμβάνει λιγότερα από 800 δείγματα που έχουν ληφθεί από 252 άτομα. Το γεγονός αυτό καθιστά απαγορευτική την χρήση πιο σύνθετων αρχιτεκτονικών, εφόσον αυτές απαιτούν μεγαλύτερο αριθμό δεδομένων για την εκπαίδευσή τους. Προκειμένου να χρησιμοποιηθούν μοντέλα με μεγαλύτερες δυνατότητες είναι απαραίτητη η συλλογή περισσότερων δεδομένων τόσο από ασθενείς που πάσχουν από την ασθένεια του Πάρκινσον όσο και από υγιή άτομα.

Αναφορικά με την μέθοδο εκπαίδευσης του TabNet, η εξασφάλιση πιο ισχυρών υπολογιστικών συστημάτων είναι απαραίτητη ώστε να ενσωματωθεί η ρύθμιση των υπερ-παραμέτρων του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Χρησιμοποιώντας τη μέθοδο grid search για την επιλογή των βέλτιστων τιμών για τις επιλεγμένες υπερ-παραμέτρους του TabNet είναι πιθανό να προκύψουν καλύτερα αποτελέσματα ταξινόμησης. Παράλληλα είναι απαραίτητο να διερευνηθεί η επίδραση της τυχαίας επιλογής της παραμέτρου seed των μοντέλων όταν το validation set παραμένει σταθερό κατά την διάρκεια ενός πλήρους κύκλου εκπαίδευσης του ταξινομητή με LOSO cross-validation.

Στις μελλοντικές επεκτάσεις της παρούσας έρευνας μπορεί να συμπεριληφθεί ο σχεδιασμός συστήματος διάγνωσης της νόσου του Πάρκινσον βασισμένου σε ένα σύνολο δεδομένων από διαφορετικές πηγές. Αναλυτικότερα, τα χαρακτηριστικά ηχητικών καταγραφών φωνής μπορούν να συνδυαστούν με δεδομένα

από αισθητήρες που καταγράφουν τη βάρδηση των υποκειμένων καθώς και τον τρόπο κατά την εκτέλεση συγκεκριμένων εργασιών ώστε να προκύψει ένα πιο γενικευμένο σύστημα ταξινόμησης. Τέλος, το TabNet μπορεί να εκπαιδευτεί με χρονοσειρές από βιομετρικά δεδομένα που έχουν συγκεντρωθεί από ασθενείς με Πάρκινσον κατά τη διάρκεια ενός δεδομένου χρονικού διαστήματος. Το μοντέλο βαθιάς μάθησης που θα προκύψει από αυτά τα δεδομένα θα μπορεί να επεκτείνει την απλή δυαδική ταξινόμηση των νέων δειγμάτων σε εκτίμηση του σταδίου που βρίσκονται τα άτομα που ταξινομούνται ως ασθενείς ή να αξιοποιηθεί για την καταγραφή και την παρακολούθηση της εξέλιξης της νόσου (telemonitoring device).

Chapter 2

Parkinson's disease

2.1 Introduction

In neurology, neurodegenerative diseases are characterized by the progressive and irreversible loss of neurons located in specific brain regions. According to studies, aging is the leading cause of these disorders, with the main scientifically proven factors being damage to the nuclear DNA of neurons, increased oxidative stress, and chronic neuroinflammation.[1] Parkinson's Disease (PD) was first described by James Parkinson in 1817 in his publication "Essay on the Shaking Palsy".[2, 4] PD is a progressive neurodegenerative disease caused by necrosis of dopaminergic neurons in substantia nigra, a brain region responsible for synthesizing the neurotransmitter dopamine.[1, 2] It's the second most common neurodegenerative disease in the elderly, after Alzheimer's disease, affecting men and women of all races and social classes.[1, 3]

2.2 Epidemiology

Incidence estimates of Parkinson's disease around the world range from 5 to >35 new cases per 100.000 individuals every year. The variance of the annual number of PD diagnoses may reflect the underlying differences in the demographics of the populations which participated in the studies or the study methods. Parkinson's disease is rare in people under 50 years old, although the incidence increases 5-10 fold from the sixth to the ninth decade of life.[2] The disease afflicts nearly 1% of the population above the age of 60.[9] Young-onset PD is usually defined by the age of onset less than 45 years, and more than 10% of those individuals have a genetic basis. Additionally, the proportion of genetically defined cases increase to >40% when the first appearance of the symptoms occurs before 30 years of age. The global prevalence is conservatively estimated at 0.3% overall, and it increases to >3% in older people (with age above 80 years).[2]

Mortality due to Parkinson's is not increased in the first decade after the disease onset. Following the mentioned period, the mortality increases and eventually doubles compared to the general statistics. The improvement of the health care system has led to prolonged survival, and therefore the increase of prevalence of Parkinson's disease over time in a 20-year study. The number of people living with PD is expected to double between 2005 and 2030.[2, 4] In the two decades between 1990 and 2010, the years lived with disability and disability-adjusted life years due to PD increased. As a result, as the world population ages, the personal, societal and economic burden

associated with Parkinson's disease is expected to progressively increase.[2, 4]

In the majority of populations, Parkinson's disease is twice as common in men as in women, except for few populations, including one study from Japan, where no difference or a slight female excess was observed. The protective effect of female sex hormones, a sex-associated genetic mechanism or sex-specific differences when a person is exposed to environmental risk factors might explain the male preponderance observed in PD. The incidence also varies within subgroups defined by race, ethnicity, genotype or environment. Heritable forms of Parkinson's disease are estimated to be only 5-10% of all cases.[2, 5] The interactions between environment and genes modify the risk for sporadic PD.

The incidence of Parkinson's disease increases significantly in individuals exposed to certain environmental factors, like pesticides and traumatic brain injury, and it is lower in smokers or caffeine users.[2, 4, 5] Greater dietary intake of dairy products has also been linked to a higher risk of Parkinson's disease, which might result from the concentration of toxicants in milk.[5] On the contrary, physical activity and healthy diets with high amounts of fruits, vegetables and grain are associated with a lower risk of PD.[5]

2.3 Pathophysiology

The loss of pigmented dopaminergic neurons in specific areas of the substantia nigra pars compacta (SNc) in the midbrain and the presence of Lewy bodies and Lewy neurites in the remaining neurons are the main characteristic features of Parkinson's disease.[2, 4, 6, 9] Even though these two neuropathologies are not specific to PD, they are crucial for a definite diagnosis of the disease.[2]

Lewy bodies and Lewy neurites are defined as abnormal protein inclusions, mainly composed of insoluble α -synuclein fibrils that accumulate in neurons in PD.[2, 4, 8] In comparison with classical Lewy bodies found in the remaining dopaminergic neurons of the SNc that are described as intraneuronal, round, eosinophilic inclusions with a hyaline core and a pale peripheral halo that are always positive for α -synuclein and ubiquitin, Lewy neurites are abnormal neurites that contain granular material and α -synuclein filaments, similar to those found in Lewy bodies.[4] The Lewy pathology at first appears in cholinergic and monoaminergic brainstem neurons and neurons in the olfactory system.[2, 41] During the disease's progression, it can also be found in limbic and neocortical brain regions.[2] Lewy bodies seem to have a greater impact on cellular physiology since they appear in later stages of PD.[4] Lewy neurites may disrupt axonal transport and other vital cellular processes and therefore compromise neural function and survival.[4]

At the early stages of PD, the loss of dopaminergic neurons is mainly restricted to the ventrolateral substantia nigra, but becomes more widespread by end-stage disease.[2, 9] According to several recent clinicopathological studies, the loss of dopaminergic neurons at the early stages of the disease suggests that the degradation in this region of the brain begins before the onset of motor symptoms.[2]

Mitochondrial dysfunction appears to be essential in the pathogenesis of Parkinson's disease.[2, 9] The cellular changes observed together in degenerating neurons in PD could be explained by the existence of a vicious cycle in which mitochondrial dysfunction and α -synuclein aggregation exacerbate each other.[2] Mitochondrial dysfunction results in an increase of oxidative stress in the brain of PD patients, even though it's still debatable whether it occurs early or late during the necrosis of

neurons.[2] According to research, nigral dopaminergic neurons are vulnerable to metabolic and oxidative stress for several reasons, including their long, unmyelinated axons and their autonomous pacemaking activity.[2]

Neuroinflammation is also a salient feature of PD, which may not be the initial trigger of the disease but is considered a vital contributor to pathogenesis.[2]

2.4 Diagnosis

Parkinsonism is the result of decreased dopaminergic transmission in the motor region of the striatum with opposing effects on the direct and indirect pathways.[2] Clinically, Parkinson's disease is defined by the presence of bradykinesia and at least one more cardinal motor feature (rigidity, rest tremor), with additional supporting and exclusionary criteria.[2] Motor symptoms also include lower-limb stiffness, weakness and heaviness with a gait pattern characterized by small, shuffling steps and therefore reduced velocity. As the disease progresses, gait becomes frequently interrupted by start hesitation and freezing, especially when encountering doorways or obstacles.[6] The symptoms are typically unilateral at the beginning of motor dysfunction, and the asymmetry persists during the disease.[2]

Except for cardinal motor features, most PD patients also experience non-motor symptoms, which involve a variety of functions, including sleep disorders (sleep-wake cycle regulation), cognitive impairment (frontal executive dysfunction, memory retrieval deficits, dementia, hallucinosis), disorders of mood and affect, autonomic dysfunction (orthostatic hypotension, urogenital dysfunction, constipation, hyperhidrosis), sensory symptoms (hyposmia) and pain.[2] Some non-motor symptoms appear years or even decades before the onset of classic motor ones.[2]

In a long-term study with subjects that had been diagnosed with Parkinson's disease over 20 years ago, dementia was present in 83%, hallucinosis in 74%, symptomatic orthostatic hypotension in 48%, constipation in 40%, and urinary incontinence in 71% of those patients.[2] At later stages of the disease, PD features included treatment-resistant motor symptoms, such as freezing of gait (81%), postural instability and falling (87%, with fractures in 35%) and choking (48%).[2]

Although the clinical diagnosis of patients with classic motor features of Parkinson's disease is relatively straightforward, error rates for the clinical diagnosis approach 20% at an early stage of the disease, even in specialized centers.[2] Clinical trials and observational studies mainly focus on the progression of motor impairment as classified by the Unified Parkinson's Disease Rating Scale (UPDRS), which is the most common scale for monitoring motor disability associated with PD.[2] The accuracy of clinical diagnoses of PD can be improved by the strict use of standard clinical criteria, such as the UK Parkinson's Disease Society Brain Bank (UKPDSBB).[2] Available diagnostic tests that can be used for PD diagnosis include imaging (visualization of striatal dopamine depletion with PET, SPECT or MRI techniques), genetic tests, and cerebrospinal fluid and blood tests.[2]

2.4.1 Vocal disorder based diagnosis

The vocal disorders associated with Parkinson's disease are considered critical since they usually precede motor symptoms in 90% of PD patients at the early stages of the disease. The most common vocal disorders include dysphonia (defective voice), hypophonia (reduced volume), monotone (reduced pitch) and dysarthria (difficulty with articulation).[3, 32] Vocal features are often used for speech-based teliagnosis

of PD using machine learning and signal processing techniques, because the method is non-invasive, low in cost and sufficiently accurate.[32] The most common ones are vocal features known as baseline features that include jitter, shimmer, fundamental frequency parameters, harmonicity parameters, recurrence period density entropy (RPDE), detrended fluctuation analysis (FDA) and pitch period entropy (PPE).[3, 32]

2.5 Management

As mentioned before, the loss of dopaminergic neurons in the substantia nigra pars compacta that results in striatal dopamine depletion is the primary mechanism underlying the cardinal motor features of PD.[2] At the moment, there are not any available neuroprotective or disease-modifying therapies for Parkinson's disease.[8] Therefore, the main treatment for Parkinson's disease aims to substitute striatal dopamine loss with systemic administration of the dopamine precursor amino acid L-DOPA or dopamine agonists.[2, 9] It is believed that if the treatment begin as early as possible, interventions that aim to slow or stop neurodegeneration will be more effective.[8] According to clinical observations, postural instability and gait disorder in early-mid PD can be efficiently treated with dopaminergic therapy.[6] Some additional drugs are also prescribed to reverse the side effects of L-DOPA, such as dyskinesia.[2]

In some PD patients, a variety of non-motor symptoms (pain, anxiety, panic, depression, restlessness) appear to fluctuate due to dopaminergic therapy and therefore be equally or even more disabling than motor symptoms.[2] Non-motor issues like cognitive dysfunction, depression and autonomic failure are treated with antipsychotic and antidepressant medications, such as clozapine.[2] Even though surgical treatments are considered effective in reducing motor fluctuations, dyskinesia, tremor and rigidity, the improvement in gait and postural stability varies and is also less durable.[6] Since 1993, Deep Brain Stimulation (DBS) is also used as a treatment for PD. High-frequency electrical stimulation is used in specific brain targets. The treatment also involves the implantation of an electrode in brain tissue.[2]

Most patients with Parkinson's disease have to deal with residual motor disabilities that affect gait, mobility, postural control, balance, speech and swallowing functions, which are usually insufficiently responsive to drugs and DBS.[2] These symptoms seem to improve with various exercise-based strategies.[2, 6]

Chapter 3

Deep Learning

3.1 Introduction

AI has been one of the oldest fields in computer science.[10] By definition, artificial intelligence (AI) is “a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behavior, and with the creation of artifacts that exhibit such behavior”.[11, 12] Alan Turing is considered one of the founders of modern computer science and AI, as he defined intelligent behavior in a computer as the ability to achieve human-level performance in cognitive tasks, later referred to as the “Turing test”.[12]

Diagnosis, treatment, and predicting the outcome of various clinical situations depend on a combination of many clinical, biological and pathological variables. This fact has led to a growing need for using analytical tools in medicine.[12] For over 50 years, researchers have been exploring the potential applications of AI in all fields of medicine.[12] Medical artificial intelligence is used to develop AI programs to help clinicians formulate a diagnosis, make therapeutic decisions and predict an outcome.[12] These systems, including Artificial Neural Networks (ANNs), fuzzy expert systems, evolutionary computation and hybrid intelligent systems, aim to support healthcare workers in everyday tasks related to the manipulation of data and knowledge.[12] Gunn was the first who successively investigated the application of AI in surgery. Specifically, he attempted to diagnose acute abdominal pain with computer analysis.[12] Furthermore, more than 100 studies have been published since the outbreak of the SARS-Cov-2 virus pandemic using Artificial Intelligence to screen, diagnose and make a prognosis on the coronavirus disease. [42]

3.2 Artificial Neural Networks

ANNs are computational analytical tools, the concept of which was based on the human biological nervous system and the structure of the brain.[12, 13] Neural Networks are nowadays used in a wide variety of problems like pattern recognition, classification, regression, clustering, dimensionality reduction, computer vision, natural language processing (NLP) and predictive analysis.[13] The prototype of the now known as neural networks was created in 1957 by Frank Rosenblatt, and it was named perceptron.[13]

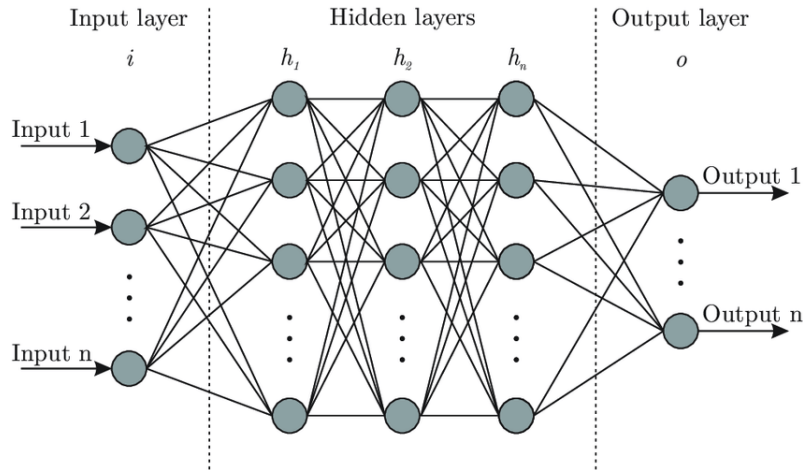


Figure 3.1: Artificial neural network architecture.[43]

ANNs are built by networks of highly interconnected processing units, known as “neurons”.[12, 13] Neurons can perform parallel computations related to data processing and knowledge representation.[12] These processing units are organized in input, hidden, and output layers, while every neuron is connected to other neurons in adjacent layers via a corresponding weight value.[13] The sum is then transformed as it passes through a selected activation function, which usually is a sigmoid, tan hyperbolic, or rectified linear unit (ReLU) function.[13] The aforementioned functions are easily differentiable. This property makes them ideal for easy calculation of partial derivatives of the error delta with respect to individual weights.[13] Activation functions transform the input value into a narrow output range, usually $[0,1]$ or $[-1,1]$.[13] The output of every activation function is then passed as input to the subsequent unit in the following layer.[13] The output of the last layer (output layer) is considered as the solution to the initial problem.[13]

The abilities of ANNs, such as learning from historical examples, analyzing non-linear data, generalizing, classifying and recognizing patterns, have made them an alluring analytical tool in medicine.[12] ANNs have been successfully applied in clinical diagnosis, image analysis in radiology and histopathology, data interpretation in intensive care setting, diagnosis of carotid atherosclerosis and waveform analysis.[12, 14, 15, 16]

3.3 Deep Neural Networks

Deep Neural Networks (DNNs) are a type of multilayer perceptrons (MLP), meaning that they consist of a higher or deeper number of processing layers.[13] DNNs are trained with algorithms to learn representations from input data without the need for any manual design of feature extractors, which happened in earlier machine learning techniques.[13] The advent of DNNs made it possible to map complex, non-linear functions which shallow neural networks could not efficiently map.[13]

The most significant breakthrough of DNNs happened in the 1970s with the invention of the backpropagation learning algorithm, although the method was not widely used nor understood until the mid-1980s.[13] The evolution of DNNs was also significantly supported by the proliferation of cheaper processing units, like general-purpose graphic processing unit (GPGPU), and the availability of large training data sets (big data).[13]

In recent times, deep learning has been among the most significant developments in computer science.[13] Deep learning has already surpassed human-level capability and performance in many fields, including predicting movie ratings, decisions related to approving loan applications, time taken by car delivery, etc.[13] It also seems to have the ability to improve human lives, providing more accurate diagnosis of diseases like cancer, discovering new drugs, or predicting natural disasters.[13] Esteva et al. reported that a deep neural network had the ability to learn from 129.450 images of 2.032 different diseases and diagnosed at the same level as 21 board-certified dermatologists.[17] Additionally, Google AI managed to exceed the average accuracy of US board-certified general pathologists in grading prostate cancer by 9%.[18]

3.4 TabNet

In 2019, S. Arik and T. Pfister from Google Cloud AI proposed a new canonical DNN architecture for tabular data that can be used for classification or regression, called TabNet. TabNet inputs raw tabular data without any previous preprocessing and it's trained using gradient descent-based optimization and end-to-end learning. TabNet's design uses DNN building blocks to implement Decision Tree - like output manifold.[19]

The model uses sequential attention and performs soft feature selection to obtain decision boundaries in hyperplane form. More specifically, the output is a linear combination of the input features with coefficients that determine the proportion of each feature in the final result. TabNet uses sparse instance-wise feature selection, with controlled sparsity learned from data, and constructs a sequential architecture with one or more steps that jointly determine the final decision and mimic ensembling. In this way, the model selects the best features for every individual input, unlike many other methods.[19]

With instance-wise feature selection, the model enables interpretability and the learning capacity of the model is improved by the non-linear processing of the selected features. The model uses the raw numerical features (in tabular form) without any previous preprocessing and includes mapping of categorical features with trainable embeddings. It also performs only batch normalization (BN) and not global feature normalization.[19]

TabNet uses Xavier Initialization to initialize all of its weights so that the variance of the activations is the same across every layer, using the uniform distribution in the interval $\left[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right]$, where n is the size of the previous layer. This constant variance helps prevent the gradient from exploding or vanishing.[20]

3.4.1 TabNet Encoder architecture

As mentioned before, TabNet is based on sequential multi-step processing. The encoder consists of N_{steps} decision steps where the i^{th} step decides the salient features that should be used in the decision, with information from the $(i - 1)^{th}$ step, and produces a processed feature representation that is later aggregated into the final output. The sequential form of top-down attention is inspired by similar applications applied in visual and text data and reinforcement learning while searching for subsets of relevant information in bigger datasets.[19]

TabNet’s encoder mainly consists of a Feature transformer, an Attentive transformer and a trainable feature Mask at each decision step. Every decision step inputs the same $B \times D$ feature matrix, where B is the batch size and D is the dimension of the features.[19]

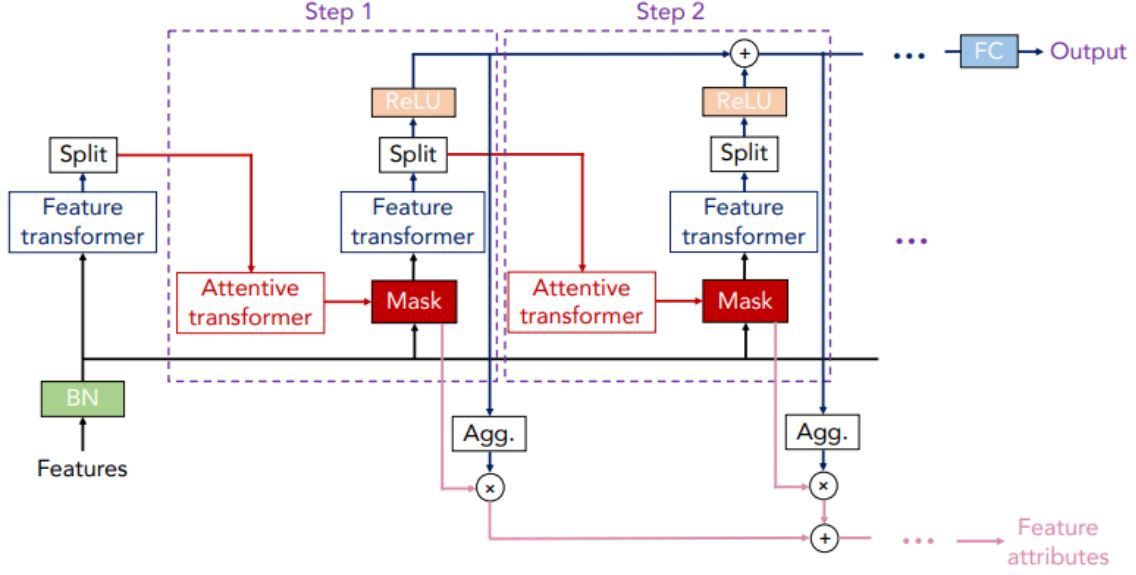


Figure 3.2: TabNet Encoder architecture.

Every batch $X \in \mathbb{R}^{B \times D}$ of the input feature matrix, given as input to TabNet is multiplied with the Mask $M[i]$ of the current decision step and then passed to the Feature transformer for the necessary processing, denoted as f_i . The output tensor of the Feature transformer is sent to the Split module, where it is divided into two separate tensors which contain a different representation of the initial features. The first part, $d[i] \in \mathbb{R}^{B \times n_d}$, is passed through a ReLU (or another activation function) and then is given as the output of the current decision step. The second one, $a[i] \in \mathbb{R}^{B \times n_a}$, is used as input to the Attentive transformer of the following decision step ($i + 1$) to construct its Mask, $M[i + 1]$. [19]

$$[d[i], a[i]] = f_i(M[i] \times X) \quad (3.1)$$

The output tensor of the encoder, which is the aggregation of the outputs of the decision steps, contains the processed features that will determine the overall decision, following the logic of the Decision Trees.[19]

$$d_{out} = \sum_{i=1}^{N_{steps}} ReLU(d[i]) \quad (3.2)$$

Finally, a linear mapping $W_{final}d_{out}$ is applied with a Fully Connected layer in order to get the final output of TabNet. Several layers can be added in the output, creating a neural network with any number of hidden layers based on the requirements of the problem to be solved.[19]

3.4.1.1 Feature processing

As shown in Figure 3.3, the Feature transformer block consists of layers. In order to achieve robust learning with high capacity and parameter-efficiency, the Feature transformer comprises one or more layers that are shared across all decision steps, since every decision step inputs the same feature matrix, and decision step-dependent layers. Each layer is composed of a fully connected (FC) layer, a batch normalization (BN) layer and a Gated Linear Unit (GLU) layer. Every layer is also connected to a normalized residual connection with normalization with the factor $\sqrt{0.5}$ that helps in stabilizing the learning process and ensuring that the variance throughout the network doesn't change significantly. For faster training, a large batch size is used. Additionally, except for the BN applied to the input features, TabNet uses ghost batch normalization, using a virtual batch size B_V and momentum m_B . [19]

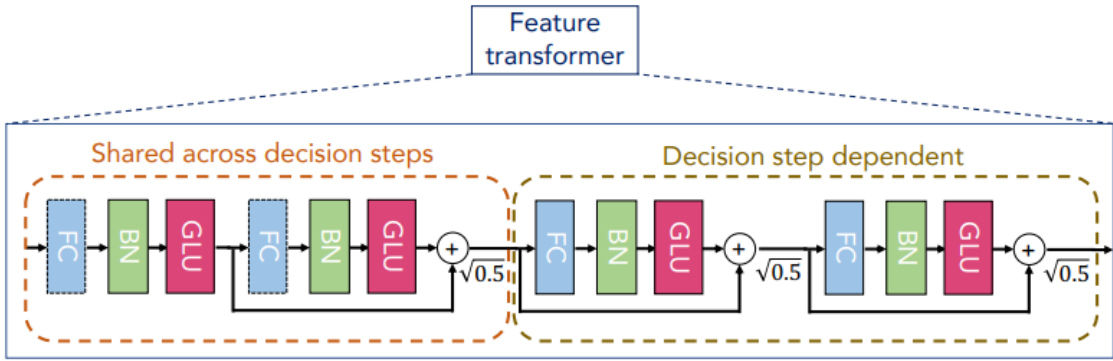


Figure 3.3: Topological structure of the Feature transformer block.

In the first shared layer of the Feature transformer, the FC layer maps, with a trainable linear transformation, the input tensor $X \in \mathbb{R}^{B \times D}$ to a new tensor $X \in \mathbb{R}^{B \times 2(n_d + n_a)}$ that is passed to the BN layer for the normalization and then to the GLU layer. Inside the GLU layer (eq. 3.3), the second half of the tensor acts as a gate that controls (by element-wise multiplication) the information passed from its first half to the output, $X \in \mathbb{R}^{B \times (n_d + n_a)}$, that is given as input to the following layers that function in the same way. All the fully connected layers of the Feature transformer except the first one, as expected, have $(n_d + n_a)$ inputs and $2(n_d + n_a)$ outputs. [19]

$$GLU(x) = x[:, : (n_d + n_a)] \otimes \sigma(x[:, (n_d + n_a) :]) \quad (3.3)$$

Gated Linear Units reduce the vanishing gradient problem that occurs when implementing deep architectures with other models, like LSTMs, by having linear units coupled to the gates. Unlike other recurrent neural networks, GLUs have no temporal dependencies, so they simplify parallelization. The gradient of the gated linear unit (eq. 3.4) has a path $\nabla X \otimes \sigma(X)$ without downscaling for the activated gating units in $\sigma(X)$, which can be interpreted as a multiplicative skip connection that helps gradients flow through the sequential layers. [21]

$$\nabla[X \otimes \sigma(X)] = \nabla X \otimes \sigma(X) + X \otimes \sigma'(X) \nabla X \quad (3.4)$$

3.4.1.2 Feature selection

The selection of the salient features at each step is realized by its Mask, $M[i]$. The Attentive transformer of the i^{th} decision step selects the best features by leaning

a mask. More specifically, the tensor $a[i-1]$ from the Split module of the $(i-1)^{th}$ step that is given as input to the Attentive transformer, is passed through the h_i layer, representing a fully-connected layer and a BN layer. The h_i layer makes a linear combination of the previously processed features by mapping $a[i-1] \in \mathbb{R}^{B \times n_a}$ to a new higher-dimensional and sparser tensor with dimensions $B \times D$. The output of the h_i layer is multiplied by the Prior scale, $P[i-1]$, of the previous decision step. The Prior scale is a representation of the usage of every feature in the preceding steps. The desired mask, $M[i]$, is generated with the Sparsemax() function. Equation 3.5 represents the mentioned procedure.[19]

$$M[i] = \text{Sparsemax}(P[i-1] \times h_i(a[i-1])) \quad (3.5)$$

Sparsemax maps the Euclidean projection into the probabilistic simplex. This easily evaluated and cheaply differentiable function can return sparse posterior distributions, which means it can assign precisely zero probability to some of its output variables. This property makes Sparsemax suitable for use as a filter to predict multiple labels or as a component to determine a group of variables or features that may be relevant to a decision, resulting in a more interpretable model. Sparsemax can also be used as a loss function.[22] Sparsemax generates weights for each feature, j , for each sample, b , and makes the weights for each sample sum to 1, $\sum_{j=1}^D M_{b,j}[i] = 1$, thus realizing instance-wise feature selection of the most beneficial features for every decision step. TabNet uses the sparse regular term L_{sparse} to control the sparsity of the selected features, given by the equation:

$$L_{sparse} = \sum_{i=1}^{N_{steps}} \sum_{b=1}^B \sum_{j=1}^D \frac{-M_{b,j}[i] \log(M_{b,j}[i] + \epsilon)}{N_{steps} \cdot B} \quad (3.6)$$

where ϵ is a small number for numerical stability. The sparsity regularization is added to the overall loss with a coefficient λ_{sparse} . [19]

The prior scale term $P[i]$ denotes the usage of each feature and is updated at every decision step by equation 3.7, where γ is a relaxation parameter. Setting $\gamma = 1$ enforces a feature to be used at precisely one decision step, while values greater than one provide more flexibility to the feature selection process. $P[0]$ is initialized as all ones.[19]

$$P[i] = \prod_{j=1}^i (\gamma - M[j]) \quad (3.7)$$

$$P[0] = 1^{B \times D}$$

The trained mask $M[i]$ is multiplied with the input tensor $X \in \mathbb{R}^{B \times D}$ in order to obtain a sparse tensor with the selected features of the current decision step, which is then inputted to the Feature transformer.[19]

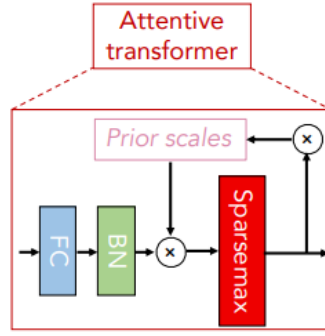


Figure 3.4: Topological structure of the Attentive transformer block.

3.4.2 Interpretability of the model

The feature selection mask of every decision step determines the most important features that participate in the decision-making. The value of the weight $M_{b,j}[i]$ indicates the contribution of the j^{th} feature of the b^{th} sample, $f_{b,j}$, with the features with zero weights being ignored. The final trained mask of every decision step is available, giving the model local interpretability. Due to the non-linear processing of the features in every decision step, an additional process is needed in order to estimate the feature importance of $f_{b,j}$. The global interpretability of the model is achieved by a linear combination of the masks of different steps. TabNet quantifies an aggregate feature importance which weights the relative importance of each step in the final prediction using the coefficient $\eta_b[i] = \sum_{c=1}^{n_d} \text{ReLU}(d_{b,c}[i])$ to denote the aggregate decision contribution at the i^{th} decision step for the b^{th} sample. When $d_{b,c}[i]$ has only negative values, no features at the i^{th} decision step contribute to the overall decision. The aggregate importance mask is derived from scaling the decision mask at each step with the coefficient $\eta_b[i]$ (eq. 3.8). Normalization for $M_{\text{agg}-b,j}$ is also used to ensure that $\sum_{j=1}^D M_{\text{agg}-b,j} = 1$. [19]

$$M_{\text{agg}-b,j} = \frac{\sum_{i=1}^{N_{\text{steps}}} \eta_b[i] M_{b,j}[i]}{\sum_{j=1}^D \sum_{i=1}^{N_{\text{steps}}} \eta_b[i] M(b, j)[i]} \quad (3.8)$$

3.4.3 TabNet Decoder architecture

The encoded representation of the initial input matrix is the aggregated vector of the encoder decision steps' outputs without the Fully Connected layer, and it's used as the input of the Decoder. The Decoder uses the Feature transformer of each decision step to reconstruct its representation vector into features and finally outputs the reconstructed input matrix. [19]

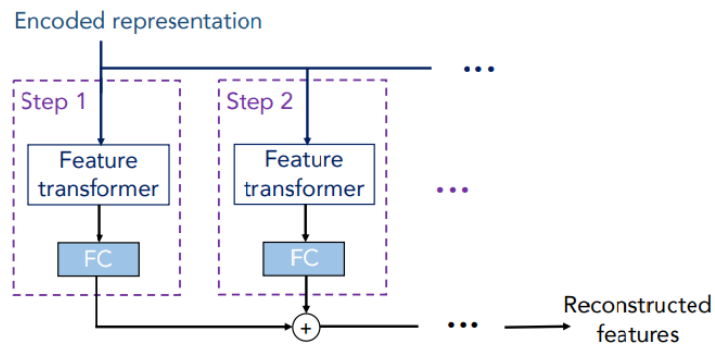


Figure 3.5: TabNet Decoder architecture.

Chapter 4

Materials and methods

4.1 Data

The model was trained with data available in the UCI Machine Learning Repository of the University of California Irvine.[23] The dataset consists of extracted features from voice recordings from 188 patients with Parkinson’s disease (107 men and 81 women) with ages between 33 to 87 (65.1 ± 10.9), gathered at the Department of Neurology in Cerrahpaşa Faculty of Medicine, Istanbul University. Furthermore, 64 healthy individuals (23 men and 41 women) aged 41 to 82 years (61.1 ± 8.9) were included in the control group.

Table 4.1: Extracted features from spectrograms of speech signals.[3, 32]

Feature set	Description	Total features
Baseline features	Jitter and shimmer variants. Mean, median, standard deviation, minimum and maximum values of the vocal fold vibration frequencies. Harmonicity parameters. Recurrence period density entropy, detrended fluctuation analysis, and pitch period entropy.	32
Time-frequency features	Mean, minimum and maximum speech signal power in dB. First four formant frequencies in the vocal tract. First four bandwidths of formant frequencies.	11
MFCCs	Mean and standard deviation of the original 13 MFCCs and their first-second derivatives. Log- energy of the signal.	84
Wavelet transform	Discrete wavelet transform extracted at ten levels. The Shannon’s and the log energy entropy are computed. The Teager–Kaiser energy of the approximation and detailed coefficients is computed.	182
Vocal fold features	Periodicity of glottis movements. Glottal to Noise Excitation. Amount of noise due to the pathological vocal fold vibration. Empirical Mode Decomposition.	22
TQWT	The TQWT algorithm relies on Q (Q-factor), r (redundancy) and J (number of levels) parameters. After decompositions with the TQWT, energy/ entropy values of each level were computed.	423

Every subject was initially examined by a physician, and then the sustained phonation of the vowel /a/ was collected with three repetitions. During the data collection process, the microphone was set to 44.1 kHz. All subjects who participated in the study were informed about the data collection process, signed informed consent, and attended the test voluntarily, according to the approval of the Clinical Research Ethics Committee of Bahcesehir University.[3]

As shown in Table 4.1, the dataset includes baseline features, time frequency features, Mel-Frequency Cepstral Coefficients (MFCCs), vocal fold features, and additional features extracted with the tunable Q-factor wavelet transform (TQWT) related with the fundamental frequency, proposed by Sakar et. al.[3] After the feature extraction, the obtained feature vectors were standardized so that every feature has a zero mean and unit variance.[3]

4.2 Training method

We used leave-one-subject-out (LOSO) cross-validation for the training of the model. At every iteration, the three samples of one individual were left out of the training set and used to test the classifier's generalization ability. In order to eliminate the problem of the unbalanced dataset used for the training of the model, we performed ensemble classification. More specifically, after the selection of the test samples, the remaining dataset was divided into two sub-datasets. The control samples were common in both sub-datasets, while the samples of the PD patients were randomly split, including half samples at each sub-dataset. The emerging sub-datasets still had more PD samples than control, but splitting in more than two resulted in incomplete training due to the small training data size.

Following the split into the ensemble sub-datasets, each of them was further divided into training and validation sets. The validation set consisted of randomly selected samples, ensuring that no subject's samples were included in the validation and training sets simultaneously, keeping the model unbiased. Furthermore, it was always balanced and its size was a specific percentage of the total samples. The remaining samples were used for the training of the model. The validation set was also used for the hyperparameter tuning and the early stopping in the training of TabNet. This method of splitting data into train and validation sets ensures that the ensemble classifiers have validation sets consisting of different samples, which could otherwise add bias to the model.

4.3 Model

This research uses the PyTorch implementation of Google AI's TabNet, available at [24]. Pytorch-TabNet is an open-source library that provides a scikit-like interface for training a TabNetClassifier or TabNetRegressor. The project was created by Dreamquark, with the main contributors being Sebastien Fischman, Eduardo Carvahlo Pinto, and Bazire Houssin.

After some first experiments with the dataset for Parkinson's classification, we observed that the training stopped by the early stopping method during the first few epochs before the model started fitting. This fact may result from the small validation set's size since the model can accidentally achieve a slightly higher Area Under the ROC Curve (AUC) score at one or more epochs during the first epochs and trigger the early stopping rules. To avoid this issue, we modify TabNet's code to ignore a

specific number of epochs from the beginning of training. We chose the number of these epochs to be the double of the early stopping's patience.

4.4 Hyperparameter tuning

Initially, we performed hyperparameter tuning using the grid search method. We created a grid of possible values for some basic parameters that included:

- the width of the decision prediction layer (n_d)
- the width of the attention embedding of each mask (n_a)
- the number of steps in the architecture (n_{steps})
- Number of samples per batch ($batch_size$)
- Size of the mini batches used for Ghost Batch Normalization ($virtual_batch_size$)
- the coefficient of feature reusage in the masks ($gamma$)
- the number of independent GLU layers at each step ($n_{independent}$)
- the number of shared GLU layers at each step (n_{shared})
- the sparsity loss coefficient of the generated masks ($lamda_sparse$)
- the initial learning rate used for training with Adam ($learning_rate$)
- the parameters of the Pytorch's Scheduler that change the learning rates during training ($scheduler_params$)
- Number of consecutive epochs without improvement before performing early stopping ($patience$)

However, after some trials, we noticed no significant variation in the selected parameters from the grid search. Thus, we decided to set the hyperparameters manually based on the observations from the grid search results to reduce the computational needs of training the models. Table 4.2 shows the selected values.

In Pytorch-TabNet's implementation, the seed for the model's initialization is manually set in order to add reproducibility. We decided to test TabNet's performance, stability and accuracy by training the model multiple times with both constant and randomly selected seed.

Table 4.2: Selected hyperparameters and training parameters for TabNet.

Hyperparameter	Value
n_d	8
n_a	8
n_{steps}	1
$batch_size$	64
$virtual_batch_size$	32
$gamma$	1.3 (default)
$n_{independent}$	2
n_{shared}	2
$lamda_sparse$	$1e - 3$ (default)
$learning\ rate$	0.01
$scheduler_params$	{ $step_size : 10, gamma : 0.9$ }
$patience$	40

4.5 Classification predictions

TabNet has the ability to output both binary and probability-like predictions. To increase the stability and reliability of the model, we decided to use soft voting of the probability predictions from the ensemble classifiers and then binarize the result to classify every sample. Finally, following the method often used in research for Parkinson’s disease classification, we perform majority voting of the samples to make a final prediction of every subject since every person who participated in the study provided three recording samples.

4.6 Evaluation metrics

The generalization ability of the trained classifiers in unseen data is estimated by the most common evaluation metrics. After the LOSO cross-validation is completed, we create the confusion matrix from the predictions in both sample and subject classification cases. The row of the table represents the predicted class, while the column represents the actual class. The parameters tp and tn denote the number of positive and negative instances that are correctly classified. In contrast, fp and fn denote the number of misclassified negative and positive instances, respectively.

We also compute the accuracy, auc (area under the ROC curve), specificity, precision, recall, F1-score, and MCC (Matthews correlation coefficient) as shown in Table 4.3.

Table 4.3: Evaluation metrics.[44]

Metric	Formula	Evaluation focus
accuracy	$\frac{tp + tn}{tp + fp + tn + fn}$	The accuracy metric measures the ratio of correct predictions over the total number of instances evaluated.
specificity	$\frac{tn}{fp + tn}$	This metric is used to measure the fraction of negative patterns that are correctly classified.
precision	$\frac{tp}{tp + fp}$	Precision is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class.
recall	$\frac{tp}{tp + fn}$	Recall is used to measure the fraction of positive patterns that are correctly classified.
F1-score	$\frac{2 \cdot precision \cdot recall}{precision + recall}$	This metric represents the harmonic mean between recall and precision values.
Area Under the ROC Curve (auc)	area under the ROC curve (true positive rate (recall) against false positive rate (1 - specificity) at various threshold settings)	AUC is a performance measurement for the classification problems at various threshold settings and represents the degree of separability between classes.
Matthews correlation coefficient (MCC)	$\frac{tp \cdot tn - fp \cdot fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$	The MCC is a correlation coefficient between the observed and predicted binary classifications with range between -1 and +1.

Chapter 5

Experimental Results

5.1 Classification performance of TabNet

This section presents the results obtained by training TabNet with the dataset for Parkinson's classification. For every experiment, we calculated the selected metrics for each sample as an independent test case. Following the method often used in research for Parkinson's disease classification, we also performed majority voting of the three available samples to make a final prediction for every subject since every person who participated in the study provided three recording samples.

5.1.1 Best results

Figure 5.1 present the best results that we obtained with TabNet from the inference of the trained models during the LOSO cross-validation. More precisely, the model was set with the selected hyperparameters (Table 4.2) and validation set's size equal to 10% of the total training samples of every ensemble classifier. TabNet realized the accuracy of 95.24% for samples' predictions and 97.22% for subjects' predictions. TabNet reached maximum scores of 98.92% for precision, 97.34% for recall, 96.88% for specificity, 98.12% for F1-score and 92.83% for MCC when classifying all the

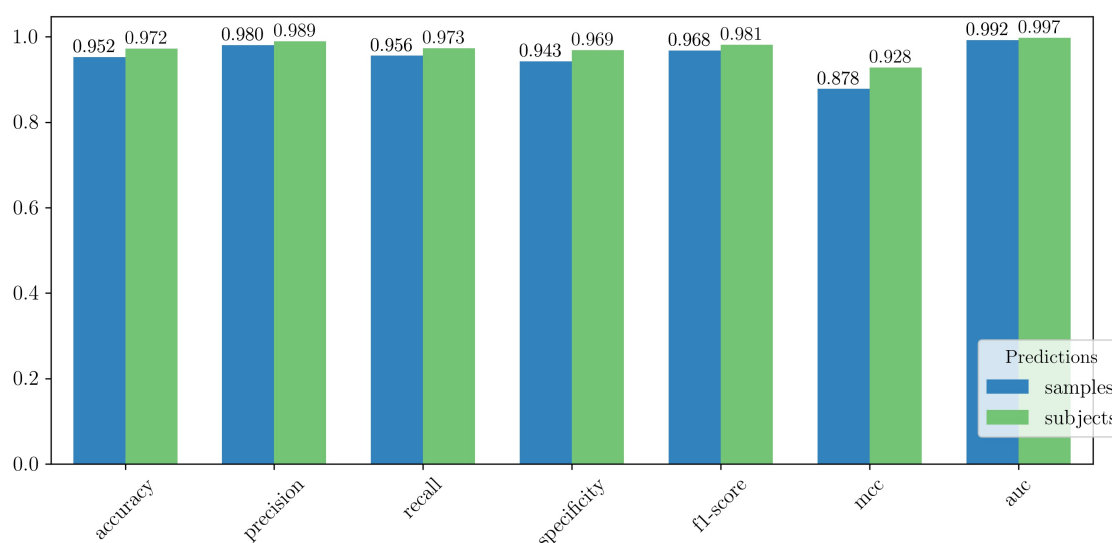


Figure 5.1: Best results obtained with TabNet in the cases of Parkinson's disease classification for every individual sample and each subject using its three provided voice recordings.

subjects using the three samples that they provided, with the metrics being slightly lower when treating every sample as an individual classification case. From the 64 individuals in the control group, 2 were wrongly predicted as PD patients, while 5 subjects diagnosed with Parkinson’s disease by the physicians were mistakenly classified as healthy (Figure 5.3).

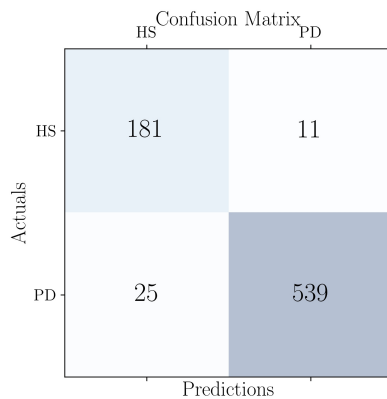


Figure 5.2: Confusion matrix that presents the correct and incorrect classification results for every sample from the iteration that achieved the highest accuracy score.

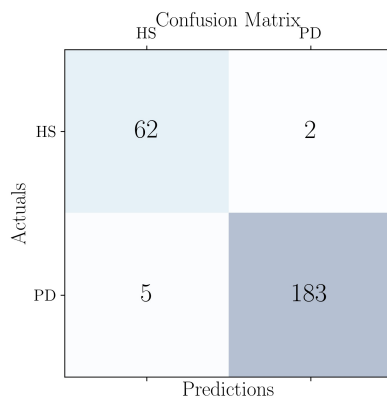


Figure 5.3: Confusion matrix that presents the correct and incorrect classification results for every subject from the iteration that achieved the highest accuracy score.

5.1.2 Hyperparameter tuning

As mentioned in Section 4.4, we performed hyperparameter tuning to achieve the best possible results for the classification with the dataset used. Figures 5.4, 5.5 show the change in the average performance of the model in the indicative cases of using a single-step and a three-step architecture. According to the obtained predictions, setting the number of steps equal to three and thus using a deeper architecture resulted in a slight drop in the metrics. This may be a consequence of the extremely small datasets currently available for Parkinson’s disease classification since bigger models need a large amount of data to train adequately. The same effect was observed when we changed the hyperparameters in order to create a more complex model. For that reason, we kept the architecture of the models as simple as possible, with single-step architecture, a couple of independent and shared decision steps for the feature transformer, small batch sizes and widths for both the decision

prediction layer and the attention embedding of each mask.

In order to ensure that the size of the sub-datasets used for validation did not affect the hyperparameter-tuning phase, we included the proportion of the samples that TabNet used for validation in the grid search. The model's performance was tested with validation sets' sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier. Minor changes in the parameters resulted in slightly higher or lower classification scores. However, as mentioned before, we decided not to use grid search for hyperparameter tuning and manually set the values so that we could reduce the computational needs of the training.

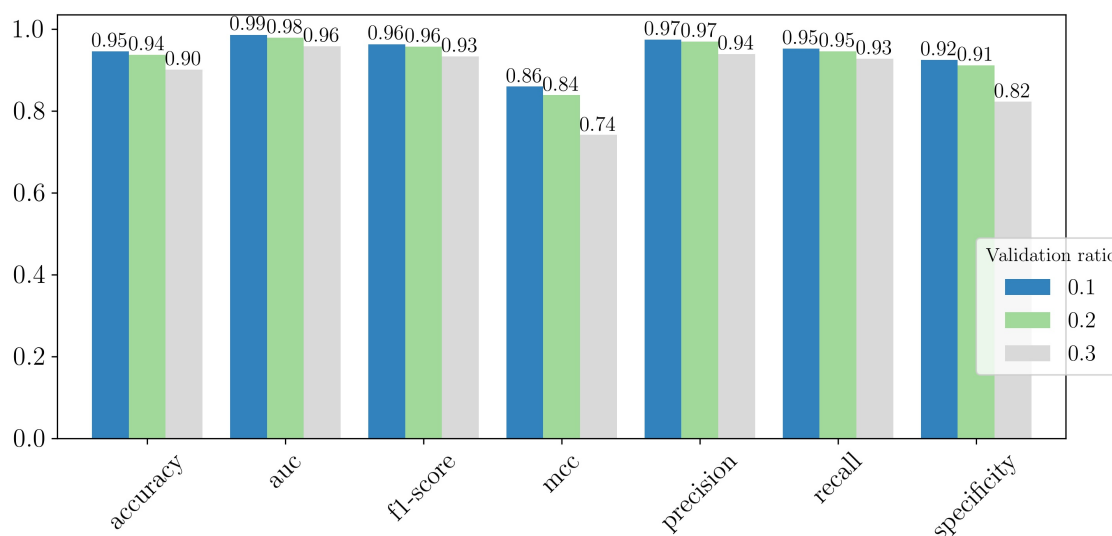


Figure 5.4: TabNet's classification results with single-step architecture and different validation sets' sizes. The model's performance was tested with validation sets' sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.

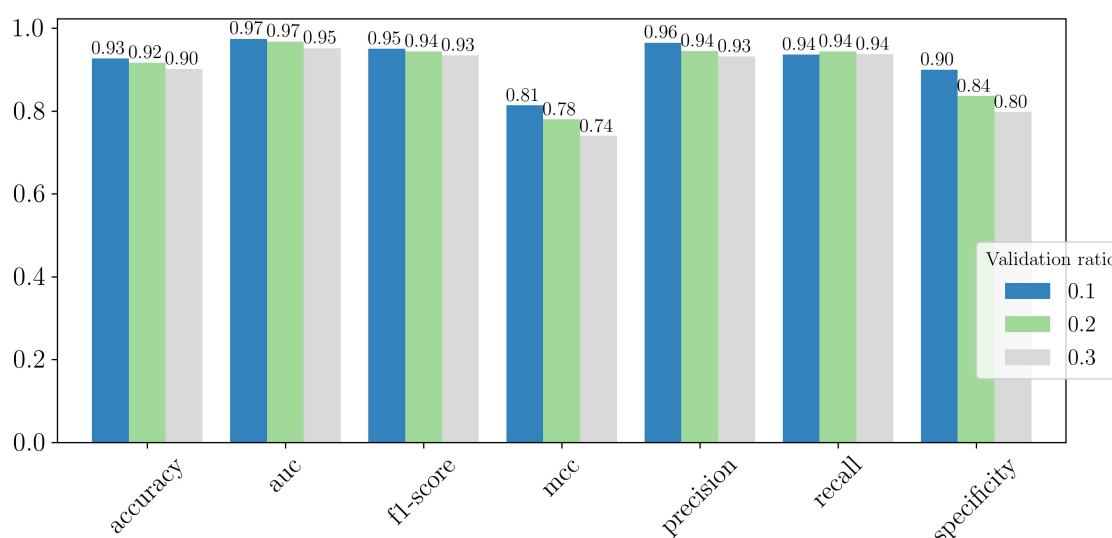


Figure 5.5: TabNet's classification results with three-step architecture and different validation sets' sizes. The model's performance was tested with validation sets' sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.

For every combination of the hyperparameters in the grid search, we repeated the LOSO cross-validation training four or more times in order to obtain a more general estimation of TabNet’s performance. In Figures 5.6, 5.7 we present the box plots from the classification results using single-step and three-step architecture. The deepening of the model results in a less stable classifier with worse average metrics and more considerable variations between the trained models with the same parameters.

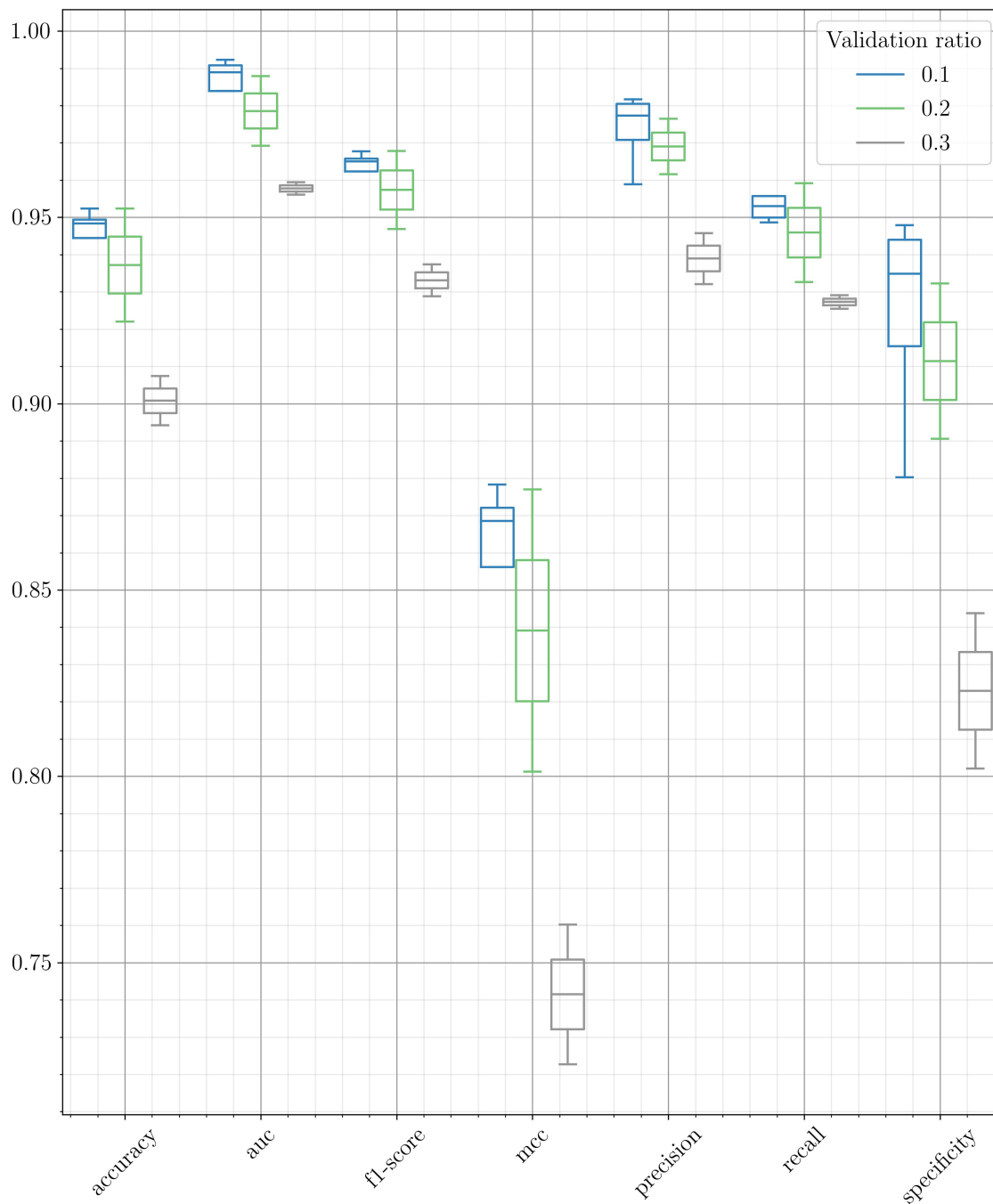


Figure 5.6: TabNet’s classification box plots with single-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.

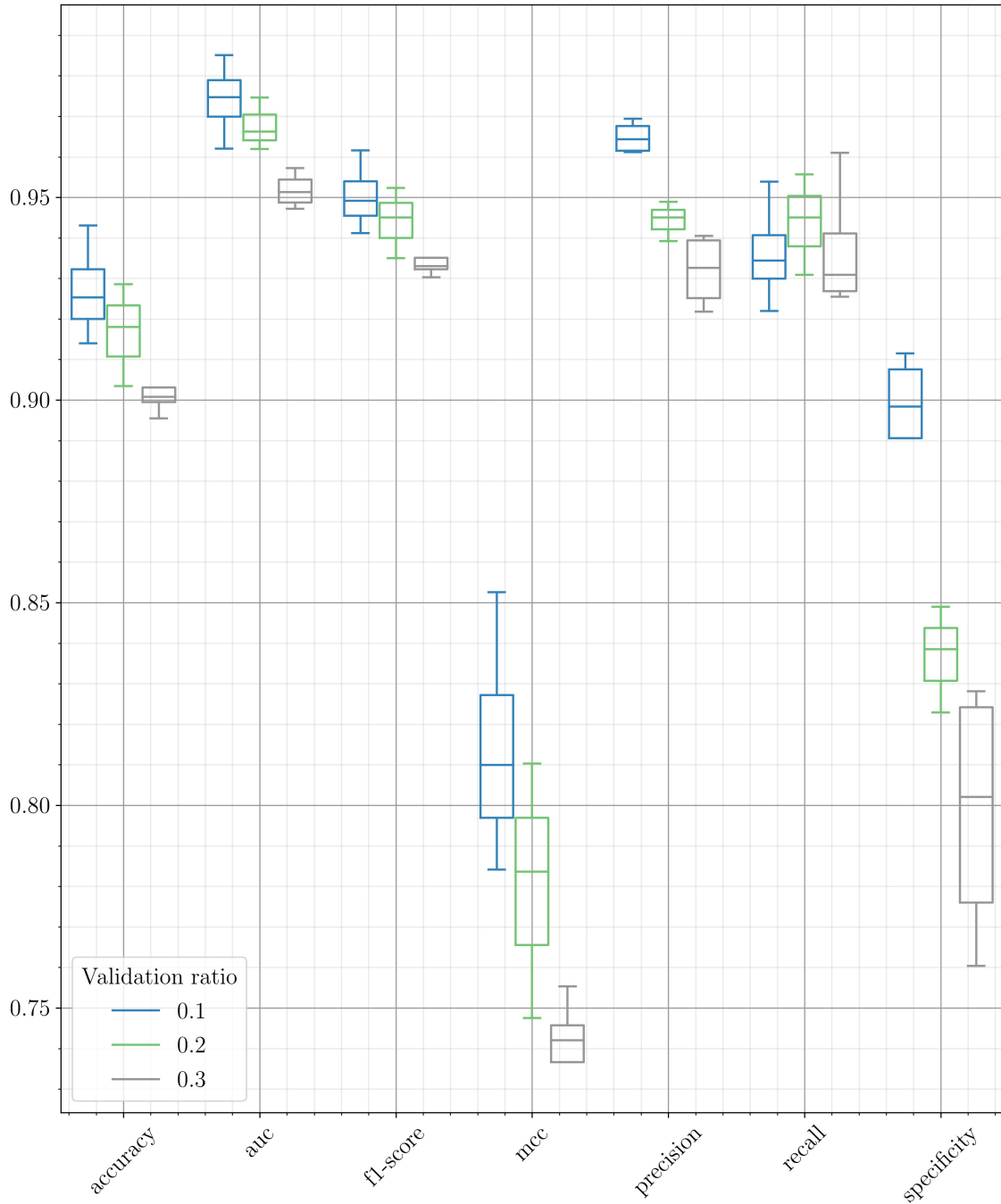


Figure 5.7: TabNet’s classification box plots with three-step architecture and different validation sets’ sizes. The model’s performance was tested with validation sets’ sizes equal to 10%, 20%, and 30% of the total training samples of every ensemble classifier.

5.1.3 TabNet’s initialization method

As mentioned before, Pytorch-TabNet’s implementation includes setting the seed for the model’s initialization manually in order to ensure reproducibility. Considering that the dataset was randomly split into two sub-datasets for the ensemble classifiers and later into train and validation sets, we tested the effect of the validation sets’ selection in TabNet’s performance. The validation set’s size was set to 10% of the samples of every ensemble classifier. The samples of every validation set were

randomly chosen, ensuring that it was always balanced. Figure 5.8 includes the box plots of TabNet’s predictions for the test samples with constant and randomly selected seeds for every model’s initialization. The results indicate that the validation set affects the classification ability of the model, considering that TabNet’s outputs are set to be entirely reproducible for a specific seed.

When the seed is randomly selected for every single model during the LOSO cross-validation, the model’s performance in terms of stability and accuracy varies significantly. For example, the accuracy varies between 91.4% and 94.31%, while the values of other metrics such as MCC differ up to 6.85%.

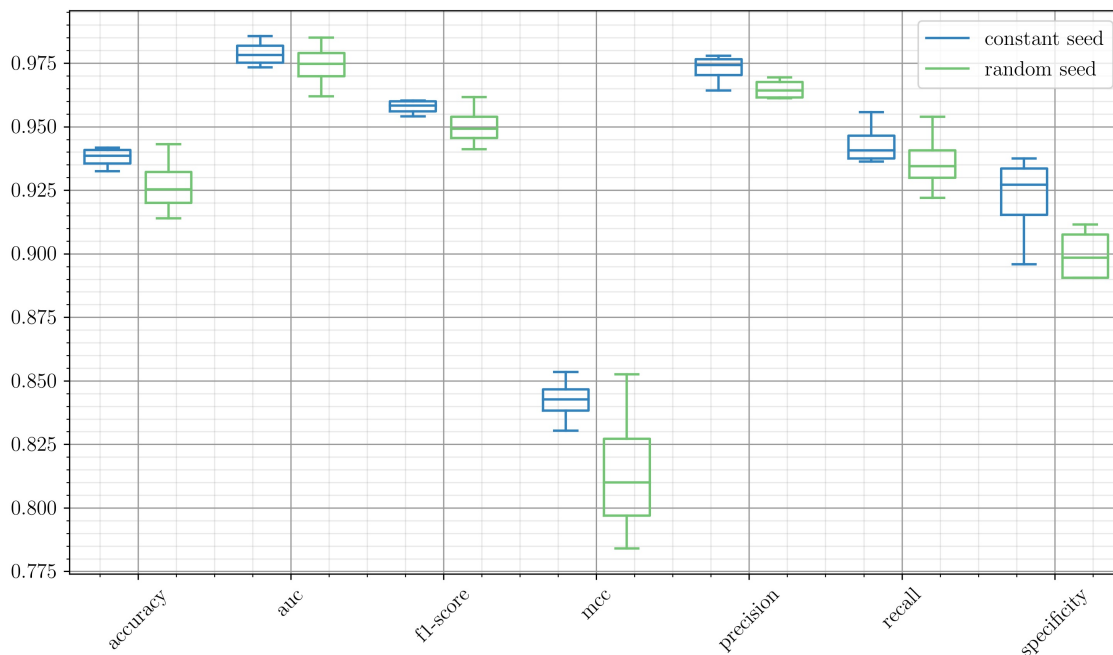


Figure 5.8: Box plots of TabNet’s predictions for the test samples with constant and randomly selected seeds for every model’s initialization. The validation set’s size was set to 10% of the samples of every ensemble classifier. The samples of every validation set were randomly chosen, ensuring that it was always balanced.

5.1.4 Feature reduction

Sakar et al. [3] created the currently biggest dataset for Parkinson’s classification in both samples and features, with the other available datasets containing less than 50 features. In the final part of this section, we tried to investigate the effect of the initial number of features that TabNet must choose from in order to select the salient ones for the classification. For that reason, we reduced the number of features from 753 to 40 and 18. In both cases, we eliminated most features, keeping only the common ones between dataset [3] and two other frequently used and available datasets in the bibliography.

Dataset [25] includes a total of 195 voice samples with 18 features from 23 PD patients and 8 healthy individuals, and dataset [26] consists of 40 features from 40 PD patients and 40 people in the control group.

Figures 5.9, 5.10 show a significant drop in all the evaluation metrics as the number of the initially available features reduces, with the accuracy of the predictions for subjects decreasing from 96% to 89% for 18 features. During this phase, we used

random initialization and validation sets' size equal to 10% of the total training samples of every ensemble classifier. The hyperparameters were set to the values mentioned in Table 4.2.

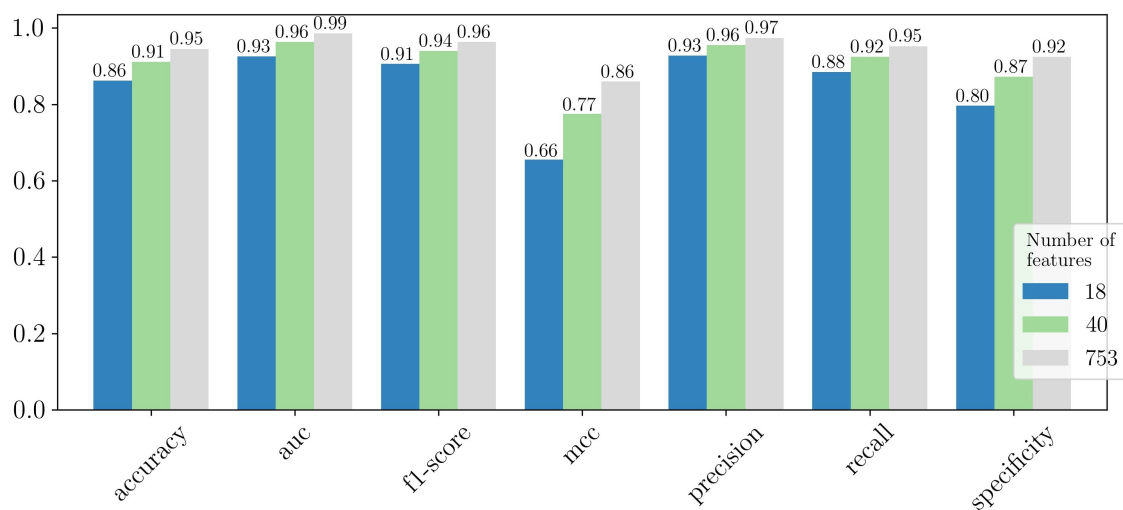


Figure 5.9: Average TabNet's evaluation metrics for classification of individual samples during feature reduction.

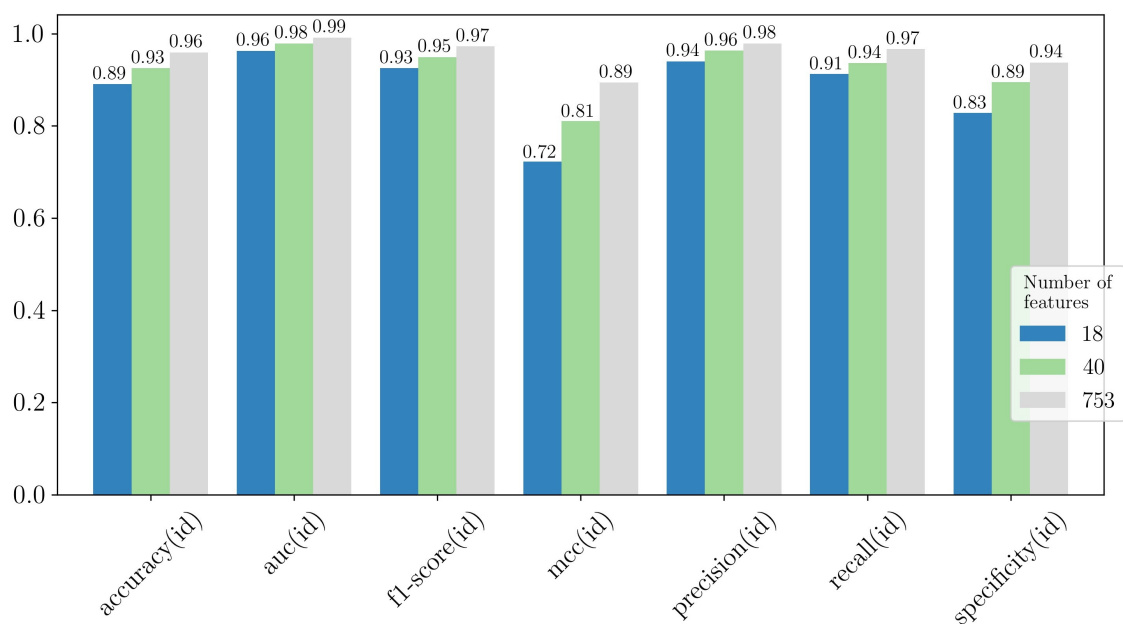


Figure 5.10: Average TabNet's evaluation metrics for classification of subjects during feature reduction.

5.2 Interpretability of the model

TabNet creates a feature selection mask for every decision step, which determines the salient features that will participate in the decision-making. The model realizes instance-wise feature selection, meaning that the masks select the best combination of features for every individual sample in contrast with most other dimensionality reduction techniques. The final masks of every decision step for every sample are available at the end of the training, giving the model local interpretability. By aggregating all the masks, TabNet creates an explain matrix that summarizes the feature selection of the model. Furthermore, the global interpretability of the model is achieved by a linear combination of the masks of different steps. TabNet quantifies the aggregate feature importance, which is weighted by the relative importance of each step in the final decision.

In order to present TabNet’s interpretability results, a two-step architecture was selected, with the rest of the hyperparameters being the same as before (Table 4.2). We extracted the masks, explain matrices and feature importances of the two ensemble classifiers for two randomly selected subjects, one diagnosed with Parkinson’s disease (subject with ID 100) and one from the control group (subject with ID 251). We present the results in the case of 18 initial features in order to keep the graphs legible.

Figures 5.11, 5.12 and 5.15 belong to subject No.100 and Figures 5.13, 5.14 and 5.16 to subject No.251. It is important to note that TabNet selects different features based on the validation sets used to train the ensemble classifiers. Additionally, the features’ contribution to the final prediction varies considerably when the input samples belong to a healthy subject or a PD patient. All the samples from these two subjects were classified correctly.

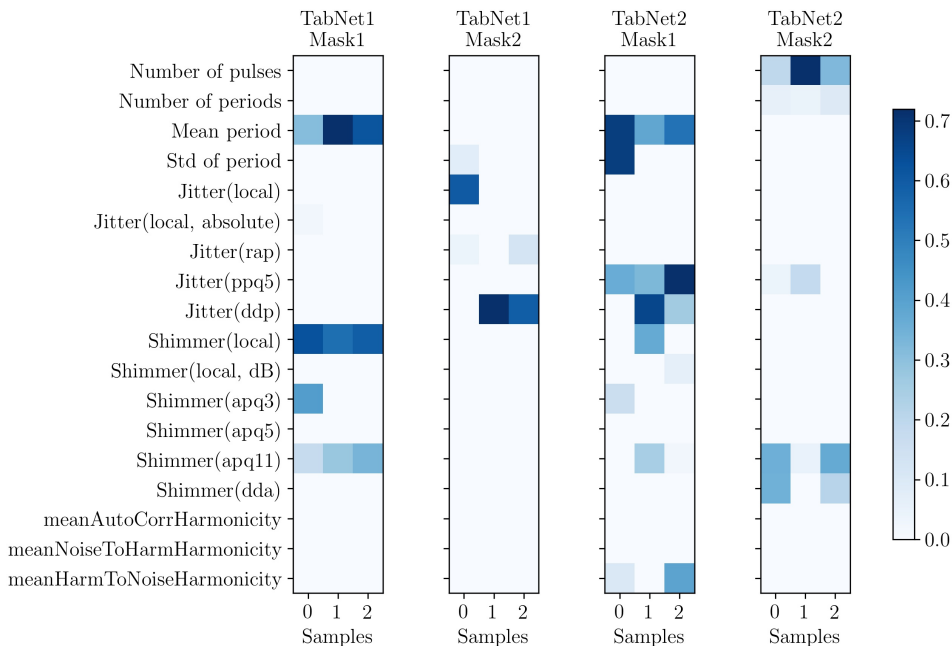


Figure 5.11: Masks for the samples of a subject with Parkinson’s disease (ID:100).

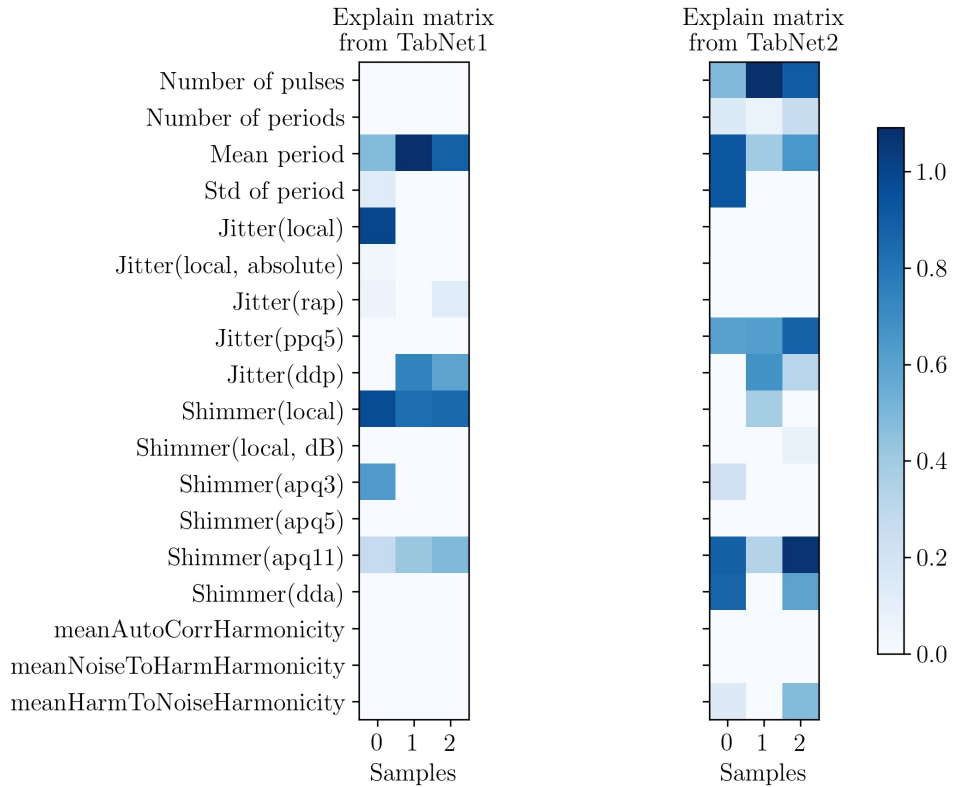


Figure 5.12: Explain matrices for the samples of a subject with Parkinson's disease (ID:100).

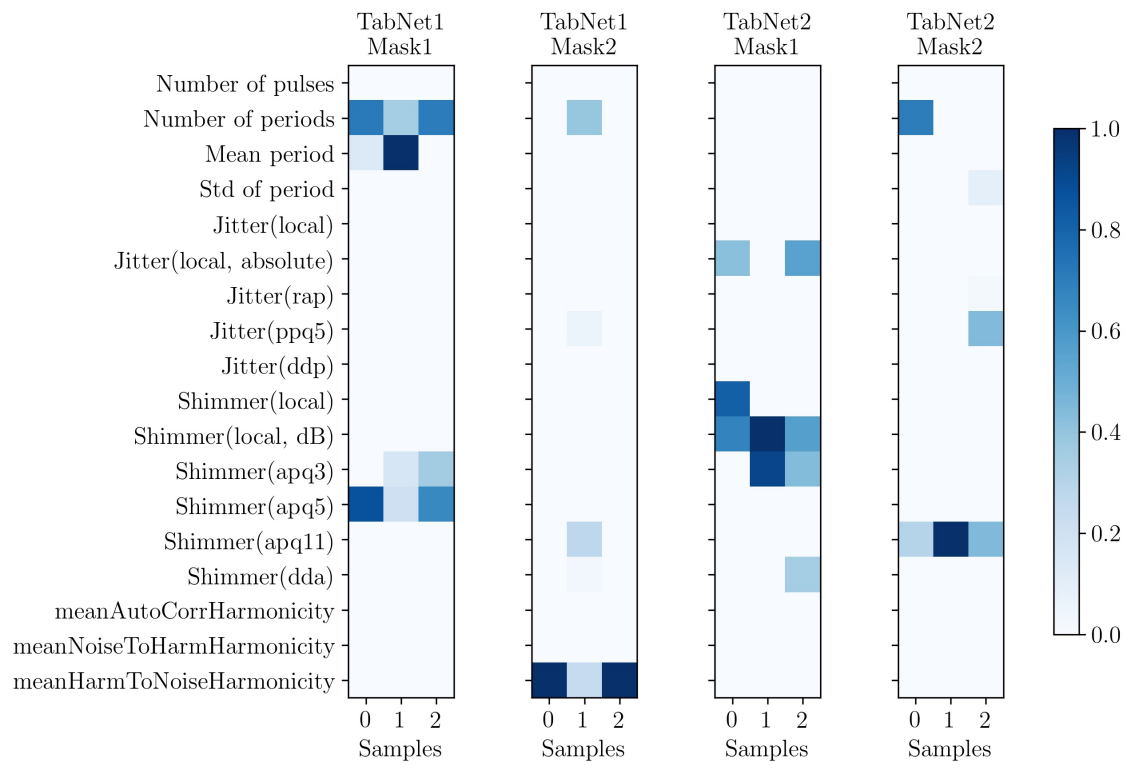


Figure 5.13: Masks for the samples of a subject from the control group (ID:251).

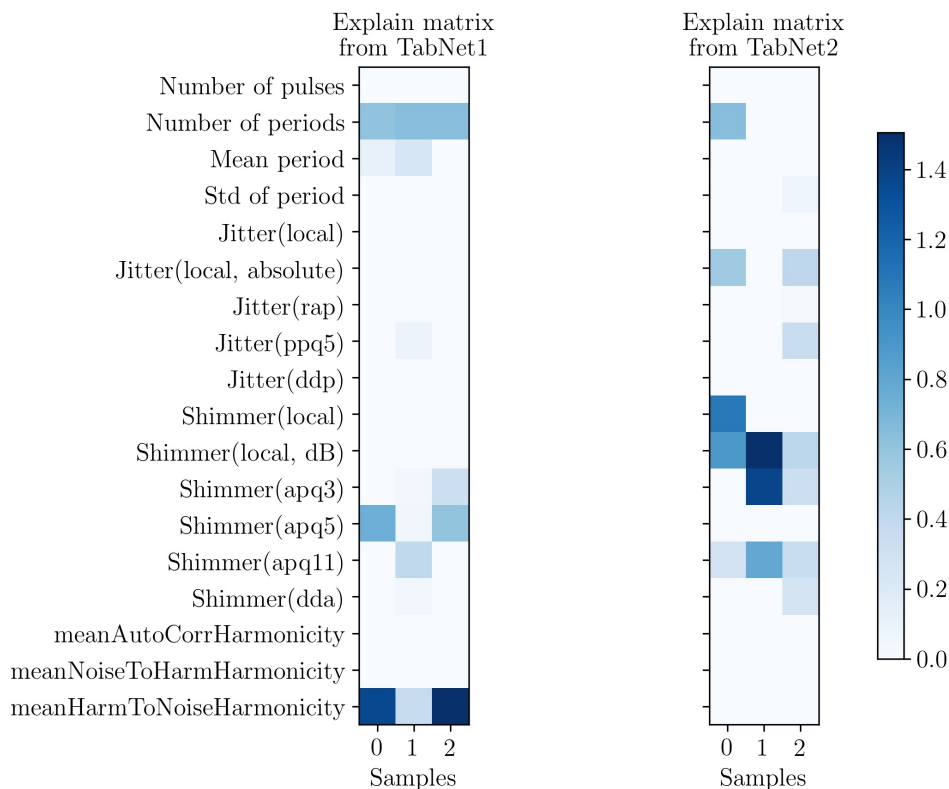


Figure 5.14: Explain matrices for the samples of a subject from the control group (ID:251).

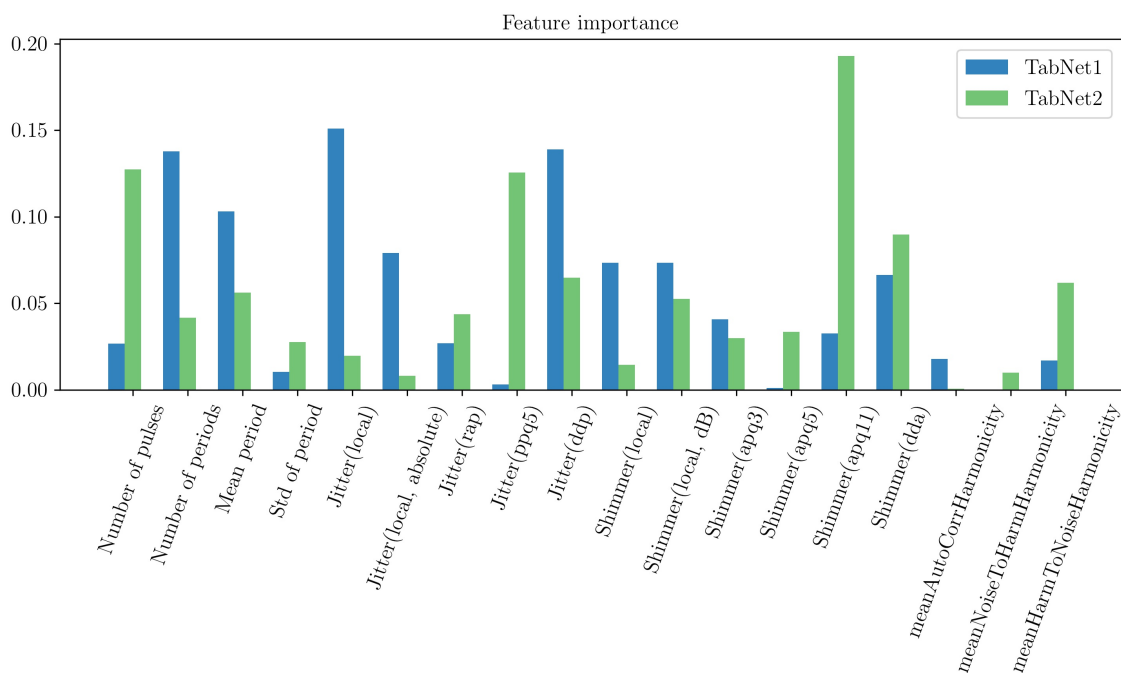


Figure 5.15: Feature importances for the samples of a subject with Parkinson's disease (ID:100).

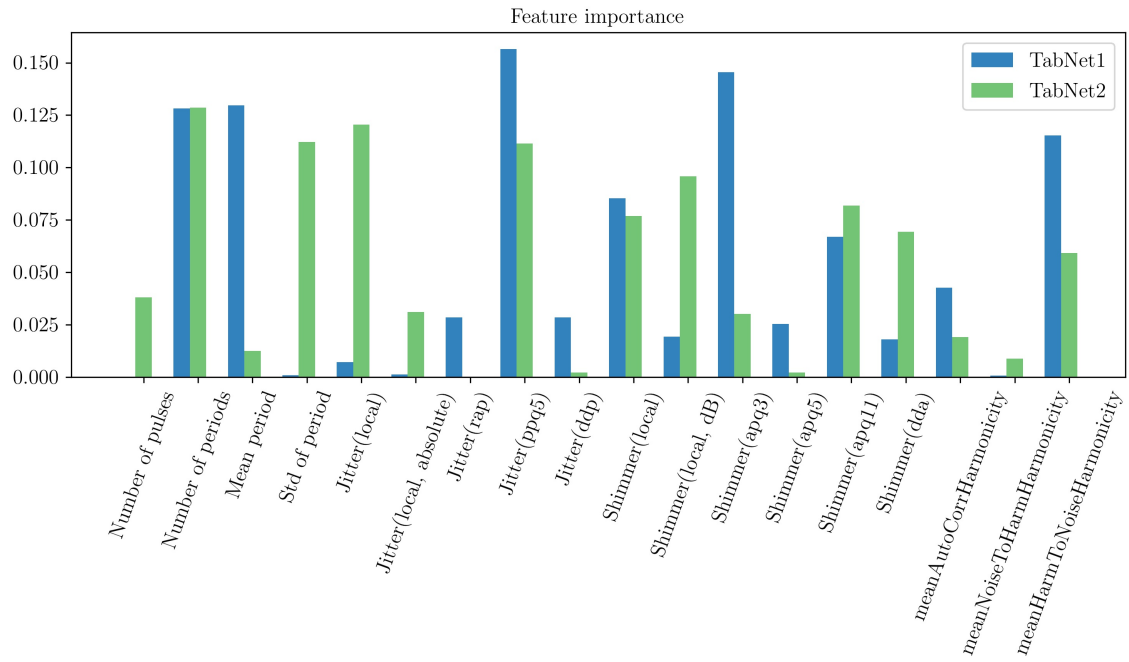


Figure 5.16: Feature importances for the samples of a subject from the control group (ID:251).

Chapter 6

Discussion

There have been numerous studies in the literature (Table 6.1) about PD classification based on features extracted from voice records. According to the results of PD telemedicine studies, the choice of feature extraction and learning algorithms is directly related to the accuracy and reliability of the proposed systems.[3] Many studies dedicated to distinguishing healthy subjects from PD patients use the dataset [3], that is also used in this study, which is the bigger publicly available dataset at the time. Other commonly used and available datasets are [25] that includes a total of 195 voice samples from 23 PD patients and 8 healthy individuals, [27] consisting of multiple speech recordings of 20 PD and 20 healthy subjects, and [26] with 40 PD patients and 40 people in the control group. All datasets have similar features including vocal fundamental frequency, measures of variation in fundamental frequency, measures of variation in amplitude, measures of the ratio of noise to tonal components, nonlinear dynamical complexity measures, and nonlinear measures of fundamental frequency variation.[3]

Some literature studies propose classifiers that achieve accuracy scores of nearly 100% on the [3] dataset. Tuncer et al. presented a novel octopus-based feature extraction network (multiple pooling method). They used many classifiers in order to evaluate the ability to distinguish healthy control from PD patients, reaching 97.62% using a 1 Nearest Neighbor (1NN) classifier.[28] Olivares et al. used the Bat Algorithm in the Extreme Learning Machine (BA-ELM), resulting in a 96.7% maximum accuracy.[29] Another study uses SMOTE (Synthetic Minority Over-sampling Technique) and Random Forest achieving 94.9% accuracy.[30] In another study, the AdaBoost classifier was used with grid search hyperparameter tuning and preprocessing with Standard Scaler, with the highest accuracy of 94%.[31] Solana-Lavalle et al. proposed PD classification with Wrappers feature subset selection that showed an accuracy of 94.7% with the SVM classifier with RBF kernel.[32].

However, even though these studies, among others, reported very high classification rates, the cross-validation techniques used by the authors cause biased results. This is caused by the fact that every subject has multiple recordings, and both training and test sets used to train the models included voice samples of the same subject.[3, 33] Properly splitting the dataset into training/test (or training/validation/test) sets using the Leave-One-Subject-Out cross-validation technique results in a dramatic drop in the evaluation metrics.[3, 33]

Some published studies follow the proper methods for creating the train/test sets, thus construct an unbiased model for both train and inference stages.[3] Wijaya et al. used Gradient Boosted Tree (GBT) with the weight threshold equal to or bigger than 9, resulting in 92% accuracy, 94% F1-score and 78% MCC scores.[34] Sakar et al.

used the LOSO cross-validation to validate the generalization ability of the model, achieving 86% accuracy, 84% F1-score and 59% MCC with the SVM (RBF) classifier and the mRMR feature selection algorithm to select the top-50 features from all feature subsets.[3] Gunduz created a 9-layered CNN (1 input layer, 6 convolution layers (each of two convolution layers followed by max-pooling operation), 1 fully-connected (dense) layer, and 1 output layer) with 86.9% accuracy, 91.7% F1-score, and 63.2% MCC.[35]

Similar results were published in studies that used the other datasets with both k-fold and LOSO cross-validation. For example, Gürüler et al. created a classification system by combining k-means clustering-based feature weighting and Complex Valued Artificial Neural Network (CVANN), with which they reached an accuracy of 99.5% with 10-fold cross-validation with the [27] dataset.[45]

It is important to note that, since every subject of the study provided three recording samples, most studies that use LOSO cross-validation perform majority voting to make a final prediction of every individual. This results in slightly higher metrics during the inference of the model.

According to our results, TabNet outperforms the classifiers in literature with average and maximum accuracy for every sample reaching 94.54% and 95.24%, respectively, when using all the available features in the dataset [3] and LOSO cross-validation. The average and maximum values for F1-score and MCC were 96.31%, 96.77%, and 85.97%, 87.83%, respectively. In the case of predictions for every subject, the average and maximum observed results were 95.93% and 97.22% for accuracy, 97.26% and 98.12% for F1-score, 89.43% and 92.83% for MCC metric. Furthermore, TabNet internally selects the best features for the classification of every subject, realizing interpretability (global and local) in contrast to the studies in the bibliography.

Table 6.1: Parkinson’s disease classification results in the literature. All the included studies use the Sakar et al. (2019) [3] dataset to train and inference the proposed models.

Author, Year	Methods	Train/test split	Best results
Sakar et al.(2019) [3]	SVM(RBF)	LOSO cv	Accuracy = 0.86 F1 score = 0.84 MCC = 0.59
Solana-Lavalle et al. (2020) [32]	SVM(RBF) with Wrappers feature subset selection	10-fold cv	Accuracy = 0.95
Tuncer et al.(2019) [28]	INN classifier with octopus-based feature extraction network	10-fold cv	Accuracy = 0.98 F1 score = 0.98
Olivares et al.(2020) [29]	Bat Algorithm in the Extreme Learning Machine (BA-ELM)	80 – 20	Accuracy = 0.96
Polat(2019) [30]	SMOTE (Synthetic Minority Over-sampling Technique) and Random Forest	10-fold cv	Accuracy = 0.95 F1 score = 0.95
Anisha et al.(2020) [31]	AdaBoost with grid search hyperparameter tuning and preprocessing with Standard Scaler	10-fold cv	Accuracy = 0.94 F1 score = 0.94
Wijaya et al.(2019) [34]	Gradient Boosted Tree (GBT)	LOSO cv	Accuracy = 0.92 F1 score = 0.94 MCC = 0.78
Gunduz(2019) [35]	9-layered CNN	LOSO cv	Accuracy = 0.87 F1 score = 0.82 MCC = 0.63
Xiong et al.(2020) [36]	Adaptive Grey Wolf Optimization (GWO) with sparse autoencoder and Linear Discriminant Analysis(LDA)	10-fold cv	Accuracy = 0.95
Ashour et al.(2020) [37]	SVM(cubic), weighted-hybrid selected features that	70-30	Accuracy = 0.94
Yücelbaş(2020) [38]	Simple Logistic hybrid system(SLGS)	10-fold cv	Accuracy = 0.88 F1 score = 0.87 MCC = 0.63
Nissar et al.(2019) [39]	XGBoost with Minimum Redundancy - Maximum Relevance (mRMR) feature selection	not mentioned	Accuracy = 0.95 F1 score = 0.95
Behir(2020) [40]	Gaussian mixture model (GMM) with mRMR feature selection	10-fold cv	Accuracy = 0.95 MCC = 0.70
Proposed approach	TabNet	LOSO cv	Accuracy = 0.97 F1 score = 0.98 MCC = 0.93

Chapter 7

Conclusion and future work

7.1 Summary

This thesis investigated the potential of using deep learning in Parkinson's disease classification from features extracted from voice recordings. Pytorch-TabNet, the selected DNN model, outperformed the other classifiers in the literature using the leave-one-subject-out cross-validation technique. Unlike all the available models, Pytorch-TabNet provides local and global interpretability of the predictions.

7.2 Future work

One of the most significant drawbacks in Parkinson's disease classification using machine learning and neural networks is the small size of the available datasets. The largest publicly available dataset includes less than 800 samples belonging to 252 individuals. This fact limits the possibilities of using deeper architectures that need significantly more training data. As a result, collecting more data from patients with PD and healthy subjects is vital for more extensive research with models of higher capabilities.

Considering the methods used to train TabNet, hyperparameter tuning should be performed during the LOSO cross-validation training if more computational resources are available. Using the grid search method to find the best parameters for every model might lead to better results. More extended research can also be done on the effects of the samples that constitute the validation set. The validation set should be kept constant while randomly setting the classifiers' seed parameters to inspect the performance of the trained models.

A more complex Parkinson's disease classification system can be developed in the future, following a multimodal approach. Along with features extracted from voice recordings, other motor symptoms such as gait characteristics and resting tremor can also be considered in the classification process, using a single model or multiple classifiers with ensemble learning. Additionally, a time series of biomedical data gathered from people with Parkinson's disease over a predefined period can be used to train TabNet. The resulting deep learning architecture could estimate the stage of new subjects suffering from PD or be used as a telemonitoring device for remote symptom progression monitoring.

Bibliography

- [1] Bianca L B Marino et al. "Parkinson's Disease: A Review from Pathophysiology to Treatment." eng. In: *Mini reviews in medicinal chemistry* 20.9 (2020), pp. 754–767. issn: 1875-5607 (Electronic). doi: [10.2174/1389557519666191104110908](https://doi.org/10.2174/1389557519666191104110908).
- [2] Werner Poewe et al. "Parkinson disease". In: *Nature Reviews Disease Primers* 3.1 (2017), p. 17013. issn: 2056-676X. doi: [10.1038/nrdp.2017.13](https://doi.org/10.1038/nrdp.2017.13). url: <https://doi.org/10.1038/nrdp.2017.13>.
- [3] C. Okan Sakar et al. "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform". In: *Applied Soft Computing Journal* 74 (Jan. 2019), pp. 255–263. issn: 15684946. doi: [10.1016/j.asoc.2018.10.022](https://doi.org/10.1016/j.asoc.2018.10.022).
- [4] Rocha EM, De Miranda B, and Sanders LH. "Alpha-synuclein: Pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease". In: *Neurobiology of disease* 109.Pt B (Jan. 2018), pp. 249–257. issn: 1095-953X. doi: [10.1016/J.NBD.2017.04.004](https://doi.org/10.1016/J.NBD.2017.04.004). url: <https://pubmed.ncbi.nlm.nih.gov/28400134/>.
- [5] David K. Simon, Caroline M. Tanner, and Patrik Brundin. *Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology*. Feb. 2020. doi: [10.1016/j.cger.2019.08.002](https://doi.org/10.1016/j.cger.2019.08.002). url: <https://pubmed.ncbi.nlm.nih.gov/31733690/>.
- [6] Samuel D. Kim et al. "Parkinson disease". In: *Handbook of Clinical Neurology*. Vol. 159. Elsevier B.V., Jan. 2018, pp. 173–193. doi: [10.1016/B978-0-444-63916-5.00011-2](https://doi.org/10.1016/B978-0-444-63916-5.00011-2).
- [7] Ronald B Postuma et al. "MDS clinical diagnostic criteria for Parkinson's disease." eng. In: *Movement disorders : official journal of the Movement Disorder Society* 30.12 (Oct. 2015), pp. 1591–1601. issn: 1531-8257 (Electronic). doi: [10.1002/mds.26424](https://doi.org/10.1002/mds.26424).
- [8] Lorraine V. Kalia and Anthony E. Lang. *Parkinson disease in 2015: Evolving basic, pathological and clinical concepts in PD*. Feb. 2016. doi: [10.1038/nrneurol.2015.249](https://doi.org/10.1038/nrneurol.2015.249). url: <https://pubmed.ncbi.nlm.nih.gov/26782330/>.
- [9] Garrett E. Alexander. *Biology of Parkinson's disease: Pathogenesis and pathophysiology of a multisystem neurodegenerative disorder*. 2004. doi: [10.31887/dcns.2004.6.3/galexander](https://doi.org/10.31887/dcns.2004.6.3/galexander). url: [/pmc/articles/PMC3181806/%20/pmc/articles/PMC3181806/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181806/](https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC3181806/).
- [10] Andreas Holzinger et al. *Causability and explainability of artificial intelligence in medicine*. July 2019. doi: [10.1002/widm.1312](https://doi.org/10.1002/widm.1312).
- [11] Stuart Charles. Shapiro. *Encyclopedia of Artificial Intelligence*. 2nd ed. Vol. 1. Wiley, 2011, p. 35. isbn: 0471503053. doi: [10.4018/978-1-59904-849-9](https://doi.org/10.4018/978-1-59904-849-9). url: https://books.google.com/books/about/Encyclopedia_of_Artificial_Intelligence.html?hl=el&id=fKURAQAAMAAJ.
- [12] A N Ramesh et al. *Artificial intelligence in medicine*. 2004. doi: [10.1308/147870804290](https://doi.org/10.1308/147870804290).

- [13] Ajay Shrestha and Ausif Mahmood. “Review of deep learning algorithms and architectures”. In: *IEEE Access* 7 (2019), pp. 53040–53065. ISSN: 21693536. DOI: [10.1109/ACCESS.2019.2912200](https://doi.org/10.1109/ACCESS.2019.2912200).
- [14] John Stoitsis et al. “Computer aided diagnosis based on medical image processing and artificial intelligence methods”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 569.2 SPEC. ISS. (Dec. 2006), pp. 591–595. ISSN: 01689002. DOI: [10.1016/j.nima.2006.08.134](https://doi.org/10.1016/j.nima.2006.08.134).
- [15] Stavroula Gr Mougiakakou et al. “Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image statistics, laws’ texture and neural networks”. In: *Ultrasound in Medicine and Biology* 33.1 (Jan. 2007), pp. 26–36. ISSN: 03015629. DOI: [10.1016/j.ultrasmedbio.2006.07.032](https://doi.org/10.1016/j.ultrasmedbio.2006.07.032).
- [16] Marios Skevofilakas et al. “A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus”. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC’10*. Vol. 2010. Annu Int Conf IEEE Eng Med Biol Soc, 2010, pp. 6713–6716. ISBN: 9781424441235. DOI: [10.1109/IEMBS.2010.5626245](https://doi.org/10.1109/IEMBS.2010.5626245). URL: <https://pubmed.ncbi.nlm.nih.gov/21096083/>.
- [17] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *Nature* 542.7639 (Feb. 2017), pp. 115–118. ISSN: 14764687. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056). URL: <https://pubmed.ncbi.nlm.nih.gov/28117445/>.
- [18] Kunal Nagpal et al. “Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer”. In: *npj Digital Medicine* 2.1 (June 2019), pp. 1–10. ISSN: 23986352. DOI: [10.1038/s41746-019-0112-2](https://doi.org/10.1038/s41746-019-0112-2). arXiv: [1811.06497](https://arxiv.org/abs/1811.06497). URL: <https://www.nature.com/articles/s41746-019-0112-2>.
- [19] Sercan Ö. Arik and Tomas Pfister. “TabNet: Attentive Interpretable Tabular Learning”. In: (2019). arXiv: [1908.07442](https://arxiv.org/abs/1908.07442). URL: [www.aaai.org%20http://arxiv.org/abs/1908.07442](http://www.aaai.org/20http://arxiv.org/abs/1908.07442).
- [20] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Journal of Machine Learning Research*. Vol. 9. 2010, pp. 249–256. URL: <http://www.iro.umontreal..>
- [21] Yann N Dauphin et al. “Language modeling with gated convolutional networks”. In: *34th International Conference on Machine Learning, ICML 2017*. Vol. 2. 2017, pp. 1551–1559. ISBN: 9781510855144. arXiv: [1612.08083](https://arxiv.org/abs/1612.08083).
- [22] Andre F.T. Martins and Ramon F Astudillo. “From softmax to sparsemax: A sparse model of attention and multi-label classification”. In: *33rd International Conference on Machine Learning, ICML 2016*. Vol. 4. 2016, pp. 2432–2443. ISBN: 9781510829008. arXiv: [1602.02068](https://arxiv.org/abs/1602.02068).
- [23] *Parkinson’s Disease Classification Data Set*. URL: <https://archive.ics.uci.edu/ml/datasets/Parkinson%5C%27s+Disease+Classification#>.
- [24] URL: <https://github.com/dreamquark-ai/tabnet>.
- [25] Max Little et al. “Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease”. In: *Nature Precedings 2008* (Sept. 2008), pp. 1–1. ISSN: 1756-0357. DOI: [10.1038/npre.2008.2298.1](https://doi.org/10.1038/npre.2008.2298.1). URL: <https://www.nature.com/articles/npre.2008.2298.1>.
- [26] Lizbeth Naranjo et al. “Addressing voice recording replications for Parkinson’s disease detection”. In: *Expert Systems with Applications* 46 (Mar. 2016), pp. 286–292. ISSN: 09574174. DOI: [10.1016/j.eswa.2015.10.034](https://doi.org/10.1016/j.eswa.2015.10.034).

- [27] Betül Erdogdu Sakar et al. “Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings”. In: *IEEE Journal of Biomedical and Health Informatics* (2013). issn: 21682194. doi: [10.1109/JBHI.2013.2245674](https://doi.org/10.1109/JBHI.2013.2245674).
- [28] Turker Tuncer and Sengul Dogan. “A novel octopus based Parkinson’s disease and gender recognition method using vowels”. In: *Applied Acoustics* 155 (Dec. 2019), pp. 75–83. issn: 1872910X. doi: [10.1016/j.apacoust.2019.05.019](https://doi.org/10.1016/j.apacoust.2019.05.019). url: <https://doi.org/10.1016/j.apacoust.2019.05.019>.
- [29] Rodrigo Olivares et al. “An optimized brain-based algorithm for classifying parkinson’s disease”. In: *Applied Sciences (Switzerland)* 10.5 (Mar. 2020), p. 1827. issn: 20763417. doi: [10.3390/app10051827](https://www.mdpi.com/2076-3417/10/5/1827). url: <https://www.mdpi.com/2076-3417/10/5/1827>.
- [30] Kemal Polat. “A hybrid approach to Parkinson disease classification using speech signal: The combination of SMOTE and random forests”. In: *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*. Institute of Electrical and Electronics Engineers Inc., Apr. 2019. isbn: 9781728110134. doi: [10.1109/EBBT.2019.8741725](https://doi.org/10.1109/EBBT.2019.8741725).
- [31] C. D. Anisha and N. Arulanand. “Early Prediction of Parkinson’s Disease (PD) Using Ensemble Classifiers”. In: *2020 International Conference on Innovative Trends in Information Technology, ICITIIT 2020*. Institute of Electrical and Electronics Engineers Inc., Feb. 2020. isbn: 9781728142104. doi: [10.1109/ICITIIT49094.2020.9071562](https://doi.org/10.1109/ICITIIT49094.2020.9071562).
- [32] Gabriel Solana-Lavalle, Juan Carlos Galán-Hernández, and Roberto Rosas-Romero. “Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features”. In: *Biocybernetics and Biomedical Engineering* 40.1 (Jan. 2020), pp. 505–516. issn: 02085216. doi: [10.1016/j.bbe.2020.01.003](https://doi.org/10.1016/j.bbe.2020.01.003). url: <https://doi.org/10.1016/j.bbe.2020.01.003>.
- [33] C. Okan Sakar and Olcay Kursun. “Teliagnosis of parkinson’s disease using measurements of dysphonia”. In: *Journal of Medical Systems* 34.4 (Aug. 2010), pp. 591–599. issn: 01485598. doi: [10.1007/s10916-009-9272-y](https://doi.org/10.1007/s10916-009-9272-y). url: <https://pubmed.ncbi.nlm.nih.gov/20703913/>.
- [34] Adi Wijaya, Kharis, and Wini Prastuti. “Gradient boosted tree based feature selection and parkinson’s disease classification”. In: *Proceedings - 2019 5th International Conference on Science and Technology, ICST 2019*. Institute of Electrical and Electronics Engineers Inc., July 2019, pp. 1–5. isbn: 9781728123691. doi: [10.1109/ICST47872.2019.9166264](https://doi.org/10.1109/ICST47872.2019.9166264). url: <https://ieeexplore.ieee.org/document/9166264/>.
- [35] Hakan Gunduz. “Deep Learning-Based Parkinson’s Disease Classification Using Vocal Feature Sets”. In: *IEEE Access* 7 (Aug. 2019), pp. 115540–115551. issn: 21693536. doi: [10.1109/ACCESS.2019.2936564](https://doi.org/10.1109/ACCESS.2019.2936564).
- [36] Yanhao Xiong and Yaohua Lu. “Deep Feature Extraction from the Vocal Vectors Using Sparse Autoencoders for Parkinson’s Classification”. In: *IEEE Access* 8 (2020), pp. 27821–27830. issn: 21693536. doi: [10.1109/ACCESS.2020.2968177](https://doi.org/10.1109/ACCESS.2020.2968177).
- [37] Amira S. Ashour et al. “A Novel Framework of Two Successive Feature Selection Levels Using Weight-Based Procedure for Voice-Loss Detection in Parkinson’s Disease”. In: *IEEE Access* 8 (2020), pp. 76193–76203. issn: 21693536. doi: [10.1109/ACCESS.2020.2989032](https://doi.org/10.1109/ACCESS.2020.2989032).
- [38] Şule Yücelbaş. “Simple Logistic Hybrid System Based on Greedy Stepwise Algorithm for Feature Analysis to Diagnose Parkinson’s Disease According to Gender”. In: *Arabian Journal for Science and Engineering* 45.3 (Mar. 2020), pp. 2001–2016. issn: 21914281. doi: [10.1007/s13369-020-04357-1](https://doi.org/10.1007/s13369-020-04357-1). url: <https://link.springer.com/article/10.1007/s13369-020-04357-1>.

- [39] Iqra Nissar et al. "Voice-based detection of parkinson's disease through ensemble machine learning approach: A performance study". In: *EAI Endorsed Transactions on Pervasive Health and Technology* 5.19 (2019). issn: 24117145. doi: [10.4108/eai.13-7-2018.162806](https://doi.org/10.4108/eai.13-7-2018.162806).
- [40] Ouiem Bchir. "Parkinson's disease classification using Gaussian mixture models with relevance feature weights on vocal feature sets". In: *International Journal of Advanced Computer Science and Applications* 11.4 (2020), pp. 413–419. issn: 21565570. doi: [10.14569/IJACSA.2020.0110456](https://doi.org/10.14569/IJACSA.2020.0110456). URL: www.ijacsa.thesai.org.
- [41] Lisa Klingelhofer and H. Reichmann. "Parkinson's disease as a multisystem disorder". In: *Journal of Neural Transmission* 124.6 (June 2017), pp. 709–713. doi: [10.1007/S00702-017-1692-0](https://doi.org/10.1007/S00702-017-1692-0).
- [42] Eleni S. Adamidi, Konstantinos Mitsis, and Konstantina S. Nikita. *Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review*. Jan. 2021. doi: [10.1016/j.csbj.2021.05.010](https://doi.org/10.1016/j.csbj.2021.05.010).
- [43] Facundo Bre, Juan M. Gimenez, and Víctor D. Fachinotti. "Prediction of wind pressure coefficients on building surfaces using artificial neural networks". In: *Energy and Buildings* 158 (Jan. 2018), pp. 1429–1441. issn: 03787788. doi: [10.1016/j.enbuild.2017.11.045](https://doi.org/10.1016/j.enbuild.2017.11.045).
- [44] Hossin M and Sulaiman M.N. "A Review on Evaluation Metrics for Data Classification Evaluations". In: *International Journal of Data Mining Knowledge Management Process* 5.2 (2015), pp. 01–11. issn: 2231007X. doi: [10.5121/ijdkp.2015.5201](https://doi.org/10.5121/ijdkp.2015.5201).
- [45] Hüseyin Gürüler. "A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method". In: *Neural Computing and Applications* 28.7 (July 2017), pp. 1657–1666. issn: 09410643. doi: [10.1007/s00521-015-2142-2](https://doi.org/10.1007/s00521-015-2142-2).