

**Δράση «Εμβληματικές δράσεις σε διαθεματικές επιστημονικές περιοχές με ειδικό ενδιαφέρον για την σύνδεση με τον παραγωγικό ιστό» ID 16618**

Εθνικό δίκτυο έρευνας για την ανάδειξη της γενετικής βάσης των νευροεκφυλιστικών νόσων Alzheimer και Parkinson, την ανίχνευση αξιόπιστων βιοδεικτών και την ανάπτυξη καινοτόμων υπολογιστικών τεχνολογιών και θεραπευτικών στρατηγικών στη βάση της ιατρικής ακριβείας (BRAIN PRECISION, TAEDR-0535850)

**ΤΙΤΛΟΣ ΠΑΡΑΔΟΤΕΟΥ: Ανάλυση δεδομένων ακτιγράφων για τη έγκαιρη πρόγνωση νοητικής έκπτωσης**

**ΕΝΟΤΗΤΑ ΕΡΓΑΣΙΑΣ 5: Ανάπτυξη νέων υπολογιστικών μοντέλων και τεχνολογιών για την έγκαιρη διάγνωση των νευροεκφυλιστικών νόσων Alzheimer και Parkinson και των πρόδρομων μορφών τους.**

**ΥΠΕΥΘΥΝΗ ΕΡΕΥΝΗΤΙΚΗ ΟΜΑΔΑ (ΦΟΡΕΑΣ): ΗΛΙΑΣ ΜΑΝΩΛΑΚΟΣ (ΕΚΠΑ)**

## Ανάλυση δεδομένων ακτιγράφων για τη έγκαιρη πρόγνωση νοητικής έκπτωσης

**Φορέας Υλοποίησης:** Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

**Υπεύθυνος:** Καθ. Ηλίας Μανωλάκος

**Συμμετέχων:** Καθ. Νικόλαος Σκαρμέας

**Ε.Ε. :**5

### ΠΕΡΙΛΗΨΗ

Η ακτιγραφία αναδεικνύεται ως μια εξαιρετικά πολύτιμη, χαμηλού κόστους και μη επεμβατική μέθοδος για τη συνεχή παρακολούθηση της ανθρώπινης δραστηριότητας, προσφέροντας κρίσιμα δεδομένα για την έγκαιρη πρόγνωση της Ήπιας Γνωσιακής Εξασθένησης (MCI) και άλλων νευροεκφυλιστικών καταστάσεων. Η συγκεκριμένη έρευνα, που υλοποιήθηκε σε συνεργασία με τη μελέτη ALBION του Αιγινήτειου Νοσοκομείου, αντιμετώπισε την πρόκληση της ανάλυσης περιορισμένων και μη ισορροπημένων συνόλων δεδομένων, όπου τα θετικά δείγματα MCI ή θετικών βιοδεικτών Εγκεφαλονωτιαίου Υγρού (CSF) είναι σημαντικά λιγότερα από τα δείγματα ελέγχου. Κύριος στόχος ήταν η δημιουργία μοντέλων αναφοράς (baseline models) που θα λειτουργούν ως σημείο εκκίνησης για τη μελλοντική έρευνα, επιτρέποντας την αξιολόγηση της προγνωστικής ισχύος των δεδομένων ακτιγραφίας πριν από την εφαρμογή πιο σύνθετων τεχνικών.

Στο επίπεδο της μεθοδολογίας, η έρευνα πρωτοτύπησε χρησιμοποιώντας ταξινομητές χρονοσειρών (όπως ο TSVC και ο 1NNTS) απευθείας πάνω στα δεδομένα, χωρίς την παραδοσιακή και χρονοβόρα διαδικασία της εξαγωγής χαρακτηριστικών (feature engineering). Μέσα από χιλιάδες υπολογιστικά πειράματα σε περιβάλλον cloud, εξετάστηκαν διάφορες τεχνικές προεπεξεργασίας, όπως η εξομάλυνση και η αποσύνθεση σήματος (STL decomposition) σε τάση και εποχικότητα. Ένα από τα σημαντικότερα ευρήματα ήταν ότι οι «φειδωλές» (parsimonious) προσεγγίσεις—όπως η χρήση μόνο της μέσης ημερήσιας ακτιγραφίας ή η εστίαση στο δεύτερο οκτάωρο της ημέρας—προσέφεραν συγκρίσιμη ή και καλύτερη απόδοση, καθώς βοηθούσαν τα μοντέλα να αποφύγουν την υπερπροσαρμογή και να μάθουν πιο γενικευμένα μοτίβα.

Τα αποτελέσματα των μοντέλων κρίθηκαν ενθαρρυντικά, επιτυγχάνοντας διάμεσες τιμές AUC μεταξύ 0,61 και 0,71 για τη πρόβλεψη MCI και 0,58 έως 0,66 για την πρόβλεψη θετικότητας σε βιοδείκτες CSF. Παραταύτα, η ανάλυση του Ποσοστού Σωστής Ταξινόμησης (CCR) ανά δείγμα αποκάλυψε ένα εγγενές «ταβάνι» στην απόδοση των μοντέλων που βασίζονται αποκλειστικά σε χρονοσειρές, κυρίως λόγω της δυσκολίας διαχωρισμού υγιών ατόμων με χαμηλή δραστηριότητα από ασθενείς με υψηλή δραστηριότητα. Τα συμπεράσματα της μελέτης υποδεικνύουν ότι η βελτίωση της διαγνωστικής ακρίβειας απαιτεί τη δημιουργία μεγαλύτερων και πιο ισορροπημένων συνόλων δεδομένων, καθώς και την ενσωμάτωση εξειδικευμένων χαρακτηριστικών (όπως οι μέθοδοι multicoinor) και άλλων βιολογικών δεικτών.

## Contents

ΛΙΣΤΑ ΕΙΚΟΝΩΝ .....	5
ΛΙΣΤΑ ΠΙΝΑΚΩΝ.....	7
1. ΕΙΣΑΓΩΓΗ.....	9
1.1 Η Ακτιγραφία ως μια χαμηλού κόστους προσέγγιση για την παρακολούθηση κύκλων ανάπαυσης και δραστηριότητας .....	9
1.2 Συσκευές ακτιγραφίας .....	9
1.3 Τυπικές περιπτώσεις χρήσης ακτιγραφίας.....	10
1.4 Μηχανική Μάθηση και ακτιγραφία .....	11
1.5 Εφαρμογή μεθόδων εξαγωγής χαρακτηριστικών ακτιγραφίας.....	12
1.6 Κίνητρο και σκοπός της έρευνας μας.....	13
2. ΥΠΟΒΑΘΡΟ.....	14
3. ΜΕΘΟΔΟΙ .....	17
3.1 Επισκόπηση της διεργασίας ταξινόμησης αναφοράς.....	17
3.2 Δεδομένα εκπαίδευσης και διαμόρφωση χώρου λύσεων .....	18
3.3 Μετρικές απόδοσης, σύνολα μοντέλων και αποδεκτά όρια απόδοσης. ....	19
3.4 Ανάλυση Ποσοστού Σωστής Ταξινόμησης (Correct Classification Rate - CCR) .....	20
3.5 Επικύρωση επιλεγμένων μοντέλων .....	20
3.6 Πρόβλεψη άγνωστων δεδομένων .....	21
4. ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ MCI.....	23
4.1 Αποτελέσματα τυχαίου ταξινομητή.....	23
4.2 Αποτελέσματα για τα δεδομένα 6-ημερών .....	23
4.2.1 Αποτελέσματα εξερεύνησης χώρου λύσεων .....	23
4.2.2 Επιλεγμένα μοντέλα .....	25
4.2.3 Επικύρωση επιλεγμένων μοντέλων .....	27
4.2.4 Αξιολόγηση ικανότητας πρόβλεψης σε άγνωστα δεδομένα.....	28
4.2.5 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 6-ημερών .....	30
4.3 Αποτελέσματα για τα δεδομένα 1-ημέρας.....	31
4.3.1 Αποτελέσματα εξερεύνησης χώρου λύσεων .....	31
4.3.2 Επιλεγμένα μοντέλα .....	34
4.3.3 Επικύρωση επιλεγμένων μοντέλων .....	36
4.3.4 Αξιολόγηση ικανότητας πρόβλεψης σε άγνωστα δεδομένα.....	37
4.3.5 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 1-ημερας .....	39
5. ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ CSF .....	41
5.1 Αποτελέσματα τυχαίου ταξινομητή .....	41
5.2 Αποτελέσματα για τα δεδομένα 6-ημερών .....	41
5.2.1 Αποτελέσματα εξερεύνησης χώρου λύσεων .....	41
5.2.2 Επιλεγμένα μοντέλα .....	42
5.2.3 Επικύρωση επιλεγμένων μοντέλων .....	45

5.2.4 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 6-ημερών .....	45
5.3 Αποτελέσματα για τα δεδομένα 1-ημέρας.....	46
5.3.1 Αποτελέσματα εξερεύνησης χώρου λύσεων .....	46
5.3.2 Επιλεγμένα μοντέλα .....	47
5.3.3 Επικύρωση επιλεγμένων μοντέλων .....	48
5.3.4 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 1-ημέρας .....	48
6. ΣΥΜΠΕΡΑΣΜΑΤΑ .....	49
ΠΑΡΑΡΤΗΜΑ.....	50
ΑΝΑΦΟΡΕΣ .....	51

## ΛΙΣΤΑ ΕΙΚΟΝΩΝ

<b>Εικόνα 1:</b> Παραδείγματα Αντιγράφων, τροποποιήθηκε από [11].....	10
<b>Εικόνα 2:</b> α) Παράδειγμα χρονοσειράς ακτιγραφίας ενός συμμετέχοντος, β) γράφημα ράβδων (του μέσου αριθμού σημείων δεδομένων ανά ημέρα για όλες τις ακτιγραφίες στο σύνολο δεδομένων της μελέτης ALBION, γ) Ακατέργαστες ακτιγραφίες για όλους τους συμμετέχοντες της μελέτης ALBION του Ιανουαρίου 2024.....	14
<b>Εικόνα 3:</b> Επισκόπηση διεργασία ανάλυσης απόδοσης ταξινόμησης αναφοράς δεδομένων MCI, CSF.....	17
<b>Εικόνα 4:</b> Κατανομή των διάμεσων μετρικών απόδοσης σε 50 επαναλήψεις mCV (10 γύροι x 5 εξωτερικοί βρόγχοι) για μοντέλα χρονοσειρών 6 ημερών MCI-NC που πέτυχαν διάμεσο ανάκλησης $\geq 0.5$ και ειδικότητας $\geq 0.7$ (21 φιλτραρισμένα πειράματα), ομαδοποιημένα ανάλογα με την προσέγγιση βελτιστοποίησης υπερπαραμέτρων που χρησιμοποιήθηκε για την επιλογή του καλύτερου μοντέλου. Τα μοντέλα που χρησιμοποιούν βελτιστοποίηση υπερπαραμέτρων εμφανίζονται με κόκκινο χρώμα, ενώ τα μοντέλα με προεπιλεγμένες υπερπαραμέτρους εμφανίζονται με μπλε χρώμα. Τα διαγράμματα παρουσιάζουν: α) Ανάκληση (Recall), β) Ειδικότητα (Specificity), γ) AUC και δ) MCC.....	24
<b>Εικόνα 5:</b> Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας MCI-NC για τα πειράματα 6 ημερών που πέτυχαν τουλάχιστον διάμεσο Recall $\geq 0.5$ και διάμεσο Specificity $\geq 0.7$ (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός   χρονική περίοδος   συστατικά αποσύνθεσης".....	25
<b>Εικόνα 6:</b> Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Σύνολο Ελέγχου (N=32, MCI=14, NC=18) και τα μοντέλα του Πίνακα 6. α) ιστόγραμμα που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. β), γ) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI (κατηγορία 1, μπλε γραμμή) και NC (κατηγορία 0, κίτρινη γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 6 MCI και 3 NC "δύσκολων δειγμάτων" με CCR $\leq 0.2$ , αντίστοιχα. δ), ε) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 6 MCI και 12 NC "εύκολων δειγμάτων" (πράσινη γραμμή) με CCR $\geq 0.8$ . .....	29
<b>Εικόνα 7:</b> Κατανομή των διάμεσων μετρικών απόδοσης σε 50 επαναλήψεις mCV (10 γύροι x 5 εξωτερικοί κύκλοι) για μοντέλα χρονοσειρών MCI-NC 1- ημερών και 1-ημερών-χωρίς-αφύσικες-ημέρες που πέτυχαν διάμεσο ανάκλησης $\geq 0.5$ και ειδικότητας $\geq 0.7$ (166 φιλτραρισμένα πειράματα), ομαδοποιημένα ανάλογα με την προσέγγιση βελτιστοποίησης υπερπαραμέτρων που χρησιμοποιήθηκε για την επιλογή του καλύτερου μοντέλου. Τα μοντέλα που χρησιμοποιούν βελτιστοποίηση υπερπαραμέτρων εμφανίζονται με κόκκινο χρώμα, ενώ τα μοντέλα με προεπιλεγμένες υπερπαραμέτρους εμφανίζονται με μπλε χρώμα. Τα διαγράμματα παρουσιάζουν: α) Ανάκληση (Recall), β) Ειδικότητα (Specificity), γ) AUC και δ) MCC.....	32
<b>Εικόνα 8:</b> Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας για τα πειράματα 1- ημερών MCI-NC που πέτυχαν τουλάχιστον διάμεση ανάκληση $\geq 0.5$ και διάμεση ειδικότητα $\geq 0.7$ (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός   χρονική περίοδος   συστατικά αποσύνθεσης. α) Με χρήση τομής χρόνου 08:00-16:00 (61 πειράματα), β) Με χρήση ολόκληρων χρονοσειρών (106 πειράματα).....	33
<b>Εικόνα 9:</b> Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Σύνολο Ελέγχου (N=32, MCI=14, NC=18) και τα μοντέλα του Πίνακα 9. α) ιστόγραμμα που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. β), γ) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI (κατηγορία 1, μπλε γραμμή) και NC (κατηγορία 0, κίτρινη γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 6 MCI και 2 NC "δύσκολων δειγμάτων" με CCR $\leq 0.2$ , αντίστοιχα. δ), ε) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 6 MCI και 12 NC "εύκολων δειγμάτων" (πράσινη γραμμή) με CCR $\geq 0.8$ . .....	39
<b>Εικόνα 10:</b> Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας CSF test positive-CSF test negative για τα πειράματα 6 ημερών που πέτυχαν τουλάχιστον διάμεσο Recall $\geq 0.5$ και διάμεσο Specificity $\geq 0.667$ (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός   χρονική περίοδος   συστατικά αποσύνθεσης".....	42
<b>Εικόνα 11:</b> Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Αρχικά CSF δεδομένα Ελέγχου (N=76, CSF test positive=16, CSF test negative=70) και τα μοντέλα του Πίνακα 11. α) ιστόγραμμα	

που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. b), c) Μέσες τιμές ακτιγραφίας για τις κατηγορίες CSF test positive (κατηγορία 1, πορτοκαλί γραμμή) και CSF test negative (κατηγορία 0, μπλε γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 8 CSF test negative και 4 CSF test positive "δύσκολων δειγμάτων" με  $CCR \leq 0.2$ , αντίστοιχα. d), e) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 30 CSF test negative και 3 test positive "εύκολων δειγμάτων" (πράσινη γραμμή) με  $CCR \geq 0.8$ . ..... 44

## ΛΙΣΤΑ ΠΙΝΑΚΩΝ

<b>Πίνακας 1:</b> Σύνολο δεδομένων MCI .....	15
<b>Πίνακας 2:</b> Σύνολο δεδομένων CSF .....	16
<b>Πίνακας 3:</b> Αποτελέσματα τυχαίου ταξινομητή για το σύνολο δεδομένων MCI .....	23
<b>Πίνακας 4:</b> Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις mCV για τα επιλεγμένα πειράματα 6 ημερών χρονοσειρών MCI-NC, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης (recall), χαμηλού ορίου 95% CI του διάμεσου ανάκλησης και χαμηλού ορίου 95% CI του διάμεσου ειδικότητας (specificity). Ο πίνακας χωρίζεται σε τρεις ζώνες, διαχωρισμένες με οριζόντιες γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος mCV .....	26
<b>Πίνακας 5:</b> Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 6 ημερών MCI-NC. Τα αποτελέσματα της απόδοσης mCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο. ....	27
<b>Πίνακας 6:</b> Απόδοση των επιλεγμένων μοντέλων 6-ημέρων που αξιολογήθηκαν σε ένα μικρό σύνολο ελέγχου αποτελούμενο από μόλις 32 δείγματα (MCI = 14, NC = 18). Τα μοντέλα είναι ταξινομημένα κατά φθίνουσα σειρά ανάκλησης και ειδικότητας. Εμφανίζονται μόνο τα μοντέλα που πέτυχαν ανάκληση $\cong 0.5$ και ειδικότητα $\cong 0.7$ . Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε μοντέλου. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο .....	28
<b>Πίνακας 7:</b> Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις mCV για τα επιλεγμένα πειράματα με μέσες ακτιγραφίες μιας ημέρας χρονοσειρών MCI-NC, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου ορίου 95% CI του διάμεσου ανάκλησης και χαμηλότερου ορίου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε τέσσερις ζώνες, διαχωρισμένες με γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος mCV .....	35
<b>Πίνακας 8:</b> Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 1 ημέρας MCI-NC. Τα αποτελέσματα της απόδοσης mCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο. ....	37
<b>Πίνακας 9:</b> Απόδοση των επιλεγμένων μοντέλων 1-ημέρας MCI-NC που αξιολογήθηκαν σε ένα μικρό σύνολο ελέγχου αποτελούμενο από μόλις 32 δείγματα (MCI = 14, NC = 18). Τα μοντέλα είναι ταξινομημένα κατά φθίνουσα σειρά ανάκλησης και ειδικότητας. Εμφανίζονται μόνο τα μοντέλα που πέτυχαν ανάκληση $\cong 0.5$ και ειδικότητα $\cong 0.7$ . Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε μοντέλου. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.....	37
<b>Πίνακας 10:</b> Αποτελέσματα ταξινομητή αναφοράς για το σύνολο δεδομένων CSF .....	41
<b>Πίνακας 11:</b> Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις mCV για τα επιλεγμένα πειράματα 6 ημερών χρονοσειρών CSF test negative-CSF test positive, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου 95% CI του διάμεσου ανάκλησης και χαμηλότερου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε δύο ζώνες, διαχωρισμένες με γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος mCV .....	43
<b>Πίνακας 12:</b> Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 6 ημερών CSF test positive – CSF test negative. Τα αποτελέσματα της απόδοσης mCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.....	45
<b>Πίνακας 13:</b> Διάμεσες μετρικές απόδοσης από τις 40 επαναλήψεις mCV για τα επιλεγμένα πειράματα 1- ημέρας χρονοσειρών CSF test negative-CSF test positive, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου 95% CI του διάμεσου ανάκλησης και χαμηλότερου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε δύο ζώνες, διαχωρισμένες με γραμμές, με βάση τον τύπο του ταξινομητή που χρησιμοποιείτε. ....	47
<b>Πίνακας 14:</b> Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους	

αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 1-ημέρας CSF test positive – CSF test negative. Τα αποτελέσματα της απόδοσης rCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο..... 48

**Πίνακας 15:** Χώρος υπερπαραμέτρων ταξινομητών που χρησιμοποιήθηκαν στην ταξινόμηση αναφοράς..... 50

**Πίνακας 16:** Προεπιλεγμένες υπερπαραμέτροι ταξινομητών που χρησιμοποιήθηκαν στην ταξινόμηση αναφοράς ..... 50

## 1. ΕΙΣΑΓΩΓΗ

Σε αυτό το εδάφιο παρουσιάζεται ολοκληρωμένη επισκόπηση της ακτιγραφίας, ξεκινώντας από την ιστορική της εξέλιξη ως μεθόδου καταγραφής κύκλων δραστηριότητας και ανάπαυσης. Επιπλέον παρουσιάζουμε λεπτομερώς τις κύριες συσκευές ακτιγραφίας, αναλύοντας τις τεχνικές προδιαγραφές και τον τρόπο δημιουργίας των σχετικών συνόλων δεδομένων (χρονοσειρών). Στο επίκεντρο της ανάλυσης βρίσκονται οι εφαρμογές της ακτιγραφίας στον χώρο της υγειονομικής περιθαλψής, ενώ ιδιαίτερη έμφαση δίνεται στον ρόλο της μηχανικής μάθησης και των τεχνικών εξαγωγής χαρακτηριστικών κατά την επεξεργασία και ανάλυση των χρονοσειρών ακτιγραφίας. Το κεφάλαιο ολοκληρώνεται με την παρουσίαση του κεντρικού ερευνητικού κινήτρου και σκοπού της μελέτης.

### 1.1 Η Ακτιγραφία ως μια χαμηλού κόστους προσέγγιση για την παρακολούθηση κύκλων ανάπαυσης και δραστηριότητας

Η ακτιγραφία είναι μη επεμβατική μέθοδος καταγραφής δραστηριότητας μέσω της αξιολόγησης της κίνησης, χρησιμοποιώντας μια μικρή συσκευή που συνήθως φοριέται σαν ρολόι. Η ακτιγραφία δεν αντικαθιστά τις κλινικές συνεντεύξεις, τα ημερολόγια ύπνου ή την πολυυπνογραφία (PSG), η οποία είναι το "χρυσό πρότυπο" για τη μελέτη του ύπνου. Ωστόσο, σε σύγκριση με την PSG, η οποία απαιτεί μια εκτενή εγκατάσταση για την παρακολούθηση των φάσεων ύπνου, των κινήσεων των άκρων, της ροής του αέρα, της αναπνευστικής προσπάθειας, του καρδιακού ρυθμού, των επιπέδων οξυγόνου και της θέσης του σώματος σε ένα εργαστήριο με τεχνικό προσωπικό, η ακτιγραφία προσφέρει πολύτιμες πληροφορίες για τα μοτίβα δραστηριότητας/ξεκούρασης σε συνθήκες ελεύθερης ζωής, σε εκτεταμένες χρονικές περιόδους και με πολύ χαμηλότερο κόστος [1].

Η καθορισμός των μοτίβων ανάπαυσης- δραστηριότητας σε μη εργαστηριακό περιβάλλον αναφέρθηκε για πρώτη φορά ως "ακτιγραφία" το 1978, με μια μελέτη που συγκρίνει τις δυνατότητες μιας συσκευής που χρησιμοποιεί πιεζοηλεκτρικούς μετατροπείς με την ηλεκτροεγκεφαλογραφία (EEG) και άλλες τεχνικές για την αξιολόγηση κύκλων ύπνου-εγρήγορσης, όπου η συσκευή ακτιγραφίας παρουσίασε μόνο ελαφρώς χειρότερα αποτελέσματα [2]. Το χαμηλό κόστος της μεθόδου και η πρώτη επιτυχία οδήγησαν στην πρώτη μεγάλη μελέτη για την επικύρωση της ακτιγραφίας με 102 συμμετέχοντες, όπου οι ερευνητές δεν κατάφεραν να δείξουν αξιόλογα αποτελέσματα από την ακτιγραφία σε ηλικιωμένα άτομα ή άτομα με διαταραχές ύπνου [3]. Το 1982, η αποτελεσματικότητα και η ακρίβεια της μεθόδου αυξήθηκαν με τη δημιουργία και ενσωμάτωση ενός υπολογιστικού αλγορίθμου που ταξινομεί τις καταστάσεις ύπνου/εγρήγορσης, πετυχαίνοντας ακρίβεια 93% σε σύγκριση με την EEG [4]. Ένα ακόμα σημαντικό επίτευγμα στην ανάπτυξη αυτόματων αλγορίθμων καταμέτρησης για την ακτιγραφία σημειώθηκε το 1994, όταν οι ερευνητές έδειξαν συσχέτιση 91-93% μεταξύ των δεδομένων ακτιγραφίας και των αποτελεσμάτων της PSG [5]. Τέλος, το 1995, το Διοικητικό Συμβούλιο της Αμερικανικής Ένωσης Διαταραχών Ύπνου ενέκρινε μια έκθεση από κορυφαίους ερευνητές για την ακτιγραφία ως εργαλείο έρευνας για τη μέτρηση των μοτίβων ανάπαυσης- δραστηριότητας σε όλες τις ηλικιακές ομάδες [6]. Μετά από αυτή την έγκριση, η ακτιγραφία είχε ευρεία αναγνώριση ως αξιόπιστη και οικονομική μέθοδος παρακολούθησης της δραστηριότητας.

### 1.2 Συσκευές ακτιγραφίας

Υπάρχουν πολλές συσκευές ακτιγραφίας ή "ακτιγράφοι" διαθέσιμες σήμερα. Ορισμένες δημοφιλείς επιλογές είναι το Actrust 2 από την Condor Instruments [7], το Actiwatch 2 από την Philips [8], το GENEActiv από την Activinsights [9] και το ActiGraph wGT3X-BT από την ActiGraph [10]. Όλοι οι ακτιγράφοι μοιάζουν με ρολόι καρπού, όπως φαίνεται στην Εικόνα

1. Διαφέρουν όμως ως προς το διάστημα καταγραφής και τις προδιαγραφές της μπαταρίας τους. Οι περισσότερες καταγράφουν την κίνηση χρησιμοποιώντας τριαξονικό μετρητή επιτάχυνσης, με το Actiwatch 2 να χρησιμοποιεί μονοαξονικό. Είναι σημαντικό να σημειωθεί ότι οι συσκευές με περισσότερους άξονες είναι πιο ευαίσθητες στην κίνηση [11]. Η επιλογή της κατάλληλης συσκευής ακτιγραφίας είναι κρίσιμη και συνήθως γίνεται βάση του πρωτόκολλου της ερευνητικής μελέτης. Για παράδειγμα, μελέτες που απαιτούν αραιά δεδομένα δραστηριότητας για εκτεταμένες χρονικές περιόδους μπορεί να επωφεληθούν από συσκευές με μεγάλη διάρκεια ζωής μπαταρίας και μεγάλες περιόδους καταγραφής.



**Εικόνα 1:** Παραδείγματα Αντιγράφων, τροποποιήθηκε από [11]

Όλες οι συσκευές ακτιγραφίας διαθέτουν ενσωματωμένο λογισμικό που αναλύει την κίνηση που καταγράφεται σε περιόδους και την μετατρέπει σε μετρήσεις δραστηριότητας, ποσοτικοποιώντας την επιτάχυνση σε κάθε περίοδο. Αν και οι αλγόριθμοι που χρησιμοποιούνται στις σύγχρονες συσκευές δεν είναι ανοιχτά διαθέσιμοι, μπορούμε να αναφερθούμε στις αρχικές μεθόδους που δημοσιεύτηκαν για τα παλαιότερα μοντέλα.

Οι αισθητήρες επιταχυνσιόμετρου ανιχνεύουν αλλαγές στην επιτάχυνση και τις μετατρέπουν σε ηλεκτρικά σήματα. Εφαρμόζεται ένα φίλτρο χαμηλής διέλευσης, συνήθως με όριο 0.21 Hz, για να φιλτραριστούν οι ελάχιστες κινήσεις και να μην καταγραφούν. Εάν η διαδικασία μέτρησης είναι αρχικά αναλογική, το σήμα ψηφιοποιείται χρησιμοποιώντας έναν αναλογικό-ψηφιακό μετατροπέα. Για παράδειγμα, αν χρησιμοποιούνται 8 bits, μπορούν να καταγραφούν  $2^8=256$  μοναδικά επίπεδα σήματος, και έτσι να διακριθούν 256 επίπεδα ακτιγραφίας. Τέλος, τα δεδομένα λαμβάνονται με μια καθορισμένη συχνότητα (Hz), η οποία αποφασίζει το διάστημα για την καταγραφή των μετρήσεων δραστηριότητας [12].

Οι περισσότερες συσκευές ακτιγραφίας χρησιμοποιούν δεδομένα μέτρησης δραστηριότητας για να δημιουργήσουν αναφορές, περιλαμβάνοντας και μια οπτική αναπαράσταση των μοτίβων δραστηριότητας με την πάροδο του χρόνου, όπου φαίνονται τα χρονικά σημεία έναρξης και λήξης του ύπνου. Επιπλέον, παρέχουν και βασικούς δείκτες ύπνου όπως ο Συνολικός Χρόνος Ύπνου (TST), η Εγρήγορση Μετά την Έναρξη Ύπνου (WASO), η Καθυστέρηση Έναρξης Ύπνου (SOL), ο Αναλογικός Δείκτης Ύπνου-Ξυπνήματος (SWR) και η Καθυστέρηση Ύπνου (SL) [11].

### 1.3 Τυπικές περιπτώσεις χρήσης ακτιγραφίας

Η τυπική χρήση της ακτιγραφίας στον τομέα της υγειονομικής περίθαλψης είναι ως μια οικονομική, μη επεμβατική συμπληρωματική μέθοδο για την παρακολούθηση διαταραχών που επηρεάζουν τον ύπνο και τον κερκάρδιο ρυθμό.

Τα χαρακτηριστικά της αϋπνίας μπορούν να καταγραφούν αξιόπιστα με την ακτιγραφία, με τα αποτελέσματα για το WASO, TST και την αποδοτικότητα ύπνου να ταιριάζουν στενά με αυτά της PSG. Οι ασθενείς με αϋπνία εμφανίζουν σημαντικά χειρότερες μετρικές ύπνου, όπως SOL, TST, WASO, αποδοτικότητα ύπνου και αριθμό αφυπνήσεων, καθιστώντας αυτούς τους δείκτες αποτελεσματικούς στο να διακρίνουν τους ασθενείς με αϋπνία και να επιτυγχάνουν υψηλά ποσοστά ακρίβειας κατά την πρόβλεψη τους [13]. Μια άλλη χρήση της

ακτιγραφίας είναι στη διάγνωση διαταραχών αναπνοής κατά τον ύπνο, όπως η άπνοια, με τα αποτελέσματα να είναι παρόμοια με την PSG στην εκτίμηση των μοτίβων ύπνου/εγρήγορης [14].

Η ακτιγραφία παίζει επίσης σημαντικό ρόλο στη μελέτη ψυχικών διαταραχών, όπως η κατάθλιψη και η διπολική διαταραχή. Μια μετα-ανάλυση για το θέμα ανασκόπησε 38 μελέτες. Κατέληξε στο συμπέρασμα ότι τα άτομα με κατάθλιψη ήταν λιγότερο σωματικά δραστήρια και περνούσαν περισσότερο χρόνο ξύπνια μετά την αρχική έναρξη του ύπνου σε σύγκριση με τα υγιή άτομα. Εκείνοι σε ύφεση ή με σταθεροποιημένα συμπτώματα κοιμόντουσαν περισσότερο, χρειάζονταν περισσότερο χρόνο για να κοιμηθούν και παρέμεναν ξύπνιοι για περισσότερο χρόνο κατά τη διάρκεια της νύχτας σε σχέση με τα υγιή άτομα. Μετά τη θεραπεία, παρατηρήθηκαν σημαντικές βελτιώσεις στην συνολική αποδοτικότητα ύπνου τους [15].

Τέλος, η ακτιγραφία είναι σημαντικό εργαλείο και στη μελέτη της έκπτωσης της γνωστικής λειτουργίας. Χαμηλότερα μέση επίπεδα δραστηριότητας, μειωμένο μέγεθος δραστηριότητας και νωρίτερα και αργότερα χρονικά σημεία κορυφής στο μέγεθος της δραστηριότητας έχουν συνδεθεί με υψηλότερο κίνδυνο έκπτωσης της γνωστικής λειτουργίας. Οι δείκτες ύπνου επίσης υποδεικνύουν ότι ο υψηλότερος κίνδυνος γνωστικής έκπτωσης και άνοιας σχετίζεται με χαμηλότερη αποδοτικότητα ύπνου, μεγαλύτερη καθυστέρηση στην έναρξη του ύπνου και μεγαλύτερη μεταβλητότητα στην συνολική ποσότητα ύπνου [16].

#### 1.4 Μηχανική Μάθηση και ακτιγραφία

Όπως συζητήθηκε στην προηγούμενη ενότητα, η ακτιγραφία παρέχει σημαντικές πληροφορίες σε διάφορους τομείς υγείας. Αυτές συνοδεύονται, όμως και από αρκετές επιτυχημένες πρωτοβουλίες για τη συλλογή, αποθήκευση και διανομή μεγάλων συνόλων δεδομένων ακτιγραφίας. Για παράδειγμα, η UK Biobank συνέλεξε δεδομένα δραστηριότητας 7 ημερών από 100.000 συμμετέχοντες το 2012 [17]. Επιπλέον, η National Sleep Research Resource φιλοξενεί καταγραφές ακτιγραφίας για πάνω από 26.800 άτομα από την ίδρυσή της [18]. Οι μέθοδοι Μηχανικής Μάθησης (MM) και Βαθιάς Μάθησης (BM) μπορούν φυσικά να εκμεταλλευτούν αυτή τη μαζική αύξηση στη συλλογή δεδομένων ακτιγραφίας.

Η πλειοψηφία των ερευνητικών εργασιών που συνδυάζουν την ακτιγραφία με τη Μηχανική/Βαθιά Μάθηση σχετίζεται με την αύξηση της αποτελεσματικότητας της ακτιγραφίας στην αξιολόγηση του ύπνου, αντικαθιστώντας τους παραδοσιακούς αλγόριθμους με προγνωστικά μοντέλα. Ωστόσο, αν και αυτά τα μοντέλα εκπαιδεύονται σε μεγάλα σύνολα δεδομένων, τα αποτελέσματα που προσφέρουν συχνά, ξεπερνιούνται από απλές ευριστικές μεθόδους και κλασικούς αλγόριθμους, με τους ερευνητές να θεωρούν ότι οι διαφορές στους υπολογισμούς δραστηριότητας ανάμεσα σε ακτιγράφους είναι το κύριο ζήτημα πίσω από αυτά τα αποτελέσματα [19].

Τα μοντέλα MM μπορούν επίσης να χρησιμοποιηθούν ως εργαλεία ελέγχου για να βοηθήσουν τους επαγγελματίες υγείας να εντοπίσουν άτομα που διατρέχουν κίνδυνο για ορισμένες καταστάσεις ή ασθένειες πριν από τη διενέργεια πιο λεπτομερών διαγνωστικών εξετάσεων, βοηθώντας έτσι στην αποδοτική κατανομή πόρων και στη βελτίωση της θεραπείας των ασθενών. Τα δεδομένα ακτιγραφίας έχουν χρησιμοποιηθεί για την εκπαίδευση αλγορίθμων MM για την ταξινόμηση των σταδίων κατάθλιψης σε ήπια, μέτρια και σοβαρή, με σκορ Εμβαδού Κάτω από την Καμπύλη (Area Under the Curve - AUC) 0.99 [20]. Οι αλγόριθμοι MM έχουν δείξει την ικανότητά τους να διακρίνουν την οξεία από τη χρόνια αϋπνία και τους συμμετέχοντες με αϋπνία από υγιή άτομα, με AUC 0.81 και 0.62, αντίστοιχα [24]. Οι μέθοδοι βαθιάς μάθησης (BM) έχουν εφαρμοστεί για να επιτευχθεί AUC 0.86 για την ταξινόμηση των ατόμων με φυσιολογική γνωστική λειτουργία (Normal Controls - NC) από ασθενείς με Αλτσχάιμερ (AD) και AUC 0.61 για την ταξινόμηση των NC ατόμων

από ασθενείς με Ήπια Γνωσιακή Εξασθένηση (MCI), η οποία είναι το στάδιο που χωρίζει την άνοια από τη φυσιολογική μείωση της γνωστικής απόδοσης λόγω γήρανσης [21].

### 1.5 Εφαρμογή μεθόδων εξαγωγής χαρακτηριστικών ακτιγραφίας

Η εξαγωγή χαρακτηριστικών αποτελεί κρίσιμο βήμα στην ανάλυση των δεδομένων ακτιγραφίας, καθώς επιτρέπει την απομόνωση σχετικών πληροφοριών από τα δεδομένα κίνησης με πολύ μικρότερο αριθμό χαρακτηριστικών σε σχέση με το σύνολο των χρονικών σημείων της ακτιγραφίας. Τα χαρακτηριστικά αυτά χρησιμοποιούνται για την αξιολόγηση ή την ταξινόμηση της υγειονομικής κατάστασης του ατόμου. Η επιλογή των κατάλληλων μεθόδων εξαγωγής χαρακτηριστικών εξαρτάται από τον στόχο της ανάλυσης και τις απαιτήσεις της εκάστοτε εφαρμογής. Δυο από τις πιο γνωστές μεθόδους ανάλυσης είναι η μέθοδος *cosinor* και η μη-παραμετρική ανάλυση της ακτιγραφίας.

Η πρώτη οδηγεί στην εξαγωγή χαρακτηριστικών από το ταίριασμα καμπυλών ημίτονων στα δεδομένα χρονοσειράς της ακτιγραφίας. Για κάθε καμπύλη εξάγονται 3 χαρακτηριστικά: η μέση τιμή της καμπύλης (*mesor*), η χρονική στιγμή της μέγιστης τιμής της καμπύλης (*acrophase*), η διαφορά (*amplitude*) μεταξύ της μέγιστης δραστηριότητας και του *mesor*. Η *cosinor* ανάλυση είναι ερμηνεύσιμη και δίνει αξιόλογα αποτελέσματα όταν ο ρυθμός της ακτιγραφίας προσεγγίζει ημιτονοειδείς καμπύλες [22].

Η μη-παραμετρική ανάλυση, αποφεύγει το να βασίζεται σε προκαθορισμένες υποθέσεις σχετικά με τη μορφή της χρονοσειράς της ακτιγραφίας. Σε αντίθεση με τη μέθοδο *cosinor*, η οποία προσφέρει ευελιξία στην ανάλυση διαφορετικών περιόδων, η μη-παραμετρική ανάλυση εστιάζει συνήθως σε μια συγκεκριμένη περίοδο, όπως οι 24 ώρες, για την αξιολόγηση των ημερήσιων ρυθμών. Η μη-παραμετρική προσέγγιση επιτρέπει την εξαγωγή πέντε βασικών χαρακτηριστικών [23]:

- Ενδοημερήσια Μεταβλητότητα (*Intradaily Variability - IV*): Μετράει τη μεταβλητότητα του ρυθμού εντός της ίδιας ημέρας.
- Διαημερήσια Σταθερότητα (*Interdaily Stability - IS*): Αξιολογεί τη σταθερότητα και τη συνοχή του ρυθμού από τη μια μέρα στην άλλη.
- Μέσος Όρος Δραστηριότητας κατά τις 5 Ώρες με τη Χαμηλότερη Δραστηριότητα (*L5*): Υπολογίζει τον μέσο όρο της δραστηριότητας κατά τη διάρκεια των 5 ωρών με τη μικρότερη δραστηριότητα.
- Μέσος Όρος Δραστηριότητας κατά τις 10 Ώρες με την Υψηλότερη Δραστηριότητα (*M10*): Υπολογίζει τον μέσο όρο της δραστηριότητας κατά τη διάρκεια των 10 ωρών με την υψηλότερη δραστηριότητα.
- Μετρική RA (*Relative Amplitude*): Είναι ο λόγος  $(M10-L5) / (M10+L5)$  που κοιμάται μεταξύ 0 και 1.

Η ανάλυση της ακτιγραφίας, χρησιμοποιώντας τόσο την μέθοδο *cosinor* όσο και την μη-παραμετρική μέθοδο, έχει εφαρμοστεί ευρέως σε διάφορους τομείς υγείας, αποκαλύπτοντας σημαντικές πληροφορίες σχετικά με διαταραχές ύπνου, καρδιαγγειακές παθήσεις, διαβήτη και γνωσιακές διαταραχές.

Στην έρευνα για τις διαταραχές ύπνου, μελέτες των Fossion et al. (2017) και Kim et al. (2019) δεν βρήκαν σημαντικές διαφορές στις μη-παραμετρικές και *cosinor* παραμέτρους (π.χ. *IV*, *Mesor*, *Amplitude*) μεταξύ ασθενών με αϋπνία και NC ατόμων, αν και οι ασθενείς με αϋπνία έδειξαν καθυστερημένα *Acrophase* και διαφορετικές των φυσιολογικών μετρικές ύπνου όπως το *WASO* και το *SE*. Αυτά τα ευρήματα υπογραμμίζουν τη σημασία της προσαρμογής των μεθόδων εξαγωγής χαρακτηριστικών σε συγκεκριμένους ερευνητικούς στόχους, όπως η εστίαση σε μετρικές ύπνου για μελέτες της αϋπνίας [24] [25].

Στην καρδιολογία, οι Yeung et al. (2023) συνέδεσαν χαμηλότερο Amplitude στην ανάλυση cosinor με υψηλότερο κίνδυνο υπέρτασης, ενώ οι Yang et al. (2023) συνέδεσαν μη-παραμετρικά χαρακτηριστικά όπως χαμηλότερο IS και RA με καρδιαγγειακή θνησιμότητα σε ασθενείς με διαβήτη τύπου 2. Ομοίως, οι Ferreira et al. (2017) και Xiao et al. (2020) τόνισαν διαταραχές στον κirkάδιο ρυθμό στον διαβήτη τύπου 2, με χαμηλότερο Amplitude και καθυστερημένη Acrophase να αυξάνουν τον κίνδυνο διαβήτη [26], [27], [28], [29].

Στην έρευνα για άνοια και ήπια γνωστική διαταραχή (MCI), μελέτες συνέδεσαν χαμηλότερο Amplitude, υψηλότερο IV και καθυστερημένη Acrophase με αυξημένο κίνδυνο MCI και νόσου Alzheimer (AD). Οι Musiek et al. (2018) και Lee et al. (2021) επεσήμαναν αυτές τις συσχετίσεις, ενώ μια συστηματική ανασκόπηση των Cote et al. (2021) επιβεβαίωσε χαμηλότερο IS, RA και Amplitude σε ασθενείς με άνοια σε σύγκριση με NC άτομα [30], [31], [32].

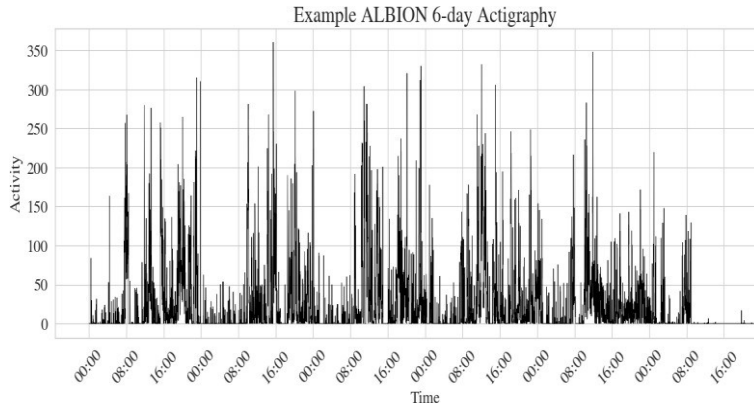
## 1.6 Κίνητρο και σκοπός της έρευνας μας

Μεγάλο μέρος της ανάλυσης ακτιγραφίας βασίζεται στις δύο προαναφερθείσες μεθόδους. Στην παρούσα έρευνα, δουλεύοντας με ένα σχετικά μικρό σύνολο δεδομένων από την χώρα μας που παρουσιάζεται αναλυτικά στο επόμενο εδάφιο, αποφασίσαμε να ξεκινήσουμε την ανάλυσή μας από τα δεδομένα ακτιγραφίας, χωρίς την εξαγωγή χαρακτηριστικών, ώστε να δημιουργήσουμε μια βάση αναφοράς (baseline) πάνω στην οποία οι μελλοντικοί ερευνητές θα μπορούν να στηριχθούν και να την εξελίξουν.

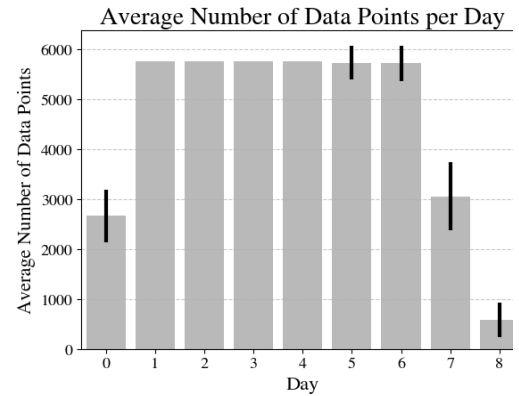
Πιστεύουμε ότι η εξέταση του σήματος της ακτιγραφίας καθαυτού, χωρίς τη χρήση εξαγωγής διανύσματος χαρακτηριστικών, για είσοδο σε τεχνικές μηχανικής μάθησης προσφέρει μια πρωτοποριακή προσέγγιση που δεν έχει εξεταστεί επαρκώς. Σε συνδυασμό με τη χρήση ταξινομητών χρονοσειρών αποκαλύπτει σημαντικές πληροφορίες σχετικά με το σύνολο δεδομένων, οι οποίες μπορούν να συνεισφέρουν ιδέες για τον σχεδιασμό μετέπειτα αναλύσεων πέραν της παρούσας ανάλυσης αναφοράς (baseline study).

## 2. ΥΠΟΒΑΘΡΟ

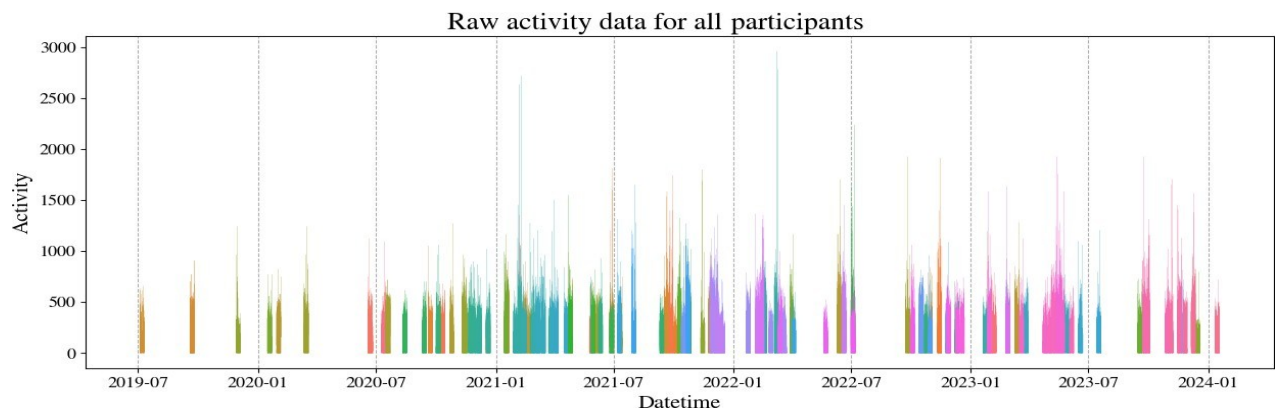
a)



b)



c)



**Εικόνα 2:** a) Παράδειγμα χρονοσειράς ακτιγραφίας ενός συμμετέχοντος, b) γράφημα ράβδων (του μέσου αριθμού σημείων δεδομένων ανά ημέρα για όλες τις ακτιγραφίες στο σύνολο δεδομένων της μελέτης ALBION, c) Ακατέργαστες ακτιγραφίες για όλους τους συμμετέχοντες της μελέτης ALBION του Ιανουαρίου 2024.

Η έρευνα αυτή πραγματοποιήθηκε σε συνεργασία με την ερευνητική ομάδα Βιοδεικτών Νευροεκφύλισης του Αιγινήτειου Νοσοκομείου (ALBION) [33], [34] που καθοδηγεί ο Καθηγητής Νικόλαος Σκαρμέας. Η μελέτη ALBION σχεδιάστηκε για να διερευνήσει τα προδρομικά στάδια της νόσου Alzheimer (AD) πριν την πλήρη εκδήλωση των συμπτωμάτων. Αξιολογεί τους συμμετέχοντες συλλέγοντας ένα ευρύ φάσμα δεδομένων, όπως βιολογικά δείγματα, εργαστηριακές εξετάσεις, ερωτηματολόγια και πληροφορίες ιατρικές εξετάσεις, σε διάστημα δέκα ετών με ετήσια επαναλαμβανόμενη αξιολόγηση. Με την ανάλυση αυτών των δεδομένων κατά τη διάρκεια του χρόνου, η μελέτη στοχεύει να ανακαλύψει τις βιολογικές διαδικασίες που οδηγούν στην έναρξη της νόσου Alzheimer. η κατανόηση αυτών των πρώιμων αλλαγών θα μπορούσε να οδηγήσει σε καλύτερες θεραπείες και στρατηγικές πρόληψης της νόσου.

Οι συμμετέχοντες στη μελέτη ALBION είναι άτομα άνω των 40 ετών με θετικό οικογενειακό ιστορικό Alzheimer ή υποψία νοητικής έκπτωσης. Οι συμμετέχοντες μπορούν να ανήκουν σε μία από τις τρεις κατηγορίες: α) γνωσιακά φυσιολογικοί (NC) ηλικιωμένοι, β) άτομα με υποκειμενική νοητική έκπτωση και γ) άτομα με ήπια γνωσιακή διαταραχή (MCI) που δεν πληρούν τα κριτήρια για διάγνωση άνοιας. Οι συμμετέχοντες αποκλείονται από τη μελέτη

αν έχουν 1) διάγνωση άνοιας, 2) νευρολογικές, ψυχιατρικές ή ιατρικές παθήσεις με αυξημένο κίνδυνο γνωσιακής διαταραχής, 3) εμφυτευμένες συσκευές που ενδέχεται να αποτελούν κίνδυνο ή να επηρεάζουν την ποιότητα των μαγνητικών τομογραφιών (MRI), ή 4) λαμβάνουν αντιπηκτική αγωγή.

Οι μετρήσεις ακτιγραφίας, που αποτελούν το βασικό ενδιαφέρον της παρούσας ανάλυσης, συλλέγονται χρησιμοποιώντας το μονοαξονικό Actiwatch 2 (Phillips-Respironics) [8]. Οι συμμετέχοντες καλούνται να φορούν τον ακτιγράφο στον καρπό του μη κυρίαρχου χεριού τους για 7 ημέρες μετά την πρώτη τους επίσκεψη στην κλινική, με καταγραφές δραστηριότητας κάθε 15 δευτερόλεπτα. Τα αρχεία κάθε συμμετέχοντος εξάγονται από τον ακτιγράφο και αποθηκεύονται προσωπικό σε μορφή Excel. Νέοι συμμετέχοντες προστίθενται στη μελέτη ALBION κάθε τέσσερις μήνες, στην παρούσα εργασία χρησιμοποιήθηκε το σύνολο δεδομένων ALBION του Ιανουαρίου 2024. Από τις 136 ακτιγραφίες που περιλαμβάνονται σε αυτό το σύνολο δεδομένων, οι 122 προέρχονται από την πρώτη αξιολόγηση των συμμετεχόντων, ενώ μόνο 14 προέρχονται από δεύτερη αξιολόγηση. Ένα δείγμα ακτιγραφίας απεικονίζεται στην Εικόνα 2 α). Κατά μέσο όρο, οι ακτιγραφίες των συμμετεχόντων περιλαμβάνουν μόνο 6 ημέρες (ημέρα 1 έως ημέρα 6) πλήρους 24ωρης δραστηριότητας, από μεσάνυχτα έως μεσάνυχτα, όπως παρουσιάζεται στην Εικόνα 2 β). Οι ημερομηνίες συλλογής των ακτιγραφιών κυμαίνονται από το 2019 έως το 2024, όπως αναφέρεται λεπτομερώς στην Εικόνα 2 γ).

Η μελέτη ανάλυση που παρουσιάζεται σε αυτό το έγγραφο εστιάζει στην ανάπτυξη προγνωστικών μοντέλων ταξινόμησης αναφοράς για δύο στόχους από τη βάση δεδομένων ALBION:

- MCI έναντι Υγιών Ατόμων (NC): Οι ετικέτες (0=NC ή 1=MCI) καθορίζονται με βάση γνωστικές αξιολογήσεις και νευροψυχολογικά τεστ και χωρίς τη χρήση ακτιγραφίας.
- CSF test positive έναντι CSF test negative: Οι ετικέτες (0 ή 1) καθορίζονται από βιοδείκτες CSF (Aβ42, Tau, Phospho-Tau) χρησιμοποιώντας συγκεκριμένα εργαστηριακά πρωτόκολλα.

Τα δεδομένα ακτιγραφίας συγχρονίζονται με τις ημερομηνίες διάγνωσης εντός καθορισμένων χρονικών παραθύρων (90 ημέρες για MCI, 360 ημέρες για CSF). Εφαρμόζονται κανόνες βάσει της χρονικής σειράς της ακτιγραφίας και της διάγνωσης για την εξαγωγή ετικετών σε μη συγχρονισμένα δείγματα, διασφαλίζοντας έτσι ακριβή επισήμανση για την εφαρμογή μηχανικής μάθησης. Τα σύνολα δεδομένων αποτελούνται από 127 δείγματα NC-MCI (98 NC, 29 MCI) και 86 δείγματα για πρόβλεψη CSF (22 θετικά, 64 αρνητικά), όπως φαίνεται και στους Πίνακες 1, 2.

<i>MCI-Healthy participants dataset</i>		
<b>Actigraphy Sample Category</b>	<b>Label</b>	<b>Counts</b>
Synchronized samples	NC (0)	94
	MCI (1)	28
NC → Actigraphy → NC	NC (0)	4
MCI → Actigraphy → MCI	MCI (1)	1

**Πίνακας 1:** Σύνολο δεδομένων MCI

<i>CSF test positive - CSF test negative participants dataset</i>		
<b>Actigraphy Sample Category</b>	<b>Label</b>	<b>Counts</b>
Synchronized samples	Negative (0)	63
	Positive (1)	18
Positive test → Actigraphy	Positive (1)	4
Negative → Actigraphy → Negative	Negative (0)	1

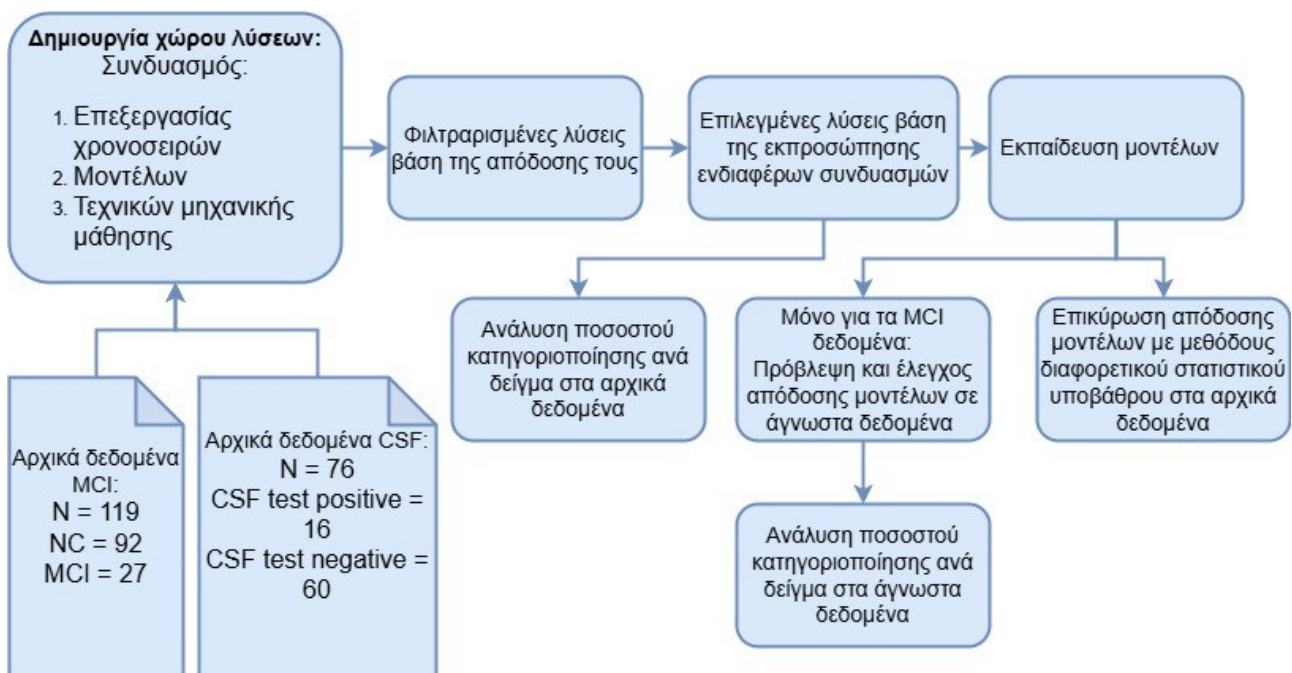
**Πίνακας 2:** *Σύνολο δεδομένων CSF*

### 3. ΜΕΘΟΔΟΙ

Σε αυτό το εδάφιο παρουσιάζεται αρχικά μια γενική εικόνα της ροής εργασιών (pipeline) για την επίτευξη ανάλυσης ταξινόμησης αναφοράς των MCI και CSF δεδομένων, ενώ στην συνέχεια ακολουθούν οι λεπτομέρειες σχετικά με κάθε μια από τις μεθόδους που χρησιμοποιήθηκαν.

#### 3.1 Επισκόπηση της διεργασίας ταξινόμησης αναφοράς

Για την καθιέρωση ταξινόμησης αναφοράς ανάμεσα σε συμμετέχοντες με ήπια νοητική εξασθένηση (Mild Cognitive Impairment - MCI) και μή (Normal Controls – NC), καθώς και σε συμμετέχοντες με θετικό CSF έλεγχο (CSF test positive) – ή αρνητικό (CSF test negative) του ερευνητικού προγράμματος ALBION παρουσιάζουμε μια ανάλυση που χρησιμοποιεί τα προ-επεξεργασμένα, δεδομένα χρονοσειρών ακτιγραφίας κάθε συμμετέχοντα χωρίς την εφαρμογή τεχνικών εξαγωγής χαρακτηριστικών, δηλαδή χρησιμοποιώντας μονάχα τις πληροφορίες τιμής της ακτιγραφίας κάθε συμμετέχοντα ανά χρονική στιγμή καθ'όλη την διάρκεια καταγραφής της. Οι δύο βασικοί μας στόχοι είναι η εύρεση συνδυασμών επεξεργασίας χρονοσειρών ακτιγραφίας και τεχνικών μηχανικής μάθησης που προσφέρουν προγνωστικές λύσεις με αξιόλογη απόδοση, καθώς και ο προσδιορισμός των χαρακτηριστικών δειγμάτων ακτιγραφίας τα οποία δυσκολεύουν την επιτυχή ταξινόμηση. Μια επισκόπηση της ροής που ακολουθήσαμε φαίνεται στην παρακάτω εικόνα.



**Εικόνα 3:** Επισκόπηση διεργασία ανάλυσης απόδοσης ταξινόμησης αναφοράς δεδομένων MCI,CSF

Για να δημιουργήσουμε και να αξιολογήσουμε τον χώρο των πιθανών λύσεων, πραγματοποιούμε συγκρίσεις ταξινομητών χρονοσειρών (time series classifiers) χρησιμοποιώντας τη μέθοδο επαναλαμβανόμενης ένθετης διασταύρωσης επικύρωσης (repeated nested Cross Validation – rncv) [4], σε συνδυασμό με διάφορες διαμορφώσεις δεδομένων (όπως μοναδικούς μετασχηματισμούς, και περιόδους εξομάλυνσης χρονοσειρών) μαζί με πληθώρα μεθόδων μηχανικής μάθησης (όπως στρατηγικών επιλογής μοντέλων και βελτιστοποίησης υπερπαραμέτρων).

Στην συνέχεια ακολουθεί φιλτράρισμα των λύσεων/μοντέλων που ευρέθησαν να πληρούν τα όρια απόδοσης που θέτουμε ως αποδεκτά για την επιλογή των τελικών μοντέλων που εκπροσωπούν συνδυασμούς επεξεργασίας δεδομένων και τεχνικών μηχανικής μάθησης που αναμένεται πως θα έχουν αξιολογη απόδοση σε νέα δεδομένα (inference).

Επιπλέον πραγματοποιούμε ανάλυση Ποσοστού Σωστής Ταξινόμησης (Correct Classification Rate – CCR) για να εντοπίσουμε ουσιώδεις πληροφορίες σχετικά με την απόδοση σε επίπεδο δείγματος αυτών των μοντέλων. Τέλος, εκπαιδεύουμε τα επιλεγμένα μοντέλα στο πλήρες σύνολο δεδομένων.

Για να επικυρώσουμε την ποιότητα των αποτελεσμάτων της mCV και να αποκτήσουμε καλύτερη εκτίμηση της απόδοσης των εκπαιδευμένων επιλεγμένων μοντέλων υπό διαφορετικές στατιστικές συνθήκες, διεξάγουμε ένα βήμα επικύρωσης, χρησιμοποιώντας 3 διαφορετικές μεθόδους αξιολόγησης, τα αποτελέσματα των οποίων συγκρίνουμε τόσο μεταξύ τους όσο και με τα αποτελέσματα της mCV.

Τέλος, αξιοποιούμε τα εκπαιδευμένα, επιλεγμένα μοντέλα για την πρόβλεψη κατηγοριών MCI-NC σε άγνωστες χρονοσειρές ακτιγραφίας, χρησιμοποιώντας ένα σύνολο δεδομένων που δεν είχε συμπεριληφθεί στην εκπαίδευση τους. Αυτή η διαδικασία μας επιτρέπει να αξιολογήσουμε την ικανότητα γενίκευσης των αποτελεσμάτων της μεθόδου mCV, καθώς και των τριών μεθόδων επικύρωσης που χρησιμοποιήθηκαν, και να πραγματοποιήσουμε μια δευτερεύουσα ανάλυση CCR στα επιλεγμένα μοντέλα βάση της απόδοσης τους στα άγνωστα δείγματα. Ο συνδυασμός αυτός επιτρέπει μια λεπτομερή αξιολόγηση τόσο της συνολικής απόδοσης όσο και της απόδοσης σε επίπεδο δείγματος.

Σημειώνεται ότι, για ορισμένα σύνολα δεδομένων, δεν περιλαμβάνονται όλα τα προαναφερθέντα βήματα στην ενότητα αποτελεσμάτων. Αντίθετα, επιλέξαμε να παρουσιάσουμε μόνο τις πιο χρήσιμες πληροφορίες.

### 3.2 Δεδομένα εκπαίδευσης και διαμόρφωση χώρου λύσεων

Το ίδιο σύνολο δεδομένων ακτιγραφίας χρησιμοποιείται για την εκπαίδευση και επικύρωση μοντέλων, συγκεκριμένα N=119 συνολικά δείγματα για την ταξινόμηση ανάμεσα σε MCI-NC (MCI=27, NC=92) καθώς και N=76 δείγματα για την ταξινόμηση σε CSF test positive, CSF test negative (CSF test positive=16, CSF test negative=70). Τα σύνολα δεδομένων διαφέρουν λίγο σε αριθμό από τα αντίστοιχα της ενότητας 2, καθώς κατά την προεπεξεργασία των χρονοσειρών ακτιγραφίας ορισμένες κρίθηκαν μη κατάλληλες για μηχανική μάθηση.

Οι τιμές για τις διάφορες παραμέτρους της επεξεργασίας χρονοσειρών, όπως και των μεθόδων μηχανικής μάθησης που διαμορφώνουν το σύνολο λύσεων παρατίθενται παρακάτω:

#### Διαμορφώσεις συνόλου δεδομένων:

1. Σύνολο δεδομένων: 6-ημέρες, 1-ημέρα, 1-ημέρα-χωρίς-αφύσικες-ημέρες (3 συνδυασμοί)
2. Εξομάλυνση: 5 λεπτά (5T), 15 λεπτά (15T), 30 λεπτά (30T), 45 λεπτά (45T), 1 ώρα (1H) (5 συνδυασμοί), χρησιμοποιώντας την μέθοδο “resample” της Python βιβλιοθήκης pandas [36].
3. Αποσύνθεση χρονοσειρών (STL decomposition): τάση και εποχικότητα, χωρίς αποσύνθεση, μόνο τάση (trend), μόνο εποχικότητα (seasonality), δηλαδή συνολικά

- 4 συνδυασμοί, χρησιμοποιώντας την μέθοδο “STL” από την Python βιβλιοθήκη statsmodels [37].
4. Κανονικοποίηση: χωρίς κανονικοποίηση, μέθοδος yeojohnson [38] (2 συνδυασμοί)
  5. Τομή χρόνου (ισχύει μόνο για τα σύνολα δεδομένων 1-ημέρα, 1-ημέρα-χωρίς-αφύσικες-ημέρες): παράθυρο 08:00:00-16:00:00, χωρίς τομή χρόνου (2 συνδυασμοί). Η τομή πραγματοποιείται με την μέθοδο “between\_time” της βιβλιοθήκης pandas.

#### Διαμορφώσεις Μηχανικής Μάθησης:

1. Αριθμός Δοκιμών βελτιστοποίησης υπερπαραμέτρων: (ισχύει μόνο για τους ταξινομητές TSVC, TSF): 0, 300 (2 συνδυασμοί), οι επιλογές υπερπαραμέτρων όταν οι δοκιμές είναι 300 φαίνονται στον Πίνακα 16, ενώ οι προεπιλεγμένες υπερπαραμέτροι όταν οι δοκιμές είναι 0 στον Πίνακα 17.
2. Επιλογή μοντέλου (ισχύει μόνο όταν ο αριθμός δοκιμών βελτιστοποίησης > 0): Best, Sem, Valid\_opt, Gap\_opt, Train\_opt (5 συνδυασμοί)
3. Εξισορρόπηση κατηγοριών, (μόνο για τους ταξινομητές TSVC, TSF, καθώς το 1NNTS δεν το υποστηρίζει): Βάρη κατηγοριών, εμπνευσμένη από το άρθρο [39].
4. Ταξινομητές χρονοσειρών: 1-Nearest Neighbor Time Series Classifier (1NNTS) [40], Time Series Support Vector Classifier (TSVC) [41], Time Series Forest (TSF) [42] (3 συνδυασμοί).

Ο παραπάνω χώρος παραμέτρων οδηγεί σε συνολικά 2.600 συνδυασμούς, το οποίο μεταφράζεται και σε 2.600 πειράματα mCV για τους στόχους MCI-NC και CSF test positive-CSF test negative. Το εκτενές πεδίο λύσεων μας επιτρέπει να αξιολογήσουμε το αντίκτυπο των στρατηγικών μηχανικής μάθησης, όπως η επιλογή υπερπαραμέτρων και να καταλάβουμε πώς παράγοντες όπως η εξομάλυνση, η αποσύνθεση και η κανονικοποίηση επηρεάζουν την ακρίβεια και την ικανότητα γενίκευσης των μοντέλων σε νέα δεδομένα. Για την επίτευξη αυτού του στόχου, χρησιμοποιούνται 10 γύροι από 5-εσωτερικούς και 5-εξωτερικούς βρόγχους της mCV.

### **3.3 Μετρικές απόδοσης, σύνολα μοντέλων και αποδεκτά όρια απόδοσης.**

Σε όλες τις περιπτώσεις, για τους σκοπούς αυτής της ανάλυσης, όταν αξιολογούμε την απόδοση της ταξινόμησης σε πολλαπλά σύνολα ελέγχου, χρησιμοποιούμε τη διάμεσο των μετρικών ευαισθησίας (recall), ειδικότητας (specificity), AUC και Mathews Correlation Coefficient (MCC). Η διάμεσος υπολογίζεται από την κατανομή των αντίστοιχων μετρικών που προκύπτουν από όλα τα 50 (10 επαναλήψεις x 5 εξωτερικοί βρόχοι) σύνολα ελέγχου (test sets) που χρησιμοποιούνται στην αξιολόγηση της ταξινόμησης με τη χρήση mCV.

Επιλέξαμε την διάμεσο τιμή καθώς επηρεάζεται λιγότερο από ακραίες τιμές απόδοσης (outliers) σε μεμονωμένες περιπτώσεις των γύρων mCV. Επομένως, όταν αναφέρεται κάποια μετρική στο πλαίσιο της συνολικής απόδοσης, εννοείται η τιμή της διάμεσου. Αυτή η προσέγγιση ισχύει και για τα 95% διαστήματα εμπιστοσύνης (Confidence Intervals – CI), τα οποία βασίζονται στη διάμεσο και υπολογίζονται ως το 2,5ο και 97,5ο εκατοστημόριο της κατανομής που προκύπτει από 10.000 Bootstrap δειγματοληψίες των τιμών των μετρικών απόδοσης, όπως προκύπτουν από τις 50 (10 επαναλήψεις x 5 εξωτερικοί βρόχοι) μετρήσεις απόδοσης κατηγοριοποίησης μέσω της mCV.

Για την απόδοση της ταξινόμησης MCI-NC θεωρούμε αποδεκτά τα μοντέλα που επιτυγχάνουν ευαισθησία  $\geq 0.5$  και ειδικότητα  $\geq 0.7$  στα mCV πειράματα. Τα μοντέλα που περνούν αυτά τα όρια χαρακτηρίζονται ως “φιλτραρισμένα”. Με πανομοιότυπο τρόπο χαρακτηρίζουμε τα μοντέλα που επιτυγχάνουν ανάκληση  $\geq 0.5$  και ειδικότητα  $\geq 0.667$  κατά την ταξινόμηση CSF test positive – CSF test negative.

Ακολουθώντας αυτό το βήμα, δημιουργούμε ομάδες βασισμένες στην απόδοση τους και εντοπίζουμε ενδιαφέροντες συνδυασμούς μοντέλων και διαμορφώσεων συνόλων δεδομένων, σχηματίζοντας το υποσύνολο των "επιλεγμένων" μοντέλων. Αυτά τα μοντέλα χρησιμοποιούνται στη συνέχεια για ανάλυση, επικύρωση, αλλά και εξαγωγή συμπερασμάτων μέσω της ανάλυσης CCR.

### 3.4 Ανάλυση Ποσοστού Σωστής Ταξινόμησης (Correct Classification Rate – CCR)

Για τα αποτελέσματα της  $mCV$ , όταν εκτελείται η ανάλυση του Ποσοστού Σωστής Ταξινόμησης (CCR), αυτή πραγματοποιείται ανά μοντέλο και ανά διαμόρφωση συνόλου δεδομένων και τεχνικών μηχανικής μάθησης, όπως περιγράφεται στην ενότητα 3.2. και προκύπτει ανά δείγμα για κάθε προαναφερθέν συνδυασμό από τον λόγο του αριθμού που το δείγμα κατηγοριοποιήθηκε σωστά κατά την  $mCV$  προς τον αριθμό επαναλήψεων της  $mCV$  (10), όπως φαίνεται στην παρακάτω εξίσωση.

$$CCR(\text{sample}) = \frac{\text{number of times correctly classified}}{\text{rounds of } rNCV}$$

Στη συνέχεια, οι μέσες τιμές CCR για κάθε δείγμα υπολογίζονται ως ο μέσος όρος όλων των επιλεγμένων μοντέλων:

$$\hat{CCR}(\text{sample}) = \sum_i^{\text{number of models}} \frac{CCR(\text{sample})_i}{\text{number of models}}$$

Για τα αποτελέσματα της ανάλυσης από την πρόβλεψη του συνόλου αγνώστων δεδομένων (holdout set), επειδή υπάρχει μία μόνο πρόβλεψη για κάθε δείγμα, η CCR ανά δείγμα υπολογίζεται μετρώντας πόσες φορές αυτό ταξινομείται σωστά από τα επιλεγμένα μοντέλα. Αυτός ο αριθμός στη συνέχεια διαιρείται με τον αριθμό των μοντέλων στο σύνολο για να λάβουμε την τελική τιμή της CCR για κάθε δείγμα του συνόλου.

Και στις δύο περιπτώσεις, ορίζουμε ως "δύσκολα δείγματα" εκείνα που έχουν  $CCR \leq 0.2$  και ως "εύκολα δείγματα" εκείνα με  $CCR \geq 0.8$ . Τα υπόλοιπα δείγματα, τα οποία βρίσκονται μεταξύ αυτών των ορίων, θεωρούμε ότι ανήκουν στην κατηγορία "μεσαίας δυσκολίας". Αυτά είναι περιπτώσεις δευτερεύοντος ενδιαφέροντος, όπου τα μοντέλα εμφανίζουν μέτρια εμπιστοσύνη στην ταξινόμηση και δεν λαμβάνονται υπόψη στην ανάλυση μας.

### 3.5 Επικύρωση επιλεγμένων μοντέλων

Για το βήμα της επικύρωσης χρησιμοποιούμε τις παρακάτω μεθόδους:

1. Τυχαία Δειγματοληψία Κατηγορίας με όμοιο διαχωρισμό κατηγοριών (Random Class Stratified Sampling - RCSS), όπου κάθε δειγματοληψία σε σύνολα δειγμάτων εκπαίδευσης και ελέγχου, με όμοιες κατανομές κατηγοριών είναι στατιστικά ανεξάρτητη.

2. Επαναλαμβανόμενη διασταυρωμένη επικύρωση (repeated Cross Validation – rCV), όπου κάθε επανάληψη είναι στατιστικά ανεξάρτητη και τα σύνολα δειγμάτων εκπαίδευσης και ελέγχου που προκύπτουν έχουν όμοια κατανομή κατηγοριών.
3. Αξιολόγηση με τη μέθοδο Bootstrapping. Σε αυτήν την περίπτωση, το σύνολο ελέγχου αποτελείται από τα δείγματα που δεν επιλέγονται στα bootstrapped δείγματα που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Αυτά τα δείγματα είναι γνωστά ως OOB (Out-Of-Bag). Αυτή η μέθοδος διαφέρει σημαντικά από τις δύο προηγούμενες, καθώς οι κατανομές των κατηγοριών στα σύνολα εκπαίδευσης και ελέγχου δεν είναι απαραίτητα όμοιες, γεγονός που μπορεί να επηρεάσει την αξιολόγηση της απόδοσης του μοντέλου.

Για κάθε μέθοδο δημιουργούνται  $B = 100$  διακριτά σύνολα δεδομένων. Σε κάθε σύνολο, το 20% των δειγμάτων λειτουργεί ως σύνολο ελέγχου, ενώ το υπόλοιπο 80% χρησιμοποιείται για την εκπαίδευση των μοντέλων. Εξαίρεση αποτελεί η μέθοδος OOB, όπου το σύνολο ελέγχου αντιστοιχεί περίπου στο 1/3 του συνολικού αριθμού των δειγμάτων.

Για να αξιολογήσουμε γρήγορα την ελάχιστη απόδοση ενός μοντέλου, από όλες τις μεθόδους επικύρωσης, προσδίδουμε ένα σκορ σε κάθε μέθοδο αξιολόγησης χρησιμοποιώντας τον ακόλουθο τύπο, ο ίδιος τύπος μπορεί επίσης να εφαρμοστεί για τον υπολογισμό του σκορ της μεθόδου mCV που χρησιμοποιήθηκε για την κατασκευή του χώρου λύσεων:

$$\text{model score} = \sum(AUC_{\text{lower CI}}, Recall_{\text{lower CI}}, Specificity_{\text{lower CI}}) - \sum(stdv AUC, stdv Recall, stdv Specificity)$$

Όλες οι μετρικές προκύπτουν από την κατανομή της διάμεσου για 10.000 Bootstrapped δείγματα των μετρικών που υπολογίζονται από τα σύνολα εκπαίδευσης-ελέγχου.

Κάθε μία από αυτές τις στρατηγικές αξιολόγησης παρέχει μια μοναδική οπτική σχετικά με την απόδοση του μοντέλου.

Είναι σημαντικό να σημειωθεί ότι τα μοντέλα που έχουν εκπαιδευτεί και βελτιστοποιηθεί ως προς τις υπερπαραμέτρους τους επαναεκπαιδούνται, για κάθε διαχωρισμό σε σύνολα εκπαίδευσης και ελέγχου, κατά τη διαδικασία της επικύρωσης, διατηρώντας όμως τις ίδιες υπερπαραμέτρους. Αυτή η προσέγγιση ενδέχεται να εισάγει έναν βαθμό διαρροής δεδομένων, τον οποίο θεωρούμε αποδεκτό συμβιβασμό για την επίτευξη των στόχων αυτής της ανάλυσης.

### 3.6 Πρόβλεψη άγνωστων δεδομένων

Σε σύγκριση με το σύνολο δεδομένων CSF, το οποίο έχει λιγότερα από 80 δείγματα, το σύνολο δεδομένων MCI, το οποίο έχει 119 δείγματα, παρέχει ένα μεγαλύτερο σύνολο δεδομένων, καταλληλότερο για την εκπαίδευση μοντέλων που προορίζονται για την πρόβλεψη άγνωστων δεδομένων. Έτσι, για τον σκοπό της ανάλυσης της ικανότητας γενίκευσης των αποτελεσμάτων της mCV και της διαδικασίας επικύρωσης των επιλεγμένων μοντέλων σε νέα άγνωστα MCI-NC δεδομένα, χρησιμοποιούμε τα επιπλέον δείγματα MCI που προστέθηκαν στο σύνολο δεδομένων ALBION μετά τον Ιανουάριο του 2024. Το σύνολο άγνωστων δεδομένων περιλαμβάνει  $N=32$  ( $NC=18$  και  $MCI=14$ ) δείγματα.

Για την πραγματοποίηση προβλέψεων τα εκπαιδευμένα μοντέλα φορτώνονται και χρησιμοποιούνται για την πρόβλεψη της κατηγορίας κάθε δείγματος (MCI-NC). Δεδομένου ότι οι πραγματικές κατηγορίες κάθε δείγματος είναι γνωστές, αυτές χρησιμοποιούνται για τον υπολογισμό των μετρικών απόδοσης ταξινόμησης, όπως και στις προηγούμενες μεθόδους

## 4. ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ MCI

Σε αυτό το εδάφιο παρουσιάζονται με ακόλουθο σχολιασμό τα αποτελέσματα της ανάλυσης ταξινόμησης αναφοράς του στόχου (endpoint) MCI-NC. Ξεκινάμε καθορίζοντας μια βάση απόδοσης (baseline) μέσω ενός απλού τυχαίου ταξινομητή, ο οποίος παρέχει τυχαίες προβλέψεις. Αυτή η τεχνική λειτουργεί ως σημείο εκκίνησης για τη σύγκριση των αποτελεσμάτων πιο σύνθετων μοντέλων. Στη συνέχεια, εφαρμόζουμε τις μεθόδους που περιγράφηκαν στο προηγούμενο εδάφιο, παρουσιάζοντας τα αποτελέσματα μας ξεχωριστά για τα αρχικά δεδομένα χρονοσειρών 6 ημερών και τα δεδομένα μέσης ακτιγραφίας μιας ημέρας.

### 4.1 Αποτελέσματα τυχαίου ταξινομητή

Όπως φαίνεται στον Πίνακα 3, ο τυχαίος ταξινομητής επιτυγχάνει συντελεστής συσχέτισης Matthews (Matthews's Correlation Coefficient - MCC)  $\cong 0$  και AUC  $\cong 0.5$ , που αντιστοιχούν σε τυχαία κατανομή προβλέψεων. Επιπλέον, η ειδικότητα (specificity) είναι σημαντικά υψηλότερη από την ευαισθησία (recall) λόγω του πολύ μεγαλύτερου αριθμού δειγμάτων NC στο σύνολο δεδομένων. Αυτά τα αποτελέσματα επιβεβαιώνουν ότι ο τυχαίος ταξινομητής συμπεριφέρεται όπως αναμένεται κατά τη διάρκεια των 50 επαναλήψεων διαχωρισμού εκπαίδευσης-ελέγχου της mCV, παρέχοντας μια χρήσιμη αρχική βάση σύγκρισης. Οποιοδήποτε μελλοντικό μοντέλο που επιτυγχάνει καλύτερα αποτελέσματα όσον αφορά το MCC και το AUC, καθώς και μια πιο ισορροπημένη σχέση ανάμεσα στην ανάκληση (recall) και την ειδικότητα θα αποδεικνύει την ικανότητα του να μαθαίνει ουσιαστικά μοτίβα στα δεδομένα και να παρέχει αξιόπιστες προβλέψεις.

Metric	Median	95% CI of the median
MCC	-0.011	(-0.242, 0.036)
Recall	0.167	(0.000, 0.200)
Specificity	0.824	(0.778, 0.833)
AUC	0.495	(0.389, 0.517)

Πίνακας 3: Αποτελέσματα τυχαίου ταξινομητή για το σύνολο δεδομένων MCI

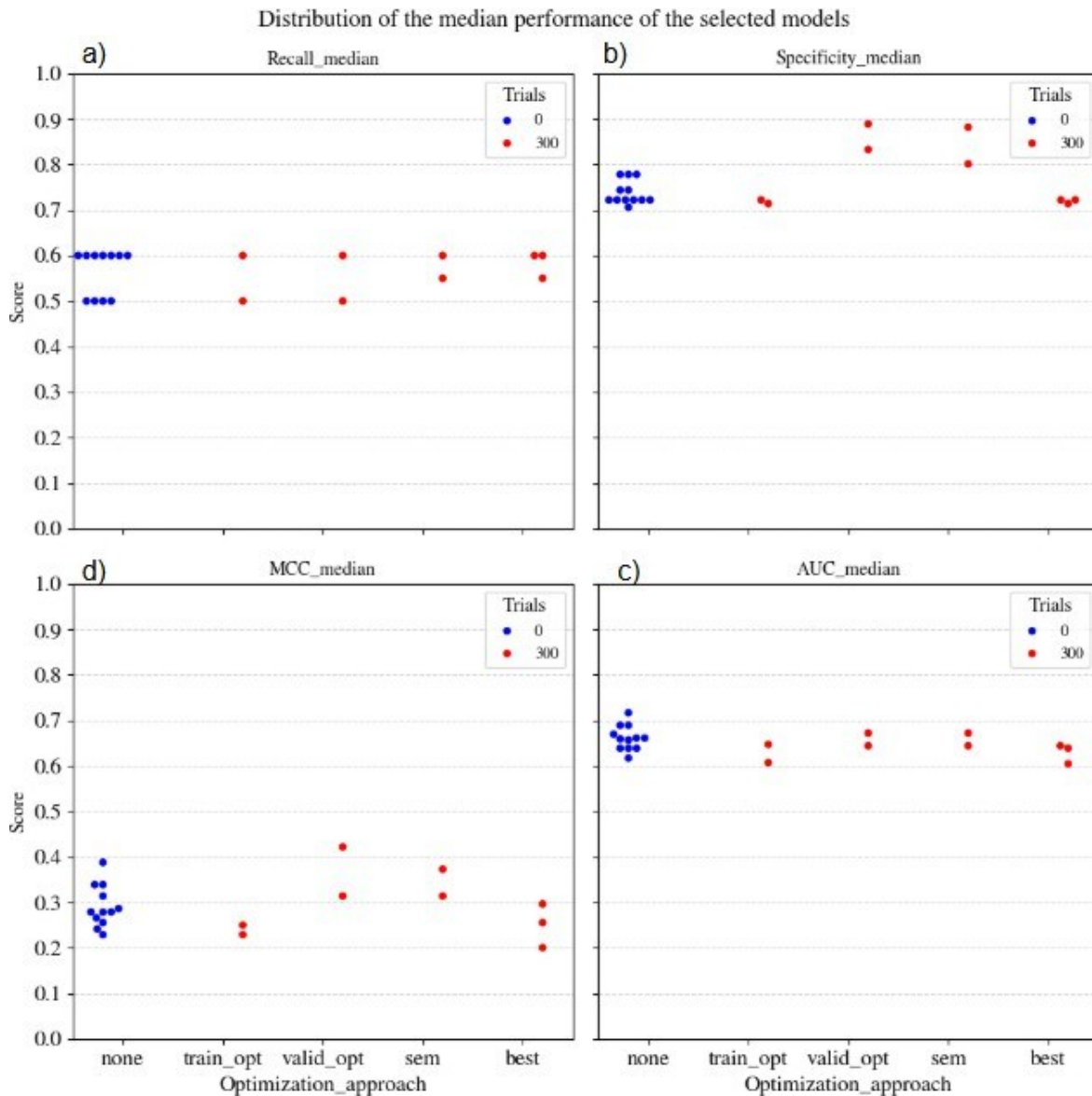
### 4.2 Αποτελέσματα για τα δεδομένα 6-ημερών

#### 4.2.1 Αποτελέσματα εξερεύνησης χώρου λύσεων

Ολοκληρώθηκαν συνολικά 432 από τα 520 mCV πειράματα που εκτελέστηκαν για τα δεδομένα των 6 ημερών ακτιγραφίας. Τα υπόλοιπα υπερέβησαν το χρόνο εκτέλεσης των 24 ωρών στο cloud και τερματίστηκαν από τον διαχειριστή εργασιών. Από τα ολοκληρωμένα πειράματα, μόνο 21 ( $\cong 5\%$ ) πληρούσαν τα ελάχιστα κριτήρια απόδοσης, δηλαδή διάμεσο ανάκλησης  $\geq 0.5$  και διάμεσο ειδικότητας  $\geq 0.7$ . Αυτά τα πειράματα αποτελούν το σύνολο που αποκαλούμε "φιλτραρισμένα". Αξιοσημείωτο είναι ότι σχεδόν όλα τα φιλτραρισμένα πειράματα χρησιμοποίησαν τον ταξινομητή TSCV (19), ενώ μόνο 2 χρησιμοποίησαν τον ταξινομητή 1NNTS. Τέλος, ο ταξινομητής TSF δεν χρησιμοποιήθηκε σε κανένα από τα φιλτραρισμένα πειράματα.

Τα δύο μοντέλα 1NNTS πέτυχαν διάμεσο ανάκλησης = 0.5, με ειδικότητα = 0.72 και 0.77 αντίστοιχα. Επιπλέον, είχαν διάμεσο MCC = 0.24 και 0.22, και διάμεσο AUC = 0.638 και 0.61, αντίστοιχα. Το εύρος της διάμεσης απόδοσης των 19 ταξινομητών TSCV συνοψίζεται στην Εικόνα 4, με ανάκληση = [0.5, 0.6], ειδικότητα = [0.7, 0.9], MCC = [0.2, 0.4] και AUC = [0.6, 0.73]. Είναι φανερό ότι η βελτιστοποίηση των παραμέτρων των μοντέλων μέσω της προσέγγισης επικύρωσης (valid-opt) και one-sem τείνει να αυξάνει την ειδικότητα. Ωστόσο,

όσον αφορά το AUC, μόνο τα μη βελτιστοποιημένα μοντέλα φτάνουν στην κορυφή του εύρους με τιμές  $> 0.7$ .

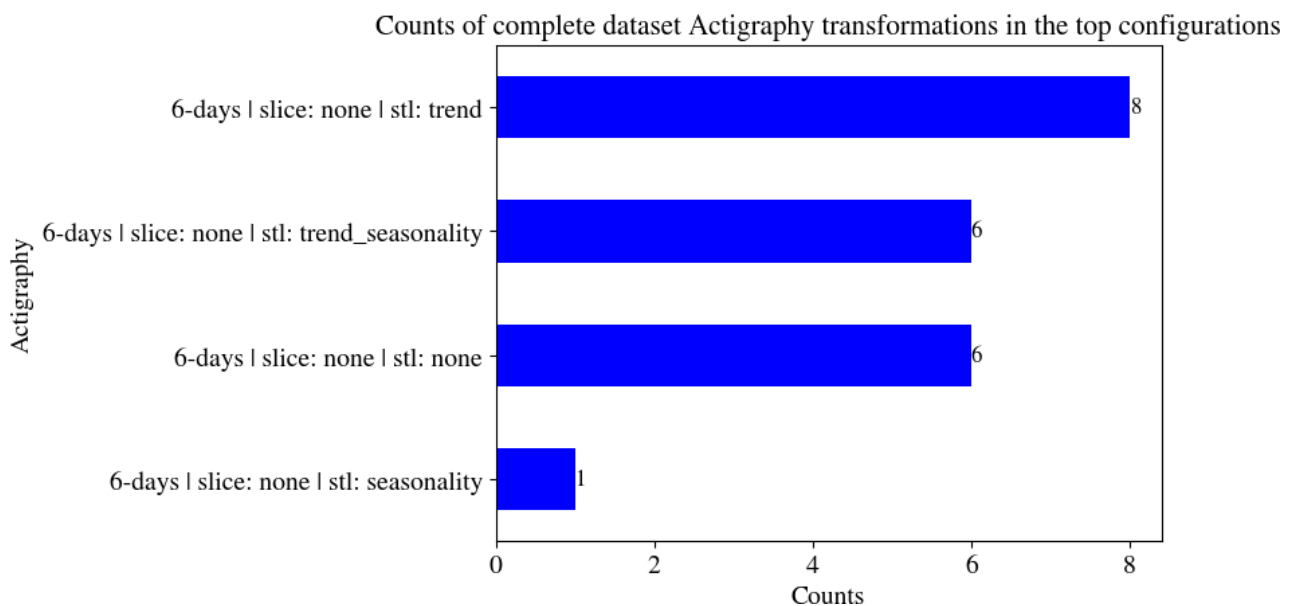


**Εικόνα 4:** Κατανομή των διάμεσων μετρικών απόδοσης σε 50 επαναλήψεις mCV (10 γύροι x 5 εξωτερικοί βρόγχοι) για μοντέλα χρονοσειρών 6 ημερών MCI-NC που πέτυχαν διάμεσο ανάκλησης  $\geq 0.5$  και ειδικότητας  $\geq 0.7$  (21 φιλτραρισμένα πειράματα), ομαδοποιημένα ανάλογα με την προσέγγιση βελτιστοποίησης υπερπαραμέτρων που χρησιμοποιήθηκε για την επιλογή του καλύτερου μοντέλου. Τα μοντέλα που χρησιμοποιούν βελτιστοποίηση υπερπαραμέτρων εμφανίζονται με κόκκινο χρώμα, ενώ τα μοντέλα με προεπιλεγμένες υπερπαραμέτρους εμφανίζονται με μπλε χρώμα. Τα διαγράμματα παρουσιάζουν: α) Ανάκληση (Recall), β) Ειδικότητα (Specificity), γ) AUC και δ) MCC.

Όσον αφορά τις πιο σημαντικές παραμέτρους ακτιγραφίας, οι μικρότερες περίοδοι εξομάλυνσης (5 και 30 λεπτά) ήταν οι πιο συχνά επιλεγμένες, οδηγώντας σε μεγαλύτερο αριθμό χαρακτηριστικών χρονοσειρών, με καθεμία να επιλέγεται 6 φορές, όπως φαίνεται στην Εικόνα 5. Περίοδοι εξομάλυνσης 15 λεπτών επιλέχθηκαν 4 φορές, ενώ μεγαλύτερες περίοδοι (45 λεπτά και 1 ώρα) επιλέχθηκαν μόνο 3 και 2 φορές αντίστοιχα. Αυτό υποδεικνύει ότι τα μοντέλα που αναπτύχθηκαν χρησιμοποιώντας το σύνολο δεδομένων χρονοσειρών 6 ημερών επωφελούνται από μια μέτρια έως λεπτομερή εξομάλυνση για την ανάλυσή τους. Ενδιαφέρον παρουσιάζει το γεγονός ότι κανένα από τα φιλτραρισμένα πειράματα δεν χρησιμοποίησε κανονικοποίηση.

Όσον αφορά επεξεργασία χρονοσειρών μέσω αποσύνθεσης, οι τάσεις (trend) είχαν τη μεγαλύτερη συχνότητα εμφάνισης (8), δείχνοντας την ικανότητά τους να καταγράφουν μοτίβα ακτιγραφίας. Οι συνδυασμοί trend και εποχικότητας (seasonality), καθώς και οι διαμορφώσεις χωρίς διάσπαση (stl: none), είχαν ίση εκπροσώπηση (6 φορές), υποδεικνύοντας ότι και οι δύο προσεγγίσεις αποτελούν βιώσιμες επιλογές. Οι διαμορφώσεις μόνο με εποχικότητα (seasonality) είχαν τη χαμηλότερη συχνότητα εμφάνισης (1).

Τέλος, όσον αφορά τα χαρακτηριστικά μηχανικής μάθησης των φιλτραρισμένων μοντέλων, τα πειράματα με και χωρίς βελτιστοποίηση παραμέτρων είχαν παρόμοια εκπροσώπηση, με 12 και 9 αντίστοιχα. Από τα 12 βελτιστοποιημένα μοντέλα, τα 3 βελτιστοποιήθηκαν με την προεπιλεγμένη προσέγγιση "best", ενώ 2 χρησιμοποιήθηκαν με τις προσεγγίσεις εκπαίδευσης, επικύρωσης και one-sem. Η μέθοδος βελτιστοποίησης "gap" δεν επιλέχθηκε σε κανένα από τα πειράματα.



**Εικόνα 5:** Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας MCI-NC για τα πειράματα 6 ημερών που πέτυχαν τουλάχιστον διάμεσο Recall  $\geq 0.5$  και διάμεσο Specificity  $\geq 0.7$  (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός | χρονική περίοδος | συστατικά αποσύνθεσης".

#### 4.2.2 Επιλεγμένα μοντέλα

Στον Πίνακα 4 φαίνεται η ταξινόμηση των μοντέλων από το σύνολο των φιλτραρισμένων πειραμάτων, με φθίνουσα σειρά διάμεσου ανάκλησης, έπειτα χαμηλότερου 95% CI του διάμεσου ανάκλησης και τέλος διάμεσου ειδικότητας, οδηγεί σε τρεις ομάδες συνδυασμών διαμόρφωσης επεξεργασίας δεδομένων-μοντέλων, όπως παρουσιάζεται παρακάτω (ο αριθμός των μοντέλων αναφέρεται σε παρένθεση):

1. Διάμεσος Ανάκλησης = 0.6, χαμηλό όριο 95% CI  $\geq 0.5$  (3)
2. Διάμεσος Ανάκλησης = 0.6, χαμηλό όριο 95% CI = 0.4 (2)
3. Διάμεσος Ανάκλησης  $\leq 0.55$ , χαμηλό όριο 95% CI = 0.4 (6)

UID	Model	Actigraphy	Num features	Trials	Optimization	Recall	Recall lower CI	Specificity	Specificity lower CI	MCC	MCC lower CI	AUC	AUC lower CI	mCV score
247	TSVC	6-days   slice: none   stl: none	144	0	none	0.6	0.6	0.722	0.714	0.2865	0.278	0.6695	0.657	1.916
224	TSVC	6-days   slice: none   stl: none	288	0	none	0.6	0.6	0.7435	0.714	0.388	0.233	0.7165	0.627	1.874
253	TSVC	6-days   slice: none   stl: trend_seasonality	1728	0	none	0.6	0.6	0.722	0.706	0.339	0.225	0.689	0.633	1.896
317	TSVC	6-days   slice: none   stl: trend_seasonality	288	300	valid_opt	0.6	0.4	0.833	0.778	0.314	0.2305	0.644	0.6	1.65
315	TSVC	6-days   slice: none   stl: trend_seasonality	288	300	sem	0.6	0.4	0.801	0.706	0.314	0.2305	0.644	0.6	1.572
310	TSVC	6-days   slice: none   stl: trend	288	300	sem	0.55	0.4	0.882	0.8055	0.373	0.233	0.672	0.63	1.7
312	TSVC	6-days   slice: none   stl: trend	288	300	valid_opt	0.5	0.4	0.889	0.833	0.422	0.279	0.672	0.637	1.743
242	TSVC	6-days   slice: none   stl: seasonality	1728	0	none	0.5	0.4	0.778	0.765	0.314	0.233	0.661	0.625	1.676
248	TSVC	6-days   slice: none   stl: none	1728	0	none	0.5	0.4	0.778	0.7435	0.266	0.229	0.6385	0.617	1.662
271	1NNTS	6-days   slice: none   stl: none	192	0	none	0.5	0.4	0.778	0.722	0.229	0.1385	0.617	0.575	1.596
287	1NNTS	6-days   slice: none   stl: trend_seasonality	1728	0	none	0.5	0.4	0.722	0.722	0.2415	0.11	0.6385	0.561	1.557

**Πίνακας 4:** Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις mCV για τα επιλεγμένα πειράματα 6 ημερών χρονοσειρών MCI-NC, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης (recall), χαμηλού ορίου 95% CI του διάμεσου ανάκλησης και χαμηλού ορίου 95% CI του διάμεσου ειδικότητας (specificity). Ο πίνακας χωρίζεται σε τρεις ζώνες, διαχωρισμένες με οριζόντιες γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος mCV.

Για κάθε ομάδα, πραγματοποιήθηκε δεύτερη ταξινόμηση βάσει του τύπου ακτιγραφίας, του αριθμού χαρακτηριστικών που χρησιμοποιήθηκαν, και τέλος του διάμεσου MCC, ώστε να εντοπιστούν αντιπροσωπευτικοί συνδυασμοί δεδομένων-μοντέλων με υψηλές τιμές MCC, που θα μπορούσαν να θεωρηθούν κατάλληλα μοντέλα για χρήση στο πεδίο. Το αποτέλεσμα είναι 11 μοντέλα που παρουσιάζονται στον Πίνακα 4, τα οποία περιλαμβάνουν βέλτιστες λύσεις ως προς την ευαισθησία για 4 διαμορφώσεις ακτιγραφίας, συνδυασμούς από 144 έως 1728 χαρακτηριστικά (πρόκειται για τα σημεία της χρονοσειράς), βελτιστοποιημένα και μη μοντέλα ως προς τις υπερπαραμέτρους τους, καθώς και 2 προσεγγίσεις βελτιστοποίησης που χρησιμοποιήθηκαν για την επιλογή των καλύτερων συνδυασμών υπερπαραμέτρων.

Τα 3 μοντέλα της πρώτης ομάδας χρησιμοποιούν χρονοσειρές μήκους 144, 288 και 1728 (που θα τις αποκαλούμε «χαρακτηριστικά»), με την αύξηση του αριθμού χαρακτηριστικών να μην επηρεάζει σημαντικά την απόδοση τους όσον αφορά την ανάκληση και την ειδικότητα. Το καλύτερο μοντέλο είναι το μοντέλο #224, το οποίο χρησιμοποιεί 288 χαρακτηριστικά χωρίς αποσύνθεση (decomposition), επιτυγχάνοντας το υψηλότερο MCC = 0.388 (χαμηλό όριο 95% CI MCC: 0.233) και AUC = 0.7165 (χαμηλό όριο 95% CI AUC: 0.627). Αξιοσημείωτο είναι ότι κανένα από τα μοντέλα της πρώτης ομάδας δεν είναι βελτιστοποιημένο ως προς τις παραμέτρους του, αλλά παρόλα αυτά επιδεικνύουν την υψηλότερη απόδοση κατηγοριοποίησης.

Τα 2 μοντέλα της δεύτερης ομάδας χρησιμοποιούν επίσης 288 χαρακτηριστικά με χρήση αποσύνθεσης τάσης-εποχικότητας και βελτιστοποίηση υπερπαραμέτρων. Επιτυγχάνουν την ίδια τιμή ανάκλησης με την πρώτη ομάδα, αλλά με μεγαλύτερη μεταβλητότητα ανάκλησης μεταξύ των επαναλήψεων mCV (χαμηλό recall 95% CI: 0.4). Αυτή η αυξημένη μεταβλητότητα αντισταθμίζεται από σημαντική βελτίωση στο διάμεσο της ειδικότητας ( $\approx 0.8$ ).

Στην τρίτη ομάδα με 6 μοντέλα, παρατηρείται ότι τα μοντέλα #310 και #312, τα οποία χρησιμοποιούν μόνο τη χρονοσειρά της τάσης (χωρίς εποχικότητα), είναι πανομοιότυπα στις υπόλοιπες διαμορφώσεις δεδομένων και μηχανικής μάθησης με τα υπόλοιπα μοντέλα της δεύτερης ομάδας. Η εξαίρεση του στοιχείου εποχικότητας μειώνει την ανάκληση κατά

0.05-0.1, αλλά επιτυγχάνει καλύτερο χαμηλό όριο 95% CI ειδικότητας ( $\cong 0.8$ ). Το μοντέλο #242 έχει παρόμοια διαμόρφωση με το μοντέλο #253 της πρώτης ομάδας, αλλά χρησιμοποιεί μόνο τη χρονοσειρά της εποχικότητας, επιτυγχάνοντας ελαφριά βελτίωση ειδικότητας ( $\cong 0.06$ ) με κόστος μείωσης 0.1 στην ανάκληση και 0.2 στο χαμηλότερο 95% CI της. Τέλος, τα μοντέλα 1NNTS είναι τα μόνα με AUC < 0.6, αναδεικνύοντας το όφελος της εξισορρόπησης κατηγοριών που υποστηρίζει ο ταξινομητής TSVC.

Συνολικά, η βελτιστοποίηση υπερπαραμέτρων οδήγησε σε καλύτερη ειδικότητα (0.80-0.89) σε σύγκριση με τα μη βελτιστοποιημένα μοντέλα (0.72-0.79), θυσιάζοντας όμως την απόδοση σε ανάκληση κάτι που υποδηλώνει πιθανή υπερπροσαρμογή των μοντέλων στην κατηγορία NC λόγω του μικρού μεγέθους της κατηγορίας MCI. Όσον αφορά το μήκος των χρονοσειρών, τα 288 χαρακτηριστικά φαίνεται να είναι ο βέλτιστος αριθμός, ενώ η αύξηση στα 1728 δεν βελτιώνει την απόδοση. Τα μοντέλα που εκπαιδεύτηκαν στο άθροισμα των στοιχείων τάσης και εποχικότητας προσφέρουν την καλύτερη ισορροπία μεταξύ ανάκλησης και ειδικότητας, επιτυγχάνοντας υψηλότερη συνολική απόδοση σε σύγκριση με τα μοντέλα που βασίζονται μόνο σε ένα μόνο στοιχείο, τα οποία εμφανίζουν σημαντικά μειωμένη ανάκληση για οριακές βελτιώσεις στην ειδικότητα.

Συνολικά, όλα τα επιλεγμένα μοντέλα TSVC επιτυγχάνουν ένα στενό εύρος χαμηλού ορίου 95% CI AUC = [0.6, 0.657], παρέχοντας μια αποδεκτή κατηγοριοποίηση αναφοράς (baseline classification) δεδομένης της δυσκολίας του προβλήματος ταξινόμησης και την μη χρήση εξαγωγής χαρακτηριστικών (feature engineering).

#### 4.2.3 Επικύρωση επιλεγμένων μοντέλων

Ο Πίνακας 5 παρουσιάζει τα αποτελέσματα της αξιολόγησης επιλεγμένων μοντέλων χρησιμοποιώντας τρεις διαφορετικές μεθόδους. Για κάθε μέθοδο χρησιμοποιήθηκαν B=100 διαχωρισμοί εκπαίδευσης-ελέγχου.

UID	rnCV Recall	rnCV Specificity	rCV Recall	rCV Specificity	RCSS Recall	RCSS Specificity	OOB Recall	OOB Specificity
317	0.6	0.833	0.6	0.833	0.5	0.833	0.417	0.886
315	0.6	0.801	0.6	0.833	0.5	0.833	0.417	0.886
224	0.6	0.7435	0.6	0.778	0.5	0.722	0.333	0.8615
247	0.6	0.722	0.6	0.722	0.5835	0.722	0.4	0.836
<b>253</b>	<b>0.6</b>	<b>0.722</b>	<b>0.6</b>	<b>0.722</b>	<b>0.667</b>	<b>0.722</b>	<b>0.5</b>	<b>0.833</b>
310	0.55	0.882	0.5	0.889	0.5	0.889	0.423	0.8875
242	0.5	0.778	0.5	0.778	0.5	0.778	0.222	0.943
248	0.5	0.778	0.5	0.778	0.5	0.778	0.25	0.926
287	0.5	0.722	0.5	0.765	0.5	0.722	0.455	0.7575

**Πίνακας 5:** Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 6 ημερών MCI-NC. Τα αποτελέσματα της απόδοσης rnCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

Από τα αποτελέσματα φαίνεται ότι 9 από τα 11 επιλεγμένα μοντέλα διατήρησαν σταθερή απόδοση πάνω από τα προκαθορισμένα όρια όταν εφαρμόστηκαν οι μέθοδοι rCV και RCSS.

Αυτά τα αποτελέσματα επιβεβαιώνουν ότι η μέθοδος επιλογής μοντέλων με βάση το rnCV εντοπίζει επιτυχώς υποψήφια μοντέλα με σταθερή βασική απόδοση σε διάφορους διαχωρισμούς δεδομένων με ισορροπία κατηγοριών στο σύνολο δεδομένων 6 ημερών. Ωστόσο, η αξιολόγηση με τη μέθοδο OOB αποκάλυψε πιθανή υπερπροσαρμογή στην NC κατηγορία, καθώς μόνο 1 από τα 11 μοντέλα (#253, με έντονη γραφή στον Πίνακα 5)

ικανοποίησε το προκαθορισμένο όριο απόδοσης Recall όταν δοκιμάστηκε σε δείγματα με διαφορετική ισορροπία κατηγοριών μεταξύ των συνόλων εκπαίδευσης και ελέγχου.

Ο υψηλός αριθμός χαρακτηριστικών στα μοντέλα 6 ημερών φαίνεται να προκαλεί υπερπροσαρμογή στα μοτίβα της κατηγορίας NC στα σύνολα εκπαίδευσης, κάτι που αποδεικνύεται από τη συνολικά χαμηλότερη διάμεσο ανάκλησης και την υψηλότερη ειδικότητα σε σύγκριση με τις άλλες δύο μεθόδους.

#### 4.2.4 Αξιολόγηση ικανότητας πρόβλεψης σε άγνωστα δεδομένα

Στον Πίνακα 6, παρουσιάζουμε την απόδοση των επιλεγμένων μοντέλων που είτε πληρούν είτε προσεγγίζουν τα όρια απόδοσης κατηγοριοποίησης στο σύνολο αγνώστων δεδομένων ελέγχου.

UID	Model	Actigraphy	Num features	Trials	Optimization	MCC	Recall	Specificity	F1	Accuracy	Balanced accuracy	PRAUC	AUC	mCV score
247	TSVC	6-days   slice: none   stl: none	144	0	none	0.228	0.5	0.722	0.611	0.625	0.611	0.51	0.611	1.916
224	TSVC	6-days   slice: none   stl: none	288	0	none	0.168	0.5	0.667	0.584	0.594	0.583	0.488	0.583	1.874
248	TSVC	6-days   slice: none   stl: none	1728	0	none	0.168	0.5	0.667	0.584	0.594	0.583	0.488	0.583	1.662
310	TSVC	6-days   slice: none   stl: trend	288	300	sem	0.221	0.429	0.778	0.6	0.625	0.603	0.507	0.603	1.7

**Πίνακας 6:** Απόδοση των επιλεγμένων μοντέλων 6-ημέρων που αξιολογήθηκαν σε ένα μικρό σύνολο ελέγχου αποτελούμενο από μόλις 32 δείγματα (MCI = 14, NC = 18). Τα μοντέλα είναι ταξινομημένα κατά φθίνουσα σειρά ανάκλησης και ειδικότητας. Εμφανίζονται μόνο τα μοντέλα που πέτυχαν ανάκληση  $\geq 0.5$  και ειδικότητα  $\geq 0.7$ . Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε μοντέλου. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

Συγκεκριμένα, τα μοντέλα που επιτυγχάνουν ανάκληση  $\geq 0.5$  με την υψηλότερη ειδικότητα είναι τα #247, #224, και #248, χρησιμοποιώντας 144, 288 και 1728 χαρακτηριστικά (χρονικά σημεία), αντίστοιχα, χωρίς αποσύνθεση (STL decomposition) των αρχικών δεδομένων ακτιγραφίας 6-ημερών. Το πιο φειδωλό μοντέλο, #247, με τον μικρότερο αριθμό χαρακτηριστικών (χρονικά διαστήματα εξομάλυνσης ανά μια ώρα), επιτυγχάνει τη μεγαλύτερη ειδικότητα (0.722) και είναι το μοναδικό που υπερβαίνει τα προκαθορισμένα όρια.

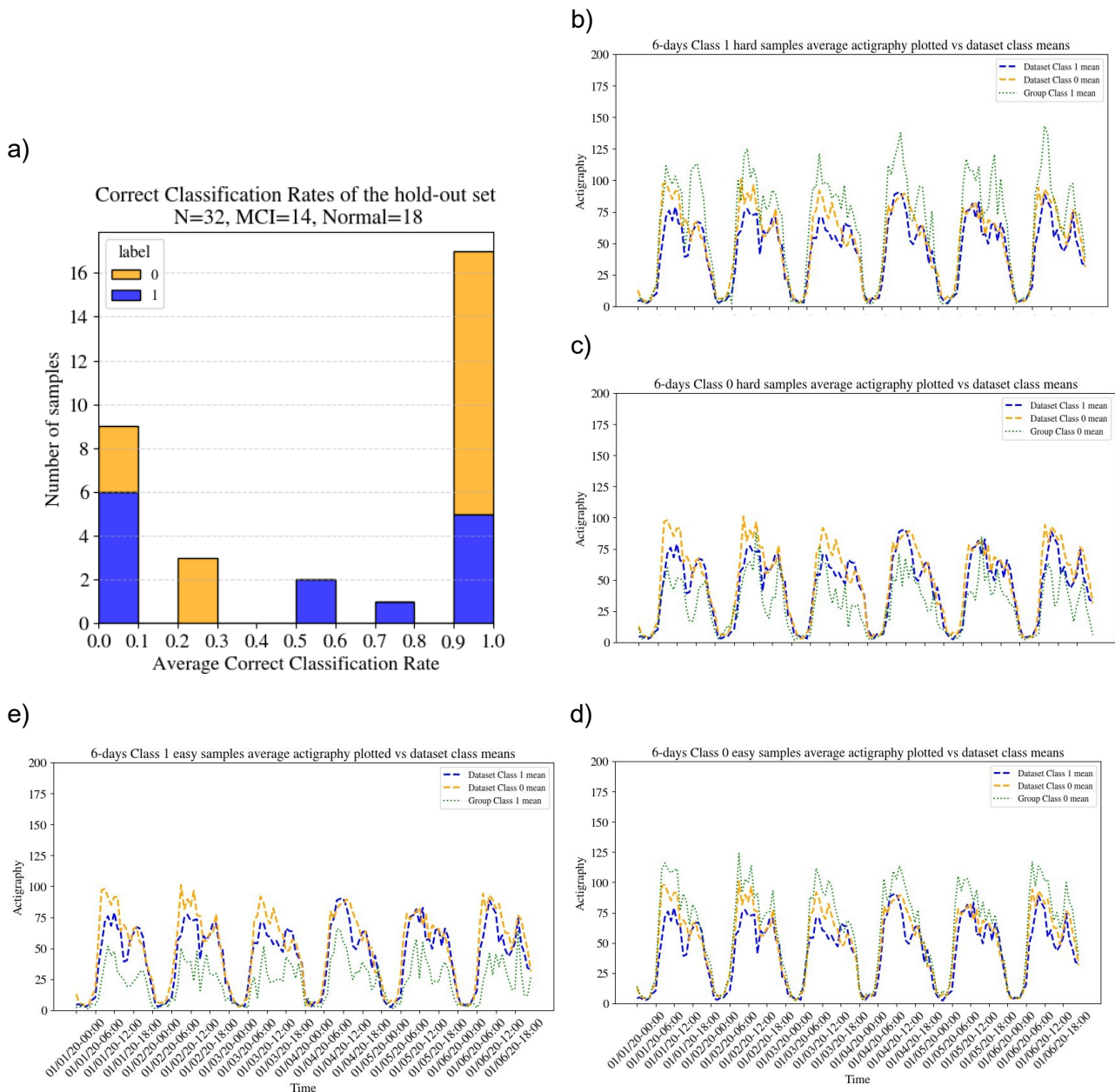
Το μοντέλο #310 είναι επίσης ενδιαφέρον, καθώς πλησιάζει το όριο της ανάκλησης με υψηλή ειδικότητα = 0.778, χρησιμοποιώντας αποκλειστικά το στοιχείο τάσης της αποσύνθεσης χρονοσειρών STL και μόνο 288 χαρακτηριστικά (χρονικά σημεία). Αντίθετα, τα μοντέλα 1NNTS, τα οποία δεν χρησιμοποιούν βάρη κατηγορίας (class weights) κατά την εκπαίδευση, αποτυγχάνουν στην πρόβλεψη της κατηγορίας MCI στο σύνολο ελέγχου.

Τα μοντέλα που εκπαιδεύτηκαν σε δεδομένα ακτιγραφίας χωρίς αποσύνθεση (decomposition) παρουσιάζουν πολύ καλύτερη απόδοση στο σύνολο ελέγχου, υποδεικνύοντας ότι τα χαρακτηριστικά της αποσύνθεσης ενδέχεται να μην συμβάλλουν στην εκμάθηση πιο γενικευμένων μοτίβων. Συνολικά, τα επιλεγμένα μοντέλα του συνόλου δεδομένων 6 ημερών παρουσιάζουν μέτρια απόδοση κατηγοριοποίησης, με 4 μοντέλα να φτάνουν  $AUC \geq 0.6$ . Το μοντέλο #247, που βρίσκεται στην κορυφή του εύρους με  $AUC = 0.611$ , θεωρείται αποδεκτό, λαμβάνοντας υπόψη τη δυσκολία της ταξινόμησης και το μικρό μέγεθος του συνόλου ελέγχου.

Τα δύο κορυφαία μοντέλα (#247 και #224), που κατατάχθηκαν υψηλότερα βάση της απόδοσης τους σε ανάκληση και ειδικότητα κατά την mCV, συνεχίζουν να διατηρούν την κορυφαία τους θέση στο σύνολο ελέγχου, όπως φαίνεται από τη σύγκριση των αποτελεσμάτων μεταξύ του συνόλου ελέγχου και του mCV στον Πίνακα 4. Οι τιμές mCV score για τα μοντέλα αυτά (1.916 και 1.874, αντίστοιχα) είναι οι υψηλότερες.

Επιπλέον, η βελτιστοποίηση υπερπαραμέτρων για την τροποποίηση του περιθωρίου του TSVC (υπερπάρаметρος C) φαίνεται να έχει μικρή επίδραση στην απόδοση, τόσο στο σύνολο δεδομένων εκπαίδευσης όσο και στο σύνολο ελέγχου. Αυτά τα αποτελέσματα ενισχύουν την εμπιστοσύνη μας στην απόδοση και τη γενίκευση της ικανότητας κατηγοριοποίησης των μοντέλων #247 και #224 σε άγνωστα δεδομένα.

Η ανάλυση της μέσης CCR με χρήση του συνόλου ελέγχου άγνωστον δεδομένων και των 4 μοντέλων του Πίνακα 4, όπως παρουσιάζεται στην Εικόνα 6, αποκαλύπτει ότι 9 από τα 32 δείγματα του συνόλου ελέγχου χαρακτηρίζονται ως "δύσκολα δείγματα" με  $CCR \leq 0.2$ . Συγκεκριμένα, πρόκειται για 6 δείγματα MCI και 3 δείγματα NC. Αντίστοιχα, 17 δείγματα χαρακτηρίζονται ως "εύκολα δείγματα", εκ των οποίων 5 είναι MCI και 12 είναι NC.



**Εικόνα 6:** Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Σύνολο Ελέγχου (N=32, MCI=14, NC=18) και τα μοντέλα του Πίνακα 6. a) Ιστόγραμμα που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. b), c) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI (κατηγορία 1, μπλε γραμμή) και NC (κατηγορία 0, κίτρινη γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 6 MCI και 3 NC "δύσκολων δειγμάτων" με  $CCR \leq 0.2$ , αντίστοιχα. d), e) Μέσες τιμές

ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 6 MCI και 12 NC "εύκολων δειγμάτων" (πράσινη γραμμή) με CCR  $\geq 0.8$ .

Μια πιο λεπτομερής εξέταση αυτών των δειγμάτων δείχνει ότι συγκεκριμένα χαρακτηριστικά χρονοσειρών επηρεάζουν σημαντικά τις λανθασμένες προβλέψεις των μοντέλων. Όλα τα δείγματα NC που είναι δύσκολο να ταξινομηθούν σωστά εμφανίζουν χαμηλή μέση δραστηριότητα, με τον μέσο όρο της ακτιγραφίας τους (πράσινη γραμμή) να μοιάζει με τον μέσο όρο της κατηγορίας MCI (μπλε γραμμή). Αντίστοιχα, τα δύσκολα δείγματα MCI παρουσιάζουν υψηλή μέση δραστηριότητα (πράσινη γραμμή), η οποία υπερβαίνει τον μέσο όρο της κατηγορίας NC (μπλε γραμμή). Τα ακριβώς αντίθετα χαρακτηριστικά παρατηρούνται στα "εύκολα δείγματα".

Συνοψίζοντας, η ανάλυση CCR στο σύνολο ελέγχου υποδηλώνει ότι το σύνολο δεδομένων δεν διαθέτει επαρκώς αντιπροσωπευτικά χαρακτηριστικά για τις δύο κατηγορίες χρονοσειρών. Αυτό είχε ως αποτέλεσμα τα μοντέλα αναφοράς μηχανικής μάθησης που χρησιμοποιούν ταξινομητές χρονοσειρών να μην μπορούν να μάθουν αποτελεσματικά ουσιαστικά μοτίβα σχετικά με την διάκριση των ακτιγραφιών (π.χ. σχετικά με το Acrophase κτλ), κάτι που συνηγορεί στην ανάγκη εξαγωγής περαιτέρω χαρακτηριστικών (feature engineering) για βελτίωση της απόδοσης.

#### 4.2.5 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 6-ημερών

Ο ταξινομητής TSVC ξεπέρασε σε απόδοση τους άλλους δύο ταξινομητές που δοκιμάσαμε, επιτυγχάνοντας AUC  $\cong [0.6, 0.7]$  κατά την διαδικασία rCV. Επιπλέον, πληρούσε ή ξεπερνούσε σταθερά τα αποδεκτά κατώφλια ανάκλησης και ειδικότητας, τόσο στην επικύρωση μέσω rCV όσο και RCSS για όλα τα επιλεγμένα μοντέλα, εκτός από δύο. Από τα 11 μοντέλα που αξιολογήθηκαν σε άγνωστα δεδομένα, τα 4 ήταν πολύ κοντά στα κατώφλι απόδοσης, γεγονός που υποδηλώνει ότι η εκπαίδευσή τους στο πλήρες σύνολο δεδομένων τους επιτρέπει ικανότητα γενίκευσης σε άγνωστα δεδομένα, με απόδοση (AUC  $\cong 0.6$ ) σε ένα δύσκολο πρόβλημα ταξινόμησης.

Τα δύο κορυφαία μοντέλα (#224, #237) παρουσιάζουν ιδιαίτερο ενδιαφέρον ως βασικά μοντέλα αναφοράς, καθώς επιτρέπουν μια γρήγορη εκτίμηση της απόδοσης χρησιμοποιώντας μια σημαντικά εξομαλυμένη εκδοχή της αρχικής 6-ήμερης χρονοσειράς.

Όσον αφορά την επίδραση των μεθόδων προεπεξεργασίας του συνόλου δεδομένων και των διαμορφώσεων μηχανικής μάθησης στην απόδοση των ταξινομητών TSVC, η αναμενόμενη απόδοση όλων των επιλεγμένων μοντέλων εμφανίζει στενή διακύμανση (περίπου  $\pm 0.1$ ) σε ανάκληση, ειδικότητα και AUC μεταξύ των διαφορετικών διαμορφώσεων του συνόλου δεδομένων. Αυτή η μικρή μεταβλητότητα υποδηλώνει ότι οι διαφορετικοί τρόποι προεπεξεργασίας προσφέρουν μόνο μικρές βελτιώσεις, κάτι που δείχνει να υπάρχει όριο στο κατώφλι απόδοσης που μπορεί να επιτευχθεί όταν χρησιμοποιούνται απευθείας τα δεδομένα χρονοσειρών, χωρίς εξαγωγή χαρακτηριστικών ή μεθόδους δημιουργίας συνθετικών δειγμάτων για την εξισορρόπηση του αριθμού δειγμάτων στις κατηγορίες NC, MCI.

Αυτό το «ταβάνι» στην απόδοσης αποδίδεται στην μη εξαγωγή χαρακτηριστικών σχετικών με το πρόβλημα γεγονός που περιορίζει την ικανότητα των μοντέλων να μάθουν αποτελεσματικά μοτίβα που χαρακτηρίζουν με ακρίβεια αυτές τις δύσκολες προς ταξινόμηση περιπτώσεις. Σε κάθε όμως περίπτωση η χρήση ταξινομητών χρονοσειρών παρέχει ένα κάτω όριο αναμενόμενης απόδοσης (baseline) και άρα είναι πολύ χρήσιμοι ως αρχική προσέγγιση ταξινόμησης.

Στην επόμενη ενότητα, επιχειρούμε μια εναλλακτική ανάλυση χρησιμοποιώντας την συμπιεσμένη εκδοχή των 6-ήμερων χρονοσειρών, δημιουργώντας μια μονοήμερη

χρονοσειρά μέσω των τιμών ακτιγραφίας για κάθε χρονικό σημείο της ημέρας. Αυτό εφαρμόζεται αρχικά χωρίς την αφαίρεση αφύσικων ημερών με ακραίες τιμές ακτιγραφίας (σύνολο δεδομένων 1-ήμερας) και στη συνέχεια κατόπιν αφαίρεσης των αφύσικων ημερών (σύνολο δεδομένων 1-ήμερας-χωρίς-αφύσικες-ημέρες).

Η ανάλυση στοχεύει να αξιολογήσει εάν η χρήση ενός συνόλου δεδομένων χρονοσειρών ακτιγραφίας με πολύ λιγότερα δείγματα επιτρέπει στους ταξινομητές χρονοσειρών να μάθουν λιγότερο πολύπλοκα και συνεπώς πιο γενικεύσιμα μοτίβα χωρίς πάλι την εξαγωγή περεταίρω χαρακτηριστικών.

### 4.3 Αποτελέσματα για τα δεδομένα 1-ήμερας

#### 4.3.1 Αποτελέσματα εξερεύνησης χώρου λύσεων

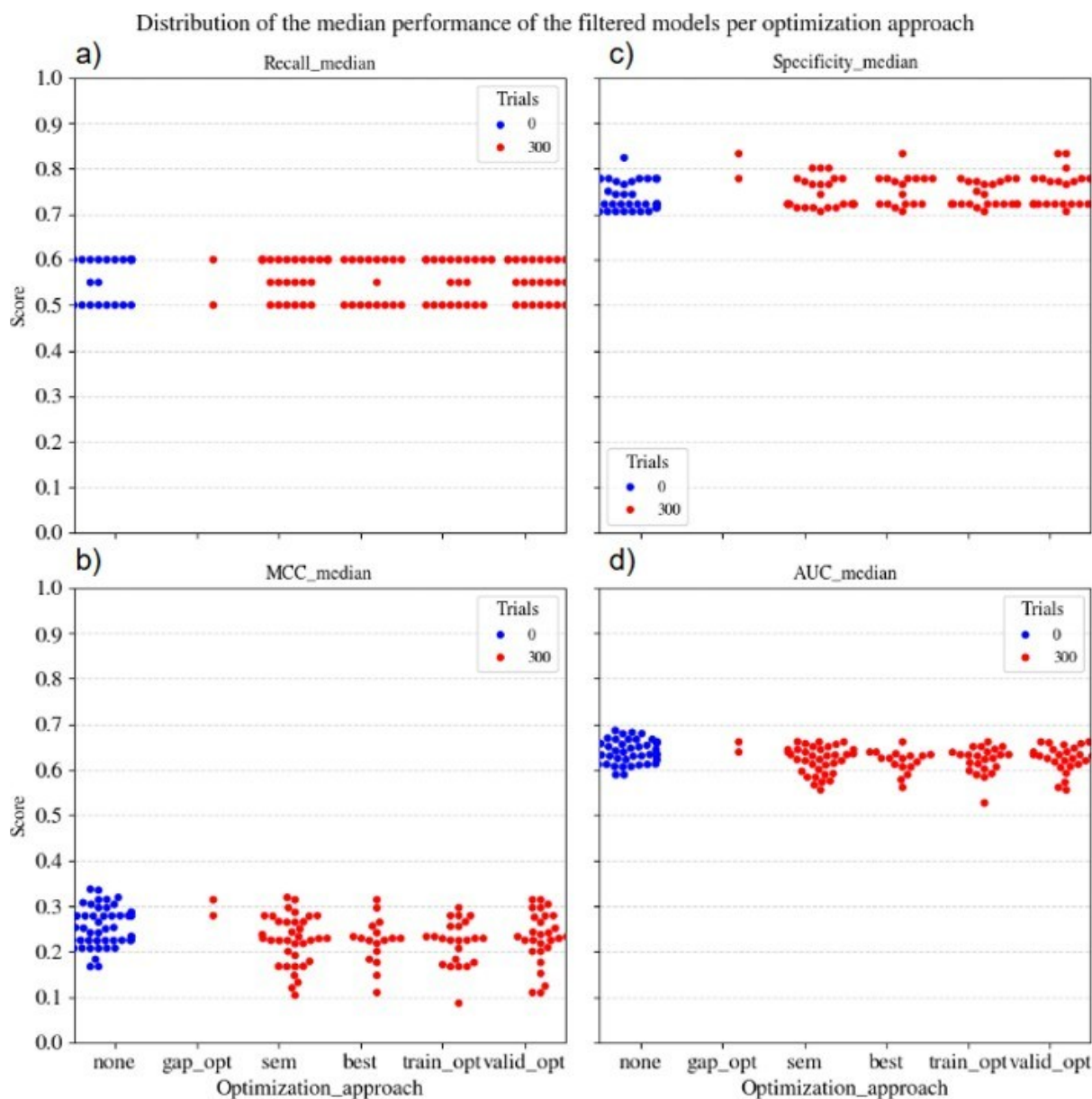
Το σύνολο δεδομένων 1-ήμερας περιλάμβανε τετραπλάσιο αριθμό πειραμάτων σε σχέση με την ανάλυση των 6-ήμερων χρονοσειρών, λόγω των επιπλέον επιλογών για την αφαίρεση αφύσικων ημερών και των μετασχηματισμών με την χρήση χρονικών παραθύρων της ημέρας. Από τα 2.080 πειράματα mCV που εκτελέστηκαν, 2.069 ολοκληρώθηκαν με επιτυχία. Χάρη στο μειωμένο αριθμό χαρακτηριστικών ανά χρονοσειρά (λιγότερα χρονικά σημεία), ελάχιστα πειράματα ξεπέρασαν το όριο εκτέλεσης των 24 ωρών που επέβαλε ο cloud server που χρησιμοποιήθηκε.

Από τα ολοκληρωμένα πειράματα, μόνο 167 ( $\cong 8\%$ ) ικανοποίησαν τις ελάχιστες απαιτήσεις απόδοσης με διάμεση ανάκληση  $\geq 0.5$  και διάμεση ειδικότητα  $\geq 0.7$ , ποσοστό που παραταύτα είναι κατά 3% υψηλότερο σε σύγκριση με την ανάλυση του συνόλου δεδομένων 6-ήμερων. Όπως προηγουμένως, θα αναφερόμαστε σε αυτά τα πειράματα ως το "φιλτραρισμένο" σύνολο επιτυχημένων πειραμάτων.

Όπως και στην ανάλυση των 6-ήμερων χρονοσειρών, ο ταξινομητής TSVC ήταν η κυρίαρχη επιλογή, καθώς εμφανίζεται σε 166/167 πειράματα του φιλτραρισμένου συνόλου, με μόνο ένα μοντέλο να ανήκει στον ταξινομητή 1NNTS.

Η Εικόνα 7 συνοψίζει την απόδοση των 166 μοντέλων TSVC με την εφαρμογή mCV. Το εύρος των διάμεσων μετρικών είναι ανάκληση = [0.5, 0.6], με μικρή διακύμανση σε όλες τις προσεγγίσεις επιλογής μοντέλων (εκτός της μεθόδου gap opt, η οποία είχε μόλις 2 περιπτώσεις), και ειδικότητα = [0.71, 0.83]. Η μετρική MCC παρουσιάζει τη μεγαλύτερη μεταβλητότητα [0.09, 0.34], με τις χαμηλότερες τιμές να επηρεάζονται ιδιαίτερα από τη στρατηγική βελτιστοποίησης των υπερπαραμέτρων. Επίσης, είναι αξιοσημείωτο ότι ορισμένα μοντέλα με βελτιστοποίηση υπερπαραμέτρων είχαν χαμηλότερες διάμεσες τιμές MCC σε σχέση με τα μη-βελτιστοποιημένα μοντέλα. Τέλος, η διάμεση AUC έχει εύρος [0.53, 0.69], με τα μη-βελτιστοποιημένα μοντέλα να επιτυγχάνουν υψηλότερο κατώ όριο απόδοσης.

Βάσει αυτών των αποτελεσμάτων, μπορούμε να συμπεράνουμε ότι, παρά τις 300 δοκιμές βελτιστοποίησης υπερπαραμέτρων, δεν παρατηρείται σημαντική βελτίωση σε καμία μετρική απόδοσης (Ανάκληση, Ειδικότητα, MCC ή AUC) σε σύγκριση με τη βασική μη-βελτιστοποιημένη εκδοχή. Αυτό υποδηλώνει ότι οι περιορισμοί απόδοσης του μοντέλου είναι εγγενείς στο ίδιο το πρόβλημα ταξινόμησης, παρά αποτέλεσμα μη-ιδανικών υπερπαραμέτρων. Επιπλέον, δείχνει ότι τα φιλτραρισμένα μοντέλα 1-ήμερας είναι ήδη αρκετά «φειδωλά» (parsimonious) σε αριθμό χαρακτηριστικών, ώστε η περαιτέρω βελτιστοποίησή τους να μην επιφέρει ουσιαστική βελτίωση απόδοσης.



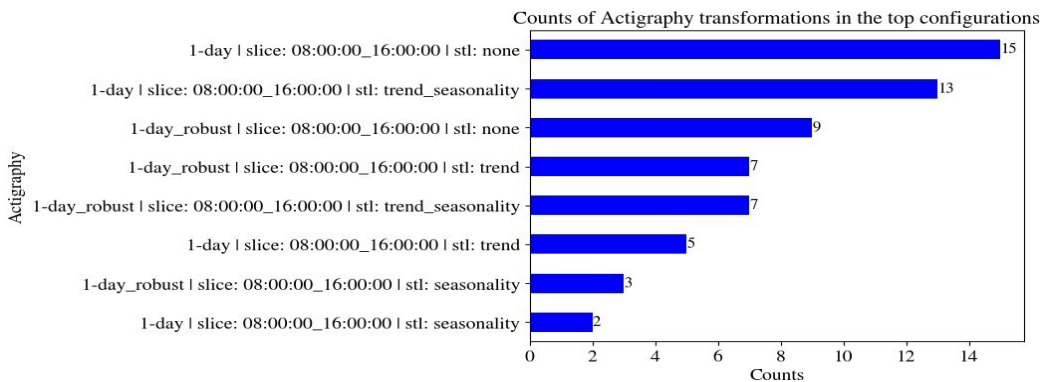
**Εικόνα 7:** Κατανομή των διάμεσων μετρικών απόδοσης σε 50 επαναλήψεις  $nCV$  (10 γύροι  $\times$  5 εξωτερικοί κύκλοι) για μοντέλα χρονοσειρών MCI-NC 1- ημερών και 1-ημερών-χωρίς-αφύσικες-ημέρες που πέτυχαν διάμεσο ανάκλησης  $\geq 0.5$  και ειδικότητας  $\geq 0.7$  (166 φιλτραρισμένα πειράματα), ομαδοποιημένα ανάλογα με την προσέγγιση βελτιστοποίησης υπερπαραμέτρων που χρησιμοποιήθηκε για την επιλογή του καλύτερου μοντέλου. Τα μοντέλα που χρησιμοποιούν βελτιστοποίηση υπερπαραμέτρων εμφανίζονται με κόκκινο χρώμα, ενώ τα μοντέλα με προεπιλεγμένες υπερπαραμέτρους εμφανίζονται με μπλε χρώμα. Τα διαγράμματα παρουσιάζουν: α) Ανάκληση (Recall), β) Ειδικότητα (Specificity), γ) AUC και δ) MCC.

Μεταξύ των φιλτραρισμένων πειραμάτων, οι περίοδοι εξομάλυνσης των χρονοσειρών σε διαστήματα 5, 15 και 30 λεπτών, καθώς και της 1 ώρας, εμφανίστηκαν συχνότερα (31 έως 46 φορές η καθεμία). Αντίθετα, η εξομάλυνση με την χρήση διαστήματος 45 λεπτών παρατηρήθηκε μόνο 8 φορές. Όσον αφορά την κανονικοποίηση, σε αντίθεση με το σύνολο δεδομένων των 6-ημερών, την χρησιμοποιούν 26 διαμορφώσεις του φιλτραρισμένου συνόλου 1-ημέρας. Αυτό είναι λογικό, καθώς η συμπίεση του συνόλου των 6 ημερών σε μία ημέρα μειώνει τη μεταβλητότητά της ακτιγραφίας και μπορεί να ενισχύσει τον ρόλο των ακραίων τιμών ή να κάνει ορισμένα χαρακτηριστικά πιο έντονα, αυξάνοντας την ασυμμετρία του συνόλου δεδομένων.

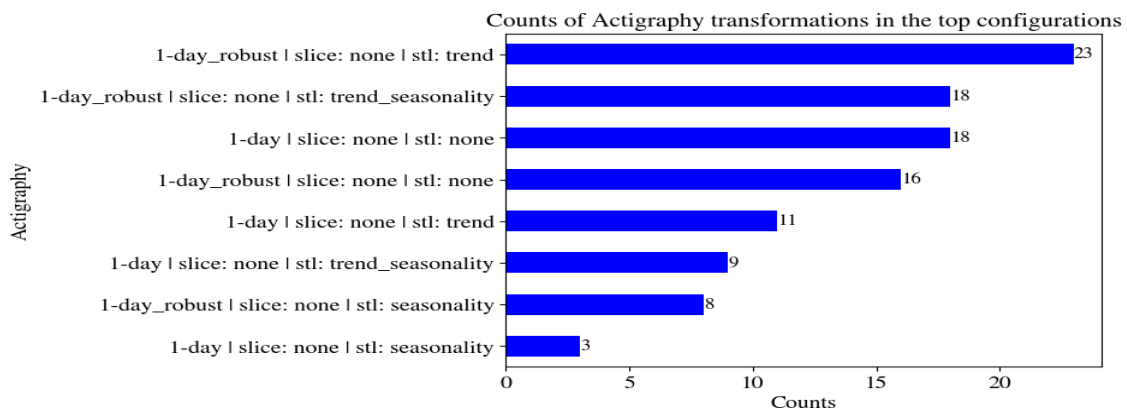
Εξετάζοντας την επιλογή των παραθύρων χρόνου 08:00:00-16:00:00 και τις επιλογές ανάμεσα στην αφαίρεση ή χρήση αφύσικων ημερών στις Εικόνες 6 α) και 6 β) αντίστοιχα, περισσότερα από το ένα τρίτο των επιλεγμένων μοντέλων (61 από τα 167) χρησιμοποιούν την ακτιγραφία στο παράθυρο 08:00:00-16:00:00, το 2<sup>ο</sup> οκτάωρο της ημέρας. Αυτό υποδηλώνει τη μεγάλη σημασία του δεύτερου δωρου και την πιθανότητα επίτευξης καλών αποτελεσμάτων ταξινόμησης αναφοράς χρησιμοποιώντας πολύ σύντομες χρονοσειρές με μόλις 8 χαρακτηριστικά (ένα δείγμα την ώρα!). Όσον αφορά την επιλογή αφαίρεσης αφύσικων ημερών (1-day\_robust), αυτή επιλέγεται περισσότερο στα κορυφαία πειράματα (91) σε σύγκριση με την εναλλακτική της (1-day) (76), κάτι που υποδηλώνει ότι η απομάκρυνση των αφύσικων ημερών πριν από τον υπολογισμό του ημερήσιου μέσου όρου ακτιγραφίας είναι γενικά επωφελής για την απόδοσης της ταξινόμησης αναφοράς σε MCI-NC κατηγοριοποίηση.

Τέλος, λαμβάνοντας υπόψη την αποσύνθεση χρονοσειρών (STL decomposition), η οποία εμφανίζεται σε 109 από τα 167 φιλτραρισμένα πειράματα, τα συστατικά της τάσης παρουσίασαν υψηλές καταμετρήσεις (46), υποδεικνύοντας την ικανότητά τους να αποτυπώνουν επαρκώς τα ημερήσια μοτίβα ακτιγραφίας. Το άθροισμα των χρονοσειρών τάσης και εποχικότητας είχε επίσης ισχυρή εκπροσώπηση (47), δείχνοντας ότι η πληροφορία της εποχικότητας μπορεί να βελτιώσει την απόδοση ενός μοντέλου όταν συνδυάζεται με την τάση. Οι χρονοσειρές μόνο με εποχικότητα είχαν τη χαμηλότερη εκπροσώπηση (16). Ωστόσο, ενδέχεται να δώσουν καλά αποτελέσματα ακόμη και όταν αναλύονται μεμονωμένα σε ορισμένες περιπτώσεις.

a)



b)



**Εικόνα 8:** Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας για τα πειράματα 1-ημερών MCI-NC που πέτυχαν τουλάχιστον διάμεση ανάκληση  $\geq 0.5$  και διάμεση ειδικότητα  $\geq 0.7$  (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός | χρονική περίοδος | συστατικά

αποσύνθεσης. a) Με χρήση τομής χρόνου 08:00-16:00 (61 πειράματα), b) Με χρήση ολόκληρων χρονοσειρών (106 πειράματα).

#### 4.3.2 Επιλεγμένα μοντέλα

Χρησιμοποιώντας την ίδια προσέγγιση όπως προηγουμένως για τα μοντέλα των 6 ημερών καταλήγουμε σε 4 ομάδες διαμορφώσεων επεξεργασίας δεδομένων-μοντέλων όπως φαίνονται στον Πίνακα 7, επιτυγχάνοντας:

1. Διάμεση Ανάκλιση = 0.6, χαμηλότερο 95% CI = 0.6
2. Διάμεση Ανάκλιση = 0.6, χαμηλότερο 95% CI = 0.5
3. Διάμεση Ανάκλιση = 0.6, Κατώτερο όριο 95% CI < 0.5
4. Διάμεση Ανάκλιση  $\leq 0.55$ , Κατώτερο όριο 95% CI < 0.5

UID	Model	Actigraphy	Num features	Trials	Optimization	Recall	Recall lower CI	Specificity	Specificity lower CI	MCC	MCC lower CI	AUC	AUC lower CI	rnCV score
2659	TSVC	1-day_robust   slice: none   stl: none	32	0	none	0.6	0.6	0.778	0.722	0.337	0.196	0.679	0.611	1.8678
2803	TSVC	1-day_robust   slice: none   stl: none	24	0	none	0.6	0.6	0.778	0.722	0.308	0.2	0.6795	0.6115	1.8551
2372	TSVC	1-day   slice: none   stl: none	96	0	none	0.6	0.6	0.778	0.714	0.335	0.2005	0.686	0.606	1.8535
2790	TSVC	1-day   slice: 08:00:00_16:00:00   stl: none	9	0	none	0.6	0.6	0.722	0.714	0.279	0.264	0.661	0.6385	1.9183
2201	TSVC	1-day_robust   slice: none   stl: none	288	0	none	0.6	0.5	0.824	0.75	0.315	0.211	0.682	0.606	1.7822
2207	TSVC	1-day   slice: none   stl: none	288	0	none	0.6	0.5	0.778	0.7498	0.304	0.218	0.6665	0.6115	1.7835
2687	TSVC	1-day   slice: none   stl: none	48	300	train_opt	0.6	0.5	0.7715	0.714	0.297	0.2005	0.6505	0.606	1.7291
2266	TSVC	1-day   slice: none   stl: trend	288	0	none	0.6	0.5	0.722	0.706	0.279	0.176	0.6505	0.602	1.7411
2405	TSVC	1-day   slice: none   stl: trend_seasonality	96	0	none	0.6	0.5	0.722	0.706	0.279	0.196	0.657	0.611	1.75
2647	TSVC	1-day   slice: none   stl: none	32	0	none	0.6	0.5	0.722	0.706	0.297	0.2045	0.661	0.6165	1.7477
2832	TSVC	1-day_robust   slice: none   stl: trend_seasonality	48	300	valid_opt	0.6	0.5	0.778	0.706	0.242	0.2245	0.6385	0.617	1.752
2828	TSVC	1-day_robust   slice: none   stl: trend_seasonality	48	300	best	0.6	0.4	0.833	0.7775	0.314	0.233	0.661	0.6305	1.7177
2587	TSVC	1-day   slice: 08:00:00_16:00:00   stl: trend_seasonality	33	300	sem	0.6	0.4	0.801	0.7715	0.314	0.176	0.661	0.606	1.6731
2413	TSVC	1-day   slice: none   stl: none	288	300	sem	0.6	0.4	0.801	0.722	0.279	0.211	0.644	0.589	1.5731
2458	TSVC	1-day_robust   slice: none   stl: none	288	300	valid_opt	0.6	0.4	0.7715	0.722	0.304	0.229	0.657	0.608	1.6053
2479	TSVC	1-day   slice: none   stl: seasonality	288	300	gap_opt	0.6	0.4	0.778	0.722	0.279	0.225	0.6385	0.607	1.619
2541	TSVC	1-day_robust   slice: none   stl: none	96	300	valid_opt	0.6	0.4	0.801	0.722	0.314	0.2329	0.661	0.617	1.6324
2839	TSVC	1-day   slice: none   stl: trend_seasonality	24	0	none	0.6	0.4	0.765	0.714	0.279	0.2245	0.644	0.617	1.6311
2415	TSVC	1-day   slice: none   stl: none	288	300	valid_opt	0.55	0.4	0.833	0.722	0.314	0.211	0.644	0.575	1.5529
2531	TSVC	1-day_robust   slice: 08:00:00_16:00:00   stl: seasonality	17	0	none	0.55	0.4	0.7435	0.714	0.208	0.1095	0.6115	0.5525	1.5187
2829	TSVC	1-day_robust   slice: none   stl: trend_seasonality	48	300	gap_opt	0.5	0.4	0.833	0.801	0.314	0.233	0.661	0.617	1.7077
2487	TSVC	1-day   slice: none   stl: trend_seasonality	288	300	valid_opt	0.5	0.4	0.833	0.75	0.297	0.218	0.6325	0.5975	1.6132
3019	TSVC	1-day   slice: 08:00:00_16:00:00   stl: trend_seasonality	9	300	best	0.5	0.4	0.778	0.722	0.218	0.11	0.6115	0.561	1.5527
3089	TSVC	1-day   slice: none   stl: trend	24	300	best	0.5	0.4	0.778	0.706	0.229	0.154	0.633	0.574	1.536

**Πίνακας 7:** Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις rnCV για τα επιλεγμένα πειράματα με μέσες ακτιγραφίες μιας ημέρας χρονοσειρών MCI-NC, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου ορίου 95% CI του διάμεσου ανάκλησης και χαμηλότερου ορίου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε τέσσερις ζώνες, διαχωρισμένες με γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος rnCV.

Αυτές οι ομάδες ενσωματώνουν μια μεγάλη ποικιλία μοντέλων, που βασίζονται σε 9 διαφορετικές διαμορφώσεις ακτιγραφίας και συνδυασμούς αριθμού χαρακτηριστικών (από 9 έως 288 δείγματα), και μοντέλα τόσο με προεπιλεγμένες όσο και με βελτιστοποιημένες υπερπαραμέτρους.

Στην 1η ομάδα, υπάρχουν μοντέλα που χρησιμοποιούν τόσο ακτιγραφίες 1-ημέρας (1-day) όσο και ακτιγραφίες χωρίς αφύσικες μέρες (1-day\_robust), με τον αριθμό χαρακτηριστικών (μήκος χρονοσειράς) να κυμαίνεται από 9 έως 96, επιτυγχάνοντας σχεδόν ίδια απόδοση. Το μοντέλο #2790, το οποίο χρησιμοποιεί τη χρονοσειρά 2<sup>ο</sup> δωρου της ημέρας, είναι

ιδιαίτερα ενδιαφέρον, καθώς επιτυγχάνει συγκρίσιμη απόδοση χρησιμοποιώντας μόνο 9 χαρακτηριστικά (ένα δείγμα κάθε ώρα)! Είναι σημαντικό να σημειωθεί ότι κανένα από αυτά τα μοντέλα δεν χρησιμοποιεί αποσύνθεση χρονοσειράς (STL decomposition) ή βελτιστοποίηση υπερπαραμέτρων.

Όσον αφορά τη 2η ομάδα, τα μοντέλα #2201 και #2207 έχουν πανομοιότυπες διαμορφώσεις, εκτός από το γεγονός ότι το #2201 χρησιμοποιεί ακτιγραφία χωρίς αφύσικες ημέρες. Αυτό οδηγεί σε υψηλότερη ειδικότητα (specificity) 0.824 (κατώτερο 95% CI: 0.75) σε σύγκριση με το #2207 (ειδικότητα = 0.778, κατώτερο 95% CI: 0.74), ενώ και τα δύο έχουν την ίδια ανάκληση. Το ίδιο ισχύει και για το μοντέλο #2659. Τα μοντέλα της 2ης ομάδας με βελτιστοποίηση υπερπαραμέτρων φτάνουν την ειδικότητα των μη βελτιστοποιημένων μοντέλων με προεπιλεγμένες υπερπαραμέτρους της 1ης ομάδας, αλλά με μείωση στο κατώτερο 95% CI της ανάκλησης κατά 0.1. Επιπλέον, σε αυτή την ομάδα εμφανίζονται ακτιγραφίες που έχουν υποστεί αποσύνθεση, αλλά δεν φαίνεται να προσφέρουν σαφές πλεονέκτημα σε σχέση με άλλες διαμορφώσεις ακτιγραφίας.

Τα μοντέλα της 3ης και 4ης ομάδας είναι σχεδόν όλα βελτιστοποιημένα ως προς τις υπερπαραμέτρους τους, και περισσότερα από τα μισά χρησιμοποιούν συστατικά αποσύνθεσης χρονοσειράς. Αυτός ο συνδυασμός οδηγεί σε ελαφρώς αυξημένη ειδικότητα (~0.8) σε σύγκριση με τις προηγούμενες ομάδες, αλλά με μείωση στην ανάκληση.

Αποδεικνύεται ότι τα πιο απλά μοντέλα όσον αφορά την βελτιστοποίηση υπερπαραμέτρων, διαμορφώσεις της ακτιγραφίας και αριθμό χαρακτηριστικών αποδίδουν καλύτερα. Οι χρονοσειρές χωρίς αφύσικες ημέρες αποδίδουν καλύτερα από αυτές που τις περιλαμβάνουν στον υπολογισμό της μέσης ακτιγραφίας. Τέλος, οι διαμορφώσεις μηχανικής μάθησης και η ρύθμιση υπερπαραμέτρων δεν προσφέρουν σημαντικά πλεονεκτήματα στο σύνολο δεδομένων 1- ημέρες, σε αντίθεση με αυτό των 6-ημερών, όπου παρατηρείτε βελτίωση στην ειδικότητα με ίδια επίπεδα ανάκλησης, όταν πραγματοποιείται βελτιστοποίηση υπερπαραμέτρων.

Ο αριθμός δειγμάτων κυμαίνεται μεταξύ [9, 288], και σχεδόν όλα τα μοντέλα έχουν διάμεση AUC στην περιοχή [0.611, 0.679], γεγονός που δείχνει ικανοποιητική προγνωστική απόδοση, δεδομένης της δυσκολίας της κατηγοριοποίησης, του μικρού και μη ισορροπημένου ως προς τις κατηγορίες MCI-NC συνόλου δεδομένων που χρησιμοποιήθηκε, και το γεγονός ότι χρησιμοποιούνται αποκλειστικά τα δεδομένα ακτιγραφίας χωρίς εξαγωγή από αυτά επιπλέον χαρακτηριστικών.

#### 4.3.3 Επικύρωση επιλεγμένων μοντέλων

Ο Πίνακας 8 παρουσιάζει τα αποτελέσματα αξιολόγησης των επιλεγμένων μοντέλων 1-ημέρας, τα οποία ξεπέρασαν τα καθορισμένα όρια σε τουλάχιστον δύο από τις τρεις μεθόδους επικύρωσης, χρησιμοποιώντας την ίδια προσέγγιση όπως στην ανάλυση των δεδομένων ακτιγραφίας 6 ημερών.

Διαπιστώνεται ότι 16 από τα 24 επιλεγμένα μοντέλα κατάφεραν να ικανοποιήσουν τα όρια ανάκλησης και ειδικότητας και στις δύο μεθόδους αξιολόγησης rCV και RCCS. Επιπλέον, 6 από αυτά τα 16 μοντέλα, ήταν εντός ορίων απόδοσης χρησιμοποιώντας και την OOB μέθοδο (σημειώνονται με έντονη γραφή στον Πίνακα 8).

Παρόμοια με τα μοντέλα 6-ημερών, αυτό υποδηλώνει ότι η επιλογή μοντέλων που βασίζεται στο mCV για το σύνολο δεδομένων 1-ημέρας μπορεί να εντοπίσει αποτελεσματικά μοντέλα με αποδεκτή προγνωστική απόδοση σε δεδομένα με διαφορετική ισορροπία κατηγοριών NC-MCI μεταξύ των συνόλων εκπαίδευσης και ελέγχου. Ωστόσο, σε αντίθεση με το 6-ήμερο σύνολο δεδομένων, τα απλούστερα μοντέλα με την χρήση δεδομένων 1-ημέρες επιδεικνύουν επίσης ισχυρές ικανότητες γενίκευσης, καθώς συναντούν επαρκώς τα όρια και

στη ΟΟΒ μέθοδο αξιολόγησης, η οποία δεν διατηρεί ισορροπία κατηγοριών ανάμεσα στα σύνολα εκπαίδευσης και ελέγχου.

UID	rnCV Recall	rnCV Specificity	rCV Recall	rCV Specificity	RCSS Recall	RCSS Specificity	OOB Recall	OOB Specificity
2201	0.6	0.824	0.6	0.778	0.5	0.778	0.4365	0.836
<b>2659</b>	<b>0.6</b>	<b>0.778</b>	<b>0.6</b>	<b>0.778</b>	<b>0.5</b>	<b>0.722</b>	<b>0.545</b>	<b>0.769</b>
<b>2803</b>	<b>0.6</b>	<b>0.778</b>	<b>0.6</b>	<b>0.778</b>	<b>0.5</b>	<b>0.722</b>	<b>0.5</b>	<b>0.767</b>
2479	0.6	0.778	0.6	0.778	0.5	0.778	0.4	0.85
2372	0.6	0.778	0.6	0.722	0.5	0.722	0.481	0.786
2207	0.6	0.778	0.6	0.778	0.5	0.778	0.444	0.846
<b>2405</b>	<b>0.6</b>	<b>0.722</b>	<b>0.6</b>	<b>0.722</b>	<b>0.5</b>	<b>0.722</b>	<b>0.5415</b>	<b>0.742</b>
<b>2647</b>	<b>0.6</b>	<b>0.722</b>	<b>0.6</b>	<b>0.722</b>	<b>0.5</b>	<b>0.722</b>	<b>0.5</b>	<b>0.767</b>
<b>2415</b>	<b>0.55</b>	<b>0.833</b>	<b>0.6</b>	<b>0.722</b>	<b>0.667</b>	<b>0.722</b>	<b>0.6</b>	<b>0.712</b>
2829	0.5	0.833	0.5	0.882	0.5	0.833	0.423	0.8545
<b>2487</b>	<b>0.5</b>	<b>0.833</b>	<b>0.6</b>	<b>0.722</b>	<b>0.5835</b>	<b>0.722</b>	<b>0.6</b>	<b>0.702</b>

**Πίνακας 8:** Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας  $B=100$  διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, ΟΟΒ) για τα επιλεγμένα μοντέλα 1 ημέρας MCI-NC. Τα αποτελέσματα της απόδοσης rnCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

#### 4.3.4 Αξιολόγηση ικανότητας πρόβλεψης σε άγνωστα δεδομένα

Μόνο 3 από τα επιλεγμένα μοντέλα πέτυχαν το όριο της ανάκλησης όταν αξιολογήθηκαν στο ίδιο σύνολο άγνωστων δεδομένων με τα επιλεγμένα μοντέλα 6-ημερών, ενώ ταυτόχρονα πλησίασαν αρκετά τα όρια απόδοσης της ειδικότητας, όπως φαίνεται στον Πίνακα 9.

UID	Model	Actigraphy	Num features	Trials	Optimization	MCC	Recall	Specificity	F1	Balanced accuracy	PRAUC	AUC
2531	TSVC	1-day_robust   slice: 08:00:00_16:00:00   stl: seasonality	17	0	none	0.181	0.571	0.611	0.59	0.591	0.492	0.591
2201	TSVC	1-day_robust   slice: none   stl: none	288	0	none	0.168	0.5	0.667	0.584	0.583	0.488	0.583
2487	TSVC	1-day   slice: none   stl: trend_seasonality	288	300	valid_opt	0.168	0.5	0.667	0.584	0.583	0.488	0.583

**Πίνακας 9:** Απόδοση των επιλεγμένων μοντέλων 1-ημέρας MCI-NC που αξιολογήθηκαν σε ένα μικρό σύνολο ελέγχου αποτελούμενο από μόλις 32 δείγματα ( $MCI = 14$ ,  $NC = 18$ ). Τα μοντέλα είναι ταξινομημένα κατά φθίνουσα σειρά ανάκλησης και ειδικότητας. Εμφανίζονται μόνο τα μοντέλα που πέτυχαν ανάκληση  $\geq 0.5$  και ειδικότητα  $\geq 0.7$ . Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε μοντέλου. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

Δεδομένου ότι προτιμούμε λύσεις με τη μέγιστη δυνατή ειδικότητα όταν η ανάκληση είναι τουλάχιστον 0.5, τα καλύτερα μοντέλα είναι το #2487 και το #2201, τα οποία χρησιμοποιούν μεγάλο αριθμό χαρακτηριστικών (288) και την πλήρη χρονοσειρά 1-ημέρας (χωρίς τομές). Ωστόσο, και τα δύο μοντέλα πέτυχαν ειδικότητα = 0.667, κάτι που αντίκειται στις προσδοκίες βάσει των αποτελεσμάτων του rnCV. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι το σύνολο άγνωστων δεδομένων είναι πιο ισορροπημένο ως προς τις κατηγορίες MCI-NC σε σχέση με το σύνολο που χρησιμοποιήθηκε στην αξιολόγηση των μοντέλων μέσω rnCV.

Ένα ενδιαφέρον εύρημα είναι το μοντέλο #2531, το οποίο χρησιμοποιεί μια χρονοσειρά με μόνο 17 δεδομένα και το 2<sup>ο</sup> 8ωρο της ημέρας, καταφέρνει να ξεπεράσει το ελάχιστο όριο ανάκλησης, πετυχαίνοντας ειδικότητα = 0.611, ενώ παράλληλα έχει το υψηλότερο AUC. Αξίζει επίσης να σημειωθεί ότι 2 από τα 3 κορυφαία μοντέλα χρησιμοποιούν αποσύνθεση, κάτι που δείχνει ότι, σε αντίθεση με το σύνολο δεδομένων των 6 ημερών, για τις

συντομότερες χρονοσειρές της μιας ημέρας, η χρήση αποσύνθεσης μπορεί να συμβάλει στη βελτίωση της απόδοσης. Μάλιστα το πολύ «φειδωλό» μοντέλο #2531 χρησιμοποιεί μόνο το τη χρονοσειρά seasonality του 2<sup>ου</sup> οκταώρου της ημέρας και μόνο ένα σημείο αυτής ανά ημίωρο. Τέτοια μοντέλα μπορεί να είναι πολύ χρήσιμα για τη συνεχή παρακολούθηση της εξέλιξης της γνωσιακής έκπτωσης σε ασθενείς με αρχική διάγνωση MCI με χρήση συσκευών wearables στο μέλλον.

Η σύγκριση των αποτελεσμάτων της mCV από τον Πίνακα 7 με την απόδοση στο άγνωστο σύνολο δεδομένων στον Πίνακα 9 αποκαλύπτει ότι μπορεί να επιτευχθεί ισχυρή απόδοση ανεξάρτητα από τον αριθμό χαρακτηριστικών ή τη χρήση αποσύνθεσης. Όσον αφορά την ανάκληση, και τα 3 μοντέλα του Πίνακα 9 ξεπερνούν το κατώτερο όριο 95% CI που αναμενόταν από τα mCV αποτελέσματα. Ωστόσο, και τα 3 μοντέλα αποτυγχάνουν να φτάσουν το αναμενόμενο κατώτερο 95% CI για τη ειδικότητα. Αυτό σημαίνει ότι, αν και το μικρό σύνολο άγνωστων δεδομένων, των μόλις 32 δειγμάτων παρέχει μόνο μια ενδεικτική εικόνα της αναμενόμενης απόδοσης των μοντέλων σε νέα δεδομένα, το γεγονός ότι όλα τα μοντέλα πετυχαίνουν τα όρια ανάκλησης αλλά αποτυγχάνουν στο κατώτερο 95% CI της ειδικότητας, υποδηλώνει ότι τα επιλεγμένα μοντέλα αναφοράς 1-ημέρας μπορεί να έχουν δυσκολία στη μείωση των ψευδώς θετικών προβλέψεων.

Τέλος, αξίζει να σημειωθεί ότι από τα 3 καλύτερα μοντέλα στο άγνωστο σύνολο δεδομένων, μόνο το μοντέλο #2487 πληροί τα κριτήρια απόδοσης σε όλες τις μεθόδους αξιολόγησης. Το μοντέλο #2201 τα πληροί για mCV και RCCS, αλλά όχι για OOB, ενώ το μοντέλο #2531 πληροί τα κριτήρια μόνο για τη μέθοδο RCCS.

Αυτά τα αποτελέσματα ενισχύουν την εμπιστοσύνη μας στο μοντέλο #2487, καθώς είναι το μόνο που διατηρεί σταθερή απόδοση σε όλα τα κριτήρια αξιολόγησης.

Η ανάλυση CCR στο σύνολο άγνωστων δεδομένων, χρησιμοποιώντας τα τρία μοντέλα του Πίνακα 9, παρουσιάζεται στην Εικόνα 9. Τα αποτελέσματα δείχνουν ότι 8 από τα 32 δείγματα του hold-out συνόλου χαρακτηρίζονται ως "δύσκολα δείγματα", με μέσο CCR  $\leq 0.2$ . Συγκεκριμένα, αυτά περιλαμβάνουν 6 δείγματα MCI και 2 δείγματα Normal (NC).

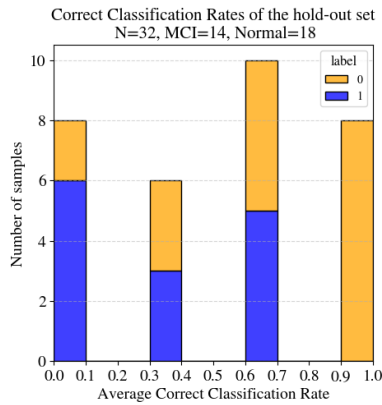
Επιπλέον, άλλα 8 από τα 32 δείγματα αναγνωρίζονται ως "εύκολα", με μέσο CCR  $\geq 0.8$ . Αυτά αποτελούνται αποκλειστικά από 8 NC δείγματα.

Παρότι τα 3 μοντέλα που χρησιμοποιούνται στην ανάλυση CCR βασίζονται σε διαφορετικά χαρακτηριστικά ακτιγραφίας, όπως την αφαίρεση αφύσικων ημερών και την χρήση του 2<sup>ου</sup> 8-ωρου της ημέρας, παρατηρείται παρόμοιο μοτίβο με αυτό των μοντέλων 6- ημερών.

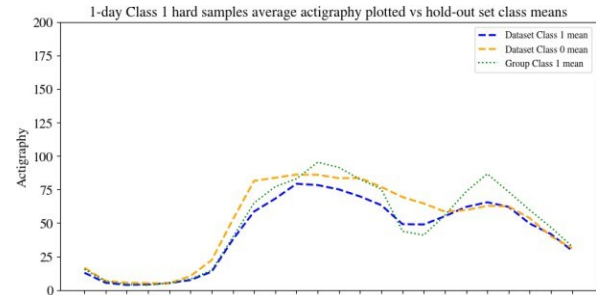
Συγκεκριμένα: Τα NC δείγματα που είναι δύσκολα στην ταξινόμηση παρουσιάζουν χαμηλά επίπεδα ακτιγραφίας, με τον μέσο όρο της ομάδας τους (πράσινη γραμμή) να είναι παρόμοιος με τον μέσο όρο της κλάσης MCI (μπλε γραμμή). Αντίθετα, τα NC "εύκολα" δείγματα διαφέρουν εμφανώς από τα MCI, διευκολύνοντας την ταξινόμησή τους. Τα MCI "δύσκολα" δείγματα εμφανίζουν υψηλότερα επίπεδα ακτιγραφίας κατά τις πρωινές και απογευματινές ώρες σε σύγκριση με τον μέσο όρο ακτιγραφίας των NC δειγμάτων στο σύνολο δεδομένων.

Αυτές οι παρατηρήσεις επιβεβαιώνουν ότι η διακριτότητα μεταξύ των κλάσεων επηρεάζεται από τα επίπεδα τιμής της ακτιγραφίας, καθώς ορισμένα MCI δείγματα εμφανίζουν μοτίβα πιο κοντά στα NC και αντίστροφα, καθιστώντας την ταξινόμηση πιο δύσκολη χωρίς την εξαγωγή περαιτέρω χαρακτηριστικών.

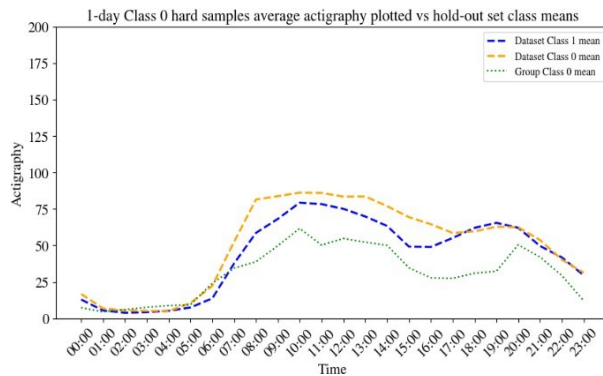
a)



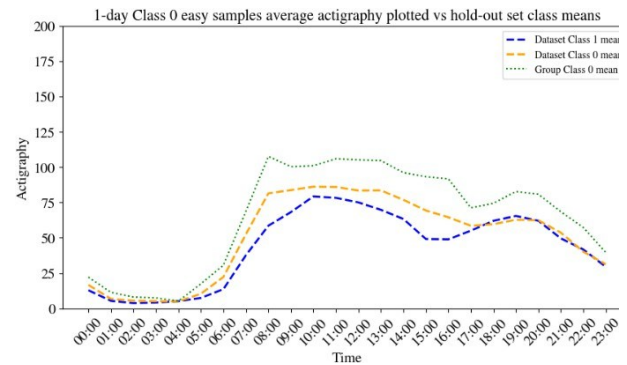
b)



c)



d)



**Εικόνα 9:** Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Σύνολο Ελέγχου (N=32, MCI=14, NC=18) και τα μοντέλα του Πίνακα 9. a) Ιστόγραμμα που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. b), c) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI (κατηγορία 1, μπλε γραμμή) και NC (κατηγορία 0, κίτρινη γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 6 MCI και 2 NC "δύσκολων δειγμάτων" με CCR  $\leq 0.2$ , αντίστοιχα. d), e) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 6 MCI και 12 NC "εύκολων δειγμάτων" (πράσινη γραμμή) με CCR  $\geq 0.8$ .

#### 4.3.5 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 1-ημερας

Τα επιλεγμένα μοντέλα μέσω της  $mCV$  για τα σύνολα δεδομένων ακτιγραφίας 6-ημερών και 1-ημέρας πέτυχαν διάμεση ανάκληση στο εύρος  $\cong [0.5, 0.6]$  και ειδικότητα  $\cong [0.7, 0.81]$ , με AUC  $\cong [0.6, 0.7]$ . Αυτό δείχνει ότι, ανεξάρτητα από τους μετασχηματισμούς χρονοσειράς και τη ρύθμιση υπερπαραμέτρων, η ανώτερη απόδοση των μοντέλων αναφοράς (baseline models) με αποκλειστική χρήση ταξινομητών χρονοσειρών και χωρίς εξαγωγή επιπλέον χαρακτηριστικών και για τα δύο σύνολα δεδομένων φαίνεται να πιάνει τα όρια των 0.5 - 0.7 για ανάκληση και ειδικότητα που θέσαμε.

Όσον αφορά τα αποτελέσματα αξιολόγησης, τα μοντέλα του 1-ήμερου συνόλου εμφανίζουν κατά μέσο όρο χαμηλότερη ειδικότητα (κατά  $\cong 0.1$ ) σε όλες τις μεθόδους αξιολόγησης. Επιπλέον, αν και έχουν την ίδια μέση ανάκληση (0.5) στις μεθόδους  $rCV$  και RCSS, η μέση ανάκληση στην OOB αξιολόγηση είναι κατά  $\cong 1.8$  υψηλότερη.

Αξιοσημείωτο είναι ότι μόνο 1 από τα 11 επιλεγμένα μοντέλα του 6-ήμερου συνόλου ξεπερνά τα όρια ειδικότητας και ανάκλησης στην OOB αξιολόγηση, σε σύγκριση με 7 από τα 24 μοντέλα του 1-ήμερου συνόλου.

Αυτά τα αποτελέσματα υποδηλώνουν ότι, αν και τα μοντέλα του 1-ήμερου συνόλου έχουν γενικά χαμηλότερη απόδοση από τα 6-ήμερα μοντέλα, προσφέρουν πιο σταθερή και φειδωλή εκμάθηση για τις κλάσεις MCI και NC. Αυτό επιτρέπει στα μοντέλα να διατηρούν

συνεπή απόδοση, ακόμη και όταν εκπαιδεύονται και αξιολογούνται σε μη ισορροπημένα ως προς την κατανομή τάξεων MCI-NC σύνολα εκπαίδευσης και ελέγχου, όπως συμβαίνει στην ΟΟΒ αξιολόγηση.

Αν και τα μοντέλα 6-ημερών φαίνεται να αποδίδουν ελαφρώς καλύτερα στο σύνολο άγνωστων δεδομένων, δεν μπορούμε να συμπεράνουμε ότι υπερέρχουν συνολικά έναντι των μοντέλων 1 ημέρας. Στην πραγματικότητα, τα μοντέλα 1 ημέρας ενδέχεται να γενικεύουν καλύτερα σε νέα δεδομένα, ακόμα και αν παρατηρείται μια μικρή μείωση της απόδοσης, βάση των παρατηρήσεων μας από την διαδικασία επικύρωσης, ειδικότερα με τον ΟΟΒ μέθοδο. Τα μοντέλα 6 ημερών μπορεί να έχουν επωφεληθεί από ευνοϊκές συνθήκες στο άγνωστο σύνολο δεδομένα, αντί να παρουσιάζουν πραγματικά ανώτερη απόδοση, μιας και το σύνολο αυτό είναι τόσο μικρό (32 δείγματα).

Τέλος, τόσο τα μοντέλα αναφοράς (baseline) 1-ημέρας όσο και 6-ημερών φαίνεται να περιορίζονται από την μη εξαγωγή χαρακτηριστικών σχετικών με το πρόβλημα κατηγοριοποίησης για τον καλύτερο διαχωρισμό των κλάσεων που δεν θα στηρίζονται κυρίως στα επίπεδα δραστηριότητας της ακτιγραφίας αλλά και σε επιμέρους μοτίβα που αυτή εμφανίζει. Η χρήση μόνο των χρονοσειρών και σχετικών ταξινομητών έχει αξία για να παρέχει μόνο μια αρχική ένδειξη (baseline) της αναμενόμενης απόδοσης και να παρέχει μοντέλα αναφοράς με τα οποία να μπορούν να συγκριθούν και να αξιολογηθούν στο μέλλον πιο σύνθετα μοντέλα MM που θα χρησιμοποιούν εξαγωγή κατάλληλων για το πρόβλημα χαρακτηριστικών (feature engineering).

## 5. ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ CSF

Στο εδάφιο αυτό, παρουσιάζουμε μοντέλα που χρησιμοποιούν αποκλειστικά δεδομένα χρονοσειρών ακτιγραφίας για την κατηγοριοποίηση ατόμων στις κλάσεις CSF test positive - CSF test negative του που αναπτύξαμε χρησιμοποιώντας τα σύνολα δεδομένων ακτιγραφίας και βιοδεικτών CSF της μελέτης ALBION. Για τον σχεδιασμό των ταξινομητών ακολουθούμε βασικά την ίδια μεθοδολογία που περιγράφεται στην ενότητα 3, αλλά με κάποιες διαφορές. Για την ανάπτυξη και σύγκριση μοντέλων πρόβλεψης χρησιμοποιούμε εδώ διαδικασία 10 γύρων mCV με 5-εσωτερικούς και 5-εξωτερικούς βρόγχους για ακτιγραφίες 6 ημερών, ενώ καταφεύγουμε σε 10 γύρους με 4-εξωτερικούς και 4-εσωτερικούς βρόγχους για μέσες ακτιγραφίες μιας ημέρας, προκειμένου να διερευνήσουμε πώς η μείωση μεγέθους του συνόλου εκπαίδευσης, ενώ αυξάνεται το μέγεθος του συνόλου ελέγχου σε κάθε βρόγχο, επηρεάζει την απόδοση της διαδικασίας mCV. Η αύξηση του μεγέθους των test sets κρίθηκε απαραίτητη λόγω του μικρού μεγέθους του συνόλου δεδομένων (μόλις 76 δείγματα, με μόνο 16 για την κατηγορία CSF θετική).

### 5.1 Αποτελέσματα τυχαίου ταξινομητή

Όπως παρουσιάζεται στον Πίνακα 10, ο τυχαίος ταξινομητής (dummy classifier) επιτυγχάνει  $MCC \cong 0$  και  $AUC \cong 0,5$ , που αντιστοιχούν σε τυχαίες προβλέψεις, όπως ήταν αναμενόμενο. Επιπλέον, η ειδικότητα είναι πολύ μεγαλύτερη από την ανάκληση λόγω του μεγαλύτερου αριθμού δεδομένων κατηγορίας CSF test negative έναντι CSF positive. Αυτά τα ευρήματα προσφέρουν μια χρήσιμη βάση αναφοράς, επιβεβαιώνοντας ότι ο ταξινομητής λειτουργεί όπως αναμενόταν κατά τη διάρκεια των 50 επαναλήψεων εκπαίδευσης-ελέγχου που προέρχονται από τους 10 γύρους επανάληψης της mCV με 5 εξωτερικούς βρόγχους. Οποιοδήποτε μοντέλο επιτύχει υψηλότερους δείκτες MCC και AUC, διατηρώντας παράλληλα μια βελτιωμένη ισορροπία μεταξύ ανάκλησης και ειδικότητας, προσφέρει δυνατότητες μάθησης και πρόβλεψης.

Metric	Median	95% CI of the median
MCC	0.0375	(-0.302, 0.075)
Recall	0.2915	(0, 0.33)
Specificity	0.75	(0.667, 0.75)
AUC	0.521	(0.33, 0.542)

*Πίνακας 10:* Αποτελέσματα ταξινομητή αναφοράς για το σύνολο δεδομένων CSF

### 5.2 Αποτελέσματα για τα δεδομένα 6-ημερών

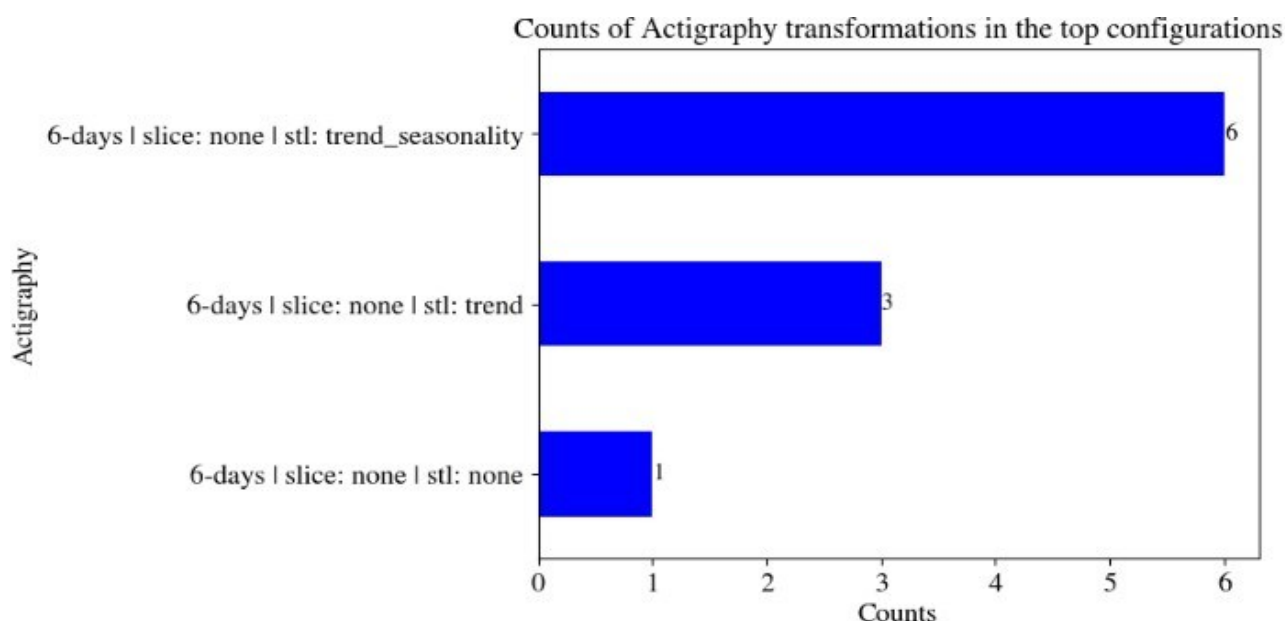
#### 5.2.1 Αποτελέσματα εξερεύνησης χώρου λύσεων

Από τα 520 πειράματα mCV, τα 471 ολοκληρώθηκαν, ενώ τα υπόλοιπα τερματίστηκαν λόγω υπέρβασης του χρονικού ορίου εκτέλεσης. Από όλα τα πειράματα που πραγματοποιήθηκαν, μόνο τα 13 (περίπου 2%) πληρούσαν τα κριτήρια επιτυχίας που θέσαμε για τον στόχο CSF: διάμεση ανάκληση (Recall)  $\geq 0,5$  και ειδικότητα (Specificity)  $\geq 0,667$ . Επιπλέον, 3 από αυτά τα πειράματα, παρουσίασαν διάμεσο  $MCC=0$  και εξαιρέθηκαν από περαιτέρω ανάλυση. Με μόνο 10 μοντέλα να παραμένουν στο "φιλτραρισμένο σύνολο", αυτά ανατίθενται άμεσα στο "επιλεγμένο σύνολο", το οποίο χρησιμεύει ως βάση για την συνέχεια της ανάλυσής μας.

Όσον αφορά τις επικρατούσες ρυθμίσεις του συνόλου δεδομένων, η εξομάλυνση 1 ώρας με 9 μετρήσεις είναι η επικρατούσα περίοδος εξομάλυνσης για το σύνολο δεδομένων CSF 6 ημερών, με μόνο 1 άλλη μέτρηση για εξομάλυνση 30 λεπτών. Αυτό μπορεί να οφείλεται στο μικρότερο μέγεθος του συνόλου δεδομένων CSF σε σχέση με το σύνολο δεδομένων MCI,

το οποίο δίνει προτεραιότητα στον ελάχιστο δυνατό αριθμό χαρακτηριστικών. Η κανονικοποίηση δεν χρησιμοποιήθηκε σε κανένα από τα 10 καλύτερα πειράματα, αλλά η αποσύνθεση της χρονοσειράς είχε σημαντική παρουσία, όπως συνοψίζεται στην Εικόνα 10. Τα συστατικά τάσης της χρονοσειράς με προστιθέμενη την εποχικότητα είχαν την μεγαλύτερη παρουσία (6 διαμορφώσεις), υποδεικνύοντας τη σημασία της εποχικότητας για την ανίχνευση προτύπων που σχετίζονται με μοτίβα σχετικά με την διάκριση CSF test positive – CSF test negative ακτιγραφιών. Η τάση μόνο συγκέντρωσε τις μισές παρουσίες (3), ενώ η μη αποσύνθεση είχε μόνο μία παρουσία στα καλύτερα πειράματα.

Όσον αφορά τα επικρατούντα χαρακτηριστικά Μηχανικής Μάθησης (MM) στα 10 επιλεγμένα μοντέλα, και τα 10 χρησιμοποίησαν τον ταξινομητή χρονοσειρών TSVC, καθιστώντας τον την καλύτερη επιλογή και για τον στόχο CSF, πέρα από τον στόχο MCI. Ενδιαφέρον έχει ότι τα μοντέλα με βελτιστοποιημένες υπερπαραμέτρους είχαν περισσότερη από διπλάσια εκπροσώπηση σε σχέση με τα μη βελτιστοποιημένα μοντέλα, με 7 και 3 παρουσίες, αντίστοιχα, τονίζοντας ότι η ρύθμιση της υπερπαραμέτρου περιθωρίου του TSCV (C) βοηθά στην βελτίωση της απόδοσης σε αυτό το μικρό σύνολο δεδομένων.



**Εικόνα 10:** Κατανομή του αριθμού των συνόλων δεδομένων χρονοσειρών ακτιγραφίας CSF test positive-CSF test negative για τα πειράματα 6 ημερών που πέτυχαν τουλάχιστον διάμεσο Recall  $\geq 0.5$  και διάμεσο Specificity  $\geq 0.667$  (σύνολο φιλτραρισμένων πειραμάτων), παρουσιασμένα στη μορφή "μετασχηματισμός | χρονική περίοδος | συστατικά αποσύνθεσης".

### 5.2.2 Επιλεγμένα μοντέλα

Τα επιλεγμένα μοντέλα μπορούν να χωριστούν σε δύο ομάδες με βάση τα χαμηλότερα αποτελέσματα του χαμηλότερου ορίου 95% CI της διάμεσου ανάκλησης που προέκυψαν από τις επιδόσεις της διαδικασίας mCV, όπως παρουσιάζονται στον Πίνακα 11.

UID	Model	Actigraphy	Num features	Trials	Optimization	Recall	Recall lower CI	Specificity	Specificity lower CI	MCC	MCC lower CI	AUC	AUC lower CI	rnCV score
34	TSVC	6-days   slice: none   stl: trend_seasonality	144	0	none	0.667	0.667	0.667	0.583	0.272	0.2	0.667	0.625	1.814
239	TSVC	6-days   slice: none   stl: trend_seasonality	144	300	sem	0.667	0.667	0.667	0.5415	0.209	0.134	0.625	0.583	1.718
240	TSVC	6-days   slice: none   stl: trend_seasonality	144	300	train_opt	0.667	0.5835	0.667	0.583	0.253	0.134	0.625	0.583	1.666
251	TSVC	6-days   slice: none   stl: trend	144	300	valid_opt	0.667	0.5835	0.667	0.583	0.158	0.075	0.583	0.542	1.606
241	TSVC	6-days   slice: none   stl: trend_seasonality	144	300	valid_opt	0.667	0.5	0.667	0.667	0.253	0.158	0.625	0.583	1.661
250	TSVC	6-days   slice: none   stl: trend	144	300	train_opt	0.667	0.5	0.7085	0.5415	0.134	0.068	0.583	0.542	1.445
237	TSVC	6-days   slice: none   stl: trend_seasonality	144	300	best	0.667	0.4165	0.75	0.667	0.272	0.167	0.646	0.583	1.547
22	TSVC	6-days   slice: none   stl: trend_seasonality	288	0	none	0.667	0.333	0.667	0.583	0.1835	0.075	0.583	0.542	1.306
247	TSVC	6-days   slice: none   stl: trend	144	300	best	0.5835	0.333	0.75	0.667	0.2	0.092	0.625	0.542	1.364
32	TSVC	6-days   slice: none   stl: none	144	0	none	0.5	0.333	0.667	0.667	0.2	0.068	0.625	0.542	1.368

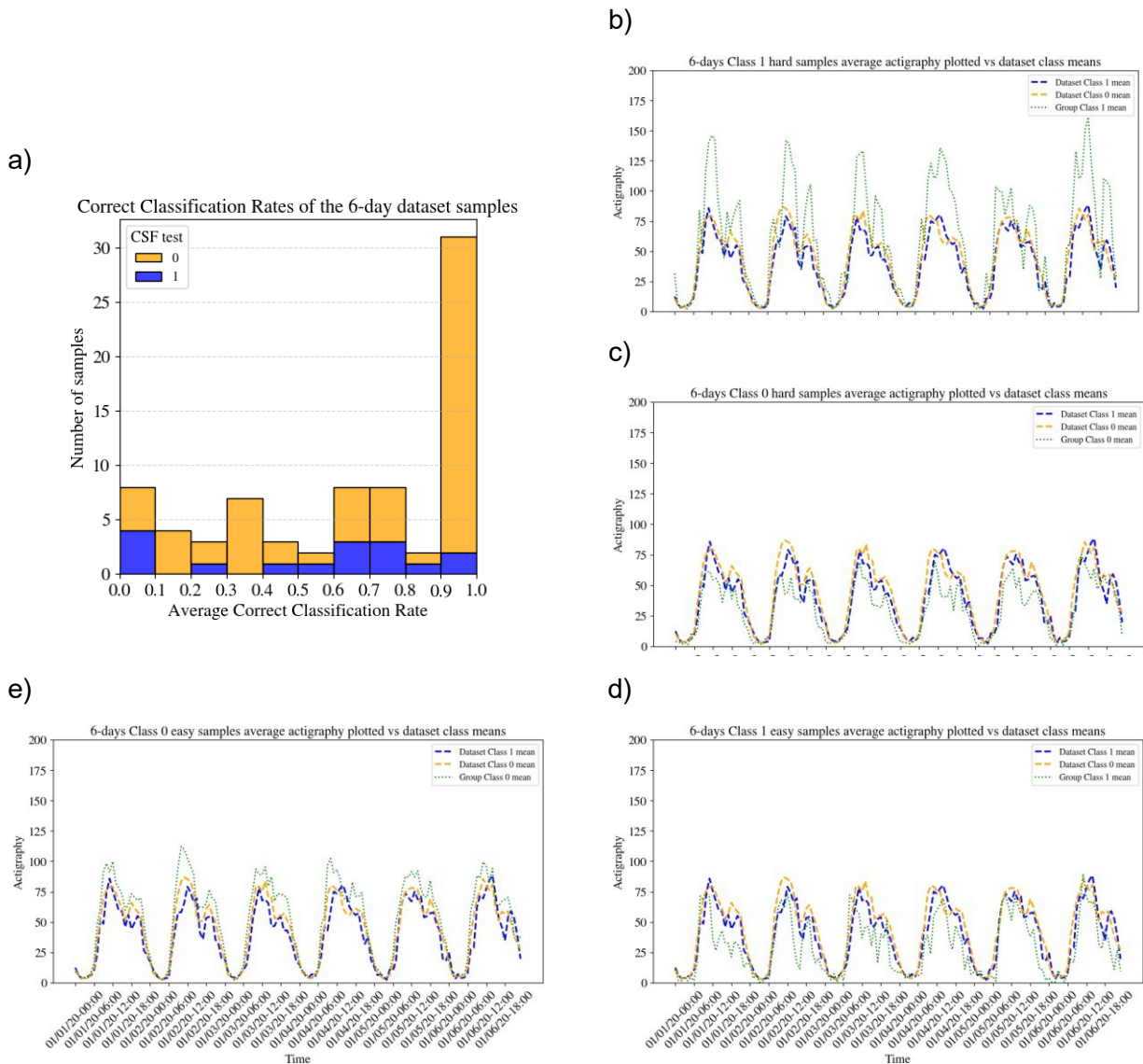
**Πίνακας 11:** Διάμεσες μετρικές απόδοσης από τις 50 επαναλήψεις rnCV για τα επιλεγμένα πειράματα 6 ημερών χρονοσειρών CSF test negative-CSF test positive, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου 95% CI του διάμεσου ανάκλησης και χαμηλότερου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε δύο ζώνες, διαχωρισμένες με γραμμές, βασισμένες στις φθίνουσες τιμές των δύο πρώτων μετρικών. Η στήλη UID αντιπροσωπεύει το μοναδικό αναγνωριστικό κάθε πειράματος rnCV

Συγκεκριμένα, η ομάδα 1 έχει ανάκληση με χαμηλότερο όριο 95% CI = [0.4, 0.667], ενώ η ομάδα 2 έχει ανάκληση με χαμηλότερο όριο 95% CI  $\leq$  0.4.

Στην 1η ομάδα επιδόσεων, όλα τα μοντέλα χρησιμοποίησαν 144 χαρακτηριστικά, που αντιστοιχούν στην πιο εκτενή επιλογή εξομάλυνσης της 1 ώρας, και αποσύνθεση της χρονοσειράς σε τάση με ή χωρίς την προσθήκη εποχικότητας. Για την ομάδα 2, η χρήση 288 χαρακτηριστικών στο μοντέλο #22 επιδεινώνει το χαμηλότερο όριο 95% CI της ανάκλησης στο ήμισυ της τιμής (=0.333) του ίδιου μοντέλου #34 (=0.667), με 144 χαρακτηριστικά.

Τα 10 επιλεγμένα μοντέλα 6 ημερών ως ομάδα παρουσιάζουν μέσο AUC > 0.6 και κάνουν φειδωλή επιλογή μοντελοποίησης μειώνοντας τον αριθμό των χαρακτηριστικών που χρησιμοποιούνται και με την χρήση πιο απλών αποσυντεθειμένων συστατικά χρονοσειρών. Το χαμηλότερο όριο 95% CI της ανάκλησης για την πρώτη ομάδα (περίπου [0.4, 0.667]) είναι σημαντικά καλύτερο από την απόδοση του τυχαίου ταξινομητή (=0.29). Αυτό το αποτέλεσμα συμφωνεί με τις προσδοκίες για μια ανάλυση απόδοσης ταξινόμησης βάσης που πραγματοποιήθηκε με βάση ένα πολύ μικρό σύνολο δεδομένων με μόνο 16 δείγματα θετικής κλάσης.

Δεδομένου ότι δεν προχωρήσαμε σε πρόβλεψη με άγνωστα δεδομένα για το σύνολο δεδομένων CSF, η ανάλυση CCR βασίζεται στα αποτελέσματα της rnCV. Παρουσιάζουμε τα αποτελέσματα αυτά στην Εικόνα 11.



**Εικόνα 11:** Ανάλυση Μέσου CCR ανά Δείγμα χρησιμοποιώντας το Αρχικά CSF δεδομένα Ελέγχου (N=76, CSF test positive=16, CSF test negative=70) και τα μοντέλα του Πίνακα 11. α) Ιστογράμμο που δείχνει τον αριθμό δειγμάτων ανά κατηγορία μέσου CCR, με χρωματική κωδικοποίηση βάσει της κατηγορίας. β), γ) Μέσες τιμές ακτιγραφίας για τις κατηγορίες CSF test positive (κατηγορία 1, πορτοκαλί γραμμή) και CSF test negative (κατηγορία 0, μπλε γραμμή) σε σύγκριση με τον μέσο όρο ακτιγραφίας (πράσινη γραμμή) των 8 CSF test negative και 4 CSF test positive "δύσκολων δειγμάτων" με CCR  $\leq 0.2$ , αντίστοιχα. δ), ε) Μέσες τιμές ακτιγραφίας για τις κατηγορίες MCI και NC σε σύγκριση με τον μέσο όρο ακτιγραφίας των 30 CSF test negative και 3 test positive "εύκολων δειγμάτων" (πράσινη γραμμή) με CCR  $\geq 0.8$ .

Οι τιμές CCR για τα CSF test negative δείγματα καλύπτουν όλη την περιοχή των κατηγοριών με μια εξαιρετικά ασύμμετρη κατανομή προς τα δεξιά (υψηλές τιμές). Η κυρίαρχη συγκέντρωση με 30 δείγματα βρίσκεται στην περιοχή των κατηγοριών 90%-100% CCR, ενώ μια μικρότερη ομάδα με 8 δείγματα βρίσκεται στην περιοχή των κατηγοριών 0%-20% CCR.

Τα CSF test positive δείγματα παρουσιάζει μια σχετικά ομοιόμορφη κατανομή των τιμών CCR, με 4 δείγματα να συγκεντρώνονται στην περιοχή 0%-10% και μόνο 3 δείγματα να πετυχαίνουν CCR  $\geq 80\%$ .

Παρόμοια με το σύνολο δεδομένων MCI, παρατηρούμε διακριτές υποομάδες με τιμές CCR  $\leq 20\%$  και  $\geq 80\%$ , υποδεικνύοντας δείγματα που είναι ιδιαίτερα δύσκολα ή εύκολα να ταξινομηθούν σωστά από τα baseline μοντέλα. Ενδιαφέρον είναι ότι τα CSF test positive δείγματα που είναι δύσκολο να ταξινομηθούν παρουσιάζουν ασυνήθιστα υψηλά επίπεδα ακτιγραφίας, ξεπερνώντας και τις δύο μέσες τιμές των κλάσεων. Αντίθετα, τα δείγματα CSF test negative που είναι δύσκολα να ταξινομηθούν δείχνουν ασυνήθιστα χαμηλά επίπεδα δραστηριότητας, πέφτοντας κάτω από τις δύο μέσες τιμές των κλάσεων. Το αντίθετο μοτίβο ισχύει για τα "εύκολα" δείγματα και στις δύο κλάσεις.

Αυτά τα ευρήματα από την ανάλυση CCR συμφωνούν με τις παρατηρήσεις μας από την ανάλυση CCR για το σύνολο δεδομένων MCI. Αυτό ενισχύει την υπόθεσή μας ότι η απόδοση των baseline μοντέλων επηρεάζεται σημαντικά από την έλλειψη χαρακτηριστικών που έχουν εξαχθεί από την ακτιγραφία για το συγκεκριμένο πρόβλημα και τη χρήση κατάλληλων ταξινομητών που να τα αξιοποιούν.

### 5.2.3 Επικύρωση επιλεγμένων μοντέλων

Μετά από εκπαίδευση με όλα τα διαθέσιμα δεδομένα, τα 10 επιλεγμένα μοντέλα αξιολογήθηκαν χρησιμοποιώντας  $B=100$  κατανομές εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης μοντέλων που αναφέρονται στην ενότητα 3.5. Από τα 10 μοντέλα, τα 5 πληρούν τα όρια ανάκλησης και ειδικότητας για τις μεθόδους rCV και το RCSS, συγκεκριμένα τα: #32, #239, #251, #241 και #237, όπως φαίνεται στον Πίνακα 12. Όσον αφορά την αξιολόγηση OOB, αυτά τα μοντέλα επιτυγχάνουν διάμεση ανάκληση στην περιοχή Recall = [0.118, 0.4] και Ειδικότητα = [0.805, 0.92], αποτυγχάνοντας να περάσουν το όριο ανάκλησης.

Αυτά τα ευρήματα δείχνουν ότι η επιλογή μοντέλων βασισμένη στον χώρο λύσεων της mCV εντοπίζει αποτελεσματικά μοντέλα ταξινόμησης αναφοράς από το σύνολο δεδομένων των 6-ημερών. Ενώ αυτά τα μοντέλα επιτυγχάνουν καλές επιδόσεις σε δεδομένα με όμοια κατανομή των δειγμάτων CSF test positive – CSF test negative στα σύνολα εκπαίδευσης-ελέγχου, τα μοντέλα #239 και #241 δείχνουν ικανότητα να μάθουν πιο γενικεύσιμα πρότυπα, αν και οι βαθμοί ανάκλησης τους (0.4) στην αξιολόγηση OOB είναι κοντά στο όριο.

UID	rnCV Recall	rnCV Specificity	rCV Recall	rCV Specificity	RCSS Recall	RCSS Specificity	OOB Recall	OOB Specificity
32	0.5	0.667	0.333	0.75	0.25	0.75	0.118	0.92
239	0.667	0.667	0.667	0.667	0.5	0.667	0.4	0.821
251	0.667	0.667	0.667	0.667	0.5	0.75	0.3665	0.805
241	0.667	0.667	0.667	0.667	0.5	0.667	0.4	0.821
237	0.667	0.75	0.5835	0.75	0.5	0.75	0.333	0.85

**Πίνακας 12:** Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας  $B=100$  διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 6 ημερών CSF test positive – CSF test negative. Τα αποτελέσματα της απόδοσης mCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

### 5.2.4 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 6-ημερών

Όπως και με τα αποτελέσματα της ανάλυσης ταξινόμησης αναφοράς για το MCI σύνολο δεδομένων, ο ταξινομητής TSVC ήταν ο βέλτιστος ταξινομητής χρονοσειρών από αυτούς που αξιολογήσαμε. Συγκεκριμένα για το μικρό σύνολο των CSF δεδομένων η

βελτιστοποίηση της υπερπαράμετρου C του ταξινομητή είναι ουσιώδης για τη βελτίωση της απόδοσης.

Εφικτά μοντέλα ταξινόμησης αναφοράς μπορούν να επιτευχθούν σχεδόν αποκλειστικά χρησιμοποιώντας 144 χαρακτηριστικά (ένα δείγμα ανά ώρα), που είναι η μεγαλύτερη (δηλαδή με τον μικρότερο αριθμό χαρακτηριστικών) περίοδος εξομάλυνσης χρονοσειράς που εξετάσαμε για την ανάλυση των δεδομένων χρονοσειρών 6 ημερών, δείχνοντας τη σημασία της λιτής μοντελοποίησης (parsimonious modeling) για ένα τόσο μικρό σύνολο δεδομένων.

Όπως και με τα δεδομένα MCI, η αναμενόμενη απόδοση όλων των επιλεγμένων μοντέλων εμφανίζει ένα μικρό εύρος (περίπου  $\pm 0.1$ ) στην ανάκληση, την ειδικότητα και το AUC παρά των διαφορετικών διαμορφώσεων του συνόλου δεδομένων. Αυτή η περιορισμένη μεταβλητότητα υποδηλώνει ότι διαφορετικοί μετασχηματισμοί δεδομένων παρέχουν μόνο μερικές βελτιώσεις, υποδεικνύοντας ένα όριο απόδοσης όταν χρησιμοποιούνται άμεσα τα δεδομένα χρονοσειρών χωρίς επεξεργασία χαρακτηριστικών, οδηγώντας σε απόδοση στην περιοχή AUC = [0.583, 0.667].

Συνολικά, λαμβάνοντας υπόψη το μικρό μέγεθος του συνόλου δεδομένων και τη σοβαρή ανισότητα κλάσεων του, η ανάλυση βασισμένη στο mCV, και αποκλειστικά στα δεδομένα (δείγματα) ακτιγραφίας και αντίστοιχους ταξινομητές χρονοσειρών εντοπίζει υποσχόμενες λύσεις βασικών μοντέλων και υποδεικνύει τους τύπους μετασχηματισμών που πρέπει να γίνουν στην αρχική ακτιγραφία για να τις επιτύχουμε.

### 5.3 Αποτελέσματα για τα δεδομένα 1-ημέρας

#### 5.3.1 Αποτελέσματα εξερεύνησης χώρου λύσεων

Στη συνέχεια, πραγματοποιήσαμε πειράματα χρησιμοποιώντας το σύνολο δεδομένων 1-ημέρας. Λόγω του πολύ μικρότερου μεγέθους κάθε χρονοσειράς σε σχέση με τα δεδομένα 6-ημερών, όλα τα 2080 πειράματα που εκτελέστηκαν ολοκληρώθηκαν επιτυχώς. Ωστόσο, μόνο τα 31 (1%) πληρούσαν τα κριτήρια για να ενταχθούν στα φιλτραρισμένα μοντέλα με διάμεση ανάκληση  $\geq 0,5$  και ειδικότητα  $\geq 0,667$ . Επιλέξαμε μόνο 7 μοντέλα από αυτό το υποσύνολο, τα οποία πέτυχαν διάμεσο AUC  $> 0,5$ , για να συντάξουμε το σύνολο επιλεγμένων δεδομένων για περαιτέρω ανάλυση.

Η περίοδος εξομάλυνσης των 45 λεπτών είναι η μόνη εμφανίζεται στο επιλεγμένο σύνολο μοντέλων, και οδηγεί σε 32 χαρακτηριστικά (δείγματα χρόνου) ανά χρονοσειρά. Όπως παρατηρήθηκε και στο σύνολο δεδομένων 6-ημερών, η απόδοση των CSF μοντέλων φαίνεται να εξαρτάται σε μεγάλο βαθμό από την περίοδο εξομάλυνσης, δίνοντας προτεραιότητα σε λιτή μοντελοποίηση που μπορεί να ανιχνεύσει γενικά πρότυπα σε αυτό το μικρό σύνολο δεδομένων. Όσον αφορά την κανονικοποίηση, τα 4 από τα 7 μοντέλα χρησιμοποίησαν τη μέθοδο yeo-Johnson, ακολουθώντας την ίδια λογική όπως για το σύνολο δεδομένων MCI 1 ημέρας. Ενδιαφέρων είναι ότι όλα τα επιλεγμένα πειράματα χρησιμοποιούν χρονοσειρές με αφαίρεση αφύσικων ημερών, πριν από τον υπολογισμό της μέσης τιμής ακτιγραφίας, υποδεικνύοντας ότι αυτές ενδέχεται να επηρεάζουν σοβαρά την εκμάθηση μοτίβων σχετικών με τον στόχο ταξινόμησης αναφοράς CSF δεδομένων.

Αξιοσημείωτο είναι ότι σε αυτή την ανάλυση, δύο ταξινομητές χρονοσειρών εκπροσωπούνται στο επιλεγμένο σύνολο μοντέλων. Για τα 4 μοντέλα που βασίζονται σε 1NNTS ταξινομητές, όλες οι ρυθμίσεις αποσύνθεσης εκπροσωπούνται μία φορά, όπως φαίνεται στον Πίνακα 13. Επιπλέον οι 3 λύσεις βασισμένες στον ταξινομητή TSVC χρησιμοποιούν μόνο το συστατικό τάσης των τελευταίων 8 ωρών της ημέρας.

Οι ταξινομητές 1NNTS δεν υποστηρίζουν βελτιστοποίηση παραμέτρων, ωστόσο όλα τα επιλεγμένα μοντέλα TSVC είναι βελτιστοποιημένα ως προς τις παραμέτρους τους,

ενισχύοντας την παρατήρηση ότι αυτή βοηθά την απόδοση του των μοντέλων σε αυτό το σύνολο δεδομένων, όπως φαίνεται και στα μοντέλα των 6 ημερών. Τα 3 μοντέλα TSVC είναι πανομοιότυπα, εκτός από τη μέθοδο βελτιστοποίησης που χρησιμοποιείται για τη ρύθμιση, η οποία περιλαμβάνει τις προσεγγίσεις βελτιστοποίησης best, gap και validation.

### 5.3.2 Επιλεγμένα μοντέλα

Όπως φαίνεται στον Πίνακα 12, όλα τα επιλεγμένα μοντέλα έχουν διάμεση ανάκληση = 0.5 με χαμηλότερο 95% CI της ανάκλησης = 0.25, διαχωρισμένα σε δύο διακριτές ομάδες με βάση τον τύπο του μοντέλου και τις τιμές ειδικότητας. Η ομάδα 1 περιλαμβάνει τα 4 μοντέλα 1NNTS με ειδικότητα = 0.867, και η ομάδα 2 περιλαμβάνει τα 3 μοντέλα TSVC με ειδικότητα = 0.8.

UID	Model	Actigraphy	Num features	Trials	Optimization	Recall	Recall lower	Specificity	Specificity lower	MCC	MCC lower	AUC	AUC lower	mCV score
104	1NNTS	1-day_robust   slice: none   stl: seasonality	32	0	none	0.5	0.25	0.867	0.867	0.322	0.244	0.65	0.592	1.534
113	1NNTS	1-day_robust   slice: none   stl: none	32	0	none	0.5	0.25	0.867	0.867	0.322	0.244	0.65	0.592	1.534
120	1NNTS	1-day_robust   slice: none   stl: trend_seasonality	32	0	none	0.5	0.25	0.867	0.867	0.322	0.244	0.65	0.592	1.533
128	1NNTS	1-day_robust   slice: none   stl: trend	32	0	none	0.5	0.25	0.867	0.867	0.322	0.244	0.65	0.592	1.534
1983	TSVC	1-day_robust   slice: 08:00:00_16:00:00   stl: trend	12	300	best	0.5	0.25	0.8	0.733	0.205	0.13	0.6045	0.558	1.445
1984	TSVC	1-day_robust   slice: 08:00:00_16:00:00   stl: trend	12	300	gap_opt	0.5	0.25	0.8	0.733	0.205	0.1545	0.617	0.575	1.4217
1987	TSVC	1-day_robust   slice: 08:00:00_16:00:00   stl: trend	12	300	valid_opt	0.5	0.25	0.8	0.733	0.205	0.05	0.617	0.525	1.371

**Πίνακας 13:** Διάμεσες μετρικές απόδοσης από τις 40 επαναλήψεις mCV για τα επιλεγμένα πειράματα 1-ημέρας χρονοσειρών CSF test negative-CSF test positive, ταξινομημένα κατά φθίνουσα σειρά διάμεσου ανάκλησης, χαμηλότερου 95% CI του διάμεσου ανάκλησης και χαμηλότερου 95% CI του διάμεσου ειδικότητας. Ο πίνακας χωρίζεται σε δύο ζώνες, διαχωρισμένες με γραμμές, με βάση τον τύπο του ταξινομητή που χρησιμοποιείτε.

Στην 1η ομάδα, όλα τα μοντέλα χρησιμοποίησαν 32 χαρακτηριστικά, με αφαίρεση αφύσικων ημερών και χωρίς βελτιστοποίηση παραμέτρων. Σημαντικό είναι ότι ανεξαρτήτως ποια από τις 4 επιλογές αποσύνθεσης χρησιμοποιήθηκε, τα αποτελέσματα της διάμεσης απόδοσης ήταν ταυτόσημα. Αυτό υποδηλώνει ότι με μόνο 32 χαρακτηριστικά διαθέσιμα, τα μοντέλα ενδέχεται να μην είναι σε θέση να εκμεταλλευτούν τα λεπτομερή χαρακτηριστικά των διαφόρων μεθόδων αποσύνθεσης, καθώς λειτουργούν κοντά στο ανώτατο όριο απόδοσης τους.

Όσον αφορά την 2η ομάδα, χρησιμοποιείται το ίδιο σύνολο δεδομένων και η ίδια ρύθμιση Μηχανικής Μάθησης, εκτός από την μοναδική επιλογή μοντέλου, που επιτυγχάνει παρόμοια αποτελέσματα απόδοσης. Η βελτιστοποίηση με τη μέθοδο gap επιτυγχάνει τα υψηλότερα χαμηλότερα 95% CI της ανάκλησης και της ειδικότητας.

Συνολικά, λαμβάνοντας υπόψη το μικρό μέγεθος και την ανισορροπία των κλάσεων του συνόλου δεδομένων, φαίνεται ότι τα καλύτερα μοντέλα βάσης έχουν κάνει λιτές επιλογές. Επιπλέον, η διάμεση χαμηλότερη 95% CI ανάκληση = 0.25, η οποία είναι ίδια για όλα τα μοντέλα, υποδεικνύει ότι αυτά τα μοντέλα ενδέχεται να έχουν δυσκολίες στην επίτευξη του επιθυμητού επιπέδου ανάκλησης 0.5 σε άγνωστα δεδομένα. Μπορούμε να εξετάσουμε περαιτέρω αυτό το σημείο αναλύοντας τα αποτελέσματα αξιολόγησης των μοντέλων στην επόμενη ενότητα.

### 5.3.3 Επικύρωση επιλεγμένων μοντέλων

Από τα 7 μοντέλα, μόνο τα 3 μοντέλα TSVC (#1983, #1984, #1987) πληρούσαν τα καθορισμένα όρια για την ανάκληση και την ειδικότητα σε όλες τις μεθόδους αξιολόγησης rCV, RCSS και OOB, όπως φαίνεται στον Πίνακα 14. Τα 4 μοντέλα 1NNTS δεν πληρούσαν αυτά τα κριτήρια, πιθανότατα λόγω της έλλειψης δυνατότητας εξισορρόπηση κατηγοριών, η οποία υποστηρίζεται μόνο από τα μοντέλα ταξινομητή TSVC.

Η απόδοση των μοντέλων 1 ημέρας TSVC σε όλες τις μεθόδους αξιολόγησης και την mCV υποδεικνύει την ανθεκτικότητά τους να γενικεύουν καλά σε διάφορους διαχωρισμούς του συνόλου δεδομένων λόγω των λιτών χαρακτηριστικών σε σχέση με το σύνολο δεδομένων 6 ημερών.

UID	mCV Recall	mCV Specificity	rCV Recall	rCV Specificity	RCSS Recall	RCSS Specificity	OOB Recall	OOB Specificity
1983	0.5	0.8	0.5	0.833	0.5	0.833	0.5	0.805
1984	0.5	0.8	0.667	0.75	0.5	0.75	0.5	0.783
1987	0.5	0.8	0.667	0.75	0.5	0.75	0.5	0.783

**Πίνακας 14:** Διάμεσες μετρικές απόδοσης για ανάκληση και ειδικότητα που επιτεύχθηκαν χρησιμοποιώντας B=100 διαχωρισμούς εκπαίδευσης-ελέγχου και τρεις διαφορετικές μεθόδους αξιολόγησης (rCV, RCSS, OOB) για τα επιλεγμένα μοντέλα 1-ημέρας CSF test positive – CSF test negative. Τα αποτελέσματα της απόδοσης mCV περιλαμβάνονται επίσης ως σημείο αναφοράς. Για περισσότερες λεπτομέρειες, ανατρέξτε στο κείμενο.

### 5.3.4 Συμπεράσματα από την ανάλυση των δεδομένων χρονοσειρών 1-ημέρας

Τα CSF μοντέλα αναφοράς 1-ημέρας παρουσιάζουν λιγότερο ισορροπημένη ανάκληση και ειδικότητα σε σύγκριση με τα αντίστοιχα μοντέλα 6 ημερών στις επιδόσεις της mCV. Ωστόσο, οι δοκιμές επικύρωσης έδειξαν ότι όλα τα 3 επιλεγμένα μοντέλα TSVC υπερέβησαν τα καθορισμένα όρια απόδοσης, σε αντίθεση με τα μοντέλα των 6 ημερών. Αυτό υποδηλώνει ότι οι 40 επαναλήψεις mCV για τα δεδομένα 1-ημέρας ενδέχεται να υποεκτίμησαν τη πραγματική ικανότητα γενίκευσης.

Επιπλέον, είναι προφανές ότι η λιτή μοντελοποίηση με μεγάλες περιόδους εξομάλυνσης, δηλαδή χαμηλή ανάλυση των χρονοσειρών, προτιμάται συνεχώς και στα δύο σύνολα δεδομένων μιας ημέρας και 6 ημερών, υποδεικνύοντας ότι τα μοντέλα επιστρέφουν σε συντηρητικά γενικά πρότυπα λόγω του μικρού μεγέθους του συνόλου δεδομένων ανάπτυξής τους. Αυτό το σημείο ενισχύεται περαιτέρω από την καθολική επιλογή αφαίρεσης αφύσικων ημερών, καθώς και την επιλογή του δεύτερου 8ωρου της ημέρας με μόνο 12 χαρακτηριστικά.

## 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

**Εισαγωγή και Πλαίσιο Έρευνας** Η ακτιγραφία αναδεικνύεται σε μια πολύτιμη, χαμηλού κόστους και μη επεμβατική μέθοδο για τη συνεχή παρακολούθηση της ανθρώπινης δραστηριότητας, προσφέροντας κρίσιμα δεδομένα για την έγκαιρη πρόγνωση της **Ήπιας Γνωσιακής Εξασθένησης (MCI)** και τη διερεύνηση βιοδεικτών **Εγκεφαλονωτιαίου Υγρού (CSF)**. Η έρευνα, που υλοποιήθηκε σε συνεργασία με τη μελέτη **ALBION** του Αιγινήτειου Νοσοκομείου, επικεντρώθηκε στην ανάπτυξη **μοντέλων αναφοράς (baseline models)** χρησιμοποιώντας ταξινομητές χρονοσειρών απευθείας στα δεδομένα, χωρίς την παραδοσιακή και χρονοβόρα διαδικασία της εξαγωγής χαρακτηριστικών (feature engineering).

**Επιδόσεις Μοντέλων και Μετρικές** Παρά τις προκλήσεις των περιορισμένων δειγμάτων (119 για MCI, 76 για CSF) και την έντονη ανισορροπία των κατηγοριών (αναλογία 4:1 υπέρ των υγιών ατόμων), τα αποτελέσματα κρίνονται ελπιδοφόρα για επίπεδα αναφοράς.

- **Για τον στόχο MCI:** Τα μοντέλα πέτυχαν διάμεσο **AUC** μεταξύ **0,61 και 0,71**, με ευαισθησία (Recall) στο διάστημα [0,5 0,6] και ειδικότητα (specificity) στο διάστημα [0,7, 0,83].
- **Για τον στόχο CSF:** Παρά το εξαιρετικά μικρό δείγμα, επιτεύχθηκαν τιμές διάμεσου **AUC** μεταξύ **0,58 και 0,67**, με Recall  $\geq 0,5$  και Specificity  $> 0,667$ . Ο ταξινομητής **TSVC (Time Series Support Vector Classifier)** υπερείχε σταθερά των υπολοίπων που δοκιμάστηκαν (1NNTS, TSF), παρουσιάζοντας την καλύτερη ικανότητα μάθησης ουσιαστικών μοτίβων.

**Τεχνικά Ευρήματα και «Φειδωλή» Μοντελοποίηση** Η έρευνα κατέδειξε ότι οι «φειδωλές» (parsimonious) προσεγγίσεις συχνά προσφέρουν την καλύτερη ισορροπία μεταξύ ακρίβειας και γενίκευσης, αποφεύγοντας την υπερπροσαρμογή. Συγκεκριμένα, βέλτιστες επιδόσεις παρατηρήθηκαν όταν τα μοντέλα εστίαζαν:

1. Στη **μέση ημερήσια ακτιγραφία** (συμπύεση των 6 ημερών σε μία).
2. Στο **δεύτερο οκτάωρο** της ημέρας (08:00-16:00), γεγονός που υποδηλώνει ότι σύντομα αλλά χαρακτηριστικά παράθυρα δραστηριότητας αρκούν για την ταξινόμηση.
3. Αποκλειστικά στη **χρονοσειρά τάσης (trend)** ή στο συνδυασμό χρονοσειράς **τάσης και εποχικότητας (seasonality)** που προκύπτει από την αποσύνθεση STL του σήματος.

**Περιορισμοί και Μελλοντικές Κατευθύνσεις** Η ανάλυση του Ποσοστού Σωστής Ταξινόμησης (**CCR**) ανά δείγμα αποκάλυψε ένα εγγενές «ταβάνι» στην **απόδοση** των μοντέλων που βασίζονται αποκλειστικά σε τιμές δραστηριότητας. Διαπιστώθηκε δυσκολία στον διαχωρισμό υγιών ατόμων με χαμηλή δραστηριότητα από ασθενείς με υψηλή δραστηριότητα. Ως εκ τούτου, η μελλοντική βελτίωση της διαγνωστικής ακρίβειας απαιτεί:

- Την εξαγωγή **εξειδικευμένων χαρακτηριστικών** (όπως μέθοδοι **cosinor**, ενδοημερήσια μεταβλητότητα **IV** και διαημερήσια σταθερότητα **IS**).
- Την εφαρμογή τεχνικών εξισορρόπησης δεδομένων (π.χ. **SMOTE**) και την ενσωμάτωση **πολυτροπικών δεδομένων** (βιοδείκτες, στατιστικά ύπνου).
- Τη χρήση μεγαλύτερων και πιο ισορροπημένων συνόλων δεδομένων που θα επιτρέψουν την αξιοποίηση τεχνικών **βαθιάς μάθησης**.

Τα παρόντα μοντέλα αναφοράς αποτελούν το απαραίτητο θεμέλιο για τη σύγκριση και αξιολόγηση μελλοντικών, πιο σύνθετων υπολογιστικών προσεγγίσεων στην ιατρική ακριβείας των νευροεκφυλιστικών νόσων.

## ΠΑΡΑΡΤΗΜΑ

Μοντέλο	Χώρος Υπερπαραμέτρων	Λόγος τιμής υπερπαραμέτρου / πολυπλοκότητας
1NNTS	'n_neighbors': 1	N/A
TSF	'n_estimators': [50, 200], 'n_windows': [3, 20], 'max_depth': [3, 12], 'min_samples_split': [2, 10], 'min_samples_leaf': [1, 5], 'ccp_alpha': [0.001, 0.1]	'n_estimators': Increases, 'n_windows': Increases, 'max_depth': Increases, 'min_samples_split': Increases, 'min_samples_leaf': Increases, 'ccp_alpha': Decreases
TSVC	'C': [0.0001, 1000.0]	'C': Increases

**Πίνακας 15:** Χώρος υπερπαραμέτρων ταξινομητών που χρησιμοποιήθηκαν στην ταξινόμηση αναφοράς.

Μοντέλο	Προεπιλεγμένες υπερπαραμέτροι
1NNTS	'n_neighbors': 1
TSF	'n_estimators': 500, 'n_windows': 1., 'max_depth': None, 'min_samples_split': 2, 'min_samples_leaf': 1, 'ccp_alpha': 0.
TSVC	'C': 1.0

**Πίνακας 16:** Προεπιλεγμένες υπερπαραμέτροι ταξινομητών που χρησιμοποιήθηκαν στην ταξινόμηση αναφοράς.

## ΑΝΑΦΟΡΕΣ

- [1] J. L. Martin and A. D. Hakim, "Wrist Actigraphy," *Chest*, vol. 139, no. 6, pp. 1514-1527, Jun. 2011, doi: 10.1378/chest.10-1872.
- [2] D. F. Kripke, D. J. Mullaney, S. Messin, and V. G. Wyborney, "Wrist actigraphic measures of sleep and rhythms," *Electroencephalogr. Clin. Neurophysiol.*, vol. 44, no. 5, pp. 674-676, May 1978, doi: 10.1016/0013-4694(78)90133-5.
- [3] D. J. Mullaney, D. F. Kripke, and S. Messin, "Wrist-Actigraphic Estimation of Sleep Time," *Sleep*, vol. 3, no. 1, pp. 83-92, Mar. 1980, doi: 10.1093/sleep/3.1.83.
- [4] J. B. Webster, D. F. Kripke, S. Messin, D. J. Mullaney, and G. Wyborney, "An Activity-Based Sleep Monitor System for Ambulatory Use," *Sleep*, vol. 5, no. 4, pp. 389-399, Sep. 1982, doi: 10.1093/sleep/5.4.389.
- [5] A. Sadeh, M. Sharkey, and M. A. Carskadon, "Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues," *Sleep*, vol. 17, no. 3, pp. 201-207, May 1994, doi: 10.1093/sleep/17.3.201.
- [6] A. Sadeh, P. J. Hauri, D. F. Kripke, and P. Lavie, "The Role of Actigraphy in the Evaluation of Sleep Disorders," *Sleep*, vol. 18, no. 4, pp. 288-302, Jun. 1995, doi: 10.1093/sleep/18.4.288.
- [7] "Actigraphy Watches for Sale| Actigraphy Monitoring Devices - Condor Instruments," Condor. Accessed: Oct. 02, 2024. [Online]. Available: <https://condorinst.com/en/acttrust-two-actigraph/>
- [8] "Philips - Actiwatch 2 Activity monitor," Philips. Accessed: Oct. 02, 2024. [Online]. Available: <https://www.philips.co.uk/healthcare/product/HC1044809/actiwatch-2-activity-monitor>
- [9] "GENEActiv," Activinsights. Accessed: Oct. 02, 2024. [Online]. Available: <https://activinsights.com/technology/geneactiv/>
- [10] "wGT3X-BT | ActiGraph Wearable Devices." Accessed: Oct. 02, 2024. [Online]. Available: <https://theactigraph.com/actigraph-wgt3x-bt>
- [11] K. V. Danilenko *et al.*, "Wearable Light-and-Motion Dataloggers for Sleep/Wake Research: A Review," *Appl. Sci.*, vol. 12, no. 22, p. 11794, Nov. 2022, doi: 10.3390/app122211794.
- [12] W. W. Tryon and R. Williams, "Fully proportional actigraphy: A new instrument," *Behav. Res. Methods Instrum. Comput.*, vol. 28, no. 3, pp. 392-403, Sep. 1996, doi: 10.3758/BF03200519.
- [13] M. M. Sánchez-Ortuño, J. D. Edinger, M. K. Means, and D. Almirall, "Home Is Where Sleep Is: An Ecological Approach to Test the Validity of Actigraphy for the Assessment of Insomnia," *J. Clin. Sleep Med.*, vol. 06, no. 01, pp. 21-29, Feb. 2010, doi: 10.5664/jcsm.27706.
- [14] V. Natale, G. Plazzi, and M. Martoni, "Actigraphy in the Assessment of Insomnia: A Quantitative Approach," *Sleep*, vol. 32, no. 6, pp. 767-771, Jun. 2009, doi: 10.1093/sleep/32.6.767.
- [15] Y. Tazawa *et al.*, "Actigraphy for evaluation of mood disorders: A systematic review and meta-analysis," *J. Affect. Disord.*, vol. 253, pp. 257-269, Jun. 2019, doi: 10.1016/j.jad.2019.04.087.
- [16] S. J. Diem *et al.*, "Measures of Sleep-Wake Patterns and Risk of Mild Cognitive Impairment or Dementia in Older Women," *Am. J. Geriatr. Psychiatry*, vol. 24, no. 3, pp. 248-258, Mar. 2016, doi: 10.1016/j.jagp.2015.12.002.
- [17] A. Doherty *et al.*, "Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study," *PLOS ONE*, vol. 12, no. 2, p. e0169649, Feb. 2017, doi: 10.1371/journal.pone.0169649.
- [18] G.-Q. Zhang *et al.*, "The National Sleep Research Resource: towards a sleep data commons," *J. Am. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351-1358, Oct. 2018, doi: 10.1093/jamia/ocy064.
- [19] M. R. Patterson *et al.*, "40 years of actigraphy in sleep medicine and current state of the art algorithms," *Npj Digit. Med.*, vol. 6, no. 1, p. 51, Mar. 2023, doi: 10.1038/s41746-023-00802-1.

- [20] J.-G. Choi, I. Ko, and S. Han, "Depression Level Classification Using Machine Learning Classifiers Based on Actigraphy Data," *IEEE Access*, vol. 9, pp. 116622-116646, 2021, doi: 10.1109/ACCESS.2021.3105393.
- [21] J. Li, Y. Rong, H. Meng, Z. Lu, T. Kwok, and H. Cheng, "TATC: Predicting Alzheimer's Disease with Actigraphy Data," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London United Kingdom: ACM, Jul. 2018, pp. 509-518. doi: 10.1145/3219819.3219831.
- [22] G. Cornelissen, "Cosinor-based rhythmometry," *Theor. Biol. Med. Model.*, vol. 11, no. 1, p. 16, Dec. 2014, doi: 10.1186/1742-4682-11-16.
- [23] C. Gao *et al.*, "Approaches for Assessing Circadian Rest-Activity Patterns Using Actigraphy in Cohort and Population-Based Studies," *Curr. Sleep Med. Rep.*, vol. 9, no. 4, pp. 247-256, Oct. 2023, doi: 10.1007/s40675-023-00267-4.
- [24] R. Fossion, A. L. Rivera, J. C. Toledo-Roy, J. Ellis, and M. Angelova, "Multiscale adaptive analysis of circadian rhythms and intradaily variability: Application to actigraphy time series in acute insomnia subjects," *PLOS ONE*, vol. 12, no. 7, p. e0181762, Jul. 2017, doi: 10.1371/journal.pone.0181762.
- [25] S. J. Kim, Y. C. Lim, H. J. Kwon, and J. H. Lee, "Association of rest-activity and light exposure rhythms with sleep quality in insomnia patients," *Chronobiol. Int.*, vol. 37, no. 3, pp. 403-413, Mar. 2020, doi: 10.1080/07420528.2019.1696810.
- [26] C. H. C. Yeung, C. Bauer, and Q. Xiao, "Associations between actigraphy-derived rest-activity rhythm characteristics and hypertension in UNITED STATES adults," *J. Sleep Res.*, vol. 32, no. 5, p. e13854, Oct. 2023, doi: 10.1111/jsr.13854.
- [27] L. Yang *et al.*, "Association of circadian rest-activity rhythms with cardiovascular disease and mortality in type 2 diabetes," *Diabetes Res. Clin. Pract.*, vol. 197, p. 110262, Mar. 2023, doi: 10.1016/j.diabres.2023.110262.
- [28] P. Cavalcanti-Ferreira *et al.*, "A nonparametric methodological analysis of rest-activity rhythm in type 2 diabetes," *Sleep Sci.*, vol. 11, no. 04, pp. 281-289, Aug. 2018, doi: 10.5935/1984-0063.20180044.
- [29] Q. Xiao *et al.*, "Cross-sectional and Prospective Associations of Rest-Activity Rhythms With Metabolic Markers and Type 2 Diabetes in Older Men," *Diabetes Care*, vol. 43, no. 11, pp. 2702-2712, Nov. 2020, doi: 10.2337/dc20-0557.
- [30] A. C. Cote, R. J. Phelps, N. S. Kabiri, J. S. Bhangu, and K. "Kip" Thomas, "Evaluation of Wearable Technology in Dementia: A Systematic Review and Meta-Analysis," *Front. Med.*, vol. 7, p. 501104, Jan. 2021, doi: 10.3389/fmed.2020.501104.
- [31] E. S. Musiek, M. Bhimasani, M. A. Zangrilli, J. C. Morris, D. M. Holtzman, and Y.-E. S. Ju, "Circadian Rest-Activity Pattern Changes in Aging and Preclinical Alzheimer Disease," *JAMA Neurol.*, vol. 75, no. 5, p. 582, May 2018, doi: 10.1001/jamaneurol.2017.4719.
- [32] P. H. Lee, A. Neishabouri, A. C. Y. Tse, and C. C. Guo, "Comparative Analysis and Conversion between ActiWatch and ActiGraph Open-Source Counts," In Review, preprint, Dec. 2022. doi: 10.21203/rs.3.rs-2367595/v1.
- [33] F. Kalligerou *et al.*, "Aiginition Longitudinal Biomarker Investigation Of Neurodegeneration (ALBION): study design, cohort description, and preliminary data," *Postgrad. Med.*, vol. 131, no. 7, pp. 501-508, Oct. 2019, doi: 10.1080/00325481.2019.1663708.
- [34] N. Scarmeas *et al.*, "Initial Data and a Clinical Diagnosis Transition for the Aiginition Longitudinal Biomarker Investigation of Neurodegeneration (ALBION) Study," *Medicina (Mex.)*, vol. 58, no. 9, p. 1179, Aug. 2022, doi: 10.3390/medicina58091179.
- [35] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, pp. 2079-2107, 2010.

- [36] J. Reback *et al.*, *pandas-dev/pandas: Pandas 1.0.5*. (Jun. 17, 2020). Zenodo. doi: 10.5281/ZENODO.3898987.
- [37] S. Seabold and J. Perktold, “statsmodels: Econometric and statistical modeling with python,” in *9th Python in Science Conference*, 2010.
- [38] S. Weisberg, “Yeo-Johnson power transformations,” *Dep. Appl. Stat. Univ. Minn. Retrieved June*, vol. 1, p. 2003, 2001.
- [39] G. King and L. Zeng, “Logistic Regression in Rare Events Data,” *Polit. Anal.*, vol. 9, no. 2, pp. 137-163, 2001, doi: 10.1093/oxfordjournals.pan.a004868.
- [40] “KNeighborsTimeSeriesClassifier — sktime documentation.” Accessed: Oct. 19, 2024. [Online]. Available: [https://www.sktime.net/en/stable/api\\_reference/auto\\_generated/sktime.classification.distance\\_based.KNeighborsTimeSeriesClassifier.html](https://www.sktime.net/en/stable/api_reference/auto_generated/sktime.classification.distance_based.KNeighborsTimeSeriesClassifier.html)
- [41] M. Cuturi, J.-P. Vert, O. Birkenes, and T. Matsui, “A Kernel for Time Series Based on Global Alignments,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Honolulu, HI: IEEE, Apr. 2007, p. II-413-II-416. doi: 10.1109/ICASSP.2007.366260.
- [42] “pyts.classification.TimeSeriesForest – pyts 0.13.0 documentation.” Accessed: Oct. 19, 2024. [Online]. Available: <https://pyts.readthedocs.io/en/stable/generated/pyts.classification.TimeSeriesForest.html>