

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ  
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΗΣ ΔΙΟΙΚΗΣΗΣ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΟΙΚΟΝΟΜΙΑΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

"

:

: 231/01061

2014

## **A. ΣΚΟΠΟΣ ΕΡΓΑΣΙΑΣ**

Το αντικείμενο της παρούσας διπλωματικής είναι η ανάλυση ιατρικών δεδομένων με μεθόδους υπολογιστικής νοημοσύνης. Συγκεκριμένα, μελετάται η ταξινόμηση και διάγνωση του τύπου Αφασίας των ασθενών, βασιζόμενοι στα αποτελέσματα του Aachen Aphasia Test. Για την ανάλυση χρησιμοποιείται το πρόγραμμα WEKA το οποίο παρέχει μια μεγάλη συλλογή ετοιμών αλγόριθμων που μπορούν να χρησιμοποιηθούν σε πολλά προβλήματα, όπως για παράδειγμα προβλήματα ταξινόμησης. Επιπλέον γίνεται προσπάθεια εντοπισμού των ταξινομητών εκείνων που παρέχουν ακρίβεια αλλά και απλότητα, έτσι ώστε τα αποτελέσματα να μπορούν εύκολα να γίνουν κατανοητά αλλά και αξιοποιήσιμα σε πολλές ιατρικές εφαρμογές.

Στην εργασία αυτή χρησιμοποιήθηκε η βάση δεδομένων 256 ασθενών με αφασία η οποία βρίσκεται στην διεύθυνση <http://fuzzy.iau.dtu.dk/aphasia.nsf> και αποτελεί εργαλείο εκμάθησης για την κατηγοριοποίηση της αφασίας με την χρήση υπολογιστικής νοημοσύνης.

# ΚΕΦΑΛΑΙΟ 1 ΙΑΤΡΙΚΟ ΠΡΟΒΛΗΜΑ

## 1.1 Γενικά- Τι είναι η Αφασία ;

Η λέξη αφασία (aphasia) προέρχεται από την ελληνική γλώσσα ,από το στερητικό α- και – φέτος από το ρήμα φημι που σημαίνει λέγω. Αναφέρεται σε κάθε μερική ή ολική απώλεια γλωσσικών ικανοτήτων σε ενήλικες και παιδιά και γενικότερα ανικανότητα λόγου. Η αφασία επισημαίνεται όταν εξαιτίας της εγκεφαλικής βλάβης ένα ή περισσότερα μέρη του λόγου δεν λειτουργούν σωστά. Ο λόγος είναι το κύριο μέσο του ανθρώπου ώστε να επικοινωνήσει με το περιβάλλον του. Στην περίπτωση της αφασίας το άτομο δεν μπορεί να μιλήσει,να αποδώσει το σωστό νόημα στις λέξεις, να κατανοεί και πολλές φορές να διαβάζει. Εκτός από την αφασία μπορεί να προκληθεί και παράλυση ή και άλλα προβλήματα σχετικά με:

- τις συνειδητές/εκούσιες κινήσεις,
- την παρατήρηση του περιβάλλοντος,
- την αυτοσυγκέντρωση, την πρωτοβουλία, και την μνήμη.

Οι άνθρωποι που υποφέρουν από αφασία αντιμετωπίζουν προβλήματα επικοινωνίας καθημερινά. Δεν υπάρχουν δυο άνθρωποι που υποφέρουν από αφασία στον ίδιο βαθμό, η αφασία είναι διαφορετική για τον καθένα. Ο βαθμός της εξέλιξης της αφασίας εξαρτάται, από πολλούς παράγοντες, όπως είναι η τοποθεσία, και βαθμός της εγκεφαλικής βλάβης, η γλωσσολογική ικανότητα του ασθενή ακόμα και η προσωπικότητά του. Ορισμένοι άνθρωποι με αφασία μπορούν να καταλάβουν το λόγο σωστά, αλλά έχουν πρόβλημα να βρουν τις σωστές λέξεις ή να φτιάξουν προτάσεις. Κάποιοι άλλοι, ωστόσο, μπορούν να μιλήσουν άνετα, αλλά ό,τι λένε είναι δυσνόητο. Αυτοί οι ασθενείς συνήθως έχουν μεγάλο πρόβλημα να καταλάβουν το λόγο. Η γλωσσολογική ικανότητα των ανθρώπων με αφασία βρίσκεται συνήθως ανάμεσα σε αυτά τα δύο άκρα. Πρέπει να αναφέρουμε ότι η αφασία δεν επηρεάζει την νοημοσύνη του ασθενούς η νοημοσύνη του παραμένει σε γενικές γραμμές άθικτη. Εκείνο που διαταράσσεται στα άτομα με αφασία είναι η ικανότητα προσέγγισης ιδεών και σκέψεων μέσω του λόγου και όχι οι ιδέες και οι σκέψεις αυτές καθαυτές.

Σχεδόν πάντα γίνεται κάποια αυτόματη αποκατάσταση του λόγου όταν παρουσιάζεται η αφασία. Η συγκεκριμένη αποκατάσταση ωστόσο είναι σπάνια ή δεν είναι ολοκληρωτική. Παρ' όλα 'αυτά με πολλή εξάσκηση, προσπάθεια και επανάληψη μπορεί να επιτευχθεί κάποια βελτίωση. Η Αφασία προσβάλλει περίπου 1 στους 250 ανθρώπους – και είναι πιο συχνή από τη νόσο Πάρκινσον, την εγκεφαλική παράλυση ή τη μυική δυστροφία. Κάθε χρόνο, αυξάνεται ο αριθμός των ατόμων που εκδηλώνουν τη διαταραχή. Ωστόσο, οι περισσότεροι άνθρωποι δεν γνωρίζουν τη διαταραχή αυτή. Η Αφασία είναι πιο συχνή σε ηλικιωμένα άτομα, αλλά μπορεί να εμφανιστεί σε ανθρώπους κάθε ηλικίας, εθνικότητας και φύλου.

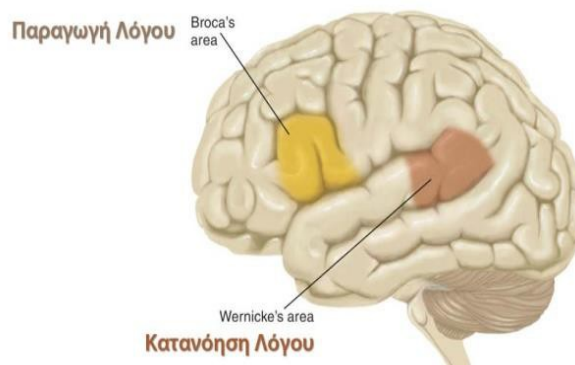
Υπάρχουν διάφορα κέντρα και σύλλογοι που στηρίζουν τα αφασιακά άτομα τα πιο γνωστά είναι το Κέντρο Ανταπόκρισης της Εθνικής Εταιρείας Αφασίας (National Aphasia Association – NAA με την αντίστοιχη ιστοσελίδα [www.aphasia.org](http://www.aphasia.org) . Η NAA παρέχει πληροφορίες σχετικά με τη διαταραχή καθώς και μέσα διαβίωσης για τους ασθενείς με αφασία, όπως εθνικό κατάλογο υποστηρικτικών ομάδων, ενημερωτικό δελτίο, εθνικό συνέδριο και άλλες εκδηλώσεις. Στην Ελλάδα υπάρχει ο Ελληνικός Σύλλογος για την Αφασία <http://www.aphasia.gr/> που διαθέτει υποστηρικτικό υλικό, βιβλία με ασκήσεις , λεξικά τσέπης και οτιδήποτε είναι απαραίτητο για τον συγκεκριμένο ασθενή. Με την εξέλιξη της τεχνολογίας υπάρχουν και ηλεκτρονικές εφαρμογές για την αντιμετώπιση της αφασίας [1].

## 1.2 Λίγη ιστορία

Το μυστήριο της δημιουργίας του λόγου έγινε αντικείμενο όχι μόνο θαυμασμού αλλά και μελέτης από τους επιστήμονες ήδη από τον 19ο αιώνα. Στον 20ο αιώνα στις μελέτες εμπλέκεται και ένας νέος κλάδος της επιστήμης, η γλωσσολογία, η οποία προσπαθεί από κοινού με την νευρολογία να απαντήσει κι αυτή στα ερωτήματα: πώς μιλάμε; και σε ποιο σημείο του εγκεφάλου εντοπίζεται η συγκεκριμένη λειτουργία του λόγου;

Την πρώτη απάντηση στο ερώτημα έδωσε το 1864 ο Pierre Paul Broca: "nous parlons avec l'hémisphère gauche!". Ο Broca βασίστηκε σε μελέτες ασθενών με αδυναμία ομιλίας, αλλά επαρκή κατανόηση του λόγου. Η νεκροτομή έδειξε βλάβη στην αριστερή οπίσθια περιοχή του μετωπιαίου λοβού σε όλους τους ασθενείς. Ο όρος «αφασία» εμφανίζεται για πρώτη φορά στις αρχές του 20ου αιώνα. μέχρι τότε αναφερόταν με τον όρο «αφημία» τον οποίο εισήγαγε ο Paul Broca από το 1861. Ο όρος αυτός αντικαταστάθηκε μετά από πολλές διαμάχες μεταξύ των περισσότερων ερευνητών. Οι πιο πολλοί είχαν δημιουργήσει το δικό τους ορισμό.

Το 1876, ο Karl Wernicke, σε ηλικία 26 ετών, δημοσίευε την εργασία του με τίτλο «το σύμπλεγμα συμπτωμάτων της αφασίας: ψυχολογική μελέτη επί ανατομικής βάσεως», η οποία έγινε η βάση και για τις κατοπινές θεωρίες της γλώσσας. Ο Karl Wernicke υποστήριξε ότι η γλώσσα εμπεριέχει ξεχωριστά κινητικά και αισθητικά προγράμματα που ρυθμίζονται από ξεχωριστές περιοχές: το κινητικό πρόγραμμα, που ρυθμίζει τις κινήσεις στόματος κατά την ομιλία, εντοπίζεται στην περιοχή Broca (Σχήμα 1), που είναι τοποθετημένη μπροστά από την κινητική περιοχή που ελέγχει το στόμα, τη γλώσσα, την υπερώα, τις φωνητικές χορδές, ενώ το αισθητικό πρόγραμμα, που ρυθμίζει την αντίληψη των λέξεων, εντοπίζεται (στην ονομαζόμενη σήμερα περιοχή Wernicke) στον κροταφικό λοβό - περιβαλλόμενη από τον ακουστικό φλοιό και τις περιοχές που ολοκληρώνουν την ακουστική, οπτική και σωματική αισθητικότητα σε σύνθετες αντιλήψεις (συνειρμικός φλοιός). Διατύπωσε, ένα συνεκτικό γλωσσικό μοντέλο, το οποίο χρησιμοποιείται, με κάποιες τροποποιήσεις, ακόμη και σήμερα.



Το 1950, ο Wilder Penfield επιβεβαίωσε την ύπαρξη γλωσσικών περιοχών που είχαν περιγράψει οι Broca και Wernicke. Ο Penfield προκαλούσε διέγερση περιοχών που σχετίζονταν με το λόγο σε ασθενείς με επιληψία κατά την περιεγχειρητική φάση, ώστε να επιβεβαιώσει ότι οι δομές αυτές δεν βλάπτονταν κατά τη διάρκεια της επέμβασης.

Σύμφωνα με τον Brookshire «η αφασία ορίζεται ως η διαταραχή της ικανότητας κάποιου ασθενή να επεξεργάζεται σύμβολα σε εκφραστικό και αντιληπτικό επίπεδο. Η αφασία εμπεριέχει μία ποικιλία από πολλές παθολογικές καταστάσεις, π.χ. από μία κωματώδη κατάσταση σε γλωσσικές διαταραχές ή ακόμη και παντελή έλλειψη του λόγου. Μιλάμε για αφασία μόνο όταν οι διαταραχές του λόγου οφείλονται σε βλάβες συγκεκριμένων φλοιϊκών και υποφλοιϊκών περιοχών του Κεντρικού Νευρικού Συστήματος. Αυτές οι δυσλειτουργίες παρατηρούνται μετά την ολοκληρωμένη απόκτηση της γλώσσας. [2]

Την δεκαετία του 1980, οι Michael Posner, Marcus Raichle και συνεργάτες, με τη βοήθεια της PET (Τομογραφίας Εκπομπής Ποζιτρονίων) πραγματοποίησαν μελέτες σε υγιή άτομα προκειμένου να διακρίνουν εάν η νευρική αντιπροσώπευση μιας λέξης την οποία διαβάζουμε μεταφέρεται επίσης κατά μήκος των ακουστικών οδών πριν αποκτήσει νόημα ή εάν μπορεί να μεταδοθεί απευθείας στην περιοχή Broca. Οι μελέτες αυτές έδειξαν ότι: α) όταν ακούμε λέξεις ενεργοποιείται η περιοχή Wernicke ενώ όταν διαβάζουμε λέξεις, χωρίς να τις ακούμε ή να τις εκφέρουμε, δεν υπάρχει ενεργοποίηση της περιοχής Wernicke και β) ότι η οπτική πληροφορία από τον ινιακό φλοιό μεταφέρεται απευθείας στην περιοχή Broca χωρίς να μετατραπεί σε ακουστική αντιπροσώπευση. Συμπεράναν έτσι ότι χρησιμοποιούνται διαφορετικές οδοί για γραπτές ή προφορικές λέξεις, ενώ ακόμη και όταν κάποιος αναζητά τη σημασία μιας λέξης ενεργοποιείται μια τελείως διαφορετική περιοχή του αριστερού μετωπιαίου φλοιού. Επομένως, το μοντέλο Wernicke-Geschwind δεν κατέστη ικανό να εξηγήσει την πολυπλοκότητα των διεργασιών του λόγου, καθώς η επεξεργασία της γλώσσας γίνεται τόσο εν σειρά όσο και εν παραλλήλω. [3]

Το 1985, γίνεται η ταξινόμηση των αφασικών διαταραχών από τους Geschwind και Damasio. Η ταξινόμηση βασίζεται στο μοντέλο Wernicke-Geschwind που αποτελεί τον πυρήνα των κλινικών και ακαδημαϊκών μελετών της αφασίας. Σύμφωνα με αυτό:

1. Όταν ακούμε μια λέξη, η πληροφορία μεταφέρεται από το όργανο της ακοής μέσω του ακουστικού νεύρου στον έσω γονατώδη πυρήνα κι από κει στον πρωτοταγή ακουστικό φλοιό (πεδίο 41 κατά Brodmann), κατόπιν στον ακουστικό φλοιό ανώτερης τάξης (πεδίο 42), εν συνεχεία στην γωνιώδη έλικα (περιοχή του βρεγματο-κροταφο-ινιακού συνειρμικού φλοιού-πεδίο 39) και στην περιοχή Wernicke (πεδίο 22). Από την περιοχή αυτή, μέσω της τοξοειδούς δεσμίδας, η πληροφορία φτάνει στην περιοχή Broca (πεδίο 45) και μετά στην κινητική περιοχή του φλοιού η οποία ελέγχει τη φώνηση (Σχήμα 2)[10].
2. Όταν διαβάζουμε μια λέξη, η πληροφορία μεταφέρεται από τον αμφιβληστροειδή προς το έξω γονατώδες σώμα κι από κει στον πρωτοταγή οπτικό φλοιό (πεδίο 17) και στο κέντρο ανώτερης τάξης (πεδίο 18). Μετά η πληροφορία φτάνει στη γωνιώδη έλικα, κατόπιν στην περιοχή Wernicke - όπου μετασχηματίζεται σε ακουστική αντιπροσώπευση- και μέσω της τοξοειδούς δεσμίδας στην περιοχή Broca. Όταν ο τελευταίος μετασχηματισμός δεν γίνεται, χάνεται η ικανότητα γλωσσικής έκφρασης. Το μοντέλο αυτό βοήθησε τον Wernicke να προβλέψει έναν νέο τύπο αφασίας (αφασία αγωγής- Leitungsaphasie) που περιγράφηκε αργότερα κλινικώς. [3]

Ο Hegde στα βιβλία του αναφέρει πολλούς ορισμούς για την αφασία. Αναφέρει ότι «η αφασία είναι μία αιτία γλωσσικής διαταραχής από κάποιο τραυματισμό του εγκεφάλου σε α) όλους τους τομείς της γλωσσικής κατανόησης και παραγωγής, οι οποίοι είναι εξασθενημένοι σε διαφορετικό βαθμό και β) ένας ή πολλοί τομείς αυτής της γλωσσικής κατανόησης και παραγωγής μπορούν να επηρεάσουν τη γλωσσική εξασθένηση της αφασίας. Η αιτία που προκαλεί αφασία είναι κάποιο χτύπημα ή ένα εγκεφαλοαγγειακό ατύχημα που στερεί το οξυγόνο και προκαλεί τραύματα στον εγκέφαλο» [5].

Ανακαιφαλαίνοντας μπορούμε να βγάλουμε τον συγκεκριμένο πίνακα [6]

Η αφασία είναι	Η αφασία δεν είναι
-Νευρογενής	-Διαταραχή Ομιλίας
-Επίκτητη	-Αίσθητηριακή διαταραχή
-Διαταραχή του λόγου (προφορικού και γραπτού)	-Διαταραχή Νοημοσύνης
-Επηρεάζει ψυχογλωσσολογικές διαδικασίες	-Άνοια
-Η λογοθεραπευτική παρέμβαση ασθενών με αφασία δεν περιλαμβάνει μόνο θεραπεία για τα γλωσσικά προβλήματα	-Σχιζοφρένια
	-Συγχυσμένος Λόγος

### 1.3 Αιτίες

Το κύριο σύμπτωμα στις αφασίες είναι μία αποδιοργάνωση των γλωσσολογικών λειτουργιών του ατόμου από βλάβη του δεξιού ή του αριστερού εγκεφαλικού ημισφαιρίου. Η συνηθέστερη αιτιολογία της αφασίας είναι το εγκεφαλικό επεισόδιο (περίπου 25 – 40% του πληθυσμού που υπέστη εγκεφαλικό επεισόδιο εκδηλώνει αφασικά χαρακτηριστικά). Υπάρχουν όμως και άλλες βλάβες οι οποίες προκαλούν την αφασία αυτές είναι :

1. εγκεφαλική θρόμβωση
2. εγκεφαλική εμβολή
3. εγκεφαλική αιμορραγία
4. κρανιοεγκεφαλικές κακώσεις
5. εγκεφαλικοί όγκοι
6. εκφυλιστικές- μεταβολικές- λοιμώδεις ασθένειες.

Ο εγκέφαλός μας χρειάζεται οξυγόνο και γλυκόζη προκειμένου να λειτουργήσει. Εάν, εξαιτίας εγκεφαλικού ή άλλων αιτιών, η κυκλοφορία του αίματος στον εγκέφαλο παρουσιάσει επιπλοκές, τότε τα εγκεφαλικά κύτταρα θα νεκρωθούν στην συγκεκριμένη περιοχή. Στον εγκέφαλο υπάρχουν διάφορες περιοχές με διαφορετικές λειτουργίες. Στους περισσότερους ανθρώπους η περιοχή για τη χρήση της γλώσσας βρίσκεται στο αριστερό μέρος του εγκεφάλου. Σε περίπτωση τραύματος σε αυτές τις περιοχές τότε αναφερόμαστε στην αφασία. Συνήθως όταν ο ασθενής πάσχει από αφασία, επηρεάζονται και άλλα μέρη του εγκεφάλου με αποτέλεσμα να παρουσιαστούν επιπλέον προβλήματα που είναι τα εξής : ημιπληγία - δυσλειτουργία μέρους των μυών το σώματος, ημιοψία-απώλεια της μισής όρασης, απραξία - άγνοια σχετικά με κάποιες διαδικασίες πως πραγματοποιούνται π.χ. το ντύσιμο, δυσφαγία - προβλήματα με την κατάποση τροφών και υγρών, προβλήματα μνήμης, διαφορετικές αντιδράσεις, επιληψία.

Η Αφασία προσβάλλει περίπου 1 στους 250 ανθρώπους – και είναι πιο συχνή από τη νόσο Πάρκινσον, την εγκεφαλική παράλυση ή τη μυική δυστροφία. Κάθε χρόνο, αυξάνεται ο αριθμός των ατόμων που εκδηλώνουν τη διαταραχή. Ωστόσο, οι περισσότεροι άνθρωποι δεν γνωρίζουν τη διαταραχή αυτή.

### 1.4 Ταξινόμηση Αφasiών

Οι αφασίες ταξινομούνται, σύμφωνα με τους Adams & Victor [7], ως εξής:

1. Αφασία Broca
2. Αφασία Wernicke
3. Σφαιρική αφασία (καθολική)
4. Σύνδρομο γλωσσικού διαχωρισμού που με τη σειρά της υποδιαιρείται σε άλλες κατηγορίες:
  - α) αφασία αγωγής
  - β) αμιγής λεκτική κώφωση
  - γ) αμιγής λεκτική τύφλωση (οπτική λεκτική αγνωσία ή αλεξία χωρίς αγραφία)
  - δ) αμιγής λεκτική βωβότητα
  - ε) κατονομαστική αφασία
  - στ) απομόνωση των περιοχών του λόγου (διαφλοιώδεις αφασίες):
    - i) διαφλοιώδης αισθητική
    - ii) διαφλοιώδης κινητική
5. Αγραφίες
6. Υποφλοιώδεις αφασίες

Στο σχήμα (Σχήμα 1) παρακάτω βλέπουμε τα κριτήρια του διαχωρισμού των κατηγοριών της αφασίας [8].

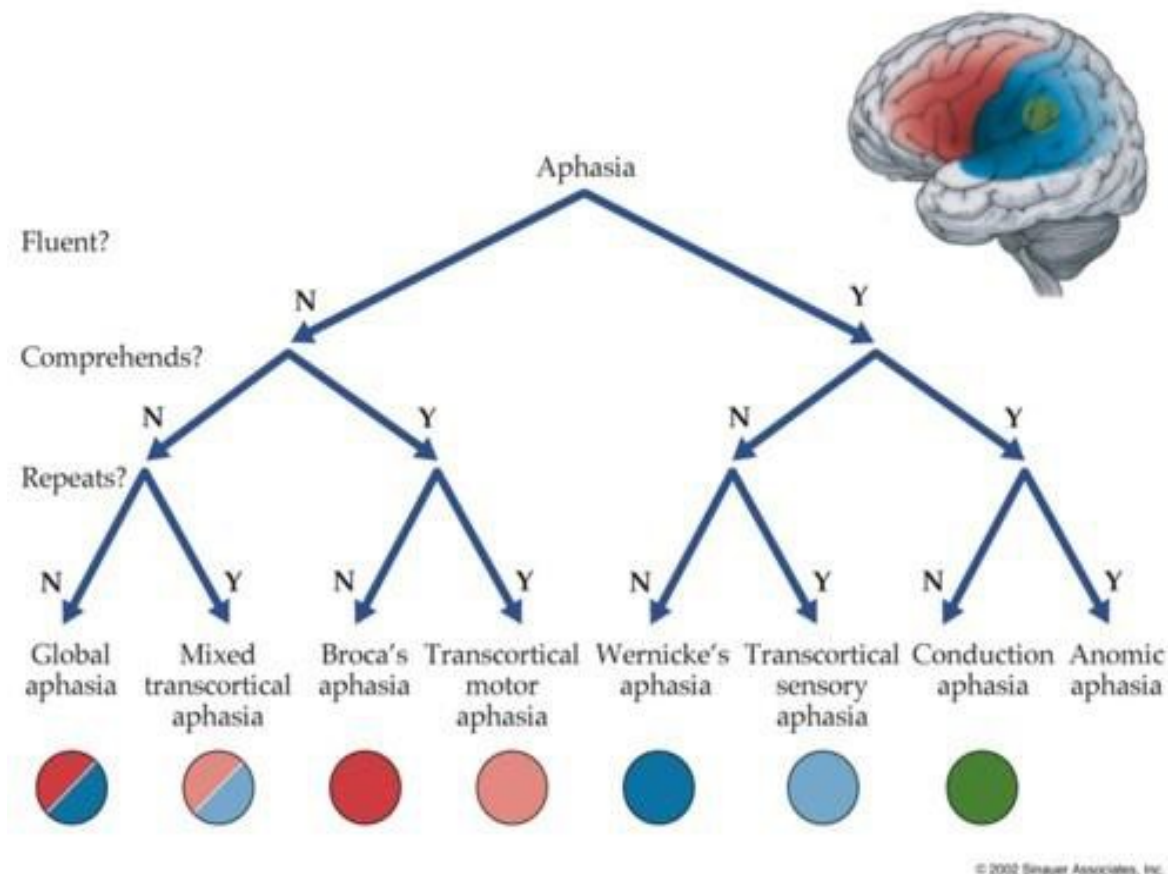


Figure 1: Είδη αφασίας.



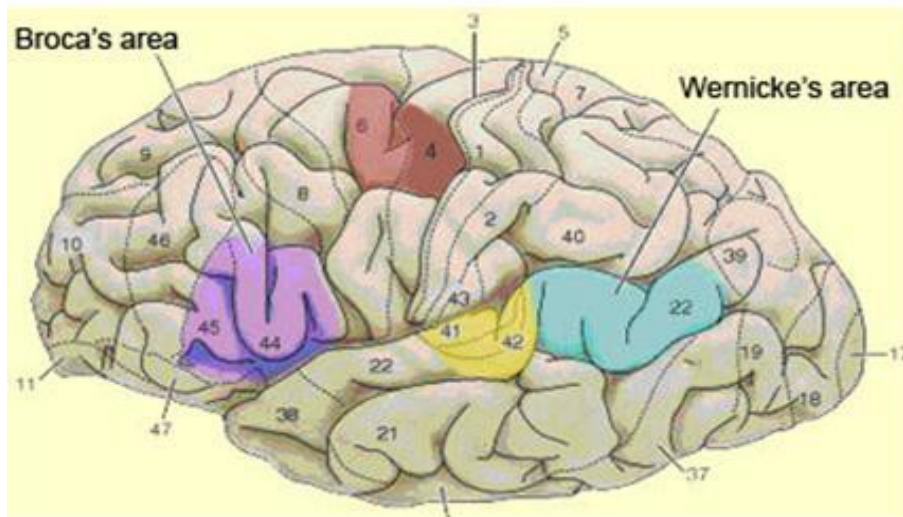


Figure 2 Πεδία του εγκεφάλου

### 1.4.1 Αφασία Broca

Το κύριο χαρακτηριστικό αυτής της αφασίας είναι η μείωση της παραγωγής γλώσσας με διατήρηση της κατανόησης της γλώσσας. Η βλάβη εντοπίζεται στον κινητικό συνειρμικό φλοιό του μετωπιαίου λοβού, εκτεινόμενη στην οπίσθια μοίρα της τρίτης μετωπιαίας έλικας (πεδία 44 και 45 κατά Brodmann - περιοχή Broca) (Σχήμα 2) [8]. Σε βαριές βλάβες καταστρέφονται και οι παρακείμενες προκινητική και προμετωπιαία περιοχή (πεδία 6,8,9,10 και 46). Οι αιτίες είναι κυρίως αγγειακές (απόφραξη άνω κύριου κλάδου μέσης εγκεφαλικής αρτηρίας), όγκοι, επιληπτικές κρίσεις κ.α. Σε ηπιότερη αφασία, στη λεγόμενη «mini Broca», η ένδεια της ομιλίας είναι ήπια, ενώ η αντίληψη και η ικανότητα γραφής παραμένει ακέραια, η μελωδικότητα χάνεται και οι λέξεις αρθρώνονται αργά. Επίσης, σε πιο ελαφριές καταστάσεις ή κατά την περίοδο της επανόδου, η ομιλία εμφανίζει βραδύτητα και δισταγμό, ποσοτική μείωση, αγραμματισμό (ομιλία σε τηλεγραφικό στυλ), κακή άρθρωση και δυσπροσωδία. Μάλιστα ο αγραμματισμός έχει επισημανθεί ως το κύριο χαρακτηριστικό της αφασίας αυτού του τύπου. Το αγραμματικό προφίλ του ασθενούς με αφασία Broca συνίσταται στη δημιουργία προτάσεων που αποτελούνται από δυο με τρεις λέξεις που ενώ ταιριάζουν μεταξύ τους σημασιολογικά δεν έχουν καμία συντακτική δομή [9]. Παράλληλα, οι ασθενείς αυτού του τύπου παράγουν συχνά φωνηματικές παραφασίες αντικαθιστώντας, προσθέτοντας ή παραλείποντας φθόγγους, ενώ σπανιότερα παρατηρούνται σημασιολογικές παραφασίες. Έτσι από τις απαντήσεις που δίνει ο ασθενής φαίνεται ότι χρησιμοποιεί τις σωστές λέξεις και ο συνομιλητής μπορεί να βγάλει νόημα, παρά το γεγονός ότι οι προτάσεις είναι κακοσχηματισμένες και απογυμνωμένες από συνθετικά στοιχεία και ρήματα.

Σε πιο βαριές αφασίες οι ασθενείς δεν μπορούν να μιλήσουν δυνατά, ενώ μπορεί να λείπει τελείως η ομιλία, χωρίς ωστόσο να παρατηρείται δυσκολία στην κατάποση και στην άρθρωση άλλων ήχων πλην των λέξεων. Ασθενείς με αφασία Broca εμφανίζουν τυπικά ομιλία αραιή, κατά την οποία παραλείπονται άρθρα, επίθετα, επιρρήματα (παραλείψεις), ενώ υπάρχει αποδόμηση της σύνταξης (τηλεγραφικός λόγος), χωρίς ωστόσο να παραλείπονται σημαντικά γνωρίσματα της γλώσσας. Οι ασθενείς με αφασία Broca δεν παρουσιάζουν σχεδόν καμία δυσκολία στην αντιληπτική τους ικανότητα και μπορούν να συμμετέχουν σε συζήτηση κατανοώντας μεμονωμένες λέξεις, αλλά δυσκολεύονται στην κατανόηση πολύπλοκων γραμματικών τύπων και συνήθως δε μπορούν να απαντήσουν.

Επίσης υπάρχει διαταραχή στην επανάληψη που είναι τόσο χαρακτηριστική, ώστε εάν ο ασθενής δεν την παρουσιάζει δεν τίθεται διάγνωση αφασίας Broca αλλά διαφλοιώδους κινητικής αφασίας.



Συνυπάρχει σοβαρή διαταραχή και της γραφής. Η κατανόηση προφορικού και γραπτού λόγου είναι ελάχιστα διαταραγμένη καθώς δεν έχει καταστραφεί η περιοχή Wernicke, ενώ η κατονομασία είναι εσφαλμένη. Οι ασθενείς έχουν επίγνωση των σφαλμάτων τους [10].

#### **1.4.2 Αφασία Wernicke**

Αυτός ο τύπος αφασίας χαρακτηρίζεται κυρίως από σημαντική ελάττωση της κατανόησης. Η βλάβη βρίσκεται στην περιοχή Wernicke ( πεδίο 22 ) (σχήμα 2) και επεκτείνεται και στην άνω περιοχή του κροταφικού λοβού (πεδία 39 και 40) καθώς και προς τα κάτω, στο πεδίο 37 (συμμετοχή ακουστικών συνειρμικών περιοχών ή αποσύνδεσή τους από τον πρωτογενή ακουστικό φλοιό -έλικες Heschl). Οι αιτίες περιλαμβάνουν απόφραξη του κατώτερου κλάδου της αριστερής μέσης εγκεφαλικής αρτηρίας, κατάληψη της περιοχής από όγκο ή απόστημα, αιμορραγία στην περιοχή κ.α. Ασθενείς με αφασία Wernicke εμφανίζουν ομιλία, φυσιολογική ως προς τον ρυθμό και τη μουσικότητα.

Οι ασθενείς μπορεί να δυσκολεύονται να βρουν την κατάλληλη λέξη ή να χρησιμοποιούν εσφαλμένη λέξη ή εσφαλμένους συνδυασμούς λέξεων ή να δημιουργούν νέες λέξεις. Εμφανίζονται δηλαδή παραφασίες και νεολογισμοί. Έτσι, παρά την ευφράδεια και την φυσιολογική προσωδία, η ομιλία στερείται νοήματος. Η λογόρροια, δηλαδή η υπεραφθονία λέξεων με λίγα νοήματα, και ο κενός λόγος συμπληρώνουν τη διαταραχή της ομιλίας. Διαταραχές σημειώνονται και στην ανάγνωση, στη γραφή, στην κατονομασία και στην επανάληψη. Οι ασθενείς δεν μπορούν να διαβάσουν, να γράψουν αυθόρμητα, να αντιληφθούν το νόημα, να γράψουν καθ' υπαγόρευση, ενώ μπορούν να αντιγράψουν αργά, και τέλος δεν μπορούν να κατονομάσουν. Οι ασθενείς αυτοί δεν καταλαβαίνουν τι τους λένε οι άλλοι ενώ δεν αντιλαμβάνονται την ανικανότητα αυτή.

Το κατεξοχήν σύμπτωμα της αφασίας αυτής θεωρείται ο παραγραμματισμός κατά τον οποίο παρουσιάζονται ταυτολογίες, ασάφειες, συχνές επαναλήψεις λέξεων και τμήματα προτάσεων που δε συνδέονται ούτε σημασιολογικά ούτε συντακτικά μεταξύ τους. Παρά το γεγονός ότι οι ασθενείς αυτοί δεν μπορούν να κατανοήσουν τα λόγια των άλλων και συχνά αναγκάζονται να μαντέψουν το νόημα τους, εξακολουθούν να διατηρούν την ικανότητα διεξαγωγής διαλόγου καθώς κατανοούν τότε τους απευθύνεται μια ερώτηση και τότε πρέπει να ζητήσουν το λόγο. Στη γραφή και στην ανάγνωση παρουσιάζουν διπλασιασμό των συλλαβών, των λέξεων ή ακόμη και των προτάσεων. Συνήθως συνυπάρχουν και άλλα εστιακά νευρολογικά σημεία όπως δεξιά ομώνυμη ημιανοψία.

#### **1.4.3 Σφαιρική αφασία (global or total)**

Σε αυτό τον τύπο αφασίας η βλάβη βρίσκεται σε μεγάλο τμήμα της γλωσσικής περιοχής (περιοχή Broca & Wernicke) (Σχήμα 3)[11]. Οι αιτίες περιλαμβάνουν απόφραξη αριστερής εσω κερωτιδας η έσης εγκεφαλικής αρτηρίας, αιμορραγία, όγκος στην περιοχή η εμφανίζεται παροδικά μετά απο επιληπτική κρίση. Η αφασία είναι σοβαρή καθώς όλες οι παράμετροι του λόγου επηρεάζονται: οι ασθενείς αρθρώνουν μόνο μερικές λέξεις ή συλλαβές (δεν παρουσιάζουν αλαλία), δεν μπορούν να διαβάσουν, να γράψουν ή να επαναλάβουν. Συνυπάρχουν δεξιά ημιπάρεση, ημιϋπαισθησία και ομώνυμη ημιανοψία [10].

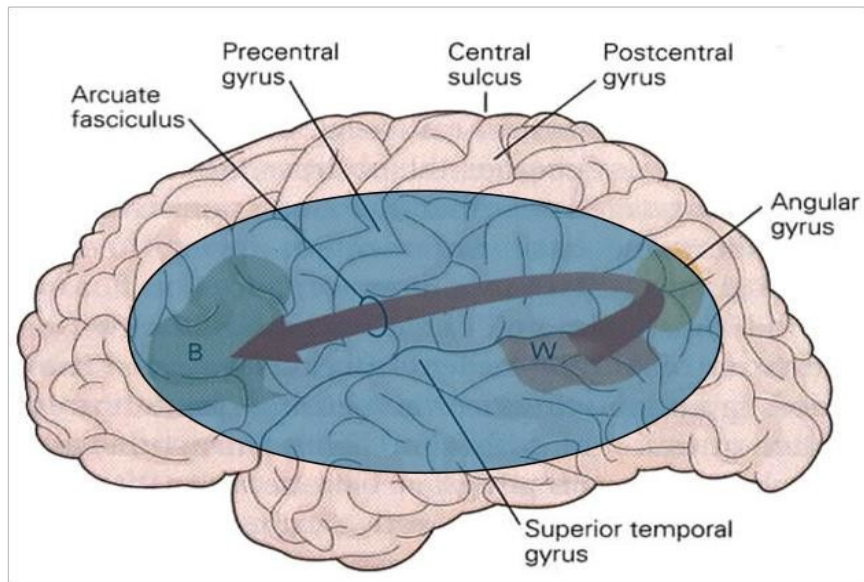


Figure 3 Καθολική Αφασία

#### 1.4.4 Σύνδρομο αποσύνδεσης

##### 4.α. Αφασία αγωγής

Η βλάβη βρίσκεται στην τοξοειδή δεσμίδα (Σχήμα 4) [11]. Τέτοιου είδους βλάβη παρατηρείται σε κάκωση της υπερχειλίας έλικας του αριστερού βρεγματικού λοβού ή (λιγότερο συχνά) σε κάκωση της οπίσθιας και άνω επιφάνειας του αριστερού κροταφικού λοβού. Κύριο αίτιο είναι η απόφραξη του ανιόντος βρεγματικού κλάδου της μέσης εγκεφαλικής αρτηρίας. Μοιάζει με την αφασία Wernicke ως προς την ευφράδεια, την παραφασία, την διαταραχή επανάληψης και την μεγαλόφωνη ανάγνωση και τη γραφή, ενώ η ευχέρεια στην παραγωγή γλώσσας είναι μειωμένη σε σχέση με τους ασθενείς με αφασία Wernicke. Η ικανότητα κατανόησης του λόγου παραμένει ανέπαφη, ενώ ο αυθόρμητος λόγος παρουσιάζει κανονική ροή και καλή άρθρωση αλλά περιορίζεται σε σύντομες και απλές προτάσεις. Η διχοτόμηση αυτή είναι το κυρίαρχο χαρακτηριστικό αυτής της μορφής αφασίας. Στους ασθενείς με αφασία αγωγής υπάρχει επίγνωση της διαταραχής και για αυτό προσπαθούν να μειώσουν τα πολλά φωνημικού τύπου παραφασικά λάθη [10].

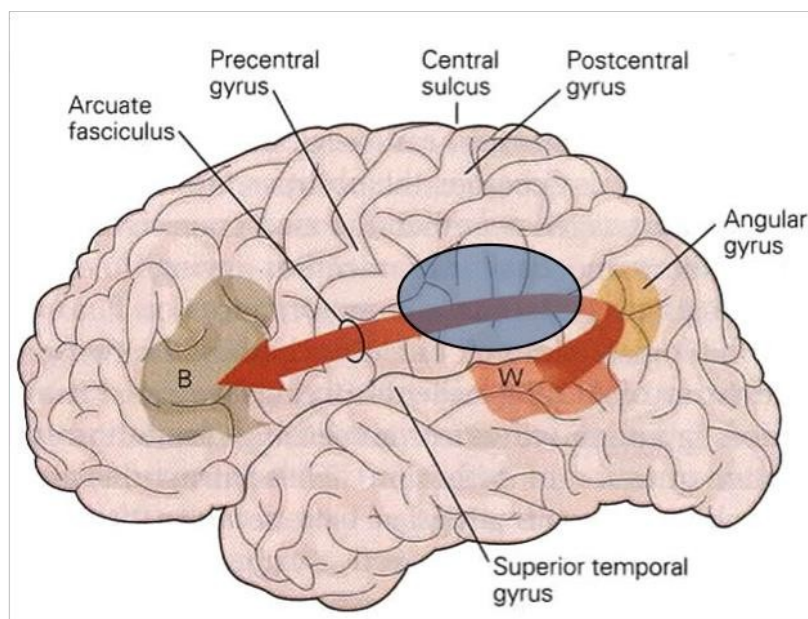


Figure 4 :Αφασία αγωγής

#### 4.β. Κατονομαστική αφασία (anomic)

Η βλάβη βρίσκεται στον οπίσθιο κροταφικό λοβό ή στη μέση κροταφική έλικα, εκεί που διακόπτονται οι συνδέσεις ανάμεσα στις αισθητηριακές γλωσσικές περιοχές και τις ιπποκάμπιες, που ενέχονται στη μνήμη και μάθηση, συνήθως λίγο πιο πίσω από την περιοχή Wernicke και σχετικά βαθιά μεταξύ της οπίσθιας μοίρας της άνω και κάτω κροταφικής έλικας και της γωνιάδους έλικας (Σχήμα 5) . Οι κύριες αιτίες είναι όγκοι της περιοχής, ερπητική εγκεφαλίτιδα, αποστήματα.

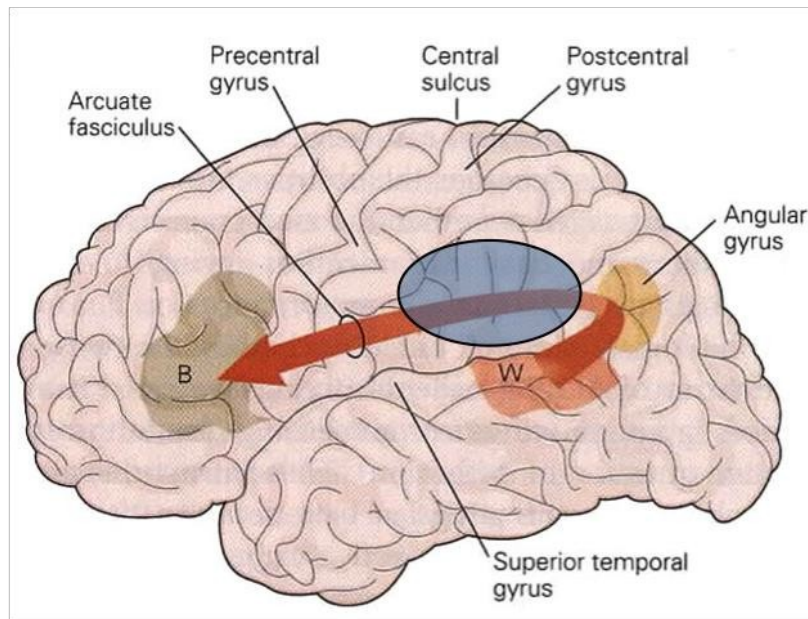


Figure 5 Ανομική Αφασία

Αναφέρεται και ως ανομία ή αμνησιακή αφασία. Ο ασθενής έχει δυσκολία μόνο στο να βρει τις σωστές λέξεις. Όταν του παρουσιαστεί ένα αντικείμενο προς κατονομασία δεν βρίσκει τη λέξη, αλλά μπορεί να το περιγράψει ή να δείξει τη χρήση του. Όταν του ζητηθεί εκλογή της λέξης που ταιριάζει με ένα αντικείμενο, αναγνωρίζει συνήθως ορθά μεταξύ των εναλλακτικών λέξεων τη σωστή για το αντικείμενο. Λόγω απώλειας της ικανότητας εύρεσης των κατάλληλων λέξεων, η ομιλία του ασθενή χαρακτηρίζεται από παύσεις, αναζήτηση λέξεων, περιφράσεις και υποκατάσταση, ενώ η αντίληψη του λόγου είναι φυσιολογική. Παρόμοια δυσκολία μπορεί να υπάρχει και στον γραπτό λόγο. Η σύνταξη και η γραμματική δεν εμφανίζουν ιδιαίτερα λάθη αν και παρατηρούνται παραφασίες φωνηματικές και συχνότερα σημασιολογικές που όμως δεν επηρεάζουν την καλή μορφή που έχει ο λόγος [12]. Η αφασία αυτή μπορεί να εμφανισθεί από την αρχή ή να ακολουθήσει, στο στάδιο βελτίωσης, την αφασία Wernicke. Οι ασθενείς είναι διστακτικοί και βραδείς όταν μιλούν με πλήρη επίγνωση της δυσκολίας τους. Συχνά χρησιμοποιούν μακριές φρασεολογίες για να πουν αυτά που θέλουν. Μια ειδική μορφή αυτής της διαταραχής είναι η απτική αφασία. Σε αυτήν οι ασθενείς δεν μπορούν να κατονομάσουν αντικείμενα που γίνονται αισθητά με τα χέρια, αλλά μπορούν να προβούν σε κατονομασία οπτικά παρουσιαζόμενων αντικειμένων [10].

#### **4.γ. Αμιγής λεκτική κώφωση**

Η βλάβη βρίσκεται σε αμφότερα τα ημισφαίρια, στο μέσο τριτημόριο άνω κροταφικών ελίκων, σε μια θέση όπου διακόπτονται οι συνδέσεις ανάμεσα στον πρωτογενή ακουστικό φλοιό και στις συνειρμικές περιοχές του άνω-οπισθίου τμήματος του κροταφικού φλοιού. Σπανίως, αμιγής λεκτική κώφωση παρατηρείται σε ετερόπλευρες βλάβες στο επικρατητικό ημισφαίριο. Η διαταραχή της ακουστικής αντίληψης, επανάληψης και της ικανότητας γραφής καθ' υπαγόρευση προεξάρχουν σε αυτή της μορφή αφασίας, ενώ η αυθόρμητη γραφή και η ικανότητα αντίληψης γραπτού λόγου διατηρούνται. Οι ασθενείς δίνουν την εντύπωση ότι είναι κωφοί. Ωστόσο, οι ακοομετρικές δοκιμασίες δεν αποκαλύπτεται ακουστικό έλλειμμα. Το σύνδρομο δεν είναι αμιγές καθώς μπορούν να βρεθούν άλλα στοιχεία αφασίας Wernicke ή να εμφανίζεται κατά τη βελτίωση αφασίας Wernicke.

#### **4.δ. Αμιγής λεκτική τύφλωση (οπτική λεκτική αγνωσία ή αλεξία χωρίς αγραφία)**

Προεξάρχουν η ανικανότητα ανάγνωσης, κατανόησης γραπτού λόγου και κατονομασίας χρωμάτων. Η ομιλία είναι τελείως φυσιολογική όπως και η αντίληψη του προφορικού λόγου, η επανάληψη όσων ακούγονται και η αυθόρμητη και καθ' υπαγόρευση γραφή. Η ικανότητα αντιγραφής επηρεάζεται (ο ασθενής πρέπει να αντιγράψει ένα-ένα τα γράμματα). Το αξιοσημείωτο της αμιγούς λεκτικής τύφλωσης είναι η διατήρηση ικανότητας ευχερούς γραφής, την οποία μετά ο ίδιος ασθενής δεν μπορεί να διαβάσει. Συνυπάρχει δεξιά ομώνυμη ημιανοψία. Η βλάβη βρίσκεται στον αριστερό οπτικό φλοιό και στην υποκείμενη λευκή ουσία και στις συνδέσεις του δεξιού οπτικού φλοιού με τις ακέραιες γλωσσικές περιοχές του επικρατούντος ημισφαιρίου. Ο διαχωρισμός στο σπληνίο του μεσολοβίου με επακόλουθη βλάβη του σπληνίου εμποδίζει τη μεταβίβαση των πληροφοριών από το αριστερό οπτικό ημιπεδίο στο δεξιό ημισφαίριο και από κει στην αριστερή γωνιώδη θλικά. Αν η βλάβη περιορίζεται στη γωνιώδη θλικά δεν θα υπάρχει ημιανοψία, αλλά η αλεξία μπορεί να συνδυάζεται με αγραφία, σύγχυση αριστερού-δεξιού, αναριθμησία και δακτυλική αγνωσία (σύνδρομο Gerstmann).

Να σημειώσουμε εδώ ότι η χειρουργική διατομή του σπληνίου συνεπάγεται απώλεια της ικανότητας ανάγνωσης στο αριστερό οπτικό ημιπεδίο και όχι στο δεξί [10].

#### **4.ε. Αμιγής λεκτική βωβότητα (pure word mutism, αμιγής κινητική αφασία του Dejerine)**

Η ικανότητα ομιλίας χάνεται ενώ η ικανότητα γραφής διατηρείται σ' αυτήν την αφασία. Ο προφορικός και γραπτός λόγος παραμένουν κατανοητοί. Υπάρχει βωβότητα με ανεπηρέαστη την εσωτερική ομιλία και την γραφή. Χαρακτηριστικά, η γλώσσα αποκαθίσταται σε αυτή την διαταραχή. Η βλάβη εντοπίζεται στον κυρίαρχο μετωπιαίο λοβό, χωρίς να έχει καθοριστεί πλήρως η ανατομική βάση. Αυτό που συμβαίνει είναι μια αποσύνδεση του κινητικού φλοιού της ομιλίας από τα κατώτερα κέντρα. Μπορεί να συνυπάρχει ημιπάρεση [10].

#### **4.στ. Διαφλοιώδεις αφασίες- απομόνωση περιοχών του λόγου**

Στις αιτίες των διαφλοιωδών αφasiών περιλαμβάνονται έμφρακτα στις ζώνες οριακής αιμάτωσης της πρόσθιας και μέσης εγκεφαλικής αρτηρίας, όπως συμβαίνει μετά από καρδιακή παύση. Οι ασθενείς με διαφλοιώδη αφασία δε μπορούν να γράψουν και να διαβάσουν με κατανόηση γιατί υπάρχει βαριά διαταραχή της οπτικής και ακουστικής αντίληψης των λέξεων. Είναι ωστόσο δυνατή η επανάληψη λέξεων οι οποίες όμως δε γίνονται κατανοητές. Η διαφλοιώδης αφασία χωρίζεται σε 3 κατηγορίες:

- 1) στην κινητική, κατά την οποία η γλωσσική παραγωγή είναι ιδιαίτερα διαταραγμένη, αλλά διατηρείται σε ικανοποιητικά επίπεδα η γλωσσική κατανόηση. Τέτοια αφασία συναντούμε

σε μετωπιαίες βλάβες και βλάβες των βασικών γαγγλίων

- 2) στην αισθητηριακή, κατά την οποία υπάρχει έλλειμμα στην γλωσσική αντίληψη, αλλά η γλωσσική παραγωγή επιτυγχάνεται έστω και με πληθώρα παραφρασιών. Τη συναντούμε σε μεταβολικές εγκεφαλοπάθειες και σε έσω μετωποβρεγματικές, καθώς και κροταφικές και θαλαμικές βλάβες
- 3) στη μικτή, κατά την οποία είναι σχεδόν αδύνατη τόσο η γλωσσική κατανόηση όσο και η γλωσσική παραγωγή αλλά παραμένει ανέπαφη η ικανότητα επανάληψης των λέξεων.

#### **4.στ.ι. Διαφλοιώδης αισθητική**

Στη διαφλοιώδη αισθητική αφασία υπάρχει έλλειμμα ακουστικής και λεκτικής αντίληψης. Η γραφή και η ανάγνωση καθίστανται αδύνατες. Η ομιλία είναι ευχερής με παραφρασίες, κατονομαστική αφασία, και κενό λόγο. Διαφοροποιείται από την αφασία αγωγής και την αφασία του Wernicke, καθώς στη διαφλοιώδη αισθητική αφασία διατηρείται η επανάληψη. Η βλάβη εντοπίζεται στην βρεγματο-ινιακή περιοχή με διατήρηση της τοξοειδούς δεσμίδας. Λόγω αυτής της βλάβης η αισθητηριακή πληροφορία δεν φτάνει στα κέντρα ολοκλήρωσης. Συνυπάρχει οπτική αγνωσία και ημιανοψία [10].

#### **4.στ.ii. Διαφλοιώδης κινητική (σύνδρομο πρόσθιας απομόνωσης, «δυναμική αφασία του Luria»)**

Σε αυτόν τον τύπο κινητικής αφασίας υπάρχει πλήρης αδυναμία διαλόγου, με ικανότητα παραγωγής μόνο ήχων και διατήρηση της κατανόησης του λόγου. Η διαφλοιώδης αισθητική αφασία πρέπει να διαφέρει από τη λεκτική αλαλία, καθώς στην πρώτη η επανάληψη διατηρείται ακέραια. Η διαταραχή εκδηλώνεται μετά από αποκατάσταση αφασίας Broca ή καταστάσεις αβουλίας και ακινητικής αλαλίας σε καταστροφή του μετωπιαίου λοβού [10].

### **1.4.5. Αγραφίες**

Στις αγραφίες η βλάβη βρίσκεται στην οπίσθια περισιλούια περιοχή, κοντά στη γωνιώδη έλικα. Η αμιγής αγραφία είναι σπάνια διαταραχή. Συνήθως συνυπάρχει με άλλες διαταραχές του λόγου. Υπάρχουν τρεις μορφές αγραφίας:

1. αφασικές αγραφίες, όπου αφθονούν τα ορθογραφικά και γραμματικά λάθη.
2. ανωμαλίες της χωρικής αντίληψης και πραξίας -κατασκευαστική αγραφία-, όπου οι λέξεις δεν τοποθετούνται σωστά στη σελίδα και συνυπάρχουν και άλλες απραξίες.
3. απραξικές αγραφίες, στις οποίες το χέρι έχει χάσει την ικανότητα διαμόρφωσης λέξεων (υπάρχει διαταραχή και στις άλλες χειρωνακτικές ικανότητες) [10].

### **1.4.6. Υποφλοιώδεις αφασίες (θαλαμικές και ραβδωτοκαψικές)**

Κλινικά μοιάζουν με τις αφασίες Broca & Wernicke. Στη θαλαμική αφασία η βλάβη είναι στον θάλαμο του επικρατούντος ημισφαιρίου. Η θαλαμική αφασία δεν έχει ενιαία κλινικά χαρακτηριστικά- μπορεί να εμφανιστεί ως αλαλία, ως διαταραχή της κατανόησης του λόγου και κατόπιν ως μειωμένη αυθόρμητη ομιλία. Η επανάληψη είναι φυσιολογική. Στη ραβδωτοκαψική αφασία η βλάβη βρίσκεται στην ραβδωτοκαψική περιοχή του επικρατούντος ημισφαιρίου με επέκταση προς την υποφλοιώδη λευκή ουσία του κροταφικού λοβού. Η αφασική διαταραχή

χαρακτηρίζεται από δυσarthρική, παραφασική ομιλία, με διαταραχές της κατανόησης, της κατονομασίας και της επανάληψης. Συνυπάρχει δεξιά ημιπάρεση. Η αποκατάσταση είναι βραδύτερη της θαλαμικής αφασίας [10].

## 1.5 Συμπτώματα -Διάγνωση

Τα πιο βασικά συμπτώματα είναι η δυσκολία στην ομιλία, προσπαθεί ο ασθενής να βρει την κατάλληλη λέξη για εκφράσει μια πρόταση. Επιπλέον χρησιμοποιεί περιέργες ή ακατάλληλες λέξεις για αυτό που θέλει να πει. Μερικοί άνθρωποι με αφασία έχουν πρόβλημα κατανόησης του τι λένε οι άλλοι. Η αφασία δεν επηρεάζει τις δεξιότητες της σκέψης όπως έχουμε ξαναφέρει, όμως το άτομο μπορεί να έχει προβλήματα κατανόησης του γραπτού λόγου και δυσκολία στη γραφή. Μερικοί άνθρωποι έχουν πρόβλημα με τη χρήση αριθμών ή ακόμα δυσκολεύονται να κάνουν απλούς υπολογισμούς.

Συνήθως αφασία αναγνωρίζεται για πρώτη φορά από τον ιατρό που θεραπεύει το άτομο που έχει υποστεί τραυματισμό. Αυτός συνήθως είναι ο νευρολόγος. Ο γιατρός πραγματοποιεί κάποιες δοκιμές στον ασθενή. Πιο συγκεκριμένα απαιτεί από το άτομο να ακολουθήσει τις εντολές του , να απαντήσει σε ερωτήσεις, να ονομάσει τα αντικείμενα, και να συμμετέχει σε μία συνομιλία. Εάν ο γιατρός διαγνώσει την αφασία, ο ασθενής απευθύνεται στον λογοθεραπευτή, ο οποίος εξετάζει τις δυνατότητες επικοινωνίας του ατόμου. Εκεί εξετάζεται πιο αναλυτικά η ικανότητα του ατόμου να μιλήσει, να εκφράσει τις ιδέες του, να κατανοήσει τη γλώσσα, να διαβάσει και να γράψει, καθώς και η ικανότητα να καταπιεί [13].

Για την διάγνωση της αφασίας ο γιατρός χρησιμοποιεί μια σειρά νευρολογικών και άλλων εξετάσεων για να προσδιορίσει τη ύπαρξη ή όχι αφασίας αλλά και τη σοβαρότητά της. Επίσης έχουν αναπτυχθεί διάφορα τεστ, τα πιο γνωστά και αποδοτικά είναι :

1. Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1972)[14]
2. Neurosensory Center Comprehensive Examination for Aphasia (Spreen and Benton, 1969) [15]
3. Western aphasia Battery (Kertesz, 1980)[16]
4. Aachener Aphasia Test (Huber et al., 1982)[17]

## 1.6 Η κατάσταση ενός αφασικού ασθενούς

Το πιο συνηθισμένο πρόβλημα που συνοδεύει την αφασία είναι η αδυναμία ή παράλυση συνολικά της αριστερής πλευράς του σώματος. Συχνά μπορεί να επηρεαστούν τα άκρα η ακόμα και το πρόσωπο της δεξιάς ή της αριστερής μεριάς του ασθενή. Με τον καιρό, συχνά, και το πόδι και το χέρι μπορούν να ξαναλειτουργήσουν, αλλά η αδυναμία στο χέρι μπορεί να παραμείνει πιο πολύ απ' ό,τι στο πόδι. Μερικοί αφασικοί ασθενείς έχουν ισχυρούς πονοκεφάλους ή παθαίνουν σπασμούς. Ορισμένοι έχουν δυσκολία στην όραση ως αποτέλεσμα της εγκεφαλικής βλάβης. Αυτή η δυσκολία παίρνει τη φόρμα της αδυναμίας να δει αντικείμενα προς τα δεξιά ή προς τα αριστερά και όχι ολική αδυναμία όρασης. Εξαιτίας της εγκεφαλικής βλάβης, ο αφασικός ασθενής έχει μικρό έλεγχο πάνω στα συναισθήματα του. Συχνά είναι «αδικαιολόγητα» εριστικός, γελά και κλαίει εύκολα, δρα παρορμητικά και ενίοτε χωρίς να λαμβάνει υπόψη του τις συνθήκες. Η διάθεση του μπορεί να αλλάζει εύκολα, χωρίς λόγο, από ώρα σε ώρα. Οι αφασικοί μπορεί ακόμη να παρουσιάζονται ληθαργικοί, αποτραβηγμένοι και απρόθυμοι να ενδιαφερθούν για τους εαυτούς τους. Συχνά παρουσιάζουν κατάθλιψη, αντιλαμβανόμενοι την κατάσταση τους. Η προσοχή τους είναι διασπασμένη και αδύναμη να κάνει αφηρημένες αναγωγές στο μέλλον` είναι συχνά επικεντρωμένη σε θέματα συγκεκριμένα, όπως στο φαί και στα φάρμακα. Ο αφασικός ασθενής είναι



εγκλωβισμένος μέσα στην αδυναμία του να επικοινωνήσει, μέσα στη φυσική του αναπηρία, μέσα στην έλλειψη ελέγχου πάνω στα συναισθήματα του.

## **1.7 Η Θεραπεία της αφασίας**

Η θεραπεία για κάποιον με αφασία εξαρτάται από παράγοντες όπως είναι η ηλικία, αιτία της εγκεφαλικής βλάβης, είδος της αφασίας, θέση και μέγεθος της βλάβης του εγκεφάλου.

Γενικά υπάρχουν δύο υποκατηγορίες θεραπειών κατά της αφασίας και κλινικοί γιατροί υιοθετούν και τις δυο :

Οι θεραπείες που βασίζονται στις δυσλειτουργίες (Impairment-based therapies) και αποσκοπούν στη βελτίωση των γλωσσικών λειτουργιών. Η θεραπεία αυτή αποτελείται από διαδικασίες που όλες τονώνουν τις ικανότητες της ακρόασης, ομιλίας, ανάγνωσης και γραφής.

Οι θεραπείες που βασίζονται στην επικοινωνία (Communication-based therapies) αυτές προορίζονται για την ενίσχυση της επικοινωνίας με κάθε μέσο. Αυτές οι θεραπείες συχνά αποτελούνται από περισσότερες φυσικές αλληλεπιδράσεις και προκλήσεις που αφορούν την πραγματική ζωή του ασθενούς.

Η επιλογή της θεραπείας εξαρτάται από τις ανάγκες και τις επιθυμίες του ατόμου και την κατάσταση του. Επίσης, η θεραπεία αλλάζει με την πάροδο του χρόνου καθώς το άτομο με αφασία βελτιώνεται.

### **1.7.1 Οι θεραπείες που βασίζονται στις δυσλειτουργίες**

Η αρχική επιθυμία του ασθενή με την αφασία είναι να μιλήσει καλύτερα και να κατανοήσει τι λένε οι άνθρωποι που είναι γύρω από αυτόν. Για αυτό και οι λογοθεραπευτές προσπαθούν να αποκαταστήσουν την βλάβη που υπάρχει. Οι θεραπείες επικεντρώνονται στις ενέργειες που επιτρέπουν σε αυτόν να κατανοήσει και να μιλήσει με την ίδια επιτυχία. Επίσης στον ασθενή διατίθεται και υλικό υποστήριξης με ασκήσεις και εφαρμογές για τον υπολογιστή που τον βοηθούν στην θεραπεία. Οι εφαρμογές για τον υπολογιστή έχουν σχεδιαστεί με τέτοιο τρόπο ώστε ο ασθενής να εξασκεί την εύρεση των λέξεων, την κατανόηση και προβλήματα που μπορεί να αντιμετωπίσει καθημερινά π.χ μια αγορά. Επίσης στην θεραπεία περιλαμβάνεται ένας εικονικός γιατρός που μιλάει στον ασθενή μέσω του υπολογιστή. Οι επιστήμονες έχουν αναπτύξει θεραπείες που αφορούν συγκεκριμένο κομμάτι της δυσλειτουργίας του λόγου. Αυτό μπορεί να είναι η εξαγωγή των λέξεων για την ανάπτυξη μιας πρότασης.

### **1.7.2 Οι θεραπείες που βασίζονται στην επικοινωνία**

Η αποκατάσταση του λόγου σε έναν αφασικό ασθενή μπορεί να είναι μόνιμη και για αυτό ο ασθενής πρέπει να βρει εναλλακτικό τρόπο για την μετάδοση των μηνυμάτων και των συναισθημάτων στο περιβάλλον του. Γίνεται χρήση στρατηγικών οι οποίες βοηθούν τον ασθενή στην πιο εύκολη αντιμετώπιση των συμπτωμάτων αφασίας που αντιμετωπίζει. Ο λογοθεραπευτής προγραμματίζει τις ενέργειες του ασθενούς για να τον διευκολύνει να συμμετέχει στην καθημερινή

ζωή .

### **Παραδείγματα θεραπειών**

Οι θεραπείες που βασίζονται στις δυσλειτουργίες

Constraint-induced therapy (CIT)- αυτές οι θεραπείες αφορούν την χρήση του μέρους του σώματος που είναι παράλυτο, σε μια προσπάθεια να επαναλειτουργήσει σωστά.

Melodic Intonation Therapy (MIT)-οι θεραπείες αυτές αφορούν τον κυματισμό της φωνής , σύμφωνα με έρευνες ο ασθενής πιο εύκολα κυματίζει την λέξη, από το να την εξάγει τελείως.

Tele-rehabilitation- Η απομακρυσμένη θεραπεία είναι μια μεθοδολογία που γίνεται μεταξύ του ασθενούς και του θεραπευτή από απόσταση μέσω του υπολογιστή. Υπάρχει κάμερα που επιτρέπει την μεταξύ τους επικοινωνία .Ο τρόπος αυτός της θεραπεία δεν εφαρμόζεται πολύ συχνά.

Οι θεραπείες που βασίζονται στην επικοινωνία

PACE therapy (Promoting Aphasics' Communicative Effectiveness)- Αυτή η μεθοδολογία είναι ευρέως χρησιμοποιημένη στην Ευρώπη και περιλαμβάνει την αναπαράσταση του λόγου σε εικόνες.

Conversational coaching- Εξάσκηση της συζήτησης είναι η πιο σίγουρη μεθοδολογία θεραπείας στην βελτίωση της ομιλίας του ασθενούς.

Supported conversation- Είναι η μεθοδολογία θεραπείας που ο ασθενής κάνει συζητήσεις με ομάδες εθελοντών που έχουν εκπαιδευτεί για τον βοηθήσουν στην βελτίωση της ομιλίας [19].

Ο ασθενής χρειάζεται την οικογένεια στο πλευρό του που έχει μια μοναδική ευκαιρία να τον βοηθήσει να βγει από τον εγκλωβισμό του και να ζήσει ξανά, όσο γίνεται πιο αρμονικά με το περιβάλλον του. Βέβαια, η οικογένεια είναι φυσικό να νιώθει στενοχώρια, υπερβολικό άγχος, κούραση, ίσως και απόρριψη. Ο ασθενής ωστόσο δεν πρέπει να λαμβάνει αυτά τα μηνύματα. Η αποκατάσταση είναι μια συνεχής διαδικασία προόδου. Μετά την πρώτη περίοδο ανάρρωσης, οι αλλαγές έρχονται μετά από δουλειά και προσπάθεια. Η εξάσκηση μπορεί να δημιουργήσει «μικρά θαύματα» σε περιστατικά αφασίας που ίσως φαίνονταν απελπίδα. Χρειάζεται ψυχραιμία, κουράγιο, επίμονη και συστηματική εξάσκηση. Όσον αφορά στην εξέλιξη του ασθενούς πρέπει να λάβουμε υπόψη μας τα εξής: Πόσο μπορεί ο ασθενής να επικοινωνήσει (ομιλία, κατανόηση, γράψιμο, διάβασμα, μαθηματικές έννοιες) καθημερινά.

Η οικογένεια πρέπει με σαφήνεια να ενημερωθεί για το ακριβή κατάσταση του ασθενούς και η θεραπεία να αρχίσει από αυτό το σημείο.

Αφιέρωση χρόνου, στον ασθενή όταν είναι διαυγής και συνεργάσιμος, με σύστημα, συνέπεια και ψυχραιμία.

Αποδοχή του ασθενούς όπως είναι (με καλό ή κακό επίπεδο επικοινωνίας, για το πότε, πόσο, πως μπορεί να συνεργαστεί και εξασκηθεί).

Δημιουργία μιας λίστας ενδιαφερόντων του ασθενούς και χρήση υλικών που του αρέσουν για την εξάσκηση της επικοινωνίας του.

Σταδιακή ανεξαρτησία του ασθενούς με κάθε ευκαιρία

Συμμετοχή του αφασικού ασθενούς, κατά το δυνατό, στις δραστηριότητες και ασχολίες

της οικογένειας.[18]

## ΚΕΦΑΛΑΙΟ 2 - ΔΕΔΟΜΕΝΑ

### 2.2 Aachen Aphasia Test (AAT)

Γενικά η διάγνωση του τύπου της αφασίας γίνεται από τον γιατρό όταν αρχίζει μια ελεύθερη ομιλία με τον ασθενή. Μετά την εκτίμηση της διάγνωσης του γιατρού ακολουθεί η θεραπεία του ασθενή. Για τον προσδιορισμό της διάγνωσης πρέπει να ονομάσουμε τα κλινικά συμπτώματα του ασθενή για την ενίσχυση της κατηγοριοποίησης της αφασίας. Παρακάτω θα δούμε τις βασικές ορολογίες που χρησιμοποιούνται για να περιγράψουν κάποιο σύμπτωμα.

Πίνακας 1 : Ιατρικές ορολογίες που σχετίζονται με την ικανότητα του λόγου[1]

Συμπτώματα σχετικά με τον λόγο	Επεξήγηση
Λεκτικές (φωνητικές) παραφασίες	Λέξεις λέγονται λάθος
Προφορικές (σημασιολογικές) παραφασίες	Λάθος ή ανεπαρκείς λέξεις
Νεολογισμοί	Λέξεις που δεν υπάρχουν.
Ανομία	Δυσκολία στην εύρεση της λέξης
Αγραμματισμός	Συντακτικά ατελείς προτάσεις, με ύφος σύντομου μηνύματος
Στερεότυπα	Σύνολο επαναλαμβανόμενων φράσεων.
Δυσαρθρία	Διαταραχή στην άρθρωση του λόγου.
Δυσπροσωδία	Διαταραχή της μελωδίας η του ρυθμού της φωνής.
Αγραφία	Διαταραχή της ανάγνωσης.
Αλεξία	Διαταραχή της γραφής.

Το τεστ AAT περιλαμβάνει 6 υποκατηγορίες τεστ που αφορούν την αυθόρμητη ομιλία , τεστ συμβόλων, την επανάληψη, τον γραπτό λόγο, την confrontation naming, και τέλος την κατανόηση [20].

#### 2.2.1 Αυθόρμητη ομιλία (Spontaneous Speech)

Το πρώτο μέρος της εξέτασης AAT περιλαμβάνει την εκτίμηση της αυθόρμητης ομιλίας. Έξι υποκατηγορίες του τεστ χρησιμοποιούνται για να χαρακτηρίσουν έξι διαφορετικά επίπεδα της αυθόρμητης ομιλίας. Τα είδη της αφασίας παρουσιάζουν διαφορετικά χαρακτηριστικά όσο αφορά την βαθμολογία.. Η αυθόρμητη ομιλία αποτελεί έναν γρήγορο τρόπο διάγνωσης της μορφής της αφασίας. Οι έξι υποκατηγορίες (από 0 έως 5 βαθμούς ) είναι:

P0: επικοινωνιακή συμπεριφορά (communicative behaviour)

P1: άρθρωση και προσωδία (μελωδία του λόγου) (articulation and prosody (melody of speech))

P2: αυτοματοποιημένος λόγος (π.χ στερεότυπα, αυτοματισμοί)(automatized language (e.g. stereotypes, automatisms))

P3: σημασιολογική δομή (π.χ. λεκτικές παραφασίες, δυσκολίες ανάκτησης λέξης)( semantic structure (e.g. verbal paraphasias, word retrieval difficulties) )

P4: δομή της φωνής (π.χ. λεκτικές παραφασίες, νεολογισμοί)(phonologic structure (e.g. literal paraphasias, neologisms))

P5: συντακτική δομή (δομή των προτάσεων, γραμματικής)( P5: syntactic structure (structure of sentences, grammar)).

### **2.2.2 Τεστ Συμβόλων (Token Test)**

Το τεστ συμβόλων (T0, παίρνει τιμές από 0 έως 100) είναι ένα γενικό τεστ για την κατανόηση του λόγου. Ο ασθενής πρέπει να διαλέξει το σωστό σύμβολο από ένα μεγάλο σύνολο συμβόλων που διαφέρουν σε σχήμα, χρώμα και μέγεθος. Το τεστ σύμβολων περιλαμβάνει τεστ με πέντε υποκατηγορίες. Τα επίπεδα δυσκολίας αυξάνονται από T1 έως T5, και βαθμολογούνται με τιμές από 0-10.

### **2.2.3 Επανάληψη (Repetition)**

Το τρίτο τεστ είναι τεστ επανάληψης (N0, παίρνει τιμές από 0 έως 100). Στο τεστ αυτό ο ασθενής πρέπει να επαναλαμβάνει διάφορα είδη ήχων, λέξεων και προτάσεων. Το τεστ αυτό αποτελείται από τεστ πέντε κατηγοριών (με βαθμούς από 0 έως 30).

N1: μονά φωνήματα

N2: μονοσύλλαβα ουσιαστικά

N3: ξένες λέξεις

N4: σύνθετες λέξεις

N5: προτάσεις

### **2.2.4 Γραπτός λόγος (Writing Language)**

Το τεστ του γραπτού λόγου (C0, παίρνει τιμές από 0 έως 100) και είναι η εκτίμηση της γραφής και της ανάγνωσης. Αποτελείται από τεστ 3 κατηγοριών (με βαθμούς από 0 έως 30).

C1: προφορική ανάγνωση

C2: επιλογή / συνδυασμός στην ορθογραφία

C3: ορθογραφία

### **2.2.5 Αντιπαράθεση Ονομασίας (Confrontation Naming)**

Η αντιπαράθεση ονομασίας (B0, παίρνει τιμές από 0 έως 100) είναι η εκτίμηση της ικανότητας του ασθενή για να περιγράψει αντικείμενα, καταστάσεις, ενέργειες με κατάλληλες λέξεις. Αποτελείται από τεστ τεσσάρων κατηγοριών (με βαθμούς από 0 έως 30).

B1: ουσιαστικά

B2: χρώματα

B3: σύνθετα ουσιαστικά

B4: προτάσεις

### **2.2.6 Κατανόηση του λόγου**

Το τεστ της κατανόησης του λόγου (V0, παίρνει τιμές από 0 έως 100) προσδιορίζει την πιθανότητα κατανόησης των λέξεων και των προτάσεων από τον ασθενή επακριβώς. Αποτελείται από δύο τεστ που αποτιμούν την επεξεργασία της ακοής και της ανάγνωσης των λέξεων ή των προτάσεων (με βαθμούς από 0 έως 60).

V1: ακουστικό για τις λέξεις και για τις προτάσεις

V2: ανάγνωση των λέξεων και προτάσεων

### 2.3 Η Βάση Δεδομένων της Αφασίας στο Διαδίκτυο - Καταγραφή Δεδομένων

Τα προφίλ του AAT των αφασικών ασθενών είχαν καταγραφεί ήδη από το 1986 με συμμετοχή 256 ασθενών. Η βάση δεδομένων της AAT περιλαμβάνει ονομαστικά δεδομένα όπως τον τύπο της διάγνωση της αφασίας, την ασθένεια, τοπολογικά δεδομένα κ.α. και τις βαθμολογίες AAT (Πίνακας 2). Η διάγνωση της αφασίας προσδιορίστηκε από τον ειδικό γιατρό. Δεδομένα όπως όνομα, ηλικία και το φύλο των ασθενών δεν αναγράφονται στη βάση για λόγους προστασία των ευαίσθητων δεδομένων του κάθε ασθενή. Όλοι οι ασθενείς εξετάστηκαν την στιγμή όταν η αφασία και η εγκεφαλική βλάβη είχαν σταθεροποιηθεί. Τα πρώτα συμπτώματα καθώς και το μέγεθος της βλάβης μπορεί να αλλάξουν μετά την εγκεφαλική διαταραχή, επειδή τμήματα του διαταραγμένου ιστού μπορεί να ανακτηθούν ή μπορεί να είναι ανεπανόρθωτα κατεστραμμένα. Επιπλέον στο συγκεκριμένο τεστ αναλύθηκαν τα προφίλ της ανατομικής κάκωσης 146 ασθενών με αξονική τομογραφία. Τα όρια των κακώσεων είχαν καταχωρηθεί ψηφιακά και οι 3D συντεταγμένες των μεμονωμένων βλαβών μεταφέρθηκαν σε ένα στατιστικό τρισδιάστατο εγκεφαλικό μοντέλο (το Aachen Voxel Model -AVM). Το 3D AVM μοντέλο βασίστηκε σε 80 καλά άθικτα εγκεφαλικά πτώματα. Το σύστημα συντονισμού αναφοράς είναι η ACPC - γραμμή (σχήμα παρακάτω), η οποία είναι η σύνδεση μεταξύ της πρόσθιας και της οπίσθιας συνένωσης τμημάτων. Με αυτό τον τρόπο τα δεδομένα της βλάβης γίνονται ανεξάρτητα από το επιμέρους μέγεθος του εγκεφάλου και των τεχνικών ρυθμίσεων του συστήματος λήψης εικόνας (π.χ. θέση και διάμετρο CT-μέρους, γωνία CT). Τα προφίλ των κακώσεων συγκρίθηκαν με αυτά τα προφίλ των εγκεφάλων των υγιών ανθρώπων [20].

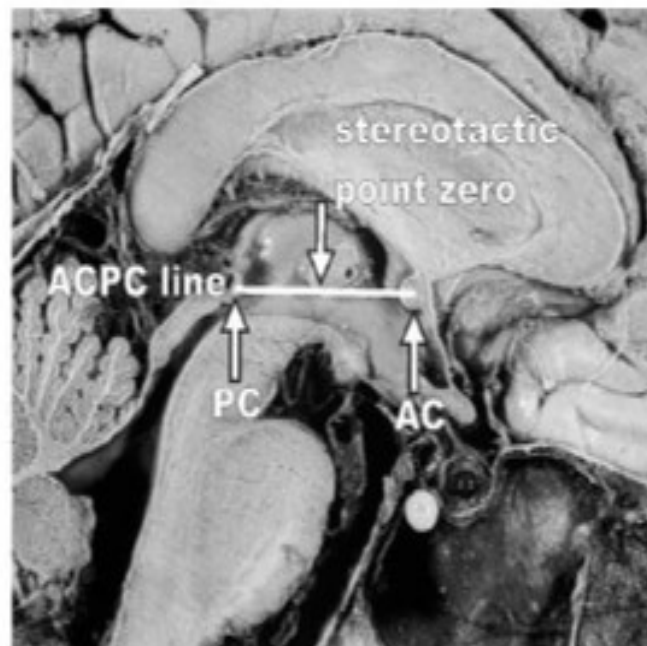


Figure 1: Η όψη της διχοτόμησης του νεκρού εγκεφάλου. Η γραμμή ACPC ορίζεται ως ενδοεπικοινωνιακή γραμμή που συνδέει τα κέντρα της πρόσθιας συνδέσμου (AC) και οπίσθιας συνδέσμου (PC)[1]

Η βάση δεδομένων της αφασίας βρίσκεται στην διεύθυνση <http://fuzzy.iau.dtu.dk/aphasia.nsf> και αποτελεί εργαλείο εκμάθησης για την κατηγοριοποίηση της αφασίας με την χρήση υπολογιστικής νοημοσύνης.. Τα ποικίλα αποτελέσματα διάγνωσης της βάσης φαίνονται στον Πίνακα 2.



Πίνακας 2: Δεδομένα που υπάρχουν στη βάση

<b>Ονομαστικά Δεδομένα (Nominal)</b>	Κατηγοριοποίηση του είδους της αφασίας (από τον ειδικό) (Classification of aphasia (made by an expert)) Ασθένεια (που προκάλεσε την αφασία) (Disease (cause of the aphasia)) Αριστερόχειρας/Δεξιόχειρας (Right/left handedness) Συμπτώματα (Symptoms) Κινητική ανεπάρκεια (Motor impairment) Ειδικοί που έκαναν τη διάγνωση (Experts, who made the diagnosis)
<b>Τοπολογικά (Topological)</b>	ΑΑΤ βαθμολογίες (AAT scores)
<b>Εικονικά Δεδομένα (Image Data Sets)</b>	Τρισδιάστατα προφίλ βλάβης (Three-dimensional lesion profiles)

Τα διαφορετικά συμπτώματα της αφασίας περιγράφονται παρακάτω Πίνακας 3:

Διάγνωση	N	%
Αφασία Wernicke	47	17.74
Αφασία Broca	42	15.85
Ολική αφασία	33	12.45
Κατονομαστική αφασία	24	9.06
Αφασία αγωγής	19	7.17
Αδιευκρίνιστη	51	19.25
Διαφλοιώδης αφασία	6	2.26
Υπολλειπτική αφασία	24	9.06
Όχι αφασία	19	7.17
Σύνολο	256	100

## 2.4 Το πρόβλημα της Ταξινόμησης της Αφασίας

Η ταξινόμηση της αφασίας βασίζεται στο ανατομικό μοντέλο του Wernicke και Lichtheim [22]. Παρ' όλα αυτά, τα σύνδρομα είναι αυστηρά εμπειρικά. Η κλασική ταξινόμηση της αφασίας είναι πολυτυπική (polytypic) που σημαίνει ότι κάθε διαταραχή μπορεί να είναι μέρος περισσότερων συμπτωμάτων. Επίσης οι προσδιορισμοί των συνδρόμων είναι πιθανολογικοί και δεν ορίζονται επακριβώς. Τα σύνδρομα δεν μπορούν να θεωρηθούν ούτε ως ομοιογενείς ομάδες, ούτε ως λειτουργικές οντότητες. Τέλος σύνδρομα μπορεί να είναι αποτελέσματα που παράγονται από την αγγείωση των περιοχών που σχετίζονται με τη γλώσσα. Τα χαρακτηριστικά σύνδρομα που ορίζονται στην κλασική ταξινόμηση (Broca, Wernicke, Καθολική Κατονομαστική) μπορούν να βρεθούν μόνο στο 60-75% όλων των αφασικών ασθενών. Περίπου το 19% των περιπτώσεων στη βάση δεδομένων δεν είχε χαρακτηριστεί ως κάποια από την κατηγορία της κλασικής αφασίας [20].

## 2.5 Σύστημα για το Soft Computing στην Ιατρική η Βάση Δεδομένων της Αφασίας

Γενικά μπορούμε να πούμε ότι η ταξινόμηση του είδους της αφασίας είναι αμφιλεγόμενη. Το Soft Computing μπορεί να διευκολύνει την ταξινόμηση των αλγορίθμων και να αποδώσει μεγάλη προσφορά στα ιατρικά ζητήματα. Επιπλέον, η επικοινωνία μεταξύ των ιατρών και μηχανικών υπολογιστών μπορεί να οδηγήσει σε ένα διεπιστημονικό πλεονέκτημα στην ανάλυση των αντιφάσεων σε ιατρικές ταξινομήσεις. Η βάση δεδομένων της αφασίας μπορεί να χρησιμοποιηθεί ως πρότυπο για τη σύγκριση διαφορετικών μεθόδων του Soft Computing [20].

Ιδανικά στοχεύουμε να φτιάξουμε ένα σύστημα το οποίο μπορεί να μάθει από την βάση δεδομένων της αφασίας και να διαγνώσει σωστά την ασθένεια. Στο παρακάτω σχήμα βλέπουμε μια σύντομη αναπαράσταση ενός τέτοιου συστήματος. Οι είσοδοι θα είναι βαθμολογίες από το AAT και η έξοδος θα είναι η διάγνωση (π.χ Αφασία Broca, Wernicke κτλ.)[23].

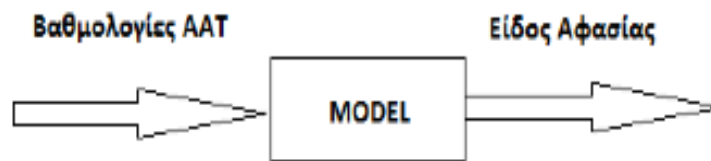


Figure 2: Το επιθυμητό σύστημα που αποφασίζει τον σωστό τύπο της αφασίας[4].

## 2.6 Βιβλιογραφική επισκόπηση

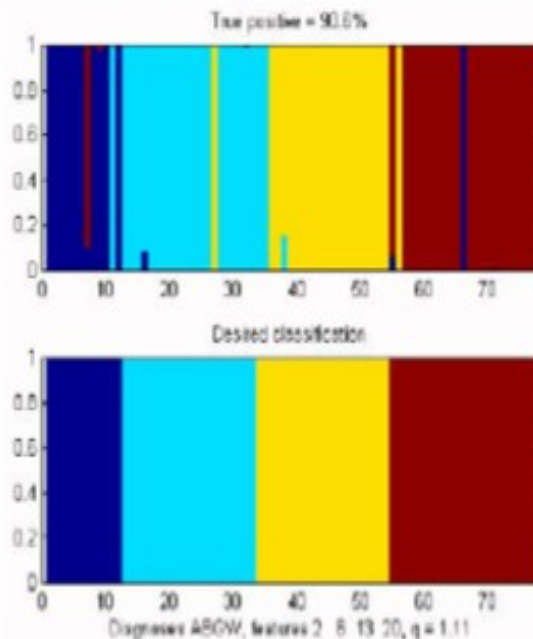
Διάφορες μεθοδολογίες έχουν εφαρμοστεί για την ταξινόμηση προφίλ ασθενών σε τύπους αφασίας. Ο γενετικός προγραμματισμός, τα νευρωνικά δίκτυα και η ασαφής λογική είναι μερικές από αυτές.

### 2.6.1 Ταξινόμηση του Είδους της Αφασίας με Χρήση των Νευρωνικών Δικτύων.

Για αρχή μπορούμε να αναφέρουμε περιληπτικά ότι η έρευνα [24] Axer, Hubertus, et al. "Aphasia classification using neural networks." European Symposium on Intelligent Techniques Aachen. Vol. 111. 2000 που είχε πραγματοποιηθεί για την ταξινόμηση της αφασίας με χρήση των νευρωνικών δικτύων χρησιμοποιούσε δύο πολυεπίπεδους αισθητήρες (multilayer perceptrons). Πρώτα πραγματοποιήθηκε η ταξινόμηση χρησιμοποιώντας την αξιολόγηση της αυθόρμητης ομιλίας του ασθενούς. Ο συγκεκριμένος ταξινομητής (classifier) εξήγαγε σωστά αποτελέσματα κατά 87% των περιπτώσεων δοκιμής (test cases). Για το δεύτερο τεστ χρησιμοποιήθηκαν εργαλεία ανάλυσης δεδομένων για να επιλέξουν τέσσερα χαρακτηριστικά ανάμεσα από 30 διαθέσιμα με σκοπό την αποδοτικότερη διάγνωση. Ο δεύτερος ταξινομητής εξήγαγε αποτελέσματα σωστά αποτελέσματα κατά 92% των περιπτώσεων. Το τεστ αυτό απαιτούσε τέσσερις AAT βαθμολογίες ως είσοδο του συστήματος για τον πολυεπίπεδο αισθητήρα. Στην πράξη το δεύτερο τεστ ήταν πιο αργό από το πρώτο.

Η επιλογή της χρήσης των νευρωνικών δικτύων δεν έγινε τυχαία, αλλά επειδή η χρήση τους δεν προϋποθέτει τον προσδιορισμό των κανόνων (rules), ούτε τις περιγραφές του συνδρόμου. Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks) είναι ικανά να δώσουν αποτελέσματα βρίσκοντας αυτόματα πρότυπα.

### 2.6.2 Ταξινόμηση του Είδους της Αφασίας με Χρήση Ασαφούς Λογικής



Στη μελέτη [26] οι Axer, H., et al. αναφέρουν πως αρχικά ο στόχος ήταν ο εντοπισμός και η σύγκριση περιπτώσεων που ήταν παρόμοιες στη βάση δεδομένων. Υποθέτοντας ότι οι ασθενείς με παρόμοιες διαγνώσεις λαμβάνουν παρόμοιες AAT βαθμολογίες, έγινε η δημιουργία ομάδων και η απλή αναπαράσταση τους σε γραφικές παραστάσεις, έτσι ώστε να συγκριθούν οι βαθμολογίες. Για να συγκριθούν τα δεδομένα, αυτά είχαν κλιμακωθεί με τη χρήση των ορίων τους, για να είναι όλοι οι αριθμοί στο διάστημα  $[0, 1]$ .

Η Ασαφής Λογική είναι ένας τρόπος προσαρμοσης ορισμένων μελών από διάφορες κατηγορίες ταυτόχρονα. Αυτό είναι κατάλληλο για την περίπτωση της κατηγοριοποίησης της αφασίας, λόγω του ακαθόριστης φύσης του προβλήματος. Ο αλγόριθμος ασαφών c-μέσων (fuzzy c-means- FCM) παρέχει μια ταξινόμηση ανάλογα με βαθμό συμμετοχής (degrees of membership). Ο αλγόριθμος εφαρμόζεται συνήθως για να βρεθούν συστάδες (ομάδες) (clusters) των δεδομένων, δηλαδή, δίνεται ένας αριθμός των συστάδων, και θα βρεθούν τα κέντρα της συστάδας και η συμμετοχή του κάθε αντικείμενου σε κάθε κατηγορία.

Στην συγκεκριμένη περίπτωση οι συστάδες έχουν ήδη δοθεί. Οι εγγραφές δεδομένων εκπαίδευσης περιέχουν τη σωστή διάγνωση, έτσι ώστε οι μέσες τιμές των βαθμολογιών εντός μιας κατηγορίας να σχηματίσουν το κέντρο της συστάδας. Ένας απλός ταξινομητής τότε θα αναθέσει κάθε νέο αντικείμενο, από τα δεδομένα των δοκιμών στο πλησιέστερο κέντρο της συστάδας[26]

Με την εκπαίδευση και την δοκιμή των δεδομένων η ακρίβεια είναι 91%, σχεδόν το ίδιο με τον ταξινομητή του νευρωνικό δικτύου. Η παράμετρος ασάφειας είναι  $q = 1,1$ , η οποία είναι ισοδύναμη με μια στενή γειτονιά, γιατί προφανώς οι γείτονες επηρέασαν τον ταξινομητή. Αυτό εξηγεί επίσης γιατί ο ταξινομητής αναθέτει μια διάγνωση μόνο στην πλειοψηφία των περιπτώσεων, και όχι τον αναμενόμενο συνδυασμό των διαγνώσεων.

### 2.6.3 Γενετικός Προγραμματισμός

Ο γενετικός προγραμματισμός (genetic programming-GP) αποτελείται από σύνολα διαφόρων συναρτήσεων, και έχει αποδειχθεί ότι είναι ικανός στην εύρεση βέλτιστων λύσεων για συγκεκριμένο χρονικό διάστημα για διάφορα ζητήματα της ταξινόμησης [27]. Ο γενετικός προγραμματισμός είναι μια επέκταση των γενετικών αλγορίθμων (genetic algorithms -GA), όπου το κύριο πρόβλημα τους που σχετίζεται με τον ορισμό των σταθερών προβλημάτων, αποφεύγεται με τη χρήση δέντρων μεταβλητού μήκους αντί του σταθερού. Επιπλέον, η θεωρία του GP επέτρεψε τη χρήση της λειτουργίας των κόμβων των δέντρων που αποτελεί ένα ισχυρό εργαλείο, όπως είναι για το πρόβλημα της επίλυσης της συμβολική παλινδρόμησης (symbolic regression)[27].

### 2.6.3.1 Απλός γενετικός προγραμματισμός (Standart Genetic Programming)

Κατά την εφαρμογή του απλού γενικού προγραμματισμού σχηματίστηκαν δέντρα απόφασης συγκεκριμένου μεγέθους σε συνδυασμό με διάφορες λειτουργίες/συναρτήσεις με τέτοιο τρόπο ώστε η έξοδος να αντιπροσωπεύει με ακρίβεια όλα τα δεδομένα δοκιμής της αφασίας που χρησιμοποιήθηκαν στον απλό γενετικό προγραμματισμό. Τα δεδομένα ήταν από 146 περιπτώσεις, και περιλάμβαναν τις τέσσερις κατηγορίες της αφασίας .Αυτές οι κατηγορίες ήταν Κατονομαστική Αφασία, Αφασία Broca, Ολική Αφασία και Αφασία Wernicke πιο συγκεκριμένα ήταν:

Κλάση 1:Κατονομαστική αφασία (καταχωρήσεις 1-24)

Κλάση 2: Αφασία Broca ( καταχωρήσεις 25-66)

Κλάση 3: Ολική Αφασία (καταχωρήσεις 67-99)

Κλάση 4: Αφασία Wernicke (καταχωρήσεις 100-146)

Τα δεδομένα για την εκπαίδευση και την δοκιμή ήταν περίπου ίδια 74 και 72 αντίστοιχα. Ο σκοπός ήταν η δημιουργία τριών κανόνων σε μαθηματική έκφραση, οι εκφράσεις αυτές αναφέρονται παρακάτω:

Κανόνας #1 θα εξάγει την Κατονομαστική Αφασία (κλάση 1) από τις δύο υπόλοιπες κατηγορίες (κλάση 2 - κλάση 4 ).

Κανόνας #2 θα εξάγει την Αφασία Broca (κλάση 2 ) από την Ολική Αφασία (κλάση 3) και την Αφασία Wernicke (κλάση 4).

Κανόνας #3 θα εξάγει την Ολική Αφασία (κλάση 3 ) από την Αφασία Wernicke (κλάση 4) [28][29].

Ο γιατρός με την εφαρμογή των κανόνων ξεκινώντας από τον Κανόνα #1, μέχρι να βρεθεί θετικό αποτέλεσμα μπορεί θεωρητικά να αποφασίσει για το πρόβλημα του ασθενή. Σε περίπτωση που δεν βρεθεί θετικό αποτέλεσμα η περίπτωση του ασθενούς τότε θεωρείται ότι ανήκει στην κλάση 4 (Αφασία Wernicke). Αυτή η προσέγγιση προσφέρει πλεονεκτήματα στο στάδιο της εκπαίδευσης, ειδικά για τον Κανόνα #2 και τον Κανόνα #3.

Οι Κανόνες #2 και #3 θα χρησιμοποιήσουν δεδομένα που δεν θα περιέχουν περιπτώσεις από την κλάση 1. Η ιδέα είναι ότι πρώτα η διαδικασία της εκπαίδευσης επιταχύνεται με τη χρήση ενός μικρότερου συνόλου δεδομένων εκπαίδευσης γιατί με αυτή την προσέγγιση είναι πολύ πιθανός ο σχηματισμός ενός απλού κανόνα[28][29].

Η ταξινόμηση εφαρμόστηκε χρησιμοποιώντας το αποτέλεσμα της μαθηματική έκφραση. Αν το αποτέλεσμα ήταν μηδέν (0) ή θετικό, τότε θεωρήθηκε ότι το αποτέλεσμα ήταν true, ενώ αρνητικό αποτέλεσμα υποδηλώνεται λάθος έξοδο. Από τα δεδομένα εκπαίδευσης εξασφαλίστηκε ότι μια συνολική ακρίβεια ταξινόμησης 100%. Σε λεπτομέρεια, η αντίστοιχη ακρίβεια ταξινόμηση κάθε κανόνας για το δεδομένα δοκιμής φαίνεται στον πίνακα 12 .

Πίνακας 12:Αποτελέσματα ταξινόμησης για τον από GP[9]

Κανόνας #	Σωστή ταξινόμηση στα δεδομένα δοκιμής
1	96.49%
2	80%
3	95.83%
Συνολικά (Χειρότερη περίπτωση)	90.77%

## 2.6.3.2 CRBS-GP: Genetic programming for the production of crisp rule-based systems

### 2.6.3.2.α First GP-model: CRBS-GP1

Στην συγκεκριμένη μεθοδολογία πραγματοποιήθηκαν εννέα εκτελέσεις για κάθε κανόνα, με τη χρήση ενός πληθυσμού από 10.000 προγράμματα (π.χ. δέντρα). Η εκτέλεση του αλγορίθμου τερματίστηκε όταν η τιμή καταλληλότητας γινόταν μεγαλύτερη από ό,τι (0,999). Κάθε 1/10 της κάθε δημιουργίας καταγράφεται καλύτερη λύση του πληθυσμού, και εκτελείται ένας έλεγχος στο σύνολο δοκιμής. Καθ' όλη την εκτέλεση του αλγορίθμου, επιλέγεται η έκφραση που έχει καλές επιδόσεις στην εκπαίδευση και στα δεδομένα δοκιμής. Στις περισσότερες περιπτώσεις, τα πειράματα για την αφασία έχουν δείξει ότι η αλγοριθμική εκτέλεση τελειώνει με μια πολύπλοκη έκφραση, η οποία αποδίδει εξαιρετικά στο σύνολο δεδομένων της εκπαίδευσης, αλλά έχει κακή απόδοση στο σύνολο δεδομένων της δοκιμής. Οι παρακάτω απλοί κανόνες προτείνουν την εξαγωγή μεταξύ των τεσσάρων τύπων της αφασίας [28]:

Κανόνας #1 : αν το N5 ( Επανάληψη των Προτάσεων ) είναι μεγαλύτερη ή ίση με το 23 τότε η αφασία είναι κατονομαστική , που ταξινομεί σωστά το 97.33% των περιπτώσεων εκπαίδευσης και το 95,15% των περιπτώσεων δοκιμής.

Κανόνας #2 : αν το P5 ( Συντακτική Δομή ) είναι μεγαλύτερη ή ίση με το P3 (Σημασιολογική Δομή) , τότε η αφασία είναι Αφασία Broca, που ταξινομεί σωστά το 84.13% των περιπτώσεων εκπαίδευσης και το 86,2% των περιπτώσεων δοκιμής

Κανόνας #3 : αν P5 (Συντακτική Δομή) είναι μικρότερη ή ίση με το 1, τότε υπάρχει Ολική Αφασία, που ταξινομεί σωστά το 100% των περιπτώσεων εκπαίδευσης και το 97.30% των περιπτώσεων δοκιμής.

Πρέπει να σημειωθεί ότι εκτός από αυτούς τους κανόνες υπάρχει μια πιο περίπλοκη σχέση για τον Κανόνα #2, η οποία κατατάσσει τρεις κατηγορίες αφασίας Broca , Wernicke, Ολική αφασία. Η πιο ακριβής μορφή του Κανόνα #2, που κατατάσσει σωστά το 100% των περιπτώσεων εκπαίδευσης και το 91.38% των περιπτώσεων δοκιμής είναι :

Κανόνας #2β: (IfLTE (IfGT (IfBTE (IfLT T0 V0) (IfGTE (IfBT P5 B4 B3) (IfGTE B3 T2)) B3) (IfBTE P3 P5 P1)) (IfBTE (IfGT T3 P3) (NOT (IfBT B3 V0 81)) (NOT C3))). Όταν το αποτέλεσμα είναι μηδέν, τότε η αφασία που υπάρχει είναι Αφασία Broca.

Μια άλλη λιγότερο περίπλοκη μορφή του συγκεκριμένου κανόνα είναι:

Κανόνας 2γ: (OR (IfBTE T4 C3 T4) (OR (IfBTE P3 P5 P1) (OR (IfBTE 3 P5 P1) (IfLTE C3 P2))))), που κατατάσσει σωστά το 100% των περιπτώσεων εκπαίδευσης και το 86.20% των περιπτώσεων δοκιμής. Επίσης και εδώ όταν το αποτέλεσμα είναι μηδέν, τότε η αφασία που υπάρχει είναι Αφασία Broca [28][29].

### 2.6.3.2.β Second GP-model: CRBS-GP2

Το δεύτερο σύστημα αποτελεί την επέκταση του πρώτου συστήματος, γιατί προσπαθεί να παρέχει μια ολοκληρωμένη ταξινόμηση ανάμεσα σε όλους τους τύπους της αφασίας. Τα διαθέσιμα δεδομένα διαχωρίστηκαν και πάλι σε δύο ίσα μέρη. Το πρώτο μέρος χρησιμοποιήθηκε για την εκπαίδευση του συστήματος και το δεύτερο μέρος για την δοκιμή. Μετά από κάθε διακοπή της εκπαιδευτικής διαδικασίας, επιλέχθηκε η έκφραση που είχε τα καλύτερα αποτελέσματα και στις δύο περιπτώσεις, στην εκπαίδευση και στην δοκιμή των δεδομένων. Παρόλο που η κατηγορία της Διαφλοιώδους Αφασίας είναι σπάνια, επιχειρήθηκε η δημιουργία κανόνα και για την συγκεκριμένη περίπτωση [28].



Μετά τη διαδικασία της εκπαίδευσης, οι κανόνες επιτυγχάνοντας την καλύτερη επίδοση τόσο στην εκπαίδευση και σε δεδομένα δοκιμών παρουσιάζονται στον Πίνακα 16. Οι τιμές της ευαισθησίας (sensitivity) και η ειδικότητας (specificity) ίσες με την μονάδα (1) υποδηλώνουν την τέλεια απόδοση, ενώ η τιμή με το μηδέν (0) υποδηλώνει την μικρότερη απόδοση. Οι τιμές απλότητας (simplicity) κανονικοποιούνται στο  $[0,1]$ , με τιμή μηδέν (0) που υποδηλώνει μια έκφραση 150-κόμβων και με τιμή ίση με την μονάδα (1) που υποδηλώνει μια έκφραση με έναν κόμβο.

Πίνακας 16: Τα αποτελέσματα της αφασίας για το CRBS-GP2, με τιμή καταλληλότητας ίση με 1 που υποδηλώνει την τέλεια καταλληλότητα της συγκεκριμένης επιλογής[9].

Rule #	Training data				Test data				Simplicity
	Fitness	Sensitivity	Specificity	Missed cases	Fitness	Sensitivity	Specificity	Missed cases	
1	0.982175	1	0.988095	1	0.924842	0.875	0.974684	4	1
2	1	1	1	0	0.912479	0.9	0.955556	3	0.375839
3	1	1	1	0	0.787879	0.909091	0.666667	2	0.986577
4	1	1	1	0	1	1	1	0	0.95302
5	1	1	1	0	0.91986	0.846154	1	2	0.912752
6	1	1	1	0	1	1	1	0	0.939597
7	1	1	1	0	0.465497	0.95	0	3	0.959732
8	1	1	1	0	0.968807	1	0.923077	1	0.959732
9	0.963134	1	0.909091	1	0.920635	0.952381	0.808089	2	0.926175
10	1	1	1	0	0.845782	1	0.5	1	0.959732
11	1	1	1	0	0.898862	1	0.954545	1	0.986577
12	1	1	1	0	0.936594	0.956522	0.916667	2	0.926175
13	0.960985	0.96	1	1	0.838887	0.88	0.857143	4	0.852349
14	1	1	1	0	1	1	1	0	0.966443

### 2.6.3.3 FRBS-GP: Genetic programming for the production of fuzzy rule-based systems.

Η λύση που λήφθηκε από την προσέγγιση FRBS-GP επιτεύχθηκε μετά 1786 δημιουργίες και αποτελείται από ένα σύνολο έντεκα (11) ασαφών κανόνων σε μορφή if-else, η οποία ταξινομεί σωστά το 88,5% (69 από τις 78 περιπτώσεις) των δεδομένων εκπαίδευσης. Το όνομα του Rx ασαφές σύνολο μπορεί να μεταφραστεί ως 'είναι περίπου X'. Για παράδειγμα, το ασαφές σύνολο R40 έχει έννοια 'είναι περίπου 40', το ασαφές σύνολο R0 έχει έννοια 'είναι περίπου 0' κ.λπ. Οι παραπάνω κανόνες ανήκουν σε έναν ασαφές κανόνα-βάσης. Για κάθε νέα περίπτωση ασθενών που εξετάστηκαν, το πρώτο βήμα αναπαριστά τη δημιουργία ασάφειας των δεδομένων εισόδου. Στη συνέχεια σε κάθε ένα από τους κανόνες δίνεται ένα βάρος και ο κανόνας με το μέγιστο βάρος βγαίνει σαν αποτέλεσμα του συστήματος [28].

Έτσι, οι κανόνες αποτελούν μέρος ενός ασαφούς μηχανισμού και δεν πρέπει να χρησιμοποιούνται χωρίς τον ταξινομητή ασάφειας. Οι συγκεκριμένοι κανόνες περιγράφονται παρακάτω.

1. If B2 is R40 and B1 is R40 and V0 is R60 and C3 is R0 then Broca aphasia
2. If B4 is R0 and D B1 is R40 and C1 is R0 then Broca aphasia
3. If B4 is R0 and C1 is R0 and C1 is R0 and B2 is R50 then Broca aphasia
4. If B1 is R40 and C1 is R0 then Wernicke aphasia
5. If B4 is R0 and B2 is R40 and N0 is R100 then Wernicke aphasia
6. If V0 is R60 and V0 is R60 and B2 is R50 then Wernicke aphasia
7. If N2 is R40 and B2 is R50 and V0 is R60 and B2 is R40 and P1 is medium and C1 is R0 and V0 is R40 and C1 is R0 then Wernicke aphasia
8. If B2 is R40 and V0 is R40 then Wernicke aphasia
9. If C0 is R0 and P1 is Medium and N2 is R40 then Global aphasia



10. If N0 is R100 then Anomic aphasia
11. If B2 is R40 and B1 is R40 and C1 is R0 then Anomic aphasia

Ο παραπάνω ταξινομητής κατηγοριοποίησε σωστά το 79.1% των δομένων δοκιμής.

#### 2.6.3.4 Συγκριτική αξιολόγηση μεθόδων στην υπάρχουσα βιβλιογραφία

Συμπερασματικά μπορούμε να πούμε ότι τα αποτελέσματα των διαφόρων μεθοδολογιών του γενετικού προγραμματισμού διαφέρουν μεταξύ τους. Η ακρίβεια των κανόνων κατά το στάδιο της εκπαίδευσης των δεδομένων φτάνει το 100% και κυμαίνεται μεταξύ 79.8% και του 88.5% όταν γίνεται η δοκιμή των δεδομένων. Τα αποτελέσματα των γενετικών μεθοδολογιών είναι σημαντικά χαμηλότερα σε σχέση με τα αποτελέσματα νευρωνικών δικτύων[28].

Πίνακας 19 :Σύγκριση μεταξύ των διαφόρων NN και GP μεθοδολογιών για την εξέταση της αφασίας [9].

Σύστημα	Ακρίβεια της ταξινόμησης
NN (αυθόρμητη ομιλία)	86%
NN (κατανόηση)	92.4%
Απλός GP	90.8%
CRBS-GP1	79.8%
CRBS-GP2	84.6%
FRBS	79.1%

## ΚΕΦΑΛΑΙΟ 3 ΑΛΓΟΡΙΘΜΟΙ

### Α) ΤΟ ΠΡΟΓΡΑΜΜΑ WEKA

Το πρόγραμμα WEKA, το οποίο αποτέλεσε το εργαλείο που χρησιμοποιήθηκε για τα πειράματα αυτής της εργασίας, είναι ένα ανοιχτού κώδικα πακέτο λογισμικού, γραμμένο εξολοκλήρου στη γλώσσα προγραμματισμού Java, και υλοποιημένο στο πανεπιστήμιο του Waikato στη Νέα Ζηλανδία (Witten & Frank 2000). Το WEKA αποτελεί ένα εργαλείο χρήσιμο τόσο για πειράματα, όσο και για ενσωμάτωση μηχανισμών μηχανικής μάθησης σε καθημερινές εφαρμογές.

Για την εκπόνηση της παρούσας διπλωματικής, χρησιμοποιήθηκαν μερικά έτοιμα σετ δεδομένων

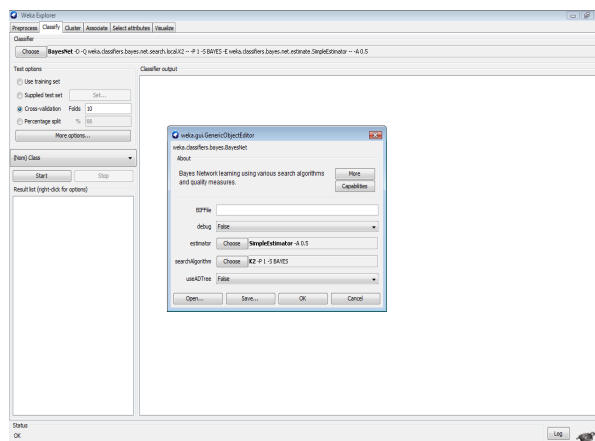
( datasets ) από τη βάση δεδομένων ΑΑΤ.

### 3.1 ΑΛΓΟΡΙΘΜΟΙ ΒΑΣΙΣΜΕΝΟΙ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ ΤΥΠΟΥ ΒΑΥΕΣ

#### 3.1.1 Αλγόριθμος ‘BayesNet’

Ένα δίκτυο “Bayes” είναι η αναπαράσταση μιας δομής δεδομένων για πλήρεις συνδυασμένες κατανομές πιθανοτήτων. (Εργαστήριο Προγραμματισμού & Τεχνολογίας Ευφυών Υπολογιστικών Συστημάτων)<sup>1</sup> Εάν έχουμε ένα σύνολο δεδομένων πχ.  $X = \{\chi_1, \chi_2\}$  το πρόβλημα για την εξόρυξη γνώσης είναι να ανακαλύψει τις ιδιότητες της κατανομής από την οποία προέρχεται το σύνολο. Ο αλγόριθμος “Bayes Net” έχει την ιδιότητα να μαθαίνει από τα δίκτυα τύπου Bayes, υπό την υπόθεση ότι υπάρχουν εκτός από αριθμητικά και ονομαστικά χαρακτηριστικά.

Για ευκολία του αναγνώστη στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγόριθμου καθώς αναφέρονται και οι παράμετροι που χρησιμοποιήθηκαν στη διπλωματική αυτή:



Εικόνα 3.1.1: Ανοικτές παράμετροι του αλγόριθμου ‘BayesNet’

- Estimator: χρησιμοποιείται για να επιλέξει τη μέθοδο για τον υπολογισμό του πίνακα δεσμευμένων πιθανοτήτων του δικτύου τύπου Bayes.

Πιο συγκεκριμένα, χρησιμοποιήθηκε η μέθοδος BMAestimator αντί την προεπιλεγμένη Simple estimator, όσον αφορά την παράμετρο Estimator. Η μέθοδος Simple Estimator χρησιμοποιεί έναν απλό εκτιμητή των δεσμευμένων πιθανοτήτων για τα δίκτυα τύπου Bayes. Η μέθοδος BMAestimator χρησιμοποιεί έναν εκτιμητή για τις δεσμευμένες πιθανότητες, ο οποίος βασίζεται στο υπόδειγμα Bayes Model Averaging, το οποίο αυξάνει την ακρίβεια του εκτιμητή.

- SearchAlgorithm: χρησιμοποιείται για να γίνει επιλογή μιας μεθόδου για να αναζητήσει δομών δικτύων.

Σε αυτήν την παράμετρο SearchAlgorithm χρησιμοποιήσαμε τις ακόλουθες μεθόδους:

1

<sup>1</sup> Εργαστήριο Προγραμματισμού & Τεχνολογίας Ευφυών Υπολογιστικών Συστημάτων  
Τμήμα Ηλεκτρονικών Μηχανικών & Μηχανικών Υπολογιστών - [www.intelligence.tuc.gr](http://www.intelligence.tuc.gr)

- i. Hill-Climber, είναι η μέθοδος αναζήτησης Hill Climbing
- ii. LAGDHill-Climber, εφαρμόζει μια τεχνική κατασκευής δικτύων τύπου Bayes κατά την οποία εκτελείται ένας αριθμός διαδικασιών ‘προς-τα-εμπρός’ (Look Ahead), ενώ πρέπει να αναφέρουμε ότι δεν υλοποιείται η στρατηγική ‘άπληστης’ αναζήτησης. Η μέθοδος αυτή βασίζεται στον αλγόριθμο Hill Climbing
- iii. Repeated Hill-Climber, εφαρμόζει επαναληπτικά τον αλγόριθμο Hill Climbing, με αποτέλεσμα να παράγει τυχαία δομές δικτύου Bayes κι επιστρέφει την καλύτερη
- iv. TabuSearch, αποτελεί μια τεχνική τύπου hill Climbing διαφέροντας στην εκτέλεση του αλγορίθμου. Το βασικό της χαρακτηριστικό είναι ότι σταματά μόλις βρεθεί μια βέλτιστη, σε σχέση με τις προηγούμενες, λύση
- v. TAN, ο συγκεκριμένος αλγόριθμος μάθησης καθορίζει ένα δέντρο απόφασης με το μέγιστο βάθος, κι επιστρέφει ένα δίκτυο απλού τύπου Bayes (Naïve Bayes), το οποίο ενισχύεται με το αντίστοιχο δέντρο απόφασης

BayesNet	CORRECTLY CLASSIFIED
<b>Parameter: estimator</b>	
<i>SimpleEstimator</i>	45.4545
BMAEstimator	-
MultinomialBMAEstimator	-
<b>Parameter: searchAlgorithm</b>	
<i>K2</i>	45.4545
LAGDHillClimber	47.8469
HillClimber	44.0191
TAN	38.2775
TabuSearch	45.933

Πίνακας 3.1.1: Πίνακας αποτελεσμάτων του αλγορίθμου ‘BayesNet’

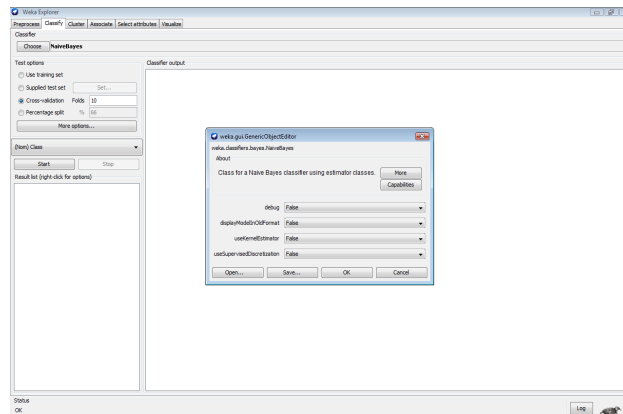
Όπως μπορούμε να παρατηρήσουμε από τον πίνακα 3.1.1, το μεγαλύτερο ποσοστό ορθούς ταξινόμησης επιτεύχθηκε όταν χρησιμοποιήθηκε ο αλγόριθμος LAGDHillClimber για αναζήτηση δομών δικτύου,(47,85%).

### 3.1.2 Αλγόριθμος ‘NaiveBayes’

Ο αλγόριθμος “Naïve Bayes” εφαρμόζει τη θεωρία πιθανοτήτων για την ταξινόμηση των δεδομένων του μοντέλου και βασίζεται στο κανόνα του Bayes (George H. John & Pat Langley, 1995). Ο ταξινομητής που βασίζεται στην μέθοδο αυτή υποθέτει ότι η παρουσία ενός συγκεκριμένου χαρακτηριστικού μιας κλάσης δε μπορεί να συσχετιστεί με την παρουσία των υπόλοιπων χαρακτηριστικών. Πιο επεξηγηματικά, για την ταξινόμηση μιας περίπτωσης σε μια συγκεκριμένη κλάση, ο αλγόριθμος λαμβάνει υπόψη του ότι το κάθε χαρακτηριστικό της περίπτωσης συνεισφέρει ανεξάρτητα στην ταξινόμηση της.

Για ευκολία του αναγνώστη στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγορίθμου

καθώς αναφέρονται και οι παράμετροι που χρησιμοποιήθηκαν στη διπλωματική αυτή:



Εικόνα 3.1.2: Ανοικτές παράμετροι του αλγόριθμου ‘NaiveBayes’

- useKernelEstimator: είναι η παράμετρος η οποία χρησιμοποιείται για τα αριθμητικά χαρακτηριστικά και όχι για την κανονική κατανομή, βασισμένος σε έναν εκτιμητή πυρήνων.
- useSupervisedDiscretization: είναι η παράμετρος η οποία μετατρέπει αριθμητικά χαρακτηριστικά χρησιμοποιώντας εποπτευόμενες διακριτοποιήσεις.

NaiveBayes	CORRECTLY CLASSIFIED
<b>Parameter: useKernelEstimator</b>	
<i>FALSE</i>	44.0191
TRUE	44.0191
<b>Parameter: useSupervisedDiscretization</b>	
<i>FALSE</i>	44.0191
TRUE	44.0191

Εικόνα 3.1.2: Πίνακας αποτελεσμάτων του αλγόριθμου ‘NaiveBayes’

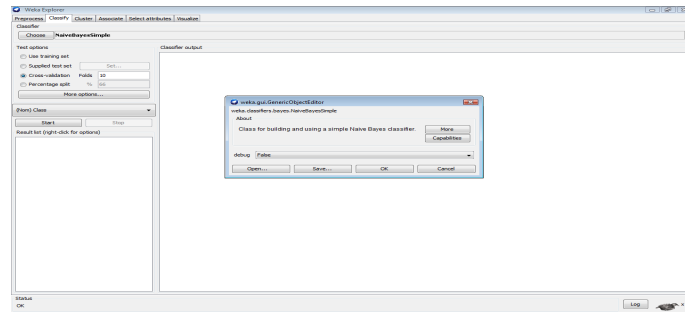
Παρατηρούμε ότι η χρήση του κεντρικού εκτιμητή τύπου Kernel δεν επιρέασε τα ποσοστά ταξινόμησης. Το ίδιο παρατηρείται και με τη χρήση του SupervisedDiscretization.

### 3.1.3 Αλγόριθμος Naive Bayes Simple

Ο ταξινομητής “Naive Bayes Simple” δημιουργεί απλά μοντέλα ταξινόμησης που βασίζονται στον κανόνα του Bayes και τα αριθμητικά χαρακτηριστικά μοντελοποιούνται μέσω κανονικής κατανομής (Richard Duda & Peter Hart, 1973).

Η προσέγγιση της απλής κατηγοριοποίησης κατά Bayes έχει αρκετά πλεονεκτήματα αφού είναι εύκολο να χρησιμοποιηθεί, απαιτείται μόνο ένα πέρασμα των δεδομένων εκπαίδευσης, αντίθετα με άλλες προσεγγίσεις κατηγοριοποίησης, ενώ τέλος στις τεχνικές που υπάρχουν απλές συσχετίσεις, η τεχνική συνήθως δίνει καλά αποτελέσματα (Margaret H.Dunham,2004).

Ακολουθεί ο πίνακας παραμέτρων καθώς φαίνεται και η παράμετρος ‘debug’ η οποία είναι η μοναδική για τον αλγόριθμο αυτόν:



Εικόνα 3.1.3: Ανοικτές παράμετροι του αλγόριθμου ‘NaiveBayesSimple’

NaiveBayesSimple	CORRECTLY CLASSIFIED
Default Values	44.0191

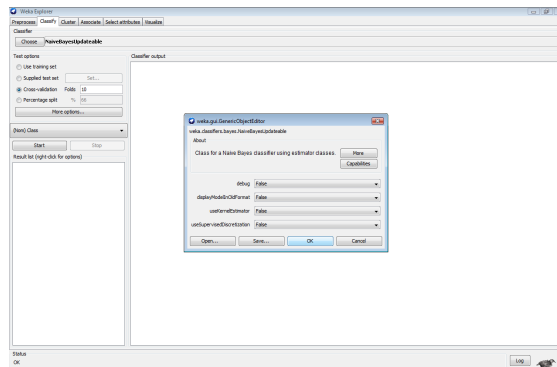
Πίνακας 3.1.3: Πίνακας αποτελεσμάτων του αλγόριθμου ‘NaiveBayesSimple’

Ο αλγόριθμος αυτός πέτυχε ποσοστό, 44,02%. Οι ρυθμίσεις που κρατήσαμε είναι οι προεπιλεγμένες από το πρόγραμμα.

### 3.1.4 Αλγόριθμος Αλγόριθμος Naive Bayes Updateable

Ο αλγόριθμος “Naive Bayes Updateable” αποτελεί μια αναβαθμισμένη παραλλαγή του αλγόριθμου Naive Bayes υποβάλλοντας σε λειτουργία μια περίπτωση του δείγματος, την φορά (George H. John & Pat Langley, 1995). Επιπλέον, η λειτουργία του συγκεκριμένου αλγορίθμου διαφοροποιείται στον τρόπο με τον οποίο χειρίζεται τις περιπτώσεις όπου δεν παρέχονται δεδομένα εκπαίδευσης.

Για ευκολία του αναγνώστη στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγόριθμου καθώς αναφέρονται και οι παράμετροι που χρησιμοποιήθηκαν:



Εικόνα 3.1.4: Ανοικτές παράμετροι του αλγόριθμου ‘NaiveBayesUpdateable’

- KernelEstimator ο οποίος χρησιμοποιεί έναν κεντρικό εκτιμητή για τα αριθμητικά χαρακτηριστικά αντί για χρήση κανονικής κατανομής.
- UseSupervised Discretization δεν επιδέχεται μεταβολή σε true από την αρχική της ρύθμιση που είναι false για αυτό και δεν συμπεριλήφθηκε στο πίνακα μας η συγκεκριμένη μεταβολή.

NaiveBayesUpdateable	CORRECTLY CLASSIFIED
<b>Parameter: useKernelEstimator</b>	
FALSE	44.0191
TRUE	44.0191
<b>Parameter: useSupervisedDiscretization</b>	
FALSE	44.0191

Πίνακας 3.1.4: Πίνακας αποτελεσμάτων του αλγόριθμου ‘NaiveBayesUpdateable’

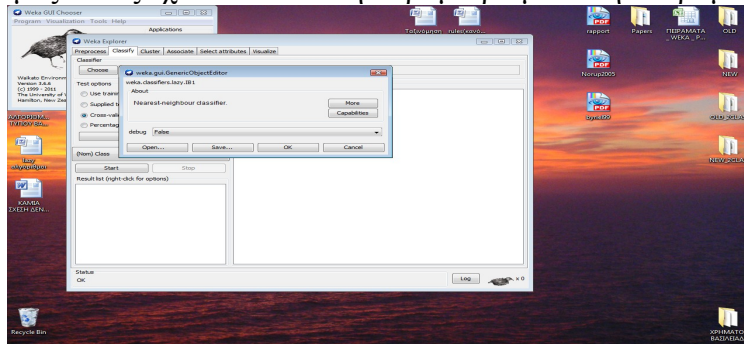
Μπορούμε να παρατηρήσουμε ότι ο εκτιμητής τύπου Kernel και η παράμετρος useSupervisedDiscretization δίνουν ακριβώς τις ίδιες τιμές.

## 3.2 ΑΛΓΟΡΙΘΜΟΙ ΒΑΣΙΣΜΕΝΟΙ ΣΤΗΝ ΑΠΟΣΤΑΣΗ (lazy)

### 3.2.1 Αλγόριθμος ‘IB1’

Ο IB1 είναι ένας αλγόριθμος απόστασης (Aha & Kibler, 1991), ο οποίος είναι βασισμένος στα παραδείγματα και χρησιμοποιεί την κανονικοποιημένη ευκλείδεια απόσταση για να βρει τα παραδείγματα εκπαίδευσης, που βρίσκονται πιο κοντά στα δεδομένα έλεγχο. Επίσης προβλέπει την ίδια κλάση με αυτή του παραδείγματος εκπαίδευσης. Εάν υπάρχουν περισσότερα του ενός παραδείγματα με την ίδια ελάχιστη απόσταση τότε επιλέγεται το πρώτο κατά σειρά που θα βρεθεί.

Ο αλγόριθμος αυτός έχει σαν ανοικτή παράμετρο μόνο την παράμετρο debug:



Εικόνα 3.2.1: Ανοικτές παράμετροι του αλγόριθμου IB1

IB1	CORRECTLY CLASSIFIED
Default Values	41.6268

Πίνακας 3.2.1: Πίνακας αποτελεσμάτων του αλγόριθμου IB1

Παρατηρούμε ότι η μέθοδος αυτή δίνει ποσοστά χειρότερα από αυτά των αλγόριθμων τύπου Bayes.

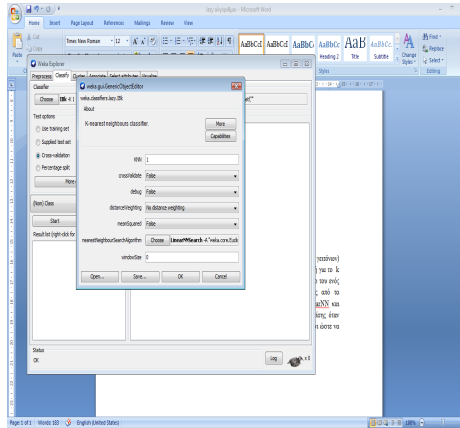
### 3.2.2 Αλγόριθμος ‘IBK’

Ο αλγόριθμος IBK είναι ένας k-πλησιέστερων γειτόνων ταξινομητής (K-nearest neighbors) (Aha & Kibler, 1991). Μπορεί να επιλεγεί την κατάλληλη τιμή για το k μέσω της διαδικασίας του ‘cross validation’. Οι προβλέψεις των περισσότερων του ενός γειτόνων μπορούν να σταθμίζονται σύμφωνα με τις αποστάσεις τους από τα παραδείγματα δοκιμής και δυο διαφορετικές μορφές



γειτόνων, τις LinearNN και KDTree, εφαρμόζονται για την μετατροπή της απόστασης σε βάρος. Επίσης όταν προστίθενται νέα παραδείγματα εκπαίδευσης τα παλιά απομακρύνονται έτσι ώστε να διατηρηθεί ο αριθμός των παραδειγμάτων εκπαίδευσης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.2.2: Ανοικτές παράμετροι του αλγόριθμου IBK

- KNN: Χρησιμοποιείται για να επιλέξουμε τον αριθμό γειτόνων που θα χρησιμοποιηθούν.
- DistanceWeighting: Ορίζει τη μέθοδο που θα χρησιμοποιηθεί για distance weighting. Δίδονται τρεις επιλογές.
  1. No Distance Weighting (Μη χρήση της παραμέτρου )
  2. Weight by 1/Distance (Το βάρος των γειτόνων χρησιμοποιώντας το αντίστροφο της απόστασης τους )
  3. Weight by 1-Distance (βάρος των γειτόνων χρησιμοποιώντας το ένα μείον την απόστασή τους)
- NearestNeighbourSearchAlgorithm: Καθορίζει την μέθοδο αναζήτησης του πλησιέστερου γείτονα. Δίνονται δυο επιλογές. Οι LinearNN και KDTree.

lbk	CORRECTLY CLASSIFIED
Parameter: KNN	
1	43.5407
10	42.5837
20	42.5837
<b>Parameter: distanceWeighting</b>	
<i>No distance Weighting</i>	43.5407
<i>Weight by 1/distance</i>	43.5407
<i>Weight by 1-distance</i>	43.5407
<b>Parameter: nearestNeighbourSearchAlgorithm</b>	
<i>LinearNNSearch</i>	43.5407
<i>BallTree</i>	43.5407
<i>CoverTree</i>	41.6268
<i>KDTree</i>	43.5407

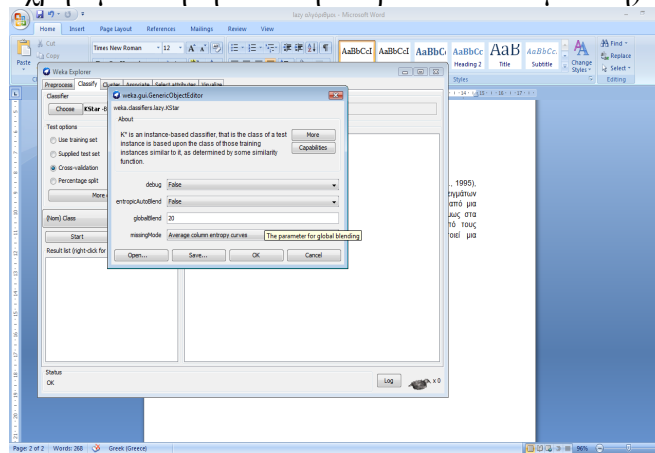
Πίνακας 3.2.2: Πίνακας αποτελεσμάτων του αλγόριθμου IBK

Η μέθοδος αυτή χαρακτηρίζεται από ποσοστά μέτριας απόδοσης όσον αφορά την ακρίβεια . Μετά από αρκετά πειράματα καταλήξαμε στο συμπέρασμα ότι μόνο η παράμετρος CoverTree έχει επίδραση στην ακρίβεια του μοντέλου και μάλιστα αρνητική.

### 3.2.3 Αλγόριθμος ‘Kstar’

Ο αλγόριθμος “KStar” είναι ένας instance-based ταξινομητής (John et al., 1995), στον οποίο ένα παράδειγμα έλεγχου βασίζεται στην κλάση των παραδειγμάτων εκπαίδευσης που είναι όμοια σε αυτό, πράγμα το οποίο καθορίζεται από μια συνάρτηση ομοιότητας (similarity function) η οποία, δεν καθορίζεται όμως στα πλαίσια αυτού του προγράμματος. Ο αλγόριθμος αυτός διαφέρει από τους υπόλοιπους ταξινομητές του ίδιου τύπου στο γεγονός ότι χρησιμοποιεί μια συνάρτηση απόστασης βασισμένη στην εντροπία.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.2.3: Ανοικτές παράμετροι του αλγόριθμου Kstar

- GlobalBlend: Αναφέρεται στο global blending (σφαιρικό μείγμα) και έχει πεδίο τιμών από 0 έως 100.

Kstar	CORRECTLY CLASSIFIED
<b>Parameter: globalBlend</b>	
20	39.2344
5	39.2344
80	39.7129

Πίνακας 3.2.3: Πίνακας αποτελεσμάτων του αλγόριθμου Kstar

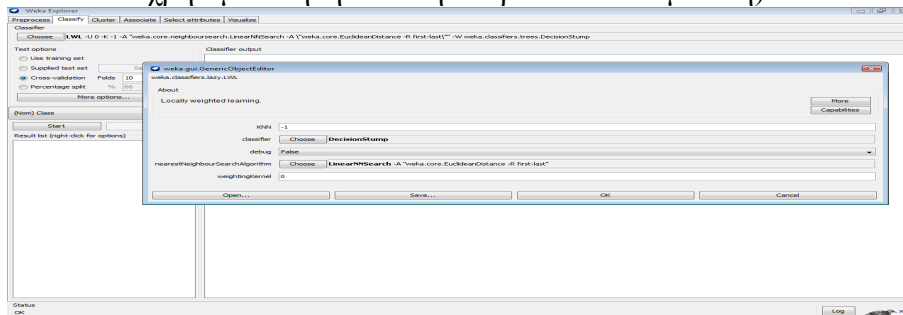
Παρατηρούμε στον παραπάνω πίνακα ότι δεν πετύχαμε αξιόλογες ακρίβειες και μάλιστα δεν απέχουν σημαντικά τα ποσοστά μεταξύ τους τόσο στον βαθμό των επαναλήψεων όσο και στις μετατροπές των παραμέτρων.

### 3.2.4 Αλγόριθμος ‘Τοπικής Σταθμισμένης Εκμάθησης’ (LWL)

Ο αλγόριθμος LWL (Locally Weighted Learning) (Atkenson et al., 1992) χρησιμοποιεί έναν αλγόριθμο βασισμένο στα στιγμιότυπα στον οποίο αναθέτει τον καθορισμό των βαρών των παραδειγμάτων τα οποία εν συνεχεία χρησιμοποιούνται από μια ειδική μέθοδο χειρισμού των

βαρών. Επίσης ο αλγόριθμος δίνει την επιλογή ορισμού του επιθυμητού ταξινομητή.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.2.4: Ανοικτές παράμετροι του αλγόριθμου LWL

- KNN: Ο αλγόριθμος k-NN (Mitchell 1997) αποτελεί έναν από τους πιο χαρακτηριστικούς αλγόριθμους για μάθηση βασισμένη σε παραδείγματα. Η πλήρης ονομασία του είναι αλγόριθμος των k κοντινότερων γειτόνων (Nearest Neighbor – NN). Κατά την εκπαίδευση, ο αλγόριθμος k-NN απλά αποθηκεύει σε μια μνήμη όλα τα διανύσματα των αντικειμένων του συνόλου TrS και τις σωστές κατηγορίες τους. Η κατάταξη νέων αντικειμένων, των οποίων δεν είναι γνωστές οι κατηγορίες, γίνεται ως εξής: Υπολογίζεται η απόσταση του διανύσματος του νέου αντικειμένου από τα διανύσματα όλων των αντικειμένων εκπαίδευσης. Επιλέγονται τα k αντικείμενα εκπαίδευσης με τις μικρότερες αποστάσεις (οι k κοντινότεροι γείτονες) και το νέο αντικείμενο κατατάσσεται στην κατηγορία που πλειοψηφεί μεταξύ των k αντικειμένων.
- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NearestNeighbourSearchAlgorithm: Καθορίζει την μέθοδο αναζήτησης του πλησιέστερου γείτονα. Δίδονται δυο επιλογές, οι LinearNN και KDTree.

LWL	CORRECTLY CLASSIFIED
<b>Parameter: KNN</b>	
-1	47.8469
10	40.1914
20	44.4976
<b>Parameter: classifier</b>	
<i>Decision Stump</i>	47.8469
J48	44.4976
random tree	27.7512
PART	45.4545
Jrip	38.9423
<b>Parameter: NearestNeighbourSearchAlgohythm</b>	
<i>LinearNNSearch</i>	47.8469
BallTree	47.8469
CoverTree	47.8469
KDTree	47.8469

**Πίνακας 3.2.4: Πίνακας αποτελεσμάτων του αλγόριθμου LWL**

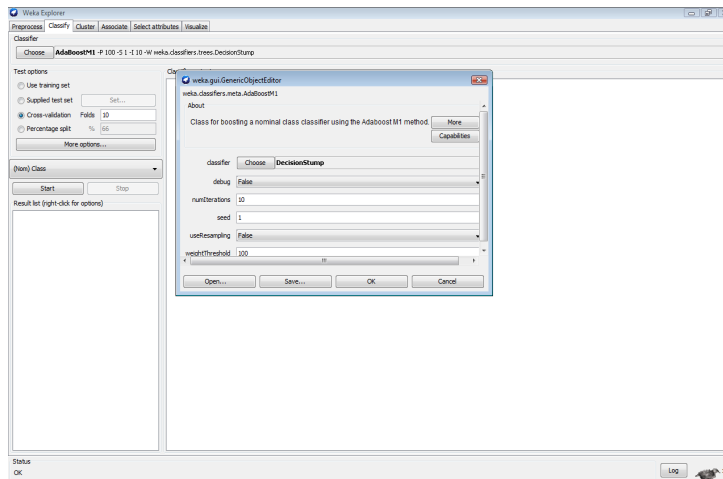
Παρατηρώντας τον πίνακα 3.2.4 βλέπουμε ότι αυτή η μέθοδος ήταν η πιο αποδοτική από τους αλγόριθμους βασισμένους στη απόσταση. Όπως Παρατηρούμε η παράμετρος ‘nearestneighboursearchalgorithm’ δεν προσφέρει κάτι στο πρόβλημα, αφού δεν υπάρχει καμία μεταβολή ακόμη και αν αλλάξουμε τις τιμές των μεταβλητών. Επίσης πειραματιστήκαμε με την επιλογή classifier όπου το πρόγραμμα δίνει την επιλογή να αλλάξει ο προεπιλεγμένος ταξινομητής (Decision Stump) και να χρησιμοποιηθεί κάποιος άλλος από όλους τους ταξινομητές του WEKA. Μετά από αρκετές δόκιμες παρατηρήθηκε ότι τα αποτελέσματα είχαν μεγάλη διακύμανση με μέγιστη απόδοση όταν χρησιμοποιήθηκε ο ταξινομητής DecisionStump (47,85%) και ελάχιστη με τη χρήση του ταξινομητή random tree (27,75%).

### 3.3 ΑΛΓΟΡΙΘΜΟΙ ΜΕΤΑΜΑΘΗΣΗΣ (meta)

#### 3.3.1 Ο αλγόριθμος ‘AdaBoostM1’

Ο αλγόριθμος “AdaBoostM1” είναι ένας αλγόριθμος που βασίζεται στη μέθοδο “boosting” (Freund & Schapire, 1996). Δίνεται η πιθανότητα επιλογής ενός ταξινομητή (προεπιλεγμένος είναι το δένδρο τύπου Decision Stump) και σε ορισμένα προβλήματα ο αλγόριθμος μπορεί να βοηθήσει τον ταξινομητή να βελτιώσει την απόδοση του. Πολλές φορές όμως ο αλγόριθμος μπορεί να υποπέσει στην «παγίδα» του υπερπαιριάσματος (overfitting). Να σημειωθεί ότι ο αλγόριθμος λειτουργεί καλύτερα σε προβλήματα διακριτών μεταβλητών και επιπλέον χρησιμοποιεί επαναλαμβανόμενα δείγματα για την περίπτωση που ο βασικός ταξινομητής που έχει επιδεχθεί δε μπορεί να χειριστεί παραδείγματα που περιέχουν βάρη.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.1: Ανοικτές παράμετροι του αλγόριθμου AdaBoostM1**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

AdaBoostM1	CORRECTLY CLASSIFIED
<b>Parameter: numIterations</b>	
10	29.6651
5	29.6651
20	29.6651
<b>Parameter: classifier</b>	
DecisionStump	29.6651
J48	48.8038
random forest	27.2727
PART	45.4545
Jrip	39.7129

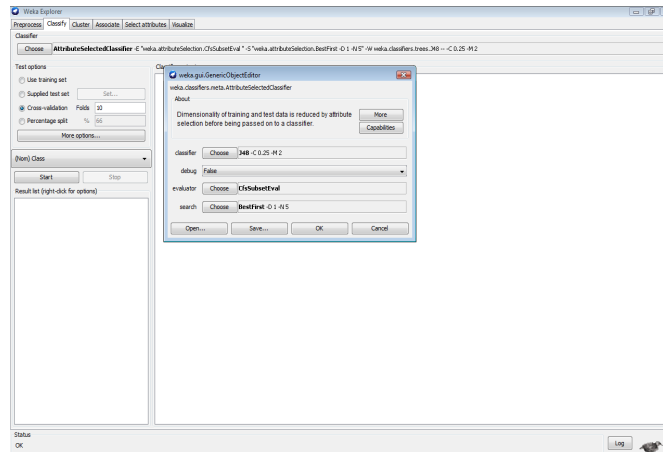
**Πίνακας 3.3.1: Πίνακας αποτελεσμάτων του αλγόριθμου AdaBoostM1**

Τα αποτελέσματα δεν δείχνουν να επηρεάζονται από τον αριθμό των επαναλήψεων. Επιπλέον, στα πειράματα που πραγματοποιήθηκαν δοκιμάσαμε αρκετούς ταξινομητές, ισχυρούς και αδύναμους και παρατηρήσαμε μεγάλη διακύμανση. Το καλύτερο ποσοστό ορθής ταξινόμησης επιτεύχθηκε με τον ταξινομητή J48 (48,80%).

### 3.3.2 Αλγόριθμος Ταξινόμησης Επιλεγμένων Χαρακτηριστικών (Attribute Selected Classifier)

Ο αλγόριθμος AttributeSelectedClassifier έχει την ιδιότητα να επιλεγεί χαρακτηριστικά, ελαττώνοντας τις διαστάσεις των δεδομένων πριν αυτά εκτελεστούν στον ταξινομητή. Αυτό μπορεί σε ορισμένα προβλήματα να βελτιώσει την απόδοση του βασικού ταξινομητή.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.2: Ανοικτές παράμετροι του αλγόριθμου Attribute Selected Classifier

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- Evaluator: Καθορίζει τον εκτιμητή που θα χρησιμοποιηθεί κατά τη διάρκεια επιλογής χαρακτηριστικού πριν το ξεκίνημα της ταξινόμησης.
- Search: Καθορίζει τη μέθοδο που θα χρησιμοποιηθεί κατά τη διάρκεια επιλογής χαρακτηριστικού πριν το ξεκίνημα της ταξινόμησης.

AttributeSelectedClassifier	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
DecisionStump	29.6651
J48	49.2823
random forest	27.7512
PART	46.89
Jrip	47.8469
<b>Parameter: evaluator</b>	
CfsSubsetEval	49.2823
ClassifierSubsetEval	19.6172
ConsistencySubsetEval	29.1866
<b>Parameter: search</b>	
BestFit	49.2823
GeneticSearch	50.2392
RandomSearch	49.7608

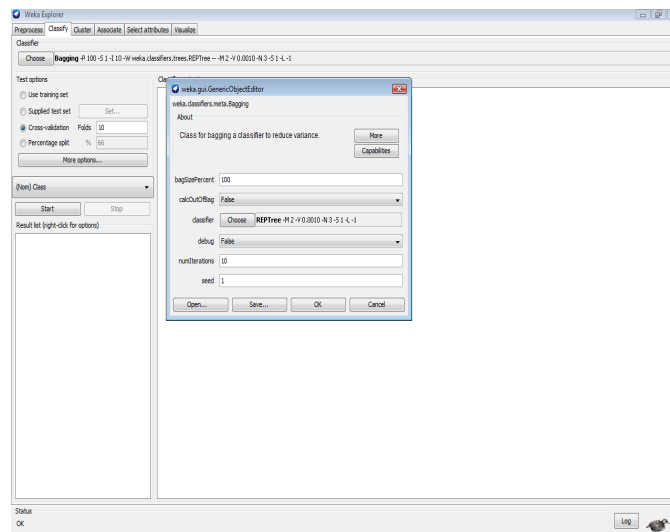
Πίνακας 3.3.2: Πίνακας αποτελεσμάτων του αλγόριθμου Attribute Selected Classifier

Μετά από πειράματα που έγιναν με αυτή τη μέθοδο μεταμάθησης παρατηρήθηκε ότι μπορεί να βελτιώσει, ελάχιστα όμως τις επιδόσεις του βασικού ταξινομητή. Εμείς παραθέτουμε τα αποτελέσματα για βασικούς ταξινομητές τους J48, CfsSubsetEval, BestFirst. Μεταβολές που πραγματοποιήθηκαν στην μέθοδο και στον εκτιμητή που θα χρησιμοποιηθούν κατά τη διάρκεια επιλογής χαρακτηριστικού πριν το ξεκίνημα της ταξινόμησης φάνηκε να επηρεάζουν την ακρίβεια του μοντέλου τόσο αρνητικά όσο και θετικά όπως φαίνεται παραπάνω. Το καλύτερο ποσοστό που επιτεύχθηκε, ήταν (50,24%).

### 3.3.3 Ο Αλγόριθμος ‘Bagging’

Ο αλγόριθμος “Bagging” (ή αλλιώς Bootstrap Aggregating)(Breiman,2001) αποτελεί έναν μέτα-αλγόριθμο μάθησης ο οποίος βελτιώνει την ακρίβεια ταξινόμησης. Επιπλέον, μειώνει τη διασπορά και ενισχύει την αποφυγή του υπέρ-ταιριάσματος. Η συγκεκριμένη μέθοδος λειτουργεί ως ακολούθως. Έστω ένα σετ δεδομένων εκπαίδευσης D μεγέθους n, τότε η μέθοδος παράγει m μικρότερου μεγέθους σετ δεδομένων εκπαίδευσης με δειγματοληψία (με επανάθεση) από το αρχικό σετ D, η οποία ακολουθεί την ομοιόμορφη κατανομή. Σε επόμενο στάδιο, κατασκευάζονται m υποδείγματα, χρησιμοποιώντας τα παραπάνω υπό-σετ δεδομένων, και συνδυάζονται είτε λαμβάνοντας υπόψη το μέσο όρο των επιμέρους αποτελεσμάτων (στην περίπτωση της παλινδρόμησης) είτε την ψηφοφορία (στην περίπτωση της ταξινόμησης).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.3: Ανοικτές παράμετροι του αλγόριθμου Bagging

- BagSizePercent: Καθορίζει το μέγεθος του κάθε «σάκου» (bag) ως ποσοστό του μεγέθους των δεδομένων εκπαίδευσης.
- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν

Bagging	CORRECTLY CLASSIFIED
<b>Parameter: bagSizePercent</b>	
100	27.2727
50	23.9234
25	22.488
<b>Parameter: classifier</b>	
REPTree	27.2727
PART	46.89
MultilayerPerceptron	-
<b>Parameter: numIterations</b>	
10	27.2727
5	28.2297
20	24.4019

Πίνακας 3.3.3: Πίνακας αποτελεσμάτων του αλγόριθμου Bagging

Παρατηρώντας τον παραπάνω πίνακα 3.1.3 βλέπουμε ότι τα αποτελέσματα που μας δίνει ο

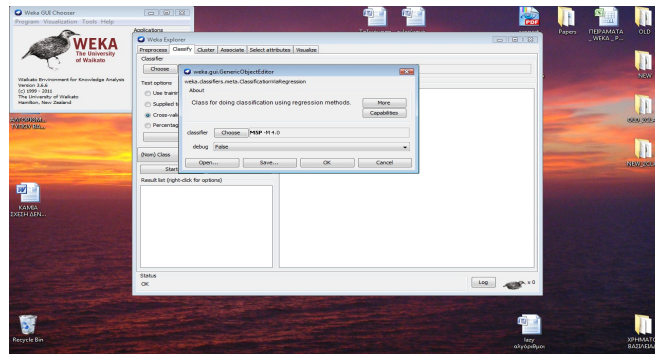


αλγόριθμος Bagging βελτιώνονται αρκετά με την χρήση του ταξινομητή PART (46,89%).

### 3.3.4 Αλγόριθμος Ταξινόμησης Μέσω Παλινδρόμησης (Classification Via Regression)

Ο αλγόριθμος Classification Via Regression (Frank et al., 1998) κάνει ταξινόμηση χρησιμοποιώντας την μέθοδο της παλινδρόμησης μετατρέποντας την κλάση σε δυαδική μορφή και δημιουργώντας ένα παλινδρομικό μοντέλο για κάθε τιμή της κλάσης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.4: Ανοικτές παράμετροι του αλγόριθμου Classification Via Regression

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.

ClassificationViaRegression	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
M5P	23.445
ConjunctiveRule	41.1483
Kstar	39.2344

Πίνακας 3.3.4: Πίνακας αποτελεσμάτων του αλγόριθμου Classification Via Regression

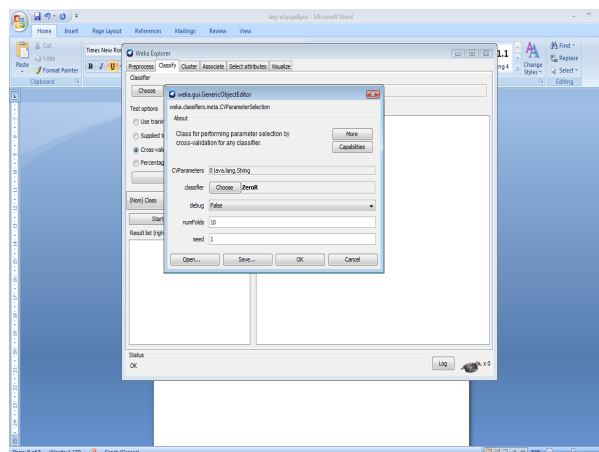
Με τη χρήση του αλγορίθμου classification via regression και τον ταξινομητή ConjunctiveRule πήραμε την μέγιστη ακρίβεια (41,15%)

### 3.3.5 Αλγόριθμος Επιλογής Παραμέτρων (CV Parameter Selection)

Ο αλγόριθμος CV Parameter Selection (Kohavi, 1995) έχει την ικανότητα να βελτιώνει την διαδικασία της ταξινόμησης χρησιμοποιώντας τη μέθοδο cross-validation για να επιλέξει τις παραμέτρους της κάθε ταξινόμησης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.5: Ανοικτές παράμετροι του αλγόριθμου CV Parameter Selection

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumFolds: Ορίζει τον αριθμό των Folds για την μέθοδο cross validation.

CVParameterSelection	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
ZeroR	19.6172
BFTree	36.8421
Jrip	41.6268
Logistic	26.7943
<b>Parameter: numFolds</b>	
10	19.6172
5	19.6172
20	19.6172

Πίνακας 3.3.5: Πίνακας αποτελεσμάτων του αλγόριθμου CV Parameter Selection

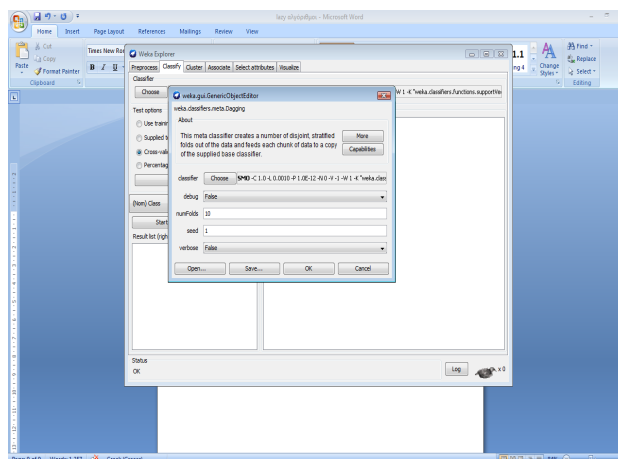
Με την μέθοδο αυτή έγιναν αρκετά πειράματα και δοκιμάσαμε αρκετούς ταξινομητές του WEKA. Παρατηρήσαμε μεγάλη διακύμανση στην επίδοση του κάθε ταξινομητή. Οι παράμετροι που εξετάστηκαν, έδωσαν καλύτερα αποτελέσματα σε σύγκριση με τις προκαθορισμένες τιμές. Πιο συγκεκριμένα, το μεγαλύτερο ποσοστό ορθής ταξινόμησης ήταν 41,63% με τον ταξινομητή Jrip.

### 3.3.6 Αλγόριθμος ‘Dagging’

Ο ταξινομητής “Dagging” (Ting & Witten, 1997) δημιουργεί έναν αριθμό από στρωματοποιημένα δεδομένα και κατόπιν τα τροφοδοτεί στον βασικό ταξινομητή. Οι προβλέψεις υπολογίζονται βάσει του μέσου όρου των αποτελεσμάτων, καθώς όλα τα επιμέρους αποτελέσματα του ταξινομητή χρησιμοποιούνται ως δεδομένα εισόδου στον μετά-ταξινομητή τύπου Vote. Αποτελεί μια χρήσιμη μέθοδο για τους βασικούς ταξινομητές που έχουν πολύ μεγάλο χρόνο εκτέλεσης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.6: Ανοικτές παράμετροι του αλγόριθμου Dagging**

- NumFolds: Ορίζει τον αριθμό των Folds για την μέθοδο cross validation.
- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.

Dagging	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
SMO	40.1914
Jrip	33.4928
J48	44.9761
<b>Parameter: numFolds</b>	
10	40.1914
5	41.1483
20	29.6651

**Πίνακας 3.3.6: Ανοικτές παράμετροι του αλγόριθμου Dagging**

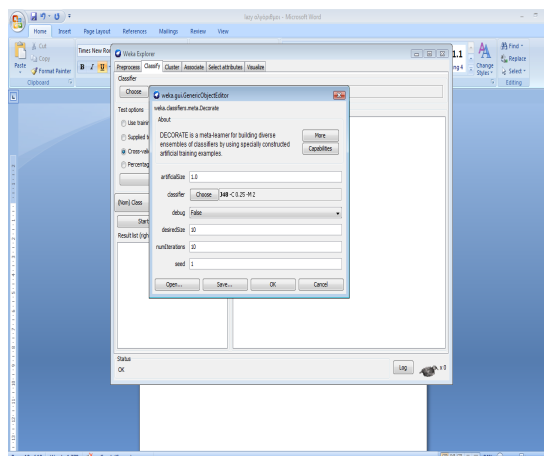
Παρατηρούμε από τον παραπάνω πίνακα 3.2.7 ότι το καλύτερο ποσοστό ορθούς ταξινόμησης επιτεύχθηκε με τον ταξινομητή J48 (44,98%).

### 3.3.7 Αλγόριθμος ‘Decorate’

Η τεχνική ‘Decorate’ (Melville & Mooney, 2005) αποτελεί έναν αλγόριθμο μεταμάθησης ο οποίος κατασκευάζει διαφόρους ταξινομητές χρησιμοποιώντας ειδικά κατασκευασμένα παραδείγματα εκπαίδευσης. Αναλυτικά πειράματα έχουν δείξει ότι αυτή η μέθοδος είναι αρκετά πιο ακριβής σε σχέση με τον βασικό ταξινομητή και τις μεθόδους ‘Bagging’ και ‘Random Forest’. Ο αλγόριθμος ‘Decorate’ μπορεί επίσης να επιτύχει μεγαλύτερη ακρίβεια σε μικρά συνήθως σετ δεδομένων σε σχέση με τη μέθοδο ‘Boosting’ αλλά και σε μεγαλύτερα σετ δεδομένων μπορεί να σημειώσει αξιόλογη ακρίβεια.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.7: Ανοικτές παράμετροι του αλγόριθμου Decorate

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν. Κάθε επανάληψη δημιουργεί έναν ταξινομητή αλλά δε τον προσθέτει απαραίτητα στο δείγμα. Η λειτουργία του ολοκληρώνεται όταν επιτευχθεί το επιθυμητό μέγεθος που έχει οριστεί από την παράμετρο DesiredSize.

Decorate	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
J48	29.6651
PART	27.7512
Logistic	25.3589
<b>Parameter: numIterations</b>	
10	29.6651
5	31.5789
20	29.1866

Πίνακας 3.3.7: Πίνακας αποτελεσμάτων του αλγόριθμου Decorate

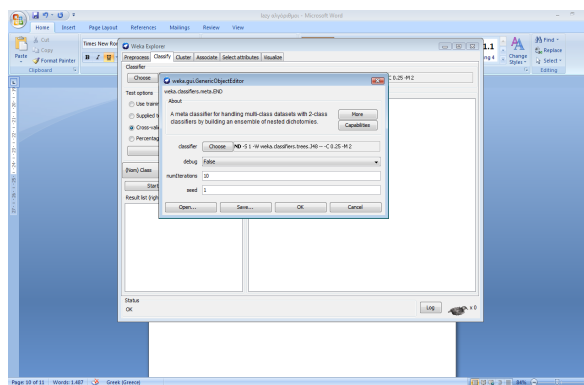
Μετά από πειράματα παρατηρήσαμε ότι η μέθοδος αυτή δίνει κακά αποτελέσματα. Με την τροποποίηση μερικών τιμών των παραμέτρων του συστήματος βλέπουμε ότι αυτές επηρεάζουν αρνητικά το αποτέλεσμα. Το μεγαλύτερο ποσοστό ορθούς ταξινομήσης ήταν 31,58%.

### 3.3.8 Αλγόριθμος 'END'

Ο End (Dong et al., 2005) αποτελεί έναν ταξινομητή μεταμάθησης ο οποίος χρησιμοποιείται για τον χειρισμό δεδομένων που αποτελούνται από πολλαπλές κλάσεις και με τη βοήθεια ταξινομητών δυο κλάσεων δημιουργούν ένα σύνολο από φωλιασμένες διχοτομήσεις.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.8: Ανοικτές παράμετροι του αλγόριθμου END

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

END	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
<i>ND</i>	49.7608
ClassBalancedND	50.7177
DataNearBalancedND	53.11
<b>Parameter: numIterations</b>	
10	49.7608
5	46.4115
20	48.8038
<b>seed</b>	
2	
<b>ClassBalancedND</b>	
<b>numIterations</b>	<b>30</b>
	61.9048

Πίνακας 3.3.8: Πίνακας αποτελεσμάτων του αλγόριθμου END

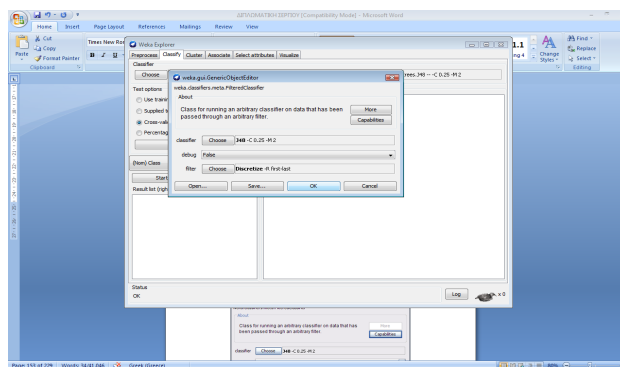
Μετά από τα πειράματα που πραγματοποιήθηκαν στον αλγόριθμο αυτό καταλήξαμε στο συμπέρασμα ότι αποτελεί έναν αρκετά γρήγορο ταξινομητή. Να σημειωθεί ότι οι μονοί ταξινομητές που μπορούσαν να τρέξουν με αυτή τη μέθοδο ήταν οι ND Classifier, ClassBalanceND, DataNearBalanceND οι οποίοι είχαν και την μεγαλύτερη ακρίβεια μέχρι τώρα. Το καλύτερο ποσοστό ακρίβειας ήταν το 61,91%, για τριάντα επαναλήψεις.

### 3.3.9 Αλγόριθμος 'Filtered Classifier'

Ο αλγόριθμος Filtered Classifier έχει την ιδιότητα να εκτελεί έναν αυθαίρετο ταξινομητή πάνω σε δεδομένα που έχουν φιλτραριστεί με μη-καθορισμένο τρόπο. Η δομή του φίλτρου βασίζεται αποκλειστικά στα δεδομένα εκπαίδευσης και στα παραδείγματα δόκιμης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.9: Ανοικτές παράμετροι του αλγόριθμου Filtered Classifier

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί
- Filter: Καθορίζει το φίλτρο που θα χρησιμοποιηθεί.

(Χρησιμοποιούνται διάφορα φίλτρα όπως τα, Add Classification, AttributeSelection, ClassOrder, Discretize, NominalToBinary, PLSFilter, Resample, SpreadSubSample, StratifiedRemoveFolds.)

FilteredClassifier	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
<i>J48</i>	48.8038
Jrip	45.933
Logistic	26.7943
<b>Parameter: filter</b>	
<i>Discretize</i>	48.8038
SMOTE	46.89
Normalize	48.8038
Resample	45.933

Πίνακας 3.3.9: Πίνακας αποτελεσμάτων του αλγόριθμου Filtered Classifier

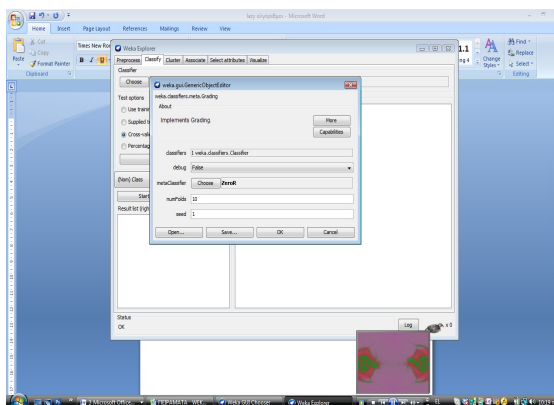
Εξετάστηκαν διάφοροι ταξινομητές και φίλτρα, και τα αποτελέσματα ήταν πολύ κοντά με τους αμιγείς ταξινομητές και κάποιες φορές χειρότερα. Παραθέτουμε ενδεικτικά τα αποτελέσματα που έγιναν για τα πειράματα για τους ταξινομητές J48, Jrip, Logistic, Discretize, SMOTE, Normalize. Να σημειώσουμε ότι η χρησιμοποίηση του φίλτρου 'Normalize' και 'Discretize' είχε την μεγαλύτερη απόδοση (48,80%).

### 3.3.10 Αλγόριθμος Grading

Στον αλγόριθμο Grading (Seewald & Fueemkranz, 2001), ως είσοδο στον ταξινομητή αποτελούν θεμελιώδεις προβλέψεις που έχουν σημειωθεί ως σωστές ή ως λανθασμένες. Για κάθε βασικό ταξινομητή, ένας αλγόριθμος μαθαίνει αυτές τις προβλέψεις.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που

χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.10: Ανοικτές παράμετροι του αλγόριθμου Grading

- Metaclassifier: Καθορίζει τον αλγόριθμο μετά-μάθησης που θα χρησιμοποιηθεί.
- NumFolds: Ορίζει τον αριθμό των Folds για την μέθοδο cross validation.

Grading	CORRECTLY CLASSIFIED
<b>Parameter: metaClassifier</b>	
ZeroR	19.6172
J48	19.6172
Nnge	19.6172
Kstar	19.6172
<b>Parameter: numFolds</b>	
10	19.6172
5	19.6172
20	19.6172

Πίνακας 3.3.10: Πίνακας αποτελεσμάτων του αλγόριθμου Grading

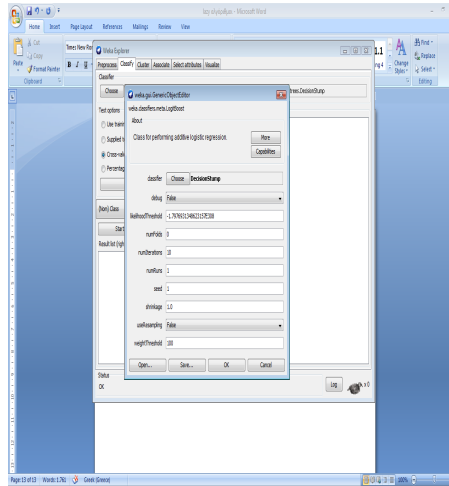
Η αλλαγή των τιμών των βασικών παραμέτρων δε σημείωσε αλλαγές στα ποσοστά ορθούς ταξινόμησης. Τα ποσοστά ακρίβειας ήταν σε πολύ χαμηλά επίπεδα (19,62%).

### 3.3.11 Αλγόριθμος ‘LogitBoost’

Ο αλγόριθμος μετά μάθησης “LogitBoost” (Friedman et al., 1998) αποτελεί έναν πολύ γνωστό αλγόριθμο. Πιο συγκεκριμένα ο αλγόριθμος αυτός εκτελεί επιπρόσθετη λογιστική παλινδρόμηση. Παρόμοια με τον ταξινομητή AdaBoostM1, υπάρχει και εδώ δυνατότητα επιτάχυνσης της εκτέλεσης της ταξινόμησης, προσδιορίζοντας ένα κατώφλι για το τμηματοποιημένο βάρος. Παρέχεται επίσης και μια παράμετρος που βοηθά στην αποφυγή του ταιριάσματος (overfitting).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):





Εικόνα 3.3.11: Ανοικτές παράμετροι του αλγόριθμου LogitBoost

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumFolds: Ορίζει τον αριθμό των Folds για την εσωτερική cross validation μέθοδο που θα πραγματοποιηθεί. Η προεπιλεγμένη τιμή 0 ορίζει ότι δε θα πραγματοποιηθεί cross validation.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

LogitBoost	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
<i>DecisionStump</i>	46.89
MultilayerPerceptron	-
Kstar	39.7129
<b>Parameter: numFolds</b>	
0	46.89
10	47.3684
20	41.1483
<b>Parameter: numIterations</b>	
10	46.89
5	46.4115
20	43.5407

Πίνακας 3.3.11: Πίνακας αποτελεσμάτων του αλγόριθμου LogitBoost

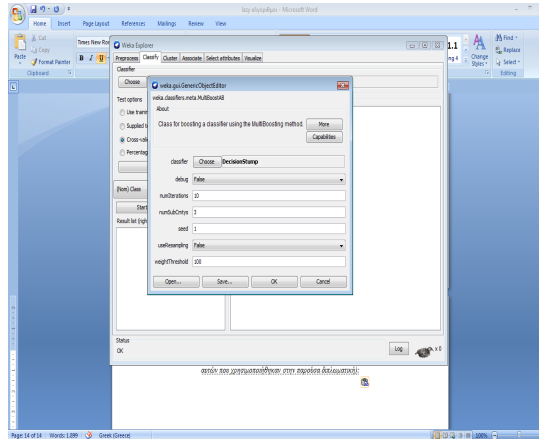
Τα πειράματα που πραγματοποιήθηκαν, έδειξαν ότι με την αλλαγή των τιμών των παραμέτρων επιτεύχθηκε μεγαλύτερο ποσοστό ορθούς ταξινόμησης, σε μια περίπτωση. Το μεγαλύτερο ποσοστό ακρίβειας ήταν 47,37%.

### 3.3.12 Αλγόριθμος ‘MultiBoostAB’

Ο αλγόριθμος MultiBoostAB (Geoffrey & Webb, 2000) αποτελεί μια προέκταση του δημοφιλούς αλγορίθμου AdaBoost. Ο MultiBoostAB μπορεί να θεωρηθεί ως συνδυασμός του AdaBoost με τη μέθοδο ‘Wagging’ και μπορεί να επιτύχει σημαντική μείωση της διακύμανσης των

αποτελεσμάτων συνδυάζοντας τα δυνατά σημεία και από τους δυο προαναφερθέντες αλγόριθμους. Χρησιμοποιώντας τον C4.5 ως βασικό αλγόριθμο μάθησης, η μέθοδος multi boosting μπορεί να βοηθήσει την λήψη αποφάσεων με το μικρότερο σφάλμα.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.12: Ανοικτές παράμετροι του αλγόριθμου MultiBoostAB**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.
- WeightThreshold: Καθορίζει το κατώφλι βάρους, για το τμηματοποιημένο βάρος. (προτείνεται μείωση στην τιμή 50 για να επιταχυνθεί η διαδικασία μάθησης)

MultiBoostAB	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
<i>DecisionStump</i>	29.6651
PART	44.9761
BayesNet	47.8469
<b>Parameter: numIterations</b>	
10	29.6651
5	29.6651
20	29.6651
<b>Parameter: weightThreshold</b>	
100	29.6651
50	29.6651
25	29.6651

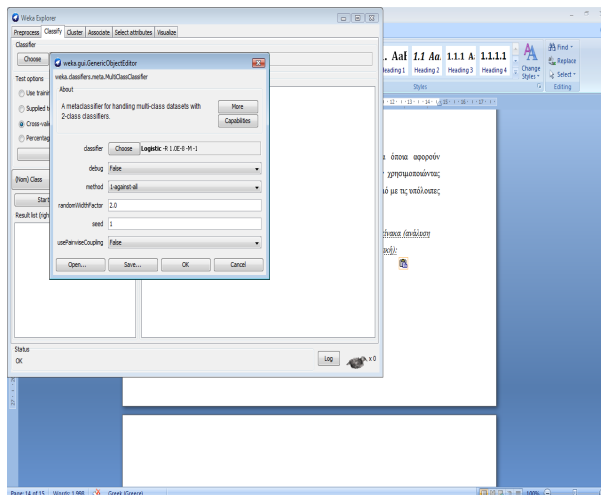
**Πίνακας 3.3.12: Πίνακας αποτελεσμάτων του αλγόριθμου MultiBoostAB**

Παρατηρώντας τον παραπάνω πίνακα 3.3.12 βλέπουμε αξιοσημείωτες διαφορές των αποτελεσμάτων ανάλογα τον ταξινομητή που χρησιμοποιήθηκε ενώ οι υπόλοιπες παράμετροι δεν είχαν καμιά επίδραση στην ακρίβεια του μοντέλου. Τα καλύτερα αποτελέσματα τα έδωσε ο ταξινομητής BayesNet (47,85%).

### 3.3.13 Αλγόριθμος ‘MultiClassClassifier’

Αυτός ο ταξινομητής μπορεί να χειριστεί προβλήματα τα οποία αφορούν πολλαπλές κλάσεις, χρησιμοποιώντας ταξινομητές δυο κλάσεων, καθώς επίσης και κάποια από τις προτεινόμενες τεχνικές από την παράμετρο 'method', σε συνδυασμό με τις υπόλοιπες παραμέτρους.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.13: Ανοικτές παράμετροι του αλγόριθμου Multi Class Classifier**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- Method: Καθορίζει τη μέθοδο που θα χρησιμοποιηθεί για τη μετατροπή των προβλημάτων πολλών κλάσεων, σε προβλήματα δυο κλάσεων.

(Δίδονται οι επιλογές: 1-against-all, random correction code, exhaustive correction code, 1-against-1).

MultiClassClassifier	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
<i>Logistic</i>	28.7081
BayesNet	44.9761
J48	35.4067
<b>Parameter: method</b>	
<i>1-against-all</i>	28.7081
Random correction code	22.488
Exhaustive correction code	26.3158

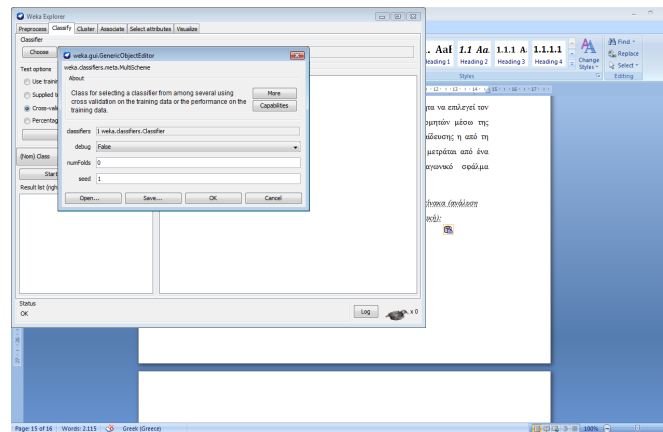
**Πίνακας 3.3.13: Πίνακας Αποτελεσμάτων του αλγόριθμου Multi Class Classifier**

Παρατηρώντας τον παραπάνω πίνακα μπορούμε να καταλήξουμε στο συμπέρασμα ότι υπάρχει σημαντική επίδραση στην ακρίβεια του μοντέλου, για διάφορες τιμές των παραμέτρων με την ακρίβεια όμως να κυμαίνεται σε χαμηλά επίπεδα. Τα καλύτερα αποτελέσματα τα έχουμε με τον ταξινομητή BayesNet (44,98%).

### 3.3.14 Αλγόριθμος 'MultiScheme'

Ο αλγόριθμος μεταμάθησης MultiScheme έχει την ικανότητα να επιλεγεί τον καλύτερο ταξινομητή μέσα από ένα σύνολο υπονήφιων ταξινομητών μέσω της διαδικασίας του cross validation που γίνεται στα δεδομένα εκπαίδευσης ή από τη μετρούμενη απόδοση στα δεδομένα εκπαίδευσης. Η απόδοση μετράται από ένα ποσοστό ακριβείας (ταξινόμηση) η από το μέσο τετραγωνικό σφάλμα (παλινδρόμηση).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.14: Ανοικτές παράμετροι του αλγόριθμου MultiScheme

- NumFolds: Ορίζει τον αριθμό των Folds για την εσωτερική cross validation μέθοδο που θα πραγματοποιηθεί. Η προεπιλεγμένη τιμή 0 ορίζει ότι δε θα πραγματοποιηθεί cross validation.

MultiScheme	CORRECTLY CLASSIFIED
<b>Parameter: numFolds</b>	
0	19.6172
10	19.6172
20	19.6172

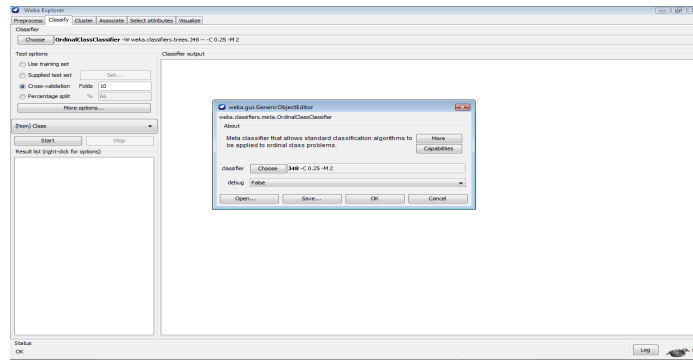
Πίνακας 3.3.14: Πίνακας Αποτελεσμάτων του αλγόριθμου MultiScheme

Τα αποτελέσματα αυτής της τεχνικής κυμαίνονται σε χαμηλά επίπεδα.

### 3.3.15 Αλγόριθμος ‘OrdinalClassClassifier’

Ο ταξινομητής ‘Ordinal Class Classifier’ (Eibe Frank & Mark Hall, 2001), αποτελεί έναν αλγόριθμο μεταμάθησης ο οποίος εφαρμόζει τυπικούς αλγόριθμους ταξινόμησης. Ουσιαστικά η μονή επιλογή που παρέχεται είναι η επιλογή ταξινομητή.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα:



Εικόνα 3.3.15: Ανοικτές παράμετροι του αλγόριθμου OrdinalClassClassifier

OrdinalClassClassifier	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
J48	37.799
PART	35.4067
Kstar	39.2344
RBFNetwork	34.9282

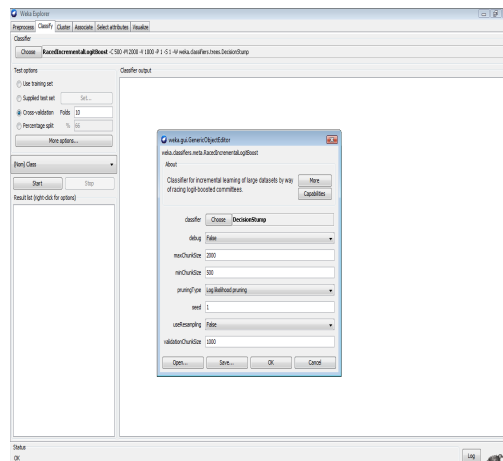
Πίνακας 3.3.15: Πίνακας Αποτελεσμάτων του αλγόριθμου OrdinalClassClassifier

Η αλλαγή του βασικού ταξινομητή μεταβάλλει την ακρίβεια ταξινόμησης περίπου κατά πέντε μονάδες. Παρόλαυτα τα ποσοστά ακρίβειας είναι σχετικά χαμηλά. Το καλύτερο ποσοστό δίνεται με τον ταξινομητή Kstar (39,23%).

### 3.3.16 Αλγόριθμος ‘RacedIncrementalLogitBoost’

Ο αλγόριθμος Raced Incremental Logit Boost (Frank et al., 2002) έχει την ικανότητα να εκπαιδεύεται σταδιακά και «αυξητικά» σε μεγάλες παρτίδες δεδομένων (datasets) με το να κάνει «τρεξίματα» χρησιμοποιώντας την μέθοδο ‘Logit Boost’. Το μέγεθος της παρτίδας ξεκίνα από ένα δεδομένο ελάχιστο αριθμό και διπλασιάζεται μέχρι να φτάσει στο μέγιστο. Συνοψίζοντας αναφέρουμε ότι η μέθοδος αυτή ενδείκνυται για μεγάλα σύνολα δεδομένων.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.16: Ανοικτές παράμετροι του αλγόριθμου RacedIncrementalLogitBoost

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.

- MaxChunkSize: Καθορίζει τον μέγιστο αριθμό των παραδειγμάτων τα όποια θα χρησιμοποιήσει ο βασικός ταξινομητής για την εκπαίδευση του. Το μέγεθος της παρτίδας που θα χρησιμοποιηθεί ξεκίνα με τον αριθμό που έχουμε επιλέξει στην παράμετρο MinChunkSize και διπλασιάζεται έως ότου να γίνει ο αριθμός ίσος με την τιμή που έχει τεθεί στην παράμετρο MaxChunkSize.
- MinChunkSize: Καθορίζει τον ελάχιστο αριθμό των παραδειγμάτων τα όποια θα χρησιμοποιήσει ο βασικός ταξινομητής για την εκπαίδευση του.

RacedIncrementalLogitBoost	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
DecisionStump	19.6172
LinearRegression	19.6172
Kstar	19.6172
<b>Parameter: maxChunkSize</b>	
2000	19.6172
1000	19.6172
5000	19.6172
<b>Parameter: minChunkSize</b>	
500	19.6172
200	19.6172
1000	19.6172

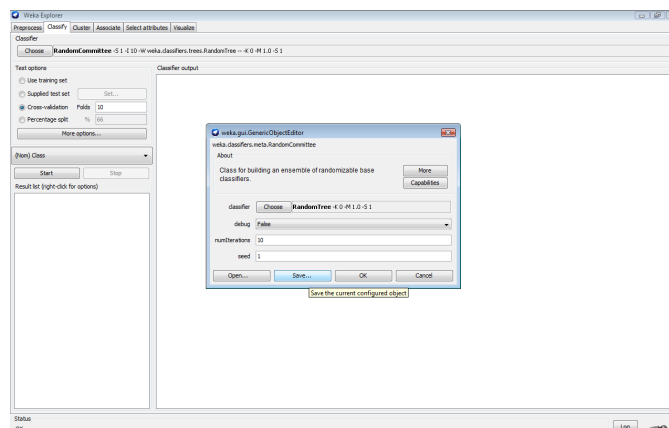
**Πίνακας 3.3.16: Πίνακας Αποτελεσμάτων του αλγόριθμου RacedIncrementalLogitBoost**

Από τα αποτελέσματα του παραπάνω πίνακα, μπορούμε να δούμε ότι η λειτουργία του αλγόριθμου είναι αρκετά ‘ανελαστική’, δηλαδή δεν επηρεάζονται τα ποσοστά ορθούς ταξινόμησης (τα οποία παραμένουν και σε χαμηλά επίπεδα), όταν μεταβληθούν οι τιμές ορισμένων παραμέτρων.

### 3.3.17 Αλγόριθμος ‘RandomCommittee’

Ο αλγόριθμος ταξινόμησης Random Committee αποτελεί έναν αρκετά απλό αλγόριθμο ο οποίος δημιουργεί ένα σύνολο από βασικούς τυχαίους ταξινομητές οι όποιοι προκύπτουν από διαφορετικούς σπόρους τυχαίων αριθμών (βασισμένοι όμως στα ίδια δεδομένα).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.17: Ανοικτές παράμετροι του αλγόριθμου RandomCommittee**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.

- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

RandomCommittee	CORRECTLY CLASSIFIED
<b>Parameter:</b> classifier	
RandomTree	33.4928
MultilayerPerceptron	-
<b>Parameter:</b> numIterations	
	10
	5
	20
	33.4928
	31.1005
	27.7512

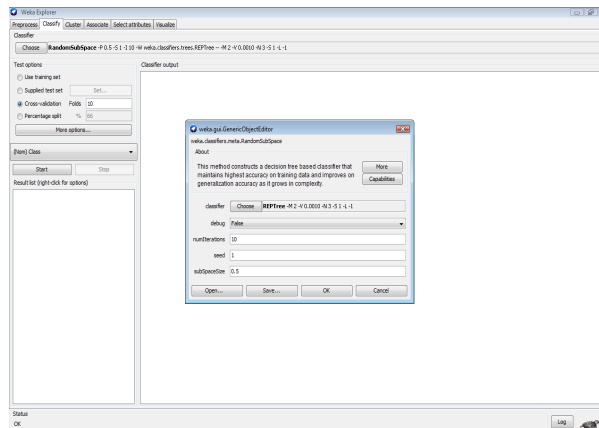
**Πίνακας 3.3.17: Πίνακας Αποτελεσμάτων του αλγόριθμου RandomCommittee**

Αλλάζοντας τον αριθμό των επαναλήψεων βλέπουμε τα αποτελέσματα να επηρεάζονται αρνητικά ενώ κυμαίνονται και σε χαμηλά επίπεδα.

### 3.3.18 Αλγόριθμος ‘RandomSubSpace’

Ο αλγόριθμος μετά-μάθησης Random Sub Space (Tim Kam Ho, 1998) κατασκευάζει ένα δένδρο απόφασης του βασικού ταξινομητή ο οποίος διατηρεί υψηλή ακρίβεια στα δεδομένα εκπαίδευσης και βελτιώνει τη γενικευμένη ακρίβεια στα πιο περίπλοκα δεδομένα. Ο ταξινομητής αποτελείται από πολλαπλά δένδρα συστηματικά κατασκευασμένα από ψευδοτυχαία επιλογή υποσυνόλων των συστατικών του διανύσματος.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.18: Ανοικτές παράμετροι του αλγόριθμου RandomSubSpace**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.



RandomSubSpace	CORRECTLY CLASSIFIED
<b>Parameter:</b> classifier	
<i>REPTree</i>	25.8373
J48	45.4545
Logistic	-
BayesNet	42.1053
<b>Parameter:</b> numIterations	
	10
	20
	5
	25.8373
	24.8804
	23.445

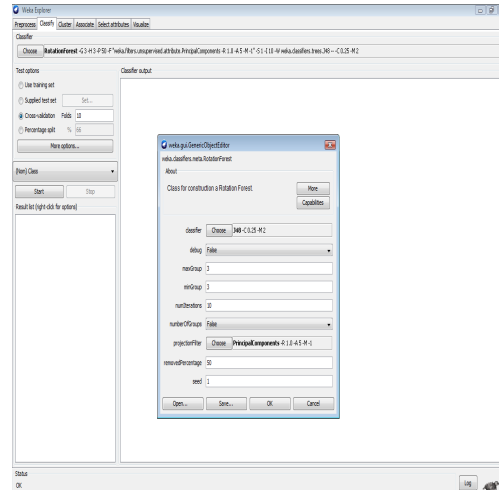
**Πίνακας 3.3.18: Πίνακας Αποτελεσμάτων του αλγόριθμου RandomSubSpace**

Μετά από αρκετά πειράματα παρατηρήσαμε ότι οι ταξινομητές που χρησιμοποιήθηκαν επηρέασαν σημαντικά τα αποτελέσματα. υψηλότερα ποσοστά ορθούς ταξινόμησης είχαμε στο με τον ταξινομητή J48 (45,46%).

### 3.3.19 Αλγόριθμος ‘RotationForest’

Είναι ο αλγόριθμος ο οποίος μπορεί να κάνει την ταξινόμηση και την οπισθοδρόμηση, ανάλογα με την εκπαιδευόμενη βάση.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.19: Ανοικτές παράμετροι του αλγόριθμου RotationForest**

- Classifier: Καθορίζει τον βασικό ταξινομητή που θα χρησιμοποιηθεί.
- NumIterations: Καθορίζει τον αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

RotationForest	CORRECTLY CLASSIFIED
<b>Parameter: classifier</b>	
J48	42.1053
Jrip	47.3684
SimpleLogistic	48.8038
<b>Parameter: numIterations</b>	
10	42.1053
20	46.89
5	40.1914

**Πίνακας 3.3.19: Πίνακας Αποτελεσμάτων του αλγόριθμου RotationForest**

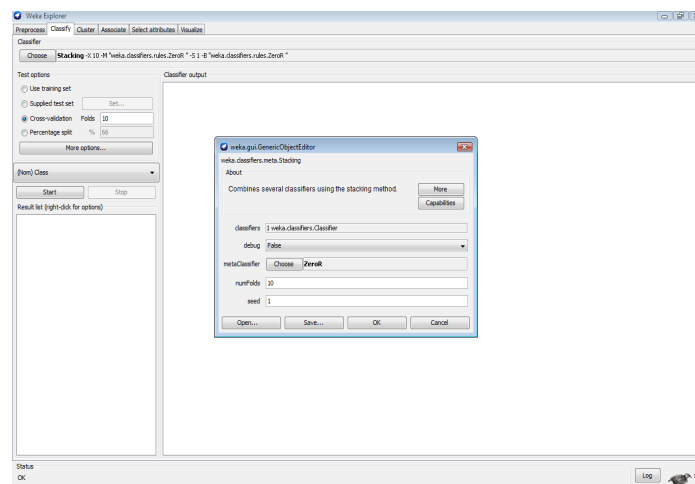
Βάσει των αποτελεσμάτων του πίνακα 3.3.19, παρατηρούμε μεταβολές με την εφαρμογή εναλλακτικών τιμών για τις παραμέτρους του αλγορίθμου. Τα καλύτερα αποτελέσματα ήταν (48,80%).

### 3.3.20 Αλγόριθμος ‘Stacking’

Ο αλγόριθμος Stacking (David & Wolpert, 1992) έχει την ικανότητα να συνδυάζει ταξινομητές χρησιμοποιώντας τη μέθοδο ‘stacking’, για τα προβλήματα ταξινόμησης και παλινδρόμησης.

Αν και έχουν ξαναπαρουσιαστεί, παρακάτω παρουσιάζουμε τις ανοικτές παραμέτρους του αλγορίθμου ξανά, για λόγους ευκολίας του αναγνώστη.

Οι παράμετροι του αλγορίθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.20: Ανοικτές παράμετροι του αλγορίθμου ‘Stacking’**

- Metaclassifier: Καθορίζει τον αλγόριθμο μετά-μάθησης που θα χρησιμοποιηθεί.
- NumFolds: Ορίζει τον αριθμό των Folds για την εσωτερική cross validation μέθοδο που θα πραγματοποιηθεί. Η προεπιλεγμένη τιμή 0 ορίζει ότι δε θα πραγματοποιηθεί cross validation.

Stacking	CORRECTLY CLASSIFIED
<b>Parameter: metaClassifier</b>	
ZeroR	19.6172
J48	19.6172
BayesNet	19.6172
<b>Parameter: numFolds</b>	
10	19.6172
5	19.6172
20	19.6172

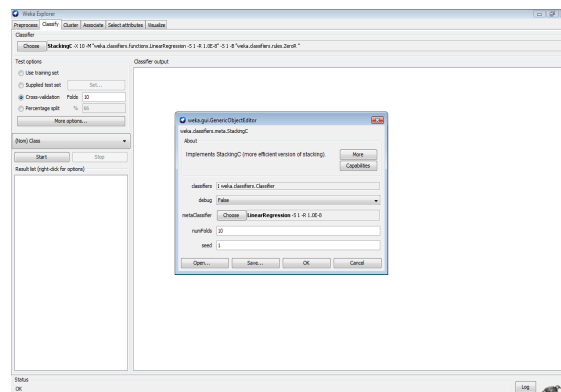
**Πίνακας 3.3.20: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Stacking’**

Στον πίνακα 3.3.20 παρατηρώντας τα αποτελέσματα με βασικό ταξινομητή τον ZeroR και διαφόρους άλλους μετά- ταξινομητές το αποτέλεσμα παρέμεινε ίδιο.

### 3.3.21 Αλγόριθμος ‘StackingC’

Ο αλγόριθμος Stacking C (Seewald, 2002) αποτελεί μια πιο βελτιωμένη έκδοση του αλγορίθμου ‘Stacking’. Σε αυτή την εκδοχή, απαιτείται ο αλγόριθμος μεταμάθησης να έχει αριθμητικό σχέδιο πρόβλεψης. Οι ανοικτές παράμετροι όπως παρατηρούμε είναι ίδιοι με αυτές του απλού αλγορίθμου ‘Stacking’ που παρουσιάστηκε παραπάνω. Τα αποτελέσματά ήταν ίδια με τον παραπάνω αλγόριθμο.

Οι παράμετροι του αλγορίθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.3.21: Ανοικτές παράμετροι του αλγόριθμου ‘StackingC’**

- Metaclassifier: Καθορίζει τον αλγόριθμο μεταμάθησης που θα χρησιμοποιηθεί.
- NumFolds: Ορίζει τον αριθμό των Folds για την εσωτερική cross validation μέθοδο που θα πραγματοποιηθεί. Η προεπιλεγμένη τιμή 0 ορίζει ότι δε θα πραγματοποιηθεί cross validation.

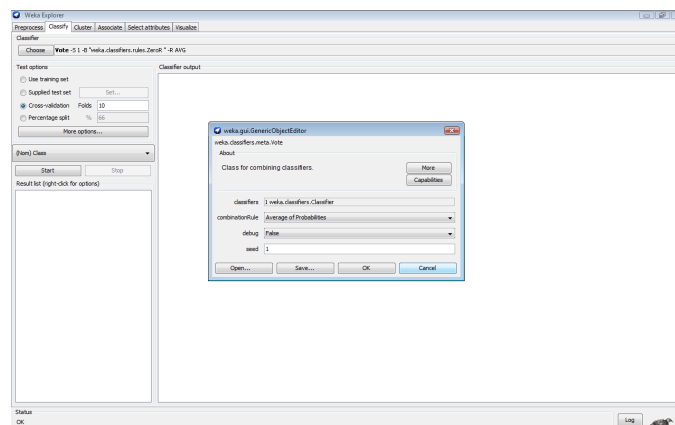
StackingC	CORRECTLY CLASSIFIED
<b>Parameter:</b> metaClassifier	
<i>Linearregression</i>	19.6172
J48	-
<b>Parameter:</b> numFolds	
10	19.6172
5	19.6172
20	19.6172

Πίνακας 3.3.21: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘StackingC’

### 3.3.22 Αλγόριθμος ‘Vote’

Ο αλγόριθμος ‘Vote’ (Ludmila Kuncheva, 2004) αποτελεί μια βασική μέθοδο για τον συνδυασμό ταξινομητών χρησιμοποιώντας μέσους όρους των πιθανοτήτων εκτίμησης για την ταξινόμηση ή με αριθμητικές προβλέψεις για την παλινδρόμηση. Οι παράμετροι που προσφέρει ο αλγόριθμος είναι η επιλογή ταξινομητών και η επιλογή συνδυασμών κανόνων που χρησιμοποιούνται. Πιο συγκεκριμένα αυτή η παράμετρος μας δίνει τις επιλογές που περιγράφονται παρακάτω.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.3.22: Ανοικτές παράμετροι του αλγόριθμου ‘Vote’

- **Combination Rules:** Καθορίζει την επιλογή συνδυασμών κανόνων που χρησιμοποιούνται με τις επιλογές
  - Average Of Probabilities (Μέσος Όρος πιθανοτήτων)
  - Product Of Probabilities (Παραγωγή πιθανοτήτων)
  - Majority Voting (Πλειοψηφία πληροφορίας)
  - Minimum Probability (Ελάχιστη πιθανότητα)
  - Maximum Probability (Μεγίστη πιθανότητα)
  - Median (Διάμεσος)

Vote	CORRECTLY CLASSIFIED
<b>Parameter:</b> combinationRule	
<i>Average of Probabilities</i>	19.6172
Product of Probabilities	19.6172
Majority Voting	19.6172

Πίνακας 3.3.22: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Vote’

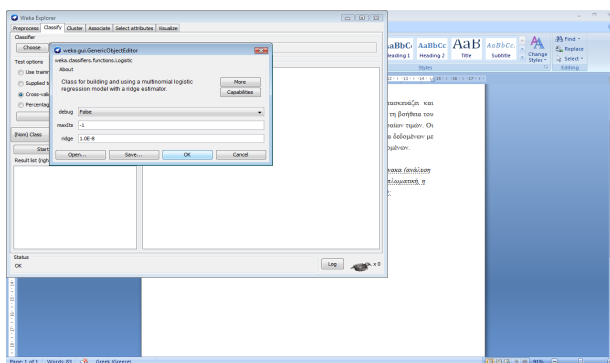
Η μέθοδος αυτή χρησιμοποιώντας ένα ταξινομητή η ένα συνδυασμό ταξινομητών δε φάνηκε να προσφέρει κάτι παραπάνω από το να χρησιμοποιηθεί απλά ένας αμιγής ταξινομητής. Τα αποτελέσματα είναι παρόμοια με εκείνα παραπάνω αλγορίθμων.

### 3.4 ΣΥΝΑΡΤΗΣΕΙΣ ΠΡΟΣΕΓΓΙΣΗΣ (functions)

#### 3.4.1 Λογιστική Παλινδρόμηση (Logistic)

Ο αλγόριθμος αυτός (Cessie & Houwelingen, 1992), κατασκευάζει και χρησιμοποιεί ένα γραμμικό υπόδειγμα λογιστικής παλινδρόμησης, με τη βοήθεια του εκτιμητή ridge estimator, ο οποίος είναι ένας ακριβής εκτιμητής ακραίων τιμών. Οι ακραίες τιμές (Margaret H.Dunham, 2004) ή σημεία είναι δείγματα δεδομένων με τιμές πολύ διαφορετικές από τις τιμές του υπολοίπου συνόλου δεδομένων.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών των παραμέτρων που χρησιμοποιήθηκαν στην παρούσα διπλωματική, η επανάληψη αναφέρεται προς ευκολία του αναγνώστη):



Εικόνα 3.4.1: Ανοικτές παράμετροι του αλγόριθμου ‘logistic’

- MaxIts: Ορίζει το μέγιστο αριθμό των επαναλήψεων που θα πραγματοποιηθούν.

Logistic	CORRECTLY CLASSIFIED
<b>Parameter: maxIts</b>	
-1	-
5	29.1866
10	27.7512

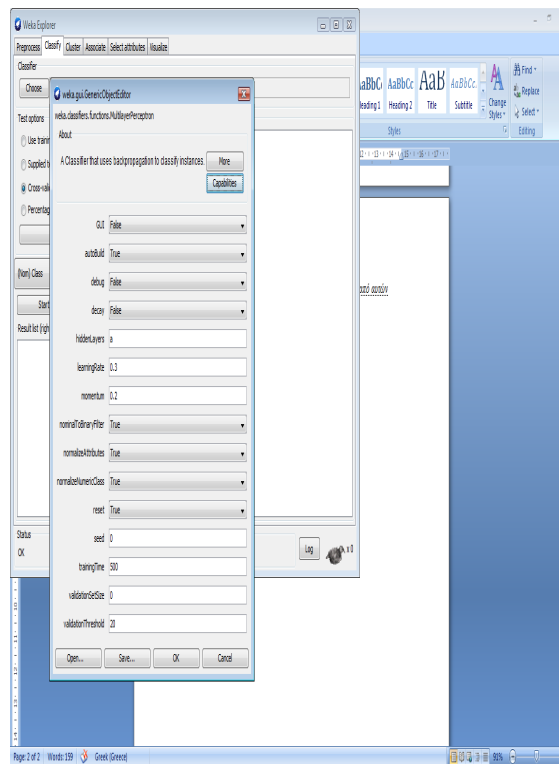
Πίνακας 3.4.1: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘logistic’

Παρατηρώντας τον παραπάνω πίνακα 3.4.1 βλέπουμε ότι η μεταβολή της παραμέτρου maxIts ίση με 5 και 10, επίσης μας έδωσε χειρότερα αποτελέσματα.

#### 3.4.2 Νευρωνικό Δίκτυο τύπου ‘Multilayer Perceptron’

Το συγκεκριμένο δίκτυο λειτουργεί ως ταξινομητής που χρησιμοποιεί ‘backpropagation’ να ταξινομήσει τις περιπτώσεις. Το δίκτυο αυτό μπορεί να κατασκευαστεί με το χέρι, που δημιουργήθηκε από έναν αλγόριθμο ή και τα δύο. Οι κόμβοι στο δίκτυο αυτό είναι σιγμοειδής.

Οι παράμετροι του αλγόριθμου αυτού και η ανάλυση κάποιων από αυτών παρατίθενται στον παρακάτω πίνακα:



Εικόνα 3.4.2: Ανοικτές παράμετροι του αλγόριθμου ‘Multilayer Perceptron’

- LearningRate: Δείχνει το ποσοστό μάθησης για τον backpropagation αλγόριθμο. Το πεδίο τιμών είναι από 0 έως 1.
- Momentum: Ορίζει τον ρυθμό ταχύτητας που εφαρμόζεται στα βάρη κατά τη διάρκεια της ενημέρωσης για τον backpropagation αλγόριθμο. Το πεδίο τιμών είναι από 0 έως 1.
- TrainingTime: Καθορίζει τον αριθμό των «εποχών» που θα πραγματοποιηθεί η εκπαίδευση.

MultilayerPerceptron	CORRECTLY CLASSIFIED
<b>Parameter: learningRate</b>	
0.3	48.3254
0.05	-
0.8	-
<b>Parameter: momentum</b>	
0.2	48.3254
0.05	-
0.8	-
<b>Parameter: trainingTime</b>	
500	48.3254
250	-
1000	-

Πίνακας 3.4.2: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Multilayer Perceptron’

όπως βλέπουμε και στον παραπάνω πίνακα, με την αλλαγή των παραμέτρων δεν καταφέραμε να πάρουμε κάποιο αποτέλεσμα. Η ακρίβεια που μας έδωσε ο αλγόριθμος με τις αρχικές ρυθμίσεις

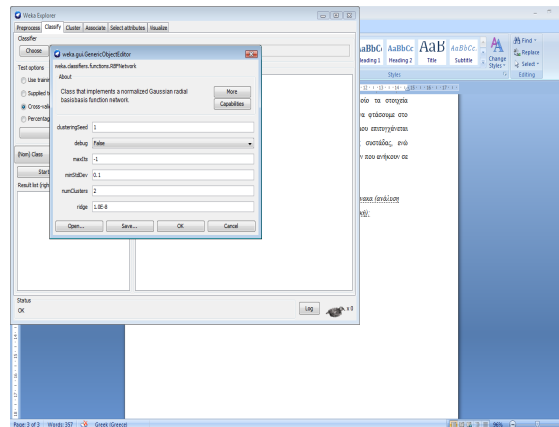
ήταν 48,33%.

### 3.4.3 Νευρωνικό δίκτυο τύπου RBF Network

Ο αλγόριθμος "RBFNetwork" του Bishop είναι ένα δίκτυο που εφαρμόζει μια γκαουσιανή συνάρτηση ακτινικής διάταξης (είναι μια συνάρτηση πραγματικών τιμών, της οποίας η τιμή εξαρτάται μόνο από την απόσταση από το αρχικό σημείο). Χρησιμοποιεί τον αλγόριθμο ταξινόμησης k-means για να ορίσει τις ομάδες (clusters) και 'μαθαίνει' είτε μια λογιστική παλινδρόμηση (προβλήματα διακριτού χώρου) είτε μια γραμμική παλινδρόμηση (προβλήματα συνεχούς χώρου).

Ο k-mean είναι ένας επαναληπτικός αλγόριθμος στον οποίο τα στοιχεία μετακινούνται μεταξύ των διαφόρων συνόλων συστάδων μέχρι να φτάσουμε στο επιθυμητό σύνολο συστάδων. Μέσω του συγκεκριμένου αλγορίθμου επιτυγχάνεται μεγάλος βαθμός ομοιότητας μεταξύ των στοιχείων της ίδιας συστάδας, ενώ ταυτόχρονα επιτυγχάνεται και μεγάλη διαφορά μεταξύ των στοιχείων που ανήκουν σε διαφορετικές συστάδες.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.4.3: Ανοικτές παράμετροι του αλγόριθμου 'RBF Network'

- MaxIts: Ορίζει τον μέγιστο αριθμό των επαναλήψεων που θα πραγματοποιήσει η λογιστική παλινδρόμηση.
- NumClusters: Ορίζει τον αριθμό των clusters που παράγονται για τον k-mean.

RBFNetwork	CORRECTLY CLASSIFIED
<b>Parameter: maxIts</b>	
-1	39.7129
5	41.6268
10	43.5407
<b>Parameter: numClusters</b>	
2	39.7129
5	39.2344
10	40.6699

Πίνακας 3.4.3: Πίνακας Αποτελεσμάτων του αλγόριθμου 'RBF Network'

Πραγματοποιώντας διάφορα πειράματα με τη μέθοδο αυτή παρατηρήσαμε πως οι μεταβολές των

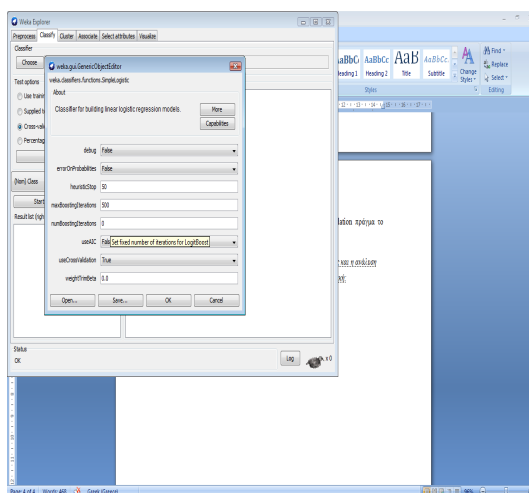


παραμέτρων επηρεάζουν σε κάποιο βαθμό τα αποτελέσματα. Όσον αφορά τον χρόνο εκτέλεσης είναι από τις πιο γρήγορες στην κατηγορία της (functions) αλλά συγκρινόμενη με άλλες κατηγορίες μπορεί να θεωρηθεί σχετικά αργή. Για παράδειγμα ο δενδρικός ταξινομητής J48 αποδίδει παρόμοιες και πολλές φορές καλύτερα αποτελέσματα σε πολύ λιγότερο χρόνο. Επίσης ένα ακόμα μειονέκτημα για τη μέθοδο μας είναι ότι δεν εξάγει κάποιο οπτικοποιήσιμο αποτέλεσμα που θα μπορούσε να βοηθήσει στην παράγωγη κάποιων κανόνων. Η μέγιστη ακρίβεια του αλγορίθμου "RBFNetwork" ήταν 43,54%.

### 3.4.4 Απλή Λογιστική Παλινδρόμηση (Simple Logistic)

Ο αλγόριθμος Simple Logistic κατασκευάζει γραμμικά λογιστικά μοντέλα παλινδρόμησης (Niels Landwehr et al.,2005). Η μέθοδος 'LogitBoost' μαζί με συναρτήσεις απλής παλινδρόμησης χρησιμοποιούνται ως βασικοί αλγόριθμοι μάθησης για τα μοντέλα. Ο ιδανικός αριθμός επαναλήψεων του 'LogitBoost' που θα πραγματοποιηθούν προκύπτει από τη διαδικασία του cross validation πράγμα το οποίο οδηγεί σε αυτόματη επιλογή χαρακτηριστικών.

Στον πίνακα φαίνονται οι παράμετροι του αλγορίθμου αυτού καθώς και η ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



Εικόνα 3.4.4: Ανοικτές παράμετροι του αλγορίθμου 'Simple Logistic'

- MaxBoostingIterations: Καθορίζει τον μέγιστο αριθμό των επαναλήψεων του LogitBoost. Προεπιλεγμένη τιμή είναι η τιμή 500. Ανάλογα με το μέγεθος των δεδομένων μικρότερες ή μεγαλύτερες τιμές μπορούν να δοκιμαστούν.
- WeightTrimBeta: Καθορίζει την τιμή beta για το βάρος του τμήματος της μεθόδου LogitBoost. Θέτοντας την τιμή 0 δεν χρησιμοποιείται βάρος τμήματος.

SimpleLogistic	CORRECTLY CLASSIFIED
<b>Parameter: maxBoostingIterations</b>	
500	42.5837
200	42.5837
1000	42.5837
<b>Parameter: weightTrimBeta</b>	
0	42.5837
0.1	43.0622
0.5	34.9282

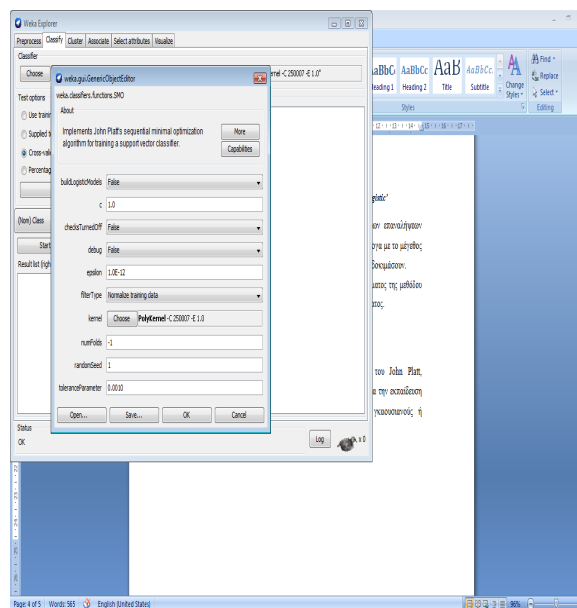
**Πίνακας 3.4.4: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Simple Logistic’**

Μετά από διάφορους πειραματισμούς καταλήξαμε στο συμπέρασμα ότι τυχών μεταβολές στις ανοικτές παραμέτρους του αλγορίθμου όπως έτειναν να μειώσουν την ακρίβεια του μοντέλου ή να την κρατούν απλά σταθερή.

### 3.4.5 Αλγόριθμος Ελάχιστης Διαδοχικής Βελτιστοποίησης SMO

Ο αλγόριθμος “SMO” (Sequential Minimal Optimization) του John Platt, εφαρμόζει τον αλγόριθμο διαδοχικής ελαχίστης βελτιστοποίησης για την εκπαίδευση ενός ταξινομητή διανύσματος υποστήριξης, χρησιμοποιώντας γκαουσσιανούς ή πολυωνμικούς πυρήνες (Platt 1998 & Keerthi et al.,2001).

Στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγόριθμου αυτού καθώς και η ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



**Εικόνα 3.4.5: Ανοικτές παράμετροι του αλγόριθμου ‘SMO’**

- NumFolds: Ορίζει τον αριθμό των Folds για την μέθοδο cross validation.
- Kernel: Καθορίζει το ποιος πυρήνας θα χρησιμοποιηθεί. Δίδονται κάποιες επιλογές σε πυρήνες όπως οι πυρήνες PolyKernel, NormalizedPolyKernel, RBFkernel, StringKernel, MIPolykernel.

SMO	CORRECTLY CLASSIFIED
<b>Parameter: numFolds</b>	
-1	44.9761
5	44.9761
10	44.9761
<b>Parameter: kernel</b>	
<i>PolyKernel</i>	44.9761
NormalizedPolyKernel	42.1053
RBFKernel	34.4498

Πίνακας 3.4.5: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘SMO’

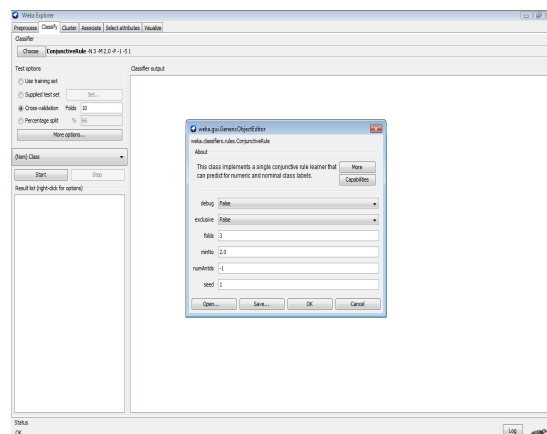
Παρατηρώντας τον πίνακα 3.4.5 βλέπουμε ότι στην επιλογή ‘numFolds’ τα δεδομένα έχουν σταθερές τιμές ότι τιμή και αν βάλουμε εμείς. Το μεγαλύτερο ποσοστό ήταν 46,98%. Στον πυρήνα Kernel, η ανοιχτή παράμετρος ‘Normalized polykernel’ απέδωσε μικρότερα ποσοστά ακρίβειας από εκείνα που έδωσε το πρόγραμμα WEKA σαν προεπιλεγμένα ενώ, η ανοιχτή παράμετρος ‘RBF Kernel’ έδωσε πολύ μικρότερα σε σχέση με τις παραπάνω δύο παραμέτρους όσες φορές και αν επαναλήφθηκε το πείραμα.

### 3.5 ΑΛΓΟΡΙΘΜΟΙ ΕΞΑΓΩΓΗΣ ΚΑΝΟΝΩΝ (rules)

#### 3.5.1 Συζευκτικός Κανόνας ‘Conjunctive Rule’

Ο αλγόριθμος “Conjunctive Rule” εφαρμόζει έναν απλό συνδυαστικό κανόνα. Ο κανόνας αυτός μπορεί να κάνει προβλέψεις για αριθμητικές και ονομαστικές κλάσεις. Ένας κανόνας αποτελείται από πρόγονους και την αξία της κλάσης για την ταξινόμηση. Σε αυτή την περίπτωση η συνέπεια είναι η διανομή των διαθεσίμων κατηγοριών στο σύνολο των διαθεσίμων δεδομένων (ή ο μέσος για αριθμητικά δεδομένα). Έτσι ο αλγόριθμος αυτός επιλεγεί έναν πρόγονο υπολογίζοντας το information gain του κάθε προγόνου, και κλαδεύει τον παραγόμενο κανόνα χρησιμοποιώντας τη μέθοδο reduced error pruning ή τη μέθοδο pre-pruning. Για την ταξινόμηση, οι πληροφορίες ενός προγόνου είναι ο σταθμισμένος μέσος όρος των εντροπιών των στοιχείων που καλύπτονται όπως και αυτών που δεν καλύπτονται από τον κανόνα. Για την παλινδρόμηση, οι πληροφορίες είναι ο σταθμισμένος μέσος όρος των μέσων τετραγωνικών σφαλμάτων των στοιχείων που καλύπτονται όπως και αυτών που δεν καλύπτονται από τον κανόνα, ομοίως.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών των παραμέτρων που χρησιμοποιήθηκαν στην παρούσα διπλωματική, η επανάληψη αναφέρεται προς ευκολία του αναγνώστη):



Εικόνα 3.5.1: Ανοικτές παράμετροι του αλγόριθμου ‘conjunctive rule’

- Folds: Προσδιορίζει το ποσό των δεδομένων που χρησιμοποιούνται για κλάδεμα. Το ένα fold χρησιμοποιείται για κλάδεμα και τα υπόλοιπα για ανάπτυξη των κανόνων.
- NumAntds: Χρησιμοποιείται για να ορίσει τον αριθμό των προγονών που επιτρέπονται σε έναν κανόνα εάν χρησιμοποιείται η τακτική pre-pruning. Εάν στη παράμετρο αυτή τεθεί τιμή διαφορετική του -1, τότε θα χρησιμοποιηθεί η τακτική κλαδέματος pre-pruning, διαφορετικά θα χρησιμοποιηθεί η τεχνική reduced error pruning.

ConjunctiveRule	CORRECTLY CLASSIFIED
<b>Parameter: folds</b>	
3	27.2727
10	26.3158
20	22.0096
<b>Parameter: numAntds</b>	
-1	27.2727
5	29.6651

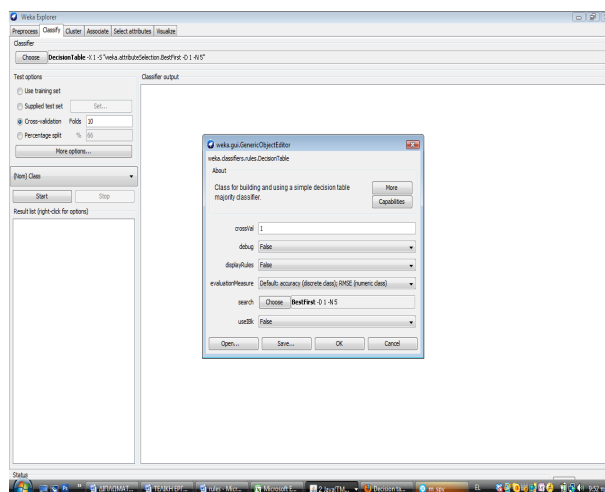
**Πίνακας 3.5.1: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘conjunctive rule’**

Όπως παρατηρούμε από τους παραπάνω πίνακες η μέθοδος αυτή υστερεί πολύ σε ικανότητα ακριβούς ταξινόμησης. Οι μεταβολές των ανοικτών παραμέτρων αύξησαν την ακρίβεια του μοντέλου από 27,27% σε 29,57%. Το μοναδικό θετικό στοιχείο αυτής της μεθόδου είναι η ταχύτατη εκτέλεση του αλγορίθμου.

### 3.5.2 Αλγόριθμος κανόνων τύπου ‘DecisionTable’

Ο αλγόριθμος κανόνων τύπου ‘DecisionTable’ είναι ένας δισδιάστατος πίνακας που δείχνει τη δράση που πρέπει να παρθεί μετά από μια σειρά σχετικών αποφάσεων. Πιο συγκεκριμένα όταν ένας αλγόριθμος περιλαμβάνει περισσότερες από δύο ή τρεις ένθετες αποφάσεις, τότε ο αλγόριθμος δίνει μια σαφή και συνοπτική εικόνα της λογικής. Πριν από τη δημιουργία ενός ‘DecisionTable’, πρέπει να κατανοηθεί ο αλγόριθμος ή η διαδικασία. Οι απαραίτητες πληροφορίες θα πρέπει να καταρτιστούν από την άμεση παρατήρηση, που προέρχονται από υπάρχοντα έγγραφα, ή που προέρχεται από τον ορισμό του προβλήματος ( Brown, Kahr and Peterson, 1974).

Οι παράμετροι του αλγορίθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών των παραμέτρων που χρησιμοποιήθηκαν στην παρούσα διπλωματική, η επανάληψη αναφέρεται προς ευκολία του αναγνώστη):



**Εικόνα 3.5.2: Ανοικτές παράμετροι του αλγορίθμου ‘decision table’**

- search: χρησιμοποιείται για να επιλέξει τη μέθοδο για την αναζήτηση δομών δικτύων.

DecisionTable	CORRECTLY CLASSIFIED
<b>Parameter: search</b>	
<i>BestFit</i>	48.8038
GeneticSearch	34.9282
LinearForwardSelection	48.8038
<b>Parameter: displayRules</b>	
TRUE	48.8038

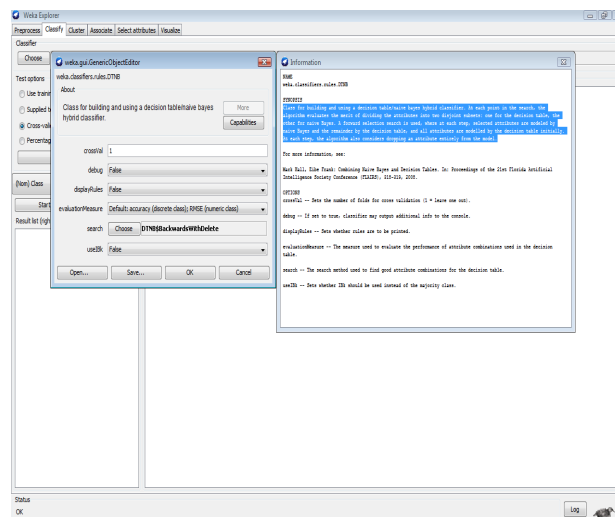
Πίνακας 3.5.2: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘decision table’

Παρατηρώντας τον πίνακα 3.5.2 βλέπουμε πως η αλλαγή των τιμών των παραμέτρων δεν επηρέασε τα αποτελέσματα ή τα επηρέασε αρνητικά. Το καλύτερο αποτέλεσμα της μεθόδου ήταν 48,80%

### 3.5.3 Αλγόριθμος κανόνων τύπου ‘DTNB’

Ο αλγόριθμος κανόνων τύπου ‘DTNB’ αξιολογεί την αξία της διαίρεσης των χαρακτηριστικών σε δύο υποσύνολα ασυνεχή: ένα για το ‘Decision Table’, το άλλο για Naive Bayes. Αποτέλεσμα αυτού σε κάθε βήμα, επιλέγονται χαρακτηριστικά έτσι ώστε να μοντελοποιηθεί με Naive Bayes και το υπόλοιπο από το ‘Decision Table’ (Hall and Frank, 2008).

Για ευκολία του αναγνώστη στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγόριθμου και η ανάλυση των παραμέτρων που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



Εικόνα 3.5.3: Ανοικτές παράμετροι του αλγόριθμου ‘DTNB’

- Search: Η μέθοδος ‘αναζήτησης’ χρησιμοποιείται για να βρούμε τους πιο ιδανικούς συνδυασμούς.
- DisplayRules: Ορίζει τους κανόνες που θα προκύψουν.

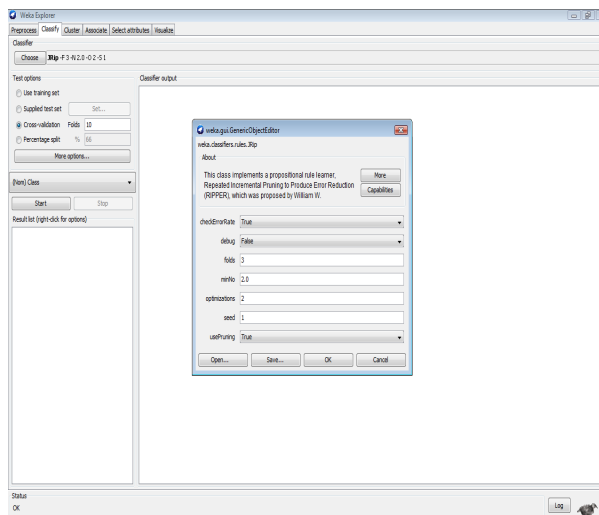
Πίνακας 3.5.3: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘DTNB’

DTNB	CORRECTLY CLASSIFIED
<b>Parameter: displayRules</b>	
<i>TRUE</i>	
<b>Parameter: search</b>	
<i>DTNB\$BackwardsWithDelete</i>	45.933

### 3.5.4 Αλγόριθμος κανόνων τύπου ‘JRip’

Ο αλγόριθμος αυτός εφαρμόζει την μέθοδο RIPPER (Repeated Incremental Pruning to Produce Error Prediction) που προτάθηκε από τον William W.Cohen το 1995 και συμπεριλαμβάνει εφευρετικές βελτιστοποιήσεις από σύνολα κανόνων. Αποτελείται από 4 βασικά σταδία. Φάση κατασκευής, φάση επέκτασης, φάση κλαδέματος και φάση βελτιστοποίησης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.5.4: Ανοικτές παράμετροι του αλγόριθμου ‘JRip’

- Folds: Καθορίζει τον αριθμό των δεδομένων που χρησιμοποιούνται για ‘κλάδεμα’. Κάθε φορά χρησιμοποιείται μόνο ένα για το κλάδεμα, και τα υπόλοιπα για την εξαγωγή των κανόνων.
- Optimizations: Καθορίζει τον αριθμό των optimization runs (επαναλήψεων βελτιστοποίησης).

Jrip	CORRECTLY CLASSIFIED
<b>Parameter: folds</b>	
3	45.933
10	39.2344
20	40.6699
<b>Parameter: optimizations</b>	
2	45.933
5	42.5837
10	44.0191

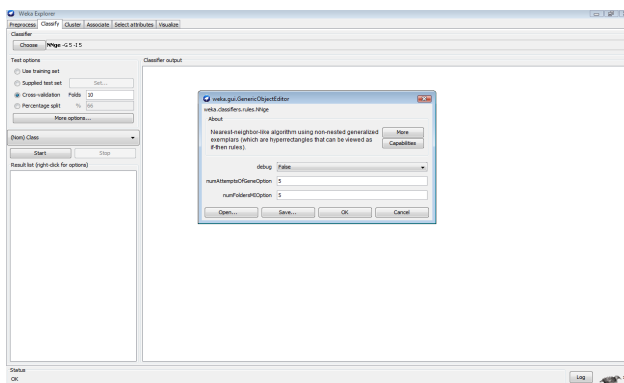
Πίνακας 3.5.4: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘JRip’

Ο JRip αποτελεί έναν αρκετά γρήγορο ταξινομητή ο οποίος όμως μειώνει τα ποσοστά όταν αλλάζουμε τις τιμές των παραμέτρων.

### 3.5.5 Αλγόριθμος κανόνων τύπου ‘NNge’

Ο αλγόριθμος NNge είναι μια nearest-neighbor μέθοδος η οποία παράγει κανόνες της μορφής if-then χρησιμοποιώντας γενικευμένα υποδείγματα κανόνων (Brent,1995).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα και γίνεται αναφορά αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



Εικόνα 3.5.5: Ανοικτές παράμετροι του αλγόριθμου ‘NNge’

- NumAttemptsOfGeneOption: Θέτει τον αριθμό των προσπαθειών για γενίκευση
- NumFoldersMIOption: Θέτει τον αριθμό των πληροφοριών για τις κοινές πληροφορίες.

Nnge	CORRECTLY CLASSIFIED
<b>Parameter: numAttemptsOfGeneOption</b>	
5	29.1866
10	29.1866
20	29.1866
<b>Parameter: numFoldersMIOption</b>	
5	29.1866
10	29.1866
20	29.1866

Πίνακας 3.5.5: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Nnge’

Παρατηρήθηκε ότι ο NNge αποτελεί έναν ταξινομητή ο οποίος χαρακτηρίζεται από σχετικά μεγάλη ταχύτητα εκτέλεσης σε σχέση με τους υπόλοιπους ταξινομητές της κατηγορίας του. Όσον αφορά το θέμα των ποσοστών είχαμε κακά αποτελέσματα όπως μπορούμε να παρατηρήσουμε στον παραπάνω πίνακα. Όσον αφορά τις ανοικτές παραμέτρους, δεν υπήρξε καμιά βελτίωση στην ακρίβεια του μοντέλου. Να σημειωθεί ότι αποτελεί μειονέκτημα η μη εξαγωγή από το πρόγραμμα σε αυτόν τον αλγόριθμο, ευκατανόητων κανόνων.

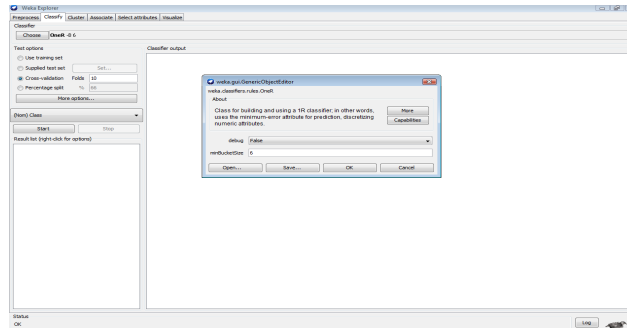
### 3.5.6 Αλγόριθμος κανόνων τύπου ‘OneR’

Ο αλγόριθμος OneR χρησιμοποιεί ταξινομητές τύπου OneR. Για την ακρίβεια χρησιμοποιεί το



χαρακτηριστικό ελάχιστου σφάλματος για πρόβλεψη διαχωρίζοντας τα αριθμητικά χαρακτηριστικά (Holte,1993).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα και γίνεται αναφορά αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



Εικόνα 3.5.6: Ανοικτές παράμετροι του αλγόριθμου ‘OneR’

- **MinBucketSize:** Καθορίζει το ελάχιστο Bucket Size για τον διαχωρισμό των αριθμητικών χαρακτηριστικών.

OneR	CORRECTLY CLASSIFIED	
<b>Parameter: minBucketSize</b>		
	3	29.6651
	6	29.6651
	15	29.6651
	25	29.6651

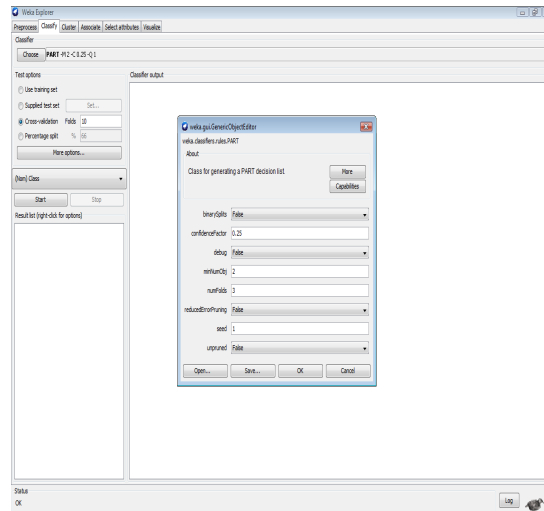
Πίνακας 3.5.6: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘OneR’

Ο ταξινομητής OneR αποτελεί έναν αρκετά γρήγορο ταξινομητή ο οποίος όμως έχει δυο σοβαρά μειονεκτήματα. Πρώτον έχει πολύ φτωχή απόδοση και δεύτερον δεν εξάγει σαφής και ευνόητους κανόνες. Τέλος παρατηρούμε ότι δεν έχουμε κάποια μεταβολή των ποσοστών στις περιπτώσεις που αλλάζουμε και δώσουμε τιμές 3,15,25 στην παράμετρο minBucketSize.

### 3.5.7 Αλγόριθμος κανόνων τύπου ‘PART’

Ο αλγόριθμος αυτός, κατασκευάζει μια PART λίστα απόφασης. Βασίζεται στην αρχή «διαίρει και βασίλευε» δηλαδή στην πραγματικότητα φτιάχνει ένα δένδρο C4.5 αλλά σε μορφή κανόνων (Eibe & Witten,1998). (Εξάγει σε κανόνες τα καλύτερα φύλλα).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.5.7: Ανοικτές παράμετροι του αλγόριθμου ‘PART’

- **Confidence Factor:** Η παράμετρος αυτή δείχνει τον ‘παράγοντα εμπιστοσύνης’ που χρησιμοποιείται κατά την διαδικασία του “pruning” (κλαδέματος). Μελέτες δείχνουν ότι η προεπιλεγμένη τιμή 0.25, δουλεύει καλά στις περισσότερες των περιπτώσεων. Εάν μικρύνουμε την τιμή αυτή, θα προκαλέσουμε ένα πιο δραστικό “pruning”.

PART	CORRECTLY CLASSIFIED
<b>Parameter: confidenceFactor</b>	
0.25	39.2344
0.5	41.1483
0.9	32.0574
<b>Parameter: numFolds</b>	
3	39.2344
10	39.2344
20	39.2344

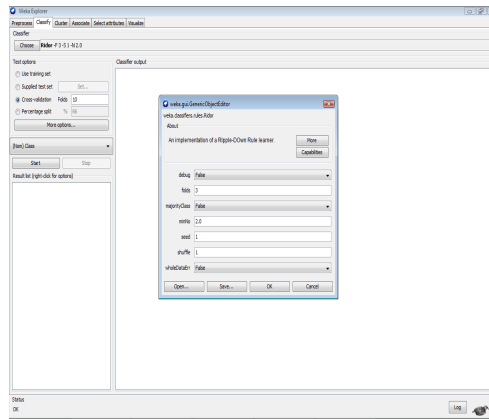
Πίνακας 3.5.7: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘PART’

Ο ταξινομητής PART αποτελεί έναν ταξινομητή ο οποίος είναι αρκετά γρήγορος στην εκτέλεση. Επίσης το πρόγραμμα μέσω αυτού του ταξινομητή μπορεί να εξάγει σαφής και απλούς κανόνες. Όσον αφορά τις ανοικτές παραμέτρους, όπως βλέπουμε και στον πίνακα, ο confidenceFactor επηρεάζει το αποτέλεσμα και θετικά και αρνητικά σε σχέση με το αποτέλεσμα που δίνει το πρόγραμμα στην αρχική τιμή παραμέτρου. Το μεγαλύτερο ποσοστό ήταν 41,15%.

### 3.5.8 Αλγόριθμος κανόνων τύπου ‘Ridor’

Ο αλγόριθμος Ridor, παράγει πρώτα έναν κανόνα προεπιλογής και στη συνέχεια τις εξαιρέσεις για τον κανόνα αυτόν, με το λιγότερο ποσοστό σφάλματος. Στο τέλος παράγει την καλύτερη εξαίρεση και από κάθε εξαίρεση επαναλαμβάνει αυτή τη διαδικασία έως ότου βρει την πιο καλή και απλή εξαίρεση. Εκτελεί, δηλαδή, μια επέκταση δένδρου με εξαιρέσεις. Οι εξαιρέσεις είναι ένα σύνολο κανόνων που προβλέπουν κλάσεις εκτός της προεπιλεγμένης (Gaines & Compton, 1995). Εδώ πρέπει να σημειωθεί ότι για να παραγάγει τις εξαιρέσεις, χρησιμοποιείτε η μέθοδος IREP.

Στον παρακάτω πίνακα φαίνονται οι παράμετροι του αλγόριθμου αυτού καθώς και η ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική:



Εικόνα 3.5.8: Ανοικτές παράμετροι του αλγόριθμου ‘Ridor’

- **Shuffle:** Καθορίζει πόσο συχνά τα δεδομένα ανακατεύονται πριν από έναν κανόνα που θα επιλεγεί. Εάν είναι μεγαλύτερο του 1 τότε ένας κανόνας χρησιμοποιείται πολλές φορές και ο πιο ακριβής κανόνας επιλέγεται.
- **Folds:** Καθορίζει την ποσότητα των δεδομένων που χρησιμοποιούνται για ‘κλάδεμα’. Το ένα fold χρησιμοποιείται για κλάδεμα και τα υπόλοιπα για ανάπτυξη των κανόνων.

Ridor	CORRECTLY CLASSIFIED
<b>Parameter: shuffle</b>	
1	41.6268
5	38.756
<b>Parameter: folds</b>	
3	41.6268
10	43.0622
20	40.6699

Πίνακας 3.5.8: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘Ridor’

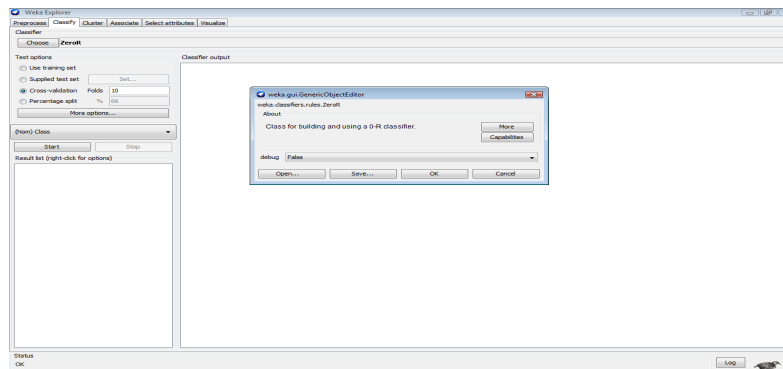
Ο αλγόριθμος Ridor αποτελεί έναν μέτριο ταξινομητή για το πρόβλημα μας. Είναι αρκετά γρήγορος και πετυχαίνει αρκετά υψηλές ακρίβειες αλλά ο τύπος κανόνων (κανόνες με εξαίρεση) που εξάγει δεν είναι εύκολα κατανοήσιμος. Για την παράμετρο ‘Shuffle’, ο αλγόριθμος έτρεξε για άλλη τιμή πλην της προεπιλεγμένης δηλαδή για 5, όπου μείωσε τα αποτελέσματα. Τα ποσοστά κυμαίνονται χαμηλά με μέγιστο το 43,06%.

### 3.5.9 Αλγόριθμος μηδενικού κανόνα ‘ZeroR’

Η μέθοδος του μηδενικού κανόνα, ZeroR, θεωρείται μια αρκετά απλή μέθοδος. Προτείνει μια τιμή για την εξαρτημένη μεταβλητή  $y$  η οποία είναι ίση με τη μέση τιμή της μεταβλητής στο σύνολο των δεδομένων, όταν πρόκειται για συνεχή αριθμητική τιμή. Όταν πρόκειται για κατηγορική μεταβλητή, είναι ίση με την τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα.

Ο αλγόριθμος αυτός έχει σαν ανοικτή παράμετρο μόνο την παράμετρο debug:

- Εάν έχει οριστεί σε true, ο ταξινομητής μπορεί να παράγει επιπλέον πληροφορίες για το πρόβλημα.



Εικόνα 3.5.9: Ανοικτές παράμετροι του αλγόριθμου 'ZeroR'

ZeroR	CORRECTLY CLASSIFIED
	19.6172

Πίνακας 3.5.9: Πίνακας Αποτελεσμάτων του αλγόριθμου 'ZeroR'

όπως βλέπουμε στον πίνακα 3.5.9 τα αποτελέσματα του αλγόριθμου 'ZeroR' είναι πολύ χαμηλά όμοια με αυτά που έδωσαν οι αλγόριθμοι Grading, MultiScheme, RacedIncrementalLogitBoost, Stacking, StackingC και Vote, 19,62%.

### 3.6 ΔΕΝΔΡΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ (trees)

#### 3.6.1 Δενδρικός αλγόριθμος Δυαδικού Καταμερισμού BF trees

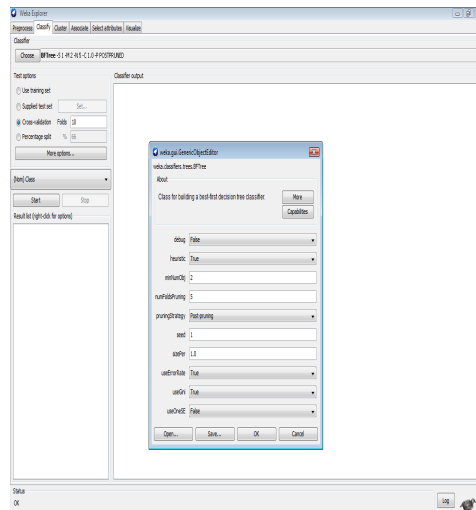
Ο αλγόριθμος "BF Trees" (Haijan Shi,2007), χρησιμοποιεί δυαδικό καταμερισμό για τα αριθμητικά και ονομαστικά χαρακτηριστικά (numeric & nominal attributes).

Όταν μιλάμε για δυαδικό καταμερισμό εννοούμε ότι, ο συγκεκριμένος αλγόριθμος χρησιμοποιεί δυαδικά δένδρα, τα οποία εύκολα δημιουργούνται, άλλα τείνουν να είναι πιο βαθιά. Η τελική απόδοση μπορεί να είναι συγκρίσιμη, επειδή ακριβώς οι συγκρίσεις περιορίζονται και είναι πιο απλές από εκείνες που απαιτούν πολλαπλές διακλαδώσεις. (Margaret H.Dunham, 2004)

Ο αλγόριθμος λειτουργεί ως εξής. Σε πρώτη φάση επιλέγεται ένα χαρακτηριστικό, το οποίο τοποθετείται στον αρχικό κεντρικό κόμβο, και επίσης δημιουργούνται κάποια κλαδιά βάσει κάποιων κριτηρίων.

Σε δεύτερη φάση, διαχωρίζεται το σετ δεδομένων εκπαίδευσης σε μικρότερα σετ, καθένα από τα οποία αντιστοιχεί σε κάθε παραγόμενο κλαδί από τον αρχικό κόμβο. Έπειτα, η συγκεκριμένη διαδικασία επαναλαμβάνεται για ένα επιλεγμένο κλαδί, χρησιμοποιώντας μόνο εκείνες τις περιπτώσεις όπου αντιστοιχούν σε αυτό. Σε κάθε βήμα επιλέγεται το καλύτερο υπό-σετ δεδομένων, ανάμεσα σε όλα τα διαθέσιμα υπό-σετ δεδομένων που χρησιμοποιούνται για την κατασκευή του δέντρου, στον επιλεγμένο κλάδο πάντα (Haijan Shi,2007).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.6.1: Ανοικτές παράμετροι του αλγόριθμου ‘BFTree’

- Num Folds Pruning: Στην παράμετρο αυτή, επιλέγουμε τον αριθμό των επαναλήψεων εσωτερικά της μεθόδου αξιολόγησης, μέσω επαναλαμβανόμενων δοκιμών σε άγνωστα δεδομένα.
- Pruning Strategy: Εδώ, επιλέγουμε την στρατηγική κλαδέματος. Η προεπιλεγμένη είναι η εκ των υστέρων κλάδεμα (post pruning). Εκτός από αυτή, υπάρχουν άλλες δυο επιλογές, οι: όχι κλάδεμα (no pruning) και η εκ των πρότερων κλάδεμα (pre pruning)

BFTree	CORRECTLY CLASSIFIED
<b>Parameter: numFoldsPruning</b>	
5	36.8421
10	35.4067
15	36.8421
<b>Parameter: pruningStrategy</b>	
Post-pruning	36.8421
Pre-pruning	31.1005

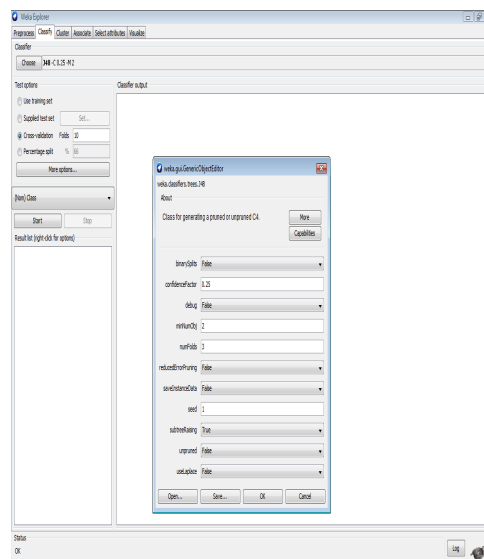
Πίνακας 3.6.1: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘BFTree’

Στον παραπάνω πίνακα παρατηρούμε τα διάφορα αποτελέσματα που προέκυψαν από τις μεταβολές στις ανοικτές παραμέτρους του συγκεκριμένου αλγορίθμου. Το μεγαλύτερο ποσοστό επιτυχίας που παρατηρήθηκε ήταν 36,84% και σημειώθηκε στις παραμέτρους numFoldsPruning: 5 και pruningStrateg: Post-pruning. Επίσης τα ποσοστά με τις αλλαγές στις τιμές των παραμέτρων πέφτουν.

### 3.6.2 Αλγόριθμος ‘J48’

Ο αλγόριθμος ταξινόμησης “J48” (Ross Quinlan,1993) δημιουργεί ένα C4.5 δένδρο το οποίο μπορεί να είναι είτε κλαδεμένο είτε ολόκληρο. Η μέθοδος κατηγοριοποίησης C4.5 εφαρμόζει μια απλή κατά βάθος διαδικασία για την κατασκευή του δένδρου. Τα γνωρίσματα κάθε κόμβου του δένδρου μπορούν να έχουν συνεχείς τιμές. Ωστόσο για να λειτουργήσει σωστά χρειάζεται ολοκληρωμένα δεδομένα. Ο C4.5 κρίνεται ακατάλληλος για μεγάλα σύνολα δεδομένων, αφού η ακρίβεια που παρουσιάζει είναι πολύ μικρή (Μ. Χαλκίδη-Μ. Βαζιργιάννης, 2005).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.6.2: Ανοικτές παράμετροι του αλγόριθμου ‘J48’

- **Confidence Factor:** Η παράμετρος αυτή δείχνει τον παράγοντα εμπιστοσύνης που χρησιμοποιείται κατά την διαδικασία του “pruning” (κλαδέματος). Μελέτες δείχνουν ότι η προεπιλεγμένη τιμή 0.25, δουλεύει καλά στις περισσότερες των περιπτώσεων. Εάν μικρύνουμε την τιμή αυτή, θα προκαλέσουμε ένα πιο δραστικό “pruning”.

J48	CORRECTLY CLASSIFIED
<b>Parameter: confidenceFactor</b>	
0.25	48.8038
0.5	44.0191
0.9	42.5837
<b>Parameter: numFolds</b>	
3	48.8038
10	48.8038
20	-

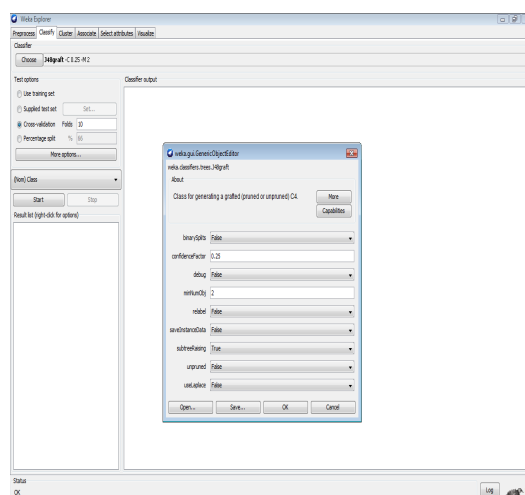
Πίνακας 3.6.2: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘J48’

Ο ταξινομητής J48 που βασίζεται στον γνωστό αλγόριθμο C4.5 είχε μεγάλη ταχύτητα εκτέλεσης. Στον παραπάνω πίνακα παρατηρούνται οι αλλαγές που έγιναν μεταβάλλοντας τις ανοικτές παραμέτρους του ταξινομητή. Όσον αφορά τις αλλαγές της παραμέτρου Confidence Factor δοκιμάσαμε δύο τιμές μεγαλύτερες (0.5) και (0.9) από αυτήν που δίνει το πρόγραμμα WEKA με τα αποτελέσματα να μειώνονται. Ενώ η μεταβλητή NumFolds δεν άλλαξε τα ποσοστά που έδιναν οι αρχικές παράμετροι.

### 3.6.3 Αλγόριθμος ‘J48 graft’

Ο συγκεκριμένος αλγόριθμος ανήκει στην κατηγορία αλγορίθμων που χρησιμοποιούνται για τη δημιουργία ενός C4.5 δέντρου απόφασης.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.6.3: Ανοιχτές Παράμετροι του αλγόριθμου ‘J48 graft’**

- **Confidence Factor:** Η παράμετρος αυτή δείχνει τον παράγοντα εμπιστοσύνης που χρησιμοποιείται κατά την διαδικασία του “pruning” (κλαδέματος). Μελέτες δείχνουν ότι η προεπιλεγμένη τιμή 0.25, δουλεύει καλά στις περισσότερες των περιπτώσεων. Εάν μικρύνουμε την τιμή αυτή, θα προκαλέσουμε ένα πιο δραστικό “pruning” (ακριβώς όπως και στον αλγόριθμο J48).

J48graft	CORRECTLY CLASSIFIED
<b>Parameter: confidenceFactor</b>	
0.25	47.8469
0.5	43.5407
0.9	41.6268
<b>Parameter: minNumObj</b>	
2	47.8469
10	47.8469
20	48.3254

**Πίνακας 3.6.3: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘J48 graft’**

Στον παραπάνω πίνακα παρατηρούνται οι αλλαγές που έγιναν μεταβάλλοντας τις ανοικτές παραμέτρους του ταξινομητή. Όσον αφορά τις αλλαγές της παραμέτρου Confidence Factor δοκιμάσαμε δύο τιμές μεγαλύτερες (0.5) και (0.9) από αυτήν που δίνει το πρόγραμμα WEKA με τα αποτελέσματα όπως και στον αλγόριθμο J48 να μειώνονται. Ενώ η μεταβλητή minNumFolds αύξησε την ακρίβεια του αλγόριθμου σε σχέση με τις αρχικές παραμέτρους.

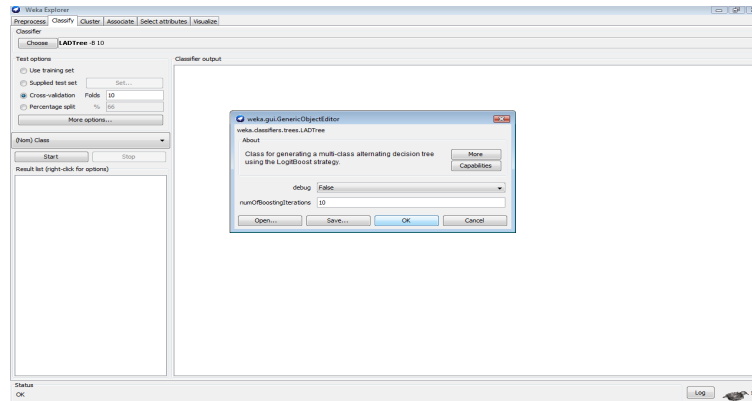
### 3.6.4 Αλγόριθμος ‘LADTree’

Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται για τη δημιουργία μιας πολυ-κατηγορίας εναλλασσόμενου δέντρου απόφασης σύμφωνα με τη στρατηγική LogitBoost.

Ο αλγόριθμος αυτός έχει σαν ανοικτή παράμετρο μόνο την παράμετρο debug, παρακάτω



δείχνουμε έναν πίνακα σαν παράδειγμα:



Εικόνα 3.6.4: Ανοικτές παράμετροι του αλγόριθμου ‘LADTree’

LADTree	CORRECTLY CLASSIFIED
<b>Parameter: numOfBoostingIterations</b>	
10	42.1053
20	43.0622
30	40.1914

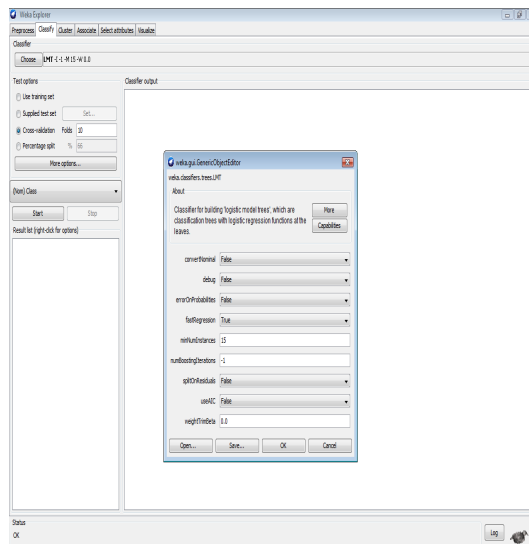
Πίνακας 3.6.4: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘LADTree’

### 3.6.5 Αλγόριθμος ‘LMT’

Ο αλγόριθμος ‘LMT’ (Landwehr, 2005) χρησιμοποιείται για προβλήματα ταξινόμησης και χρησιμοποιεί λογιστική παλινδρόμηση αντί για γραμμική παλινδρόμηση.

Η λογιστική παλινδρόμηση χρησιμοποιεί μια λογιστική καμπύλη, αντί να ταιριάζει τα δεδομένα σε ευθεία γραμμή. Η λογιστική καμπύλη χρησιμοποιεί τιμές από 0 έως 1 έτσι ώστε να μπορεί να θεωρηθεί σαν την πιθανότητα συμμετοχής σε κάποια κατηγορία. Μπορεί επίσης όπως και η γραμμική παλινδρόμηση να χρησιμοποιηθεί και όταν απαιτείται κατηγοριοποίηση σε δύο κατηγορίες. (Margaret H. Dunham, 2004)

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.6.5: Ανοικτές παράμετροι του αλγόριθμου ‘LMT’

- **MinNumInstances:** Θέτει τον ελάχιστο αριθμό παραδειγμάτων που χρησιμοποιούνται σε κάθε κόμβο για να γίνει διαμερισμός.
- **NumBoostinIterations:** θέτει έναν συγκεκριμένο αριθμό επαναλήψεων για τον LogitBoost. Εάν είναι θετική τότε ορίζεται ένας συγκεκριμένος αριθμός επαναλήψεων, που ακολουθείται σε όλο το δένδρο. Εάν είναι αρνητικός τότε ο αριθμός αυτός ακολουθεί τη μέθοδο cross-validation.

LMT	CORRECTLY CLASSIFIED
<b>Parameter: minNumInstances</b>	
15	43.0622
30	44.0191
50	44.0191
<b>Parameter: numBoostingIterations</b>	
-1	43.0622
5	43.0622
10	46.4115

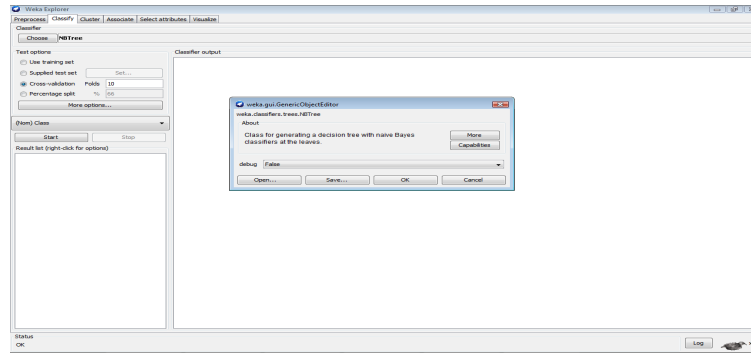
Πίνακας 3.6.5: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘LMT’

Ο αλγόριθμος LMT (Logistic Model Trees) έχει σαν μειονέκτημα τον μεγάλο χρόνο εκτέλεσης (πάνω από μια ώρα) και την μη δυνατότητα κάποιας αναπαράστασης των αποτελεσμάτων (με την μορφή κάποιου δένδρου). Στην τελική διάγνωση οι αλλαγές στις ανοικτές παραμέτρους δεν είχαν κάποια μεγάλη επίδραση, και παρατηρήθηκε μικρή αύξηση. Το καλύτερο ποσοστό ήταν 46,41%.

### 3.6.6 Αλγόριθμος NBTree

Ο αλγόριθμος NB tree (Ron Kohavi, 1996) αποτελεί ένα υβριδικό μοντέλο μεταξύ δένδρων απόφασης και του αλγόριθμου Naïve Bayes. Δημιουργεί δένδρα που τα φύλλα τους βασίζονται σε Naïve Bayes ταξινομητές

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.6.7: Ανοικτές παράμετροι του αλγόριθμου ‘NMTree’

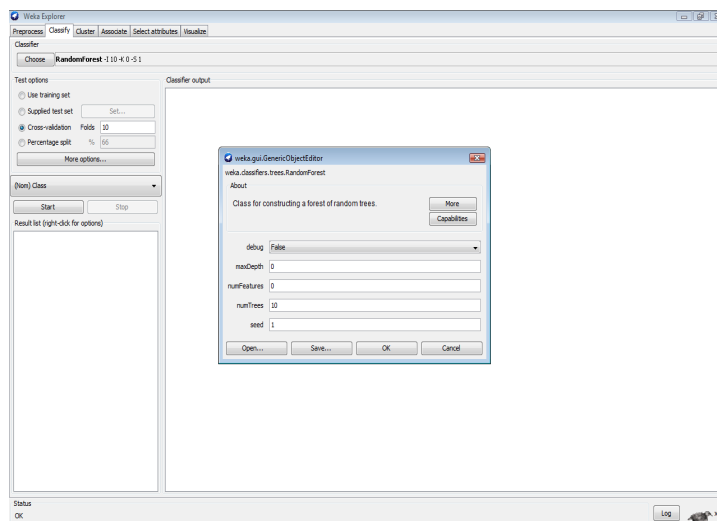
NBTrees	CORRECTLY CLASSIFIED
	48.8038

Πίνακας 3.6.7: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘NMTree’

### 3.6.7 Αλγόριθμος ‘Random Forest’

Ο αλγόριθμος αυτός δημιουργεί “random forests” (Leo Breiman, 2001) με το να κάνει Bagging σε ένα σύνολο από “random trees.”

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



Εικόνα 3.6.8: Ανοικτές παράμετροι του αλγόριθμου ‘RandomForest’

- MaxDepth: καθορίζει το μέγιστο βάθος του δένδρου
- NumTrees: Καθορίζει τον αριθμό των δένδρων που θα δημιουργηθούν.

RandomForest	CORRECTLY CLASSIFIED	
<b>Parameter: maxDepth</b>		
	0	29.6651
	2	30.622
	5	29.6651
<b>Parameter: numTrees</b>		
	5	26.3158
	10	29.6651
	20	30.1435

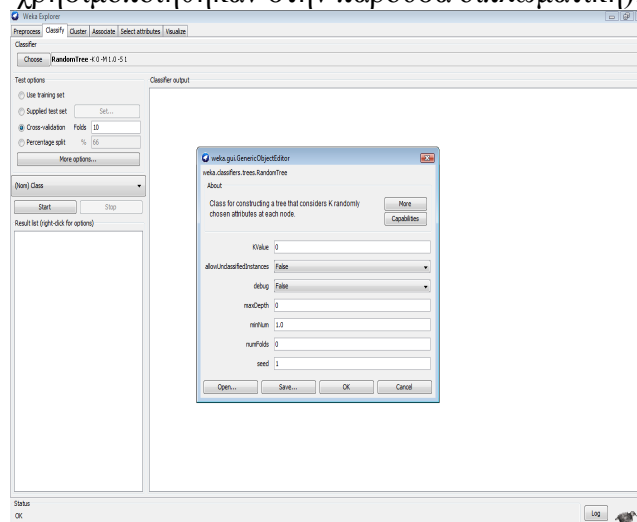
**Πίνακας 3.6.8: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘RandomForest’**

Ο ταξινομητής Random Forest είχε μέτρια απόδοση όσον αφορά τον χρόνο εκτέλεσης και κακή απόδοση όσον αφορά τα αποτελέσματα. Ορίζοντας διαφορετικές τιμές μετά από δοκιμαστικά τρεξίματα, παρατηρούμε και μειώσεις και αυξήσεις στα αποτελέσματα.

### 3.6.8 Random Tree

Ο ταξινομητής “Random Tree” δημιουργεί ένα δένδρο που επιλέγει k τυχαία χαρακτηριστικά σε κάθε κόμβο. Να σημειωθεί ότι ο αλγόριθμος δεν πραγματοποιεί κλάδεμα (pruning).

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.6.9: Ανοιχτές Παράμετροι του αλγόριθμου ‘RandomTree’**

- MaxDepth: καθορίζει το μέγιστο βάθος του δένδρου
- Num Folds: Στην παράμετρο αυτή, επιλέγουμε τον αριθμό των επαναλήψεων εσωτερικά της μεθόδου αξιολόγησης, μέσω επαναλαμβανόμενων δοκιμών σε άγνωστα δεδομένα.

RandomTree	CORRECTLY CLASSIFIED	
<b>Parameter: maxDepth</b>		
	0	22.488
	2	27.2727
	5	22.488
<b>Parameter: numFolds</b>		
	0	22.488
	5	19.1388
	10	19.1388

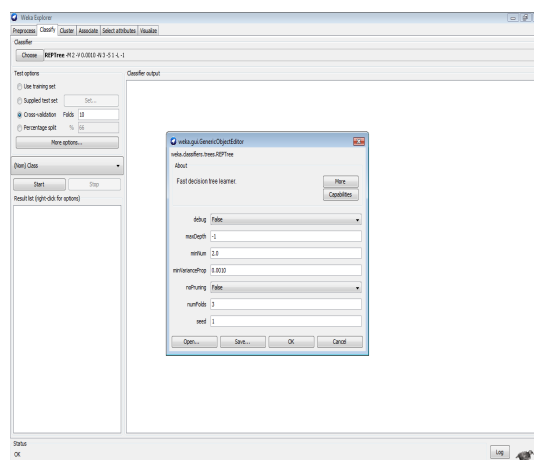
**Πίνακας 3.6.9: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘RandomTree’**

Παρατηρώντας τον παραπάνω πίνακα 3.6.9 γίνεται ασαφής η κακή απόδοση του αλγόριθμου στο πρόβλημα μας.

### 3.6.9 Αλγόριθμος ‘REPTree’

Η μέθοδος κατασκευής δενδρικών αλγορίθμων REP Trees, κατασκευάζει ένα δένδρο απόφασης/παλινδρόμησης, χρησιμοποιώντας πληροφορίες της μορφής gain/variance και κάνει κλάδεμα χρησιμοποιώντας τη μέθοδο reduced error pruning. Αποτελεί μια γρήγορη μέθοδο κατασκευής δένδρων.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.6.10: Ανοικτές παράμετροι του αλγόριθμου ‘RepTree’**

Ο αλγόριθμος ‘RepTree’ χρησιμοποιεί τις ίδιες παραμέτρους με αυτές που αναλύθηκαν παραπάνω για τον αλγόριθμο ‘RandomTree’

REPTree	CORRECTLY CLASSIFIED
<b>Parameter: maxDepth</b>	
-1	27.2727
2	27.2727
5	27.2727
<b>Parameter: numFolds</b>	
3	27.2727
10	22.488
20	21.5311

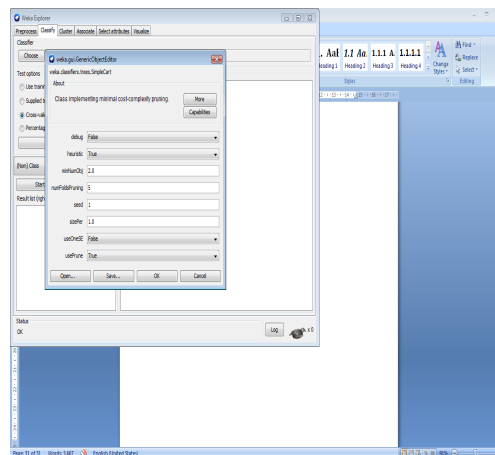
**Πίνακας 3.6.10: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘RepTree’**

Παρατηρώντας τα αποτελέσματα του παραπάνω πίνακα, βλέπουμε την απόδοση του αλγόριθμου να μειώνεται όταν μεταβληθούν οι τιμές των καθορισμένων παραμέτρων. Το καλύτερο ποσοστό ορθούς ταξινόμησης επιτεύχθηκε, ήταν (27,27%).

### 3.6.10 Αλγόριθμος ‘SimpleCart’

Ο αλγόριθμος αυτός εφαρμόζεται για την ελάχιστη τιμή σε ένα πρόβλημα πολυπλοκότητας. Εδώ επίσης θα πρέπει να τονίσουμε ότι όταν έχουμε τιμές που λείπουν, η χρήση "κλασματικού περίπτωσης" προτιμάται αντί της παρένθετης μέθοδο διαχωρισμού.

Οι παράμετροι του αλγόριθμου αυτού φαίνονται στον παρακάτω πίνακα (ανάλυση αυτών που χρησιμοποιήθηκαν στην παρούσα διπλωματική):



**Εικόνα 3.6.11: Ανοικτές παράμετροι του αλγόριθμου ‘SimpleCart’**

- Num Folds Pruning: Στην παράμετρο αυτή, επιλέγουμε τον αριθμό των επαναλήψεων εσωτερικά της μεθόδου αξιολόγησης, μέσω επαναλαμβανόμενων δοκιμών σε άγνωστα δεδομένα.

SimpleCart		CORRECTLY CLASSIFIED
<b>Parameter: numFoldsPruning</b>		
	5	36.8421
	10	38.756
	20	38.756
<b>Parameter: minNumObj</b>		
	2	36.8421
	10	38.756
	20	33.9713

**Πίνακας 3.6.11: Πίνακας Αποτελεσμάτων του αλγόριθμου ‘SimpleCart’**

Παρατηρώντας τον παραπάνω πίνακα 3.6.11 βλέπουμε πως οι μεταβολές των αρχικών ρυθμίσεων έχει και αρνητική και θετική επίδραση στα αποτελέσματα.

## ΚΕΦΑΛΑΙΟ 4 - ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εργασία, μελετήθηκε το πρόβλημα ταξινόμησης ιατρικών περιπτώσεων, οι οποίες αφορούσαν τον τύπο αφασίας ασθενών. Το συγκεκριμένο ιατρικό πρόβλημα είναι αρκετά σημαντικό αν αναλογιστεί κανείς πως κάθε χρόνο στην Αμερική 100.000 νέοι ασθενείς παθαίνουν αφασία μετά από εγκεφαλικό επεισόδιο. Για την διάγνωση της αφασίας ο γιατρός χρησιμοποιεί μια σειρά νευρολογικών και άλλων εξετάσεων για να προσδιορίσει τη ύπαρξη ή όχι αφασίας αλλά και τη σοβαρότητά της. Επίσης έχουν αναπτυχθεί διάφορα τεστ, αυτό που χρησιμοποιήθηκε σε αυτή την εργασία είναι το Aachener Aphasia Test. Ως αποτέλεσμα της παραπάνω εξέτασης, προκύπτει ο τύπος της αφασίας του ασθενούς. Η συνεισφορά της συγκεκριμένης μελέτης αφορά το στάδιο της διάγνωσης. Πιο συγκεκριμένα, χρησιμοποιήθηκαν μια βάση δεδομένων, η οποία αποτελείται τα σκορ που πέτυχαν αφασικοί ασθενείς στο AAT. Στην επιστήμη λήψης απόφασης, τέτοια προβλήματα διάγνωσης μπορούν να αντιμετωπιστούν ως προβλήματα ταξινόμησης. Με την εφαρμογή κατάλληλων μεθοδολογιών, μπορεί να προκύψει ένα σύνολο κανόνων ταξινόμησης περιπτώσεων αφασικών ασθενών, βάσει των οποίων μπορεί να ταξινομηθεί μια νέα περίπτωση. Αυτό μπορεί να αποτελέσει ένα υποστηρικτικό σύστημα λήψης απόφασης (Decision Support System-DSS) στη βασική διάγνωση του γιατρού. Όσον αφορά το μεθοδολογικό πλαίσιο χρησιμοποιήθηκε το λογισμικό WEKA, το οποίο περιλαμβάνει μεθοδολογίες ταξινόμησης από το χώρο της στατιστικής, της τεχνητής νοημοσύνης και της ευρύτερης περιοχής λήψης απόφασης.

Από την εφαρμογή των μεθόδων λήψης απόφασης, προέκυψαν κάποια σημαντικά ευρήματα. Όσον αφορά την αλγοριθμική ανάλυση, παρακάτω παρατίθενται οι καλύτερες και χειρότερες ακρίβειες ταξινόμησης ανά κατηγορία τεχνικών:

Ταξινομητές τύπου Bayes: καλύτερο ποσοστό 47,85% (BayesNet) και χειρότερο 44,02% (NaiveBayesUpdateable, NaiveBayesSimple, NaiveBayes).

Αλγόριθμοι βασισμένοι στην απόσταση (Lazy): καλύτερο ποσοστό 47,85% (IwI) και χειρότερο 39,23% (Kstar).

Αλγόριθμοι μετά-μάθησης (Meta-Learning): καλύτερο ποσοστό 61,90% (END) και χειρότερο 19,62% (MultiScheme, RacedIncrementalLogitBoost, Stacking, StakingC, Vote, grading, CVParameterSelection).

Συναρτησιακοί Αλγόριθμοι (Functions): καλύτερο ποσοστό 48,33% (MultilayerPerceptron) και χειρότερο 39,71% (RBFnetwork).

Αλγόριθμοι εξαγωγής κανόνων (Rules): καλύτερο ποσοστό 48,80% (DecisionTable) και χειρότερο 27,27% (ConjunctiveRule).

Δενδρικοί αλγόριθμοι (Trees): καλύτερο ποσοστό 48,80% (J48,NbTree) και χειρότερο 22,48% (random tree).

Από τα παραπάνω, μπορεί να παρατηρηθεί ότι όλοι οι αλγόριθμοι που χρησιμοποιήθηκαν, έδωσαν ακρίβειες ταξινόμησης πολύ κάτω από τα επιθυμητά επίπεδα. Για να γίνουμε πιο συγκεκριμένοι, ακόμα και το μεγαλύτερο ποσοστό που επιτεύχθηκε από τις τεχνικές μετά-μάθησης (61,90%) κρίνεται ανεπαρκές. Η υπάρχουσα βιβλιογραφία που αναλύσαμε παραπάνω έχει πετύχει ποσοστά πολύ ανώτερα από την παρούσα έρευνα.

## Αναφορές

[1] [<http://www.afasie.nl/aphasia/pdf/3/brochure1.pdf> A.I.A, Association, Internationale, Aphasie]

[2] An introduction to aphasia, Robert H. Brookshire BRK Publishers, 1973, MIT.

[3] Kandel E, Schwartz J, Jessel Th: Μέρος I, Γενική άποψη, κεφ.1: εγκέφαλος και συμπεριφορά, σελ 5-21, Μέρος IX, Γλώσσα, Μάθηση και Μνήμη, κεφ.34: Γλώσσα, σελ. 663- 680, στο: Νευροεπιστήμη και συμπεριφορά, Πανεπιστημιακές εκδόσεις Κρήτης, 2η έκδοση, Ηράκλειο 2000.

[4] A R Damasio, and N Geschwind, The Neural Basis of Language Annual Review of Neuroscience Vol. 7: 127-147 (Volume publication date March 1984) DOI: 10.1146/annurev.ne.07.030184.001015.

[5] Hedge, MN. 2001. Pocket guide to assessment in speech-language pathology. San Diego: Singular and Thompson Learning.

[6] Ευρογλωσσολογική λογοθεραπεία (Δεύτερος Τομος) Θέραπεια καρπαθίου Εμμ. Χρυσόστομος , 1999.

[7][Adams RA, Victor M: Derangements of Intellect, Behaviour and Language due to diffuse and focal cerebral disease: Disorders of Speech and Language, Section 5:413-429, In: Principles of



Neurology, 9th edition, McGraw-Hill Medical Publishing Division, 2009.

[8] Neuroanatomy Through Clinical Cases, Second Edition, Blumenfeld, Neuroanatomy Through Clinical Cases, Sinauer Associates, Inc.; Second edition (May 1, 2011).

[9] Grodzinsky, Yosef. Theoretical perspectives on language deficits. MIT press, 1990.

[10] Μπουουζιάνη Χρυσάνθη, Δερμιτζάκης Εμμανουήλ, Μαλεγιαννάκη Αμαρυλλίς, Τσίπτσιος Ιάκωβος. Διαταραχές λόγου: τα είδη των αφασιών. Μια σύνθετη νευρολογική και νευροψυχολογική προσέγγιση. Περιοδικό Νευρολογία, τόμος 17, τεύχος 5, Σεπτέμβριος-Οκτώβριος 2008:257-267.

[11] <http://www.studyblue.com/notes/note/n/language--aphasia/deck/5382278>.

[12] Goodglass, H., & Caplan, E. (2001). The Assessment of Aphasia and Related disorders.

Lippincott Williams & Wilkins, USA.

[13] NIDCD Fact Sheet Aphasia The NIDCD Information Clearinghouse is a service of the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, U.S. Department of Health and Human Services.

[14] Goodglass, Harold, and Edith Kaplan. Boston diagnostic aphasia examination booklet. Lea & Febiger, 1983.

[15] O. Spreen, A.L. Benton Neurosensory Center Comprehensive Examination for Aphasia, Edition A Manual of Instructions Neuropsychology Laboratory, University of Victoria, Victoria (1969).

[16] Kertesz, Andrew. Western aphasia battery test manual. Psychological Corp, 1982.

[17] Huber, Walter, Klaus Poeck, and Klaus Willmes. "The Aachen Aphasia Test." Advances in neurology 42 (1983):291-303

[18] <http://www.iatronet.gr/ygeia/pathologia/article/22533/afasia-kai-symptwmata.html>, Αφασία και Συμπτώματα Ρουσοχατζάκα Μαρία, Λογοθεραπευτής, 2013.

[19] <http://www.aphasia.org/content/aphasia-therapy-guide>.

[20] Axer, Hubertus, et al. "The aphasia database on the web: description of a model for problems of classification in medicine." Proc. ESIT. 2000.

[21] [http://en.wikipedia.org/wiki/Soft\\_computing](http://en.wikipedia.org/wiki/Soft_computing)

[22] Marshall, J.C., 1986, "The description and interpretation of aphasic language disorder." Neuropsychologia 24, pp. 5-24.

[23] Axer, H., et al. "Diagnosis of aphasia using neural and fuzzy techniques." Symposium on Computational Intelligence and Learning. 2000.

[24] Axer, Hubertus, et al. "Aphasia classification using neural networks." European Symposium on Intelligent Techniques Aachen. Vol. 111 2000

[25] Axer, Hubertus, Jan Jantzen, and Diedrich Graf von Keyserlingk. "An aphasia database on the internet: a model for computer-assisted analysis in aphasiology." Brain and language 75.3 (2000): 390-398.

[26] Axer, H., et al. "Diagnosis of aphasia using neural and fuzzy techniques." Symposium on Computational Intelligence and Learning. 2000.

[27] Koza JR, Bennett III FH, Andre D, Keane MA. Genetic programming III: Darwinian invention and problem solving. San Francisco: Morgan Kaufmann Publishers; 1999.

[28] Tsakonas, Athanasios, et al. "Evolving rule-based systems in two medical domains using genetic programming." Artificial Intelligence in Medicine 32.3 (2004): 195-216.

[29]Tsakonas, Athanasios, et al. "Hybrid computational intelligence for handling diagnosis of aphasia."European Symposium on Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems. Eunate, December. 2001.