Ατομική Διπλωματική Εργασία

ΥΠΟΛΟΓΙΣΤΙΚΟ ΣΥΣΤΗΜΑ ΝΟΗΜΟΣΥΝΗΣ ΓΙΑ ΔΙΑΓΝΩΣΗ ΧΡΩΜΟΣΩΜΙΚΩΝ ΑΝΩΜΑΛΙΩΝ ΚΑΤΑ ΤΗΝ ΠΡΩΤΗ ΤΡΙΜΗΝΙΑ ΚΥΗΣΗΣ

Γιώργος Ξενοφώντος

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2015

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Υπολογιστικό Σύστημα Νοημοσύνης για διάγνωση χρωμοσωμικών ανωμαλιών κατά την πρώτη τριμηνία κύησης

Γιώργος Ξενοφώντος

Επιβλέπων Καθηγητής
Χρίστος Ν. Σχίζας

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου.

Μάιος 2015

Ευχαριστίες

Θα ήθελα να ευχαριστήσω όλους όσοι συνέβαλαν στην επιτυχή εκπόνηση της διπλωματικής μου εργασίας. Θέλω να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου, κ. Χρίστο Ν. Σχίζα για την καθοδήγηση και την εμπιστοσύνη που μου έδειξε.

Επίσης θα ήθελα να ευχαριστήσω τον ερευνητή Αντρέα Νεοκλέους που με βοήθησε σε μέγιστο βαθμό με τις ιδέες, την στήριξη και την καθοδήγηση του. Μέσα στον τελευταίο χρόνο ήταν πάντα διαθέσιμος να ασχοληθεί με κάθε απορία μου σχετική με το θέμα μου.

Τέλος θέλω να ευχαριστήσω την οικογένεια μου και τους φίλους μου για την στήριξη που μου παρείχαν κατά την διάρκεια των σπουδών μου.

Περίληψη

Στην διπλωματική μου εργασία δημιούργησα έξυπνα συστήματα υπολογιστικής νοημοσύνης με σκοπό τηn πρόβλεψη της τρισωμίας 21 (Σύνδρομο Down) στο έμβρυο την πρώτη τριμηνία της εγκυμοσύνης. Χρησιμοποίησα μια βάση δεδομένων με 29792 περιπτώσεις φυσιολογικών περιστατικών και 380 περιπτώσεις τρισωμίας 21.

Αρχικά ανάλυσα τα δεδομένα έτσι ώστε να παρατηρήσω ποιοι παράμετροι είναι ισχυροί όσον αφορά την ύπαρξη της τρισωμίας 21. Μέσω της επεξεργασίας αυτής κατάληξα ότι η αυχενική διαφάνεια (NT), το bhCG και το Ductus Venosus είναι ισχυρές παράμετροι.

Μετά, δημιούργησα ένα νευρωνικό δίκτυο στο οποίο το σύνολο εκπαίδευσης του αποτελείται από όλα τα δεδομένα των φυσιολογικών περιπτώσεων και των T21. Ο διαχωρισμός των περιπτώσεων φυσιολογικών εμβρύων μαζί με τα T21 ήταν αρκετά καλός αλλά όχι ικανοποιητικός. Σκοπός μου ήταν να δημιουργήσω ένα σύστημα που να διαχωρίζει όσο πιο καλά γίνεται τις δύο κατηγορίες T21 και φυσιολογικών εμβρύων. Έτσι, αποφάσισα να κάνω πρώτα ένα γραμμικό διαχωρισμό των περιπτώσεων χρησιμοποιώντας ένα κατώφλι που διαχωρίζει τις δυο κατηγορίες χρησιμοποιώντας μία παράμετρο κάθε φορά. Ο γραμμικός διαχωρισμός έγινε για τις τρεις παραμέτρους Αυχενική Διαφάνεια (NT), χοριακή γονιαδοτροπίνη (bhCG) και ροή φλεβώδους πόρου (DV). Βάση αυτού, δημιούργησα τρία συστήματα. Για το πρώτο σύστημα πήρα τιμή για κατώφλι του NT η οποία μετά το γραμμικό διαχωρισμό επιστρέφει ορθή κατηγοριοποίηση 100% T21 και 9% φυσιολογικά. Για το δεύτερο σύστημα η τιμή του NT για κατώφλι επιστρέφει 100% φυσιολογικά και 64% T21 και για το τρίτο σύστημα, η τιμή του NT για κατώφλι είναι στο ενδιάμεσο μεταξύ της τιμής του πρώτου συστήματος και του δευτέρου συστήματος, δηλαδή 34% φυσιολογικά και 96% T21. Αντίστοιχα έγινε και για τις παραμέτρους bhCG και DV. Μετέπειτα δημιούργησα τριάντα ένα νευρωνικά δίκτυα για όλα τα συστήματα με διαφορετικά σύνολα εκπαίδευσης, βασισμένα στον γραμμικό διαχωρισμό που έγινε σε προηγούμενο στάδιο.

Μετά την επαλήθευση παρατήρησα πως σε κάποιες περιπτώσεις, οι τιμές της εξόδου από τα νευρωνικά δίκτυα, ήταν κοντά στα όρια διαχωρισμού φυσιολογικών με T21. Κατά την επεξεργασία των αποτελεσμάτων, αποφάσισα να δημιουργήσω μια τρίτη κατηγορία την αμφίβολες περιπτώσεις.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Γενική Εισαγωγή

Στην εργασία αυτή, ασχολήθηκα με το πρόβλημα της έγκαιρης πρόβλεψης της χρωμοσωμικής ανωμαλίας τρισωμία 21 ή αλλιώς Σύνδρομο Down στο έμβρυο, κατά την διάρκεια της εγκυμοσύνης στην πρώτη τριμηνία. Για την πρόβλεψη χρησιμοποίησα έξυπνα συστήματα υπολογιστικής νοημοσύνης. Τέτοια συστήματα χρησιμοποιούνται στην καθημερινή ζωή για παραδείγματος χάριν σε ιατρικές και βιολογικές έρευνες [14-17] [19-21]. Στην ιατρική, τα τεχνητά νευρωνικά δίκτυα αποδείχθηκαν να είναι μια ισχυρή μέθοδος διάγνωσης. Ως παράδειγμα, η μελέτη του Al-Shayea [17] αναφέρει ένα ιατρικό σύστημα διάγνωσης για την ασθένεια της νεφρίτιδας και της καρδιακής ασθένειας με τη χρήση του αλγόριθμου τροφοδοσίας προς τα εμπρός. Έγινε σωστή ταξινόμηση με ποσοστό 99%. Ένα άλλο σύστημα [18] για την διάγνωση των παθήσεων πόνου στο στήθος με την χρήση των τεχνητών νευρωνικών δικτύων. Σε αυτή την μελέτη η βάση δεδομένων αποτελείται από τις μετρήσεις που έγιναν στο στήθος, 357 δείγματα και έξι κατηγορίες. Οι κατηγορίες ασθενειών ήταν η φυματίωση, η COPD, η πνευμονία, το άσθμα, ο καρκίνος του πνεύμονα και η φυσιολογική. Η αναφορά που γίνεται από τους συγγραφείς ήταν κατά μέσο όρο ταξινόμηση για τις έξι κατηγορίες με ποσοστό 90,2%.

Συγκεκριμένα δημιούργησα νευρωνικά δίκτυα που για είσοδο είχαν τις μετρήσεις από τον προγεννητικό έλεγχο κατά την πρώτη τριμηνία κύησης. Οι μετρήσεις αυτές εξάγονται από το υπερηχογράφημα, τις αναλύσεις αίματος και το ιατρικό προϊστορικό της. Όλα τα δεδομένα από τις μετρήσεις χρησιμοποιήθηκαν για είσοδο στο σύστημα για να εκπαιδευτεί.

Σκοπός της έγκαιρης πρόβλεψης είναι η ενημέρωση των γονέων όπου καλούνται να αποφασίσουν στο πως θα χειριστούν την εγκυμοσύνη. Υπάρχουν διάφορες τεχνικές για την πραγματοποίηση μιας τέτοιας έγκαιρης διάγνωσης, όπου η κάθε μια μπορεί να επιφέρει θετικές ή αρνητικές επιπτώσεις. Οι τεχνικές αυτές χωρίζονται σε δύο κατηγορίες. Τις επεμβατικές και τις μη-επεμβατικές.

Η μη-επεμβατική μέθοδος είναι ένας προγεννητικός έλεγχος. Σε πρόσφατη επιστημονική δημοσίευση από τους Tsaliki et. al [1] το 2013 αναφέρεται μια νέα μη-επεμβατική μέθοδος η οποία παρουσιάζει 100% ακριβές διάγνωση για θετικού ποσοστού επιτυχίας (true positive rate). Ο ρυθμός θετικού ποσοστού επιτυχίας, είναι οι περιπτώσεις που το σύστημα προέβλεψε πως το αποτέλεσμα θα είναι T21 και το πραγματικό αποτέλεσμα ήταν T21. Επίσης η μέθοδος αυτή παρουσιάζει 99,2% διάγνωση για αρνητικού ποσοστού επιτυχίας. Ο ρυθμός αρνητικού ποσοστού επιτυχίας, είναι οι περιπτώσεις που το σύστημα προέβλεψε πως το αποτέλεσμα θα είναι φυσιολογικό και το πραγματικό αποτέλεσμα ήταν φυσιολογικό. Μελέτησαν τις 12 διαφορετικές μεθυλιωμένες περιοχές (DMR) που χρησιμοποιούνται στο μεθυλιωμένο DNA. Σε μελέτη που έγινε από τους Papageorgiou et. al [2] το 2010 ένα μη επεμβατικό τεστ αναπτύχθηκε με επιτυχία για την διάγνωση της τρισωμίας 21 στην αρχή της 10^{ης} εβδομάδας κύησης. Στην μελέτη αυτή είχαν εστιάσει την ύπαρξη περιορισμένης ποσότητας (3-5%) ελεύθερου εμβρυικού DNA στο αίμα της μητρικής κυκλοφορίας. Έγινε εφαρμογή του τεστ σε 40 γνωστά δείγματα (20 φυσιολογικά και 20 T21) και σε 40 άγνωστα δείγματα (26 φυσιολογικά και 14 T21) από εγκυμονούσες κατά τις 11,1 έως 14,4 εβδομάδων κύησης. Είχε ακριβής διάγνωση όλων των περιπτώσεων 100% για θετικού ποσοστού επιτυχίας και αρνητικού ποσοστού επιτυχίας.

Στις επεμβατικές μεθόδους, η πιο γνωστή είναι η αμνιοκέντηση. Στην μέθοδο αυτή συλλέγονται σημαντικές γενετικές πληροφορίες για το έμβρυο μέσω της αφαίρεσης μικρού δείγματος αμνιακού υγρού. Συνήθως γίνεται μεταξύ 14^{ης} εβδομάδας και 26^{ης} εβδομάδας κύησης εκεί που το έμβρυο περιβάλλεται με αρκετή ποσότητα αμνιακού υγρού. Ο ιατρός μέσω των υπέρηχων εντοπίζει το σημείο στο αμνιακό υγρό το οποίο δεν θα προκαλέσει τραυματισμό στο έμβρυο. Εισάγει σύριγγα διαμέσου του κοιλιακού τοιχώματος και εξάγει 10 με 20 χιλιοστόλιτρα υγρού. Ακολούθως εύκολα μπορεί να διαγνωστούν οι χρωμοσωμικές ανωμαλίες με αξιοπιστία στέλνοντας το αμνιακό υγρό στα εργαστήρια για ανάλυση. Η διάγνωση λαμβάνεται εντός δυο εβδομάδων. Εντούτοις όμως υπάρχει ένα μικρό ποσοστό κινδύνου αποβολής και κίνδυνοι για την μητέρα κατά την διάρκεια ή και μετά την εξέταση με πιθανότητα 1 προς 200 [13].

Έτσι λοιπόν οι γονείς αφού έχουν τις επιλογές τους, καλούνται να αποφασίσουν εάν θα συνεχίσει η εγκυμοσύνη χωρίς περεταίρω εξετάσεις, ή αν θα προχωρήσουν σε έκτρωση του εμβρύου ή να προχωρήσουν σε επεμβατική μέθοδο για μία πιο ακριβής διάγνωση.

Ο κίνδυνος αποβολής με την εφαρμογή επεμβατικής μεθόδου για την ανάλυση αίματος του εμβρύου, προβάλλει την ανάγκη για αναζήτηση μεθόδων μη επεμβατικών τρόπων διάγνωσης.

Καταληκτικά παρατηρούμε πως ενώ οι επεμβατικές μέθοδοι είναι επικίνδυνες για το έμβρυο είναι δύσκολο να αντικατασταθούν από τις μη-επεμβατικές για τον κύριο λόγο της αναξιοπιστίας. Άρα λοιπόν πρόκληση για εμάς είναι η δημιουργία ενός συστήματος που θα έχει ψηλά ποσοστά επιτυχίας και χαμηλά ποσοστά λάθους στην πρόβλεψη των χρωμοσωμικών ανωμαλιών.

1.2 Στόχος Διπλωματικής Εργασίας

Ο στόχος της διπλωματικής μου εργασίας είναι η δημιουργία ενός υπολογιστικού συστήματος νοημοσύνης, χρησιμοποιώντας νευρωνικά δίκτυα. Μέσω του συστήματος αυτού γίνεται η έγκαιρη πρόβλεψη της τρισωμίας 21 στο έμβρυο, κατά την πρώτη τριμηνία κύησης. Θα πρέπει μέσω του συστήματος, οι χρήστες (ιατροί, ερευνητές, διαχειριστής) να μπορούν να αλληλεπιδρούν με τέτοιο τρόπο ώστε να γίνεται κατανοητή η λειτουργία και ο σκοπός του συστήματος. Αναλυτικότερα η υλοποίηση έχει ως απώτερο σκοπό να χρησιμοποιηθεί από τους ιατρούς κατά την διάρκεια του προγεννητικού έλεγχου. Το σύστημα θα λειτουργεί ως μια μη-επεμβατική μέθοδος. Μετά το τέλος των προγεννητικών εξετάσεων ο ιατρός μέσω του συστήματος θα βλέπει το ποσοστό πρόβλεψης για την συγκεκριμένη περίπτωση και ανάλογα θα ενημερώνει τους γονείς. Επίσης οι ερευνητές μπορούν να το χρησιμοποιήσουν για νέες ιδέες και μελέτες γύρω από το θέμα. Μπορούν να συνεχίσουν την εργασία αυτή και θα μπορούν με την εφαρμογή τους να βελτιώσουν τα αποτελέσματα.

1.3 Δομή Διπλωματικής Εργασίας

Στο κεφάλαιο 1 γίνεται μια γενική εισαγωγή, οι στόχοι και η δομή γύρω από το θέμα το οποίο μελέτησα. Στο κεφάλαιο 2 παρουσιάζεται ορισμός του προβλήματος. Γίνεται αναφορά τι είναι ο προγεννητικός έλεγχος και πότε γίνεται. Επίσης τι είναι τα χρωμοσώματα και τι τρισωμίες υπάρχουν. Γίνεται ανασκόπηση βιβλιογραφίας και παρουσιάζονται κάποια παρόμοια συστήματα που έγιναν σε άλλες προηγούμενες μελέτες. Στο κεφάλαιο 3 παρουσιάζεται το θεωρητικό κομμάτι της διπλωματικής. Υπάρχει μια εισαγωγή στο τι είναι η Υπολογιστική Νοημοσύνη και τι είναι τα νευρωνικά δίκτυα. Γίνεται μια ιστορική αναδρομή για τα δίκτυα αυτά και μετά αναλυτικά τι είναι τα τεχνητά νευρωνικά δίκτυα. Τέλος παρουσιάζονται οι δύο κύριες μέθοδοι εκπαίδευσης των δικτύων αυτών που είναι η επιβλεπόμενη και η μη-επιβλεπόμενη μάθηση.

Στο κεφάλαιο 4 παρουσιάζονται οι απαιτήσεις που πρέπει να έχει το σύστημα σύμφωνα με τις γνώμες των ειδικών και κάποιες παρατηρήσεις από προηγούμενα συστήματα. Στο κεφάλαιο 5 παρουσιάζονται αναλυτικά οι τεχνικές που ακολουθήθηκαν και η υλοποίηση του συστήματος. Στο κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα από όλα τα νευρωνικά δίκτυα. Τέλος, στο κεφάλαιο 7 καταλήγουμε σε κάποια συμπεράσματα βάση των παρατηρήσεων που έγιναν αναλύοντας τα αποτελέσματα και σημειώνεται προτεινόμενη μελλοντική εργασία έτσι ώστε να βελτιωθούν τα αποτελέσματα.

Κεφάλαιο 2

Περιγραφή Προβλήματος και Ανασκόπηση Βιβλιογραφίας

2.1 Περιγραφή Προβλήματος

Κατά την σύλληψη, δημιουργούνται 23 ζεύγη χρωμοσωμάτων τα οποία κληρονομούνται από τους γονείς. Στην περίπτωση που για κάποιο λόγο κληρονομηθούν περισσότερα χρωμοσώματα, παρουσιάζεται γενετική διαταραχή. Αυτό ονομάζεται χρωμοσωμική ανωμαλία ή αλλιώς τρισωμία. Υπάρχουν διάφορες τεχνικές για την έγκαιρη διάγνωση των χρωμοσωμικών ανωμαλιών. Οι τεχνικές χωρίζονται σε δύο κατηγορίες, στις επεμβατικές και στις μη-επεμβατικές.

Η πιο γνωστή επεμβατική μέθοδος είναι η αμνιοκέντηση. Μέσω της μεθόδου αυτής συλλέγονται σημαντικές γενετικές πληροφορίες για το έμβρυο. Γίνεται με την βοήθεια σύριγγας που τοποθετείται για να εξάγει υγρό από το αμνιακό υγρού στο έμβρυο. Με την ανάλυση του υγρού μπορούν εύκολα και με αξιοπιστία να εντοπιστούν οι χρωμοσωμικές ανωμαλίες.

Οι γονείς καλούνται να αποφασίσουν εάν θα συνεχίσει η εγκυμοσύνη χωρίς περεταίρω εξετάσεις, ή αν θα προχωρήσουν σε έκτρωση του εμβρύου ή να προχωρήσουν σε επεμβατική μέθοδο. Στην επεμβατική μέθοδο υπάρχει ένα μικρό ποσοστό κινδύνου αποβολής και κίνδυνοι για την μητέρα. Ο κίνδυνος αποβολής προβάλλει την ανάγκη για αναζήτηση μεθόδων μη επεμβατικών τρόπων διάγνωσης.

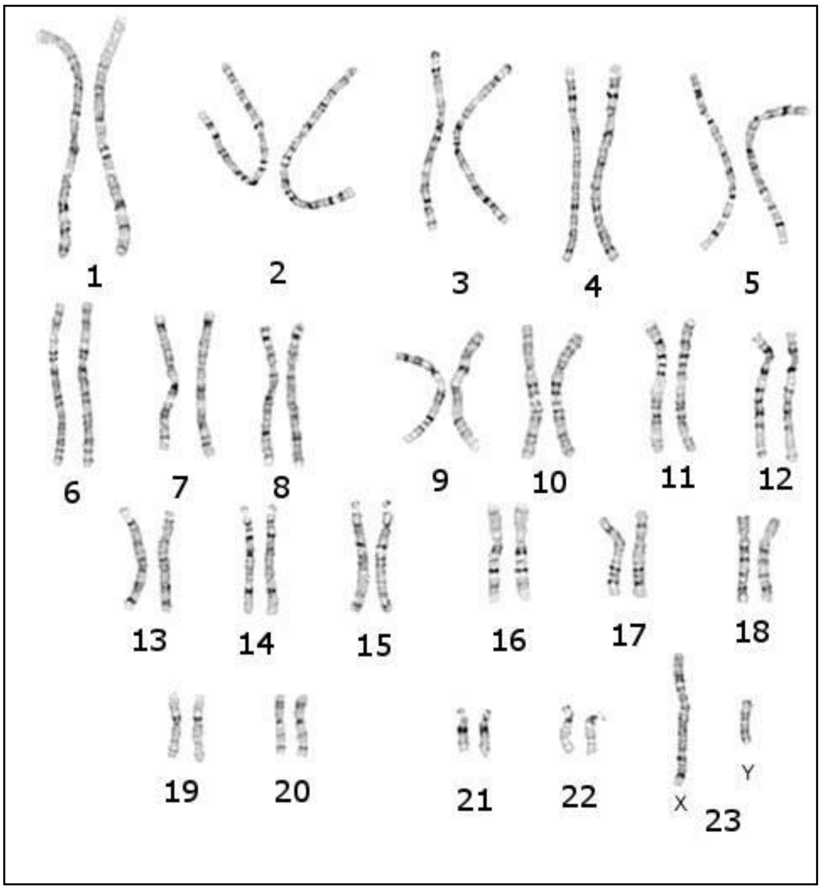
Οι μη-επεμβατικές μέθοδοι χρησιμοποιούν τα χαρακτηριστικά από τον προγεννητικό έλεγχο και μέσω διάφορων μοντέλων κάνουν πρόβλεψη για το αποτέλεσμα της εγκυμοσύνης. Ο κύριος λόγος που ακόμα χρησιμοποιούνται οι επεμβατικές μέθοδοι είναι η αναξιοπιστία των μη-επεμβατικών. Πολλές διαγνωστικές μέθοδοι ανακαλύφθηκαν αλλά χωρίς απόλυτη επιτυχία στην διάγνωση. Έστω και ένα πολύ μικρό ποσοστό λάθους κάνει τους γονείς να προβούν στην επεμβατική λύση. Άρα λοιπόν η πρόκληση για εμάς είναι η δημιουργία ενός συστήματος με πολύ ψηλά ποσοστά επιτυχίας διάγνωσης για το αποτέλεσμα της κύησης.

2.1.1 Χρωμοσώματα

Το χρωμόσωμα είναι μια οργανωμένη δομή του DNA και των πρωτεϊνών που βρίσκεται στα κύτταρα. Κάθε σωματικό κύτταρο του ανθρώπινου οργανισμού υπό φυσιολογικές συνθήκες περιέχει 23 ζεύγη χρωμοσωμάτων (Σχήμα 2.1). Εξαίρεση αποτελούν τα ωάρια και τα σπερματοζωάρια, τα οποία περιέχουν 23 χρωμοσώματα. Κατά τη σύλληψη, με τη σύζευξη σπερματοζωαρίου και ωαρίου δημιουργείται ο ζυγώτης που περιέχει 23 χρωμοσώματα από κάθε γονέα.

Τα πρώτα 22 ζευγάρια λέγονται αυτοσωμικά και είναι ίδια στους άνδρες και γυναίκες. Τα χρωμοσώματα αυτά ονομάζονται αριθμητικά από το μεγαλύτερο (νούμερο 1) ως το μικρότερο (νούμερο 22). Το εικοστό τρίτο ζευγάρι χρωμοσωμάτων αποτελείται από τα λεγόμενα φυλετικά χρωμοσώματα, και καθορίζουν το φύλο, αρσενικό ή θηλυκό , του ατόμου. Το θηλυκό έχει δύο Χ χρωμοσώματα (ΧΧ) ενώ το αρσενικό έχει ένα Χ και ένα Υ χρωμόσωμα.

Η περίπτωση απουσίας ή εμφάνισης ενός πρόσθετου χρωμοσώματος στο ίδιο ζευγάρι χρωμοσωμάτων του εμβρύου λέγεται ανευπλοειδία και δηλώνει χρωμοσωμική ανωμαλία.



Σχήμα 2.1 23 ζεύγη χρωμοσωμάτων διατεταγμένα σύμφωνα με το μέγεθος. Το χρωμόσωμα 1 είναι το μεγαλύτερο. Τα τελευταία δύο χρωμοσώματα είναι τα φυλετικά χρωμοσώματα.

2.1.2 Τρισωμία 21

Σύνδρομο Down ή αλλιώς τρισωμία 21 περιγράφει μια χρωμοσωμική ανωμαλία, που περικλείει ένα σύνολο χαρακτηριστικών, τα οποία υπάρχουν κατά την σύλληψη. Πιο αναλυτικά δημιουργείται ένα παραπάνω χρωμόσωμα στο 21ο ζευγάρι εξού και τρισωμία 21.

Αυτό προκαλεί στα άτομα που γεννιούνται ανωμαλίες στην σωματική διάπλαση, νοητική ανάπτυξη και ψυχοκοινωνική εξέλιξη τους. Τα άτομα με τρισωμία 21 έχουν συνήθως ασυνήθιστο μικρό πιγούνι, λοξά και σχιστά μάτια, κοντό λαιμό, μικρό βάρος σώματος και κοντή μύτη (Εικόνα 2.2). Έχουν μια λεκτική καθυστέρηση αφού οι γλωσσικές δεξιότητες έχουν διαφορές μεταξύ της κατανόησης της ομιλίας και της έκφρασης της ομιλίας. Συνήθως πολλές φορές πάσχουν με προβλήματα ακοής.

Η πιθανότητα εμφάνισης της τρισωμίας 21 οφείλεται από κάποιους παράγοντες όπως η ηλικία της μητέρας, αν είναι σε μεγάλη ηλικία έχει μεγαλύτερη πιθανότητα εμφάνισης της τρισωμίας σε σχέση με τις μητέρες μικρότερης ηλικίας.

Εικόνα 2.2 Μικρό κορίτσι με τρισωμία 21.

Υπάρχουν βέβαια και άλλες τρισωμίες στις οποίες το επιπλέον χρωμόσωμα βρίσκεται σε διαφορετικό χρωμόσωμα. Για παράδειγμα υπάρχει η Τρισωμία 18 (Edward’s Syndrome) και η Τρισωμία 13 (Patau Syndrome).

Η Τρισωμία 18 είναι ένα σπάνιο σύνδρομο που συνδέεται με σοβαρή αναπηρία. Όπως και στην τρισωμία 21, υπάρχει ένα επιπλέον χρωμόσωμα στο 18^{ο} ζευγάρι για αυτό και ονομάζεται Τρισωμία 18. Αυτή η τρισωμία επηρεάζει σε σημαντικό βαθμό όλα τα οργανικά συστήματα του σώματος.
Η Τρισωμία 13 είναι μια σπάνια τρισωμική διαταραχή όπου το άτομο γεννιέται με ένα επιπλέον χρωμόσωμα στο 13ο ζεύγος. Στις περισσότερες περιπτώσεις εκδηλώσεις οι παράγοντες δεν είναι κληρονομικές αλλά από διάφορα συμβάντα κατά την διάρκεια της

διαμόρφωσης των αναπαραγωγικών κυττάρων (σπερματοζωαρίων και ωαρίων). Σε πολλές περιπτώσεις το έμβρυο δεν επιζεί και αποβάλλεται από τον οργανισμό της μητέρας.

2.2 Προγεννητικός έλεγχος στις εγκυμονούσες

Ο προγεννητικός έλεγχος αναφέρεται σε μια διαδικασία διάφορων εξετάσεων που στόχο έχουν την αναγνώριση των γυναικών που έχουν αυξημένο κίνδυνο για παθολογικές διαταραχές όπως αυξημένο δείκτη γλυκόζης, απότομη αύξηση σωματικού βάρους ή ψηλή αρτηριακή πίεση. Μέσω του προγεννητικού έλεγχου μπορεί να αναγνωριστεί ο κίνδυνος εμφάνισης εμβρυικής ανωμαλίας κατά την σύλληψη ή κατά την διάρκεια της εγκυμοσύνης τους. Αυτό συνήθως παρατηρείται μετά το υπερηχογράφημα στο οποίο ελέγχεται κυρίως η αυχενική διαφάνεια και το ρινικό οστό του εμβρύου. Οι εξετάσεις που γίνονται στον προγεννητικό έλεγχο είναι η ανάλυση αίματος της μητέρας, το υπερηχογράφημα και η μελέτη του ιατρικού προϊστορικού της μητέρας.

2.3 Ανασκόπηση Βιβλιογραφίας - Άλλα παρόμοια Συστήματα

Στην μελέτη των Neocleous et. al [8] είχε γίνει δημιουργία τεχνητών νευρωνικών δικτύων για την πρόβλεψη του κινδύνου για τρισωμία 21 ή άλλων συνδρόμων μέσω του υπερηχογραφήματος κατά τις 11-13 εβδομάδες κύησης. Δημιουργήθηκαν για αυτό το σκοπό δύο συστήματα. Το πρώτο σύστημα εκπαιδεύτηκε για την αναγνώριση της τρισωμίας 21 ή της κανονικής κατάστασης. Το σύστημα αυτό προσδιόριζε σωστά όλες τις περιπτώσεις με τρισωμία 21 και κάποιο ποσοστό των άλλων συνδρόμων όπως τρισωμία 13, τρισωμία 18, Σύνδρομο Turner και Σύνδρομο Triploid με ένα ποσοστό 3.9% σφάλμα θετικού λάθους. Το δεύτερο σύστημα εκπαιδεύτηκε για την αναγνώριση των με κανονική κατάσταση με τα άλλα σύνδρομα, και στην συνέχεια την αναγνώριση αυτών με τρισωμία 21 από τα άλλα σύνδρομα. Το αποτέλεσμα του συστήματος ήταν η αναγνώριση όλων των περιπτώσεων με τρισωμία 21 και ένα ποσοστό γύρω στο 88.2% των άλλων συνδρόμων με ποσοστό σφάλματος στο 3.9%.

Στην μελέτη των Cigero et. al [9] είχε μελετηθεί ο λόγος πιθανότητας εμφάνισης τρισωμίας 21 σε έμβρυα με απουσία ρινικού οστού (nasal bone) στο υπερηχογράφημα των 11-14 εβδομάδων κύησης. Είχε παρατηρηθεί η απουσία ή παρουσία του ρινικού οστού σε 5918 έμβρυα πριν από την καρυοτυπία. Χρησιμοποιήθηκε λογιστική παλινδρόμηση ανάλυσης για να εξεταστεί η επίδραση της εθνικής προέλευσης της μητέρας, του μήκους κεφαλουραίου του εμβρύου (CRL) και της αυχενικής διαφάνειας (NT) δηλαδή την ποσότητα του υγρού στον αυχένα του εμβρύου, για την απουσία του ρινικού οστού στα κανονικά έμβρυα και στα έμβρυα με τρισωμία 21. Η ανάλυση λογιστικής παλινδρόμησης κατέδειξε ότι στα φυσιολογικά έμβρυα σημαντική ανεξάρτητη πρόβλεψη της πιθανότητας είναι η απουσία του ρινικού οστού που παρέχονται από CRL, NT και συγκεκριμένης εθνικής ομάδας, και των εμβρύων με τρισωμία 21 από το CRL και NT. Ο λόγος πιθανότητας για τρισωμία 21 που απουσιάζει ο ρινικός οστός προήλθε από τη διαίρεση της πιθανότητας στην τρισωμία 21 με τα φυσιολογικά έμβρυα. Στο τέλος συμπεραίνεται πως η απουσία ή παρουσία του ρινικού οστού σχετίζεται άμεσα από το CRL, NT και την εθνική ομάδα της μητέρας.

Στην μελέτη των Nicolaides et. al [10] μελετήθηκε για το πρώτο τρίμηνο ο έλεγχος εμφάνισης τρισωμίας 21 σε 75821 κυήσεις. Τα αποτελέσματα και οι εκτιμήσεις των πιθανών επιπτώσεων μεμονωμένων κινδύνων διαχωρίζονται σε δύο στάδια ελέγχου το πρώτο τρίμηνο. Συγκεκριμένα αξιολογείται η εμφάνιση της τρισωμίας 21 στο πρώτο τρίμηνο με συνδυασμό της ηλικίας της μητέρας με την αυχενική διαφάνεια του εμβρύου (NT), την χοριακή γονιαδοτροπίνη (b-hCG) και την πρωτεΐνη A (PAPP-A). Μετά από κάποιους υπολογισμούς ο μέσος όρος ηλικίας της μητέρας είναι στα 31 χρόνια με εύρος τιμών από 13-49, ο μέσος όρος των εβδομάδων κύησης είναι στις 12 με εύρος τιμών από 11-13 εβδομάδες και η μέση τιμή του CRL είναι 62mm με εύρος τιμών από 45-83mm. Τέλος το συμπέρασμα της μελέτης αυτής έρχεται να επιβεβαιώσει πως ο συνδυασμός των παραμέτρων αυτών στο πρώτο τρίμηνο της εγκυμοσύνης μπορεί να δώσει αποτέλεσμα ποσοστού αναγνώρισης τρισωμίας 21 στο 90% με ποσοστό λάθους στο 5% και για ατομικό έλεγχο σε δύο στάδια υπάρχει ένα ποσοστό ενδεχόμενης αναγνώρισης τρισωμίας 21 στο 90% με ποσοστό λάθους στο 2-3%.

Στην μελέτη των Persico et. al [11] είχε ως σκοπό την διερεύνηση του μήκους και σχήματος του ρινικού οστού σε έμβρυα με τρισωμία 21 στις 16-24 εβδομάδες κύησης σε τρισδιάστατη μορφή. Η μέθοδος που χρησιμοποιήθηκε ήταν η προβολή του ρινικού οστού σε τρισδιάστατη μορφή για 41 έμβρυα με τρισωμία 21, για να γίνουν ακριβές μετρήσεις του μήκους του καθώς επίσης και του σχήματος του. Καταληκτικά γνωρίζουμε πως το ρινικό οστό είναι σημαντική παράμετρος για την εμφάνιση της τρισωμίας 21 στο έμβρυο αλλά το σημείο που εστιάζεται η μελέτη αυτή είναι η μέτρηση θα πρέπει πάντα να γίνεται στην ακριβές διάμεσος του επίπεδου προσώπου και αυτό βοηθάει με τρισδιάστατη μορφή.

Στην μελέτη των Syngelaki et. al [12] γίνεται η διάγνωση των εμβρύων με μη χρωμοσωμικές ανωμαλίες στις 11-13 εβδομάδες κυήσεις. Η μέθοδος που χρησιμοποιήθηκε ήταν η αναγνώριση των εμβρύων με κάποιο σύνδρομο κατά τις 11-13 εβδομάδες κύησης σε 45191 κυήσεων. Αυτά έγιναν σύγκριση με τα έμβρυα των 20-23 εβδομάδων και σύγκριση μετά τον τοκετό δηλαδή το αποτέλεσμα της εγκυμοσύνης. Αυτή η μελέτη είχε ως συμπέρασμα ότι στις 11-13 εβδομάδες κύησης κάποιες χρωμοσωμικές ανωμαλίες είναι πάντα ανιχνεύσιμες, κάποιες ποτέ δεν μπορούν να είναι και κάποιες ανιχνεύονται ανάλογα με την αυξημένη αυχενική διαφάνεια (NT) και την φαινοτυπική έκφραση της ανωμαλίας της κύησης.

Κεφάλαιο 3

Απαιτούμενη Γνώση και Τεχνολογίες

3.1 Υπολογιστική Νοημοσύνη

Οι όροι Τεχνητή Νοημοσύνη και Υπολογιστική Νοημοσύνη χρησιμοποιούνται με τέτοιο τρόπο που έχουν την ίδια σημασιολογία. Ο όρος Τεχνητή Νοημοσύνη δηλώνει το κάθε είδος νοημοσύνης που δεν είναι βιολογική. Ο όρος Υπολογιστική Νοημοσύνη αφορά κάθε είδος νοημοσύνης που μπορεί να εκφραστεί με υπολογιστικές διαδικασίες. Είναι κλάδος της Πληροφορικής που ασχολείται με την σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων μέσω των οποίων παρουσιάζονται οι ανθρώπινες ικανότητες έχοντας κάποια χαρακτηριστικά που μιμούνται την ανθρώπινη συμπεριφορά. Μπορούμε να πούμε πως η υπολογιστική νοημοσύνη σχετίζεται άμεσα με την Τεχνητή Νοημοσύνη.

Τα χαρακτηριστικά της υπολογιστικής νοημοσύνης εντάσσονται συνήθως σε προβλήματα που μόνο οι άνθρωποι και τα ζώα μπορούν να τα λύσουν, δηλαδή προβλήματα που χρειάζονται νοημοσύνη. Αρκετοί κλάδοι της Υπολογιστικής Νοημοσύνης μετά από μελέτες και συνέδρια που έγιναν άλλαξαν κατευθύνσεις και εξελίσσονται διαφορετικά σε σχέση με τις αρχικές τους ιδέες. Η κλασσική Τεχνητή Νοημοσύνη χρησιμοποιούσε συμβολική αναπαράσταση της γνώσης και τεχνικές ανάλυσης υψηλού επιπέδου. Δηλαδή μιμείται τις απευθείας ψηλού επιπέδου ανθρώπινες συμπεριφορές, συλλογισμούς και μηχανισμούς σκέψης.

Η προσέγγιση αυτή είχε σοβαρούς περιορισμούς, και έφτασε σε ένα σημείο που δεν μπορούσε να προσφέρει τίποτε καινούργιο. Η έρευνα στράφηκε σε τεχνικές ανάλυσης χαμηλού επιπέδου. Δηλαδή να μιμείται τον τρόπο με τον οποίο συνεκτικές και δομημένες συμπεριφορές προκύπτουν από μη δομημένες και μη οργανωμένες μονάδες. Δηλαδή προκύπτουν μέσω συνεργασίας και συλλειτουργίας. Η βασική ιδέα ήταν η εξής: αντί να αναπαριστούμε την γνώση σε ένα υψηλό επίπεδο, αντικαθιστούμε τα μοντέλα αυτά με αλγορίθμους που στηρίζονται στην αριθμητική επεξεργασία της πληροφορίας. Έτσι κατασκευάστηκαν αλγόριθμοι οι οποίοι αναλύουν και μαθαίνουν από δεδομένα που ήδη υπάρχουν και περιγράφουν την εφαρμογή. Το αποτέλεσμα ήταν η μείωση αναζήτησης των αναλυτικών μοντέλων και ταυτόχρονη αύξηση αναζήτησης δεδομένων.

Κάποιες μέθοδοι που χρησιμοποιούνται στην Υπολογιστική Νοημοσύνη είναι: τα Τεχνητά Νευρωνικά Δίκτυα, η Ασαφής Λογική (Fuzzy Logic), Support Vectors Machines, K-means και K-nearest neighbors, Evolutionary algorithms, Uncertain variables, Probabilistic methods and Rough Sets.

3.2 Νευρωνικά Δίκτυα

3.2.1 Ιστορική Αναδρομή

Η ιστορία των νευρωνικών δικτύων ξεκίνησε στα τέλη του 1800 μετά από προσπάθειες των ερευνητών για την μελέτη στο πώς λειτουργεί ο ανθρώπινος εγκέφαλος. Το 1943 από τους McCulloch και Pitts [3] παρουσιάσθηκε το μοντέλο νευρωνικού δικτύου το οποίο λέει πως οι νευρώνες είναι η βασική μονάδα του δικτύου. Σε μία εργασία τους οι ερευνητές αυτοί παρουσίασαν την ιδέα ότι ένα νευρωνικό δίκτυο αποτελείται από μία συλλογή ενός μεγάλου αριθμού νευρώνων και έδειξαν πώς θα μπορούσαν να λειτουργούν οι νευρώνες με τις διασυνδέσεις τους. Αυτή θεωρείται ιστορικά ότι είναι η πρώτη εικόνα ενός νευρωνικού δικτύου.

Ο J.Von Neumann [5] χρησιμοποίησε τις πιο πάνω εργασίες ως παράδειγμα για υπολογιστικές μηχανές την εποχή που διαδόθηκε ο ηλεκτρονικός υπολογιστής. Στην εποχή αυτή έγιναν οι πρώτες προσπάθειες να πάρουν πληροφορίες σε κυτταρικό επίπεδο από τα βιολογικά δίκτυα και να παραχθούν τα πρώτα τεχνητά δίκτυα.

Το βιβλίο του D. Hebb «The organisation of behavior» [4], το οποίο εισάγει τον κανόνα μάθησης του Hebb είναι ένα άλλο έργο της εποχής που ακόμα και σήμερα έχει επιρροή. Το μοντέλο του Hebb έχει ως κεντρική ιδέα τις συνδέσεις μεταξύ μονάδων του συστήματος, δηλαδή τους νευρώνες. Ο κανόνας μάθησης λέει ότι σε κάθε προσπάθεια που το δίκτυο χρησιμοποιεί τις νευρωνικές συνδέσεις, αυτές ενισχύονται και το δίκτυο τείνει περισσότερο στο να μάθει πιο εύκολα το πρότυπο το οποίο παρουσιάζεται.

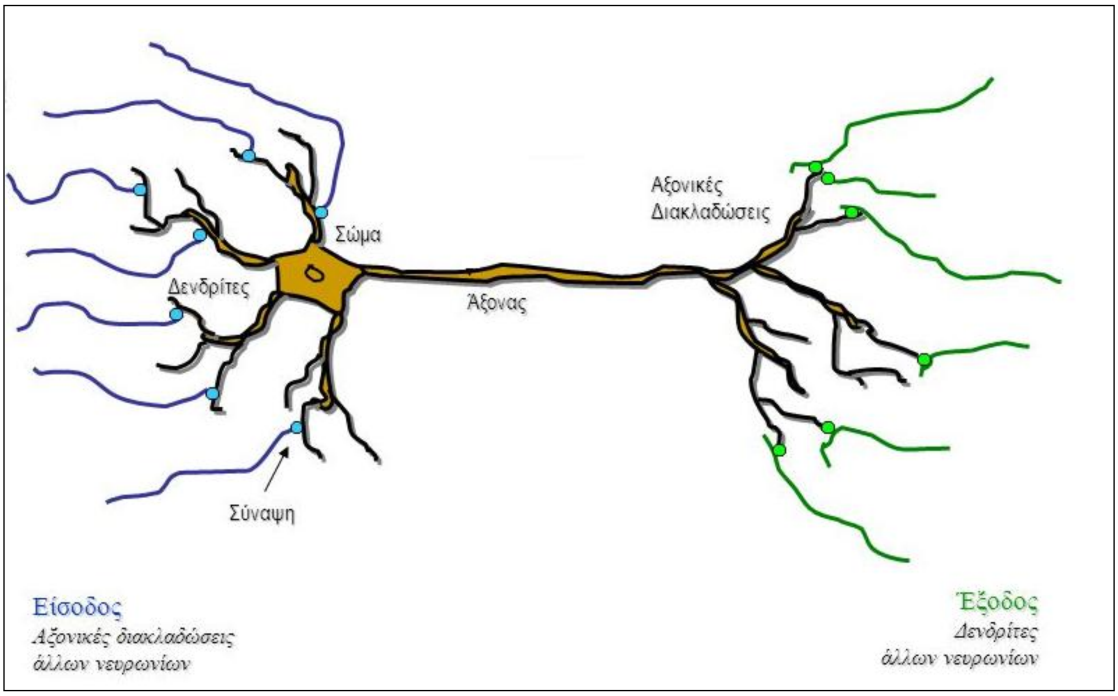
Το μοντέλο του αισθητήρα (Perceptron) παρουσιάσθηκε για πρώτη φορά από τον F. Rosenblatt το 1957 [22]. Είναι ένα απλό μοντέλο με δύο επίπεδα, της εισόδου και της εξόδου, όπου το σήμα κατευθύνεται σε μια διαδρομή από την είσοδο στην έξοδο.

Το 1986 δημοσιεύεται ένα άλλο σημαντικό έργο από τους McClelland και Rumelhart, το «Parallel Distributed Processing» [6], το οποίο ανοίγει νέους δρόμους στην εκπαίδευση των νευρωνικών δικτύων. Το έργο αυτό κάνει ένα σημαντικό βήμα πέρα από το Perceptron, με το να επιτρέπει την ύπαρξη και άλλων επιπέδων νευρώνων, εκτός από την είσοδο και την έξοδο, που αποτελούν την εσωτερική δομή του δικτύου. Προτείνουν μία νέα διαδικασία εκπαίδευσης, την μέθοδο της οπισθοδιάδοσης (back–propagation), η οποία κατέληξε να είναι η πιο χρήσιμη σήμερα τεχνική εκπαίδευσης δικτύων.

3.2.2 Βιολογικά Νευρωνικά Δίκτυα

Ο ανθρώπινος εγκέφαλος εκτελεί τους υπολογισμούς με ένα διαφορετικό τρόπο από τον υπολογιστή. Αυτό ήταν η αιτία στην ανάπτυξη του έργο σε επιστημονικό πεδίο των Τεχνητών Νευρωνικών Δικτύων. Ο νευρώνας είναι ένας τύπος κυττάρου και αποτελεί το βασικό κομμάτι στην επεξεργασία πληροφοριών για το νευρικό σύστημα του ανθρώπου. Είναι το σημαντικότερο στοιχείο του εγκεφάλου στον άνθρωπο και στα ζώα. Στον ανθρώπινο εγκέφαλο υπάρχει ο νευρώνας που έχει μια διακλαδωτική διάρθρωση εισόδων, τους δενδρίτες, και μια διακλαδωτική δομή εξόδων, τον άξονα (Σχήμα 3.1). Στο τέλος κάθε διακλάδωσης, μια δομή η οποία ονομάζεται σύναψη μετατρέπει σε ηλεκτρικούς παλμούς την ενέργεια από τον άξονα. Ένας νευρώνας στέλνει παλμό ηλεκτρικής ενέργειας κατά μήκος του άξονα σε άλλους νευρώνες που είναι συνδεδεμένοι όταν δέχεται διεγερτικές εισόδους.

Ο μέσος ανθρώπινος εγκέφαλος αποτελείται από 10^9 νευρώνες. Επιπρόσθετα δεν είναι πλήρως γνωστό πως ο εγκέφαλος κάνει την εκπαίδευση του έτσι ώστε να συλλέγει τις πληροφορίες. Η μάθηση επιτυγχάνεται με την αλλαγή του φορτίου στις συνάψεις, τέτοιο ώστε η επιρροή ενός νευρώνα σε κάποιο άλλο να αλλάζει. Οι νευρώνες αποτελούνται από το σώμα, τον άξονα και τους δενδρίτες.



Σχήμα 3.1 Διάταξη βιολογικού νευρώνα

3.2.3 Τεχνητά Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα ανήκουν στον τομέα της Υπολογιστικής Νοημοσύνης. Είναι υπολογιστικά συστήματα μάθησης που επικεντρώνονται κυρίως στα βιολογικά νευρωνικά δίκτυα. Συγκεκριμένα είναι αλγόριθμοι που προσπαθούν να αναπαραστήσουν την λειτουργία του ανθρώπινου εγκεφάλου. Το βασικό στοιχείο του ενός νευρωνικού δικτύου είναι ο νευρώνας. Ένα Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ) αποτελείται από πολλά συνδεδεμένα μεταξύ τους στοιχεία επεξεργασίας (νευρώνες), τα οποία εργάζονται για την επίλυση συγκεκριμένων προβλημάτων. Τα ΤΝΔ, όπως και οι άνθρωποι, μαθαίνουν από παραδείγματα και ρυθμίζονται έτσι ώστε να μπορούν να χρησιμοποιηθούν σε συγκεκριμένες εφαρμογές, όπως στο πρόβλημα μας δηλαδή την πρόβλεψη χρωμοσωμικών ανωμαλιών στο έμβρυο.

Σε ένα ΤΝΔ ένας αριθμός αντιστοιχεί στις συνδέσεις μεταξύ των νευρώνων ο οποίος ονομάζεται βάρος και χρησιμοποιείται για την τροποποίηση των εισόδων του νευρώνα. Κάθε νευρώνας δέχεται ένα σύνολο αριθμητικών εισόδων, τα συναπτικά βάρη των συνδέσεων, μία συνάρτηση ενεργοποίησης και μια έξοδο. Η εμπειρία και η γνώση που αποκτά το ΤΝΔ αποθηκεύεται στα βάρη του. Συγκεκριμένα το ΤΝΔ έχει τρία κύρια στοιχεία. Ένα σύνολο από συνάψεις ή συνδετικούς κρίκους, ένα αθροιστή και μια συνάρτηση ενεργοποίησης. Τα σήματα στην είσοδο της σύναψης που συνδέονται με τον νευρώνα, πολλαπλασιάζονται με το συναπτικό βάρος. Ένα σήμα Xj στην είσοδο της σύναψης j που συνδέεται στον νευρώνα k, πολλαπλασιάζεται με το συναπτικό βάρος Wkj. Μέσω του αθροιστή γίνεται η πρόσθεση των σημάτων εισόδου. Τέλος με την συνάρτηση ενεργοποίησης γίνεται η μείωση του εύρους της εξόδου του νευρώνα. Ορίζει την έξοδο ενός νευρώνα με βάση το επίπεδο ενεργοποίησης της εισόδου. Η συνάρτηση ενεργοποίησης μπορεί να είναι βηματική, στοχαστική, γραμμική και μη γραμμική. Υπάρχουν τρεις βασικοί τύποι συναρτήσεων. Η συνάρτηση κατωφλίου, η γραμμική συνάρτηση και η σιγμοειδής συνάρτηση.

Το Νευρωνικό Δίκτυο μέσω της εκπαίδευσης μαθαίνει το περιβάλλον του και βελτιώνεται σταδιακά η απόδοση του. Η εκπαίδευση επιτυγχάνεται μέσω μιας επαναληπτικής διαδικασίας ρυθμίσεων των συναπτικών βαρών.

Υπάρχει μεγάλη ποικιλία αλγορίθμων μάθησης, με τον καθένα να έχει τα πλεονεκτήματα και τα μειονεκτήματα του. Υπάρχουν τα δίκτυα εμπρός-τροφοδότησης πολλών επιπέδων που αποτελούνται από επίπεδο εισόδου που είναι ένα σύνολο αισθητήρων για είσοδο, από ένα ή περισσότερα κρυφά επίπεδα υπολογιστικών κόμβων και ένα επίπεδο εξόδου. Υπάρχει και ο αλγόριθμος πίσω διάδοσης λάθους (Back Propagation Algorithm) στον οποίο τα δίκτυα εκπαιδεύονται με έναν επιβλεπόμενο τρόπο. Ο αλγόριθμος αυτός βασίζεται στον κανόνα μάθησης διόρθωσης του λάθους. Υπάρχει μια διαδικασία της πίσω διάδοσης του λάθους που γίνεται από δύο περάσματα, διαμέσου των διαφορετικών επιπέδων του δικτύου. Το ένα προς τα εμπρός πέρασμα και το ένας προς τα πίσω πέρασμα. Ο αλγόριθμος Back Propagation έχει σαν πλεονέκτημα ότι είναι εύκολο στην χρήση λόγω του ότι μπορεί να εφαρμοστεί σε μεγάλες ποσότητες δεδομένων και σαν μειονέκτημα ότι η εκμάθηση του είναι αργή.

Τέλος η επιλογή της μεθόδου γίνεται αναλόγως του προβλήματος αν και συνήθως χρησιμοποιούνται συνδυασμοί μεθόδων για να λυθούν τα υπάρχον προβλήματα. Όλες οι μέθοδοι εκπαίδευσης των νευρωνικών δικτύων κατατάσσονται σε δύο κατηγορίες: την επιβλεπόμενη μάθηση (supervised learning) και την μη επιβλεπόμενη μάθηση (unsupervised learning).

3.3 Μέθοδοι Εκπαίδευσης Νευρωνικών Δικτύων

3.3.1 Επιβλεπόμενη Μάθηση

Επιβλεπόμενη μάθηση είναι η πιο κοινή τεχνική για την εκπαίδευση των νευρωνικών δικτύων. Βασίζεται στην εκπαίδευση ενός δείγματος δεδομένων γνωρίζοντας την είσοδο και την σωστή έξοδο. Οι τεχνικές που χρησιμοποιούνται είναι η εμπρόσθια διάδοση ή το μοντέλο του Perceptron με πολλαπλά επίπεδα (MLP). Τα MLP έχουν 3 κύρια είδη χαρακτηριστικών. Έχουν ένα ή περισσότερα κρυφά επίπεδα, που δεν αποτελούν μέρος των επιπέδων εισόδου ή εξόδου του δικτύου, και επιτρέπουν στο δίκτυο να μάθει και να λύσει τα περίπλοκα προβλήματα. Η διασύνδεση του δικτύου παρουσιάζει ψηλό βαθμό συνδεσιμότητας. Τα χαρακτηριστικά αυτά σε συνδυασμό με την μάθηση επιλύουν τα δύσκολα και ποικίλα προβλήματα. Μπορεί να γίνει με διάφορους αλγόριθμους όπως ο αλγόριθμος ανάστροφης διάδοσης σφάλματος (Back Propagation Algorithm). Η επιβλεπόμενη μάθηση είναι αποτελεσματική και βρίσκει λύσεις σε διάφορα γραμμικά και μη-γραμμικά προβλήματα όπως το πρόβλημα της ταξινόμησης ή προβλήματα πρόβλεψης.

3.3.2 Μη Επιβλεπόμενη Μάθηση

Στην μη-επιβλεπόμενη μάθηση, λόγω έλλειψης γνώσης για τις κατηγορίες του προβλήματος μας, η μάθηση γίνεται στην εύρεση της ύπαρξης κλάσεων ή συστάδων μεταξύ των δεδομένων. Πιο συγκεκριμένα ο στόχος στην μη-επιβλεπόμενη μάθηση είναι να εκπαιδευτεί το σύστημα χωρίς να του πούμε σε ποια κατηγορία ανήκει η κάθε περίπτωση. Δηλαδή το δίκτυο διαθέτει εισόδους αλλά όχι τα επιθυμητά αποτελέσματα. Η μη-επιβλεπόμενη μάθηση αναφέρεται στην ικανότητα να μαθαίνει και να οργανώνει τις πληροφορίες χωρίς να παρέχεται κάποιο σήμα σφάλματος για την αξιολόγηση της πιθανής λύσης. Το σύστημα δημιουργεί ομάδες με βάση τα πρότυπα εισόδου. Η δημιουργία τέτοιων ομάδων εξασφαλίζεται με διάφορους εξειδικευμένους αλγόριθμους, όπως για παράδειγμα με τον αλγόριθμο k-means και τον αλγόριθμο της αυτοργάνωσης (Self Organizing Map) του Kohonen [7]. Τα κύρια χαρακτηριστικά του αλγόριθμου SOM είναι πως μετατρέπει ένα εισερχόμενο σήμα σε μία ή δύο διαστάσεις χάρτη και κάνουν τον μετασχηματισμό προσαρμοστικά. Το δίκτυο παρουσιάζει την δομή εμπροσθόδρομης τροφοδοσίας (feedforward) με ένα μόνο στρώμα που αποτελείται από νευρώνες που είναι διατεταγμένοι σε σειρές και στήλες. Σε κάθε παράσταση το σήμα εισόδου διατηρείται στο σωστό πλαίσιο. Οι νευρώνες που έχουν κοινές πληροφορίες είναι κοντά μεταξύ τους και επικοινωνούν μέσω συναπτικών συνδέσεων. Το επίπεδο αυτό ονομάζεται ανταγωνιστικό επίπεδο αφού οι νευρώνες ανταγωνίζονται μεταξύ τους για να γίνουν ενεργή. Ως εκ τούτου, αυτός ο αλγόριθμος μάθησης ονομάζεται και ανταγωνιστικός αλγόριθμος.

3.4 Λογισμικά Εργαλεία

3.4.1 Matlab

Το MATLAB είναι ένα ολοκληρωμένο περιβάλλον προγραμματισμού, καθώς επίσης υψηλού επιπέδου γλώσσα προγραμματισμού για τεχνικούς υπολογιστών που χρησιμοποιείται για την μελέτη και εφαρμογή κάθε είδους μαθηματικών συναρτήσεων. Επίσης, χρησιμοποιείται για αριθμητικούς υπολογισμούς, ειδικά στην γραμμική άλγεβρα με πίνακες. Έχει γίνει ένα δυνατό εργαλείο στην μηχανική καθώς επίσης και στις επικοινωνίες. Στο περιβάλλον Matlab υπάρχουν πολλές δυνατότητες για τον χειρισμό μεγάλης ποσότητας δεδομένων. Διαθέτει ευκολία και ταχύτητα στην υλοποίηση μοντέρνων αλγορίθμων και πολλών έτοιμων συναρτήσεων που υλοποιούνται μέσω της γραμμής εργαλείων.

Στην διπλωματική μου εργασία, το Matlab με βοήθησε στην επεξεργασία των δεδομένων και στην υλοποίηση των μοντέλων των νευρωνικών δικτύων.

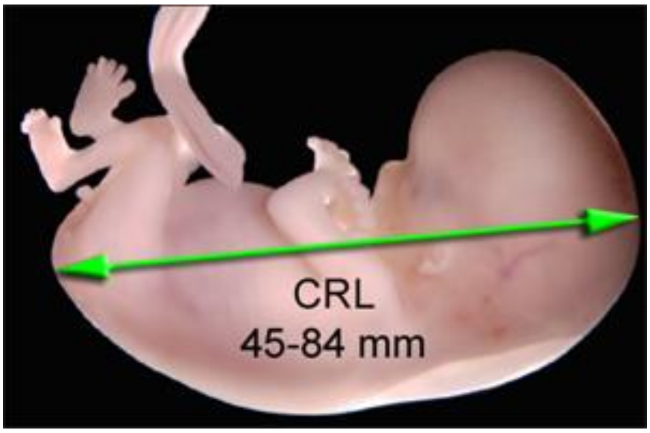
3.5 Δεδομένα

Για την μελέτη, χρησιμοποίησα μια βάση δεδομένων που συλλέχθηκε και παραχωρήθηκε από το Fetal Medicine Foundation. Η βάση δεδομένων αποτελείται από 30482 περιπτώσεις από εγκυμονούσες γυναίκες. Για κάθε γυναίκα, υπάρχουν όλες οι μετρήσεις του προγεννητικού έλεγχου, που έγιναν κατά την πρώτη τριμηνία καθώς επίσης και το αποτέλεσμα της διάγνωσης δηλαδή αν το παιδί έχει Τρισωμία 21.

Σε αυτή την βάση δεδομένων υπάρχουν 5 διαφορετικές κατηγορίες περιπτώσεων χρωμοσωμικών ανωμαλιών και μια κατηγορία με περιπτώσεις φυσιολογικών εμβρύων. Συγκεκριμένα υπάρχουν 380 περιπτώσεις με τρισωμία 21, 154 περιπτώσεις με τρισωμία 18, 49 περιπτώσεις με τρισωμία 13, 34 περιπτώσεις με Σύνδρομο Triploid και 73 περιπτώσεις με Σύνδρομο Turner. Σε αυτή την διπλωματική φτιάξαμε μοντέλα μόνο για την αναγνώριση της Τρισωμίας 21. Οι υπόλοιπες 29792 περιπτώσεις είναι χωρίς κάποια τρισωμία δηλαδή φυσιολογικά έμβρυα. Εδώ παρατηρούμε πως υπάρχει μεγάλη ανισορροπία στο πληθυσμό μεταξύ των περιπτώσεων με κάποιο σύνδρομο και των περιπτώσεων φυσιολογικών εμβρύων. Αυτό είναι ένα σημείο που κάνει το δίκτυο μας πιο δύσκολο ως προς την εκπαίδευση.

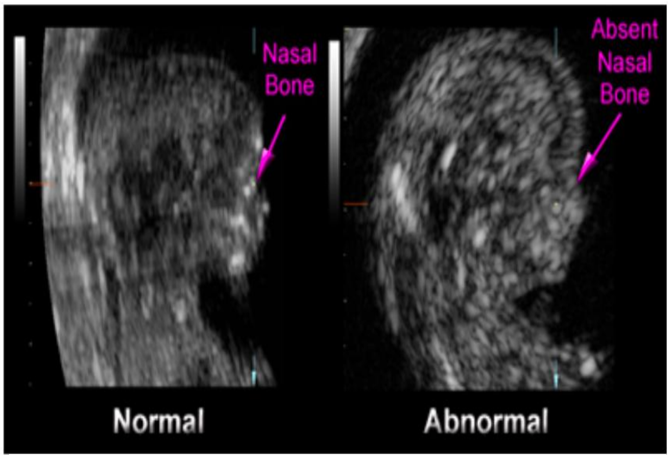
Τίτλος	**Περιγραφή**
Κεφαλουραίο μήκος (CRL)	Είναι το μήκος του εμβρύου (Σχήμα 2.2) σε εκατοστόμετρα, από την κορυφή της κεφαλής ως το κάτω μέρος των γλουτών και υπολογίζεται από τον υπέρηχο.
Αυχενική Διαφάνεια (NT)	Η αυχενική διαφάνεια (Σχήμα2.3) είναι η παρουσία υποδόριου υγρού πίσω από τον αυχένα του εμβρύου κατά τη διάρκεια του πρώτου τριμήνου της κύησης. Οι κόκκινες γραμμές μας δείχνουν τα όρια στα οποία μετριέται η αυχενική διαφάνεια στο έμβρυο.
Πρωτεΐνη A (PAPP-A)	Είναι μια πρωτεΐνη που υπάρχει στο αίμα της εγκύου.
Χοριακή γονιαδοτροπίνη (b-hCG)	Είναι μια ορμόνη που παράγεται από τον πλακούντα κατά τη διάρκεια της εγκυμοσύνης, με σκοπό να εμποδίζει την εμμηνόρροια και να προστατεύει την εγκυμοσύνη.
Καρδιακός Σφυγμός Εμβρύου	Η μέτρηση του καρδιακού σφυγμού εμβρύου.
Ρινικό Οστό (Nasal Bone)	Είναι τμήμα οστού που βρίσκεται ανάμεσα στις κόγχες των ματιών και εξέχει προς τα έξω. Μπορούμε να δούμε την διαφορά στην ύπαρξη ρινικού οστού από την πρώτη εικόνα όπως δείχνει το αριστερά τόξο σε σχέση με το δεξιό (Σχήμα 2.4). Η ύπαρξη ρινικού οστού δίνει θετικό προμήνυμα για την υγεία του εμβρύου σύμφωνα ιατρικές έρευνες.
Τριγλώχινα ροή (Tricuspid Flow)	Είναι ροή αίματος στην τριγλώχινα βαλβίδα (Σχήμα 2.5). Στην πρώτη γραφική παράσταση με γαλάζιο χρώμα βλέπουμε την μη κανονική μορφή της τριγλώχινας ροής η οποία παίρνει ψηλές αρνητικές τιμές.

Ροή φλεβώδους πόρου (Ductus Venosus Flow)	Ο φλεβώδης πόρος είναι ένα μικρό αγγείο που συνδέει την ομφαλική φλέβα με την κάτω κοίλη φλέβα και βοηθά στην μεταφορά οξυγονωμένου αίματος στον εγκέφαλο του εμβρύου.
Ηλικία Μητέρας	Η καταγραφή της ηλικίας είναι σημαντικός παράγοντας.
Βάρος Μητέρας	Η καταγραφή του βάρους με συνδυασμό με άλλους δείκτες συντελεί στους υπολογισμούς μιας ανωμαλίας.
Κάπνισμα	Δυαδικό χαρακτηριστικό που ορίζει αν η εγκυμονούσα είναι καπνίστρια ή όχι.
Κληρονομικότητα (Τρισωμία 21/ Τρισωμία 18/ Τρισωμία 13)	Είναι 3 διαφορετικά δυαδικά χαρακτηριστικά που προσδιορίζει αν υπήρξε κάποια τρισωμία σε προηγούμενη γέννηση παιδιού από τους συγκεκριμένους γονείς.
Εθνικότητα	Παράμετρος που μας δείχνει από ποια εθνικότητα προέρχεται η εγκυμονούσα.
Σύλληψη	Ο τρόπος σύλληψης του εμβρύου.
Διαβήτης	Σε περιπτώσεις αυξημένου διαβήτη μεγαλώνει η πιθανότητα για παιδιά με χρωμοσωμική ανωμαλία.

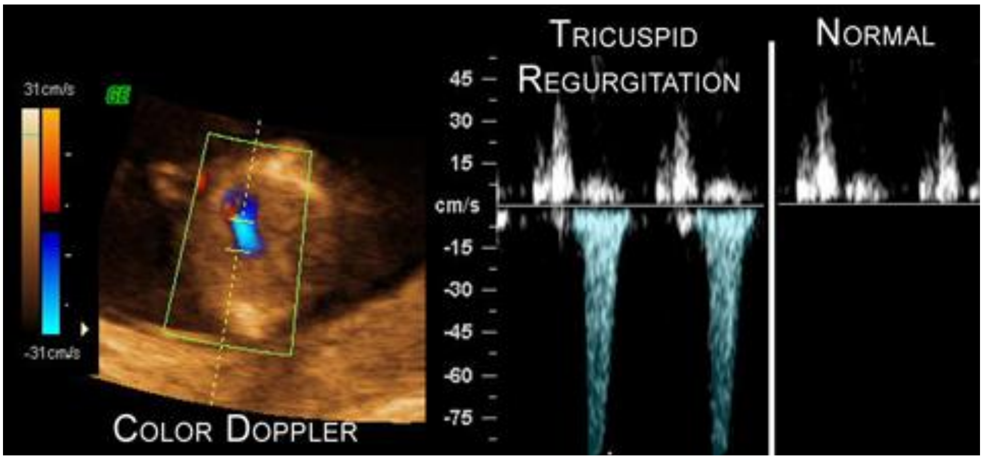


Σχήμα 2.2 Το μήκος του εμβρύου (CRL)

Σχήμα 2.3 Η αυχενική διαφάνεια του εμβρύου (NT)



Σχήμα 2.4 Φυσικό ρινικό οστό και μη φυσικό στο
έμβρυο



Σχήμα 2.5 Η τριγλωχίνα ροή σε μη κανονική μορφή και σε κανονική μορφή

Κεφάλαιο 4

Ανάλυση Απαιτήσεων και Προδιαγραφές

4.1 Απαιτήσεις και Προδιαγραφές

4.1 Απαιτήσεις και Προδιαγραφές

Η Ανάλυση Απαιτήσεων είναι η διαδικασία καθορισμού των προσδοκιών του χρήστη στην ανάπτυξη ενός νέου συστήματος. Μέρος της διαδικασίας αυτής είναι να κάνουμε τις απαιτήσεις όσο πιο ξεκάθαρες και ακριβής μπορούμε. Η διαδικασία αυτή γίνεται με διάφορες τεχνικές όπως συνεντεύξεις, ερωτηματολόγια και λίστες απαιτήσεων. Επίσης μπορεί να γίνει και με πιο σύγχρονους τρόπους όπως είναι η χρήση πρωτοτύπων. Μέσω των τεχνικών αυτών συλλέγονται αρκετές χρήσιμες πληροφορίες. Μετά την ανάλυση των πληροφοριών αυτών, εξακριβώνονται οι ακριβές απαιτήσεις των χρηστών ώστε το σύστημα να ανταποκρίνεται στις ανάγκες τους. Συνήθως οι αναλυτές χρησιμοποιούν συνδυασμό των πιο πάνω τεχνικών.

Υπάρχουν διάφορα είδη απαιτήσεων. Ένα είδος απαιτήσεων είναι οι λειτουργικές δηλαδή το τι θα πρέπει το σύστημα να μπορεί να κάνει και ποιες λειτουργίες θα πρέπει να προστεθούν. Συνήθως η διαδικασία θέσπισης απαιτήσεων επικεντρώνεται σε αυτό το είδος. Άλλο είδος είναι οι μη λειτουργικές απαιτήσεις οι οποίες περιγράφουν τις ιδιότητες του συστήματος που συνήθως εκφράζονται βάσει χαρακτηριστικών της μορφής απόδοσης, χρηστικότητας, ασφάλειας, νομιμότητας και ιδιωτικότητας. Συγκεκριμένα περιγράφουν το πώς και το πόσο καλά το σύστημα θα υποστηρίξει τις λειτουργικές απαιτήσεις. Μπορούμε να τις θεωρήσουμε ως περιορισμούς που περιορίζουν τους τρόπους με τους οποίους θα μπορούσαμε να εφαρμόσουμε τις λειτουργικές απαιτήσεις.

Επομένως το σύστημα πρέπει να είναι εύχρηστο, γρήγορο, αποδοτικό, αξιόπιστο, έγκυρο και αποτελεσματικό. Η ευχρηστία είναι η δυνατότητα ενός συστήματος, που χρησιμοποιείται από χρήστες με στόχους, υπό συνθήκες χρήσης, να παρέχει αποτελεσματικότητα, αποδοτικότητα και υποκειμενική ικανοποίηση στους χρήστες του. Επίσης η ευχρηστία είναι πολύ βασική παράμετρος της ποιότητας ενός συστήματος. Ο χρόνος διεκπεραίωσης να είναι μέσα σε λογικά πλαίσια όπως και ο αριθμός βημάτων για την ολοκλήρωση μιας λειτουργίας. Ο χρήστης που θα το χρησιμοποιήσει θα αναμένει σε πλαίσια κανονικής διάρκειας να λάβει το αποτέλεσμα του. Εντούτοις το σύστημα πρέπει να είναι αποτελεσματικό δηλαδή να κάνει αυτό που ο χρήστης επιθυμεί, στην περίπτωση μας την πρόβλεψη των χρωμοσωμικών ανωμαλιών.

Μέσω του συστήματος μου, ο χρήστης μπορεί εύκολα να βάλει τις μετρήσεις από την εγκυμονούσα και να τρέξει σύστημα. Σε πολύ λίγα δευτερόλεπτα το σύστημα θα δώσει στον χρήστη το αποτέλεσμα δηλαδή την πρόβλεψη για την εμφάνιση χρωμοσωμικής ανωμαλίας. Άρα λοιπόν το σύστημα μας είναι αποδοτικό και γρήγορο αφού μπορεί να δώσει την πρόβλεψη εντός λίγων δευτερολέπτων. Είναι εύχρηστο αφού το μόνο που θα πρέπει να κάνει ο χρήστης είναι να τοποθετήσει τις μετρήσεις της γυναίκας μέσα σε μια φόρμα. Τέλος είναι αποτελεσματικό γιατί η πρόβλεψη που δίνει είναι σε αρκετά μεγάλα ποσοστά επιτυχίας.

Σε όλους τους χώρους της επιστημονικής έρευνας όπου λαμβάνεται μία μέτρηση, τίθεται το ζήτημα της αξιοπιστίας της μέτρησης αυτής. Η αξιοπιστία αφορά το κατά πόσο μια παράμετρος αναδεικνύει το πραγματικό μέγεθος του υπό μέτρηση χαρακτηριστικού. Για παράδειγμα, κατά πόσο ένας ιατρός μετρά το πραγματικό κεφαλουραίο μήκος του εμβρύου (CRL). Στην περίπτωση αυτή επειδή το μήκος είναι σε πολύ μικρές διαστάσεις ο κάθε ιατρός μπορεί να μετράει διαφορετικά από κάποιο άλλο ιατρό. Στο μάτι του κάθε ιατρού φαίνεται διαφορετικά και είναι λογικό να μην γίνεται από όλους η ίδια μέτρηση. Αν τα παραπάνω ισχύουν τότε μπορούμε να ισχυρισθούμε ότι μετράμε κατά μεγάλο ποσοστό την πραγματική τιμή του CRL, ενώ το ποσοστό σφαλμάτων μέτρησης είναι μικρό. Δηλαδή μπορούμε να ισχυριστούμε ότι η μέτρηση μας είναι αξιόπιστη.

Αξιόλογο σημείο είναι ο καθορισμός των ομάδων χρηστών. Ομάδες χρηστών στην περίπτωση του συστήματος μου είναι οι ιατροί, οι ερευνητές και ο διαχειριστής του συστήματος.

Κεφάλαιο 5

Υλοποίηση και Σχεδιασμός Συστήματος

5.1 Ανάλυση Δεδομένων

Αρχικά δημιουργήσαμε έναν πίνακα που περιγράφει τον πληθυσμό της κάθε κατηγορίας (Πίνακας 1) που δείχνει πόσες περιπτώσεις έχουμε στην βάση δεδομένων. Στην συνέχεια δημιουργήσαμε τον πίνακα μέσων όρων (Πίνακας 2) που μας δείχνει τις τιμές των μέσων όρων για κάθε παράμετρο.

Πίνακας 1 Πληθυσμός για κάθε κατηγορία από την Βάση Δεδομένων

t_population

t_population <2x6 cell>

	1	2	3	4	5	6
1	'Normal'	'T21'	'T18'	'T13'	'Triploidi'	'Turner'
2	379	380	154	49	34	73

Πίνακας 2 Μέσοι όροι για κάθε παράμετρο ξεχωριστά ανά κατηγορία

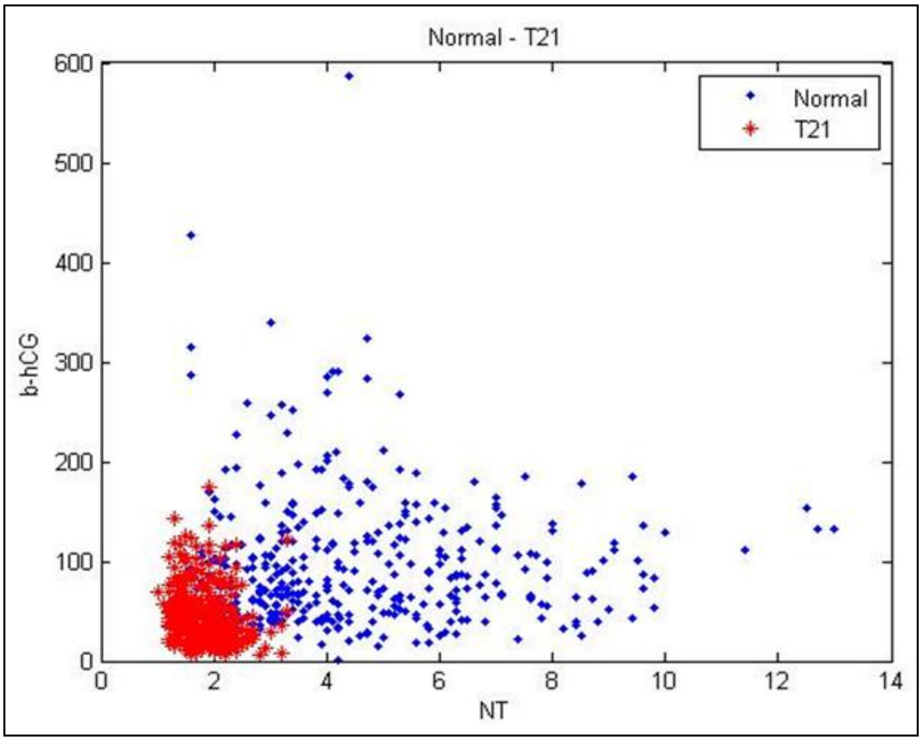
t_averages

t_averages <9x7 cell>

	1	2	3	4	5	6	7
1	[]	'Normal'	'T21'	'T18'	'T13'	'Triploidi'	'Turner'
2	'CRL'	64.0190	67.8421	59.0708	60.7122	55.8676	62.3973
3	'NT'	1.8497	4.4846	5.2155	5.2637	2.5579	9.0649
4	'PAPP-A'	3.6232	2.2629	0.5648	0.9506	1.0978	1.2802
5	'b-hCG'	44.9370	95.8217	11.5219	26.3171	139.3869	49.6464
6	'Heart rate'	159.8522	160.3105	159.5779	179.3673	154.2353	169.7808
7	'Ductus ven...	1.0676	1.5528	1.8634	1.6142	1.6826	1.8112
8	'mother''s a...	31.1092	37.0707	36.0469	35.2441	31.7577	28.4545
9	'weight'	67.8121	69.7270	70.7526	68.4286	66.8000	67.0640

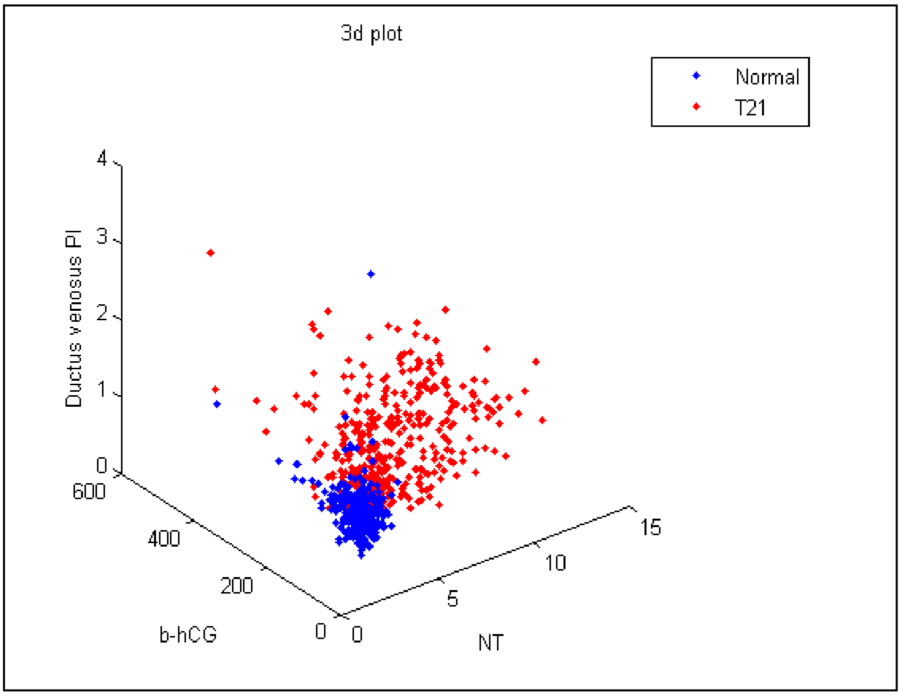
Ανάλυσα στην συνέχεια, τα δεδομένα έτσι ώστε να παρατηρήσω ποιοι παράμετροι είναι ισχυροί και έχουν ένδειξη στην ύπαρξη κάποιου συνδρόμου. Δηλαδή παρατηρήσαμε τις περιπτώσεις φυσιολογικών και Τ21 για κάθε παράμετρο ξεχωριστά και σε κάποιες περιπτώσεις υπήρχαν σημαντικές παρατηρήσεις.

Δημιούργησα δυσδιάστατες γραφικές παραστάσεις για όλους τους συνδυασμούς μεταξύ των οκτώ παραμέτρων. Συγκρίναμε τα Τ21 με μπλε χρώμα και τις φυσιολογικές περιπτώσεις με κόκκινο χρώμα (Σχήμα 1.2) ως προς τις παραμέτρους NT (άξονα X) και b-hCG (άξονα Ψ). Εδώ εξάγαμε δύο συμπεράσματα. Πρώτο συμπέρασμα είναι πως σε τιμές του NT > 2.5 (περίπου) οι περιπτώσεις είναι σχεδόν όλες φυσιολογικές και 8 περιπτώσεις Τ21 δηλαδή αρνητικού λάθους (False Negative). Το δεύτερο συμπέρασμα είναι πως σε τιμές του b-hCG >180 οι περιπτώσεις είναι σχεδόν όλες φυσιολογικές και 3 περιπτώσεις Τ21 δηλαδή αρνητικού λάθους (False Negative). Μέσω αυτή της γραφικής παράστασης βγάλαμε το συμπέρασμα πως οι δύο αυτές παράμετροι φαίνεται να παίζουν σημαντικό ρόλο για την ύπαρξη της τρισωμίας 21.



Σχήμα 1.2 Γραφική Παράσταση παραμέτρων b-hCG με το NT

Στην συνέχεια δημιουργήσαμε τρισδιάστατες γραφικές παραστάσεις μεταξύ όλων των συνδυασμών των οκτώ παραμέτρων. Συνολικά δημιουργήσαμε 56 γραφικές παραστάσεις. Αρχικός στόχος της δημιουργίας των συγκεκριμένων παραστάσεων είναι κατά πόσο ο συνδυασμός μεταξύ τριών παραμέτρων μπορεί να μας δώσει κάποιο συμπέρασμα. Σε πρώτη παρατήρηση (Σχήμα 1.4) βλέπουμε πως ο συνδυασμός των παραμέτρων Ductus Venosus, b-hCG και NT μπορεί να μας δώσει ένα διαχωρισμό μεταξύ των T21 και φυσιολογικών περιπτώσεων.



Σχήμα 1.4 Τρισδιάστατη γραφική για τις παραμέτρους b-hCG, NT και Ductus Venosus

5.2 Υλοποίηση και Σχεδιασμός Συστήματος

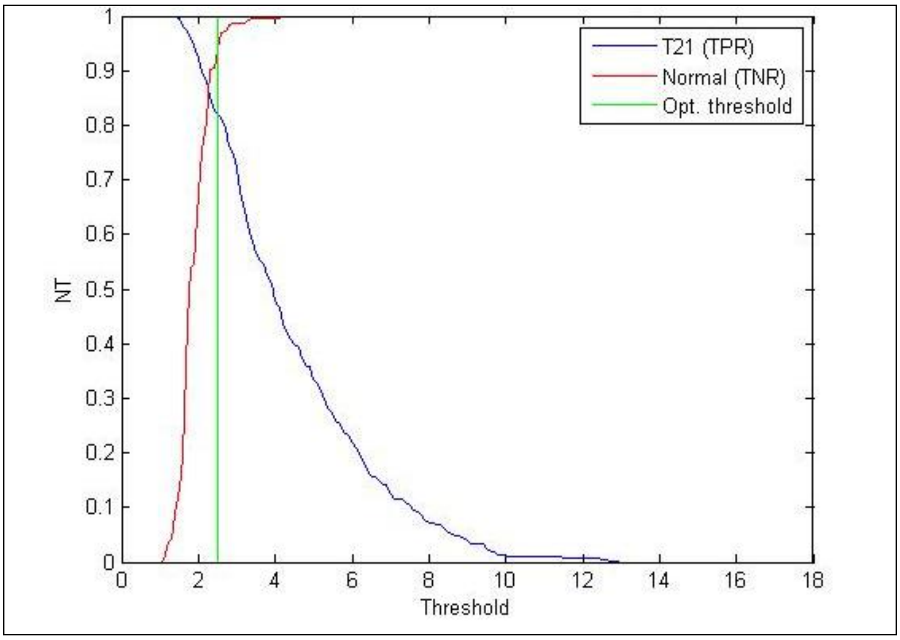
5.2.1 Γραμμική κατηγοριοποίηση

Η Γραμμική κατηγοριοποίηση είναι ο διαχωρισμός των περιπτώσεων στις δύο κατηγορίες T21 και φυσιολογικές με βάση ένα κατώφλι. Το κατώφλι είναι η τιμή, η οποία χρησιμοποιείται για σημείο αναφοράς για συγκρίσεις μεταξύ χαρακτηριστικών ή παραμέτρων. Κατώφλι στην περίπτωση μας ονομάζουμε την κατώτατη τιμή που αναφέρεται σε κάποια παράμετρο για παράδειγμα οι φυσιολογικές περιπτώσεις είναι 100%. Η τιμή αυτή του κατωφλίου που πήραμε ήταν η τιμή όπου οι περιπτώσεις των φυσιολογικών εμβρύων ήταν κοντά στο 95%. Για την τιμή αυτή δημιουργήθηκαν γραφικές παραστάσεις για όλες τις παραμέτρους.

Για τις τιμές που είναι μεγαλύτερες από το κατώφλι το σύστημα θεωρεί πως είναι μη-φυσιολογικές και αν το πραγματικό αποτέλεσμα είναι μη-φυσιολογικές τότε ονομάζονται θετικά επιτυχίας (True Positive). Σε περίπτωση που το πραγματικό αποτέλεσμα είναι φυσιολογικό τότε ονομάζονται αρνητικά λάθους (False Negative). Για τις τιμές που είναι μικρότερες από το κατώφλι το σύστημα θεωρεί πως είναι φυσιολογικές και αν το πραγματικό αποτέλεσμα είναι φυσιολογικές τότε ονομάζονται αρνητικά επιτυχίας (True Negative). Σε περίπτωση που το πραγματικό αποτέλεσμα είναι μη-φυσιολογικές τότε ονομάζονται θετικά λάθους (False Positive).

Αναλυτικότερα, επεξεργάστηκα τις παραμέτρους που έχουν μη-δυαδικές τιμές. Ο λόγος είναι ότι μέσω των μη δυαδικών αριθμών οι στατιστικοί υπολογισμοί γίνονται πιο εύκολα αντίθετα με τις παραμέτρους που έχουν μη-δυαδικές τιμές δηλαδή 1 αν είναι αληθές και 0 αν είναι ψευδές. Οι παράμετροι αυτοί είναι το CRL, το NT, το b-hCG, το PAPP-A, το Ductus Venosus, το βάρος της γυναίκας, η ηλικία της γυναίκας και ο καρδιακός ρυθμός. Χρησιμοποίησα όλες τις 380 περιπτώσεις με T21 και μετά χρησιμοποίησα τυχαία τον ίδιο αριθμό περιπτώσεων και για τα φυσιολογικά έμβρυα.

Από την γραφική παράσταση (Σχήμα 1.3) έχουμε στον άξονα Ψ τις τιμές από το μηδέν μέχρι το ένα που αναπαριστούν το ποσοστό στο οποίο κυμαίνονται οι δυο κατηγορίες, φυσιολογικές και T21, αναλόγως της τιμής του NT. Δηλαδή μπορούμε να δούμε για μια τιμή του NT πόσες περιπτώσεις σε ποσοστό είναι T21 και πόσες φυσιολογικές. Στον άξονα X έχουμε τις τιμές που μπορεί να πάρει το κατώφλι. Με κόκκινη γραμμή βλέπουμε την κατανομή των φυσιολογικών περιπτώσεων με βάση την τιμή που έχουν για την παράμετρο NT. Με μπλε γραμμή βλέπουμε αντίστοιχα για τις περιπτώσεις T21. Η πράσινη γραμμή δηλώνει το σημείο στο οποίο όρισα το κατώφλι για το NT. Παρατηρώντας την γραφική παράσταση για τιμή 2.4 του NT οι περιπτώσεις των φυσιολογικών εμβρύων κυμαίνονται γύρω στο 95% και για T21 γύρω στο 82%.



Σχήμα 1.3 Γραφική Παράσταση που καθορίζει το κατώφλι για το NT

Ακολούθως, χωρίσαμε τις περιπτώσεις για κάθε παράμετρο σε T21 και Normal αναλόγως του κατωφλίου. Έτσι υπολογίστηκαν τα True Positive, False Negative, True Negative και False Positive (Πίνακας 3). Αυτό έγινε και για τις 3 παραμέτρους.

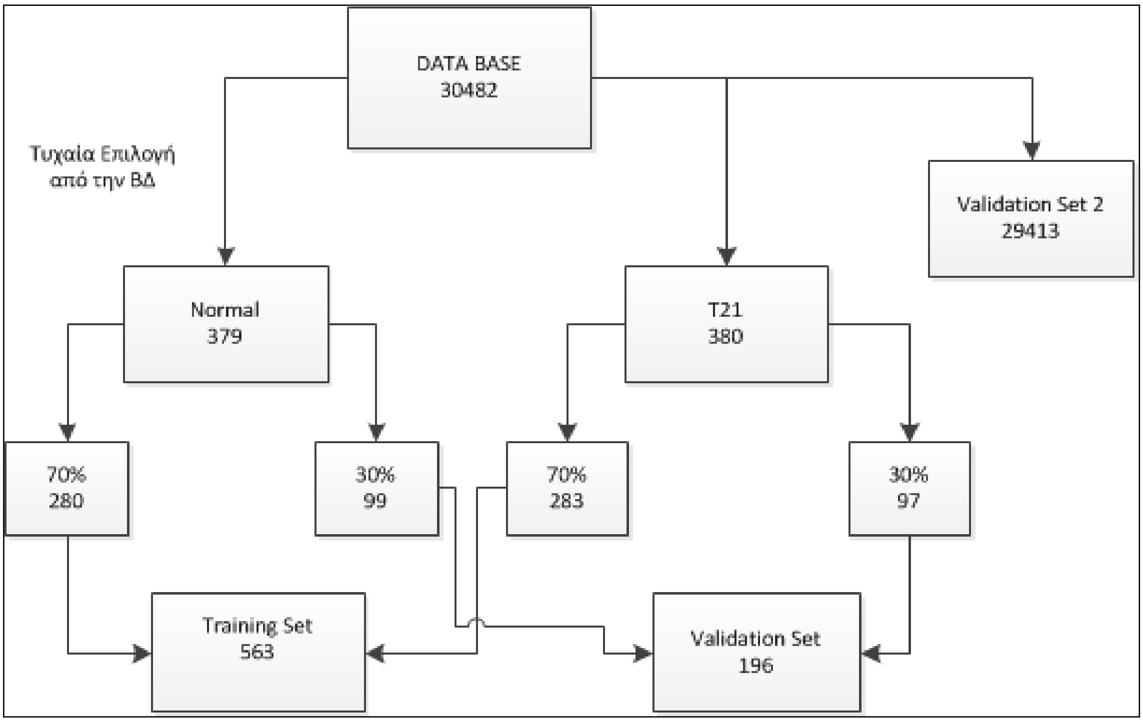
Πίνακας 3 Τα True Positive, False Negative, True Negative και False Positive για κάθε παράμετρο ξεχωριστά

t_main <27x6 cell>

	1	2	3
5	'T21-NT<3.3'	'TP'	'FN'
6	'Population'	'234 (62%)'	'135 (36%)'
7	'Id'	<234x1 dou...	<135x1 dou...
8	[]	[]	[]
9	'Normal-N...	'TN'	'FP'
10	'Population'	'374 (99%)'	'4 (1%)'
11	'Id'	<374x1 dou...	[48979;4506...
12	[]	[]	[]
13	'T21-bhCG...	'TP'	'FN'
14	'Population'	'39 (10%)'	'341 (90%)'
15	'Id'	<39x1 doub...	<341x1 dou...
16	[]	[]	[]
17	'Normal-bh...	'TN'	'FP'
18	'Population'	'372 (98%)'	'7 (2%)'
19	'Id'	<372x1 dou...	[71966;7217...
20	[]	[]	[]
21	'T21-Ductu...	'TP'	'FN'
22	'Population'	'14 (4%)'	'365 (96%)'
23	'Id'	<14x1 doub...	<365x1 dou...

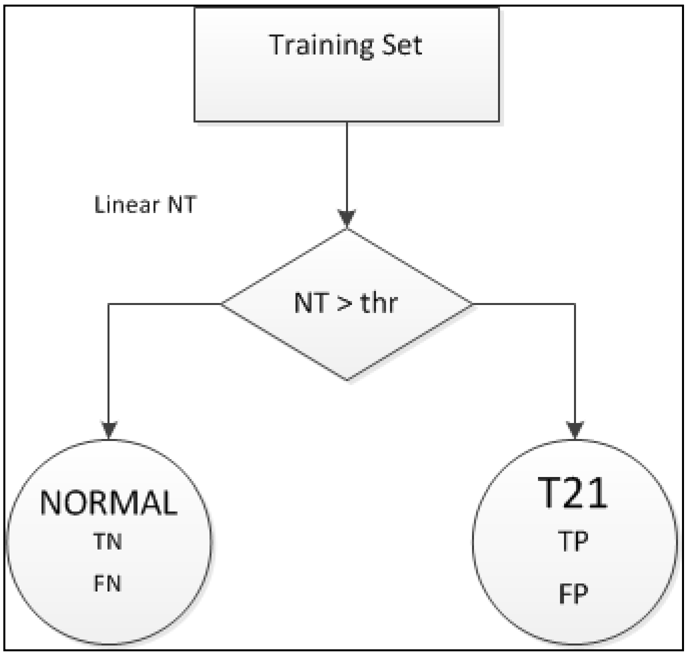
5.2.1 Μοντελοποίηση με νευρωνικά δίκτυα

Η υλοποίηση του συστήματος έγινε με δημιουργία διαφόρων μοντέλων νευρωνικών δικτύων. Οι φυσιολογικές περιπτώσεις και τα Τ21 χωρίστηκαν σε 70% και 30% (Σχήμα 5.2). Τα 70% χρησιμοποιήθηκαν για το σύνολο εκπαίδευσης και για σύνολο επαλήθευσης χρησιμοποιήθηκαν τα 30% όπως και τα υπόλοιπα φυσιολογικά έμβρυα. Από την βάση δεδομένων χρησιμοποίησα όλες τις περιπτώσεις Τ21 που είναι 380 στο σύνολο και τυχαία επέλεξα 379 περιπτώσεις φυσιολογικών εμβρύων και τις χρησιμοποίησα για σύνολο εκπαίδευσης του νευρωνικού δικτύου 1 (ΝΔ1). Πήρα επίσης και τα υπόλοιπα 29414 φυσιολογικά έμβρυα (Σχήμα 5.2) που χρησιμοποιήθηκαν στο σύνολο επαλήθευσης.



Σχήμα 5.2 Καθορισμός των Training Set και Validation Set

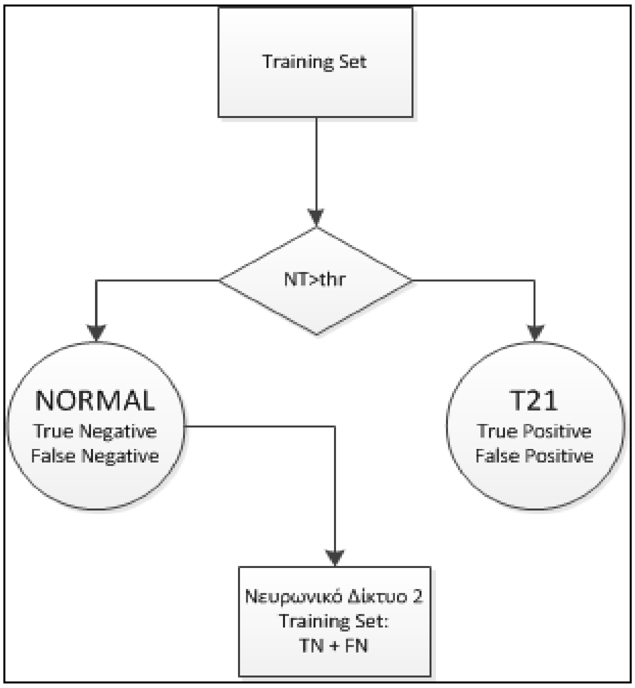
Ο διαχωρισμός των περιπτώσεων φυσιολογικών εμβρύων μαζί με τα Τ21 ήταν αρκετά καλός αλλά όχι ικανοποιητικός. Σκοπός μου ήταν να δημιουργήσω ένα σύστημα που να διαχωρίζει όσο πιο καλά γίνεται τις δύο κατηγορίες T21 και φυσιολογικών εμβρύων. Έτσι, αποφάσισα να κάνω πρώτα ένα γραμμικό διαχωρισμό των περιπτώσεων χρησιμοποιώντας ένα κατώφλι που διαχωρίζει τις δυο κατηγορίες ανάλογα με τις παραμέτρους (Σχήμα 5.9). Πήρα τρεις τιμές από κάθε παράμετρο Αυχενική Διαφάνεια (NT), Χοριακή γονιαδοτροπίνη (bhCG) και Ροή φλεβώδους πόρου (DV). Βάση αυτού, δημιούργησα τρία συστήματα. Για το πρώτο σύστημα πήρα τιμή για κατώφλι του NT η οποία μετά το γραμμικό διαχωρισμό επιστρέφει 100% T21. Για το δεύτερο σύστημα η τιμή του NT για κατώφλι επιστρέφει 100% φυσιολογικά και για το τρίτο σύστημα, η τιμή του NT για κατώφλι είναι στο ενδιάμεσο μεταξύ της τιμής του πρώτου συστήματος και του δευτέρου συστήματος.



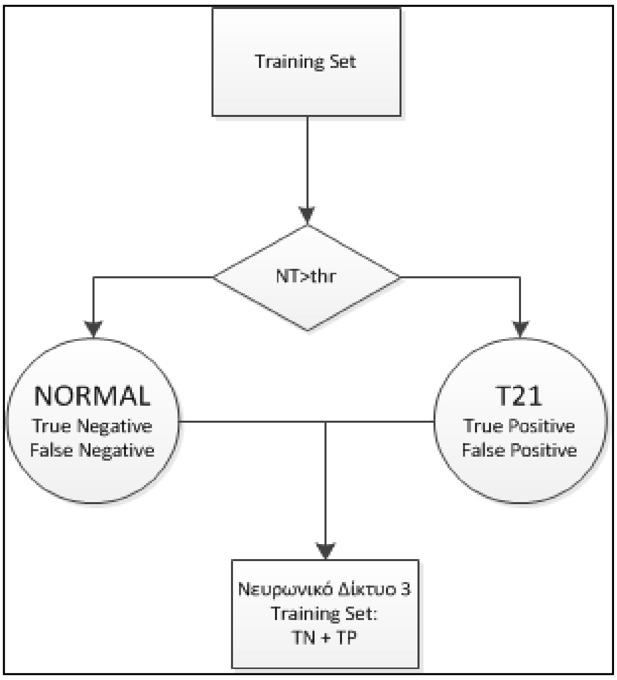
Σχήμα 5.9 Γραμμικός διαχωρισμός των περιπτώσεων βάση του κατωφλίου για το NT

Μετέπειτα δημιούργησα έντεκα νευρωνικά δίκτυα για όλα τα συστήματα. Τρία νευρωνικά για το πρώτο σύστημα, τρία νευρωνικά για το δεύτερο σύστημα, τέσσερα νευρωνικά για το τρίτο σύστημα και ένα νευρωνικό που είναι το ίδιο για όλα τα σύστημα. Το κάθε νευρωνικό δίκτυο είχε διαφορετικό σύνολο εκπαίδευσης αλλά ίδιο σύνολο επαλήθευσης. Τα νευρωνικά δίκτυα αποτελούνται από τα εξής σύνολα εκπαίδευσης και από το ίδιο σύνολο επαλήθευσης.

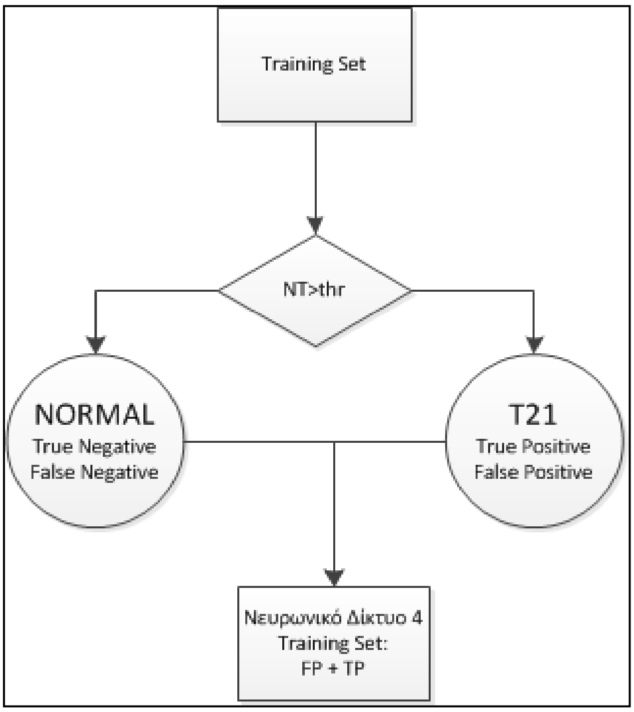
- **Νευρωνικό δίκτυο 1:** Το νευρωνικό δίκτυο 1 για σύνολο εκπαίδευσης χρησιμοποιεί όλο το 70% των περιπτώσεων και για σύνολο επαλήθευσης το υπόλοιπο 30% και τα υπόλοιπα φυσιολογικά έμβρυα, και είναι για όλα τα συστήματα το ίδιο.
- **Νευρωνικό δίκτυο 2:** Το νευρωνικό δίκτυο 2 για σύνολο εκπαίδευσης χρησιμοποιεί το True Negative και το False Negative από το γραμμικό διαχωρισμό του 70% των περιπτώσεων (Σχήμα 5.10).
- **Νευρωνικό δίκτυο 3:** Το νευρωνικό δίκτυο 3 σύνολο εκπαίδευσης χρησιμοποιεί το True Negative και το True Positive από το γραμμικό διαχωρισμό του 70% των περιπτώσεων (Σχήμα 5.11).
- **Νευρωνικό δίκτυο 4:** Το νευρωνικό δίκτυο 4 σύνολο εκπαίδευσης χρησιμοποιεί το False Positive και το True Positive από το γραμμικό διαχωρισμό του 70% των περιπτώσεων(Σχήμα 5.12).
- **Νευρωνικό δίκτυο 5:** Το νευρωνικό δίκτυο 5 για σύνολο εκπαίδευσης χρησιμοποιεί το True Negative και False Positive από το γραμμικό διαχωρισμό του 70% των περιπτώσεων (Σχήμα 5.13).



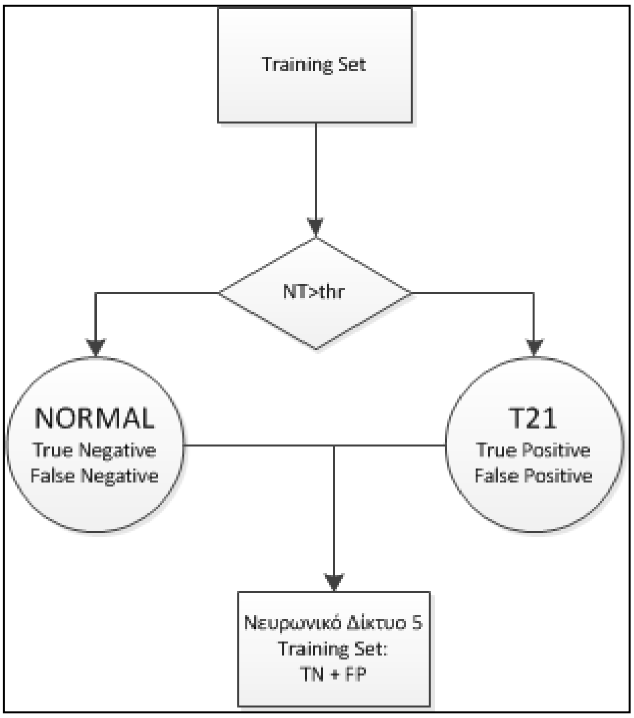
Σχήμα 5.10 Training Set για το Νευρωνικό Δίκτυο 2



Σχήμα 5.11 Training Set για το Νευρωνικό Δίκτυο 3



Σχήμα 5.12 Training Set για το Νευρωνικό Δίκτυο 4



Σχήμα 5.13 Training Set για το Νευρωνικό Δίκτυο 5

Η ίδια διαδικασία όπως πιο πάνω έγινε με την παράμετρο Ductus Venosus και την παράμετρο bhCG. Δημιουργήθηκαν ακόμα είκοσι δύο νευρωνικά δίκτυα με τον ίδιο τρόπο όπως πιο πάνω, δηλαδή έντεκα νευρωνικά αντίστοιχα. Στα νευρωνικά δίκτυα η εκπαίδευση έγινε με όλες τις παραμέτρους της βάσης δεδομένων.

Κεφάλαιο 6

Αξιολόγηση Συστήματος - Αποτελέσματα

6.1 Αξιολόγηση Συστήματος - Αποτελέσματα

Μέσα από την διπλωματική εργασία και μετά την υλοποίηση των νευρωνικών δικτύων έλαβα αποτελέσματα. Τα οποία κάποια ήταν αρκετά καλά και κάποια όχι. Στην συνέχεια παρουσιάζω αναλυτικά τα αποτελέσματα που πάρθηκαν από το γραμμικό διαχωρισμό και από την υλοποίηση των νευρωνικών, για τις τρείς παραμέτρους που χρησιμοποιήθηκαν NT, Ductus Venosus και bhCG, για κάθε σύστημα ξεχωριστά (Πίνακας 6.1-6.3).

Εδώ μπορούμε να επιβεβαιώσουμε αυτό που ανέφερα και στο προηγούμενο κεφάλαιο, πως για το σύστημα 1 έχουμε ποσοστά γύρω στο 100% για τα Φυσιολογικά έμβρυα, για το σύστημα 2 έχουμε ποσοστά κοντά στο 100% για τις περιπτώσεις T21 και για το σύστημα 3 έχουμε ένα συνδυασμό ποσοστών και για τις δυο περιπτώσεις. Επιπρόσθετα βλέπουμε και το ποσοστό των αποτελεσμάτων για τα υπόλοιπα 29414 φυσιολογικών εμβρύων.

Πίνακας 6.1 Ποσοστά Γραμμικού Διαχωρισμού για την παράμετρο NT

1	Systems	Normal	T21	Normal Rest
2	System 1	9.2784%	100%	10.4066%
3	System 2	98.9691%	64.3564%	99.4526%
4	System 3	83.5052%	97.0297%	76.0692%

Πίνακας 6.2 Ποσοστά Γραμμικού Διαχωρισμού για την παράμετρο DV

1	Systems	Normal	T21	Normal Rest
2	System 1	2.0619%	99.0099%	4.1137%
3	System 2	100%	17.8218%	99.7756%
4	System 3	92.7835%	61.3861%	93.0849%

Πίνακας 6.3 Ποσοστά Γραμμικού Διαχωρισμού για την παράμετρο bhCG

1	Systems	Normal	T21	Normal Rest
2	System 1	7.2165%	98.0198%	9.4105%
3	System 2	98.9691%	2.9703%	99.354%
4	System 3	34.0206%	96.0396%	36.1291%

Να σημειώσω ότι τα αποτελέσματα του νευρωνικού δικτυού 1 παραμένουν ίδια σε όλα τα συστήματα (Πίνακας 6.4). Στην συνέχεια, παρουσιάζονται τα αποτελέσματα για το σύνολο εκπαίδευσης και για το σύνολο επαλήθευσης. Στην πρώτη γραμμή βλέπουμε το σύνολο των περιπτώσεων που ήταν φυσιολογικά και T21, στην δεύτερη γραμμή βλέπουμε το σύνολο των περιπτώσεων που το νευρωνικό δίκτυο μας προέβλεψε σωστά το αποτέλεσμα και στην τρίτη γραμμή βλέπουμε και το αντίστοιχο ποσοστό των προβλέψεων.

Πίνακας 6.4 Αποτελέσματα από το Νευρωνικό Δίκτυο 1

ΝΔ1	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
Ποσοστό	98.5816%	98.5663%	94.8454%	93.063%	93.4963%

Στο Σύστημα 1 για την παράμετρο NT βλέπουμε τα αποτελέσματα για τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν (Πίνακας 6.5). Επίσης και για το σύστημα 2 (Πίνακας 6.6) και για το σύστημα 3 (Πίνακας 6.7) παρουσιάζονται τα αποτελέσματα από την υλοποίηση των νευρωνικών δικτύων.

Πίνακας 6.5 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο NT του Συστήματος 1

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	98,5765%	96,3636%	93,8144%	97,0297%	94,5366%
ΝΔ3	100%	100%	98,9691%	87,1287%	99,3099%
ΝΔ5	100%	100%	100%	51,4851%	99,8062%

Πίνακας 6.6 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο NT του Συστήματος 2

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ3	97,2973%	99,2832%	59,7938%	98,0198%	56,0005%
ΝΔ4	97,9592%	97,8495%	93,8144%	95,0495%	92,4084%
ΝΔ5	91,8919%	99,1837%	6,1856%	99,0099%	8,1084%

Πίνακας 6.7 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο NT του Συστήματος 3

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	100%	84,6154%	92,7835%	98,0198%	93,3399%
ΝΔ3	99,061%	99,6047%	92,7835%	95,0495%	91,2627%
ΝΔ4	98,5507%	98,419%	93,8144%	91,0891%	92,8469%
ΝΔ5	99,5305%	92,7536%	81,4433%	94,0594%	77,2523%

Επιπρόσθετα παρουσιάζονται τα αποτελέσματα από τα νευρωνικά δίκτυα των συστημάτων 1, 2 και 3 για την παράμετρο Ductus Venosus (Πίνακες 6.8-6.10) και της παραμέτρου bhCG (Πίνακες 6.11-6.13).

Πίνακας 6.8 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο DV του Συστήματος 1

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	98,9362%	96,5385%	97,9381%	94,0594%	94,8732%
ΝΔ3	100%	100%	100%	40,5941%	99,8844%
ΝΔ5	100%	0%	100%	0%	100%

Πίνακας 6.9 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο DV του Συστήματος 2

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ3	100%	100%	23,7113%	99,0099%	25,4981%
ΝΔ4	100%	96,7626%	93,8144%	87,1287%	95,152%
ΝΔ5	100%	99,2754%	1,0309%	47,5248%	2,8898%

Πίνακας 6.10 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο DV του Συστήματος 3

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	99,2188%	89,4737%	96,0972%	84,1584%	94,8868%
ΝΔ3	98,8281%	98,1818%	91,7526%	80,198%	96,4949%
ΝΔ4	57,6923%	99,3939%	59,7938%	99,0099%	62,9564%
ΝΔ5	99,2188%	80,7692%	92,7835%	38,6139%	94,5536%

Πίνακας 6.11 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο bhCG του Συστήματος 1

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ3	90,4762%	98,9209%	38,1443%	97,0297%	41,8202%
ΝΔ4	98,0843%	98,5612%	92,7835%	97,0297%	94,6862%
ΝΔ5	80,9524%	99,6169%	6,1856%	74,2574%	6,3405%

Πίνακας 6.12 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο bhCG του Συστήματος 2

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	99,2908%	99,2424%	92,7835%	97,0297%	93,7751%
ΝΔ3	100%	93,3333%	98,9691%	37,6238%	99,83%
ΝΔ5	100%	0%	100%	0%	100%

Πίνακας 6.13 Αποτελέσματα από τα Νευρωνικό Δίκτυα για την παράμετρο bhCG του Συστήματος 3

Ποσοστά	Φυσιολογικά Σύνολο Επανάληψης	T21 Σύνολο Επανάληψης	Φυσιολογικά Σύνολο Επαλήθευσης	T21 Σύνολο Επαλήθευσης	Υπόλοιπα Φυσιολογικά Σύνολο Επαλήθευσης
ΝΔ2	100%	72,2222%	86,5979%	79,2079%	87,8289%
ΝΔ3	100%	99,2337%	73,1959%	100%	75,7462%
ΝΔ4	97,7528%	98,8506%	93,8144%	97,0297%	93,1631%
ΝΔ5	89,4231%	83,1461%	44,3299%	25,7426%	46,9266%

Στην συνέχεια είχαν καταχωρηθεί όλα τα αποτελέσματα σε ένα νέο πίνακα (Πίνακας 6.14) μέσα στον οποίο μπορούμε να δούμε συγκεντρωτικά τα ποσοστά για κάθε νευρωνικό δίκτυο και για τον γραμμικό διαχωρισμό. Με τον πίνακα αυτό μπορούσα εύκολα να κάνω συγκρίσεις μεταξύ των νευρωνικών δικτύων.

Πίνακας 6.14 Αποτελέσματα νευρωνικών δικτύων και γραμμικού διαχωρισμού σε ποσοστά

'ALL ND'	'Normal'	'T21'	'Normal Rest'
'ND1'	'94.8454%'	'93.0693%'	'93.4963%'
'ND3-Sys1_NT'	'59.7938%'	'98.0198%'	'56.0005%'
'ND4-Sys1_NT'	'93.8144%'	'95.0495%'	'92.4084%'
'ND2-Sys2_NT'	'93.8144%'	'97.0297%'	'94.5366%'
'ND3-Sys2_NT'	'98.9691%'	'87.1287%'	'99.3099%'
'ND2-Sys3_NT'	'92.7835%'	'98.0198%'	'93.3399%'
'ND3-Sys3_NT'	'92.7835%'	'95.0495%'	'91.2627%'
'ND4-Sys3_NT'	'93.8144%'	'91.0891%'	'92.8469%'
'ND3-Sys1_DV'	'23.7113%'	'99.0099%'	'25.4981%'
'ND4-Sys1_DV'	'93.8144%'	'87.1287%'	'95.152%'
'ND2-Sys2_DV'	'97.9381%'	'94.0594%'	'94.8732%'
'ND3-Sys2_DV'	'100%'	'40.5941%'	'99.8844%'
'ND2-Sys3_DV'	'96.9072%'	'84.1584%'	'94.8868%'
'ND3-Sys3_DV'	'91.7526%'	'80.198%'	'96.4949%'
'ND4-Sys3_DV'	'59.7938%'	'99.0099%'	'62.9564%'
'ND3-Sys1_bhCG'	'38.1443%'	'97.0297%'	'41.8202%'
'ND4-Sys1_bhCG'	'92.7835%'	'97.0297%'	'94.6862%'
'ND2-Sys2_bhCG'	'92.7835%'	'97.0297%'	'93.7751%'
'ND3-Sys2_bhCG'	'98.9691%'	'37.6238%'	'99.83%'
'ND2-Sys3_bhCG'	'86.5979%'	'79.2079%'	'87.8289%'
'ND3-Sys3_bhCG'	'73.1959%'	'100%'	'75.7462%'
'ND4-Sys3_bhCG'	'93.8144%'	'97.0297%'	'93.1631%'
'Lnr System 1 NT'	'9.2784%'	'100%'	'10.4066%'
'Lnr System 2 NT'	'98.9691%'	'64.3564%'	'99.4526%'
'Lnr System 3 NT'	'83.5052%'	'97.0297%'	'76.0692%'
'Lnr System 1 DV'	'2.0619%'	'99.0099%'	'4.1137%'
'Lnr System 2 DV'	'100%'	'17.8218%'	'99.7756%'
'Lnr System 3 DV'	'92.7835%'	'61.3861%'	'93.0849%'
'Lnr System 1 bhCG'	'7.2165%'	'98.0198%'	'9.4105%'
'Lnr System 2 bhCG'	'98.9691%'	'2.9703%'	'99.354%'
'Lnr System 3 bhCG'	'34.0206%'	'96.0396%'	'36.1291%'

Ακολούθως μετά συγκεντρώθηκαν τα αποτελέσματα τα οποία βγήκαν από την υλοποίηση των νευρωνικών δικτύων για τις περιπτώσεις του συνόλου επαλήθευσης. Μετά από επεξεργασία των αποτελεσμάτων δημιουργήθηκαν δύο πίνακες. Ο πρώτος πίνακας έχει συγκεντρωτικά τις εξόδους από τα νευρωνικά δίκτυα (Πίνακας 6.15) και ο δεύτερος πίνακας είναι τα ίδια εξόδοι τυγχάνοντας κάποιας επεξεργασίας (Πίνακας 6.16).

Όρισα ένα νέο κατώφλι στο οποίο ελέγχονται όλες οι περιπτώσεις του συνόλου επαλήθευσης, αυτές που είναι μεγαλύτερες από το κατώφλι το αποτέλεσμα τους γίνεται 1 αντίθετα το αποτέλεσμα γίνεται ίσο με 0 (το 1 συμβολίζει τα T21 και το 0 τα φυσιολογικά).

Μετέπειτα, έδωσα βάρος στο νευρωνικό δίκτυο 1 το οποίο όπως ανέφερα και πιο πάνω το σύνολο εκπαίδευσης του είναι όλο το 70% των περιπτώσεων. Λόγω πιο πολλών δεδομένων στο σύνολο εκπαίδευσης, είχαμε καλύτερα αποτελέσματα από τα άλλα νευρωνικά. Κατά την επεξεργασία των αποτελεσμάτων, αποφάσισα να δημιουργήσω και μια τρίτη κατηγορία κατηγοριοποίησης, δηλαδή εκτός από τα φυσιολογικά έμβρυα και τα μη φυσιολογικά, μια νέα κατηγορία τις αμφίβολες περιπτώσεις (NotSure).

Τα αποτελέσματα από το νευρωνικό δίκτυο 1 πέρασαν από ένα νέο έλεγχο (Σχήμα 6.3). Η στήλη με κίτρινο χρώμα είναι οι έξοδοι από το νευρωνικό δίκτυο 1. Όσα από αυτά ήταν μεγαλύτερα από το 0 έλεγα πως ήταν μη φυσιολογικά, αν ήταν μεγαλύτερα από -0.8 και μικρότερα από το 0 έλεγα πως ανήκαν στην κατηγορία αμφίβολες περιπτώσεις και σε αντίθετη περίπτωση δηλαδή μικρότερα από το -0.8 έλεγα πως ανήκαν στην κατηγορία φυσιολογικά (Πίνακας 6.17). Για κάθε μια από τις κατηγορίες είχε γίνει και έλεγχος αν όντως το πραγματικό αποτέλεσμα είναι αυτό ή όχι που το σύστημα προέβλεψε . Αποτέλεσμα αυτού ήταν να εμφανιστούν οι περιπτώσεις False Positive αλλά και οι περιπτώσεις False Negative, όλα τα αποτελέσματα καταχωρήθηκαν σε ένα καινούργιο πίνακα (Πίνακας 6.18 – 6.19).

Πίνακας 6.15 Οι έξοδοι από τα νευρωνικά δίκτυα του συνόλου επαλήθευσης

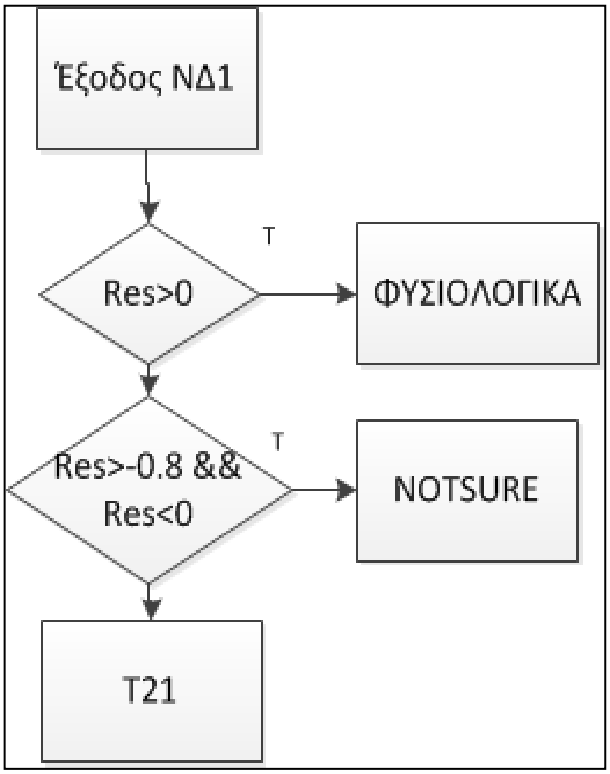
'Index'	' ID'	' Class'	'Linear NT Sys1'	'Linear NT Sys2'	'Linear NT Sys3'	'[1]ND-1 Sys3'	'[2]ND-3 Sys1'	'[3]ND-4 Sys1'	'[4]ND-5 Sys1'	'[5]ND-2 Sys2'	'[6]ND-3 Sys2'	'[7]ND-5 Sys2'	'[8]ND-2 Sys3'	'[9]ND-3 Sys3'	'[10]ND-4 Sys3'
27676	69455	'Normal'	1	0	1	0.93024	1.02542	0.26454	0.97292	0.18911	-0.9239	-1	0.57525	0.99946	-0.54522
27902	69686	'Normal'	1	0	0	-1	0.92505	-0.9319	1.07	-0.8853	-1	-1	-0.8761	-1	-0.90366
468	72684	'T21'	1	0	1	-0.4394	1.00803	-0.0754	2.23029	0.46476	1	-1	0.61611	-0.9142	0.593464

Πίνακας 6.16 Οι έξοδοι από τα νευρωνικά δίκτυα του συνόλου επαλήθευσης μετά από επεξεργασία

'Case Index'	'Case ID'	'Case Class'	Linear NT Sys1'	'Linear NT Sys2'	'Linear NT Sys3'	'[1]ND-1 Sys3'	'[2]ND-3 Sys1'	'[3]ND-4 Sys1'	'[4]ND-5 Sys1'	'[5]ND-2 Sys2'	'[6]ND-3 Sys2'	'[7]ND-5 Sys2'	'[8]ND-2 Sys3'	'[9]ND-3 Sys3'	'[10]ND-4 Sys3'
27676	69455	'Normal'	1	0	1	1	1	1	1	1	0	0	1	1	0
27902	69686	'Normal'	1	0	0	0	1	0	1	0	0	0	0	0	0
468	72684	'T21'	1	0	1	0	1	0	1	1	1	0	1	0	1

Πίνακας 6.17 Η κατηγορία αμφίβολες περιπτώσεις

Class	NotSure
1	0
0	0
0	1



Σχήμα 6.3 Ο έλεγχος της εξόδου του νευρωνικού δικτύου 1

Βλέπουμε πως το σύστημα μας προέβλεψε σωστά 91 από τα 97 φυσιολογικά έμβρυα και 94 από τα 101 μη φυσιολογικά. Επίσης, 5 από τα 97 φυσιολογικά προέβλεψε λάθος και 4 από τα 101 από τα μη φυσιολογικά. Να σημειώσω πως στην νέα κατηγορία αμφίβολες περιπτώσεις βλέπουμε πως έχουμε μία περίπτωση από τα φυσιολογικά και τρεις περιπτώσεις από τα μη φυσιολογικά.

Πίνακας 6.18 Αποτελέσματα μετά από επεξεργασία του νευρωνικού δικτύου 1

	Περιπτώσεις Συνόλου Επαλήθευσης
Πρόβλεψη Φυσιολογικών	91
Σύνολο Φυσιολογικών	97
False Positive	5
Πρόβλεψη Τ21	94
Σύνολο Τ21	101
False Negative	4
NotSure Φυσιολογικών	1
NotSure T21	3

Πίνακας 6.19 Αποτελέσματα μετά από επεξεργασία του νευρωνικού δικτύου 1 του συνόλου επαλήθευσης υπόλοιπα φυσιολογικά

	Περιπτώσεις Συνόλου Επαλήθευσης
Πρόβλεψη Φυσιολογικών	26839
Σύνολο Φυσιολογικών	29414
False Positive	1913
NotSure Φυσιολογικών	661

Κεφάλαιο 7

Συμπεράσματα και Μελλοντική Εργασία

7.1 Συμπεράσματα

Το αποτέλεσμα της διπλωματικής μου εργασίας είναι η δημιουργία του υπολογιστικού συστήματος νοημοσύνης για την πρόβλεψη της τρισωμίας 21 στην πρώτη τριμηνία κύησης. Ξεκινώντας από μια βάση δεδομένων και χωρίς άλλες πληροφορίες κατάφερα να δημιουργήσω ένα σύστημα πρόβλεψης το οποίο επιστρέφει ποσοστά επιτυχίας μεγαλύτερα από 95%. Κατά την εξέλιξη της υλοποίησης είχαμε πολλές επεξεργασίες των δεδομένων. Μέχρι το σημείο για την δημιουργία των νευρωνικών δικτύων αναλύσαμε με κάθε λεπτομέρεια όλα τα δεδομένα της βάσης δεδομένων. Είχε γίνει μια μεγάλη διαδικασία μέχρι της λήψης της τελικής απόφασης σε σχέση ποιοι παράμετροι είναι δείκτες στην εμφάνιση της τρισωμίας 21. Μετά από την απόφαση αυτή ακολούθησε η υλοποίηση των νευρωνικών δικτύων. Συγκεκριμένα προσπάθησα να δημιουργήσω όσο περισσότερα νευρωνικά δίκτυα κάνοντας χρήση όλους τους συνδυασμούς μεταξύ των περιπτώσεων. Δηλαδή το σύνολο εκπαίδευσης σε όλα τα δίκτυα ήταν διαφορετικό αντίθετα με το σύνολο επαλήθευσης που ήταν σε όλα τα δίκτυα το ίδιο. Ο λόγος ήταν η δημιουργία πολλαπλών μοντέλων για να γίνει στο τέλος η καλύτερη σε ποσοστά επιτυχίας επιλογή. Στο τέλος καταλήξαμε στο δίκτυο που γινότανε εκπαίδευση με όλα τα δεδομένα του συνόλου εκπαίδευσης. Η επεξεργασία των αποτελεσμάτων του δικτύου αυτού έγινε με σκοπό ακόμα καλύτερα ποσοστά. Συμπερασματικά μπορώ να πω ότι τα τελικά αποτελέσματα μου είναι πολύ καλά σε σχέση με προηγούμενες μελέτες στο συγκεκριμένο θέμα.

Σε προσωπικό επίπεδο μέσω της διπλωματικής μου εργασίας γνώρισα με λεπτομέρεια την διαδικασία της εγκυμοσύνης στους πρώτους τρεις μήνες. Έμαθα μέσω της μελέτης σχεδόν όλους τους παράγοντες που μπορούν να προκαλέσουν χρωμοσωμική ανωμαλία στο έμβρυο. Η ευκαιρία που μου δόθηκε να μελετήσω εκ βάθους την τρισωμία 21 ήταν κάτι πρωτόγνωρο για εμένα. Ίσως χωρίς την δημιουργία αυτής της εργασίας να μην μάθαινα ποτέ τι είναι αυτά τα παιδιά που βλέπουμε καθημερινά στην ζωή μας. Η μόνη διαφορά από τα φυσιολογικά παιδιά οφείλεται μόνο σε ένα επιπλέον χρωμόσωμα.

7.2 Μελλοντική Εργασία

Στην διπλωματική εργασία, μετά από τα ενθαρρυντικά αποτελέσματα, υπάρχει προοπτική για ακόμα καλύτερα αποτελέσματα. Με κάποιες τροποποιήσεις και αλλαγές στην επεξεργασία των αποτελεσμάτων τα αποτελέσματα μπορούν γίνουν ακόμα καλύτερα. Μια πρόταση είναι η αξιοποίηση και των υπολοίπων νευρωνικών δικτύων. Στην παρούσα διπλωματική δημιουργηθήκαν περισσότερα από τριάντα νευρωνικά δίκτυα αλλά έμφαση δόθηκε μόνο στο ένα που είχε τα ψηλότερα ποσοστά επιτυχίας. Μπορεί μέσω μιας παραλλαγής στην υλοποίηση και με χρήση κάποιων κανόνων να συνδυαστούν μεταξύ τους όλα τα νευρωνικά και να παράγουν ένα καλύτερο αποτέλεσμα.

Μια άλλη πρόταση είναι η χρησιμοποίηση μη-επιβλεπόμενης μάθησης στην επιλογή των φυσιολογικών περιπτώσεων από την βάση δεδομένων. Δηλαδή στην εργασία αυτή είχαν επιλεγεί τυχαία 379 περιπτώσεις φυσιολογικές περιπτώσεις. Μπορεί μέσω κάποιου αλγόριθμου της μη-επιβλεπόμενης μάθησης να γίνεται κατηγοριοποίηση των φυσιολογικών περιπτώσεων αναλόγως των χαρακτηριστικών του. Στην συνέχεια μπορεί να δοκιμαστεί η κάθε κατηγορία φυσιολογικών περιπτώσεων και να γίνει σύγκριση των αποτελεσμάτων.

Ακόμη μια άλλη πρόταση είναι η προ-επεξεργασία της βάσης δεδομένων. Δηλαδή κάποιες τιμές που είναι δυαδικές να γίνουν με κάποιο τρόπο επεξεργασία. Για παραδείγματος χάριν μπορεί να αφαιρεθούν όλα τα μηδέν από την βάση δεδομένων αφού όλοι γνωρίζουμε πως σε μαθηματικές πράξεις το μηδέν δεν είναι καλός αριθμός για την παραγωγή αποτελεσμάτων. Αυτά τα μηδέν μπορεί να αντικατασταθούν με μια άλλη τιμή όπως το αρνητικό ένα.

Τέλος το σύστημα αυτό μπορεί να χρησιμοποιηθεί και στον ηλεκτρονικό φάκελο. Δηλαδή μπορεί εύκολα να γίνει διασύνδεση με το ηλεκτρονικό φάκελο και να ενσωματωθεί στην κατηγορία γυναικολογικά. Στην κατηγορία αυτή ο ιατρός θα εισάγει όλες τις μετρήσεις που θα κάνει κατά τον προγεννητικό έλεγχο. Με την ενσωμάτωση αυτή θα μπορεί ο ιατρός την ίδια ώρα να πάρει την πρόβλεψη για την εγκυμοσύνη άμεσα.

Αναφορές

[1] E. Tsaliki, E. A. Papageorgiou, C. Spyrou, G. Koumbaris, E. Kypri1, S. Kyriakou, C. Sotiriou, E. Touvana, A. Keravnou, A. Karagrigoriou, K. Lamnissou, V. Velissariou and P. C. Patsalis, MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21, Prenatal Diagnosis 2012, 32, 996–1001

[2] E. A. Papageorgiou, A. Karagrigoriou, E.Tsaliki, V. Velissariou, N. P Carter & P. C Patsalis, Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21, Nature Medicine, Volume 17, Number 4, April 2011.

[3] W.S.McCulloch and W. Pitts, A logical calculus of ideas immanent in nervous activity, Bullettin of Mathematical Biophysics, 1943.

[4] D. Hebb, The organisation of behavior, New York: Wiley & Sons, 1949.

[5] J. Von Neumann, The computer and the brain, Yale University Press, New Haven 1958.

[6] J. L. McClelland and D. E. Rumelhart, Parallel Distributed Processing, Volumes 1 and 2, MIT Press, Cambridge, Mass. 1986.

[7] R. Sathya, Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification, International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013.

[8] Andreas C. Neocleous, Christos N. Schizas, Argiro Singelaki, and Kypros H. Nikolaides, First Trimester Noninvasive Prenatal Diagnosis of Aneuploidies: A Computational Intelligence Approach, unpublished.

[9] S. Cigero, G. Rembouskos, H. Vandecruys, M. Hogg and K. H. Nicolaides, Likelihood ratio for trisomy 21 in fetuses with absent nasal bone at the 11–14-week scan, Ultrasound Obstet Gynecol 2004; 23: 218–223.

[10] K. H. Nicolaides, K. Spencer, K. Avgidou, S. Faiola and O. Falcon, Multicenter study of first-trimester screening for trisomy 21 in 75 821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening, Ultrasound Obstet Gynecol 2005; 25: 221–226.

[11] Nicola Persico, Francisca Molina, Guillermo Azumendi, Luigi Fedele and Kypros H. Nicolaides, Nasal bone assessment in fetuses with trisomy 21 at 16–24 weeks of gestation by three-dimensional ultrasound, Prenatal Diagnosis 2012, 32, 240–244.

[12] Argyro Syngelaki, Teodora Chelemen, Themistoklis Dagklis, Lindsey Allan and Kypros H. Nicolaides, Challenges in the diagnosis of fetal non-chromosomal abnormalities at 11–13 weeks, Prenat Diagn 2011; 31: 90–102.

[13] A. Tabor and Z. Alfirevic, Update on Procedure-Related Risks for Prenatal Diagnosis Techniques, Fetal Diagn. and Ther., Vol. 27, No.1, Jan. 2010.

[14] J. Patel and R. Goyal, Applications of artificial neural networks in medical science, Curr. Clin. Pharmacol, Vol. 2, No. 3, pp 217-26, Sep. 2007.

[15] F. Schnorrenberg, N. Tsapatsoulis, C. Pattichis, C. Schizas, S. Kollias, M. Vassiliou, A. Adamou, K. Kyriacou, A modular neural network system for the analysis of nuclei in histopathological sections, IEEE Engineering in Medicine and Biology Magazine, Vol. 19, No. 3, pp. 48-63, Jan/Feb 2000.

[16] F. Schnorrenberg, C. Pattichis, K. Kyriacou, C. Schizas, Computeraided detection of breast cancer nuclei, IEEE Transactions on Information Technology in Biomedicine, Vol. 1, No. 2, pp. 128-40, Jun. 1997.

[17] C. Schizas and C. Pattichis, Learning systems in biosignal analysis, BioSystems, Vol. 41, No. 2, pp. 105-125, Jan. 1997.

[18] O. Er, N. Yumusak and F. Temurtas, Chest Diseases Diagnosis Using Artificial Neural Network, Exp. Sys. with App., Vol. 37, No. 12, pp.76487655, Dec. 2010.

[19] I. Khan, P. Zope and S. Suralkar, Importance of Artificial Neural Network in Medical Diagnosis Disease Like Acute Nephritis Disease and Heart Disease, Int. Jour. of Eng. Sc. and Innov. Tech., Vol. 2, No. 2, Mar. 2013.

[20] A. Maithili, K. Kumari and S. Rajamanickam, Neural Networks Towards Medical, Int. Jour. of Mod. Eng. Res., Vol. 1, No. 1, pp.57-64, Sep/Oct. 2011.

[21] C. Stephan, H. Cammann, H. Meyer, C. Mller, S. Deger, M. Lein, and K. Jung, An Artificial Neural Network for Five Different Assay Systems of Prostate-Specific Antigen in Prostate Cancer Diagnostics, BJUI Intern., Vol. 102, No. 7, pp. 799-805, Jun. 2008.

[22] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65:386–408, 1958.