Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τεχνολογία Ήχου Και Εικόνας
Κατηγοριοποίηση Καρδιακών Ήχων

9ο Εξάμηνο

Στεφανίδης Ιωάννης 9587
Μουστάκας Γεώργιος 9365
Σφυράκης Εμμανουήλ 9507
Σαρρής Αναστάσιος Λουκάς 9451

24 Φεβρουαρίου 2022

Περιεχόμενα

1 Εισαγωγή

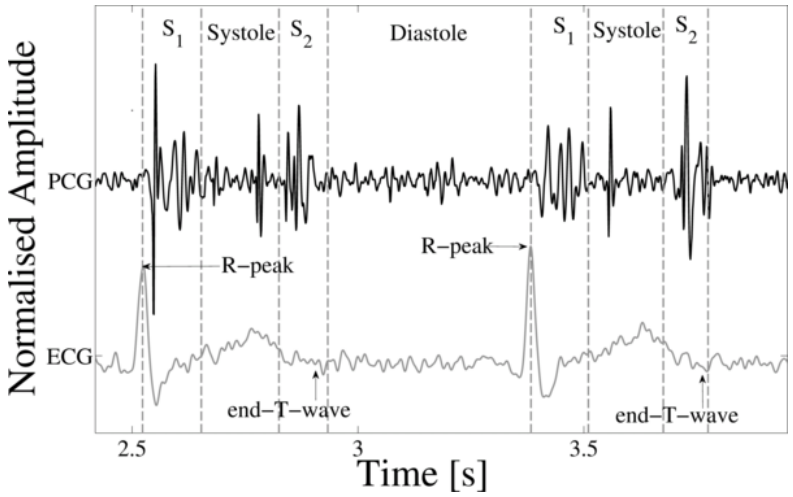
Οι καρδιοπάθειες ταλαιπωρούν, και σε πολλές περιπτώσεις οδηγούν στον θάνατο, ένα μεγάλο μέρος του πληθυσμού παγκοσμίως. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας[1] είναι υπεύθυνες για το 30% των θάνατων ανά έτος. Στις εύπορες κοινωνίες όπου υπάρχουν τα μέσα και οι πόροι η μαγνητική ακτινογραφία και η εξέταση με υπέρηχο έχουν αντικαταστήσει την διάγνωση κάποιου καρδιακού προβλήματος από μια απλή στηθοσκόπιση του ασθενούς. Όμως στις υποανάπτυκτες χώρες του πλανήτη η κατάσταση δεν είναι η ίδια, καθώς η έλλειψη εξειδικευμένου προσωπικού ή ακόμα και γιατρών σε κάποιες περιπτώσεις οδηγεί σε μη έγκαιρη διάγνωσή ενός καρδιακού προβλήματος έχοντας ως αποτέλεσμα τα προαναφερόμενα.

Μία λύση στο πρόβλημα αυτό προσπάθησε να δώσει η πρόκληση του PhysioNet το 2016 [2]. Μέσα απο αυτήν, οι υπεύθυνοι ήθελαν να ενθαρρύνουν την ανάπτυξη αλγορίθμων κατηγοριοποίησης των καρδιακών ήχων καθώς επίσης και την δημιουργία μιας μεγάλης δημόσιας βάσης δεδομένων με καρδιακούς ήχους. Κατάφεραν τη συλλογή 4430 ηχογραφήσεων από 1072 άτομα και 233.512 φωνοκαρδιογραφήματα σε όλο τον πλανήτη. Τα δείγματα αφορούν τόσο υγιείς όσο και καρδιοπαθείς ασθενείς που πάσχουν από διάφορες παθήσεις όπως στεφανιαία νόσο ή πάθηση στις βαλβίδες. Οι ηχογραφήσεις προέρχονται από ετερογενείς πηγές τόσο από κλινικό εξοπλισμό όσο και από επισκέψεις ιατρών στο σπίτι. Ακόμη παρέχονται πληροφορίες για το κάθε δείγμα όπως ηλικία, φύλλο,αριθμός ηχογραφήσεων ανά ασθενή,διάρκεια και περιοχή ηχογράφησης. Τα φωνοκαρδιογραφήματα προέρχονται από διαφορετικά μέρη του σώματος των ασθενών με τα τέσσερα πιο συχνά σημεία να είναι οι περιοχές των βαλβίδων(αορτική,μιτροειδής,τριγλωχίνα,πνευμονική). Ένα ποσοστό δεδομένων έχουν αρκετό θόρυβο ώστε να είναι ρεαλιστική η βάση δεδομένων. Ο σκοπός είναι μέσα από ένα μικρό δείγμα ήχου μερικών δευτερολέπτων έως και μερικά λεπτά, μέσω του αλγορίθμου,να διαχωρίζεται ο ήχος σε φυσιολογικό ή μη φυσιολογικό, οπότε και χρειάζεται να διαγνωστεί από κάποιον ειδικό. Θα μπορούσε να χρησιμοποιηθεί ώς κάποια εφαρμογή για κινητό ή ακόμα ως ιστοσελίδα όπου θα ανεβαίνει ένα ηχητικό αρχείο και θα ξεκινάει ο αλγόριθμος κατηγοριοποίησης. Αρκετές απόπειρες έχουν γίνει για την σχεδίαση τέτοιων αλγορίθμων στηριζόμενοι είτε σε εξαγωγή δεδομένων και μέσω μηχανκής μάθησης να γίνει ο διαχωρισμός είτε με τη χρήση νευρωνικών δικτύων. Εμείς θα προσπαθήσουμε να εξάγουμε τα αποτελέσματα μας με τον δεύτερο τρόπο όπως φαινεται αναλύτικότερα στην ενότητα 3.

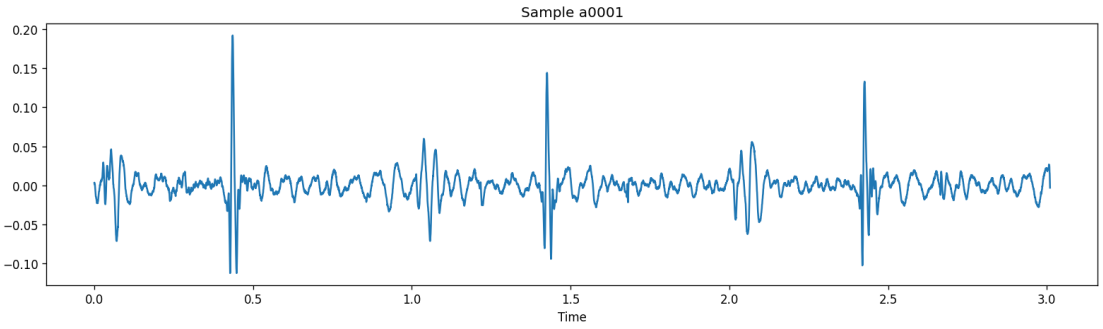
2 Προεπεξεργασία

2.1 Κατάτμηση δειγμάτων

Κατά τη διάρκεια του καρδιακού κύκλου,η καρδιά παράγει ηλεκτρική δραστηριότητα, η οποία στη συνέχεια προκαλεί κολπικές και κοιλιακές συσπάσεις. Αυτό με την σειρά του οδηγεί το αίμα γύρω από το σώμα. Το άνοιγμα και το κλείσιμο των καρδιακών βαλβίδων σχετίζεται με επιταχύνσεις επιβραδύνσεις του αίματος, προκαλώντας δονήσεις ολόκληρης της καρδιακής δομής. Αυτές οι δονήσεις ακούγονται στα θωρακικά τοιχώματα και η ακρόαση συγκεκριμένων καρδιακών ήχων μπορεί να δώσει μια ένδειξη για την υγεία της καρδιάς. Το φωνοκαρδιογράφημα (PCG) είναι η γραφική αναπαράσταση μιας εγγραφής καρδιακού ήχου. Ένα τυπικό PCG φαίνεται στην παρακάτω εικόνα 2.1:



Σχήμα 2.1: PCG



Σχήμα 2.2: Αναπαράσταση ήχου με αρχή το πρώτο S1 και διάρκεια 3 δευτερόλεπτα

Όπως φαίνεται και στην εικόνα ένας πλήρης καρδιακός κύκλος στο φωνοκαρδιογράφημα αποτελείται από τέσσερις διακριτές περιοχές. Αυτές είναι οι S1, συστολή, S2 και διαστολή. Και οι τέσσερις ήχοι που αποτελούν ένα κύκλο σχετίζονται με το κλείσιμο συγκεκριμένων βαλβίδων και την ροή αίματος από και προς τις κοιλίες.

Προκειμένου να διαχωρίσουμε τα PCG που έχουμε στην διάθεσή μας στα παραπάνω μέρη, χρησιμοποιούμε τον αλγόριθμο του Springer[3], τον οποίο παρείχε ο διαγωνισμός στους συμμετέχοντες. Ο αλγόριθμος αυτός εντοπίζει τα σημεία εκείνα στα οποία ξεκινάει κάθε ένα από τα στάδια που αναφέρθηκε παραπάνω (S1, S2, συστολή, διαστολή). Στην συνέχεια, έχοντας βρει το σημείο του πρώτου S1, κρατάμε 3 δευτερόλεπτα, από αυτό και μετά. Αυτή η διαδικασία θα γίνεται ώστε τα δείγματα με τα οποία θα εκπαιδεύσουμε το νευρωνικό δίκτυο να είναι "ευθυγραμμισμένα" μεταξύ τους.

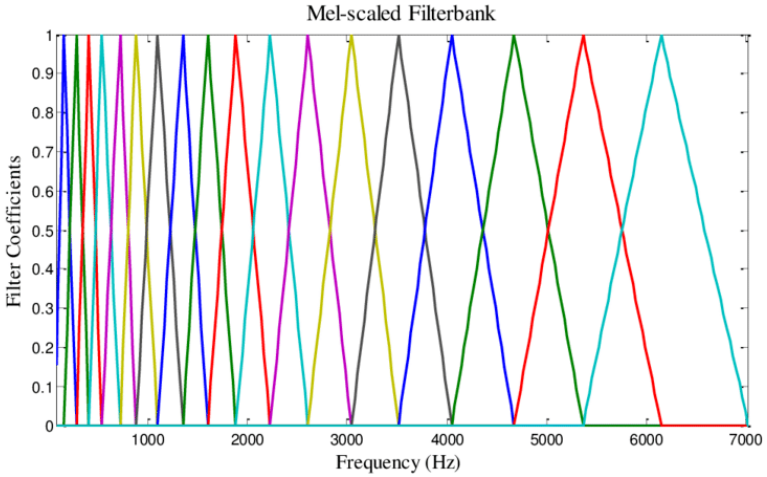
3 Εξαγωγή χαρακτηριστικών

3.1 Mel Frequency Cepstral Coefficients

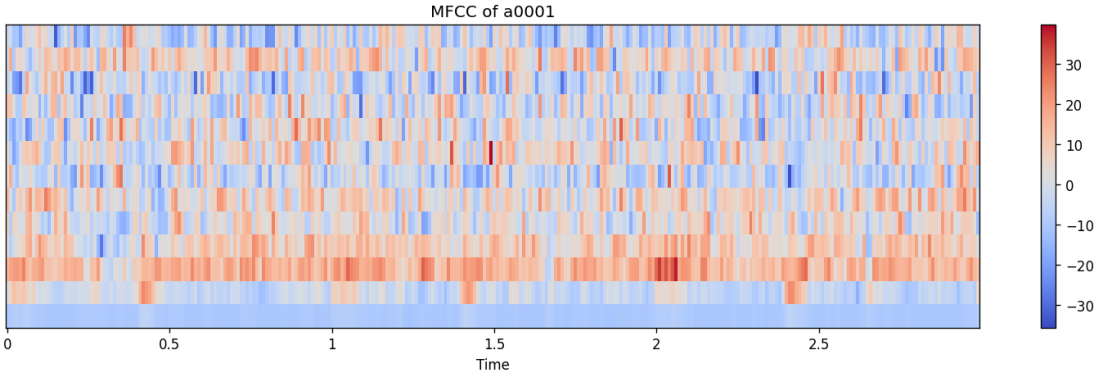
Ένα από τα χαρακτηριστικά που εξάγαμε από τα σήματα καρδιακών ήχων που είχαμε στη διάθεση μας ώστε να εκπαιδεύσουμε το νευρωνικό δίκτυο είναι τα Mel Frequency Cepstral Coefficients. Ο υπολογισμός τους έγινε στα ήδη προεπεξεργασμένα δεδομένα και συγκεκριμένα στο κάθε ένα σήμα που δημιουργήθηκε από την έξοδο των παραθύρων.Το σύνολο των συντελεστών που παρήγαγε αυτή η διαδικασία είναι 13 από τους οποίους οι 12 αναπαριστούν την περισσότερη πληροφορία της φασματικής περιβάλλουσας και ο $13^{ος}$ αναπαριστά τη συνολική ενέργεια του σήματος. Δεν επιλέχθηκαν περισσότεροι από 13 συντελεστές καθώς η αύξηση του αριθμού τους πάνω από αυτό το όριο έχει ως αποτέλεσμα την ταχεία μεταβολή των συντελεστών γεγονός που δυσχεραίνει την εκπαίδευση του νευρωνικού δικτύου. Ο κύριος λόγος που επιλέχθηκαν τα mfcc's είναι ότι αποτελούν την καλύτερη προσέγγιση της

λειτουργίας του κοχλία του ανθρώπινου αυτιού που είναι επιλεκτικός στις συχνότητες και στο πως αντιδρά σε αυτές. Επιγραμματικά η διαδικασία για τον υπολογισμό των mfcc είναι [4]

- Εφαρμογή επικαλυπτόμενων παραθύρων στο σήμα
- Υπολογισμός του φάσματος ενέργειας
- Εφαρμογή του φίλτρου Mel και άθροισμα της ενέργειας του κάθε φίλτρου
- Λογαρίθμηση του αποτελέσματος του προηγούμενου βήματος
- Εφαρμογή μετασχηματισμού συνημιτόνων



Σχήμα 3.1: Φίλτρο Mel

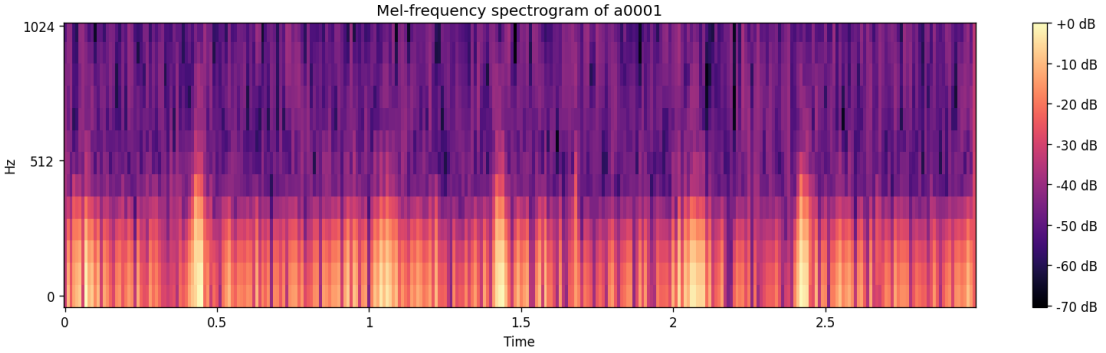


Σχήμα 3.2: Αναπαράσταση τιμών MFCC σε χάρτη θερμότητας για το δείγμα στο σχήμα 2.2

3.2 Mel Spectogram

Το δεύτερο χαρακτηριστικό που εξήχθει είναι το σπεκτόγραμμα mel που αποτελεί μια μη γραμμική αναπαράσταση του συχνοτικού περιεχομένου του σήματος. Η διαδικασίας υπολογισμού είναι παρόμοια με εκείνη για τα mfcc. Επιγραμματικά τα βήματα που ακολουθούνται είναι [5]

- Εφαρμογή επικαλυπτόμενων παραθύρων στο σήμα
- Υπολογισμός του φάσματος ενέργειας μέσω του γρήγορου μετασχηματισμού Fourrier
- Μετατροπή του συχνοτικού περιεχομένου σε dB
- Εφαρμογή της κλίμακας mel στο συχνοτικό περιεχόμενο



Σχήμα 3.3: Mel σπεκτόγραμμα για το δείγμα στο σχήμα 2.2

3.3 Εξαγωγή στην Python

Για τις ανάγκες υλοποίησης του νευρωνικού δικτύου ήταν απαραίτητη η εξαγωγή των mfcc's από τα φωνοκαρδιογραφήματα που είχαμε στη διάθεσή μας. Μετά την προεπεξεργασία στην οποία υποβλήθηκαν όταν πλέον είχαμε τα δείγματα μας χωρισμένα σε μικρότερα από επικαλυπτόμενα παράθυρα τότε σε κάθε ένα καινούριο ηχητικό σήμα το οποίο είχε δημιουργηθεί εφαρμόστηκε η συνάρτηση mfcc της βιβλιοθήκης `python_speech_features` η έξοδος της οποίας, τα 13 mfcc's, αποθηκεύτηκαν σε μορφή εικόνας που είναι και τα χαρακτηριστικά εκπαίδευσης του συνελικτικού νευρωνικού δικτύου. Όσο αφορά το σπεκτόγραμμα mel χρησιμοποιήθηκε από τη βιβλιοθήκη `librosa.feature` η συνάρτηση `melspectogram` που πάλι η έξοδος του αποθηκεύεται σε μορφή εικόνας. Η συνάρτηση εφαρμόστηκε στα ίδια ακριβώς σήματα με εκείνα που ήταν είσοδος της mfcc.

4 Νευρωνικό Δίκτυο

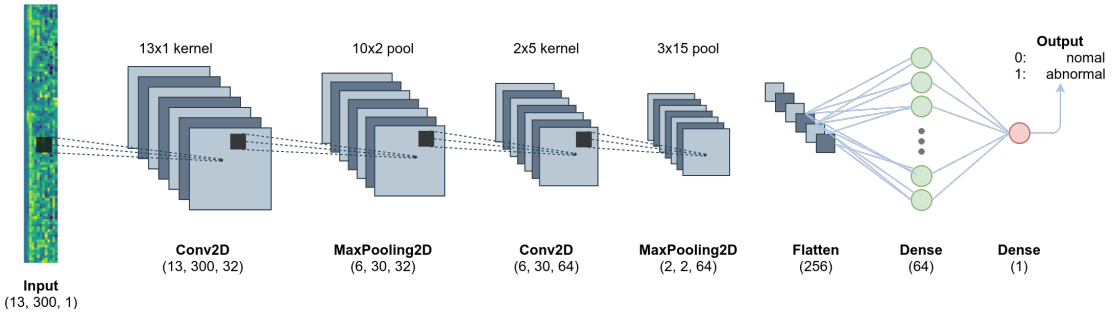
Έχοντας πλέον μετατρέψει τον καρδιακό ήχο σε έναν δισδιάστατο πίνακα από τιμές MFCC μπορούμε εύκολα να τον αναπαραστήσουμε ως εικόνα και πιο συγκεκριμένα ως χάρτη θερμότητας (παράδειγμα στο σχήμα 3.2). Για αυτό τον λόγο επιλέχτηκε η χρήση συνελικτικού νευρωνικού δικτύου CNN, καθώς είναι ένας από τους καλύτερους τύπους νευρωνικού δικτύου για την κατηγοριοποίηση εικόνων [6]. Τέλος στον διαγωνισμό της Physionet [7] υπήρχαν κι άλλες προσπάθειες με χρήση CNN με αρκετά καλά ποσοστά ακρίβειας [8], το οποίο μας ώθησε ακόμα περισσότερο στην επιλογή αυτού του τύπου νευρωνικού δικτύου.

4.1 Αρχιτεκτονική δικτύου

Αρχικά το δίκτυο μας δέχεται ως είσοδο έναν πίνακα 13×300 που είναι οι τιμές των MFCC για 3 δευτερόλεπτα ήχου με επικαλυπτόμενα παράθυρα διάρκειας 20ms και βήμα 10ms. Στην συνέχεια τα δεδομένα περνάνε από δύο συνελικτικά στρώματα για να καταλήξουν στο τέλος σε πλήρους συνδεδεμένους νευρώνες ώστε να καταφέρουμε να πάρουμε το τελικό δυαδικό αποτέλεσμα (για κάποιον με πιθανή καρδιοπάθεια 1, αλλιώς 0).

Όπως φαίνεται και στο σχήμα 4.1 μετά από κάθε συνελικτικό στρώμα έχουμε ένα στρώμα MaxPolling το οποίο είναι πολύ σημαντικό καθώς μειώνει αρκετά το μέγεθος των αρχικών δεδομένων και έτσι βοηθάει στην εκπαίδευση του δικτύου.

Τέλος είναι σημαντικό να αναφερθεί ότι σε όλα τα στρώματα εκτός από αυτά των MaxPolling χρησιμοποιείτε ως συνάρτηση ενεργοποίησης η `relu` κι ότι ως optimizer επιλέχθηκε ο AdamOptimizer. Επίσης τα μεγέθη των πυρήνων και ο αριθμός των φίλτρων και νευρώνων επιλέχτηκαν μετά από πολλές δοκιμές κρατώντας αυτά που απέδιδαν την μεγαλύτερη ακρίβεια.



Σχήμα 4.1: Σχεδιάγραμμα της αρχιτεκτονικής του CNN δικτύου μας

4.2 Εκπαίδευση

Για να εκπαιδεύσουμε τον δίκτυό μας χρησιμοποιήσαμε ένα σύνολο από 3240 φωνοκαρδιογραφήματα που παρείχε η Physionet [7] στους διαγωνιζόμενους. Από αυτά τα δείγματα χρησιμοποιήσαμε το 60% για εκπαίδευση του δικτύου και το 40% για επαλήθευση της ακρίβειάς του. Στην συνέχεια εκπαιδεύσαμε το δίκτυο για 100 εποχές με μέγεθος παρτίδας 32.

5 Αποτελέσματα

Στα παρακάτω σχήματα φαίνονται τρις σημαντικές μετρικές που καταγράφαμε κατά την διάρκεια της εκπαίδευσης.

Πρώτη είναι η μετρική **Accuracy** που μας δείχνει πόσο συχνά οι προβλέψεις του νευρωνικού δικτύου είναι σωστές.

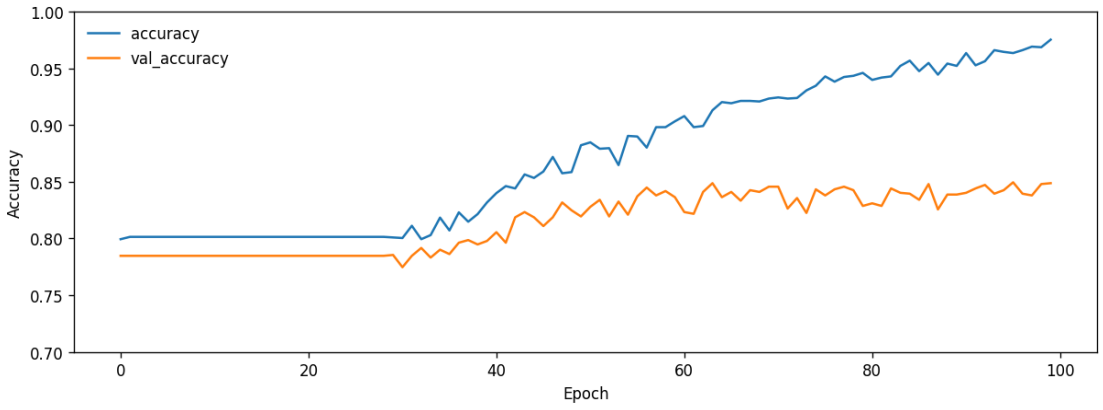
Δεύτερη είναι η μετρική **Loss** και πιο συγκεκριμένα στα σχήματα 5.2 και 5.5 χρησιμοποιήσαμε την *Binary crossentropy loss function* η οποία υπολογίζει πόσο απέχει η πρόβλεψη του νευρωνικού από την πραγματική τιμή μέσω της παρακάτω συνάρτησης 1.

$$\text{Loss} = -\frac{1}{\substack{\text{output}\\\text{size}}} \sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i) \tag{1}$$

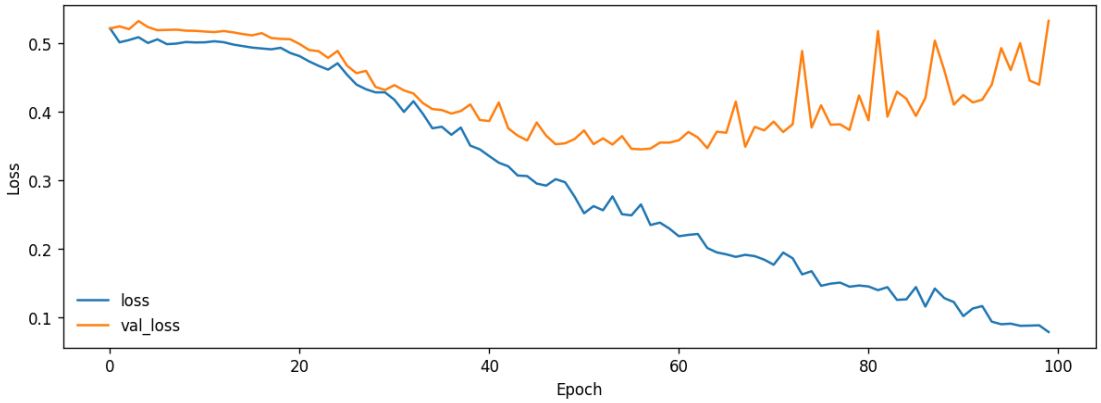
Τελευταία μετρική είναι η **Recall** την οποία θεωρήσαμε πολύ σημαντική για το συγκεκριμένο πρόβλημα, καθώς μας δείχνει το ποσοστό των σωστά κατηγοροποιημένων ατόμων με πρόβλημα, από το σύνολο όλων αυτών τον ατόμων (ιδανικά θα θέλαμε να έχουμε 100% recall, δηλαδή να μπορούμε να εντοπίσουμε κάθε άτομο με καρδιακό πρόβλημα).

5.1 Χρήση MFCC

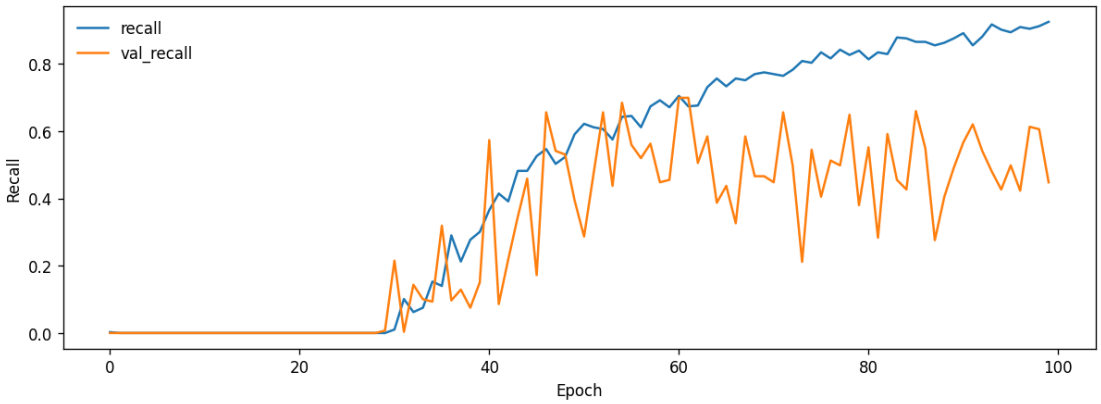
Με την χρήση των MFCC ως είσοδο στο νευρωνικό το accuracy σταθεροποιείται κοντά στο 81%, ενώ το validation_loss έχει αρκετή απόκλιση από το loss και κυμαίνετε ανάμεσα στο 0.4 με 0.5. Το recall έχει δραματικές μεταβολές από εποχή σε εποχή κάτι που θα εξηγηθεί παρακάτω, με μια μέση τιμή 45% μετά την 40^{στη} εποχή.



Σχήμα 5.1: Accuracy νευρωνικού δικτύου ανά εποχή με χρήση **MFCCs** ως είσοδο



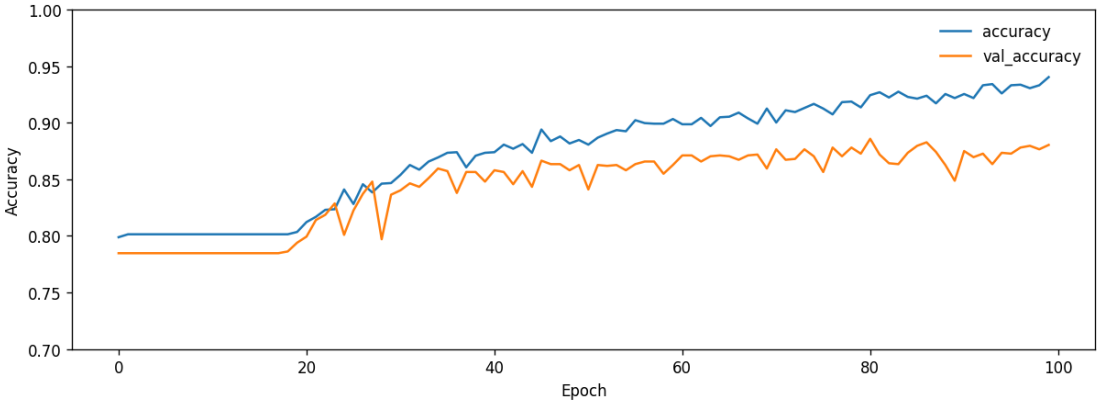
Σχήμα 5.2: Loss νευρωνικού δικτύου ανά εποχή με χρήση **MFCCs** ως είσοδο



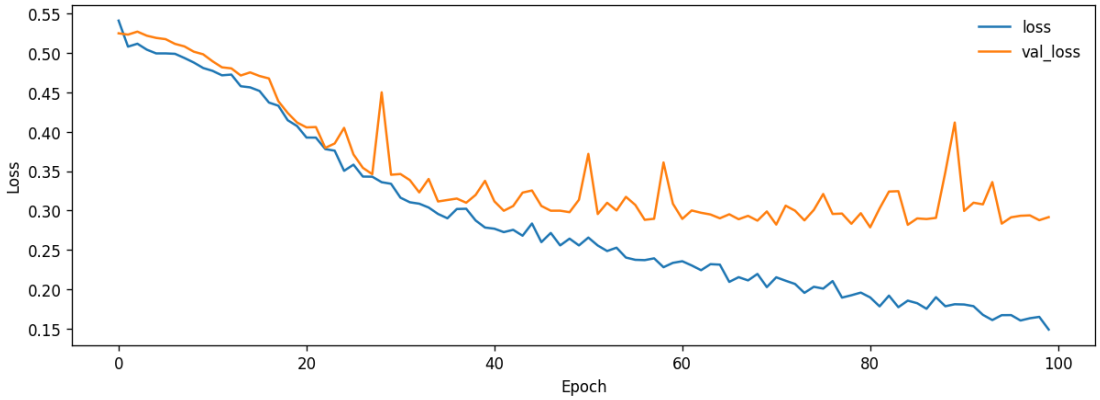
Σχήμα 5.3: Recall νευρωνικού δικτύου ανά εποχή με χρήση **MFCCs** ως είσοδο

5.2 Χρήση Mel spectogram

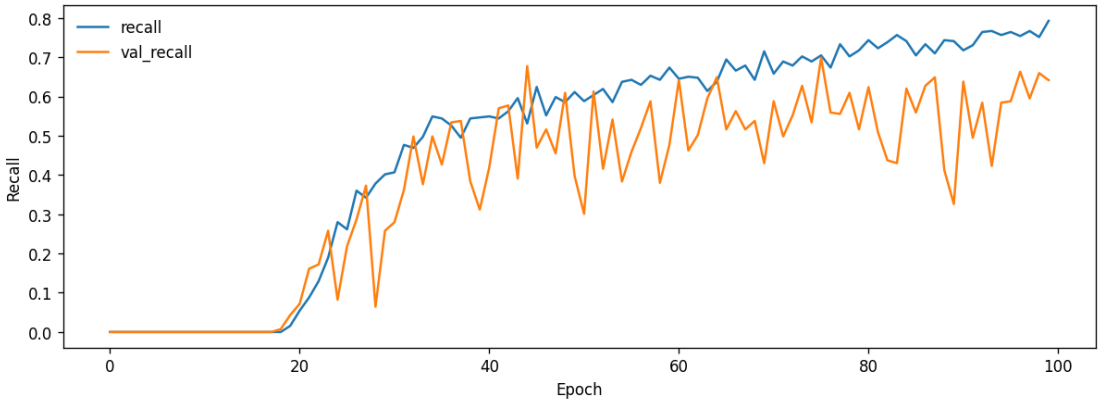
Όταν χρησιμοποιήσαμε τα mel σπεκτογράμματα ως είσοδο στο νευρωνικό δίκτυο, είδαμε ελαφρώς καλύτερα αποτελέσματα. Αφού πλέον η ακρίβεια σταθεροποιείται στο 86% και το validation_loss δεν απέχει τόσο από το loss με μέση τιμή κάτω από 0.35. Όσο αναφορά το recall, έχουμε κι εδώ μεγάλες μεταβολές από εποχή σε εποχή με την τιμή να βρίσκετε γύρω από το 57%.



Σχήμα 5.4: Accuracy νευρωνικού δικτύου ανά εποχή με χρήση **spectogram** ως είσοδο



Σχήμα 5.5: Loss νευρωνικού δικτύου ανά εποχή με χρήση **spectogram** ως είσοδο



Σχήμα 5.6: Recall νευρωνικού δικτύου ανά εποχή με χρήση **spectogram** ως είσοδο

5.3 Συμπεράσματα

Όπως είδαμε παραπάνω η καλύτερη προσπάθεια μας είχε 86% accuracy με 57% recall όμως ακόμα κι αυτή η προσπάθεια δεν είναι ικανοποιητική λόγω των άνισων δεδομένων. Τα δείγματα που παρείχε η Physionet [7] ήταν συνολικά 3228 από τα οποία μόνο τα 665 (20%) ήταν από ανθρώπους με καρδιοπάθεια. Αυτό σημαίνει ότι ακόμα κι ένα νευρωνικό που θα κατηγοροποιούσε όλα τα δείγματα ως 0 (δηλαδή χωρίς καρδιοπάθεια) θα είχε accuracy περίπου 80% αλλά με 0% recall. Αυτή είναι και η αιτία που βλέπουμε και τις μεγάλες μεταβολές από εποχή σε εποχή στο recall καθώς το δίκτυο έχει την τάση να κατηγοριοποιεί κάθε δείγμα ως άτομο χωρίς καρδιοπάθεια λόγω της αριθμητικής υπεροχής αυτών τον ατόμων στα δείγματα.

Αναφορές

[1] World Health Organisation, "Cardiovascular diseases (cvds)," 2021, last accessed 13 November 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016," in *2016 Computing in cardiology conference (CinC)*. IEEE, 2016, pp. 609–612.

[3] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2015.

[4] F. Haytham, "Speech processing for machine learning: Filter banks, mel-frequency cepstral coefficients (mfccs) and what's in-between," 2016. [Online]. Available: https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

[5] L. Roberts, "Understanding the mel spectogram," 2020. [Online]. Available: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53

[6] M. Ramprasath, M. V. Anand, and S. Hariharan, "Image classification using convolutional neural networks," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 17, pp. 1307–1319, 2018.

[7] C. Liu, D. Springer, B. Moody, I. Silva, A. Johnson, M. Samieinasab, R. Sameni, R. Mark, and G. Clifford, "Classification of heart sound recordings - the physionet computing in cardiology challenge 2016," 2016, last accessed 13 November 2021. [Online]. Available: https://physionet.org/content/challenge-2016/1.0.0/

[8] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 813–816.