



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών Και Μηχανικών Υπολογιστών

Τεχνολογία Ήχου Και Εικόνας Κατηγοριοποίηση Καρδιακών Ήχων

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

14 Νοεμβρίου 2021

Περιεχόμενα

1	Εισαγωγή	2
2	Καρδιακοί ήχοι	3
3	Η προσέγγισή μας	4
3.1	Κατάτμηση Δειγμάτων	4
3.2	Μετασχηματισμός MFCC	4
3.3	Νευρωνικό δίκτυο	5
3.3.1	Αρχιτεκτονική δικτύου	5

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 στηθοσκόπηση 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 διάγνωση 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,μित्रοειδής,τριγλώχίνα,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 ως 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 μηχανικής μάθησης να γίνει ο διαχωρισμός είτε με τη χρήση νευρωνικών δικτύων. Εμείς θα προσπαθήσουμε να εξάγουμε τα αποτελέσματα μας με τον δεύτερο τρόπο όπως φαίνεται αναλυτικότερα στο Κεφάλαιο 3.

2 Καρδιακοί ήχοι

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:

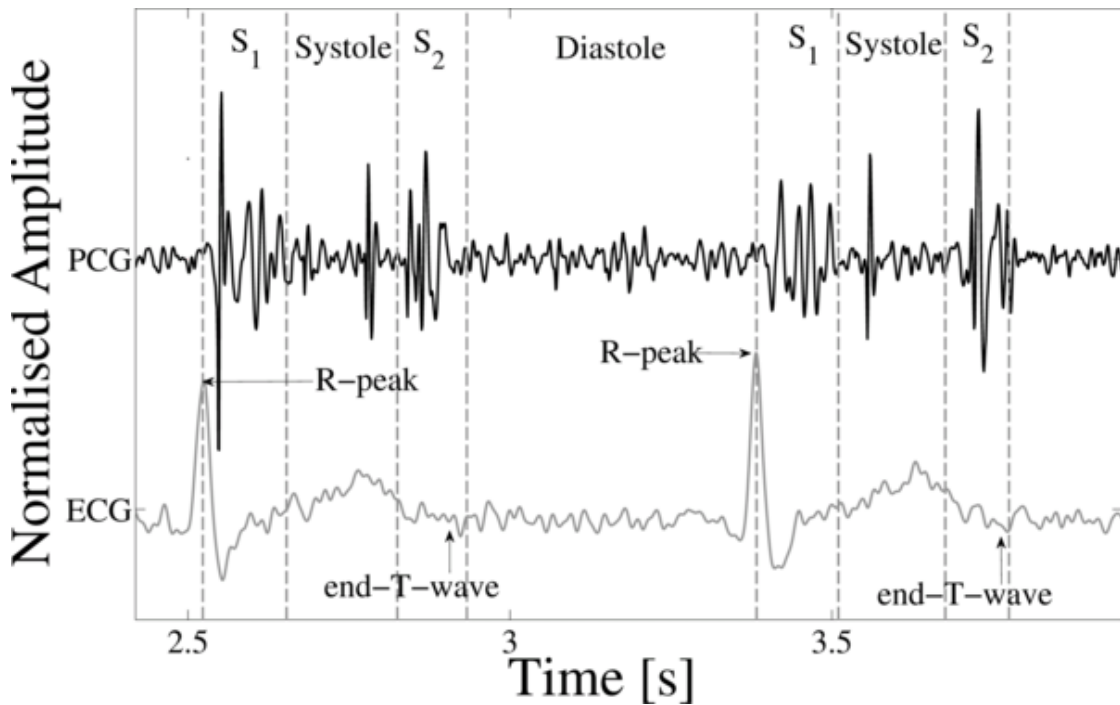


Figure 2.1: PCG και ECG [2]

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. [3]

3 Η προσέγγισή μας

Η προσέγγισή μας έχει 3 βασικά μέρη:

- **Κατάτμηση** των PCG δειγμάτων στις βασικές περιοχές του καρδιακού ήχου.
- **MFCC μετασχηματισμό** των PCG δειγμάτων σε αναπαράσταση χρόνου-συχνότητας της κατανομής της ενέργειας του σήματος.
- **Εκπαίδευση & Κατηγοροποίηση** των MFCC heat maps με την χρήση του συνελκτικού νευρωνικού δικτύου.

3.1 Κατάτμηση Δειγμάτων

Για να κάνουμε την κατάτμηση των PCG δειγμάτων όπως φαίνεται στην εικόνα 2.1 θα χρησιμοποιήσουμε τον αλγόριθμο του Springer [4] τον οποίο τον παρέχει ο διαγωνισμός στους συμμετέχοντες. Ωστόσο δεν θα χρησιμοποιήσουμε όλα τα δεδομένα που μας δίνει ο αλγόριθμος, αλλά αυτό που θα κάνουμε είναι να βρούμε που ξεκινάει το πρώτο *S1* και στη συνέχεια θα αναλύουμε τα 3 επόμενα δευτερόλεπτα. Αυτή η διαδικασία θα γίνεται ώστε τα δείγματα με τα οποία θα εκπαιδεύσουμε το νευρωνικό δίκτυο να είναι “ευθυγραμμισμένα” μεταξύ τους.

3.2 Μετασχηματισμός MFCC

Αφού πάρουμε τα 3 δευτερόλεπτα σκοπός είναι να δημιουργήσουμε μια εικόνα που θα απεικονίζει χαρακτηριστικά του ηχητικού αποσπάσματος ώστε να μπορέσουμε να προπονήσουμε το νευρωνικό μας δίκτυο. Αυτό μπορούμε να το πετύχουμε υπολογίζοντας τις MFCC τιμές του καρδιογραφήματος με την εξής διαδικασία:

1. Παίρνουμε επικαλυπτόμενα παράθυρα πάνω στο ηχητικό δείγμα
2. Υπολογίζουμε τον μετασχηματισμό Fourier για κάθε παράθυρο
3. Εφαρμόζουμε τα Mel φίλτρα και αθροίζουμε τις ενέργειες σε κάθε φίλτρο
4. Υπολογίζουμε τις λογαριθμικές τιμές των παραπάνω ενεργειών
5. Τέλος παίρνουμε τον διακριτό συνημιτονικό μετασχηματισμό των λογαριθμικών τιμών.

Έτσι θα έχουμε 12 MFCC τιμές για κάθε παράθυρο και μαζί με την ολική ενέργεια του παραθύρου ως ξεχωριστή τιμή παίρνουμε 13 τιμές. Αναπαριστώντας αυτά τα χαρακτηριστικά σε ένα γράφημα όπου ο y άξονας θα κυμαίνεται από 0 έως 12 (μία γραμμή για κάθε MFCC τιμή) και ο x άξονας από 0 έως $3000/step_size$ ($step_size$ που θα ορίσουμε για κάθε παράθυρο), θα πάρουμε ένα heat map (βλέπε 3.1) που μπορούμε να χρησιμοποιήσουμε ως εικόνα για το νευρωνικό δίκτυο.

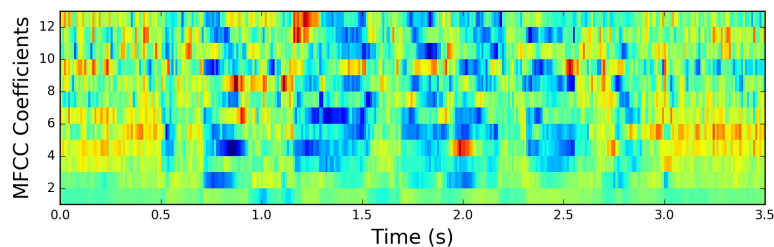


Figure 3.1: MFCC heat map [5]

3.3 Νευρωνικό δίκτυο

Η μετατροπή των δοθέντων μονοδιάστατων χρονοσειρών σε πίνακες θερμότητας δύο διαστάσεων μας επιτρέπει την επεξεργασία δειγμάτων των καρδιακών ήχων υπό τη μορφή εικόνας. Γι' αυτό κρίνεται και επιβεβλημένη η χρήση συνελικτικών νευρωνικών δικτύων λόγω των αυξανόμενων δυνατοτήτων τους στην ανάλυση εικόνων.

3.3.1 Αρχιτεκτονική δικτύου

Το νευρωνικό δίκτυο δέχεται σαν είσοδο ένα κανάλι 13x300 MFCC χάρτη θερμότητας (αν χρησιμοποιήσουμε βήμα 10ms στον MFCC μετασχηματισμό) και δίνει σαν τελική έξοδο την ζητούμενη δυαδική κατηγοριοποίηση. Ενδιάμεσα χρησιμοποιούνται δύο συνελικτικά στρώματα, το καθένα ακολουθούμενο από ένα στρώμα max-pooling, και συνολικά ακολουθούμενο από δύο πλήρως συνδεδεμένα κρυφά στρώματα.

Το πρώτο συνελικτικό στρώμα μαθαίνει 64 2x20 πυρήνες, χρησιμοποιώντας την τεχνική same-padding. Στη συνέχεια εφαρμόζουμε ένα max-pooling φίλτρο 1x20, χρησιμοποιώντας ένα οριζόντιο βήμα μήκους 5, το οποίο έχει ως αποτέλεσμα τη μείωση των διαστάσεων καθενός από τους 64 χάρτες χαρακτηριστικών σε 13x60. Ένα δεύτερο συνελικτικό στρώμα χρησιμοποιεί 64 πυρήνες 2x10 επάνω στο προηγούμενο στρώμα, ξανά με same-padding. Μετά την προαναφερθείσα διαδικασία ακολουθεί και πάλι η διαδικασία max-pooling, με την χρήση φίλτρου 1x4 και βήματος 2, μειώνοντας με αυτό τον τρόπο τη διάσταση του χάρτη χαρακτηριστικών σε 13x30. Ακολουθεί μια διαδικασία πλάτυνσης για την μετατροπή καθενός από τους 64 13x30 χάρτες χαρακτηριστικών σε διάνυσμα μεγέθους 24,960. Το διάνυσμα αυτό εισάγεται στο πρώτο πλήρως συνδεδεμένο κρυφό στρώμα μήκους 1024 κρυφών μονάδων και, ακολουθούμενο από ένα δεύτερο κρυφό στρώμα 512 κρυφών μονάδων, δίνει την δυαδική έξοδο κατηγοριοποίησης.

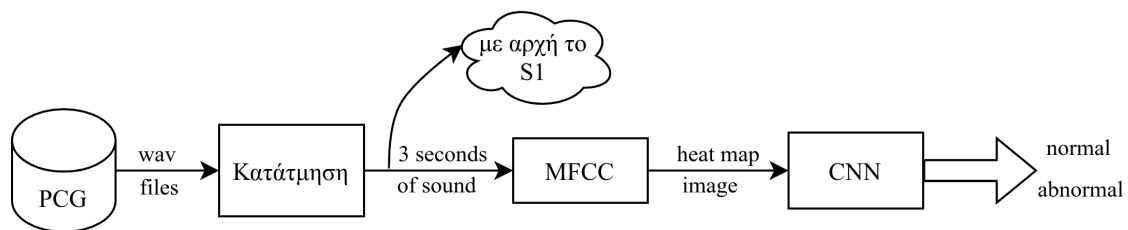


Figure 3.2: Διάγραμμα της προσέγγισής μας για την κατηγοροποίηση καρδικών ήχων

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-\(cvds\)](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] 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/>
- [4] 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.
- [5] H. M. Fayek, “Speech processing for machine learning: Filter banks, mel-frequency cepstral coefficients (mfccs) and what’s in-between,” 2016, last accessed 13 November 2021. [Online]. Available: <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>