



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τεχνολογία Ήχου Και Εικόνας Κατηγοριοποίηση Καρδιακών Ήχων

9^ο Εξάμηνο

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

22 Φεβρουαρίου 2022

Περιεχόμενα

1	Εισαγωγή	2
2	Εξαγωγή χαρακτηριστικών	3
2.1	Mel Frequency Cepstral Coefficients	3
2.2	Εξαγωγή στην Python	4
3	Νευρονικό Δίκτυο	5
3.1	Αρχιτεκτονική δικτύου	5
3.2	Εκπαίδευση	6

1 Εισαγωγή

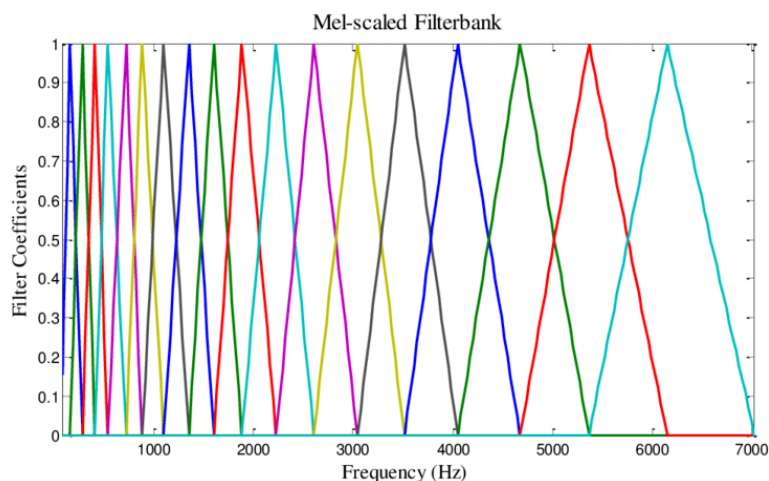
Εισαγωγή

2 Εξαγωγή χαρακτηριστικών

2.1 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 προεπεργασμένα 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 δυσχαιρένει την εκπαίδευση του νευρωνικού δικτύου. Ο κύριος λόγος που επιλέχθηκαν τα mfcc's είναι ότι αποτελούν την καλύτερη προσέγγιση της λειτουργίας του κοχλίου του ανθρώπινου αυτιού που είναι επιλεκτικός στις συχνότητες και στο πως αντιδρά σε αυτές. Επιγραμματικά η διαδικασία για τον υπολογισμό των mfcc είναι [1]

- 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 2.1: Φίλτρο Mel

Το δεύτερο χαρακτηριστικό που εξήχθει είναι το σπεκτόγραμμα mel που αποτελεί μια μη γραμμική αναπαράσταση του συχνοτικού περιεχομένου του σήματος. Η διαδικασία υπολογισμού είναι παρόμοια με εκείνη για τα mfcc. Επιγραμματικά τα βήματα που ακολουθούνται είναι [2]

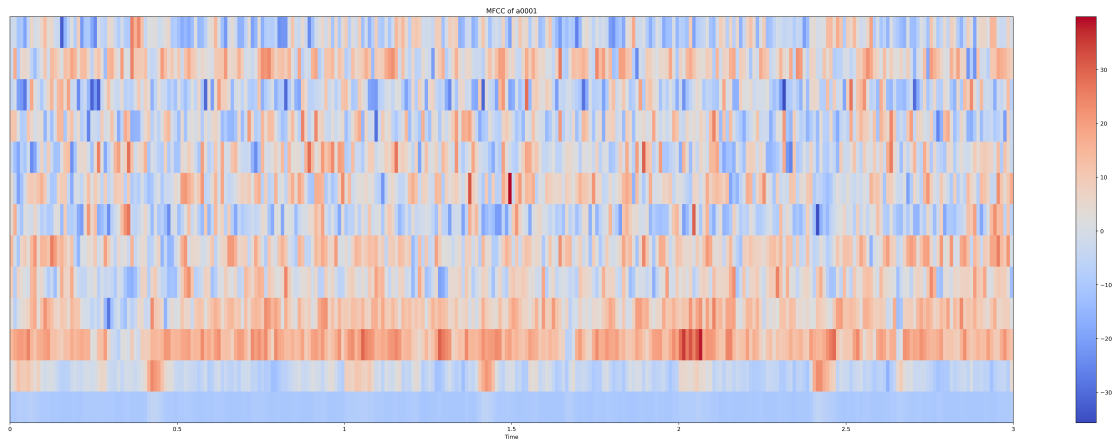
- Εφαρμογή επικαλυπτόμενων παραθύρων στο σήμα
- Υπολογισμός του φάσματος ενέργειας μέσω του γρήγορου μετασχηματισμού Fourier
- Μετατροπή του συχνοτικού περιεχομένου σε dB
- Εφαρμογή της κλίμακας mel στο συχνοτικό περιεχόμενο

2.2 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 συνελκτικού νευρωνικού δικτύου. Όσο αφορά το σπεκτόγραμμα mel χρησιμοποιήθηκε από τη βιβλιοθήκη `librosa.feature` η συνάρτηση `melspectrogram` που πάλι η έξοδος του αποθηκεύεται σε μορφή εικόνας. Η συνάρτηση εφαρμόστηκε στα ίδια ακριβώς σήματα με εκείνα που ήταν είσοδος της mfcc.

3 Νευρονικό 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 3.1). Για αυτό τον λόγο επιλέχτηκε η χρήση συνελικτικού νευρονικού δικτύου CNN, καθώς είναι ένας από τους καλύτερους τύπους νευρονικού δικτύου για την κατηγοριοποίηση εικόνων [3]. Τέλος στον διαγωνισμό της Physionet [4] υπήρχαν κι άλλες προσπάθειες με χρήση CNN με αρκετά καλά ποσοστά ακρίβειας [5], το οποίο μας ώθησε ακόμα περισσότερο στην επιλογή αυτού του τύπου νευρονικού δικτύου.



Σχήμα 3.1: Αναπαράσταση τιμών MFCC σε χάρτη θερμότητας για δείγμα καρδιακού ήχου 3ον δευτερολέπτων

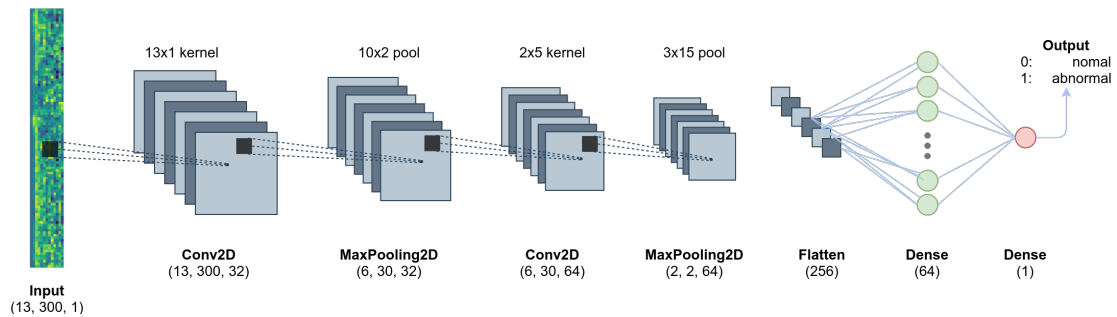
3.1 Αρχιτεκτονική δικτύου

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 3.2 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 \tanh κι ότι ως optimizer

επιλέχθηκε ο AdamOptimizer. Επίσης τα μεγέθη των πυρήνων και ο αριθμός των φίλτρων και νευρώνων επιλέχθηκαν μετά από πολλές δοκιμές κρατώντας αυτά που απέδιδαν την μεγαλύτερη ακρίβεια.



Σχήμα 3.2: Σχεδιάγραμμα της αρχιτεκτονικής του CNN δικτύου μας

3.2 Εκπαίδευση

Για να εκπαιδύσουμε τον δίκτυό μας χρησιμοποιήσαμε ένα σύνολο από 3240 φωνοκαρδιογραφήματα που παρείχε η Physionet [4] 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 50 εποχές με μέγεθος παρτίδας 32.

Αναφορές

- [1] 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>
- [2] L. Roberts, ``Understanding the mel spectrogram," 2020. [Online]. Available: <https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53>
- [3] 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.
- [4] 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/>
- [5] 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.