ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
Χειμερινό Εξάμηνο Ε.ΔΕ.Μ.Μ. 2022-23

Αναφορά 3ης Εργαστηριακής Άσκησης:
Αναγνώριση Είδους και Εξαγωγή Συναισθήματος από Μουσική

Κοτσιφάκου Κοντιλένια Μαρία 03400174
Παπακωνσταντίνου Άννα 03400187

Project εργαστηρίου

Σκοπός της άσκησης είναι η αναγνώριση του είδους και η εξαγωγή συναισθηματικών διαστάσεων από φασματογραφήματα (spectrograms) μουσικών κομματιών. Σας δίνονται 2 σύνολα δεδομένων, το Free Music Archive (FMA) genre με 3834 δείγματα χωρισμένα σε 20 κλάσεις (είδη μουσικής) και τη βάση δεδομένων (dataset) multitask music με 1497 δείγματα με επισημειώσεις (labels) για τις τιμές συναισθηματικών διαστάσεων όπως valence, energy και danceability. Τα δείγματα είναι φασματογραφήματα, τα οποία έχουν εξαχθεί από clips 30 δευτερολέπτων από διαφορετικά τραγούδια.

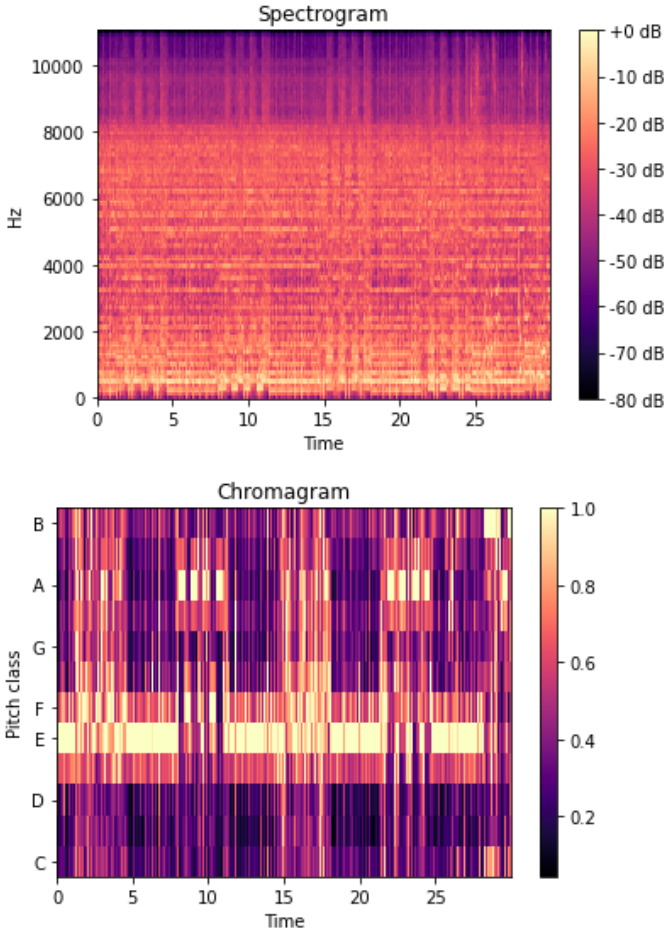
Βήμα 0

Το βήμα 0 αποτελεί εξοικείωση με το Kaggle, εξερεύνηση φακέλων, φόρτωση των dataset και απαραίτητων βιβλιοθηκών της Python.

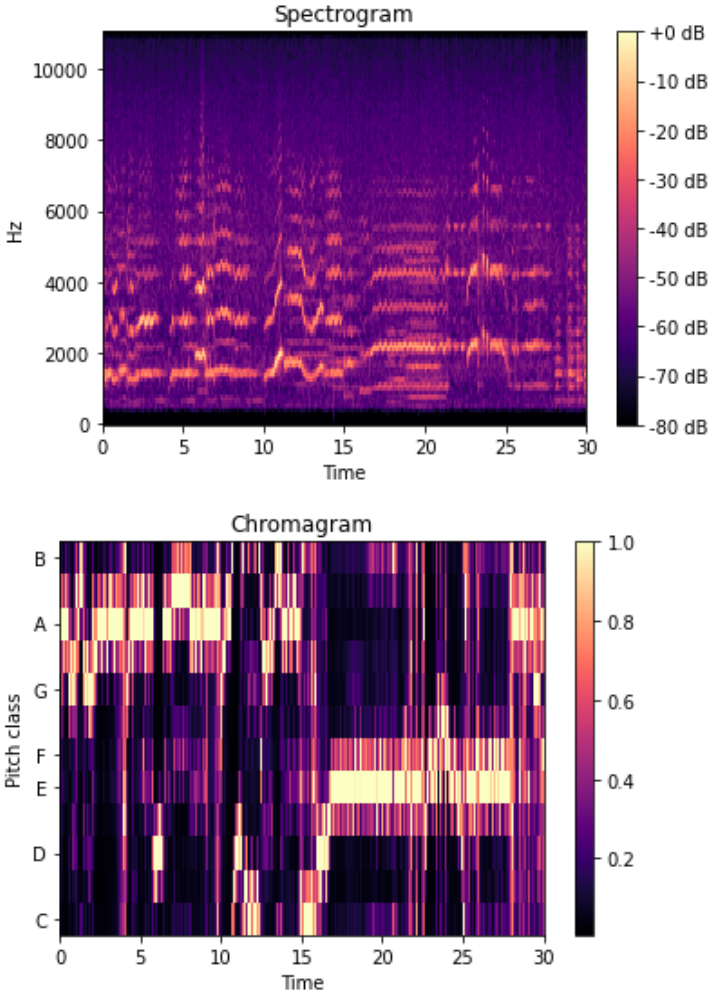
Βήμα 1-3

Αυτά τα βήματα εκτελέστηκαν ομαδοποιήθηκαν και εκτελέστημα μια φορά για τα φασματογραφήματα και χρωμογραφήματα στα μη επεξεργασμένα αρχεία, ενώ στην συνέχεια εκτελέστηκαν για τα συγχρονισμένα φασματογραφήματα και χρωμογραφήματα στο ρυθμό της μουσικής.

Αρχικά επιλέχθηκαν τυχαία δύο αρχεία με διαφορετική κατηγορία και αφού φορτώθηκαν με κατάλληλο τρόπο διαχωρίστηκε κάθε αρχείο σε δύο μέρη ένα για το φασματογράφημα και ένα για το χρωμογράφημα.



Εικόνα 1.1. *Απεικόνιση φασματογραφήματος και χρωματογραφήματος για το πρώτο αρχείο*

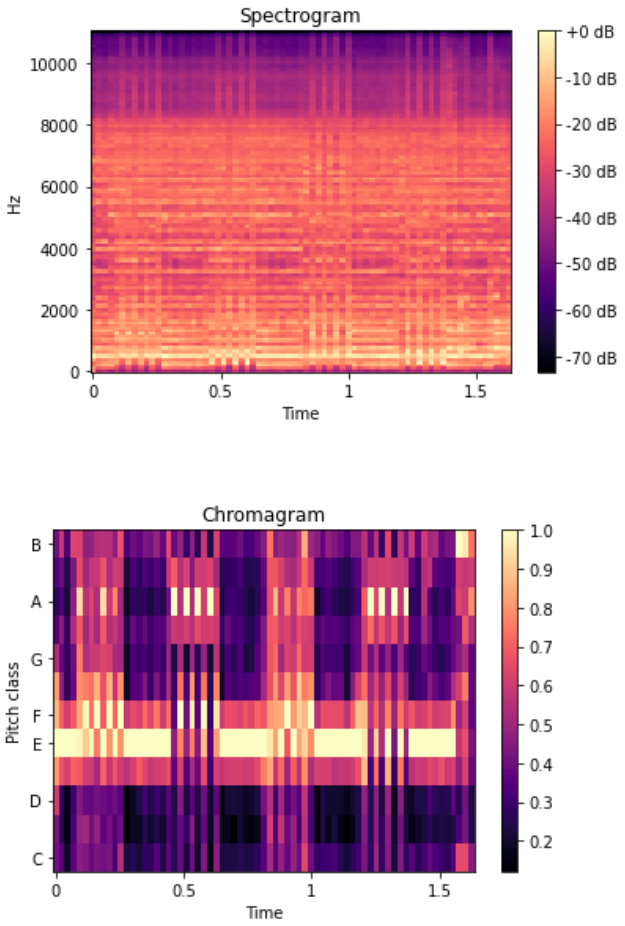


Εικόνα 1.2. *Απεικόνιση φασματογραφήματος και χρωματογραφήματος για το δεύτερο αρχείο*

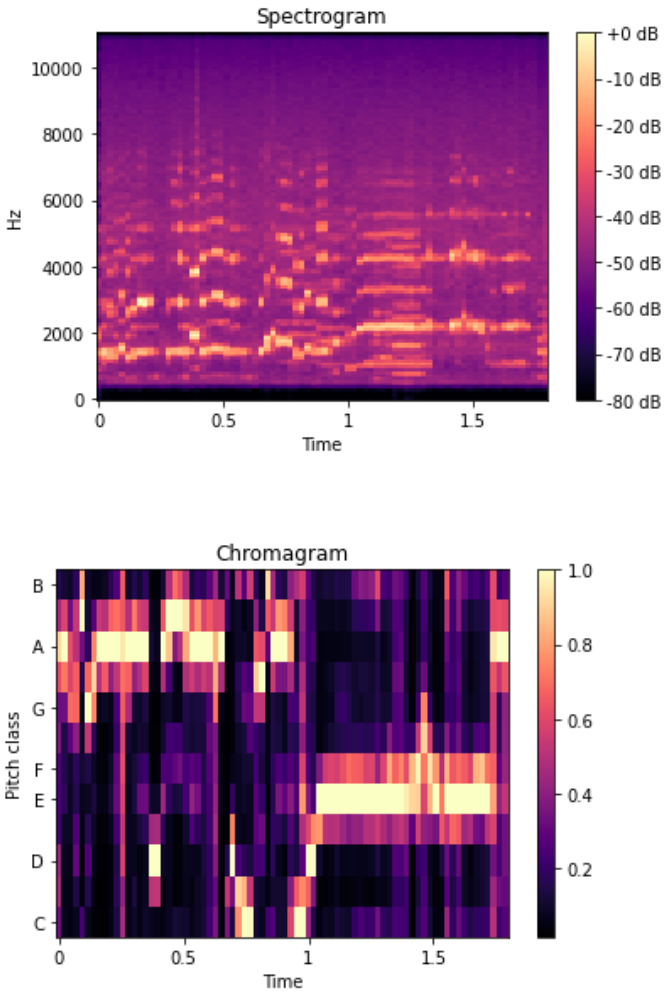
Κάθε φασματογραφήματος είναι στην ουσία αρκετοί FFT μετασχηματισμοί από επικαλυπτόμενα κινούμενα παράθυρα όπου ο ένας μετασχηματισμός είναι πάνω στον άλλον. Είναι ένας τρόπος για την οπτική αναπαράσταση της έντασης ή του πλάτους του σήματος που διαφοροποιείται στον χρόνο και σε διαφορετικές συχνότητες. Ο άξονας y έχει μετασχηματιστεί σε λογαριθμική κλίματα (ώστε να συμβαδίζει με τον τρόπο που αντιλαμβάνεται το ανθρώπινο αυτί τον ήχο) και η χρωματική κλίμακα είναι σε dB.

Οι απεικονίσεις στις εικόνες 1.1 και 1.2 διαφέρουν αρκετά για τα δύο είδη μουσικής. Φαίνεται ότι στο πρώτο ηχητικό αρχείο τα dB είναι πιο ψηλά από ότι στο δεύτερο. Επίσης, το πρώτο αρχείο φαίνεται να έχει ένα μοτίβο στο χρωμογράφημα κάτι που θα μπορούσε να υποδείξει ηχητικά μοτίβα όσον αφορά τις νότες, ενώ αυτό δεν παρατηρείται στο δεύτερο, το οποίο φαίνεται να διαχωρίζεται οπτικά σε δύο μέρη.

Στην συνέχεια, επιλέχθηκαν από το dataset με τα συγχρονισμένα στο ρυθμό της μουσικής φασματογραφήματα και χρωμογραφήματα τα ίδια τυχαία αρχεία για χάρη της σύγκρισης των αποτελεσμάτων. Αφού φορτώθηκαν με κατάλληλο τρόπο διαχωρίστηκε πάλι κάθε αρχείο σε δύο μέρη ένα για το φασματογράφημα και ένα για το χρωμογράφημα.



Εικόνα 1.3. *Απεικόνιση συγχρονισμένου φασματογραφήματος και χρωματογραφήματος για το πρώτο αρχείο*



Εικόνα 1.4. *Απεικόνιση συγχρονισμένου φασματογραφήματος και χρωματογραφήματος για το δεύτερο αρχείο*

Οι απεικονίσεις στις εικόνες 1.3 και 1.4 φαίνεται να συμπιέζουν οπτικά την πληροφορία που διακρίνεται στις Εικόνες 1.1 και 1.2.

Σε κάθε μέρος ξεχωριστά υπολογίστηκαν οι διαστάσεις των προηγούμενων χαρακτηριστικών αλλά συγκεντρώνουμε στις Εικόνες 1.5 και 1.6 τα αποτελέσματα αυτά. Γενικά οι διαστάσεις είναι διαφορετικές για κάθε ηχητικό αρχείο άρα δεν μπορούν να χρησιμοποιηθούν σε ένα LSTM καθώς απαιτείται συγκεκριμένος αριθμός ακολουθίας εισόδου. Επιπλέον, τα πρώτα έχουν πολύ μεγάλο μήκους ακολουθίας κατι που θα ήταν μη αποδοτικό ως είσοδο σε ένα LSTM αφού περιλαμβάνει υπολογιστικά ακριβές διαδικασίες και θα πάρει πολύ χρόνο για να εκπαιδευτεί. Τα αρχεία που έχουν συγχρονισμό σε beat έχουν εμφανώς μικρότερη διάσταση και φαίνεται να έχει γίνει καλή συμπίεση οπτικά, άρα δεν χάνεται κάποια πολύ σημαντική πληροφορία και μπορούν να χρησιμοποιηθούν για εκπαίδευση LSTM δεδομένου ότι χρησιμοποιηθεί μια μέθοδος padding.

```
Dimensions of spectogram 1 are: (128, 1291)
Dimensions of spectogram 2 are: (128, 1293)
Dimensions of chromagram 1 are: (12, 1291)
Dimensions of chromagram 2 are: (12, 1293)
```

Εικόνα 1.5. *Διαστάσεις των φασματογραφημάτων και των χρωματογραφημάτων για τα δύο αρχεία*

```
Dimensions of beat-synced spectogram 1 are: (128, 71)
Dimensions of beat-synced spectogram 2 are: (128, 78)
Dimensions of beat-synced chromagram 1 are: (12, 71)
Dimensions of beat-synced chromagram 2 are: (12, 78)
```

Εικόνα 1.5. *Διαστάσεις των συγχρονισμένων φασματογραφημάτων και των χρωματογραφημάτων για τα δύο αρχεία*

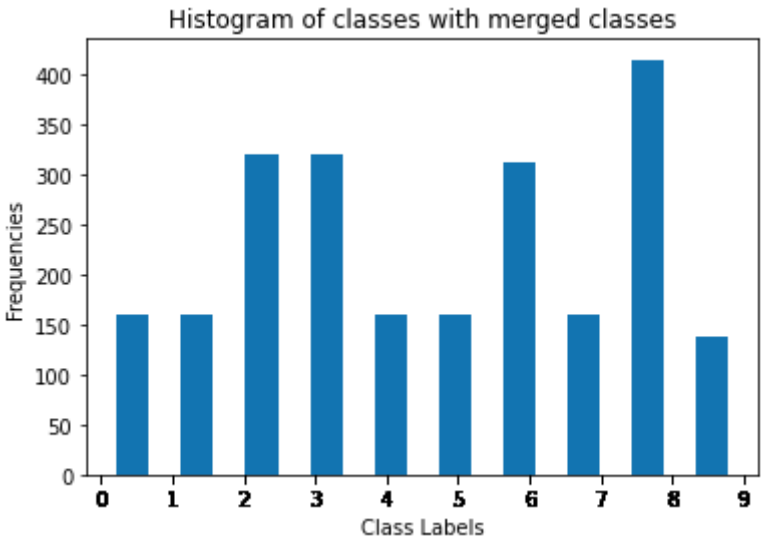
Βήμα 4

α-β) Εισαγάγαμε στο κομμάτι του κώδικα το dictionary που συγχωνεύει τις κλάσεις και την βοηθητική συνάρτηση read_spectrogram στην οποία δίνεται ως όρισμα ένα αρχείο (που έχει ήδη διαβαστεί μέσω Python) και περιέχει τα χαρακτηριστικά, ενώ στην συνέχεια ανάλογα με το όρισμα που της δόθηκε επιστρέφει τα φασματογραφήματα ή τα χρωμογραφήματα του αρχείου. Έπειτα, ενσωματώθηκαν δύο κλάσεις που περιέχουν μετασχηματισμούς, η πρώτη κλάση είναι LabelTransformer όπου εφαρμόζει ή αντιστρέφει έναν μετασχηματισμό χρησιμοποιώντας συναρτήσεις από την κλάση που κληρονομεί και η δεύτερη είναι η PaddingTransform η οποία κάνει padding στην πρώτη διάσταση δισδιάστατων δομών όταν η διάσταση αυτή είναι μικρότερη από ένα ορισμένο μέγιστο ή αφαιρεί τις τελευταίες γραμμές της δομής αν είναι η διάσταση μεγαλύτερη ώστε το τελικό αποτέλεσμα να έχει ακριβώς την διάσταση πρώτη μήκους max_length.

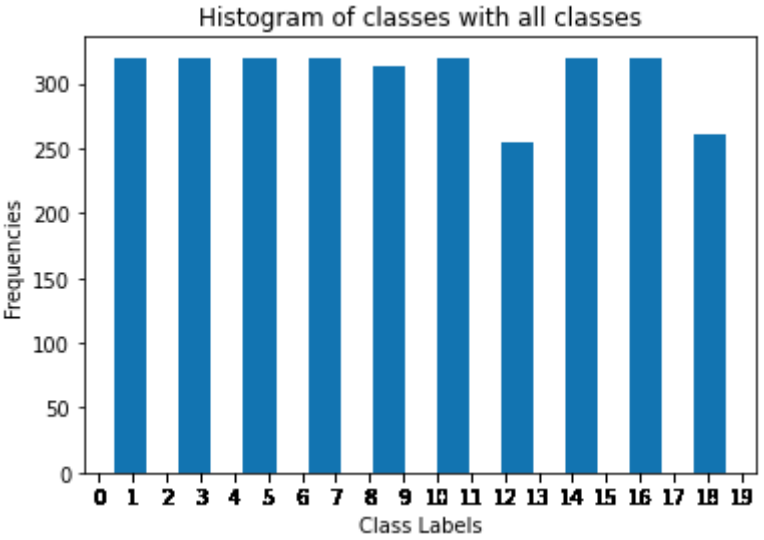
Τέλος, η κλάση που μπορεί να αξιοποιηθεί πλήρως για το συγκεκριμένο dataset είναι η SpectrogramDataset. Δέχεται ως όρισμα κατα την αρχικοποιήση της τον φάκελο που είναι αποθηκευμένα τα δεδομένα, το αν θα δημιουργηθούν δεδομένα για train ή test, τι είδους χαρακτηριστικά θα χρησιμοποιηθούν (mel spectrograms ή chromagrams) το μέγιστο μήκος για padding και αν θα γίνει regression ή όχι. Αφού διαβαστούν τα αρχεία και επιλεχθούν τα δεδομένα με βάση τα ορίσματα, αρχικοποιούνται οι κλάσεις με τους μετασχηματισμούς και μετατρέπονται ανάλογα με την περίπτωση οι ετικέτες με συγκεκριμένο τρόπο σε δομή ndarray. Επιπλέον, υπάρχει η συνάρτηση get_files_labels όπου χρησιμοποιείται στην συνάρτηση αρχικοποίησης της κλάσης και εξάγει τις ετικέτες από τα αρχεία ενώ είναι επίσης υπεύθυνη να χρησιμοποιήσει ή όχι το CLASS_MAPPING dictionary σε περίπτωση που έχει επιλεχθεί συγχώνευση κλάσεων.

γ) Χρησιμοποιώντας την κλάση του Dataset που εισάγαμε στον κώδικα, δημιουργούμε ένα dataset με τα συγχωνευμένα δεδομένα και ένα με τα ασυγχώνευτα δεδομένα. Σε αυτά χρησιμοποιούμε την συνάρτηση plot_class_histogram για να εξάγουμε το ιστόγραμμα κλάσεων και συχνοτητων σε κάθε περίπτωση (Εικόνες 4.1 και 4.2). Παρατηρείται ότι το δεύτερο dataset είναι περισσότερες κλάσεις με περισσότερα δείγματα ανα κλάση.

Επειδή τα datasets που χρησιμοποιήσαμε είναι πολύ μεγάλα και δεσμεύουν χώρο στην RAM προσπαθούμε να τα διαγράψουμε και να καλέσουμε τον garbage collector.



Εικόνα 4.1. *Ιστόγραμμα συγχωνευμένων κλάσεων και συχνοτήτων ανά κλάσης*



Εικόνα 4.2. *Ιστόγραμμα όλων των κλάσεων και συχνοτήτων ανά κλάσης*

Βήμα 5

α) Χρησιμοποιήθηκαν οι υλοποιημένες από το εργαστήριο κλάσεις LSTMBackbone και Classifier. Η πρώτη, υλοποιεί πλήρως τον σκελετό ενός LSTM δικτύου με τις ιδιαιτερότητες του dataset (padding και unpadding ακολουθιών όπου χρειάζεται), υποστηρίζει Dropout που απαιτείται αργότερα και προαιρετικά υλοποίηση Bidirectional LSTM. Η κλάση Classifier ρυθμίζει το αν θα χρησιμοποιηθεί η κλάση σκελετός LSTM ή κάποια άλλη κλάση με διαφορετικό νευρωνικό δίκτυο από επόμενα ερωτήματα.Επίσης, δέχεται ως όρισμα τον αριθμό των κλάσεων και προαιρετικά την επιλογή εκπαίδευσης κάποιου μοντέλου που εκπαιδεύτηκε σε προηγούμενο χρόνο και επιθυμούμε να συνεχίσουμε την εκπαίδευση του.

β) Για το χωρισμό των δεδομένων χρησιμοποιήθηκε η custom συνάρτηση torch_train_val_split όπου χωρίζει τυχαία σε validation και training set. Έπειτα, θέτοντας τον αριθμό των batches ως 8

δημιουργήθηκε dataset με χαρακτηριστικά τα mel spectrogram από τα μη συγχρονισμένα στο beat αρχεία και τις συγχωνευμένες κλάσεις.

Στην συνέχεια υπάρχει η συνάρτηση overfit_with_a_couple_of_batches όπου καλείται για να γίνει η υπερεκπαίδευση ενός μοντέλου χρησιμοποιώντας μόνο το πρώτο batch και εκτυπώνεται το σφάλμα. Έπειτα, υπάρχει μια πρόχειρη συνάρτηση train που χρησιμοποιείται μόνο για το overfiting. Ο εμπλουτισμός αυτής την συνάρτησης θα γίνει στο επόμενο ερώτημα. Με την χρήση αυτών των συναρτήσεων, γίνεται εκπαίδευση ενός Bidirectional LSTM πάνω στο προαναφερθέν dataset. Τα αποτελέσματα του σφάλματος της υπερεκπαίδευσης στο προαναφερθέν dataset φαίνονται στην Εικόνα 5.1. Εφόσον το σφάλμα πέφτει πολύ γρήγορα κοντά στο 0 το δίκτυο είναι ικανό να εκπαιδευτεί και γίνεται σωστά ανανέωση των βαρών.

```
Training in overfitting mode...
Epoch 1, Loss at training set: 2.3302810192108154
Epoch 20, Loss at training set: 2.113520383834839
Epoch 40, Loss at training set: 1.919742465019226
Epoch 60, Loss at training set: 1.6924514770507812
Epoch 80, Loss at training set: 1.4351270198822021
Epoch 100, Loss at training set: 1.1674225330352783
Epoch 120, Loss at training set: 0.9215145111083984
Epoch 140, Loss at training set: 0.7128852605819702
Epoch 160, Loss at training set: 0.48869168758392334
Epoch 180, Loss at training set: 0.3701286315917969
Epoch 200, Loss at training set: 0.25177136063575745
Epoch 220, Loss at training set: 0.17905089259147644
Epoch 240, Loss at training set: 0.14710302650928497
Epoch 260, Loss at training set: 0.10256366431713104
Epoch 280, Loss at training set: 0.082839235663414
Epoch 300, Loss at training set: 0.08865702897310257
Epoch 320, Loss at training set: 0.06941753625869751
Epoch 340, Loss at training set: 0.05226120352745056
Epoch 360, Loss at training set: 0.27159348130226135
Epoch 380, Loss at training set: 0.062395721673965454
Epoch 400, Loss at training set: 0.06358655542135239
```

Εικόνα 5.1. *Το σφάλμα κατά την υπερεκπαίδευση για το μοντέλο γ*

Στο επόμενο κομμάτι του κώδικα μαζί με την εκπαίδευση ομαδοποιήσαμε και την υπερεκπαίδευση όλων των dataset που ζητούνται στα ερωτήματα γ,δ και ε. Αν και οι γραμμές έχουν μπει σε σχόλια, τις εκτελέσαμε και παρουσιάζουμε τα αποτελέσματα στις εικόνες 5.2-5.4.

```
Training in overfitting mode...
Epoch 1, Loss at training set: 2.3852882385253906
Epoch 20, Loss at training set: 2.0546536445617676
Epoch 40, Loss at training set: 1.871077060699463
Epoch 60, Loss at training set: 1.5723834037780762
Epoch 80, Loss at training set: 1.3089864253997803
Epoch 100, Loss at training set: 1.1016018390655518
Epoch 120, Loss at training set: 0.8528724312782288
Epoch 140, Loss at training set: 0.5381694436073303
Epoch 160, Loss at training set: 0.31637588143348694
Epoch 180, Loss at training set: 0.20965516567230225
Epoch 200, Loss at training set: 0.15076398849487305
Epoch 220, Loss at training set: 0.09239865839481354
Epoch 240, Loss at training set: 0.08078336715698242
Epoch 260, Loss at training set: 0.05997972935438156
Epoch 280, Loss at training set: 0.056855708360672
Epoch 300, Loss at training set: 0.046782009303569794
Epoch 320, Loss at training set: 0.03004838153719902
Epoch 340, Loss at training set: 0.033626116812229156
Epoch 360, Loss at training set: 0.03298570588231087
Epoch 380, Loss at training set: 0.022292589768767357
Epoch 400, Loss at training set: 0.024184376001358032
```

Εικόνα 5.2. *Το σφάλμα κατά την υπερεκπαίδευση για το μοντέλο δ*

```
Training in overfitting mode...
Epoch 1, Loss at training set: 2.2962892055511475
Epoch 20, Loss at training set: 2.265810012817383
Epoch 40, Loss at training set: 2.2177212238311768
Epoch 60, Loss at training set: 2.1734232902526855
Epoch 80, Loss at training set: 2.072392463684082
Epoch 100, Loss at training set: 1.9262406826019287
Epoch 120, Loss at training set: 1.7522735595703125
Epoch 140, Loss at training set: 1.6385537385940552
Epoch 160, Loss at training set: 1.1903280019760132
Epoch 180, Loss at training set: 0.7781898975372314
Epoch 200, Loss at training set: 0.8739820718765259
Epoch 220, Loss at training set: 0.6823611259460449
Epoch 240, Loss at training set: 0.6157413721084595
Epoch 260, Loss at training set: 0.48051461577415466
Epoch 280, Loss at training set: 0.4477790296077728
Epoch 300, Loss at training set: 0.38986337184906006
Epoch 320, Loss at training set: 0.33095163106918335
Epoch 340, Loss at training set: 0.2823084592819214
Epoch 360, Loss at training set: 0.28123557567596436
Epoch 380, Loss at training set: 0.2309008538722992
Epoch 400, Loss at training set: 0.19860553741455078
```

Εικόνα 5.3. *Το σφάλμα κατά την υπερεκπαίδευση του μοντέλου του ε ερωτήματος για τα μη συγχρονισμένα στο beat χρωματογραφήματα*

```
Training in overfitting mode...
Epoch 1, Loss at training set: 2.2966761589050293
Epoch 20, Loss at training set: 2.2723429203033447
Epoch 40, Loss at training set: 2.2201170921325684
Epoch 60, Loss at training set: 2.1384377479553223
Epoch 80, Loss at training set: 2.0181117057800293
Epoch 100, Loss at training set: 1.8436330556869507
Epoch 120, Loss at training set: 1.7835557460784912
Epoch 140, Loss at training set: 1.746873378753662
Epoch 160, Loss at training set: 1.6856372356414795
Epoch 180, Loss at training set: 1.4694199562072754
Epoch 200, Loss at training set: 1.2690974473953247
Epoch 220, Loss at training set: 1.1278225183486938
Epoch 240, Loss at training set: 1.0316003561019897
Epoch 260, Loss at training set: 0.8500701189041138
Epoch 280, Loss at training set: 0.779327929019928
Epoch 300, Loss at training set: 0.7316762804985046
Epoch 320, Loss at training set: 0.5502145886421204
Epoch 340, Loss at training set: 0.5306561589241028
Epoch 360, Loss at training set: 0.4302818775177002
Epoch 380, Loss at training set: 0.37336504459381104
Epoch 400, Loss at training set: 0.3026898503303528
```

Εικόνα 5.4. *Το σφάλμα κατά την υπερεκπαίδευση του μοντέλου του ε ερωτήματος για τα συγχρονισμένα στο beat χρωματογραφήματα*

γ-ζ) Για τα επόμενα βήματα υπάρχει η κλάση EarlyStopper ώστε να γίνεται early dropping σε περίπτωση που αυξάνεται το σφάλμα validation, και οι συναρτήσεις train_one_epoch, validate_one_epoch και train όπου υλοποιούν την εκπαίδευση του νευρωνικού σε μια εποχή, τον υπολογισμό του validation σφάλματος και την συνολική εκπαίδευση και αποθήκευση του μοντέλου ενσωματώνοντας όλες τις προηγούμενες συναρτήσεις.

Με την χρήση των συναρτήσεων που έχουμε φτιάξει, εκπαιδεύουμε τα 6 ζητούμενα μοντέλα, για τα φασματογραφήματα, τα χρωματογραφήματα και την ένωση τους και στις δύο περιπτώσεις, δηλαδή τα συγχρονισμένα στο beat και μη συγχρονισμένα. Το batch είναι ρυθμισμένο σε 8 και έχουμε 10 κατηγορίες/κλάσεις. Κάθε μοντέλο είναι ένα Bidirectional LSTM με 2 επίπεδα μεγέθους 128 και learning rate e^{-4}. Έγινε αναζήτηση αλλά δεν βρέθηκαν υπερπαράμετροι με καλύτερα αποτελέσματα στις μετρικές του επόμενο βήματος. Τα αποτελέσματα της εκπαίδευσης φαίνονται στις Εικόνες 5.5-5.10.

```
Training started for model lstm_genre_mel...
Epoch 1/40, Loss at training set: 2.190835605452071
        Loss at validation set: 2.0757748607931465
Epoch 5/40, Loss at training set: 1.9361704438279717
        Loss at validation set: 1.9636204838752747
Epoch 10/40, Loss at training set: 1.8523254750610947
        Loss at validation set: 1.8016868788620521
Epoch 15/40, Loss at training set: 1.7813909456327364
        Loss at validation set: 1.7661850144123208
Epoch 20/40, Loss at training set: 1.866530999992833
        Loss at validation set: 1.825235703895832
Early Stopping was activated.
Epoch 22/40, Loss at training set: 1.8225362115092092
        Loss at validation set: 1.8075841644714619
Training has been completed.
```

Εικόνα 5.5. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του γ ερωτήματος*

```
Training started for model lstm_genre_beat...
Epoch 1/40, Loss at training set: 2.1750208041368624
        Loss at validation set: 2.039030309381156
Epoch 5/40, Loss at training set: 1.8681884307365912
        Loss at validation set: 1.8063695225222358
Epoch 10/40, Loss at training set: 1.767194931383257
        Loss at validation set: 1.7335552963717231
Epoch 15/40, Loss at training set: 1.7538079671013407
        Loss at validation set: 1.7084493205465119
Epoch 20/40, Loss at training set: 1.7034411213614724
        Loss at validation set: 1.7386183224875351
Early Stopping was activated.
Epoch 24/40, Loss at training set: 1.679756003282803
        Loss at validation set: 1.7581992437099587
Training has been completed.
```

Εικόνα 5.6. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του δ ερωτήματος*

```
Training started for model lstm_genre_chroma...
Epoch 1/40, Loss at training set: 2.244691852883343
       Loss at validation set: 2.2132566221829117
Epoch 5/40, Loss at training set: 2.170595138103931
       Loss at validation set: 2.1521840753226447
Epoch 10/40, Loss at training set: 2.1504929308251386
       Loss at validation set: 2.148137429664875
Epoch 15/40, Loss at training set: 2.1420736132246074
       Loss at validation set: 2.1475220672015487
Early Stopping was activated.
Epoch 17/40, Loss at training set: 2.1371061977369963
       Loss at validation set: 2.1499500069124946
Training has been completed.
```

Εικόνα 5.7. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του ε ερωτήματος για μη συγχρονισμένα στο beat χρωματογραφήματα*

```
Training started for model lstm_genre_chroma_beat...
Epoch 1/40, Loss at training set: 2.239639591861081
       Loss at validation set: 2.21094732860039
Epoch 5/40, Loss at training set: 2.160009251528488
       Loss at validation set: 2.172522588022824
Epoch 10/40, Loss at training set: 2.1469125303871186
       Loss at validation set: 2.1578103077822717
Epoch 15/40, Loss at training set: 2.1338405227248285
       Loss at validation set: 2.151394276783384
Early Stopping was activated.
Epoch 18/40, Loss at training set: 2.1257953891506443
       Loss at validation set: 2.1560826794854524
Training has been completed.
```

Εικόνα 5.8. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του ε ερωτήματος για συγχρονισμένα στο beat χρωματογραφήματα*

```
Training started for model lstm_genre_all...
Epoch 1/40, Loss at training set: 2.2134930308247025
        Loss at validation set: 2.124926408817028
Epoch 5/40, Loss at training set: 1.9252869179754546
        Loss at validation set: 1.9428672728867367
Epoch 10/40, Loss at training set: 1.8728836315534847
        Loss at validation set: 1.8388176909808456
Epoch 15/40, Loss at training set: 1.830945205378842
        Loss at validation set: 1.8336986981589218
Epoch 20/40, Loss at training set: 1.7741548003572407
        Loss at validation set: 1.824475664516975
Early Stopping was activated.
Epoch 21/40, Loss at training set: 1.751107793607753
        Loss at validation set: 1.7980439971233237
Training has been completed.
```

Εικόνα 5.9. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του ζ ερωτήματος για μη συγχρονισμένα στο beat χρωματογραφήματα*

```
Training started for model lstm_genre_all_beat...
Epoch 1/40, Loss at training set: 2.1693030941537965
        Loss at validation set: 2.0174036293194213
Epoch 5/40, Loss at training set: 1.8597366845969
        Loss at validation set: 1.848135403518019
Epoch 10/40, Loss at training set: 1.7800299084031737
        Loss at validation set: 1.7579412583647103
Epoch 15/40, Loss at training set: 1.715159102435752
        Loss at validation set: 1.7365560572722862
Early Stopping was activated.
Epoch 16/40, Loss at training set: 1.7220826696007798
        Loss at validation set: 1.7581836036567031
Training has been completed.
```

Εικόνα 5.10. *Το σφάλμα κατά την εκπαίδευση του μοντέλου του ζ ερωτήματος για συγχρονισμένα στο beat χρωματογραφήματα*

Βήμα 6

Δημιουργήθηκε η συνάρτηση test η οποία δέχεται ως είσοδο ένα μοντέλο, ένα dataset, την διαδρομή του αρχείου που αποθηκεύτηκε το εκάστοτε μοντέλο, το μέγεθος του batch και την συσκευή που θα τρέξει το μοντέλο (GPU). Αφού δημιουργηθεί ένα αντικείμενο Dataloader με το δοσμένο dataset, αρχικοποιείται το μοντέλο και φορτώνονται τα βάρη του από το αρχείο. Στην συνέχεια, το μοντέλο επιστρέφει τις προβλεπόμενες τιμές και με βάση τις τιμές των αληθινών ετικετών υπολογίζονται οι μετρικές accuracy, precision, recall, F1-score αλλά και τα macro και micro precision, recall, F1-score για όλες τις κλάσεις.

Η παραπάνω συνάρτηση καλείται 6 φορές μια για κάθε μοντέλο, αφού φορτωθεί το αντίστοιχο dataset του testing και δημιουργηθεί το αρχικό μοντέλο. Τα αποτελέσματα των μετρικών, παρ' όλη την προσπάθεια αναζήτησης βέλτιστων υπερπαραμέτων δεν βελτιώθηκε σημαντικά και φαίνονται στις Εικόνες 6.1 ως 6.6.

For lstm_genre_mel.pth model:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.55	0.42	0.48	52
2	0.60	0.29	0.40	163
3	0.55	0.38	0.45	117
4	0.07	0.33	0.12	9
5	0.05	0.11	0.07	18
6	0.26	0.51	0.34	39
7	0.00	0.00	0.00	0
8	0.54	0.32	0.40	177
9	0.00	0.00	0.00	0
accuracy			0.34	575
macro avg	0.26	0.24	0.23	575
weighted avg	0.52	0.34	0.40	575

Accuracy: 0.3391304347826087

Εικόνα 6.1. *Μετρικές για το μοντέλο με τα μη συγχρονισμένα στο beat φασματογραφήματα*

For lstm_genre_beat.pth model:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.57	0.46	0.51	50
2	0.69	0.30	0.42	182
3	0.53	0.40	0.46	104
4	0.15	0.38	0.21	16
5	0.00	0.00	0.00	6
6	0.62	0.46	0.52	105
7	0.00	0.00	0.00	0
8	0.38	0.35	0.36	112
9	0.00	0.00	0.00	0
accuracy			0.37	575
macro avg	0.29	0.23	0.25	575
weighted avg	0.55	0.37	0.43	575

Accuracy: 0.37043478260869567

Εικόνα 6.2. *Μετρικές για το μοντέλο με τα συγχρονισμένα στο beat φασματογραφήματα*

For lstm_genre_chroma.pth model:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	0
3	0.42	0.21	0.28	165
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
6	0.31	0.32	0.31	76
7	0.00	0.00	0.00	0
8	0.57	0.18	0.27	334
9	0.00	0.00	0.00	0
accuracy			0.20	575
macro avg	0.13	0.07	0.09	575
weighted avg	0.50	0.20	0.28	575

Accuracy: 0.20347826086956522

Εικόνα 6.3. *Μετρικές για το μοντέλο με τα μη συγχρονισμένα στο beat χρωματογραφήματα*

For lstm_genre_chroma_beat.pth model:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	0
3	0.71	0.20	0.31	283
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
6	0.41	0.32	0.36	100
7	0.00	0.00	0.00	0
8	0.34	0.18	0.24	192
9	0.00	0.00	0.00	0
accuracy			0.22	575
macro avg	0.15	0.07	0.09	575
weighted avg	0.54	0.22	0.30	575

Accuracy: 0.21565217391304348

Εικόνα 6.4. *Μετρικές για το μοντέλο με τα συγχρονισμένα στο beat χρωματογραφήματα*

For lstm_genre_all.pth model:

	precision	recall	f1-score	support
0	0.03	0.17	0.04	6
1	0.62	0.44	0.52	57
2	0.34	0.36	0.35	76
3	0.60	0.36	0.45	133
4	0.33	0.27	0.30	48
5	0.10	0.33	0.15	12
6	0.50	0.44	0.47	88
7	0.00	0.00	0.00	4
8	0.41	0.36	0.38	118
9	0.18	0.18	0.18	33
accuracy			0.36	575
macro avg	0.31	0.29	0.28	575
weighted avg	0.45	0.36	0.39	575

Accuracy: 0.3565217391304348

Εικόνα 6.5. *Μετρικές για το μοντέλο με τα μη συγχρονισμένα στο beat χρωματογραφήματα και φασματογραφήματα*

For lstm_genre_all_beat.pth model:

	precision	recall	f1-score	support
0	0.03	0.50	0.05	2
1	0.57	0.43	0.49	54
2	0.76	0.39	0.51	158
3	0.50	0.40	0.44	100
4	0.10	0.33	0.15	12
5	0.07	0.18	0.11	17
6	0.71	0.46	0.56	120
7	0.00	0.00	0.00	0
8	0.40	0.37	0.38	110
9	0.03	0.50	0.06	2
accuracy			0.40	575
macro avg	0.32	0.36	0.27	575
weighted avg	0.58	0.40	0.46	575

Accuracy: 0.3982608695652174

Εικόνα 6.6. *Μετρικές για το μοντέλο με τα μη συγχρονισμένα στο beat χρωματογραφήματα και φασματογραφήματα*

Τα μοντέλο που εκπαιδεύτηκαν με χαρακτηριστικά συγχρονισμένα στο beat φαίνεται να είχαν καλύτερα αποτελέσματα από ότι τα αντίστοιχα μοντέλα με χαρακτηριστικά μη συγχρονισμένα στο beat .

Στην συνέχεια ακολουθεί η θεωρητική απάντηση του συγκεκριμένου ερωτήματος. Όλες τις μετρικές που αναφέρονται χρησιμοποιούνται ως ένας δείκτης για την ορθότητα του μοντέλου. Κάθε μετρική έχει την σκοπιμότητα της και εκτιμά την ορθότητα με διαφορετικό τρόπο. Η ακρίβεια ορίζεται ως ο αριθμός των σωστά ταξινομημένων δεδομένων σε σχέση με τον αριθμό όλων των δεδομένων (Εικόνα 6.7). Ωστόσο δεν είναι καλή μετρική για προβλήματα με μη ισορροπημενα δεδομένα.

Άλλη μετρική που λαμβάνει υπόψη της τα δεδομένα που δεν έχουν ταξινομηθεί σωστά είναι το precision όπου παίρνει τιμή 1 μόνο όταν δεν υπάρχουν FP. Αντίστοιχη μετρική είναι και το recall που ονομάζεται επίσης sensitivity ή TP rate και λαμβάνει τιμή 1 όταν δεν υπάρχουν FN. Αν με βάση την φύση του προβλήματος ταξινόμησης μας ενδιαφέρει η ακρίβεια των Positive (όπως για παράδειγμα διαγνωστικά τεστ σπάνιων/σοβαρών ασθενειών) τότε προτείνεται η χρήση του recall, διαφορετικα αν απαιτείται η ορθή διάγνωση των TP προτείνεται το precision (για παράδειγμα σε διαφημιστικές καμπάνιες που χρειάζεται ορθός εντοπισμός ενδιαφέροντος των χρηστών ώστε να προβληθεί η διαφήμιση).

Σε περίπτωση που μας ενδιαφέρουν να μην υπάρχουν FP και FN, θα ήταν ιδανικός ένας ταξινομητής με precision και recall 1 ταυτόχρονα, γι' αυτό δημιουργήθηκε το F1-score που είναι ο αρμονικός μέσος των δύο προηγούμενων μετρικών και είναι 1 όταν και οι δύο είναι ταυτόχρονα 1. Επιπλέον, χρησιμοποιείται σε μη ισορροπημένα dataset, άρα είναι καλύτερη μετρική σε σχέση με το accuracy σε αυτή την περίπτωση. Γενικά, η απόκλιση των accuracy και F1-score αποδεικνύει πως το dataset δεν είναι ισορροπημένο. Άλλος ένας διαχωρισμός για την χρήση των δύο προαναφερθέντων μετρικών είναι ότι το accuracy χρησιμοποιείται όταν είναι σημαντική η έυρεση TP και TN, ενώ αν τα FP και FN θεωρούνται πιο σημαντικά χρησιμοποιείται το F1-score.

Καταλήγουμε πως είναι σημαντικός ο προσδιορισμός του στόχου ενός ταξινομητή ώστε να επιλεχθεί κατάλληλη μετρική που να βελτιστοποιεί τον αντίστοιχο στόχο και όχι η άκριτη χρήση όλων ή άλλων μετρικών.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Εικόνα 6.7. *Τύπος ακρίβειας*

$$Precision = \frac{TP}{TP + FP}$$

Εικόνα 6.8. *Τύπος precision*

$$Recall = \frac{TP}{TP + FN}$$

Εικόνα 6.9. *Τύπος recall*

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Εικόνα 6.10. *Τύπος F1-score*

Για το macro-average καθενός από τα precision, recall και f1 score, υπολογίζεται ο αριθμητικός μέσος όρος της αντίστοιχης μετρικής για όλες τις κλάσεις χωρίς να λαμβάνεται υπόψη το support αυτής. Χρησιμοποιείται για μη ισορροπημένα dataset ώστε να ελέγξουμε την απόδοση ενός ταξινομητή σε όλες τις κλάσεις.

Από την άλλη μερία, το micro-average χρησιμοποιείται για ισομοιρασμενα δεδομένα ανά κλάση, και έτσι μπορεί να εξηγηθεί η απόκλιση των micro και macro μετρικών. Το micro-average precision είναι το άθροισμα όλων των TP μιας κλάσης διαιρεμένο από το άθροισμα των προβλεπόμενων Positive όλων των κλάσεων. Ανάλογα, το micro-average recall είναι το άθροισμα των TP μιας κλάσης διαιρεμένο με το άθροισμα των TP όλων των κλάσεων. Το ενδιαφέρον έγκειται στο γεγονός ότι για προβλήματα multi-class ταξινόμησης με ένα label το micro-F1-score ισούται με το accuracy, το micro-precision και το micro-recall. Σε περίπτωση που δεν ικανοποιούνται οι προηγούμενες προϋποθέσεις, δεν ισχύει η ισότητα αυτή.

Με βάση την Εικόνα 4.2 γνωρίζουμε ότι τα dataset είναι ισορροπημένο άρα το accuracy και το F1 score και αντίστοιχα οι micro και οι macro average μετρικές θα είναι ενδεικτικές και θα κυμαίνονται σε παρόμοια επίπεδα.

$$\text{PrecisionMacroAvg} = \frac{(Prec_1 + Prec_2 + \ldots + Prec_n)}{n}$$

Εικόνα 6.11. *Τύπος macro-average precision*

$$\text{RecallMacroAvg} = \frac{(Recall_1 + Recall_2 + \ldots + Recall_n)}{n}$$

Εικόνα 6.12. *Τύπος macro-average recall*

$$\text{Macro F1 Score} = \frac{\sum_{i=1}^{n} \text{F1 Score}_i}{n}$$

Εικόνα 6.13. *Τύπος macro-average F1-score*

$$\text{PrecisionMicroAvg} = \frac{(TP_1 + TP_2 + \ldots + TP_n)}{(TP_1 + TP_2 + \ldots + TP_n + FP_1 + FP_2 + \ldots + FP_n)}$$

Εικόνα 6.14. *Τύπος micro-average precision*

$$\text{RecallMicroAvg} = \frac{(TP_1 + TP_2 + \ldots + TP_n)}{(TP_1 + TP_2 + \ldots + TP_n + FN_1 + FN_2 + \ldots + FN_n)}$$

Εικόνα 6.15. *Τύπος micro-average recall*

$$
\begin{aligned}
\text{Micro F1 Score} &= \frac{\text{Net } TP}{\text{Net } TP + \frac{1}{2}(\text{Net } FP + \text{Net } FN)} \\
&= \frac{M_{11} + M_{22}}{M_{11} + M_{22} + \frac{1}{2}\left[(M_{12} + M_{21}) + (M_{21} + M_{12})\right]} \\
&= \frac{M_{11} + M_{22}}{M_{11} + M_{12} + M_{21} + M_{22}} \\
&= \frac{TP + TN}{TP + FP + FN + TN} \\
&= \text{Accuracy}
\end{aligned}
$$

Εικόνα 6.16. *Τύπος micro-average F1-score για 2 κλάσεις*

Βήμα 7

α) Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) ή στα Αγγλικά Convolutional Neural Networks (CNN), χωρίζονται σε δύο μεγάλες κατηγορίες τα Αβαθή Νευρωνικά Δίκτυα (Shallow Neural Networks) και τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks). Ένα συνελικτικό επίπεδο (convolutional layer) είναι ουσιαστικά ένα σύνολο από νευρώνες που εκτελούν συνέλιξη των φίλτρων που έχουν προκαθοριστεί, με την εικόνα-διάνυσμα που δέχονται στην είσοδο. Κάθε επίπεδο μπορεί να περιλαμβάνει νευρώνες που εκτελούν συνέλιξη, διαδικασίες pooling, εισαγωγή μη γραμμικότητας ή ακόμη και κανονικοποίηση, ενώ έχει διακριτές εισόδους και εξόδους. Οι διαστάσεις των φίλτρων που περιλαμβάνουν, ο αριθμός τους και το βάθος τους (αριθμός καναλιών) μπορεί να διαφέρει σημαντικά ανάλογα με το πρόβλημα.

Το convolutional layer είναι το βασικό δομικό στοιχείο ενός CNN, και εκεί λαμβάνει χώρα η πλειοψηφία των υπολογισμών. Απαιτεί μερικά στοιχεία, τα οποία είναι τα δεδομένα εισόδου, ένα φίλτρο και ένας χάρτης χαρακτηριστικών. Πρακτικά σε αυτό, υπολογίζεται η συνέλιξη της εισόδου με κάποιον σταθερό πυρήνα (kernel) μικρότερης διάστασης. Αυτό κατά κάποιον τρόπο αφαιρεί το θόρυβο και ομαλοποιεί τα δεδομένα εισόδου.

Στα pooling layers, γνωστά και ως downsampling, πραγματοποιείται μείωση διαστάσεων, μειώνοντας τον αριθμό των παραμέτρων στην είσοδο. Παρόμοια με το convolutional layer, η λειτουργία pooling σαρώνει ένα φίλτρο σε ολόκληρη την είσοδο, αλλά η διαφορά είναι ότι αυτό το φίλτρο δεν έχει βάρη. Αντίθετα, ο kernel εφαρμόζει μια συνάρτηση συνάθροισης στις τιμές εντός του πεδίου υποδοχής, συμπληρώνοντας τον πίνακα εξόδου.

Υπάρχουν δύο κύριοι τύποι ομαδοποίησης:

- Max pooling: Καθώς το φίλτρο κινείται κατά μήκος της εισόδου, επιλέγει το pixel με τη μέγιστη τιμή για αποστολή στη διάταξη εξόδου. Επιπλέον, αυτή η προσέγγιση τείνει να χρησιμοποιείται πιο συχνά σε σύγκριση με τη average pooling.
- Average pooling: Καθώς το φίλτρο μετακινείται κατά μήκος της εισόδου, υπολογίζει τη μέση τιμή εντός του πεδίου λήψης για αποστολή στον πίνακα εξόδου.

Με αυτόν τον τρόπο μειώνεται η διάσταση της εισόδου, καθώς κάθε γειτονιά δεδομένων αντιπροσωπεύεται πλέον από μια μόνο τιμή. Γεωμετρικά, θα μπορούσε κανείς να πει πως, ενώ ένα convolutional layer ομαλοποιεί την εικόνα, ένα pooling layer κρατά ένα πιο πρόχειρο περίγραμμα της

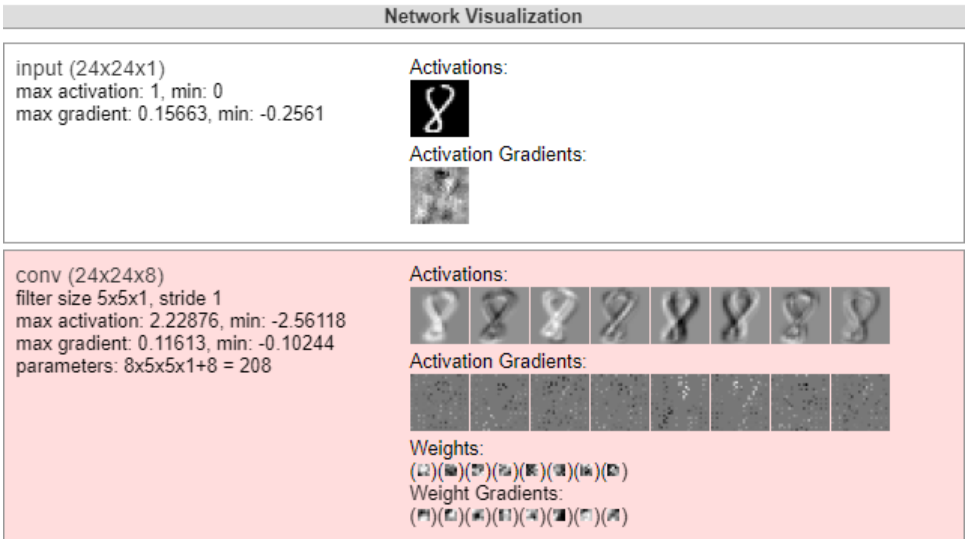
που βασίζεται στα πιο αντιπροσωπευτικά features της, ενώ παράλληλα μειώνει σημαντικά τη διάσταση της.

Ένα από τα σημαντικότερα αίτια για την επιτυχία των CNN στην αναγνώριση εικόνας εντοπίζεται στο γεγονός πως τα συνελικτικα επίπεδα τους είναι αναλλοίωτα στις μεταφορές. Έτσι, γεωμετρικά χαρακτηριστικά της εικόνας, όπως γωνίες, πλευρές, κ. α., τα οποία εμφανίζονται στην είσοδο, παραμένουν αναλλοίωτα και κατά την έξοδο τους από το επίπεδο. Το ίδιο ισχύει και για τα επίπεδα του pooling, αλλά μόνο τοπικά. Παράλληλα, έχουν σημαντικά λιγότερα βάρη, και επομένως πιο εύκολη εκπαίδευση σε σχέση με ένα πλήρως συνεκτικό δίκτυο αντίστοιχου μεγέθους.

Στα υπόλοιπα δομικά στοιχεία ενός CNN συμπεριλαμβάνονται οι συναρτήσεις ενεργοποίησης που δρουν στην έξοδο των συνελικτικων και pooling επίπεδων. Η πιο συνηθισμένη επιλογή είναι η Rectified Linear Unit (ReLU), $f(x) = \max\{0, x\}$. Το πλεονέκτημα της έναντι των σιγμοειδών συναρτήσεων είναι ο τι αποφεύγεται ο κορεσμός, καθω ς $\lim_{x\to\infty} f(x) = +\infty$, το οποίο έχει ως αποτέλεσμα τα δίκτυα να εκπαιδεύονται πολύ γρηγορότερα.

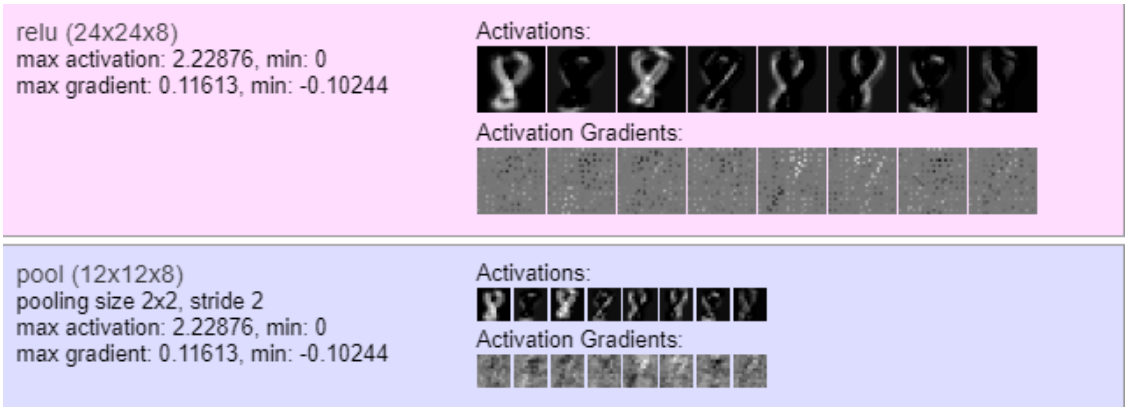
Μια ακόμα συνηθισμένη πρακτική είναι αυτή του batch normalization, κατά την οποία τα δεδομένα κανονικοποιούνται προτού εφαρμοστεί σε αυτά η συνάρτηση του κάθε επίπεδου. Ο λόγος είναι ότι, ιδίως σε βαθιά δίκτυα, μικρές αλλαγές στα βάρη των υψηλότερων στρώμα των μπορεί να δημιουργήσουν μεγάλες αλλαγές σε αυτά των υπολοίπων, αλλάζοντας επίσης και τις αντίστοιχες κατανομές. Η κανονικοποίηση εξασφαλίζει ότι τα βάρη εξακολουθούν να βλέπουν δεδομένα ίδιας τάξης μεγέθους με αυτά που έβλεπαν πριν την ανανέωση τους. Τέλος, στο προτελευταίο επίπεδο συμπεριλαμβάνεται ένα fully connected layer, το οποίο συνδυάζει τις τοπικές πληροφορίες των προηγούμενων επίπεδων για να δώσει την τελική πρόβλεψη για ολόκληρη την αρχική εικόνα.

Στη σελίδα ConvNetJS δόθηκε η δυνατότητα εκπαίδευσης ενός CNN στα δεδομένα του dataset MNIST, τα οποία περιλαμβάνουν εικόνες ψηφίων από το 0 έως το 9. Παρατίθενται ακολούθως εικόνες, στις περιγραφές των οποίων αναλύονται τα δομικά στοιχεία του CNN που χρησιμοποιήθηκε και το πως αυτά επεξεργάζονται, τροποποιούν και εν τέλει ταξινομούν ένα δεδομένο εισόδου.



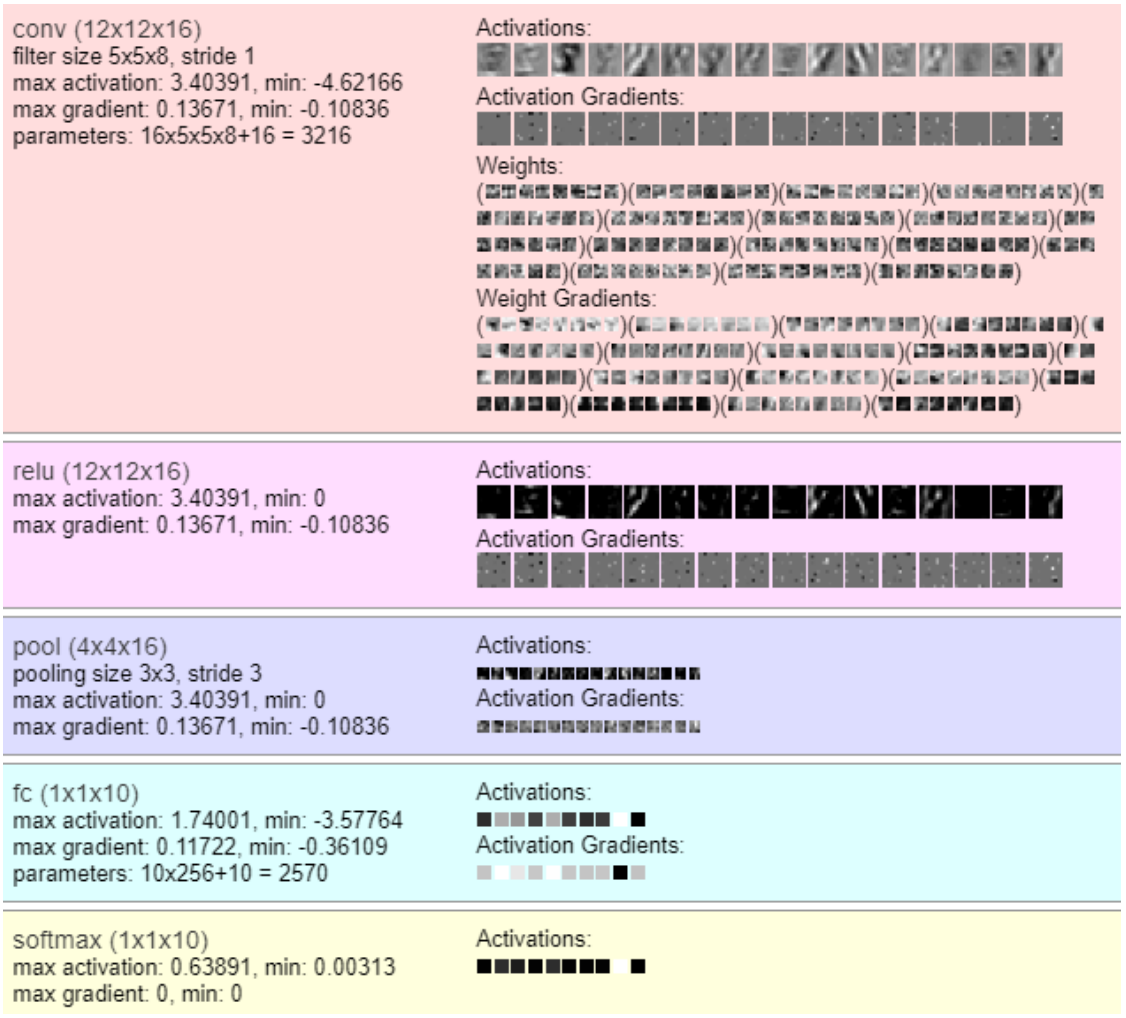
Εικόνα 7.1. *Οπτικοποίηση του δικτύου (α μέρος)*

Το δίκτυο παίρνει μια εικόνα MNIST 28x28 και περικόπτει ένα τυχαίο παράθυρο 24x24 πριν από την εκπαίδευση σε αυτό (αυτή η τεχνική ονομάζεται data augmentation και βελτιώνει τη γενίκευση). Στο πρώτο συνελικτικό επίπεδο τα γεωμετρικά χαρακτηριστικά του ψηφίου 8 διατηρήθηκαν και ταυτόχρονα ομαλοποιήθηκε και το περίγραμμά του.



Εικόνα 7.2. *Οπτικοποίηση του δικτύου (β μέρος)*

Έπειτα εφαρμόστηκε στην έξοδο του convolutional layer η συνάρτηση ενεργοποίησης ReLU η οποία διατήρησε όσες τιμές βρίσκονταν πάνω από ένα ορισμένο κατώφλι και μηδένισε όλες τις υπόλοιπες κάτωθεν του. Συνεπώς ορισμένα χαρακτηριστικά του ψηφίου διατηρήθηκαν ενώ άλλα χάθηκαν στο μαύρο φόντο. Στο επόμενο στάδιο το pooling layer, εφαρμόστηκε ένα παράθυρο 2x2 με stride 2, επομένως συνολικά έγινε λόγος για 12x12 παράθυρα. Ως αποτέλεσμα αυτής της εφαρμογής επήλθε η μείωση της διάστασης της εικόνας εξόδου αλλά και η διατήρηση των χαρακτηριστικών του ψηφίου.



Εικόνα 7.3. *Οπτικοποίηση του δικτύου (γ μέρος)*

Τέλος, προχωρώντας στην εφαρμογή και άλλων layers οι διαστάσεις προοδευτικά γίνονταν μικρότερες με αποτέλεσμα να χάνεται αρκετή πληροφορία σχετικά με τα γεωμετρικά χαρακτηριστικά του ψηφίου. Στο στάδιο FC, το δίκτυο συνδύαζε όλες τις προηγούμενες πληροφορίες για την κατάταξη του ψηφίου σε μία από τις δέκα κλάσεις (0-9) ενώ στην έξοδο softmax τελικά επιλεγόταν η κλάση με τη μεγαλύτερη πιθανότητα.

β) Για την ανάπτυξη συνελικτικού δικτύου, δημιουργήθηκε η κλάση CNNbackbone, η οποία αποτελεί τον σκελετό για ένα συνελικτικό δίκτυο με 4 επίπεδα, όπου κάθε επίπεδο έχει 2D convolution, batch normalization, συνάρτηση ενεργοποίησης ReLU και max pooling.

γ) Η παρακάτω επεξήγηση περιλαμβάνει παραδείγματα από CNN που δέχονται ως είσοδο εικόνες. Ωστόσο, όλα τα παρακάτω μπορούν να γενικευτούν και σε άλλα σήματα όπως τα φασματογραφήματα που χρησιμοποιούνται στην παρούσα την άσκηση. Η χρήση αυτών των παραδειγμάτων είναι χρήσιμη ώστε να είναι πιο περιγραφική η ανάλυση και πιο εύκολη στην κατανόηση.

Ο συνελικτικός τελεστής είναι βοηθητικός στην εξαγωγή χαρακτηριστικών μιας εικόνας για παράδειγμα, γιατί διατηρεί τις χωρικές σχέσεις των στοιχείων χρησιμοποιώντας φίλτρα. Τα φίλτρα ή kernels εφαρμόζονται σε καθένα από τα επικαλυπτόμενα τετράγωνα κομμάτια που χωρίζεται η εικόνας. Κάθε διαφορετικό φίλτρο, έχει διαφορετική λειτουργία. Τα πιο γνωστά φίλτρα που χρησιμοποιήθηκαν στο πεδίο της Αναλυσης Εικόνες έχουν ικανότητα να εντοπίσουν ακμές, να θολώσουν την Εικόνα με συγκεκριμένο τρόπο ή να την κάνουν πιο “αιχμηρή”.

Το βάθος του συνελικτικού τελεστή ορίζει το πόσα φίλτρα θα εφαρμοστούν και άρα πόσα αποτελέσματα (Feature Map) θα παραχθούν. Στην συγκεκριμένη περίπτωση επιλέχθηκε το βάθος 2. Το stride καθορίζει το πόσα pixel δεν θα είναι επικαλυπτόμενα από κάθε εφαρμογή kernel στα διαφορετικά κομμάτια που χωρίζεται η εικόνα.Η τελευταία παράμετρος του συνελικτικού τελεστή, το zero-padding, ορίζει το πόσα 0 θα προστεθούν στις άκρες του kernel και έτσι μπορεί να καθοριστεί και το μέγεθος του Feature Map.

Όπως και σε άλλα είδη νευρωνικών δικτύων, η συνάρτηση ενεργοποίησης ReLU εισάγει την μη γραμμικότητα ώστε να υπάρχει μεγαλύτερη ελευθερία μάθησης και εφαρμόζεται σε όλα τα στοιχεία του Feature Map. Για το pooling υπάρχουν 3 λειτουργίες, το max, το average και το sum όπου είναι μια μέθοδος διατήρησης της χρήσιμης πληροφορίας με μείωση των διαστάσεων. Το Feature Map μετά την εφαρμογή της ReLU, χωρίζεται σε ίσα τετραγωνικά κομμάτια χωρίς επικάλυψη και σε καθένα από αυτά ορίζεται η αντίστοιχη πράξη εύρεσης μέγιστου, μέσου ή αθροίσματος ώστε πλέον κάθε τετραγωνικό κομμάτι αναπαρίσταται με την scalar τιμή του αποτελέσματος της πράξης.

δ) Όπως πραγματοποιήθηκε και στο Βήμα 5, εκπαιδεύτηκε μέσω overfitting ένα νέο 2D CNN στα φασματογραφήματα των μουσικών κομματιών, για να ελεγχθεί πως το νευρωνικό δίκτυο αυτό είναι εκπαιδεύσιμο. Το αποτέλεσμα φαίνεται στην Εικόνα 7.4

```
Training in overfitting mode...
Epoch 1, Loss at training set: 2.4744949340820312
Epoch 20, Loss at training set: 6.350982584990561e-05
Epoch 40, Loss at training set: 5.250071990303695e-05
Epoch 60, Loss at training set: 2.656596916494891e-05
Epoch 80, Loss at training set: 1.525787047285121e-05
Epoch 100, Loss at training set: 1.047508976625977e-05
Epoch 120, Loss at training set: 7.912275577837136e-06
Epoch 140, Loss at training set: 6.317940915323561e-06
Epoch 160, Loss at training set: 5.230204806139227e-06
Epoch 180, Loss at training set: 4.425572569743963e-06
Epoch 200, Loss at training set: 3.844446382572642e-06
Epoch 220, Loss at training set: 3.367622184669017e-06
Epoch 240, Loss at training set: 2.9951020223961677e-06
Epoch 260, Loss at training set: 2.71198655354965e-06
Epoch 280, Loss at training set: 2.443770654281252e-06
Epoch 300, Loss at training set: 2.235158035546192e-06
Epoch 320, Loss at training set: 2.056347057077801e-06
Epoch 340, Loss at training set: 1.9073373778155656e-06
Epoch 360, Loss at training set: 1.7732286323735025e-06
Epoch 380, Loss at training set: 1.6689218682586215e-06
Epoch 400, Loss at training set: 1.5795161516507505e-06
```

Εικόνα 7.4. *Το σφάλμα κατά την υπερεκπαίδευση του CNN μοντέλου για τα μη συγχρονισμένα στο beat χρωματογραφήματα*

ε) Παρομοίως με το Βήμα 5, εκπαιδεύτηκαν δύο 2D CNN χρησιμοποιώντας ως χαρακτηριστικά, τα φασματογραφήματα των μουσικών κομματιών που ήταν συγχρονισμένα στο beat και σε αυτά που δεν ήταν συγχρονισμένα. Το σφάλμα εκπαίδευσης παρουσιάζεται στην Εικόνα 7.5, ενώ στην Εικόνα 7.6 φαίνονται τα αποτελέσματα των μετρικών στα μοντέλα.

```
Training started for model cnn_genre_mel...
Epoch 1/40, Loss at training set: 3.1370683707200087
        Loss at validation set: 2.622793127750528
Epoch 5/40, Loss at training set: 1.6175110275611218
        Loss at validation set: 2.0287502212771056
Epoch 10/40, Loss at training set: 0.8888490358730415
        Loss at validation set: 2.3312322302111266
Early Stopping was activated.
Epoch 11/40, Loss at training set: 0.6961749877248492
        Loss at validation set: 2.5517498295882652
Training has been completed.

Training started for model cnn_genre_beat...
Epoch 1/40, Loss at training set: 2.914639550886113
        Loss at validation set: 1.954826163834539
Epoch 5/40, Loss at training set: 1.5074269025872795
        Loss at validation set: 1.9026941975642895
Epoch 10/40, Loss at training set: 0.8992608470189107
        Loss at validation set: 2.202688192499095
Early Stopping was activated.
Epoch 11/40, Loss at training set: 0.6592061949240697
        Loss at validation set: 2.4111322483112074
Training has been completed.
```

Εικόνα 7.5. *Το σφάλμα κατά την εκπαίδευση των δύο CNN μοντέλων*

```
For cnn_genre_mel.pth model:
              precision    recall  f1-score   support

           0       0.57      0.18      0.28       127
           1       0.70      0.44      0.54        63
           2       0.26      0.84      0.40        25
           3       0.42      0.41      0.42        82
           4       0.72      0.35      0.48        82
           5       0.07      0.75      0.14         4
           6       0.47      0.54      0.50        69
           7       0.00      0.00      0.00         0
           8       0.34      0.41      0.37        85
           9       0.29      0.26      0.28        38

    accuracy                           0.38       575
   macro avg       0.39      0.42      0.34       575
weighted avg       0.51      0.38      0.40       575

Accuracy: 0.3826086956521739

For cnn_genre_beat.pth model:
              precision    recall  f1-score   support

           0       0.07      0.33      0.12         9
           1       0.72      0.43      0.54        68
           2       0.36      0.63      0.46        46
           3       0.56      0.47      0.51        96
           4       0.55      0.28      0.37        79
           5       0.00      0.00      0.00         0
           6       0.73      0.46      0.57       123
           7       0.10      0.29      0.15        14
           8       0.41      0.39      0.40       107
           9       0.29      0.30      0.30        33

    accuracy                           0.42       575
   macro avg       0.38      0.36      0.34       575
weighted avg       0.54      0.42      0.45       575

Accuracy: 0.4191304347826087
```

Εικόνα 7.6. *Μετρικές ορθότητας των δύο CNN μοντέλων*

Το μοντέλο που εκπαιδεύτηκε σε φασματογραφήματα συγχρονισμένα στο beat είχε καλύτερα αποτελέσματα από ότι το άλλο μοντέλο. Επιπλέον, αξίζει να σημειωθεί ότι τα CNN μοντέλα είχαν καλύτερη απόδοση από ότι τα αντίστοιχα LSTM μοντέλα.

Βήμα 8

α) Δημιουργήθηκε η κλάση Regressor που λειτουργεί ακριβώς όπως και η συνάρτηση Classifier με την διαφορά ότι η νέα κλάση έχει συνάρτηση κόστους το μέσο τετραγωνικό σφάλμα. Με την χρήση αυτής της κλάσης, δίνεται η δυνατότητα να χρησιμοποιηθούν οι κλάσεις LSTBackbone και CNNBackbone (άρα και τα αντίστοιχα μοντέλα) όπως και προηγουμένως.

Επειδή το νέο dataset δεν έχει ετικέτες για τα test δεδομένα, δημιουργήθηκε η συνάρτηση torch_train_val_test_split, η οποία χωρίζει το dataset σε 3 κατηγορίες (train, validation και test set) με την ίδια τεχνική που αξιοποιήθηκε και στην συνάρτηση torch_train_val_split.

β-δ) Για κάθε έναν από τους συναισθηματικούς άξονες του dataset, εκπαιδεύτηκε ένα Bidirectional LSTM μοντέλο χρησιμοποιώντας όλα τα χαρακτηριστικά ως features, δηλαδή και τα φασματογραφήματα και τα χρωμογραφήματα που ήταν συγχρονισμένα στο beat, αφού αυτό ήταν το καλύτερο μοντέλο του Βήματος 5. Αναλόγως, εκπαιδεύεται ένα CNN με χαρακτηριστικά τα φασματογραφήματα που ήταν συγχρονισμένα στο beat ως το καλύτερο από τα δύο μοντέλα του βήματος 7. Ενδεικτικά παρουσιάζουμε το στάδιο του overfitting για έλεγχο της εκπαιδευσιμότητας των μοντέλων στις Εικόνες 8.1 και 8.2.

```
Training in overfitting mode...
Epoch 1, Loss at training set: 0.21558672189712524
Epoch 20, Loss at training set: 0.035828765481710434
Epoch 40, Loss at training set: 0.008130584843456745
Epoch 60, Loss at training set: 0.002751801162958145
Epoch 80, Loss at training set: 0.0020773070864379406
Epoch 100, Loss at training set: 0.0025998682249337435
Epoch 120, Loss at training set: 0.001137365004979074
Epoch 140, Loss at training set: 0.0014442871324717999
Epoch 160, Loss at training set: 0.002154476474970579
Epoch 180, Loss at training set: 0.0020941554103046656
Epoch 200, Loss at training set: 0.004840388428419828
Epoch 220, Loss at training set: 0.0023582472931593657
Epoch 240, Loss at training set: 0.0017350269481539726
Epoch 260, Loss at training set: 0.0025863582268357277
Epoch 280, Loss at training set: 0.0018833677750080824
Epoch 300, Loss at training set: 0.0012831545900553465
Epoch 320, Loss at training set: 0.0009921093005686998
Epoch 340, Loss at training set: 0.002015325939282775
Epoch 360, Loss at training set: 0.002114384202286601
Epoch 380, Loss at training set: 0.001071615144610405
Epoch 400, Loss at training set: 0.001925529446452856
```

Εικόνα 8.1. *Το σφάλμα κατά την υπερεκπαίδευση του LSTM μοντέλου για παλινδρόμηση*

```
Training in overfitting mode...
Epoch 1, Loss at training set: 0.12742477655410767
Epoch 20, Loss at training set: 1.0704753398895264
Epoch 40, Loss at training set: 0.5594513416290283
Epoch 60, Loss at training set: 0.28035375475883484
Epoch 80, Loss at training set: 0.000706303573679179
Epoch 100, Loss at training set: 0.002665699692443013
Epoch 120, Loss at training set: 0.00037997501203790307
Epoch 140, Loss at training set: 9.660288924351335e-06
Epoch 160, Loss at training set: 4.464939138415502e-06
Epoch 180, Loss at training set: 7.284382945726975e-07
Epoch 200, Loss at training set: 1.36079520984822e-07
Epoch 220, Loss at training set: 4.309963053117372e-09
Epoch 240, Loss at training set: 6.929855300707999e-11
Epoch 260, Loss at training set: 5.718530510234743e-11
Epoch 280, Loss at training set: 9.505937703657708e-13
Epoch 300, Loss at training set: 8.999676004428636e-13
Epoch 320, Loss at training set: 1.722164078010735e-13
Epoch 340, Loss at training set: 1.7284090825242515e-13
Epoch 360, Loss at training set: 8.0629947163402e-14
Epoch 380, Loss at training set: 9.688777558025663e-14
Epoch 400, Loss at training set: 7.800704526772506e-14
```

Εικόνα 8.2. *Το σφάλμα κατά την υπερεκπαίδευση του CNN μοντέλου για παλινδρόμηση*

Έπειτα παρουσιάζεται το σφάλμα κατα την διάρκεια της εκπαίδευσης για κάθε συναισθηματικό άξονα
στις Εικόνες 8.3 και 8.4 για κάθε είδος μοντέλου.

```
Training started for model reg_1_lstm_genre_all_beat...
Epoch 1/40, Loss at training set: 0.07832193683994854
        Loss at validation set: 0.06780776152243981
Epoch 5/40, Loss at training set: 0.06341507162736809
        Loss at validation set: 0.06864608423067974
Early Stopping was activated.
Epoch 6/40, Loss at training set: 0.062072342399345795
        Loss at validation set: 0.07045437653477375
Training has been completed.

Training started for model reg_2_lstm_genre_all_beat...
Epoch 1/40, Loss at training set: 0.07558812674782846
        Loss at validation set: 0.04420183419894714
Epoch 5/40, Loss at training set: 0.03993029730475467
        Loss at validation set: 0.04038866605752936
Epoch 10/40, Loss at training set: 0.033835415274876615
        Loss at validation set: 0.057888652986058824
Epoch 15/40, Loss at training set: 0.030011464011571978
        Loss at validation set: 0.03681051308432451
Early Stopping was activated.
Epoch 18/40, Loss at training set: 0.026536732976851257
        Loss at validation set: 0.040720417916488186
Training has been completed.

Training started for model reg_3_lstm_genre_all_beat...
Epoch 1/40, Loss at training set: 0.04303144461920728
        Loss at validation set: 0.03272691125480028
Epoch 5/40, Loss at training set: 0.0252283644295581
        Loss at validation set: 0.028841457723711546
Early Stopping was activated.
Epoch 8/40, Loss at training set: 0.024474246301890715
        Loss at validation set: 0.031165581536837496
Training has been completed.
```

Εικόνα 8.3. *Το σφάλμα κατά την εκπαίδευση του LSTM μοντέλου για παλινδρόμηση*

```
Training started for model reg_1_cnn_genre_mel_beat...
Epoch 1/40, Loss at training set: 11.50638637419628
        Loss at validation set: 0.10018078567316899
Epoch 5/40, Loss at training set: 0.31629398824728056
        Loss at validation set: 0.2331665499995534
Early Stopping was activated.
Epoch 8/40, Loss at training set: 0.20618481388234575
        Loss at validation set: 0.08797052753372835
Training has been completed.

Training started for model reg_2_cnn_genre_mel_beat...
Epoch 1/40, Loss at training set: 12.14903806000948
        Loss at validation set: 0.12122594292920369
Epoch 5/40, Loss at training set: 0.09844934686856426
        Loss at validation set: 0.07191487793953946
Epoch 10/40, Loss at training set: 0.10021391380578279
        Loss at validation set: 0.1237283693268322
Early Stopping was activated.
Epoch 13/40, Loss at training set: 0.14090749065837135
        Loss at validation set: 0.052684950158716395
Training has been completed.

Training started for model reg_3_cnn_genre_mel_beat...
Epoch 1/40, Loss at training set: 12.845665124366464
        Loss at validation set: 0.11087952754818477
Epoch 5/40, Loss at training set: 0.07221368711198801
        Loss at validation set: 0.04997336131054908
Epoch 10/40, Loss at training set: 0.08505855686557681
        Loss at validation set: 0.05636154455490983
Epoch 15/40, Loss at training set: 0.19376833074442718
        Loss at validation set: 0.16174177309641471
Early Stopping was activated.
Epoch 17/40, Loss at training set: 0.164735953623186
        Loss at validation set: 0.08000605825621349
Training has been completed.
```

Εικόνα 8.4. *Το σφάλμα κατά την εκπαίδευση του CNN μοντέλου για παλινδρόμηση*

ε) Για να είναι εφικτή η αξιολόγηση του μοντέλου, παράχθηκε μια νέα συνάρτηση, η regression_test η οποία έχει παρόμοιο τρόπο λειτουργίας με την συνάρτηση test, με την διαφορά ότι είναι κατάλληλη για παλινδρόμηση αφού ελέγχει την ορθότητα με βάση το Spearman Correlation. Τα αποτελέσματα (Εικόνες 8.5 και 8.6) υποδεικνύουν ότι το LSTM μοντέλο είχε καλύτερη απόδοση και πως ο συναισθηματικός άξονας που έχει την μεγαλύτερη ακρίβεια είναι η ενέργεια (energy).

```
For reg_1_lstm_genre_all_beat.pth model:
tensor(0.2952, dtype=torch.float64)
Spearman Correlation: 0.2952411885939081

For reg_2_lstm_genre_all_beat.pth model:
tensor(0.6871, dtype=torch.float64)
Spearman Correlation: 0.6870695609602532

For reg_3_lstm_genre_all_beat.pth model:
tensor(0.3456, dtype=torch.float64)
Spearman Correlation: 0.34556648951114405
```

Εικόνα 8.5. *Ορθότητα του LSTM μοντέλου για παλινδρόμηση*

```
For reg_1_cnn_genre_mel_beat.pth model:
tensor(0.2983, dtype=torch.float64)
Spearman Correlation: 0.29834197705149534

For reg_2_cnn_genre_mel_beat.pth model:
tensor(0.6178, dtype=torch.float64)
Spearman Correlation: 0.6178056576488485

For reg_3_cnn_genre_mel_beat.pth model:
tensor(0.2475, dtype=torch.float64)
Spearman Correlation: 0.24754037959901837
```

Εικόνα 8.6. *Ορθότητα του CNN μοντέλου για παλινδρόμηση*

Βήμα 9

α) Οι ερευνητές στο συγκεκριμένο άρθρο διερεύνησαν το πως επηρεάζεται η δυνατότητα χρήσης transfer learning από την δυσκολία βελτιστοποίησης που προκύπτει απο τον σχετικά εύθραυστο διαχωρισμό του δικτύου και την εξειδίκευση των στρωμάτων του δικτύου που αφορούν το αρχικό task σε βάρος της επίδοσης στο τελικό task. Παρατήρησαν πως οποιοδήποτε από αυτά τα σενάρια μπορεί να επηρεάσει το δίκτυο σε συνάρτηση με το επίπεδο που θα γίνει το fine-tuning. Μια ακόμα περίπτωση που ερευνήθηκε είναι το πόσο δύσκολο γίνεται να χρησιμοποιηθεί Transfer Learning όταν τα δύο tasks είναι όλο και περισσότερο ασυσχέτιστα εννοιολογικά κυρίως για το fine-tuning των υψηλότερων επίπεδων του δικτύου. Τέλος, παρατηρήθηκε ότι ακόμα και στην τελευταία περίπτωση, τα αποτελέσματα ήταν καλύτερα από τυχαία αρχικοποίηση βαρών και πως αυτό μπορεί να βελτιώσει την επίδοση των βαθιών νευρωνικών δικτύων.

β-ε) Επιλέχθηκε το μοντέλο του Βήματος 7 (CNN) καθώς έχουν καλύτερη επίδοση στο αρχικό task, δηλαδή είχε μεγαλύτερη ακρίβεια (ισορροπημένο dataset), οπότε θα αρχικοποιηθεί σε βελτιστοποιημένο επίπεδο και στην συνέχεια θα γίνει fine-tuning οπότε αναμένονται καλύτερα αποτελέσματα. Επιπλέον, επιλεχθηκε να εκπαιδευτεί στα συγχρονισμένα στο beat δεδομένα, καθώς είχαν την καλύτερη ακρίβεια, αλλά και να εκπαιδευτεί σε όλα τα χαρακτηριστικά του dataset (φασματογραφήματα και χρωμογραφήματα) για να έχει γενικότερη γνώση το μοντέλο.

Πράγματι, το μοντέλο έγινε fine-tune με αλλαγή στου τελευταίου γραμμικού επιπέδου και επανεκπαίδευση όλων των βαρών πάνω στο συναισθηματικό άξονα valence. Τα αποτελέσματα είναι ενθαρρυντικά, καθώς επιβεβαιώνεται η θεωρία που αναφέρθηκε παραπάνω και φαίνεται να έχει καλή επιλογή μοντέλου.

```
Training started for model finetuned_model...
Epoch 1/40, Loss at training set: 0.3690227462545685
        Loss at validation set: 0.05276664715403548
Epoch 5/40, Loss at training set: 0.04872777058702448
        Loss at validation set: 0.07742467651573512
Epoch 10/40, Loss at training set: 0.03372867624882771
        Loss at validation set: 0.05792255370089641
Early Stopping was activated.
Epoch 12/40, Loss at training set: 0.02149735496926081
        Loss at validation set: 0.05203865303729589
Training has been completed.
```

Εικόνα 9.1. *Το σφάλμα κατά το fine-tuning του καλύτερου μοντέλου*

```
For finetuned_model.pth model:
Spearman Correlation: 0.5197280223828704
```

Εικόνα 9.2. *Ορθότητα του fine-tuned μοντέλου για valence*

ΒΙΒΛΙΟΓΡΑΦΙΑ

1) https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53
2) https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd
3) https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f
4) https://www.educative.io/answers/what-is-the-difference-between-micro-and-macro-averaging
5) https://www.v7labs.com/blog/f1-score-guide
6) https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2
7) https://www.ibm.com/topics/convolutional-neural-networks
8) https://nemertes.library.upatras.gr/server/api/core/bitstreams/d88e631b-b01e-4888-8b37-eeff37e65b6e/content