



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ

**Εκ νέου επισήμανση με βάση την σχετικότητα
διαφορετικών ετικετών:
Βασικά συναισθήματα και ετικέτες
Σθένους-Διέγερσης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΑΠΑΡΟΥ ΑΛΕΞΑΝΔΡΑΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Αθήνα, Οκτώβριος 2022



**Εκ νέου επισήμανση με βάση την σχετικότητα
διαφορετικών ετικετών: Βασικά συναισθήματα και
ετικέτες Σθένους-Διέγερσης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΚΑΠΑΡΟΥ ΑΛΕΞΑΝΔΡΑΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Οκτωβρίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Στέφανος Κόλλιας
Καθηγητής

.....
Γιώργος Στάμου
Καθηγητής

.....
Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής



Copyright © - All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
Αλεξάνδρα Καπαρού, 2022.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Αλεξάνδρα Καπαρού

18 Οκτωβρίου 2022

Περίληψη

Η παρούσα διπλωματική εργασία ασχολείται με το θέμα της Αναγνώρισης Συναισθήματος και της Ανίχνευσης Σθένους/Διέγερσης χρησιμοποιώντας μεθόδους Μηχανικής Μάθησης και εξάγοντας πληροφορίες από εικόνες προσώπων. Το πεδίο της ανάλυσης συναισθημάτων είναι ένα πολύ ενεργό ερευνητικό πεδίο με μεγάλο αριθμό εφαρμογών ακόμα και στην καθημερινή μας ζωή.

Προσεγγίσαμε αυτό το πρόβλημα κάνοντας χρήση Συνελκτικών Νευρωνικών Δικτύων εκμεταλλευόμενοι επίσης την μέθοδο της επαύξησης δεδομένων προκειμένου να έχουμε μεγαλύτερο αριθμό δεδομένων. Η μεθοδολογία που υλοποιήσαμε κάνει σύζευξη δύο προβλημάτων, ένα ταξινόμησης και ένα παλινδρόμησης, όπου εκμεταλλεύεται τις ετικέτες σθένους και διέγερσης προκειμένου να οδηγήσει σε καλύτερη απόδοση στο πρόβλημα της ταξινόμησης.

Οι ετικέτες παρέχονται από το σύνολο δεδομένων που χρησιμοποιήσαμε, το Affectnet, το οποίο αποτελείται από έναν μεγάλο αριθμό εικόνων που απεικονίζουν εκφράσεις ανθρώπων και έχουν επισημανθεί αναφορικά με το συναίσθημα, το σθένος και την διέγερση. Σχετικά με τις ετικέτες συναισθήματος, αυτές ανήκουν σε μία εκ των 8 διακριτών κατηγοριών: Neutral, Happy, Sad, Surprise, Fear, Disgust, Anger, Contempt ενώ οι ετικέτες σθένους και διέγερσης απαρτίζονται από δεκαδικούς αριθμούς στο εύρος [-1, 1].

Έχοντας εφαρμόσει πειραματικά την μεθοδολογία μας, στην συνέχεια εκπαιδεύουμε τα μοντέλα μας και ακολουθεί η τελική τους αξιολόγηση στο προαναφερθέν σύνολο δεδομένων.

Λέξεις Κλειδιά

Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μάθηση, Συνελκτικά Νευρωνικά Δίκτυα, Αναγνώριση Συναισθημάτων, Ταξινόμηση, Παλινδρόμηση, Σθένος, Διέγερση, Βασικά συναισθήματα.

Abstract

This diploma thesis deals with the topic of Emotion Recognition and Valence/Arousal Detection using Machine Learning methods and extracting information from face images. The field of sentiment analysis is a very active research field with a large number of applications in our daily life.

We approached this problem by making use of Convolutional Neural Networks by also exploiting the data augmentation method in order to get a larger amount of data. The methodology we implemented couples two problems, a classification problem and a regression problem, where it exploits the labels of valence/arousal in order to lead to better performance in the classification problem.

The labels are provided by the dataset we used, called Affectnet, which consists of a large number of images depicting human expressions that have been labeled with respect to emotion, valence and arousal. Regarding the emotion labels, they belong to one of 8 distinct categories: Neutral, Happy, Sad, Surprise, Fear, Disgust, Anger, Contempt while the valence and arousal labels are decimal numbers in the range $[-1, 1]$.

Having experimentally applied our methodology, we then train our models and finally we evaluate them on the given dataset.

Keywords

Artificial Intelligence, Machine Learning, Deep Learning, Convolutional Neural Networks, Emotion Recognition, Classification, Regression, Valence, Arousal, Basic Expressions.

Στην οικογένεια μου

Ευχαριστίες

Η παρούσα διπλωματική γράφει την τελευταία σελίδα του κεφαλαίου της ζωής μου που ονομάζεται σχολή ΗΜΜΥ. Ενός κεφαλαίου πέντε ετών, με στιγμές χαράς και ενθουσιασμού αλλά παράλληλα και άγχους και στεναχώριας, το οποίο με γέμισε γνώσεις και εμπειρίες και οδήγησε στο να γίνω ο άνθρωπος που είμαι πλέον.

Για αυτό το ταξίδι της διπλωματικής, λοιπόν, θα ήθελα να ευχαριστήσω αρχικά τον κο Στέφανο Κόλλια, ο οποίος με τίμησε με την επιλογή του να δεχτεί να είναι επιβλέπων μου σε μία τόσο σημαντική, για τις σπουδές μου, εργασία όσο αυτή. Επίσης, θα ήθελα να ευχαριστήσω θερμά την κα Παρασκευή Τζούβελη η οποία από την πρώτη στιγμή αφιέρωσε χρόνο σε εμένα και ασχολήθηκε να μου λύσει απορίες και άγχη σχετικά με το κομμάτι της διπλωματικής. Ακόμα, χρωστάω ένα μεγάλο ευχαριστώ από καρδιάς στον κο Δημήτριο Κόλλια που αποτέλεσε στήριγμα καθ'όλη την διάρκεια της εκπόνησης αυτής της εργασίας και ο οποίος με τις γνώσεις, τις ιδέες, το χιούμορ και την καλή του διάθεση βοήθησε στο να γίνει αυτή η διαδικασία όχι μόνο πιο ξεκούραστη αλλά ταυτόχρονα ευχάριστη.

Τέλος, θα ήθελα να ευχαριστήσω τα αδέρφια μου, Πηνελόπη και Νεκτάριο, που έκαναν υπομονή μαζί μου τις μέρες που η σχολή φάνταζε αδιαπέραστο βουνό καθώς και τους γονείς μου, Αντώνη και Άννα, που μου έδωσαν τα φτερά για να μπορέσω να πετάξω μόνη μου.

Αθήνα, Οκτώβριος 2022

Αλεξάνδρα Καπαρού

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	7
1 Εισαγωγή	17
1.1 Περιγραφή του προβλήματος	17
1.2 Αντικείμενο της διπλωματικής	18
1.3 Οργάνωση του τόμου	18
I Θεωρητικό Μέρος	21
2 Θεωρητικό Υπόβαθρο	23
2.1 Μηχανική μάθηση	23
2.1.1 Επιβλεπόμενη Μάθηση	23
2.1.2 Μη επιβλεπόμενη Μάθηση	24
2.1.3 Ενισχυτική Μάθηση	24
2.2 Τεχνητά νευρωνικά δίκτυα (ΤΝΔ)	25
2.2.1 Το μοντέλο του τεχνητού νευρώνα	25
2.3 Συνελκτικά νευρωνικά δίκτυα (CNN)	26
2.3.1 Επίπεδο Εισόδου (Input Layer)	27
2.3.2 Συνελκτικό Επίπεδο (Convolutional Layer)	27
2.3.3 Επίπεδο Ενεργοποίησης (Activation Layer)	29
2.3.4 Επίπεδο Υποδειγματοληψίας (Pooling Layer)	30
2.3.5 Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)	31
2.3.6 Επίπεδο Κανονικοποίησης Παρτίδας (Batch Normalization Layer)	32
2.3.7 Επίπεδο Απόσυρσης (Dropout Layer)	32
2.4 Σύγχρονα συνελκτικά νευρωνικά δίκτυα	33
2.4.1 Resnet	33
2.5 Υπερπαράμετροι νευρωνικών δικτύων	34
2.5.1 Ρυθμός Απόσυρσης (Dropout Rate)	34
2.5.2 Ρυθμός εκμάθησης (Learning rate)	35
2.5.3 Συνάρτηση ενεργοποίησης (Activation function)	35
2.5.4 Αριθμός εποχών (Number of epochs)	35

2.5.5 Μέγεθος παρτίδας (Batch size)	36
2.6 Τυπική Προεπεξεργασία Εικόνων Προσώπων	36
2.7 Συναρτήσεις απωλειών (Loss Functions)	39
2.7.1 Συνάρτηση Απωλειών Προβλήματος Ταξινόμησης	39
2.7.2 Συνάρτηση Απωλειών Προβλήματος Παλινδρόμησης	40
2.8 Μετρικές Αξιολόγησης Μοντέλων	40
2.8.1 Μετρικές Αξιολόγησης Προβλήματος Ταξινόμησης	41
2.8.2 Μετρικές Αξιολόγησης Προβλήματος Παλινδρόμησης	42
2.9 Επαύξηση Δεδομένων (Data Augmentation)	43
3 Ανάλυση συναισθημάτων	47
3.1 Θεωρία των Διακριτών Συναισθημάτων	47
3.1.1 Η Θεωρία του Δαρβίνου	48
3.1.2 Η Θεωρία των Βασικών Συναισθημάτων του Paul Ekman	48
3.1.3 Σύστημα Κωδικοποίησης Ενεργειών Προσώπου (FACS)	50
3.2 Διαστατικές Αναπαραστάσεις	51
3.2.1 Σθένος και Διέγερση (Valence and Arousal)	51
4 Σύνολα Δεδομένων	53
4.1 RAF-DB	53
4.2 Afew-VA	54
4.3 AffectNet	55
II Πρακτικό Μέρος	59
5 Παρεμφερείς εργασίες	61
6 Μεθοδολογία	63
6.1 Εκ νέου επισήμανση με βάση την σχετικότητα διαφορετικών ετικετών	63
6.2 Εκκαθάριση Δεδομένων (Data Cleaning)	65
6.2.1 Ουδέτερη έκφραση (Neutral)	65
6.2.2 Χαρούμενη έκφραση (Happy)	65
6.2.3 Λυπημένη έκφραση (Sad)	66
6.2.4 Έκπληκτη έκφραση (Surprise)	66
6.2.5 Έκφραση φόβου (Fear)	66
6.2.6 Έκφραση αηδίας (Disgust)	66
6.2.7 Ουμωμένη έκφραση (Anger)	66
6.2.8 Έκφραση περιφρόνησης (Contempt)	66
6.3 Εκκαθάριση δεδομένων στο σύνολο δοκιμής	67
7 Πειραματική Διαδικασία	69
7.1 Υπολογιστικά Συστήματα	69
7.2 Διαδικασία Εκπαίδευσης	69

8 Αποτελέσματα και Αξιολόγηση	71
8.1 Αξιολόγηση Μοντέλων Σθένους-Διέγερσης	71
8.2 Αξιολόγηση Μοντέλων Βασικών Συναισθημάτων	72
8.3 Αξιολόγηση Μοντέλων Συνδυασμού Βασικών Συναισθημάτων και Σθένους/Διέγερσης	73
8.4 Αξιολόγηση Μοντέλων εκ νέου Επισημάνσης με βάση την Σχετικότητα Διαφορετικών Ετικετών	74
8.5 Στατιστικά που αφορούν την εκκαθάριση δεδομένων	76
III Επίλογος	81
9 Επίλογος και Μελλοντικές Επεκτάσεις	83
9.1 Συμπεράσματα Διπλωματικής Εργασίας	83
9.2 Μελλοντικές Επεκτάσεις	84
Βιβλιογραφία	93

Κατάλογος Σχημάτων

2.1	Απλοποιημένη δομή Τεχνητού Νευρώνα [1]	26
2.2	Αρχιτεκτονική ενός Συνελκτικού Νευρωνικού Δικτύου [2]	27
2.3	Συνέλιξη φίλτρου ενός συνελκτικού επιπέδου με τον πίνακα εισόδου και παραγωγή ενός χάρτη ενεργοποίησης [3]	28
2.4	Zero padding [4]	28
2.5	Η συνάρτηση ενεργοποίησης ReLU, $\max(0, x)$. [5]	29
2.6	Ένα παράδειγμα εισόδου και εξόδου από μία συνάρτηση ενεργοποίησης ReLU, $\max(0, x)$. [6]	30
2.7	Επίπεδο υποδειγματοληψίας [7]	31
2.8	Τεχνική της απόσυρσης κόμβων-νευρώνων [8]	33
2.9	Γενική αναπαράσταση του δικτύου ResNet 34 επιπέδων.[9]	33
2.10	Παράδειγμα υποεκπαίδευσης, ισορροπημένου μοντέλου και υπερεκπαίδευσης αντίστοιχα. [10]	34
2.11	Από αριστερά προς τα δεξιά: (α) Ένας μικρός ρυθμός εκμάθησης απαιτεί πολλές ενημερώσεις πριν φτάσει στο ελάχιστο σημείο, (β) Ο βέλτιστος ρυθμός εκμάθησης φτάνει γρήγορα στο ελάχιστο σημείο, (γ) Πολύ μεγάλος ρυθμός εκμάθησης προκαλεί δραστικές ενημερώσεις που οδηγούν σε αποκλίνουσες συμπεριφορές. [11]	35
2.12	Παραδείγματα συναρτήσεων ενεργοποίησης [12]	36
2.13	Παράδειγμα ανίχνευσης προσώπου και οφθαλμών. [13]	37
2.14	Παράδειγμα ευθυγράμμισης προσώπου. [13]	37
2.15	Παράδειγμα περικοπής προσώπου. [13]	38
2.16	Παράδειγμα αλλαγής μεγέθους εικόνας. [13]	38
2.17	Παράδειγμα κανονικοποίησης εικόνας. [13]	38
2.18	Παράδειγμα αποθορυβοποίησης εικόνας. [14]	39
2.19	Η ερμηνεία των κελιών ενός confusion matrix.[15]	42
2.20	Από αριστερά, έχουμε την αρχική εικόνα, ακολουθούμενη από την εικόνα που έχει αναστραφεί οριζόντια, και στη συνέχεια την εικόνα που έχει αναστραφεί κάθετα. [16]	44
2.21	Οι εικόνες περιστρέφονται κατά 90 μοίρες δεξιόστροφα σε σχέση με την προηγούμενη, καθώς κινούμαστε από αριστερά προς τα δεξιά. [16]	44
2.22	Από αριστερά προς τα δεξιά, η αρχική εικόνα και η εικόνα με μεταβολή φωτεινότητας μεγαλύτερη του 1. [17]	45
3.1	Επτά Βασικά Συναισθήματα κατά Ekman [18]	49

3.2	Ενδεικτικά παραδείγματα AU [19]	50
3.3	Έξι από τα επτά βασικά συναισθήματα με σύνθεση μονάδων δράσης [20] . . .	50
3.4	Ο διαδιάστατος κύκλος συναισθημάτων [:]	52
4.1	Παραδείγματα εικόνων της RAF-DB από τις 6 βασικές κατηγορίες και από άλλες 12 σύνθετες συναισθηματικές καταστάσεις [21]	53
4.2	Κατανομή δειγμάτων σε κλάσεις για το σύνολο εκπαίδευσης (training set) και το σύνολο επικύρωσης (validation set) [22]	54
4.3	Παράδειγμα επισημειωμένων τιμών valence και arousal για ένα δείγμα βίντεο από το σύνολο δεδομένων μας μαζί με μερικά αντιπροσωπευτικά καρέ. [23] .	55
4.4	Κατανομή των τιμών valence και arousal στο σύνολο δεδομένων AFEW-VA. [23]	55
4.5	Δείγμα εικόνων από την βάση AffectNet. [24]	56
4.6	Κατανομή δειγμάτων σε κλάσεις για το σύνολο εκπαίδευσης (training set) και για το σύνολο επικύρωσης (validation set) της βάσης AffectNet. [22]	57
4.7	Κατανομή δειγμάτων valence-arousal σε τεταρτημόρια της βάσης AffectNet. [25]	57
6.1	Τα clusters που έχουν δημιουργηθεί.	63
6.2	Τα clusters που έχουν δημιουργηθεί και πού τοποθετείται σε σχέση με τα clusters μία τυχαία εικόνα με σθένος = 0.785714 και διέγερση = -0.055556.	64
6.3	Οι αποστάσεις από κάθε cluster (neutral, happy, sad, surprise, fear, disgust, anger, contempt αντίστοιχα).	64
6.4	Οι νέες αποστάσεις που υπολογίζονται από τον τύπο $\max(distance)/distance$	64
6.5	Οι αποστάσεις μετά την εφαρμογή της softmax.	65
8.1	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος παλινδρόμησης.	71
8.2	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος παλινδρόμησης με εφαρμογή επαύξησης δεδομένων.	72
8.3	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης.	73
8.4	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με εφαρμογή επαύξησης δεδομένων.	73
8.5	Τα clusters που έχουν δημιουργηθεί.	74
8.6	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\lambda = 0.7$	75
8.7	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\lambda = 0.9$ έπειτα από εκκαθάριση δεδομένων.	75
8.8	Τα νέα clusters που έχουν δημιουργηθεί σε σύγκριση με τα αρχικά. Παρατηρούμε ότι μετά την εκκαθάριση τα νέα clusters έχουν μετατοπιστεί.	75
8.9	Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\lambda = 0.7$ έπειτα από εκκαθάριση δεδομένων και εκ νέου υπολογισμό των clusters.	76

Κατάλογος Πινάκων

6.1	Αριθμός εικόνων πριν την εκκαθάριση δεδομένων και ποσοστά λάθος εικόνων ανά συναίσθημα	67
6.2	Αριθμός εικόνων και ποσοστά λάθος εικόνων ανά συναίσθημα στο σύνολο δοκιμής	68
8.1	Αποτελέσματα των μοντέλων παλινδρόμησης σθένους-διέγερσης στο σύνολο αξιολόγησης.	72
8.2	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης.	73
8.3	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων και ανίχνευσης σθένους/διέγερσης στο σύνολο αξιολόγησης.	74
8.4	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης.	76
8.5	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Neutral	77
8.6	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Happy	77
8.7	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Sad	77
8.8	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Fear	77
8.9	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Disgust	78
8.10	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Anger	78
8.11	Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα Contempt	78

Κεφάλαιο 1

Εισαγωγή

Στις επόμενες ενότητες παραθέτουμε μία γενική εισαγωγή που αφορά το αντικείμενο της διπλωματικής και το πρόβλημα το οποίο επιθυμούμε να λύσουμε μέσω της εργασίας μας. Στη συνέχεια περιγράφουμε την δομή της εργασίας προκειμένου να είναι πιο εύκολη η πλοήγηση του αναγνώστη σε αυτήν.

1.1 Περιγραφή του προβλήματος

Το πεδίο της αναγνώρισης συναισθήματος με χρήση υπολογιστικής όρασης απασχολεί μεγάλο πλήθος ερευνητών τα τελευταία χρόνια. Ο κλάδος της πληροφορικής που ασχολείται με την αναγνώριση συναισθημάτων ονομάζεται συναισθηματική υπολογιστική (Affective computing) και αφορά την αναγνώριση ανθρώπινων συναισθημάτων βασισμένων σε έναν ή περισσότερους τύπους δεδομένων που συλλέγονται από το άτομο του οποίου η συναισθηματική κατάσταση αξιολογείται.

Η συναισθηματική υπολογιστική δεν αναφέρεται σε ολοκληρωτική αναγνώριση των συναισθημάτων αλλά σε μια προσέγγιση αυτών. Οι κύριες δυσκολίες σε ένα εγχείρημα πλήρους αναγνώρισης είναι αρχικά το γεγονός ότι σύμφωνα με την ψυχολογία το συναίσθημα είναι μια σύνθετη υποκειμενική συνειδητή εμπειρία, επομένως η κατηγοριοποίηση του εκάστοτε συναισθήματος σε μία κατηγορία η οποία θα έβρισκε σύμφωνους όλους τους ανθρώπους θα ήταν πρακτικά αδύνατη, μιας και ο τρόπος που αντιλαμβάνεται ο κάθε ένας μία οπτική έκφραση συναισθήματος διαφέρει από άνθρωπο σε άνθρωπο. Επίσης, στην φύση τα συναισθήματα είναι σύνθετα, δηλαδή μείξεις διαφόρων συναισθημάτων και όχι διακριτά όπως θα διευκόλυνε την ολοκληρωτική αναγνώριση τους. Έτσι, ένας άνθρωπος μπορεί να είναι χαρούμενος αλλά παράλληλα λίγο έκπληκτος, λίγο αισιόδοξος κ.ο.κ. Η αποτύπωση αυτού του συναισθήματος σε μία εικόνα υπάρχει μεγάλη πιθανότητα να "κρύβει" τα υπόλοιπα συναισθήματα και έτσι ένα νευρωνικό δίκτυο να εντοπίζει μόνο το συναίσθημα που είναι πιο εμφανές, οδηγώντας σε μία ελλιπή αναγνώριση συναισθήματος.

Πέρα όμως από την κατηγοριοποίηση του συναισθήματος σε διακριτές κλάσεις, η συναισθηματική υπολογιστική περιλαμβάνει και άλλες εργασίες, εξίσου σημαντικές, όπως είναι η αναγνώριση του σθένους και της διέγερσης ενός συναισθήματος. Έτσι, ενώ σε μερικές εφαρμογές χρειάζεται να εντοπίζουμε επακριβώς την συναισθηματική κατάσταση του χρήστη, σε άλλες κάτι τέτοιο δεν είναι τόσο χρήσιμο όσο το πόσο θετική/αρνητική και ενεργητική/παθητική είναι η ψυχική του κατάσταση.

Οι παραπάνω εργασίες μπορούν να βρουν χρήση σε ένα ευρύ φάσμα πεδίων. Ενδεικτικά αναφέρουμε τα κάτωθι:

- **Ψυχολογία** : Η ψυχοθεραπεία μπορεί να επωφεληθεί από εφαρμογές ανίχνευσης συναισθήματος καθώς με την αναγνώριση της συναισθηματικής κατάστασης του ασθενή η συμβουλευτική του επαγγελματία υγείας θα είναι πιο στοχευμένη [26, 27, 28, 29, 30, 31, 32, 33, 34].
- **Εκπαίδευση** : Η αναγνώριση των συναισθημάτων των μαθητών μπορεί να βοηθήσει στην ανίχνευση της εμπλοκής του μαθητή στην εκπαιδευτική διαδικασία (παραδείγματος χάρι αν ένας μαθητής δεν είναι συγκεντρωμένος στο μάθημα τότε θα έχει παθητική κατάσταση άρα αρνητική διέγερση) και στην ενίσχυση του τρόπου ηλεκτρονικής μάθησης με στοιχεία που θα εξασφαλίζουν την προσοχή του μαθητή [35, 36, 37, 38].
- **Οδήγηση** : Οι αυτοκινητοβιομηχανίες μπορούν να αξιοποιήσουν την όραση υπολογιστών για την παρακολούθηση της συναισθηματικής κατάστασης των οδηγών κατά την οδήγηση. Έτσι, εάν ο οδηγός είναι πολύ κουρασμένος, αγχωμένος ή θυμωμένος, το όχημα μπορεί να παρέχει προειδοποιήσεις για μη ασφαλή οδήγηση [39, 40].

1.2 Αντικείμενο της διπλωματικής

Σκοπός της παρούσας διπλωματικής εργασίας είναι η προσέγγιση των θεμάτων της Αναγνώρισης Συναισθήματος και της Ανίχνευσης Σθένους/Διέγερσης από εικόνες, χρησιμοποιώντας τεχνικές της Μηχανικής Μάθησης. Η προσέγγιση μας περιλαμβάνει την χρήση Συνελκτικών Νευρωνικών Δικτύων προκειμένου να καταστεί εφικτή η ανίχνευση του συναισθήματος αλλά και των τιμών σθένους/διέγερσης, ανάλογα με το πείραμα που εκτελούσαμε κάθε φορά. Πιο αναλυτικά, η μεθοδολογία που παρουσιάζουμε, δεδομένης μιας εικόνας προσώπου είναι σε θέση να υπολογίσει την ετικέτα όσον αφορά το συναίσθημα, κάνοντας χρήση της γνώσης των ετικετών σθένους και διέγερσης. Επιπλέον, μέρος της εργασίας αποτελεί η μελέτη της θεωρίας των Νευρωνικών Δικτύων αποσκοπώντας στη βαθύτερη κατανόησή της και έπειτα στην εφαρμογή της για το σχεδιασμό, την εκπαίδευση και την αξιολόγηση μοντέλων που επιλύουν το παραπάνω πρόβλημα. Τέλος, σημαντικό ρόλο για την τελική απόδοση των μοντέλων παίζει η επιλογή των συνόλων δεδομένων που θα χρησιμοποιήσουμε, μιας και θα πρέπει να έχουν επαρκή δείγματα, σωστές ετικέτες κ.ο.κ καθώς και η προεπεξεργασία του συνόλου δεδομένων, το οποίο θα αποτελέσει την είσοδο των μοντέλων μας.

1.3 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε 8 Κεφάλαια, με τα Κεφάλαια 2-4 να αποτελούν το θεωρητικό κομμάτι της εργασίας και τα Κεφάλαια 5-8 το πειραματικό μέρος. Το Κεφάλαιο 2 περιέχει το θεωρητικό υπόβαθρο των πεδίων της μηχανικής μάθησης και των νευρωνικών δικτύων που είναι απαραίτητο προκειμένου κάποιος να κατανοήσει πλήρως την παρούσα εργασία. Στο Κεφάλαιο 3 γίνεται παρουσίαση της θεωρίας της ανάλυσης συναισθημάτων,

η οποία αποτελεί κεντρικό θέμα της συγκεκριμένης εργασίας ενώ στο Κεφάλαιο 4 γίνεται την περιγραφή των αντίστοιχων συνόλων δεδομένων για την εκτέλεση των απαιτούμενων εργασιών. Αναφορικά με το πρακτικό μέρος, αρχικά στο Κεφάλαιο 5 ασχολούμαστε με παρεμφερείς εργασίες της επιστημονικής κοινότητας στο πεδίο της ανάλυσης συναισθημάτων. Στην συνέχεια, αναλύεται η μεθοδολογία που ακολουθήθηκε για την εκ νέου επισήμανση του συνόλου δεδομένων με βάση την σχετικότητα διαφορετικών ετικετών ενώ στο Κεφάλαιο 7 περιγράφεται ολόκληρη η πειραματική διαδικασία με αναλυτική περιγραφή των βημάτων εκπαίδευσης. Τέλος, στο Κεφάλαιο 8 παρουσιάζονται τα αποτελέσματα της διπλωματικής εργασίας ενώ στον Επίλογο αναφέρουμε τα τελικά συμπεράσματα της διπλωματικής εργασίας και γίνονται προτάσεις για μελλοντικές επεκτάσεις.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Σε αυτή την ενότητα θα καλύψουμε το απαραίτητο θεωρητικό υπόβαθρο το οποίο θα πρέπει να γνωρίζει ο αναγνώστης προκειμένου να κατανοήσει την παρούσα εργασία.

2.1 Μηχανική μάθηση

Μηχανική μάθηση (machine learning) ονομάζεται το πεδίο της επιστήμης των υπολογιστών που μελετά τη δημιουργία αλγορίθμων οι οποίοι «μαθαίνουν» χωρίς να έχουν προγραμματιστεί με συγκεκριμένους κανόνες. Με άλλα λόγια, οι αλγόριθμοι αυτοί χρησιμοποιούν δεδομένα με σκοπό να ανακαλύψουν μοτίβα και σχέσεις ώστε να κάνουν προβλέψεις ή να πάρουν αποφάσεις. Αναφορές στη μηχανική μάθηση υπάρχουν από τη δεκαετία του 1960 όμως η χρήση των τεχνικών αυτών αυξήθηκε ραγδαία μετά τη δεκαετία του 1990 ως αποτέλεσμα της ανάπτυξης κλάδων της επιστήμης υπολογιστών όπως είναι η ψηφιοποίηση αρχείων, η εξόρυξη δεδομένων και οι υπερυπολογιστές. Εν γένει, ο τομέας της Μηχανικής Μάθησης αναπτύσσει τρεις τρόπους μάθησης, ανάλογους με τους τρόπους με τους οποίους μαθαίνει ο άνθρωπος: επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση και ενισχυτική μάθηση.

2.1.1 Επιβλεπόμενη Μάθηση

Επιβλεπόμενη (supervised) μάθηση ονομάζεται η τεχνική με την οποία ένα πρόγραμμα εκπαιδεύεται για να καταλάβει τη σχέση μεταξύ των δεδομένων που δίνουμε και ενός επιθυμητού αποτελέσματος. Δηλαδή έχουμε προκαθορισμένη είσοδο (input - δεδομένα) και έξοδο (output - αποτέλεσμα). Αλλιώς, συνηθίζεται να λέμε ότι έχουμε δεδομένα με ετικέτες (labels) που δείχνουν τη σύνδεση με την έξοδο, τις οποίες έχουν βάλει άνθρωποι ή άλλοι κώδικες (επιβλεπόμενη). Η επιβλεπόμενη μάθηση χρησιμοποιείται συνήθως σε προβλήματα :

- Ταξινόμησης (Classification) : Η ταξινόμηση είναι ένα πρόβλημα αυτόματης ανάθεσης μιας ετικέτας σε ένα παράδειγμα που δεν έχει ετικέτα, δηλαδή η τιμή της ετικέτας είναι ποιοτική, εφόσον αντιπροσωπεύει μία κλάση που περιγράφει το διάνυσμα χαρακτηριστικών. Η ανίχνευση της ανεπιθύμητης ηλεκτρονικής αλληλογραφίας είναι ένα διάσημο παράδειγμα ταξινόμησης. Σε ένα πρόβλημα ταξινόμησης μια ετικέτα είναι μέρος ενός πεπερασμένου συνόλου τάξεων. Εάν το μέγεθος του συνόλου των τάξεων είναι δύο τότε μιλάμε για μια διωνυμική ταξινόμηση (binary classification) ενώ

όταν οι τάξεις είναι περισσότερες από δύο τότε μιλάμε για προβλήματα πολυωνυμικής ταξινόμησης (multiclass classification).

- Παλινδρόμησης (Regression) : σε αυτή την περίπτωση η τιμή της ετικέτας είναι ποσοτική, δηλαδή παίρνει συνεχείς τιμές περιγράφοντας ένα μέγεθος. Η εκτίμηση της τιμής μια κατοικίας με βάση τα χαρακτηριστικά του σπιτιού, όπως η περιοχή, ο αριθμός των υπνοδωματίων κ.λπ. είναι ένα διάσημο παράδειγμα παλινδρόμησης.

Με λίγα λόγια η διαφορά των δύο ειδών προβλημάτων συνοψίζεται στις δύο παρακάτω γραμμές:

Κατηγοριοποίηση είναι το πρόβλημα επιβλεπόμενης μάθησης με διακριτές κλάσεις y .

Παλινδρόμηση είναι το πρόβλημα επιβλεπόμενης μάθησης με συνεχείς κλάσεις y .

2.1.2 Μη επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση (unsupervised learning) είναι το είδος εκπαίδευσης στο οποίο χρησιμοποιούνται πληροφορίες που δεν είναι ούτε ταξινομημένες ούτε επισημασμένες και έτσι δίνεται η δυνατότητα στον αλγόριθμο να ενεργεί σε αυτές τις πληροφορίες χωρίς καθοδήγηση. Σκοπός είναι να ομαδοποιούνται ασαφείς πληροφορίες σύμφωνα με ομοιότητες, μοτίβα και διαφορές, χωρίς προηγούμενη εκπαίδευση σε δεδομένα. [41] [42] Οι αλγόριθμοι μη επιβλεπόμενης μάθησης εφαρμόζουν τις παρακάτω τεχνικές για να περιγράψουν τα δεδομένα:

- Ομαδοποίηση (Clustering): Πραγματοποιείται μια διερεύνηση των δεδομένων, προκειμένου να διαχωριστούν σε ομάδες (clusters), βάσει κάποιων μοτίβων χωρίς προηγούμενη γνώση των χαρακτηριστικών της κάθε ομάδας. Τα χαρακτηριστικά προσδιορίζονται από την ομοιότητα των επιμέρους δεδομένων αλλά και από τις διαφορές από τα υπόλοιπα .
- Μείωση διάστασης (Dimensionality reduction): Εξαιτίας της ύπαρξης «θορύβου» στα εισερχόμενα δεδομένα, οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούν τη μείωση των διαστάσεων για την εξάλειψη αυτού του θορύβου κατά τον διαχωρισμό των σχετικών πληροφοριών. [43]

2.1.3 Ενισχυτική Μάθηση

Στην ενισχυτική μάθηση (reinforcement learning) κατασκευάζουμε ένα εικονικό «περιβάλλον» που έχει συγκεκριμένους κανόνες και αφήνουμε τον υπολογιστή να αλληλεπιδράσει με αυτό μέχρι την επίτευξη κάποιου στόχου όπως η μεγιστοποίηση ενός σκορ. Για την ώρα οι περισσότερες εφαρμογές αυτού του είδους μάθησης υπάρχουν σε παιχνίδια. Με τη χρήση ενισχυτικής μάθησης, οι υπολογιστές μαθαίνουν να παίζουν τόσο καλά μερικά παιχνίδια που πλέον οι άνθρωποι δεν μπορούν να τους ανταγωνιστούν. Στην ουσία, η ενισχυτική μάθηση αφορά στην ανάπτυξη ενός αυτοσυντηρούμενου συστήματος το οποίο, σε συνεχείς αλληλουχίες προσπάθειας και αποτυχίας, βελτιώνεται με βάση το συνδυασμό των δεδομένων και των αλληλεπιδράσεων με τα εισερχόμενα δεδομένα.

2.2 Τεχνητά νευρωνικά δίκτυα (ΤΝΔ)

Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANNs), αποτελούν απλοποιημένα μοντέλα του κεντρικού νευρικού συστήματος των ζώντων οργανισμών και ειδικότερα του ανθρώπου. Οι λειτουργίες τους, μιμούνται αυτές των βιολογικών νευρώνων του εγκεφάλου και τη δομή των βιολογικών Νευρικών Δικτύων. Επιχειρούν δηλαδή, να συνδυάσουν την λειτουργία του ανθρώπινου εγκεφάλου, με τον αυστηρά αφηρημένο μαθηματικό τρόπο σκέψης, ξεχωρίζοντας με αυτό τον τρόπο τα ΤΝΔ από την Βιολογία και την κλασσική λειτουργία των υπολογιστών. Έτσι, για παράδειγμα, ένα Νευρωνικό Δίκτυο είναι ικανό να μαθαίνει, να εκπαιδεύεται, να θυμάται ή να ξεχνά μία αριθμητική παράσταση κ.λ.π, πράξεις που μέχρι πρότινος αποδίδαμε μόνο στην ανθρώπινη σκέψη και ικανότητα. Επιπλέον, καθίσταται ικανά να χρησιμοποιούν πολύπλοκες συναρτήσεις και άλλα σύνθετα αντικείμενα της μαθηματικής Ανάλυσης.

2.2.1 Το μοντέλο του τεχνητού νευρώνα

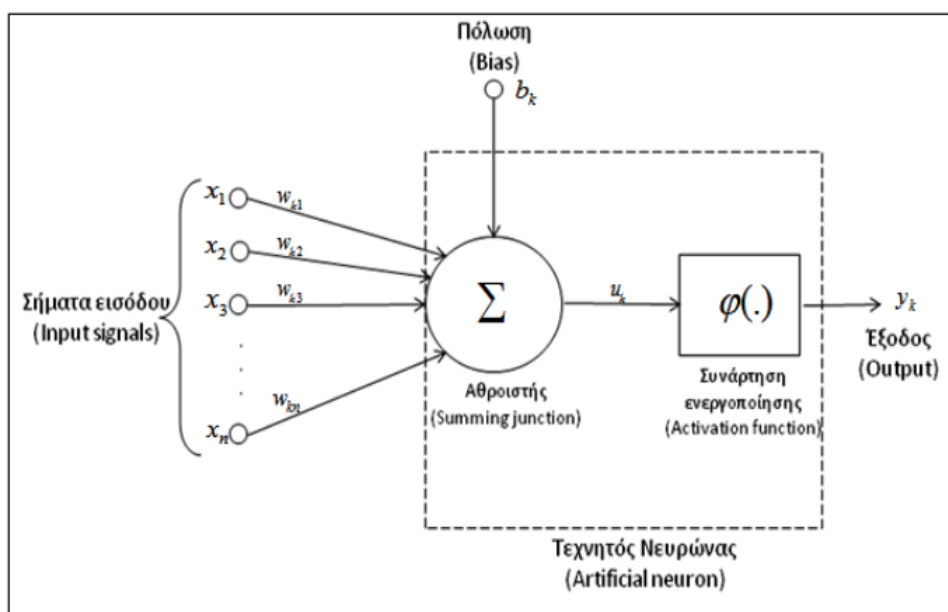
Ο Τεχνητός Νευρώνας (Artificial Neuron) είναι η βάση για τον σχεδιασμό των ΤΝΔ καθώς αποτελεί την βασική μονάδα επεξεργασίας των διαθέσιμων πληροφοριών. Για αυτό τον λόγο ένα ΤΝΔ χαρακτηρίζεται ως μια αλληλεπίδραση των απλών αυτών στοιχείων, δηλαδή των τεχνητών νευρώνων. Όπως έχει ήδη αναφερθεί ερέθισμά τους αποτέλεσαν οι νευρώνες (neurons) που συνιστούν τον ανθρώπινο εγκέφαλο για αυτό και υπάρχει αντιστοιχία μεταξύ τους στον τρόπο λειτουργία τους. Πιο συγκεκριμένα, ένας τεχνητός νευρώνας χαρακτηρίζεται από τα εξής βασικά στοιχεία :

- Τα *σήματα εισόδου* (input signals) x_j ή αλλιώς την εισερχόμενη πληροφορία (input information) η οποία μπορεί να προέρχεται είτε απευθείας από το περιβάλλον είτε από κάποιον άλλον τεχνητό νευρώνα.
- Ένα σύνολο από *συναψεις* (synapses) ή *συνδεδεμένους συνδέσμους*, καθένας από τους οποίους χαρακτηρίζεται από ένα *συναπτικό βάρος* (Synaptic Weight). Δηλαδή κάθε σήμα εισόδου x_j στην είσοδο της σύναψης j συνδεδεμένης με τον νευρώνα k πολλαπλασιάζεται με το αντίστοιχο συναπτικό του βάρος w_{kj} . Είναι σημαντικό να τονιστεί πως ο πρώτος δείκτης k αντιστοιχεί στον τεχνητό νευρώνα και ο δεύτερος δείκτης j στο τέλος της εισόδου της σύναψης στην οποία αναφέρεται το βάρος (δηλαδή πρόκειται για τον ίδιο δείκτη με εκείνον του σήματος εισόδου). Τα συναπτικά βάρη μπορούν να πάρουν οποιαδήποτε τιμή, ωστόσο πολλές φορές γίνεται η κανονικοποίηση τους απαιτώντας το άθροισμα τους να είναι ίσο με 1.
- Έναν *αθροιστή* (summing junction or adder) των σημάτων εισόδου, αφού προηγουμένως αυτά έχουν σταθμιστεί με την χρήση των συναπτικών βαρών.
- Μια *συνάρτηση ενεργοποίησης* (activation function) ή αλλιώς *συνάρτηση μεταφοράς* (transfer function) $\phi(\cdot)$ για τον περιορισμό του πλάτους της εξόδου ενός νευρώνα, σε ένα κλειστό μοναδιαίο διάστημα $[0, 1]$ ή $[-1, 1]$. Μπορεί να χρησιμοποιηθεί οποιαδήποτε συνάρτηση ως συνάρτηση ενεργοποίησης και σε ένα ΤΝΔ να υπάρχουν νευρώνες με

διαφορετικές συναρτήσεις. Υπάρχουν διάφορες περιπτώσεις συναρτήσεων ενεργοποίησης όπως η βηματική, η συνάρτηση προσήμου, η σιγμοειδής κ.ά. Τις περισσότερες φορές οι προαναφερθείσες συναρτήσεις είναι μη γραμμικές έτσι ώστε η έξοδος να μην είναι ευθέως ανάλογη με την είσοδο. Ωστόσο μπορεί να χρησιμοποιηθεί και ως συνάρτηση ενεργοποίησης η γραμμική συνάρτηση.

- Το *σήμα εξόδου* (output signal) y_k το οποίο είναι ουσιαστικά το αποτέλεσμα που παράγεται από τον τεχνητό νευρώνα k .
- Τέλος ο τεχνητός νευρώνας περιλαμβάνει και μια εξωτερικά *εφαρμοσμένη πόλωση* ή *μεροληψία* (bias) b_k της οποίας η λειτουργία βοηθά στην αποφυγή του σφάλματος στην περίπτωση που τα δεδομένα εισόδου είναι μηδενικά. Επιπλέον έχει ως αποτέλεσμα την μείωση ή την αύξηση της εισόδου της συνάρτησης ενεργοποίησης ανάλογα με το αν η τιμή της είναι αρνητική ή θετική αντίστοιχα. [1]

Το Σχήμα 2.1 απεικονίζει την απλοποιημένη δομή ενός τεχνητού νευρώνα όπως την περιγράψαμε παραπάνω.



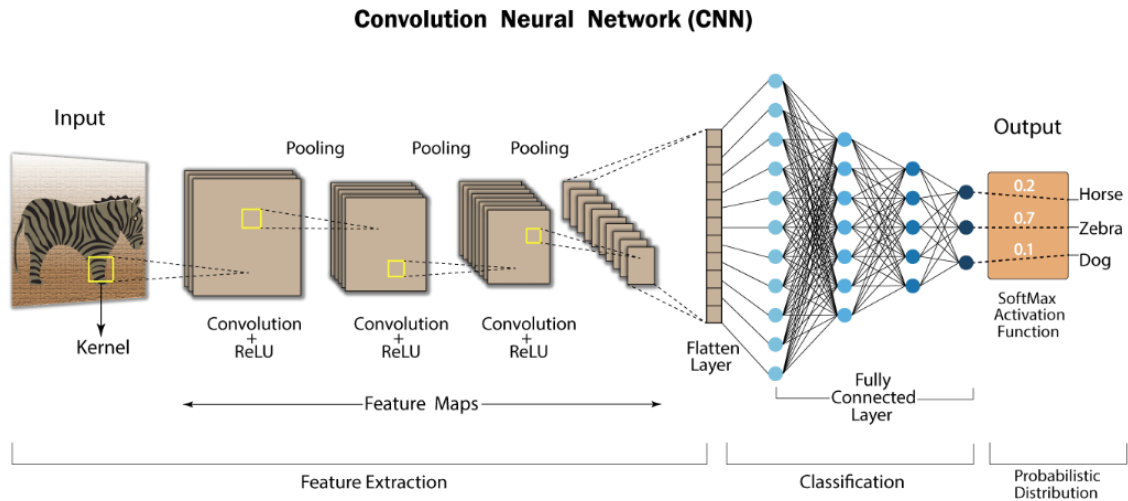
Σχήμα 2.1: Απλοποιημένη δομή Τεχνητού Νευρώνα [1]

2.3 Συνελκτικά νευρωνικά δίκτυα (CNN)

Τα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) αποτελούν δίκτυα βαθιάς μάθησης τα οποία χρησιμοποιούνται κυρίως σε ανάπτυξη εφαρμογών που απαιτούν ανάλυση εικόνας και βίντεο. Το βασικό πλεονέκτημα των συνελκτικών δικτύων έναντι των παραδοσιακών πλήρως συνδεδεμένων δικτύων είναι η διατήρηση της χωρικής συσχέτισης που υπάρχει σε μία εικόνα.

Ο όρος "συνελκτικό" προέρχεται από την γραμμική μαθηματική πράξη της συνέλιξης, η οποία εμπεριέχει τον πολλαπλασιασμό ενός συνόλου βαρών με το διάνυσμα εισόδου, όπως

και στα παραδοσιακά νευρωνικά δίκτυα. Η διαφορά είναι ότι στα συνελκτικά δίκτυα η δισδιάστατη είσοδος πολλαπλασιάζεται με έναν δισδιάστατο πίνακα από βάρη, ο οποίος ονομάζεται πυρήνας (kernel) ή φίλτρο (filter). [44]



Σχήμα 2.2: Αρχιτεκτονική ενός Συνελκτικού Νευρωνικού Δικτύου [2]

Ένα Συνελκτικό Νευρωνικό Δίκτυο αποτελείται από έναν αριθμό από συνελκτικά (convolutional) και υποδειγματοληπτικά (subsampling) επίπεδα, τα οποία, προαιρετικώς, ακολουθούνται από πλήρως συνδεδεμένα επίπεδα (fully-connected layers). Η είσοδος περνάει ακολουθιακά από μια σειρά επεξεργασιών όπου ένα βήμα επεξεργασίας, το οποίο συνήθως αποκαλείται επίπεδο, θα μπορούσε να είναι ένα συνελκτικό επίπεδο (convolutional layer), ένα συγκεντρωτικό επίπεδο (pooling layer), ένα επίπεδο κανονικοποίησης (normalization layer), ένα πλήρως συνδεδεμένο επίπεδο (fully connected layer) ή ένα επίπεδο απωλειών (loss layer). Θα περιγράψουμε λεπτομερώς τα επίπεδα αυτά στην συνέχεια.

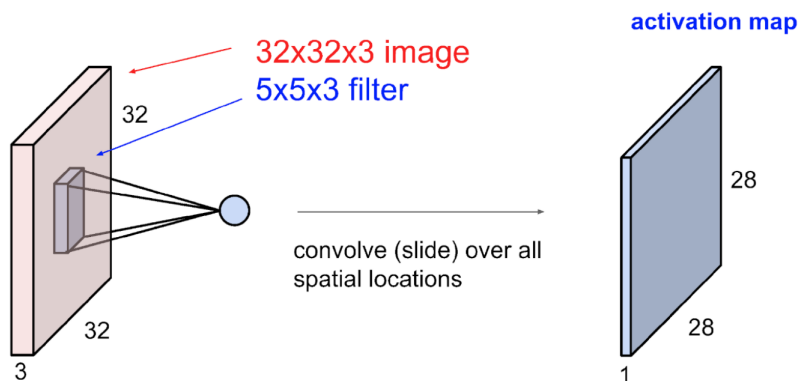
2.3.1 Επίπεδο Εισόδου (Input Layer)

Η είσοδος σε ένα τέτοιο δίκτυο είναι, συνήθως, ένας πίνακας 3 διαστάσεων, ο οποίος αποκαλείται τένσορας. Μια έγχρωμη εικόνα μπορεί να αναπαρασταθεί ως ένας τένσορας 3 διαστάσεων $H \times W \times 3$, όπου το H (height) αντιστοιχεί στον αριθμό των pixel της εικόνας στον κάθετο άξονα, το W (width) αντιστοιχεί στον αριθμό των pixel της εικόνας στον οριζόντιο άξονα και το 3 αναφέρεται στα τρία κανάλια χρωμάτων RGB (Red, Green, Blue). Σε περίπτωση που έχουμε ασπρόμαυρη εικόνα τότε η τρίτη διάσταση ισούται με 1 και η είσοδος του δικτύου είναι ένας πίνακας δύο διαστάσεων όπου κάθε τιμή του πίνακα αντιστοιχεί σε ένα εικονοστοιχείο

2.3.2 Συνελκτικό Επίπεδο (Convolutional Layer)

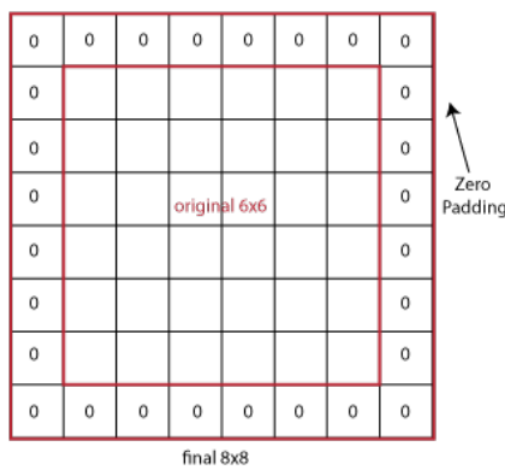
Τα επίπεδα συνέλιξης είναι ο πυρήνας των μοντέλων CNN. Οι παράμετροι ενός συνελκτικού επιπέδου είναι μία σειρά από δισδιάστατα φίλτρα τα οποία όμως εκτείνονται σε όλο το βάθος του όγκου εισόδου. Το βάθος των φίλτρων αυτών ισούται με το βάθος του τένσορα στην είσοδο. Τα επίπεδα αυτά εφαρμόζουν πράξη συνέλιξης πάνω στα δεδομένα εισόδου, το

οποίο επηρεάζει την δομή των "τοπικών" διασυνδέσεων. Η συνέλιξη ενός φίλτρου με τον όγκο εισόδου παράγει έναν χάρτη ενεργοποίησης (activation map), με τον τρόπο που φαίνεται στο σχήμα 2.3. Στο παράδειγμα αυτό εφαρμόζεται φίλτρο διαστάσεων $5 \times 5 \times 3$ σε έναν πίνακα $32 \times 32 \times 3$ και παράγεται ένας χάρτης ενεργοποίησης διαστάσεων $28 \times 28 \times 1$.



Σχήμα 2.3: Συνέλιξη φίλτρου ενός συνελκτικού επιπέδου με τον πίνακα εισόδου και παραγωγή ενός χάρτη ενεργοποίησης [3]

Η τιμή του βάθους του τένσορα στην έξοδο ενός συνελκτικού επιπέδου αντιστοιχεί στον αριθμό των φίλτρων που εφαρμόζονται στον τένσορα εισόδου, δηλαδή ο αριθμός των χαρτών ενεργοποίησης αντιστοιχεί στον αριθμό των φίλτρων, ο οποίος αποτελεί υπερπαραμέτρο του επιπέδου συνέλιξης. Το βήμα μετατόπισης (stride) του φίλτρου πάνω στην είσοδο είναι και αυτό μία υπέρ-παραμέτρος των επιπέδων συνέλιξης.



Σχήμα 2.4: Zero padding [4]

Ένα πρόβλημα που εμφανίζεται στην περίπτωση των μοντέλων CNN με μεγάλο αριθμό κρυφών επιπέδων είναι η γρήγορη μείωση των διαστάσεων μήκους και πλάτους του όγκου, το οποίο είναι αποτέλεσμα της διαδοχικής εφαρμογής πράξεων συνέλιξης. Αυτή η συμπεριφορά είναι ανεπιθύμητη αφού περιορίζει και τις διαστάσεις των φίλτρων που μπορούμε να χρησιμοποιήσουμε σε κάθε συνελκτικό επίπεδο. Η χρήση φίλτρων μεγάλων διαστάσεων φέρει σαν αποτέλεσμα την γρήγορη μείωση των διαστάσεων του όγκου. Για να αποτρέψουμε αυτή την συμπεριφορά μπορούμε να επεκτείνουμε τις διαστάσεις μήκους και πλάτους,

προσθέτοντας μηδενικά στα σύνορα του όγκου εισόδου του εκάστοτε συνελκτικού επιπέδου. Η διαδικασία αυτή ονομάζεται zero-padding και φαίνεται στο σχήμα 2.4. Το μέγεθος του συνόρου που προστίθεται είναι η τρίτη υπέρ-παράμετρος ενός επιπέδου συνέλιξης. [45]

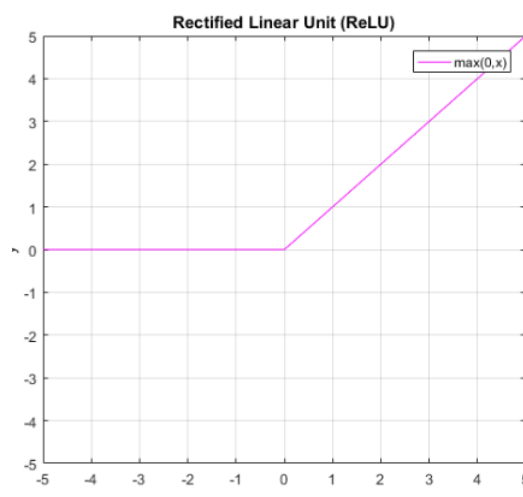
2.3.3 Επίπεδο Ενεργοποίησης (Activation Layer)

Τα επίπεδα ενεργοποίησης είναι εξαιρετικά σημαντικά για ένα τεχνητό νευρωνικό δίκτυο. Είναι τα επίπεδα τα οποία εφαρμόζουν πάνω στους χάρτες ενεργοποίησης μια μη γραμμική απεικόνιση στα στοιχεία τους. Μετά από κάθε συνελκτικό επίπεδο σε ένα CNN, εφαρμόζουμε μια μη γραμμική συνάρτηση ενεργοποίησης, όπως η ReLU (Rectified Linear Unit) η οποία εισάγει τη μη-γραμμικότητα και συγχρόνως είναι και η συνάρτηση ενεργοποίησης. [44] Μία μονάδα ReLU με κατώφλι το 0 περιγράφεται από τη σχέση:

$$Y_i^{(l)} = \max(0, Y_i^{(l-1)}) \quad (2.1)$$

Οι μονάδες αυτές παρουσιάζουν τρία σημαντικά πλεονεκτήματα στα συνελκτικά νευρωνικά δίκτυα σε σχέση με παραδοσιακές συναρτήσεις ενεργοποίησης όπως είναι η σιγμοειδής ή η υπερβολική εφαπτομένη, είναι:

- Οι ReLUs έχουν την ικανότητα να μεταδίδουν την κλίση μεταξύ των επιπέδων πιο αποδοτικά, με αποτέλεσμα να αποφεύγεται η εξαφάνιση κλίσης (vanishing gradient) που αποτελεί συχνό φαινόμενο στα βαθιά νευρωνικά δίκτυα.
- Οι ReLUs παρουσιάζουν μη αρνητικές τιμές κατωφλιού, το οποίο επιλύει το πρόβλημα της ακύρωσης και συμβάλει σε ένα σποραδικό όγκο ενεργοποίησης στην έξοδο τους. Η σποραδικότητα της εξόδου δημιουργεί ανθεκτικότητα σε μικρές διακυμάνσεις της εισόδου, που αποτελούν το θόρυβο [46].
- Οι ReLUs απαρτίζονται μόνο από απλές πράξεις όσον αφορά το υπολογιστικό τους κόστος (κυρίως συγκρίσεις), συνεπώς είναι πιο αποδοτικές στην υλοποίησή τους [5]



Σχήμα 2.5: Η συνάρτηση ενεργοποίησης ReLU, $\max(0, x)$. [5]

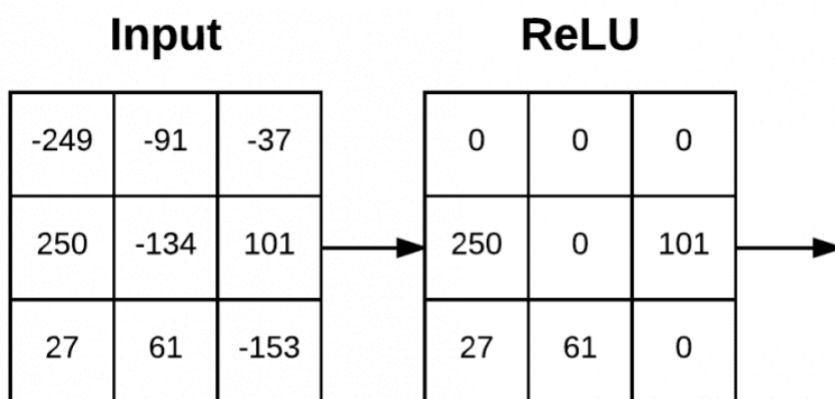
Μία εξίσου σημαντική συνάρτηση ενεργοποίησης που χρησιμοποιείται ευρέως στα επίπεδα ενεργοποίησης είναι η συνάρτηση softmax. Η Softmax συγλίνει ένα διάνυσμα τιμών σε μία κατανομή πιθανοτήτων. Τα στοιχεία του διανύσματος εξόδου έχουν εύρος (0,1) και αθροίζουν στο 1. Η Softmax περιγράφεται από την σχέση:

$$\frac{e^{z_j}}{\sum_k e^{z_k}} \tag{2.2}$$

όπου z_i οποιοσδήποτε πραγματικός αριθμός. [44]

Ένα επίπεδο ενεργοποίησης δέχεται μία είσοδο μεγέθους $W_{input} \times H_{input} \times D_{input}$ και εφαρμόζει την αντίστοιχη συνάρτηση ενεργοποίησης (Σχήμα 2.6). Καθώς η συνάρτηση ενεργοποίησης εφαρμόζεται ανά στοιχείο, η έξοδος του επιπέδου ενεργοποίησης έχει ακριβώς τις ίδιες διαστάσεις με την είσοδο.

Τα επίπεδα ενεργοποίησης δεν είναι τεχνικά "επίπεδα" (εξαιτίας του γεγονότος ότι δεν μαθαίνονται παράμετροι/βάρη μέσα σε ένα επίπεδο ενεργοποίησης) και μερικές φορές παραλείπονται από τα διαγράμματα αρχιτεκτονικής δικτύου, καθώς θεωρείται ότι αμέσως μετά από μια συνέλιξη ακολουθεί μια ενεργοποίηση.



Σχήμα 2.6: Ένα παράδειγμα εισόδου και εξόδου από μία συνάρτηση ενεργοποίησης ReLU, $\max(0, x)$. [6]

2.3.4 Επίπεδο Υποδειγματοληψίας (Pooling Layer)

Επόμενο επίπεδο μετά το τμήμα της συνέλιξης σε ένα CNN είναι το επίπεδο υποδειγματοληψίας. Μία από τις λειτουργίες του είναι η σταδιακή μείωση του μεγέθους αναπαράστασης για τη μείωση του αριθμού των παραμέτρων και των υπολογισμών στο δίκτυο και, ως εκ τούτου, για τη βελτίωση της γενίκευσης με την αύξηση παραμέτρων. [47]

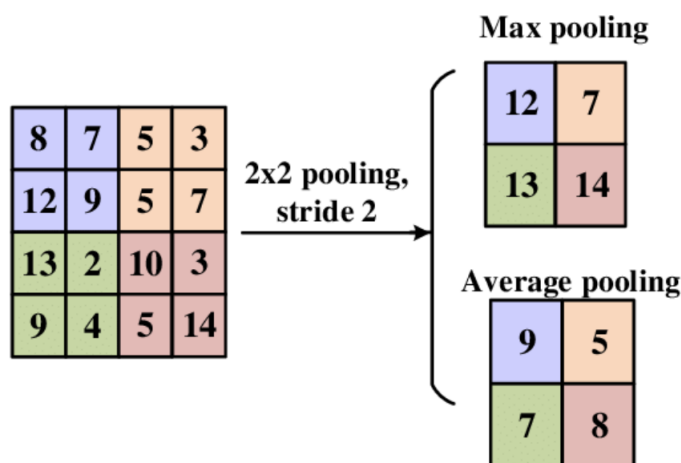
Όμως διαισθητικά ο λόγος που τοποθετείται ένα τέτοιο στρώμα είναι έτσι ώστε η εύρεση ενός χαρακτηριστικού να μην είναι συνυφασμένη με την ακριβή θέση. Αυτό επιτυγχάνεται παραδείγματος χάριν με το max pooling μιας και η ύπαρξη ενός χαρακτηριστικού στην περιοχή, ισοδυναμεί με αυξημένη απόκριση του φίλτρου, συνεπώς και με αυξημένη τιμή στο φίλτρο. Αυτό διατηρεί μόνο τη μέγιστη τιμή σε εκείνη την περιοχή χωρίς να μας ενδιαφέρουν οι μη αποκρίσεις.

Ο λόγος για τον οποίο η απλή δειγματοληψία δεν είναι αρκετά αποδοτική είναι ότι δεν

βοηθά στην προώθηση των μεγάλων ενεργοποιήσεων στο δίκτυο. Δηλαδή αν τύχει ένα εικονοστοιχείο με μεγάλη τιμή να βρίσκεται σε σημείο που προσπερνάει η υποδειγματοληψία τότε η ενεργοποίησή του θα χαθεί. Αυτό θα έχει σαν αποτέλεσμα την μείωση της απόδοσης του δικτύου. Για να αντιμετωπιστεί αυτό το πρόβλημα, δημιουργήθηκαν οι παρακάτω τεχνικές.

- Η τεχνική του average pooling : Σε αυτήν την τεχνική λαμβάνουμε τον μέσο όρο μιας περιοχής στην οποία κάνουμε υποδειγματοληψία. Για παράδειγμα αν εκτελέσουμε υποδειγματοληψία 1 : 2 θα λάβουμε τον μέσο όρο κάθε περιοχής 4 εικονοστοιχείων (2×2) (Σχήμα 2.7).
- Η τεχνική του max pooling : Σε αυτήν την τεχνική λαμβάνουμε την μέγιστη τιμή της περιοχής στην οποία κάνουμε υποδειγματοληψία. Όπως και στο παραπάνω παράδειγμα για μια αντίστοιχη υποδειγματοληψία θα λαμβάναμε την μέγιστη τιμή των 4 εικονοστοιχείων (Σχήμα 2.7).

Πειράματα έχουν δείξει ότι η τεχνική του max pooling είναι πιο αποδοτική από την τεχνική του average pooling [48], κάτι το οποίο είναι αναμενόμενο δεδομένου ότι θέλουμε να προωθούνται οι μέγιστες ενεργοποιήσεις μέσα στο δίκτυο και στην περίπτωση του average pooling οι μεγάλες τιμές μειώνονται λόγω του μέσου όρου. [49]



Σχήμα 2.7: Επίπεδο υποδειγματοληψίας [7]

2.3.5 Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer)

Το πλήρως συνδεδεμένο επίπεδο είναι μία παραδοσιακή αρχιτεκτονική πολλαπλών επιπέδων με νευρώνες, η οποία χρησιμοποιεί μία συνάρτηση ενεργοποίησης (συνήθως τη softmax) στην έξοδό της. Κάθε επίπεδο που ανήκει σε αυτή την αρχιτεκτονική έχει την ιδιότητα ότι κάθε νευρώνας που περιλαμβάνει, συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου.

Η έξοδος των συνελκτικών επιπέδων και των επιπέδων υποδειγματοληψίας αναπαριστά χαρακτηριστικά υψηλών στρωμάτων. Στη βασική περίπτωση, που το πρόβλημα ανήκει στην κατηγορία της ταξινόμησης (classification), ο σκοπός του πλήρως συνδεδεμένου επιπέδου

είναι να χρησιμοποιήσει αυτά τα χαρακτηριστικά έτσι ώστε να ταξινομήσει την εικόνα εισόδου σε διάφορες κλάσεις, βασιζόμενο στο σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση του μοντέλου.

2.3.6 Επίπεδο Κανονικοποίησης Παρτίδας (Batch Normalization Layer)

Τα επίπεδα κανονικοποίησης παρτίδας χρησιμοποιούνται για την κανονικοποίηση των ενεργοποιήσεων μιας δεδομένης εισόδου πριν την διάδοση της στο επόμενο επίπεδο του δικτύου.

Αν θεωρήσουμε ότι x είναι η είσοδος της κανονικοποίησης παρτίδας ενεργοποιήσεων, τότε μπορούμε να υπολογίσουμε το κανονικοποιημένο \hat{x} μέσω της ακόλουθης εξίσωσης:

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (2.3)$$

Κατά την εκπαίδευση, υπολογίζουμε τα μ_β και σ_β σε μία παρτίδα β , όπου:

$$\mu_\beta = \frac{1}{M} \sum_{i=1}^m x_i \quad \sigma_\beta^2 = \frac{1}{M} \sum_{i=1}^m (x_i - \mu_\beta)^2 \quad (2.4)$$

Θέτουμε το ϵ ίσο με μία μικρή θετική ποσότητα όπως πχ e^{-7} για να αποφύγουμε την διαίρεση με το μηδέν.

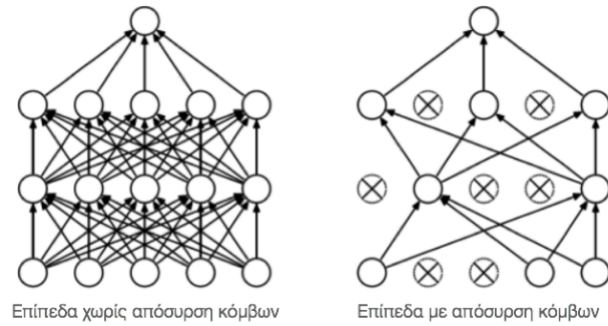
Η κανονικοποίηση παρτίδας αποδεικνύεται ότι είναι πολύ αποτελεσματική στην μείωση του αριθμού των εποχών που χρειάζεται να εκπαιδευτεί ένα νευρωνικό. Επίσης έχει το πρόσθετο πλεονέκτημα να βοηθάει στη σταθεροποίηση της εκπαίδευσης, επιτρέποντας μία μεγαλύτερη ποικιλία ρυθμών εκπαίδευσης και βαρών κανονικοποίησης.[50]

2.3.7 Επίπεδο Απόσυρσης (Dropout Layer)

Αν και δεν αποτελεί κύριο συστατικό ενός CNN, το επίπεδο απόσυρσης χρησιμοποιείται σε πολλές αρχιτεκτονικές δικτύων συνήθως πριν από το πλήρως συνδεδεμένο επίπεδο. Η τεχνική της απόσυρσης είναι μια μέθοδος αποφυγής της υπερπροσαρμογής (overfitting) στα νευρωνικά δίκτυα. Η υπερπροσαρμογή συμβαίνει όταν ένα μοντέλο ενώ φαινομενικά έχει εκπαιδευτεί εξαιρετικά καλά, στην πραγματικότητα έχει απλώς απομνημονεύσει όλα τα δεδομένα εκπαίδευσης. Το πρόβλημα αποκαλύπτεται όταν το μοντέλο δοκιμαστεί σε σύνολο δεδομένων στο οποίο δεν έχει εκπαιδευτεί ποτέ, όπου και αποτυγχάνει στη γενίκευσή του.

Η τεχνική της απόσυρσης, όπως προδίδει και η ονομασία της, αυτό που κάνει είναι ότι αποσύρει ένα ποσοστό τυχαίων νευρώνων από κάποιο επίπεδο. (Σχήμα 2.8)

Η πιθανότητα της απόσυρσης αποφασίζεται από τον χρήστη, πρόκειται δηλαδή για μια υπερπαράμετρο. Οι αποσυρόμενοι νευρώνες δε συμμετέχουν πλέον στον τρέχοντα γύρο της εκπαίδευσης και αυτό έχει ως αποτέλεσμα τον εξαναγκασμό των υπόλοιπων (ενεργών) νευρώνων του επιπέδου στον εντοπισμό νέων χαρακτηριστικών που δεν βασίζονται στους προσωρινά (καθώς μπορεί να ενεργοποιηθούν ξανά σε επόμενο γύρο) ανενεργούς νευρώνες. Όσο διαρκεί αυτή η διαδικασία, τα βάρη ανανεώνονται συνεχώς και ως αποτέλεσμα έχουμε ποιοτικότερη εκπαίδευση δηλαδή λιγότερο ευαίσθητη στην υπερπροσαρμογή. [8]



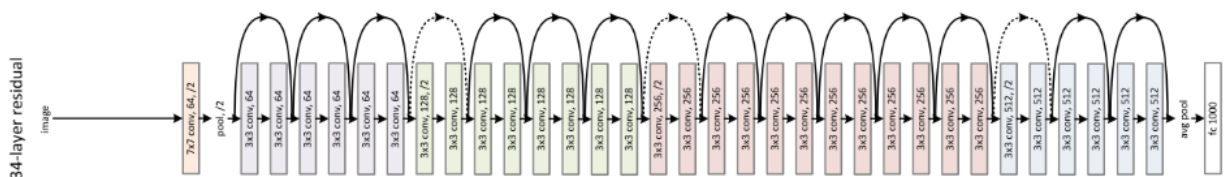
Σχήμα 2.8: Τεχνική της απόσυρσης κόμβων-νευρώνων [8]

2.4 Σύγχρονα συνελκτικά νευρωνικά δίκτυα

Στο σημείο αυτό θα παρουσιάσουμε μία γνωστή αρχιτεκτονική δικτύων που χρησιμοποιείται κατά κόρον για την επίλυση προβλημάτων ταξινόμησης και παλινδρόμησης και την οποία χρησιμοποιήσαμε και εμείς στο πειραματικό μέρος της διπλωματικής.

2.4.1 Resnet

Η ονομασία "ResNet" αποτελεί συντομογραφία της φράσης 'Residual Network'. [9] Όπως υποδηλώνει το όνομα του δικτύου, η νέα ορολογία που εισάγει είναι η «υπολειμματική» μάθηση (residual learning). Τα βαθιά συνελκτικά δίκτυα έχουν οδηγήσει σε μία σειρά ανακαλύψεων για την ταξινόμηση των εικόνων. Έτσι, με την πάροδο του χρόνου υπάρχει μία τάση να αυξάνεται το βάθος των δικτύων έτσι ώστε να επιλύουν πιο σύνθετα προβλήματα. Όμως, όσο αυξάνεται το βάθος του δικτύου η εκπαίδευση καθίσταται πιο δύσκολη. Η «υπολειμματική» μάθηση προσπαθεί να λύσει αυτό το πρόβλημα. Σε γενικές γραμμές, σε ένα βαθύ συνελκτικό νευρωνικό δίκτυο, πολλά επίπεδα (layers) στοιάζονται και εκπαιδούνται. Το δίκτυο μαθαίνει χαμηλού, μεσαίου και υψηλού επιπέδου χαρακτηριστικά στο τέλος των επιπέδων του. Στην «υπολειμματική» μάθηση πέρα από την εκμάθηση των τυπικών χαρακτηριστικών, το δίκτυο προσπαθεί να μάθει και κάποια υπολείμματα (residuals). Το υπόλειμμα μπορεί να γίνει εύκολα κατανοητό ως η αφαίρεση του χαρακτηριστικού που εξάχθηκε από την είσοδο ενός στρώματος. Το ResNet εφαρμόζει αυτή τη μορφή μάθησης χρησιμοποιώντας συνδέσεις συντόμευσης, συνδέοντας άμεσα την είσοδο του ν-οστού επιπέδου με κάποιο επόμενο επίπεδο. Έχει αποδειχθεί ότι η εφαρμογή αυτής της μορφής μάθησης στα δίκτυα καθιστά ευκολότερη την εκπαίδευσή τους και επιλύει το πρόβλημα της περιορισμένης ακρίβειας. [51]



Σχήμα 2.9: Γενική αναπαράσταση του δικτύου ResNet 34 επιπέδων.[9]

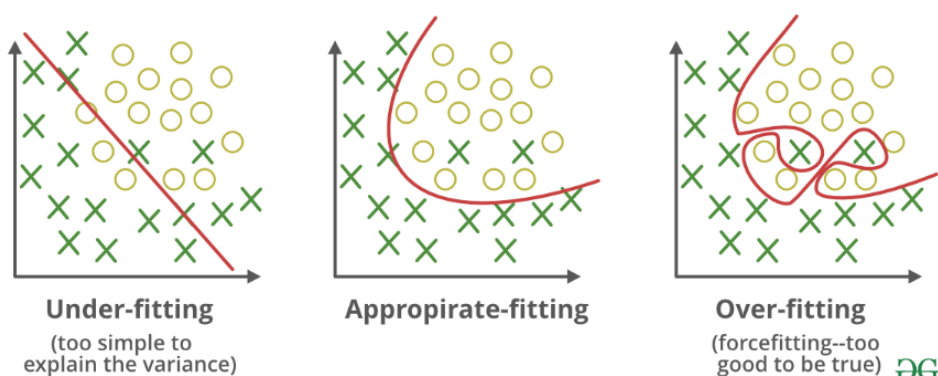
2.5 Υπερπαράμετροι νευρωνικών δικτύων

Οι υπερπαράμετροι είναι οι μεταβλητές που καθορίζουν τη δομή του δικτύου (π.χ.: αριθμός κρυφών μονάδων) και οι μεταβλητές που καθορίζουν τον τρόπο εκπαίδευσης του δικτύου (π.χ.: ρυθμός μάθησης). Ονομάζονται 'υπερ' επειδή επηρεάζουν τον τρόπο με τον οποίο θα υπολογίζονται οι παράμετροι (δηλαδή τα βάρη των συνδέσεων) και ορίζονται πριν από την εκπαίδευση του νευρωνικού. Οι βασικές υπερπαράμετροι είναι ο ρυθμός απόσυρσης (Dropout rate), ο ρυθμός εκμάθησης (Learning rate), συνάρτηση ενεργοποίησης (Activation function), ο αριθμός των εποχών (number of epochs) και το μέγεθος παρτίδας (Batch size).

Προκειμένου να γίνει πλήρως αντιληπτή η επίδραση των παρακάτω υπερπαραμέτρων, είναι σημαντικό να εξηγηθούν οι έννοιες της υπερεκπαίδευσης (overfitting) και της υποεκπαίδευσης (underfitting). Όπως γνωρίζουμε, οι παράγοντες που καθορίζουν πόσο καλά θα αποδώσει ένας αλγόριθμος μηχανικής μάθησης είναι η ικανότητά του:

- Να κάνει το σφάλμα εκπαίδευσης μικρό.
- Να μειώσει τη διαφορά μεταξύ του σφάλματος εκπαίδευσης και του σφάλματος δοκιμών.

Αυτοί οι δύο παράγοντες αντιστοιχούν στις δύο κεντρικές προκλήσεις της μηχανικής μάθησης, την υπερεκπαίδευση και την υποεκπαίδευση. Πιο συγκεκριμένα, η υπερεκπαίδευση δημιουργείται εξαιτίας της απομνημόνευσης των δεδομένων εκπαίδευσης και συνήθως οδηγεί σε κακή απόδοση στο σύνολο δοκιμών. Αυτό σημαίνει ότι η απόδοση στο σύνολο εκπαίδευσης μπορεί να είναι πολύ καλή, αλλά η απόδοση στο σύνολο δοκιμών να είναι αρκετά κακή. [52] Από την άλλη, υποεκπαίδευση παρατηρείται όταν το μοντέλο δεν είναι αρκετά σύνθετο ώστε να αποτυπώνει τη σχέση μεταξύ της εισόδου και της εξόδου, με αποτέλεσμα να μην προβλέπει σωστά την έξοδο.



Σχήμα 2.10: Παράδειγμα υποεκπαίδευσης, ισορροπημένου μοντέλου και υπερεκπαίδευσης αντίστοιχα. [10]

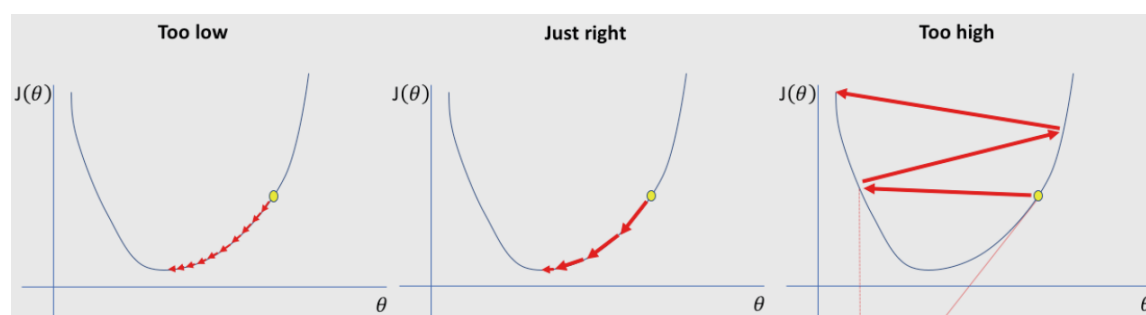
2.5.1 Ρυθμός Απόσυρσης (Dropout Rate)

Όπως έχουμε προαναφέρει στην Ενότητα 2.3.7 η απόσυρση είναι μια τεχνική κανονικοποίησης για την αποφυγή της υπερεκπαίδευσης (αύξηση της ορθότητας επικύρωσης),

αυξάνοντας έτσι τη γενικευτική ισχύ. Η πιθανότητα της απόσυρσης αποφασίζεται από τον χρήστη αν και γενικά προτιμάται μία μικρή πιθανότητα απόσυρσης της τάξης του 20-50% μιας και μια πολύ χαμηλή πιθανότητα έχει ελάχιστο αποτέλεσμα και μια πολύ υψηλή τιμή οδηγεί σε υποεκμάθηση (under-learning) από το δίκτυο. [53]

2.5.2 Ρυθμός εκμάθησης (Learning rate)

Ο ρυθμός εκμάθησης είναι μια υπερπαράμετρος που ελέγχει πόσο θα αλλάξει το μοντέλο ως απάντηση στο εκτιμώμενο σφάλμα κάθε φορά που ενημερώνονται τα βάρη του μοντέλου. Η επιλογή του ρυθμού εκμάθησης αποτελεί πρόκληση, καθώς μια πολύ μικρή τιμή μπορεί να οδηγήσει σε μια αργή διαδικασία εκπαίδευσης που μπορεί οδηγήσει στο να κολλήσει το μοντέλο σε μια σταθερή κατάσταση, ενώ μια πολύ μεγάλη τιμή μπορεί να οδηγήσει σε πολύ γρήγορη εκμάθηση ενός μη βέλτιστου συνόλου βαρών ή σε μια ασταθή διαδικασία εκπαίδευσης ή σε αδυναμία σύγκλισης στη βέλτιστη λύση. [54]



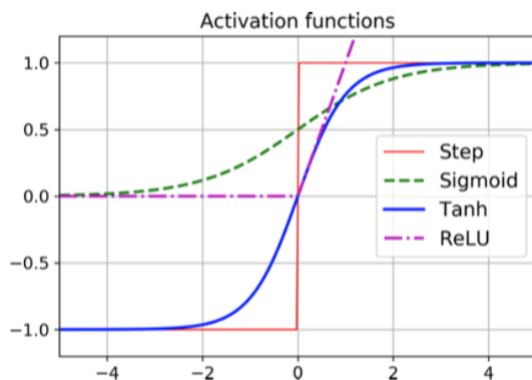
Σχήμα 2.11: Από αριστερά προς τα δεξιά: (α) Ένας μικρός ρυθμός εκμάθησης απαιτεί πολλές ενημερώσεις πριν φτάσει στο ελάχιστο σημείο, (β) Ο βέλτιστος ρυθμός εκμάθησης φτάνει γρήγορα στο ελάχιστο σημείο, (γ) Πολύ μεγάλος ρυθμός εκμάθησης προκαλεί δραστικές ενημερώσεις που οδηγούν σε αποκλινοσες συμπεριφορές. [11]

2.5.3 Συνάρτηση ενεργοποίησης (Activation function)

Με βάση όσα έχουμε αναφέρει νωρίτερα, στα επίπεδα ενεργοποίησης χρησιμοποιείται μία συνάρτηση ενεργοποίησης για την εισαγωγή μη γραμμικότητας στα μοντέλα, η οποία επιτρέπει στα μοντέλα βαθιάς μάθησης να μαθαίνουν μη γραμμικά όρια πρόβλεψης. Οι τιμές εισόδου που μετακινούνται από ένα επίπεδο σε ένα άλλο μεταβάλλονται συνεχώς σύμφωνα με τη συνάρτηση ενεργοποίησης, η οποία αποφασίζει τον τρόπο υπολογισμού των τιμών εισόδου ενός επιπέδου σε τιμές εξόδου. Οι πιο συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης είναι η ReLU, η σιγμοειδής και η softmax. (Σχήμα 2.16) [53]

2.5.4 Αριθμός εποχών (Number of epochs)

Ο αριθμός των εποχών είναι μια υπερπαράμετρος που καθορίζει τον αριθμό των φορών που ο αλγόριθμος μάθησης θα επεξεργαστεί ολόκληρο το σύνολο δεδομένων εκπαίδευσης. Μια εποχή σημαίνει ότι κάθε δείγμα στο σύνολο δεδομένων εκπαίδευσης είχε την ευκαιρία να ενημερώσει τις εσωτερικές παραμέτρους του μοντέλου. Μια εποχή αποτελείται από μία ή περισσότερες παρτίδες (batches). Συνήθως, η εκπαίδευση ενός νευρωνικού δικτύου απαιτεί



Σχήμα 2.12: Παραδείγματα συναρτήσεων ενεργοποίησης [12]

αρκετές εποχές. Ωστόσο η μεγάλη αύξηση του αριθμού των εποχών δεν σημαίνει απαραίτητα ότι το δίκτυο θα έχει καλύτερα αποτελέσματα, μιας και μία λάθος επιλογή του αριθμού των εποχών μπορεί να οδηγήσει σε υπερεκπαίδευση ή υποεκπαίδευση. [55]

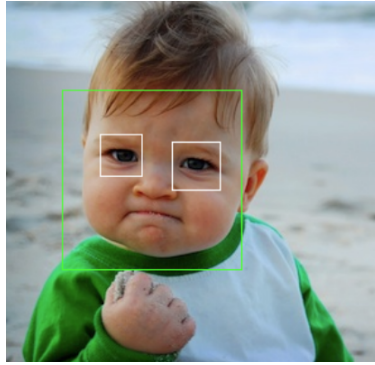
2.5.5 Μέγεθος παρτίδας (Batch size)

Το μέγεθος παρτίδας είναι μία από τις σημαντικές υπερπαραμέτρους στη μηχανική μάθηση καθώς καθορίζει τον αριθμό των δειγμάτων που θα επεξεργαστούν πριν από την ενημέρωση των εσωτερικών παραμέτρων του μοντέλου. Το μέγεθος μιας παρτίδας πρέπει να είναι μεγαλύτερο ή ίσο με 1 και μικρότερο ή ίσο με τον αριθμό των δειγμάτων στο σύνολο δεδομένων εκπαίδευσης. Όσο υψηλότερο είναι το μέγεθος παρτίδας, τόσο περισσότερο χώρο μνήμης χρειαζόμαστε.

2.6 Τυπική Προεπεξεργασία Εικόνων Προσώπων

Κατά την εκπαίδευση των νευρωνικών χρησιμοποιήσαμε αποκλειστικά εικόνες που περιέχουν πρόσωπα. Παρόλο που είναι εφικτή η εκπαίδευση και με την απευθείας τροφοδότηση των αρχικών εικόνων στο νευρωνικό, η προεπεξεργασία των εικόνων συμβάλλει στην αύξηση της ακρίβειας των αποτελεσμάτων και ως εκ τούτου τα περισσότερα σύνολα δεδομένων παραθέτουν τις εικόνες που έχουν προκύψει μετά από μερικά στάδια προεπεξεργασίας. Τα πιο συνηθισμένα στάδια περιλαμβάνουν:

- Ανίχνευση προσώπου και χαρακτηριστικών (Face and feature detection) : Όταν εργαζόμαστε σε ένα πρόβλημα ταξινόμησης προσώπου, μπορεί να θέλουμε να κάνουμε ανίχνευση προσώπου για να επικυρώσουμε (υπάρχει όντως κάποιο πρόσωπο ;), να περικυλώσουμε και να ισιώσουμε τις εικόνες μας. Η ανίχνευση προσώπου περιλαμβάνει τον εντοπισμό ενός προσώπου από μια εικόνα χρησιμοποιώντας ολόκληρη την εικόνα (προσέγγιση με βάση την εικόνα) ή την ανίχνευση ενός ή περισσότερων χαρακτηριστικών της εικόνας, όπως μύτη, μάτια, χείλη κ.λπ. (προσέγγιση βασισμένη σε χαρακτηριστικά). Η ανίχνευση προσώπου μπορεί επίσης να γίνει με βάση τα μοντέλα ενεργού σχήματος, όπως ο εντοπισμός των ορίων του κεφαλιού (bounding box). [13]
- Ευθυγράμμιση προσώπου (Face alignment) : Η ευθυγράμμιση προσώπων μπορεί



Σχήμα 2.13: Παράδειγμα ανίχνευσης προσώπου και οφθαλμών. [13]

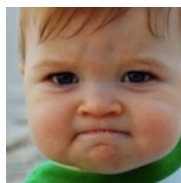
ί να οριστεί ως μια διαδικασία μετασχηματισμού διαφορετικών συνόλων σημείων από εικόνες εισόδου (συστήματα συντεταγμένων εισόδου) σε ένα άλλο σύστημα συντεταγμένων. Μπορούμε να ονομάσουμε αυτό το σύστημα συντεταγμένων ως σύστημα συντεταγμένων εξόδου και να το ορίσουμε ως το σταθερό πλαίσιο αναφοράς μας. Στόχος μας είναι να παραμορφώσουμε και να μετασχηματίσουμε όλες τις συντεταγμένες εισόδου και να τις ευθυγραμμίσουμε με τις συντεταγμένες εξόδου. Για το σκοπό αυτό, θα εφαρμόσουμε τρεις βασικούς αφινικούς μετασχηματισμούς: περιστροφή, μετατόπιση και κλιμάκωση. Με αυτόν τον τρόπο, μπορούμε να μετασχηματίσουμε τα ορόσημα του προσώπου από τα συστήματα συντεταγμένων εισόδου στο σύστημα συντεταγμένων εξόδου. Η ευθυγράμμιση προσώπων επιτρέπει τη δημιουργία αντιστοιχιών μεταξύ διαφορετικών εικόνων, έτσι ώστε οι εργασίες που θα εκτελεστούν σε διαφορετικές εικόνες να μπορούν να εκτελούνται σε κοινή βάση. [56]



Σχήμα 2.14: Παράδειγμα ευθυγράμμισης προσώπου. [13]

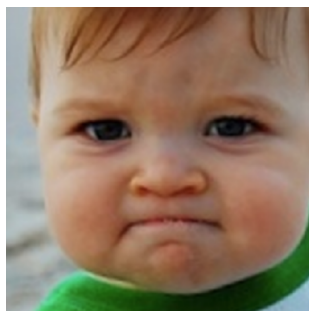
- Περικοπή προσώπου (Face cropping) : Για να βοηθήσουμε το νευρωνικό δίκτυο στην εργασία ταξινόμησης προσώπου, θα ήταν καλό να απαλλαγούμε από εξωτερικές πληροφορίες που αποσπούν την προσοχή, όπως το παρασκήνιο, τα ρούχα ή τα αξεσουάρ, το οποίο επιτυγχάνεται με την περικοπή του προσώπου από την υπόλοιπη εικόνα. Η περικοπή αυτή μπορεί χονδρικά να χωριστεί σε άνω περικοπή προσώπου και κάτω περικοπή προσώπου. Η περικοπή του άνω μέρους του προσώπου αναφέρεται στη διαδικασία φιλτραρίσματος των περιττών περιοχών από τη φωτογραφία ενός ατόμου από το κεφάλι μέχρι τους ώμους. Αντίθετα, η περικοπή κάτω μέρους του προσώπου προ-

σπαθεί να προσδιορίσει με ακρίβεια τα αυτιά και τα μάτια στην περικομμένη εικόνα ενός ατόμου για να διευκολύνει την επακόλουθη αναγνώριση προσώπου. [14]



Σχήμα 2.15: Παράδειγμα περικοπής προσώπου. [13]

- Αλλαγή μεγέθους εικόνας (Image resizing) : Ο σκοπός της αλλαγής μεγέθους της εικόνας είναι η παραγωγή μικρότερου μεγέθους δεδομένων, το οποίο επιταχύνει τον χρόνο επεξεργασίας. Προφανώς, τα διαφορετικά μεγέθη εικόνας μεταφέρουν διαφορετικές πληροφορίες γι' αυτό και πρέπει να επιλεγεί το καλύτερο μέγεθος εικόνας. [57]



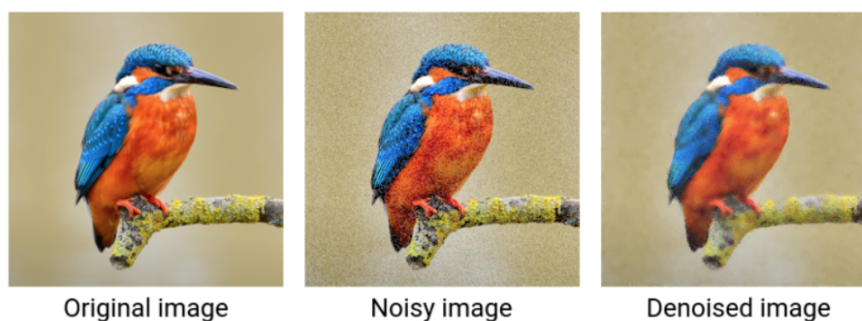
Σχήμα 2.16: Παράδειγμα αλλαγής μεγέθους εικόνας. [13]

- Κανονικοποίηση εικόνας (Image normalization) : Η διακύμανση του φωτισμού είναι μία από τις πιο σημαντικές προκλήσεις στην αναγνώριση προσώπων. Μία εικόνα με ανεξέλεγκτες συνθήκες φωτισμού περιέχει μη ομοιόμορφη αντίθεση, δηλαδή η κατανομή της έντασης/των επιπέδων του γκρι δεν είναι ίδια. Για να γίνουν αυτά τα επίπεδα ίσα ή σχεδόν ίσα χρησιμοποιούμε την τεχνική της κανονικοποίησης εικόνας, προκειμένου να διορθώσουμε τις πολύ σκούρες/ πολύ φωτεινές εικόνες. [13]



Σχήμα 2.17: Παράδειγμα κανονικοποίησης εικόνας. [13]

- Αποθορυβοποίηση και φιλτράρισμα εικόνας (Image De-noising and Filtering) : Η αποθορυβοποίηση εικόνας είναι η διαδικασία ανίχνευσης σημαντικών χαρακτηριστικών σε μια εικόνα και η βελτίωσή τους με παράλληλη καταστολή του θορύβου του παρασκηίου. Παραδείγματα χρήσιμων χαρακτηριστικών προς βελτίωση περιλαμβάνουν τις γραμμές και τις ακμές καθώς και λειτουργίες που μπορούν να συμπληρώσουν κενά για να δημιουργήσουν μια πλήρη αναπαράσταση ενός αντικειμένου στην εικόνα-στόχο. Το φιλτράρισμα εικόνας είναι κάθε τροποποίηση που μεταβάλλει κάποιο χαρακτηριστικό μιας εικόνας εισόδου για να ληφθεί μια τροποποιημένη εικόνα εξόδου. Για παράδειγμα, το χωρικό φιλτράρισμα τροποποιεί τις εντάσεις των εικονοστοιχείων. Το φασματικό φιλτράρισμα μεταβάλλει ιδιότητες όπως είναι η απόχρωση και ο κορεσμός ενώ άλλα φίλτρα μπορούν να βελτιώσουν χρονικά χαρακτηριστικά, όπως η θόλωση που προέρχεται από κίνηση. [14]



Σχήμα 2.18: Παράδειγμα αποθορυβοποίησης εικόνας. [14]

2.7 Συναρτήσεις απωλειών (Loss Functions)

Η συνάρτηση απωλειών είναι η συνάρτηση που υπολογίζει την απόσταση μεταξύ της τρέχουσας εξόδου του αλγορίθμου και της αναμενόμενης εξόδου, είναι δηλαδή μια μέθοδος αξιολόγησης του τρόπου με τον οποίο ο αλγόριθμός μας μοντελοποιεί τα δεδομένα. Μπορεί να κατηγοριοποιηθεί σε δύο ομάδες. Οι συναρτήσεις απωλειών παίζουν πολύ σημαντικό ρόλο στην εκπαίδευση των νευρωνικών δικτύων, καθώς επιβραβεύουν το δίκτυο όταν κάνει σωστές προβλέψεις και του επιβάλλουν ποινές όταν κάνει λανθασμένες. [44]. Δεδομένων των δύο διαφορετικών εργασιών που κληθήκαμε να ασχοληθούμε, χρησιμοποιήσαμε συναρτήσεις απωλειών από δύο διαφορετικές κατηγορίες, αυτές που είναι για ταξινόμηση και αυτές που είναι για παλινδρόμηση.

2.7.1 Συνάρτηση Απωλειών Προβλήματος Ταξινόμησης

Όσον αφορά το πρόβλημα ταξινόμησης δοκιμάσαμε δύο διαφορετικές συναρτήσεις απωλειών, την cross entropy και την kl divergence.

Η συνάρτηση απωλειών Cross Entropy αποτελεί συνήθη μέθοδος για την εκπαίδευση μοντέλων ταξινόμησης πολλαπλών κλάσεων. Με την μέθοδο αυτή συγκρίνεται η πιθανότητα πρόβλεψης της εξόδου του μοντέλου σε σχέση με την αναμενόμενη κλάση και υπολογίζεται το κόστος με βάση την απόσταση μεταξύ τους, το οποίο στη συνέχεια επιβάλλει την αντίστοιχη

ποινή στο μοντέλο. Το κόστος υπολογίζεται για κάθε κλάση ξεχωριστά και στη συνέχεια χρησιμοποιείται ο απλός ή ο σταθμισμένος μέσος όρος για την τελική εκτίμηση του κόστους του μοντέλου. [44] Για προβλήματα πολλαπλών (M) κλάσεων η απώλεια για κάθε κλάση ορίζεται ως εξής:

$$loss = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (2.5)$$

όπου

- y : δυαδικός δείκτης (0 ή 1) εάν η ετικέτα κλάσης c είναι η σωστή ταξινόμηση για την παρατήρηση o
- p : προβλεπόμενη πιθανότητα η παρατήρηση o να ανήκει στην κλάση c

Συγκεκριμένα, χρησιμοποιήσαμε και τις δύο μορφές της cross entropy, δηλαδή και την categorical και την sparse, ανάλογα με την μορφή των δεδομένων που είχαμε κάθε φορά (δηλαδή σε one hot encoded μορφή ή όχι αντίστοιχα).

Όσον αφορά την άλλη συνάρτηση απωλειών που χρησιμοποιήσαμε, η συνάρτηση απωλειών Kullback-Leibler Divergence ποσοτικοποιεί το πόσο διαφέρει μια κατανομή πιθανότητας από μια άλλη κατανομή πιθανότητας. Η απόκλιση KL είναι η σχετική εντροπία ή η διαφορά μεταξύ της cross entropy και της εντροπίας ή κάποια απόσταση μεταξύ της πραγματικής κατανομής πιθανοτήτων και της προβλεπόμενης κατανομής πιθανοτήτων. Είναι ίση με 0 όταν η προβλεπόμενη κατανομή πιθανότητας είναι ίδια με την πραγματική κατανομή πιθανότητας και ορίζεται ως:

$$loss = - \sum_x p(x) \log(q(x)/p(x)) \quad (2.6)$$

2.7.2 Συνάρτηση Απωλειών Προβλήματος Παλινδρόμησης

Για το πρόβλημα παλινδρόμησης χρησιμοποιήσαμε την συνάρτηση απωλειών Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error) που είναι η πιο συχνά χρησιμοποιούμενη συνάρτηση απωλειών στα προβλήματα παλινδρόμησης. Η συγκεκριμένη συνάρτηση απωλειών είναι ο μέσος όρος των τετραγωνικών διαφορών μεταξύ των πραγματικών και των προβλεπόμενων τιμών, ή πιο αναλυτικά:

$$loss = \frac{1}{N} \sum_{i=0}^N (y - \hat{y}_i)^2 \quad (2.7)$$

όπου \hat{y} είναι η προβλεπόμενη τιμή.

2.8 Μετρικές Αξιολόγησης Μοντέλων

Στο σημείο αυτό αναφέρουμε την επιλογή των μετρικών αξιολόγησης των μοντέλων μας. Οι μετρικές αξιολόγησης έχουν μεγάλη συμβολή στο έργο της αύξησης της τελικής απόδοσης του μοντέλου καθώς με χρήση αυτών παρακολουθούμε την εκπαίδευση και κάνουμε

τις απαραίτητες αλλαγές των παραμέτρων των μοντέλων έτσι ώστε να πετύχουμε το καλύτερο δυνατό αποτέλεσμα. Εξίσου σημαντική είναι η επιλογή της μετρικής να γίνεται με γνώμονα την επιλογή των μετρικών που χρησιμοποιούν και άλλες ερευνητικές ομάδες για την δεδομένη εργασία και το δεδομένο σύνολο δεδομένων προκειμένου να είναι εφικτές οι συγκρίσεις της επίδοσης των μοντέλων στο πλαίσιο της ευρύτερης ερευνητικής δραστηριότητας. [44] Στην δική μας εργασία ασχοληθήκαμε με δύο προβλήματα, ένα ταξινόμησης και ένα παλινδρόμησης, κάθε ένα εκ των οποίων απαιτούσε διαφορετική μετρική αξιολόγησης, όπως περιγράφουμε λεπτομερέστερα παρακάτω.

2.8.1 Μετρικές Αξιολόγησης Προβλήματος Ταξινόμησης

Για το πρόβλημα της ταξινόμησης των συναισθημάτων κάναμε χρήση τριών μετρικών: της ορθότητας (accuracy), του f1 score και της μέσης ορθότητας (average accuracy). Προκειμένου να γίνουν πιο εύκολα κατανοητοί οι τύποι των παραπάνω μετρικών αναφέρουμε τι σημαίνουν οι παρακάτω συντομογραφίες.

- TP: True Positive, το σύνολο των προβλέψεων που είναι σωστές και ανήκουν στην κλάση για την οποία υπολογίζουμε την ορθότητα.
- TN: True Negative, το σύνολο των προβλέψεων που είναι σωστές και δεν ανήκουν στην κλάση για την οποία υπολογίζουμε την ορθότητα.
- FP: False Positive, το σύνολο των προβλέψεων που είναι λανθασμένες και ανήκουν στην κλάση για την οποία υπολογίζουμε την ορθότητα.
- FN: False Negative, το σύνολο των προβλέψεων που είναι λανθασμένες και δεν ανήκουν στην κλάση για την οποία υπολογίζουμε την ορθότητα.

Δεδομένου των παραπάνω, ο τύπος με τον οποίο υπολογίζεται η ορθότητα έχει ως εξής:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

Η ορθότητα παίρνει τιμές στο διάστημα [0, 1] με την τιμή 1 να δείχνει την καλύτερη απόδοση και την τιμή 0 τη χειρότερη. Παρά το γεγονός ότι η ορθότητα είναι η απλούστερη και πιο συχνά χρησιμοποιούμενη μετρική για προβλήματα ταξινόμησης, δεν μπορεί να δείξει την απόδοση του μοντέλου για κάθε κλάση ξεχωριστά και ειδικά στη περίπτωση σημαντικά μη ισορροπημένου συνόλου δεδομένων, όπου μία κλάση μπορεί να έχει μεγάλο πλήθος παραδειγμάτων, δεν είναι αρκετά ενδεικτική για την επιτυχία του μοντέλου σε κλάσεις με μικρότερο αριθμό παραδειγμάτων. [44]

Η επόμενη μετρική που χρησιμοποιήσαμε είναι το f1 score. Προκειμένου να ορίσουμε αυτή την μετρική, πρέπει πρώτα να αναφέρουμε τις δύο παρακάτω μετρικές:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \quad (2.9)$$

όπου το Precision ονομάζεται Ακρίβεια και αφορά πόσα από τα παραδείγματα που ο ταξινομητής έχει ταξινομήσει ως θετικά είναι πραγματικά θετικά και το Recall ονομάζεται

Ανάκληση και αφορά πόσα από τα θετικά παραδείγματα κατάφερε να βρει ο ταξινομητής. Συνδυάζοντας αυτές τις δύο μετρικές προκύπτει το f1 score ως εξής :

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{2.10}$$

Αντίστοιχα με την ορθότητα, το F1 Score παίρνει και αυτό τιμές στο διάστημα [0, 1]. Ο τύπος F1 Score που χρησιμοποιήσαμε ονομάζεται Macro F1 Score και ουσιαστικά υπολογίζεται από τις μη σταθμισμένες τιμές των Precision και Recall, οι οποίες δεν εξαρτώνται από το πλήθος των παραδειγμάτων κάθε κλάσης. Η μετρική αυτή δεν επηρεάζεται από τον πληθυσμό της κάθε κλάσης και μας δίνει μία καλύτερη εικόνα της απόδοσης του ταξινομητή ακόμα και για τις κλάσεις που έχουν μικρό πλήθος παραδειγμάτων, συνεπώς σε μη ισορροπημένα σύνολα δεδομένων καθίσταται πιο χρήσιμο από την ακρίβεια.

Τέλος, χρησιμοποιήσαμε την μετρική της μέσης ορθότητας. Προκειμένου να υπολογιστεί η συγκεκριμένη μετρική χρησιμοποιήσαμε την έννοια του πίνακα σύγχυσης (confusion matrix) η δομή του οποίου φαίνεται στο σχήμα 2.19

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Σχήμα 2.19: Η ερμηνεία των κελιών ενός confusion matrix.[15]

Δεδομένου λοιπόν του πίνακα σύγχυσης, η μέση ορθότητα ορίζεται ως η μέση τιμή της διαγωνίου του, αφού πρώτα έχουν υποστεί όλες οι τιμές κανονικοποίηση. Η μέση ορθότητα ήταν η μετρική στην οποία δώσαμε την μεγαλύτερη βαρύτητα προκειμένου να ελέγξουμε την απόδοση του εκάστοτε μοντέλου στο πρόβλημα της ταξινόμησης.

2.8.2 Μετρικές Αξιολόγησης Προβλήματος Παλινδρόμησης

Όσον αφορά το πρόβλημα της παλινδρόμησης χρησιμοποιήσαμε τις μετρικές Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error), και τον Συντελεστή Pearson.

Αναφορικά με τις παραπάνω μετρικές, η πρώτη αντιπροσωπεύει το μέσο όρο της τετραγωνικής διαφοράς μεταξύ της αρχικής και της προβλεπόμενης τιμής στο σύνολο δεδομένων, δηλαδή μετράει τη διακύμανση των υπολειμμάτων. Ο τύπος του Μέσου Τετραγωνικού Σφάλματος έχει ως εξής :

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \tag{2.11}$$

όπου y_i η πραγματική τιμή και \hat{y} η εκτιμώμενη τιμή. Η μετρική αυτή παίρνει τιμές

στο $[0, \infty)$ με το 0 να είναι η ιδανική αφού όλες οι εκτιμώμενες τιμές αντιστοιχίζονται στις πραγματικές τιμές, αντιπροσωπεύει δηλαδή την βέλτιστη συσχέτιση.

Από την άλλη, ο συντελεστής συσχέτισης Pearson είναι ο πιο συχνά χρησιμοποιημένος δείκτης συσχέτισης και ορίζεται σύμφωνα με τον παρακάτω τύπο :

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.12)$$

Ο δείκτης συσχέτισης r του Pearson αποτελείται από τα κάτωθι στοιχεία :

1. Από ένα θετικό ή αρνητικό πρόσημο (το θετικό παραλείπεται)
 - Όταν αυξάνεται η μια μεταβλητή και ταυτόχρονα αυξάνεται και η άλλη, τότε έχουμε θετική συσχέτιση (θετικό πρόσημο)
 - Όταν μειώνεται η μια μεταβλητή και μειώνεται και η άλλη, τότε έχουμε θετική συσχέτιση (θετικό πρόσημο)
 - Όταν αυξάνεται η μια μεταβλητή και μειώνεται η άλλη, τότε έχουμε αρνητική συσχέτιση (αρνητικό πρόσημο)
2. Από μια αριθμητική τιμή που κυμαίνεται από -1.00 έως 1.00. Μία συσχέτιση θεωρείται υψηλή όταν κυμαίνεται από 0.80 μέχρι 1.00. [58]

Στην εργασία μας χρειάστηκε να υπολογίζουμε κάθε φορά δύο συντελεστές, έναν για το σθένος και έναν για την διέγερση.

2.9 Επαύξηση Δεδομένων (Data Augmentation)

Η επαύξηση δεδομένων είναι μία τεχνική που χρησιμοποιείται για την αύξηση του όγκου των δεδομένων με την προσθήκη ελαφρώς τροποποιημένων αντιγράφων ήδη υπάρχοντων δεδομένων ή νέων συνθετικών δεδομένων που έχουν δημιουργηθεί από υπάρχοντα δεδομένα [59]. Οι σύγχρονοι αλγόριθμοι βαθιάς μάθησης, όπως το συνελκτικό νευρωνικό δίκτυο (CNN), μπορούν να μάθουν χαρακτηριστικά που είναι αναλλοίωτα ως προς την θέση τους στην εικόνα. Παρ' όλα αυτά, η επαύξηση μπορεί να βοηθήσει περαιτέρω σε αυτήν την ανεξάρτητη, ως προς τον μετασχηματισμό, προσέγγιση για τη μάθηση και μπορεί να βοηθήσει το μοντέλο σε χαρακτηριστικά μάθησης που είναι επίσης αμετάβλητα σε μετασχηματισμούς όπως αριστερά-προς-δεξιά κλπ.

Η επαύξηση των δεδομένων των εικόνων συνήθως εφαρμόζεται μόνο στο σύνολο δεδομένων εκπαίδευσης και όχι στο σύνολο δεδομένων επικύρωσης ή δοκιμής. Αυτό διαφέρει από την προεπεξεργασία δεδομένων, όπως η αλλαγή μεγέθους εικόνας και η κλιμάκωση πίξελ, η οποία πρέπει να εκτελείται με συνέπεια σε όλα τα σύνολα δεδομένων που αλληλεπιδρούν με το μοντέλο. [55]

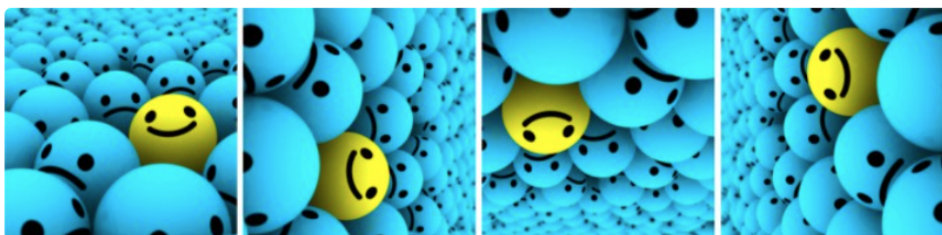
Μερικοί από τους πιο βασικούς μετασχηματισμούς στην επαύξηση των δεδομένων είναι οι παρακάτω :

- Αναστροφή (Flip) : Μπορούμε να αναστρέψουμε τις φωτογραφίες από αριστερά προς τα δεξιά και από πάνω προς τα κάτω. Θέλει προσοχή ωστόσο αυτός ο μετασχηματισμός μιας και θα πρέπει να λαμβάνουμε υπόψιν το είδος των εικόνων που θέλουμε να αντιστρέψουμε. Παραδείγματος χάρη, στην περίπτωση μας που χρησιμοποιούμε εικόνες προσώπων, μία αναστροφή από πάνω προς τα κάτω δεν θα είχε νόημα και θα δυσκόλευε την σωστή εκπαίδευση του νευρωνικού μιας και πλέον δεν θα έδειχνε ένα πρόσωπο.



Σχήμα 2.20: Από αριστερά, έχουμε την αρχική εικόνα, ακολουθούμενη από την εικόνα που έχει αναστραφεί οριζόντια, και στη συνέχεια την εικόνα που έχει αναστραφεί κάθετα. [16]

- Περιστροφή (Rotation) : Στην περιστροφή, περιστρέφουμε την εικόνα κατά έναν μικρό αριθμό μοιρών, συνήθως στο διάστημα $[-10,10]$.



Σχήμα 2.21: Οι εικόνες περιστρέφονται κατά 90 μοίρες δεξιόστροφα σε σχέση με την προηγούμενη, καθώς κινούμαστε από αριστερά προς τα δεξιά. [16]

- Μεταβολή φωτεινότητας (Brightness range) : Μπορούμε να μεταβάλλουμε την φωτεινότητα μιας εικόνας μετατρέποντας την σε πιο σκοτεινή ή πιο σκούρα, ανάλογα με το εύρος μεταβολής φωτεινότητας το οποίο θα διαλέξουμε. Το εύρος αποτελείται από δύο συνήθως δεκαδικούς αριθμούς που καθορίζουν τις επιτρεπτές τιμές μεταβολής της φωτεινότητας. Στο Keras, η τιμή 1.0 ισούται με την ουδέτερη φωτεινότητα, επομένως αν διαλέξουμε εύρος φωτεινότητας κάτω από 1, η εικόνα θα αρχίσει να σκουραίνει ενώ πάνω από το 1 θα αρχίσει να γίνεται πιο φωτεινή.



Σχήμα 2.22: Από αριστερά προς τα δεξιά, η αρχική εικόνα και η εικόνα με μεταβολή φωτεινότητας μεγαλύτερη του 1. [17]

Ανάλυση συναισθημάτων

Η μελέτη των ανθρώπινων συναισθημάτων έχει προχωρήσει σημαντικά τα τελευταία χρόνια σε σύγκριση με τις πρώτες θεωρητικές προσεγγίσεις στις αρχές του αιώνα. Παράλληλες εξελίξεις σε διαφορετικά πεδία της ψυχολογίας (π.χ. κλινική, κοινωνική, εξελικτική) συγκλίνουν στην σημασία που αποδίδουν στη μελέτη των συναισθημάτων ως αυτόνομο πεδίο έρευνας. Ένα μεγάλο σώμα κοινωνικών ψυχολογικών μελετών για τα συναισθήματα έχει δείξει την ύπαρξη ενός περιορισμένου αριθμού ('βασικών') συναισθημάτων τα οποία παρατηρεί κανείς στην έκφραση του προσώπου και τις συγκινησιακές αντιδράσεις ανθρώπων σε όλο τον πλανήτη. Οι κύριοι υποστηρικτές της θεωρίας ακολουθούν την ψυχο-εξελικτική προσέγγιση και εστιάζουν σε έρευνες που δείχνουν ότι τα βασικά συναισθήματα (όπως χαρά, λύπη, θυμός, κ.α.) χαρακτηρίζονται από μια συγκεκριμένη δομή και αποτελούν ένα υπόδειγμα που συνδυάζει συγκεκριμένες ψυχοφυσιολογικές αντιδράσεις με συγκεκριμένες εκφράσεις του προσώπου και κέντρα του φλοιού του εγκεφάλου.

Οι πρόσφατες εξελικτικές θεωρίες συναισθήματος θεωρούν, τα συναισθήματα ως έμφυτες απαντήσεις σε κάποια ερεθίσματα. Οι εξελικτικοί θεωρητικοί αναγνωρίζουν, ακόμη, την επιρροή της σκέψης και της μάθησης στο συναίσθημα. Οι σημαντικότερες προσεγγίσεις που επικρατούν στις εξελικτικές θεωρίες συναισθήματος είναι η Θεωρία των Βασικών (Διακριτών) Συναισθημάτων, όπως αυτή διατυπώθηκε πρώτα από τον Δαρβίνο και στην συνέχεια από τον Paul Ekman, η οποία περιγράφεται αναλυτικά παρακάτω, και η Διαστατική Προσέγγιση η οποία υποστηρίζει ότι το συναίσθημα έχει δύο ή και τρεις βασικές κρυμμένες διαστάσεις κατά μήκος των οποίων βρίσκονται όλα τα συναισθήματα. Οι πιο κοινές διαστάσεις είναι το σθένος (valence, που κυμαίνεται από την ευτυχία έως τη λύπη και η διέγερση arousal, που κυμαίνεται από την ηρεμία έως την ένταση). Η τρίτη διάσταση που αναφέρεται λιγότερο είναι η κυριαρχία, που κυμαίνεται από τον έλεγχο του συναισθήματος έως το μη ελεγχόμενο συναίσθημα. Μια από τις σημαντικότερες θεωρίες που μελέτησαν το συναίσθημα ως προς τις διαστάσεις του είναι η Θεωρία Συναισθημάτων ως σημεία πάνω σε διπολικούς άξονες (Τροχός Συναισθημάτων της Whissel και Θεωρία Συναισθημάτων του Plutchik). [60]

3.1 Θεωρία των Διακριτών Συναισθημάτων

Η Θεωρία αυτή στηρίζεται στις εξελικτικές θεωρίες που αντιλαμβάνονται τα συναισθήματα ως διακριτές οντότητες, ανεξάρτητες μεταξύ τους.[61] Οι εξελικτικοί θεωρητικοί πιστεύουν ότι όλοι οι άνθρωποι πολιτισμοί μοιράζονται ορισμένα συναισθήματα, τα οποία είναι διακριτά,

βασικά και παγκόσμια, όπως είναι η ευτυχία, η περιφρόνηση, η έκπληξη, η αποστροφή, ο θυμός, ο φόβος, και η θλίψη. Θεωρούν δε ότι όλα τα άλλα συναισθήματα προκύπτουν από τη μίξη και τις διαφορετικές εντάσεις αυτών των πρωταρχικών βασικών συναισθημάτων. Επίσης, το καθένα από τα βασικά συναισθήματα θεωρείται ότι έχει ένα μοναδικό πρότυπο φυσιολογικής διέγερσης, μια μοναδική έκφραση συμπεριφοράς, ένα μοναδικό τρόπο οργάνωσης των γνώσεων και της αντίληψης.

3.1.1 Η Θεωρία του Δαρβίνου

Ο Charles Darwin (1874) υπήρξε ο πρώτος εμπνευστής της θεωρίας των βασικών συναισθημάτων. Αρχικά, ασχολήθηκε με τις μορφές έκφρασης των βασικών συναισθημάτων και υποστήριξε ότι οι εκφράσεις του προσώπου δεν είναι αποτέλεσμα εκμάθησης αλλά είναι βιολογικά προκαθορισμένες. Θεώρησε ότι η έκφραση των βασικών συναισθημάτων είναι αναγκαία για την επιβίωση και εξέλιξη του ανθρώπινου είδους και επιτελείται με αντιδράσεις προσαρμοστικού χαρακτήρα. [61]

Συγκεκριμένα ο Δαρβίνος ανέλυσε λεπτομερώς γύρω στα δώδεκα βασικά συναισθήματα, θετικά και αρνητικά, περιγράφοντας για το καθένα τις χαρακτηριστικές εκφράσεις του προσώπου καθώς και τους προσαρμοστικούς μηχανισμούς αντίδρασης προς το εκάστοτε περιβαλλοντικό ερέθισμα. Η προσαρμοστική τους λειτουργία εντοπίζεται, αφενός με το να διευκολύνουν την επικοινωνία των ατόμων με το κοινωνικό τους περιβάλλον, αφετέρου με την ρύθμιση των συναισθημάτων από τα ίδια τα άτομα προκειμένου να ανταποκριθούν στις απαιτήσεις του περιβάλλοντος

3.1.2 Η Θεωρία των Βασικών Συναισθημάτων του Paul Ekman

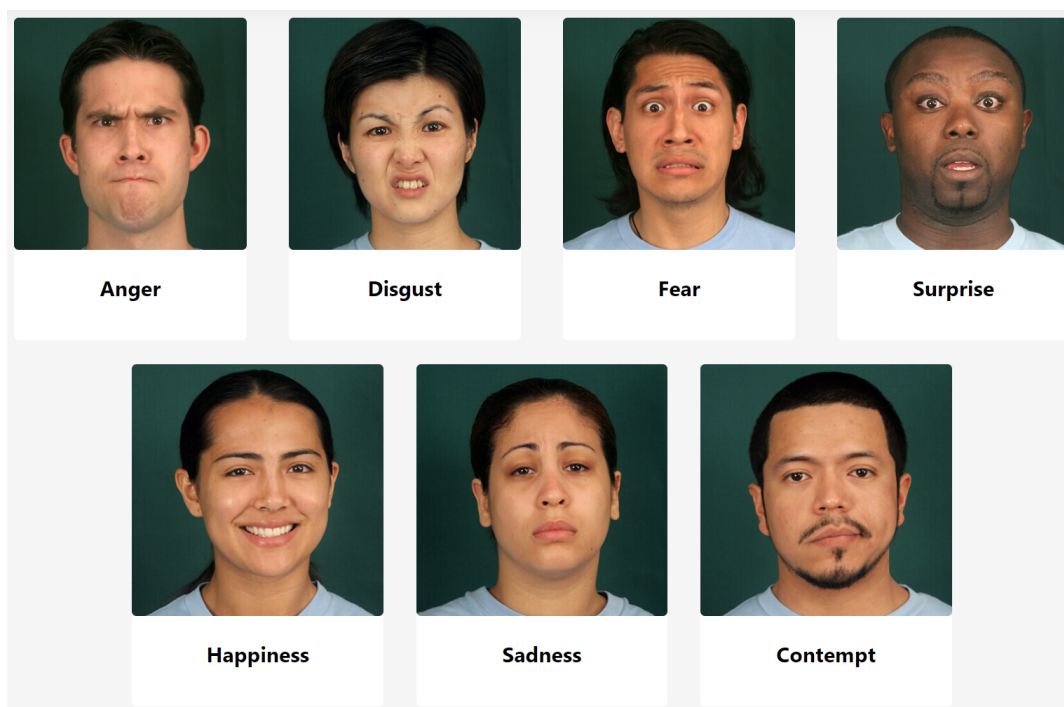
Το βιβλίο του Δαρβίνου *Η έκφραση των συγκινήσεων στους ανθρώπους και τα ζώα* (1872) γίνεται αντικείμενο εξέτασης και επανεκδίδεται με επιμέλεια του Paul Ekman [62] πρωτοπόρου μελετητή και υπέρμαχου της θεωρίας του «οικουμενικού» χαρακτήρα των συναισθημάτων. Ο Paul Ekman (1972) [63], εμπνευσμένος από την προσέγγιση του Δαρβίνου, υποστηρίζει ότι οι κινήσεις των μυών του προσώπου είναι τα βασικά στοιχεία των εκφράσεων του προσώπου και ότι η σχέση μεταξύ αυτών των κινήσεων και των συναισθημάτων είναι οικουμενική. Πιο συγκεκριμένα, την περίοδο 1967-1968 έδειξε φωτογραφίες προσώπων να εκφράζουν διάφορα συναισθήματα σε ανθρώπους από διάφορα μέρη του κόσμου, όπως Βορειοαμερικανούς, Ιάπωνες και μέλη της φυλής Φορέ στην Πάπουα Νέα Γουινέα, και βρήκε ότι όλοι αναγνώριζαν τα ίδια συναισθήματα σε κάθε πρόσωπο. [63] [64] Ακόμη, κινηματογράφησε τις εκφράσεις Αμερικανών και Ιαπώνων φοιτητών ενώ έβλεπαν ταινίες και παρατήρησε τις ίδιες εκφράσεις και στις δύο ομάδες. Θεώρησε ότι τα αποτελέσματα αυτά δείχνουν ότι η έκφραση των συναισθημάτων είναι σταθερή στους διάφορους πολιτισμούς και απέδωσε αυτή την προφανή οικουμενικότητά στην εξέλιξη, τους έμφυτους μηχανισμούς του εγκεφάλου ή σε κοινωνικές αναπτυξιακές διεργασίες. Έτσι, ο Ekman σημειώνει ότι ενώ οι καθολικές εκφράσεις δεν αποδεικνύουν απόλυτα την θεωρία του Δαρβίνου, εντούτοις παρέχονται ισχυρές ενδείξεις για την δυνατότητα αυτή. Καταλήγει λοιπόν στο συμπέρασμα ότι, η έκφραση μέσω του προσώπου μερικών βασικών συναισθημάτων είναι έμφυτη. Με αυτόν τον τρόπο, ο Ekman επιβεβαιώνει ερευνητικά την ύπαρξη τουλάχιστον έξι βασικών συναισθημάτων: Της

χαράς, της έκπληξης, του φόβου, της λύπης, του θυμού, της απέχθειας.

Ο Ekman μαζί με τον Wallace Friesen χρησιμοποίησαν τα ευρήματά τους για να δημιουργήσουν το Σύστημα Κωδικοποίησης Ενεργειών Προσώπου (FACS) , ένα πλήρες ευρετήριο εκφράσεων του προσώπου. Εφάρμοσαν το FACS στη μελέτη της αλλαγής των εκφράσεων στα άτομα με ψυχιατρικές διαταραχές και υποστήριξαν ότι όσοι πάσχουν από κατάθλιψη και σχιζοφρένεια δεν μπορούν να αναγνωρίσουν συγκεκριμένα συναισθήματα. Σήμερα, το FACS παραμένει η πιο διαδεδομένη μέθοδος για την κατηγοριοποίηση των εκφράσεων του προσώπου αν και το 2006 μια έρευνα έδειξε ότι οι Δυτικοευρωπαίοι και οι Ασιάτες χρησιμοποιούν διαφορετική στρατηγική για να αποκωδικοποιήσουν τις εκφράσεις του προσώπου.[65] [66] [67]

Όπως αναφέραμε, το μοντέλο αναπαράστασης συναισθημάτων το οποίο κατηγοριοποιεί όλα τα συναισθήματα σε επτά βασικές κατηγορίες είναι το πιο ευρέως διαδεδομένο στις έρευνες αναγνώρισης συναισθήματος. Οι 7 αυτές κατηγορίες είναι ο Θυμός, η Αηδία, ο Φόβος, η Χαρά, η Λύπη, η Έκπληξη και η Περιφρόνηση.[18] Υπάρχουν στοιχεία ότι τα 7 αυτά συναισθήματα συναντώνται σε όλους τους ανθρώπους ανεξάρτητα την ηλικία, την εθνικότητα, το φύλο και οτιδήποτε μπορεί να διαχωρίσει τους ανθρώπους. Ωστόσο, τα συναισθήματα αυτά εκφράζονται από τους ανθρώπους σε διάφορες εντάσεις και με διάφορους τρόπους στη καθημερινή ζωή, με αποτέλεσμα η κάθε κατηγορία να περιλαμβάνει μία πολύ μεγάλη γκάμα συναισθημάτων. Το γεγονός αυτό δείχνει ότι το μοντέλο αυτό δεν μπορεί να περιγράψει με ακρίβεια ένα συναίσθημα αλλά το εντάσσει σε ένα ευρύτερο πλαίσιο συναισθημάτων.[44]

















Το σχήμα 3.1 απεικονίζει τα 7 βασικά συναισθήματα με βάση την θεωρία του Ekman.



Σχήμα 3.1: Επτά Βασικά Συναισθήματα κατά Ekman [18]

3.1.3 Σύστημα Κωδικοποίησης Ενεργειών Προσώπου (FACS)

Όπως αναφέραμε πριν, ο Ekman μαζί με τον Wallace Friesen ανέπτυξαν το 'Σύστημα Κωδικοποίησης της Δράσης του Προσώπου' (Facial Action Coding System, FACS) για την περιγραφή των εκφράσεων του προσώπου από μονάδες δράσης (action units, AU). Συγκεκριμένα, οι μονάδες δράσης είναι οι θεμελιώδεις ενέργειες των μεμονωμένων μυών ή των ομάδων μυών. Από τα 44 FACS AU που όρισαν, τα 30 AU σχετίζονται ανατομικά με τις συσπάσεις συγκεκριμένων μυών του προσώπου: 12 είναι για το άνω πρόσωπο, και 18 είναι για το κάτω πρόσωπο, οι οποίες μπορούν να εμφανιστούν είτε μεμονωμένα είτε σε συνδυασμό. Αν και ο αριθμός των μονάδων ατομικής δράσης είναι σχετικά μικρός, έχουν παρατηρηθεί περισσότεροι από 7.000 διαφορετικοί συνδυασμοί AU. [55] [68]

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Σχήμα 3.2: Ενδεικτικά παραδείγματα AU [19]

Με σύνθεση μονάδων δράσης μπορεί να προκύψει κάθε ένα από τα επτά βασικά συναισθήματα που αναφέρθηκαν παραπάνω, όπως φαίνεται και από το σχήμα 3.3 που ακολουθεί.

Basic expressions	Involved Action Units
Surprise	AU 1, 2, 5, 15, 16, 20, 26
Fear	AU 1, 2, 4, 5, 15, 20, 26
Disgust	AU 2, 4, 9, 15, 17
Anger	AU 2, 4, 7, 9, 10, 20, 26
Happiness	AU 1, 6, 12, 14
Sadness	AU 1, 4, 15, 23

Σχήμα 3.3: Έξι από τα επτά βασικά συναισθήματα με σύνθεση μονάδων δράσης [20]

3.2 Διαστατικές Αναπαραστάσεις

Μία εναλλακτική μέθοδος αναπαράστασης για τον χαρακτηρισμό συναισθηματικών καταστάσεων είναι αυτή των διαστατικών αναπαραστάσεων. Κύριο χαρακτηριστικό της συγκεκριμένης αναπαράστασης είναι ότι είναι συνεχής, σε αντίθεση με τις διακριτές κατηγορίες τις οποίες έχουμε προαναφέρει. Η διαστατική αναπαράσταση εστιάζει σε μετρήσιμες παραμέτρους που παρατηρούνται σε συναισθηματικά επεισόδια και στο πώς αυτές συσχετίζονται. Η συσχέτιση επιτυγχάνεται με την απεικόνιση κάθε συναισθήματος σαν σημείο σε πολυδιάστατο χώρο. Έχουν προταθεί ποικίλα ζευγάρια ή τριάδες από διαστάσεις ως ικανές για την περιγραφή συναισθημάτων. [69]

Ιστορικά, ο πρώτος που εισήγαγε την ιδέα της διαστατικής απεικόνισης ήταν ο Wundt το 1903.[70] Την επανεισήγαγε ο Schlossberg περί τα μέσα του αιώνα προτείνοντας αρχικά δύο διαστάσεις για τον χαρακτηρισμό εκφράσεων του προσώπου, την απόλαυση/δυσάρεσκεια (pleasure/displeasure) και την προσοχή/απόρριψη (attention/rejection). [71] Στην συνέχεια εισήγαγε και μία τρίτη διάσταση, αυτή της διέγερσης (arousal). [72]

Μεταγενέστεροι ερευνητές, όπως ο Russell και η Whissell, πρότειναν την απεικόνιση των συναισθημάτων σε δύο διαστάσεις και μάλιστα σαν σημεία πάνω σε έναν κυκλικό δίσκο. [73] [74] Άλλοι ερευνητές όπως ο Plutchik [75] και ο Daly [76] πρότειναν τρισδιάστατες απεικονίσεις που οδήγησαν σε κωνικά μοντέλα. Και οι δύο εισήγαγαν ως τρίτη διάσταση αυτή της έντασης. Ο Daly στηρίχθηκε στις υπάρχουσες δύο διαστάσεις ενώ ο Plutchik πρότεινε ένα υβριδικό μοντέλο που συνδύαζε την ιδέα των βασικών συναισθημάτων με αυτή της διαστατικής αναπαράστασης.

3.2.1 Σθένος και Διέγερση (Valence and Arousal)

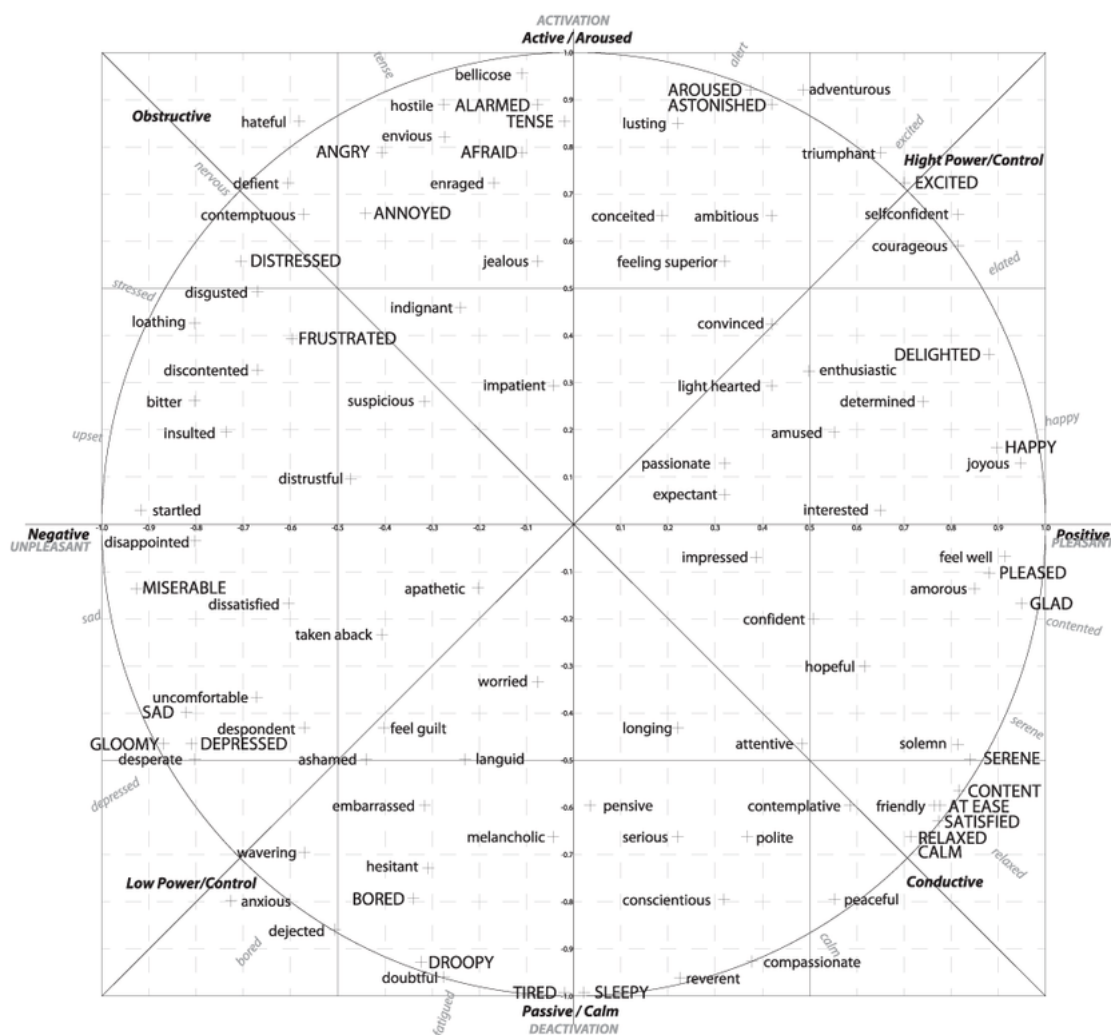
Όπως ήδη αναφέραμε, ο James Russell το 1980 διατύπωσε την αντίληψη ότι το κάθε συναίσθημα έχει δύο διαστάσεις. Το μοντέλο αυτό προτείνει ότι όλες οι συναισθηματικές καταστάσεις προκύπτουν από τις γνωστικές ερμηνείες των νευρικών αισθήσεων του πυρήνα που είναι το προϊόν δύο ανεξάρτητων νευροφυσιολογικών συστημάτων. Αυτό το μοντέλο βρίσκεται σε αντίθεση με τις θεωρίες των βασικών συναισθημάτων, οι οποίες θεωρούν ότι ένα διακριτό και ανεξάρτητο νευρικό σύστημα προάγει κάθε συναίσθημα. [55]

Σε αυτό το διδιάστατο μοντέλο η μία διάσταση (η οριζόντια) είναι το σθένος (valence) και η άλλη (η κάθετη) είναι η διέγερση (arousal) ποσότητες οι οποίες ορίζονται όπως παρακάτω.

- **Σθένος** : Κοινό χαρακτηριστικό των διαφόρων συναισθηματικών καταστάσεων αποτελεί το γεγονός ότι είναι αποτέλεσμα της επίδρασης που έχουν στο άτομο οι συνάνθρωποι του ή οι διάφορες καταστάσεις που βιώνει. Η επίδραση αυτή αποτιμάται από την μεριά του χρήστη είτε ως θετική είτε ως αρνητική, εκφράζει δηλαδή πόσο ευχάριστο ή δυσάρεστο είναι ένα ερέθισμα. Όπως φαίνεται και στο Σχήμα 3.4, τα ευχάριστα συναισθήματα είναι αντιδιαμετρικά των δυσάρεστων στον άξονα αυτό. [77] [78]
- **Διέγερση** : Ερευνητές έχουν επισημάνει το γεγονός ότι το άτομο, βιώνοντας τις διάφορες συναισθηματικές καταστάσεις, παρουσιάζει τάση συγκεκριμένων αντιδράσεων. Σύμφωνα με αυτή την παρατήρηση έγινε προσπάθεια να χαρακτηριστεί μία συναι-

σθηματική έκφραση με βάση την ενεργοποίηση του προσώπου και του σώματος. Η διάσταση αυτή εκφράζει την ενεργοποίηση του συμπαθητικού νευρικού συστήματος, όπως είναι η αύξηση των παλμών και η ετοιμότητα για δράση.[69]

Και οι δύο διαστάσεις παίρνουν τιμές στο διάστημα [-1,1] με -1 να χαρακτηρίζει την πιο αρνητική και την πιο παθητική συναισθηματική κατάσταση ενώ με 1 χαρακτηρίζεται η πιο θετική και η πιο ενεργητική συναισθηματική κατάσταση αντίστοιχα.[79]



Σχήμα 3.4: Ο διδιάστατος κύκλος συναισθημάτων [;]

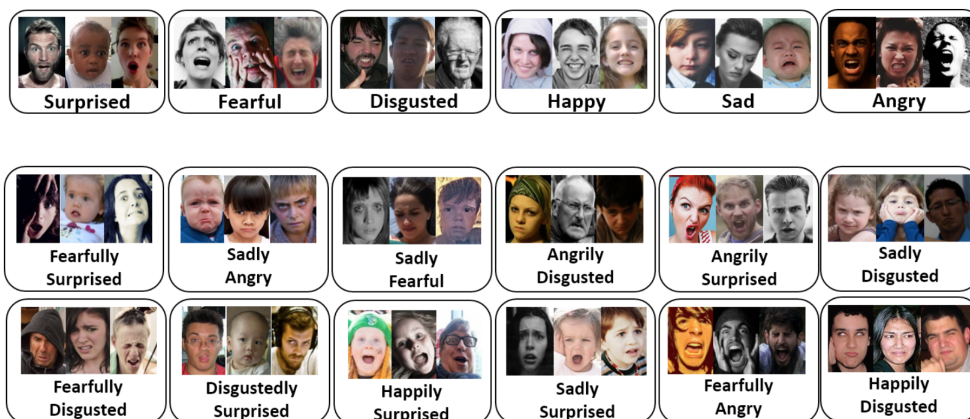
Κεφάλαιο 4

Σύνολα Δεδομένων

Στην παρούσα διπλωματική ασχοληθήκαμε με δύο διαφορετικές εργασίες. Η μία είναι η επίλυση ενός προβλήματος ταξινόμησης (classification task) καθώς ζητείται η κατηγοριοποίηση κάθε εικόνας του συνόλου δοκιμής σε μία εκ των διακριτών κατηγοριών συναισθημάτων που έχουμε προαναφέρει. Η δεύτερη είναι η επίλυση ενός προβλήματος παλινδρόμησης (regression task) καθώς ζητείται ο υπολογισμός της τιμής του σθένους και της διέγερσης για κάθε εικόνα του συνόλου δοκιμής. Για τις δύο αυτές εργασίες έπρεπε να βρούμε τα κατάλληλα σύνολα δεδομένων, τα οποία περιγράφονται αναλυτικότερα παρακάτω.

4.1 RAF-DB

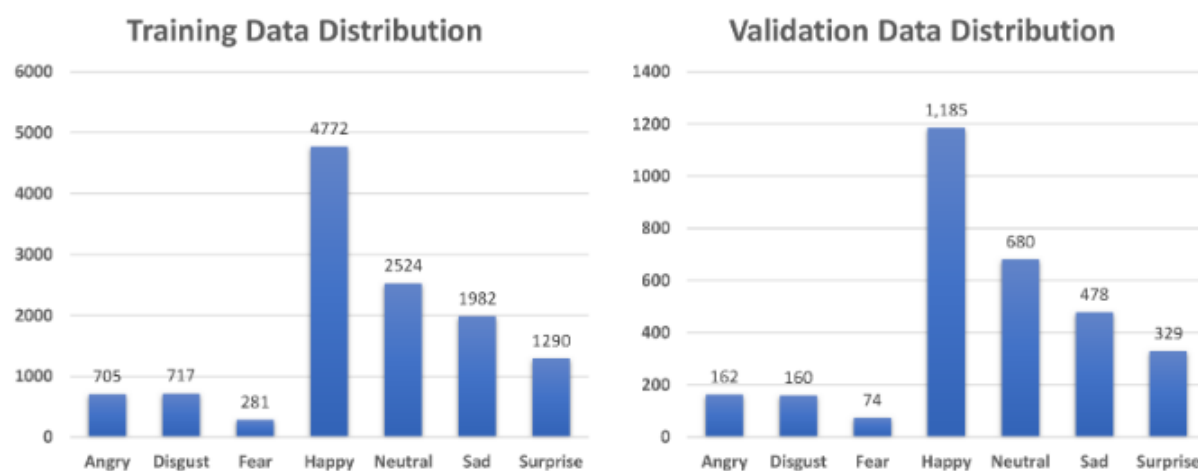
Η Real-world Affective Faces Database (RAF-DB) [21] αποτελεί μια στατική βάση δεδομένων με περίπου 30 χιλιάδες εικόνες εκφράσεων προσώπου από το διαδίκτυο. Η διαδικασία δημιουργίας ετικετών της βάσης πραγματοποιήθηκε με τη μέθοδο του πληθοπορισμού (crowdsourcing), με την κάθε εικόνα να επισημαίνεται ανεξάρτητα από περίπου 40 υπομηματιστές. Οι εικόνες σε αυτή τη βάση δεδομένων παρουσιάζουν μεγάλη ποικιλομορφία όσον αφορά την ηλικία, το φύλο και την εθνικότητα των υποκειμένων, τις στάσεις του κεφαλιού, τις συνθήκες φωτισμού, τις αποκρύψεις (π.χ. γυαλιά, τρίχες στο πρόσωπο ή ειδικά εφέ), τις διαδικασίες μετα-επεξεργασίας (π.χ. διάφορα φίλτρα και ειδικά εφέ) κ.λπ. [80] [81]



Σχήμα 4.1: Παραδείγματα εικόνων της RAF-DB από τις 6 βασικές κατηγορίες και από άλλες 12 σύνθετες συναισθηματικές καταστάσεις [21]

Το σύνολο των εικόνων της βάσης χωρίζεται σε 2 υποσύνολα. Στο πρώτο, υπάρχουν 15.339 εικόνες που έχουν επισημειωθεί ως προς ένα από τα 7 βασικά συναισθήματα (κατηγορικό μοντέλο) ενώ στο δεύτερο, οι εικόνες έχουν χαρακτηριστεί ως προς 12 σύνθετες συναισθηματικές καταστάσεις. Εμείς, στα πλαίσια της παρούσας διπλωματικής, ασχοληθήκαμε μόνο με το πρώτο υποσύνολο.

Στο σχήμα 4.2 εμφανίζεται η κατανομή δειγμάτων σε κλάσεις για το σύνολο εκπαίδευσης (training set) και το σύνολο επικύρωσης (validation set) της βάσης RAF-DB.



Σχήμα 4.2: Κατανομή δειγμάτων σε κλάσεις για το σύνολο εκπαίδευσης (training set) και το σύνολο επικύρωσης (validation set) [22]

Το σύνολο δεδομένων RAF-DB χρησιμοποιείται για την επίλυση του προβλήματος ταξινόμησης.

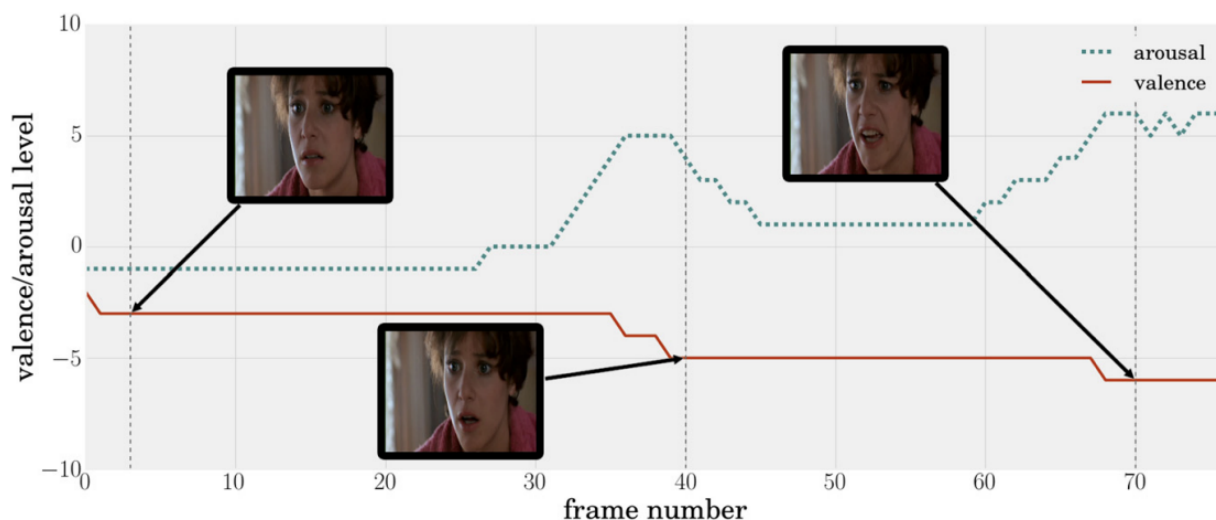
4.2 Afew-VA

Το σύνολο δεδομένων AFEW-VA είναι μια συλλογή εξαιρετικά ακριβών ανά καρτέ σχολιασμών σθένους και διέγερσης, μαζί με ανά καρτέ σχολιασμούς 68 ορόσημων του προσώπου για 600 απαιτητικά βίντεο κλιπ. Αυτά τα κλιπ προέρχονται από ταινίες μεγάλου μήκους και είχαν επίσης σχολιαστεί ως προς τις διακριτές κατηγορίες συναισθημάτων με τη μορφή της βάσης δεδομένων AFEW. [82] [83] Επομένως, το σύνολο δεδομένων αποτελείται από φακέλους όπου ο κάθε ένας αντιστοιχεί σε εικόνες που έχουν ληφθεί διαδοχικά από συγκεκριμένα καρτέ βίντεο.

Η βάση αποτελείται από 30028 εικόνες οι οποίες δεν είναι εκ των προτέρων διαχωρισμένες σε train και test set. Εμείς επιλέξαμε έναν τυχαίο διαχωρισμό 20018 εικόνων σε train set και 9910 σε test set.

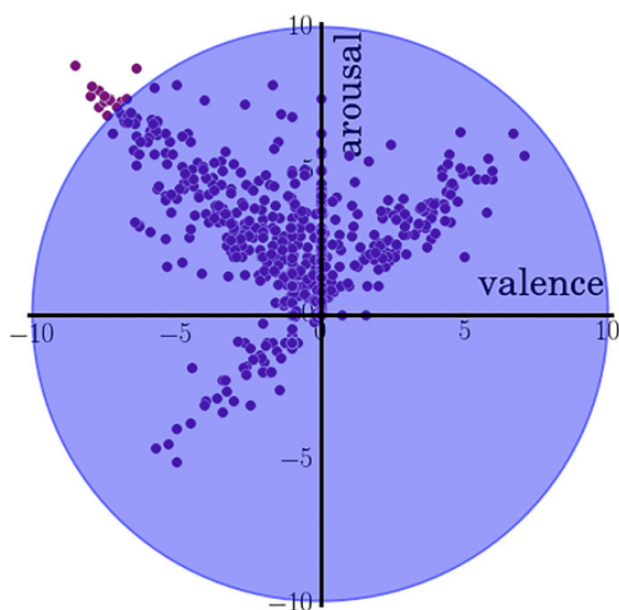
Στο Σχήμα 4.3 απεικονίζονται στιγμιότυπα που έχουν ληφθεί διαδοχικά από το ίδιο κομμάτι μίας ταινίας καθώς και οι αντίστοιχες τιμές σθένους και διέγερσης.

Όσον αφορά την κατανομή των τιμών σθένους και διέγερσης σε αυτό το σύνολο δεδομένων, με βάση το Σχήμα 4.4 παρατηρούμε ότι οι περισσότερες τιμές συγκεντρώνονται στο 2ο τεταρτημόριο με θετικό arousal και αρνητικό valence. Επίσης, βλέπουμε ότι υπάρχουν αρκετές τιμές και στο 1ο και 3ο τεταρτημόριο ενώ καμία τιμή δεν υπάρχει στο 4ο τεταρτη-



Σχήμα 4.3: Παράδειγμα επισημειωμένων τιμών *valence* και *arousal* για ένα δείγμα βίντεο από το σύνολο δεδομένων μας μαζί με μερικά αντιπροσωπευτικά καρτέ. [23]

μόριο, κάτι το οποίο είναι αναμενόμενο μιας και αντιπροσωπεύει θετικό *valence* και αρνητικό *arousal*, δηλαδή ένα συναίσθημα το οποίο είναι θετικό και παράλληλα παθητικό.



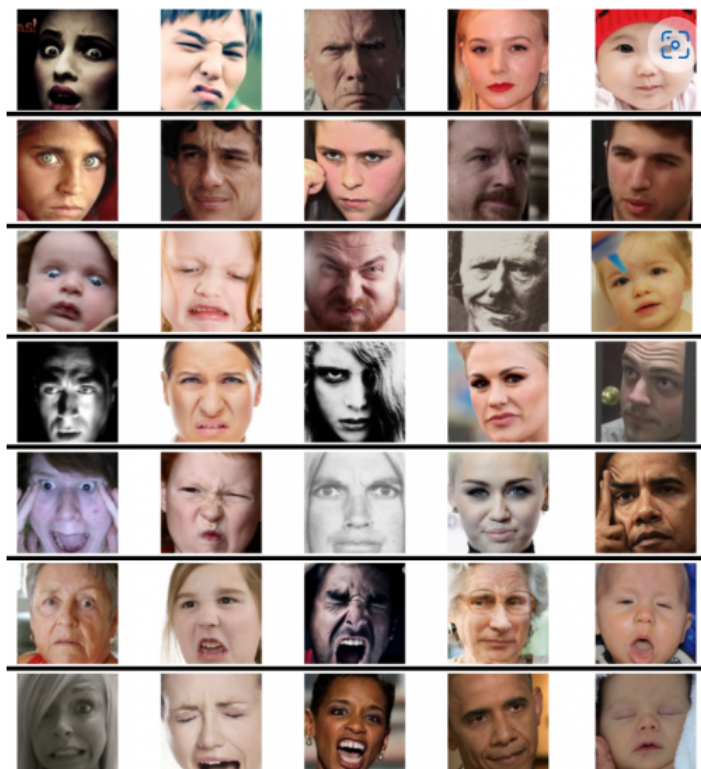
Σχήμα 4.4: Κατανομή των τιμών *valence* και *arousal* στο σύνολο δεδομένων AFEW-VA. [23]

Το σύνολο δεδομένων AFEW-VA χρησιμοποιείται για την επίλυση του προβλήματος παλινδρόμησης.

4.3 AffectNet

Η AffectNet [24] αποτελεί μια στατική βάση δεδομένων η οποία περιλαμβάνει πάνω από 1 εκατομμύριο εικόνες εκφράσεων προσώπου κάτω από μη-ελεγχόμενες συνθήκες, *in-the-wild*. Οι εικόνες αυτές συγκεντρώθηκαν από το ίντερνετ αξιοποιώντας τρεις βασικές μηχανές

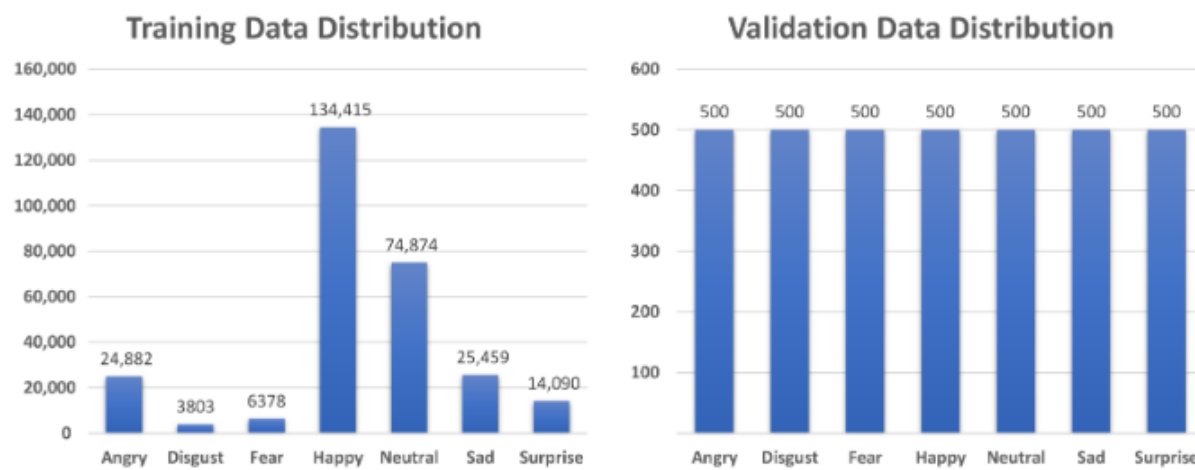
αναζήτησης και 1250 λέξεις-κλειδιά σχετικές με το συναίσθημα σε 6 διαφορετικές γλώσσες. Οι μισές περίπου από τις εικόνες (440 χιλιάδες) περιέχουν ετικέτες ως προς ένα από τα 7 βασικά συναισθήματα (συν το συναίσθημα της περιφρόνησης) καθώς και ως προς τις συνιστώσες Valence-Arousal (διανυσματικό μοντέλο). Το σύνολο εκπαίδευσης της βάσης αποτελείται από περίπου 321 χιλ. εικόνες ενώ το σύνολο επικύρωσης από 5 χιλ. Στα πλαίσια της παρούσας διπλωματικής αξιοποιούμε μία μικρότερη έκδοση αυτής της βάσης, την AffectNet8Labels στην οποία το σύνολο εκπαίδευσης της βάσης αποτελείται από περίπου 280 χιλ. εικόνες ενώ το σύνολο επικύρωσης από περίπου 4 χιλ. [25] [84]



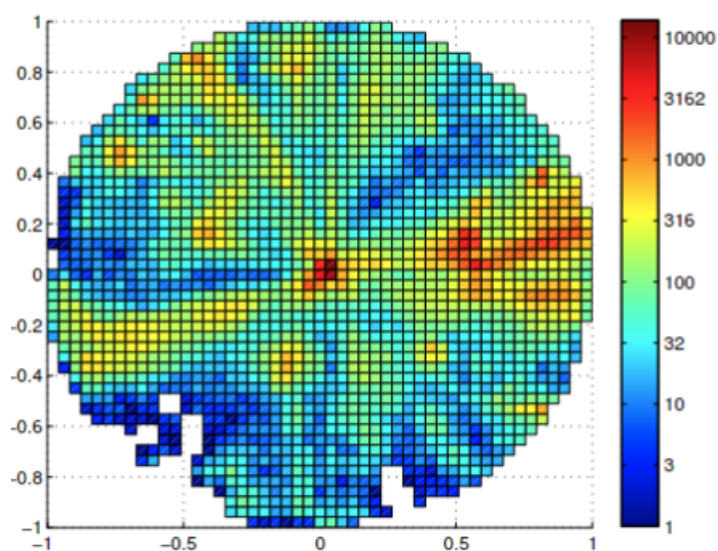
Σχήμα 4.5: Δείγμα εικόνων από την βάση AffectNet. [24]

Όπως βλέπουμε στο σχήμα 4.6, το σύνολο εκπαίδευσης είναι μη-ισορροπημένο, με την κλάση “happy” να έχει με διαφορά τα περισσότερα δείγματα. Δεύτερη βρίσκεται η “neutral” με επίσης μεγάλο αριθμό εικόνων ενώ ακολουθούν όλες οι υπόλοιπες. Το σύνολο επικύρωσης είναι πλήρως ισορροπημένο, με όλες τις κλάσεις να έχουν τον ίδιο αριθμό δειγμάτων (500).

Αναφορικά με την κατανομή των τιμών Valence-Arousal όπως απεικονίζεται στο σχήμα 4.7 υπάρχουν περισσότερα δείγματα στο κέντρο και στο δεξιό μέσο (θετική αξία και μικρή θετική διέγερση) της περιφέρειας, γεγονός που επιβεβαιώνει τον υψηλότερο αριθμό “happy” και “neutral” εικόνων στη βάση δεδομένων σε σύγκριση με άλλες κατηγορίες στο κατηγορηματικό μοντέλο. [25]



Σχήμα 4.6: Κατανομή δειγμάτων σε κλάσεις για το σύνολο εκπαίδευσης (training set) και για το σύνολο επικύρωσης (validation set) της βάσης AffectNet. [22]



Σχήμα 4.7: Κατανομή δειγμάτων valence-arousal σε τεταρτημόρια της βάσης AffectNet. [25]

Μέρος 

Πρακτικό Μέρος

Παρεμφερείς εργασίες

Στο παρόν κεφάλαιο θα ασχοληθούμε με την έρευνα που έχει ήδη γίνει γύρω από το πεδίο της ανίχνευσης συναισθήματος και της εκτίμησης σθένους/διέγερσης περιγράφοντας περιληπτικά την δουλειά που έχει γίνει από άλλους ερευνητές και τα αποτελέσματα στα οποία κατέληξαν. Θα εξερευνήσουμε τις παρεμφερείς εργασίες χρονολογικά, από τις πιο παλιές στις πιο νέες.

Αρχικά, στο πεδίο της αναγνώρισης της έκφρασης του προσώπου αρκετοί ερευνητές ήρθαν αντιμέτωποι με το θέμα της αλληλοεπικάλυψης και της παραλλαγής στάσεων, το οποίο επέφερε αλλαγή στην εμφάνιση του προσώπου και επομένως δημιουργούσε δυσκολίες στην ανίχνευση της έκφρασης. Στην εργασία "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition" [85], αρχικά επισημάνθηκαν αρκετά σύνολα δεδομένων με ετικέτες που αφορούν την στάση και την απόκρυψη και στην συνέχεια δημιουργήθηκε ένα δίκτυο για την προσαρμοστική αποτύπωση της σημασίας των περιοχών του προσώπου στην αναγνώριση της έκφρασης. Τέλος, προτάθηκε μία συνάρτηση απωλειών η οποία δίνει περισσότερη προσοχή σε σημαντικές περιοχές του προσώπου.

Στην συνέχεια, στην εργασία τους "Deep Neural Network Augmentation: Generating Faces for Affect Analysis" [86, 87] οι συγγραφείς ασχολήθηκαν με την σύνθεση εκφράσεων προσώπου. Το σύστημα τους δέχεται σαν είσοδο μία ουδέτερη εικόνα ανθρώπου και μία βασική έκφραση προσώπου ή ένα ζεύγος τιμών σθένους-διέγερσης και παράγει ως έξοδο την αρχική εικόνα με την αντίστοιχη έκφραση προσώπου. Έπειτα, οι νέες αυτές εικόνες χρησιμοποιήθηκαν για επαύξηση δεδομένων κατά την εκπαίδευση βαθιών νευρωνικών δικτύων για την αναγνώριση της έκφρασης σε άλλες βάσεις δεδομένων.

Παρόλο που το μοντέλο StarGAN έχει ξεχωρίσει για τις επιδόσεις του στην σύνθεση εκφράσεων του προσώπου, δεν είναι πάντοτε επαρκές μιας και δεν έχει την δυνατότητα να δημιουργήσει εκφράσεις δεδομένου ενός ζεύγους σθένους/διέγερσης. Αυτό αποτελεί πρόβλημα μιας και οι βασικές εκφράσεις του προσώπου τις οποίες μπορεί να δημιουργήσει το StarGAN δεν αντικατοπτρίζουν τα μεικτά συναισθήματα (διαστατική αναπαράσταση) που εμφανίζονται στις καθημερινές αλληλεπιδράσεις ανθρώπου-υπολογιστή. Γι'αυτό τον λόγο, στο "VA-StarGAN: Continuous Affect Generation" [88] δημιουργήθηκε ένα νέο μοντέλο, το VA-StarGAN το οποίο εκτελεί και τις δύο λειτουργίες, καλύπτοντας έτσι το κενό του StarGAN.

Η εργασία "Exploiting multi-CNN features in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset" [89, 90] αφορά το θέμα της εκτίμησης των τιμών σθένους και διέγερσης. Έχοντας δημιουργήσει ένα δίκτυο που αποτελείται από CNN και

RNN, χρησιμοποιούν τα χαρακτηριστικά που έχουν εξαχθεί κατά την εκπαίδευση του CNN στα σύνολα δεδομένων Aff-Wild [91, 92, 93] και Aff-Wild2 [94, 95, 96, 97, 98, 99, 100] και τα τροφοδοτούν στο υποδίκτυο RNN. Στην συνέχεια, κάνουν χρήση του παραπάνω δικτύου στο σύνολο δεδομένων One-Minute Gradual-Emotion (OMG-Emotion) για την ανίχνευση του σθένους και της διέγερσης.

Μία άλλη σημαντική εργασία που αφορά το πεδίο με το οποίο ασχολούμαστε στην παρούσα διπλωματική είναι η "Distribution Matching for Heterogeneous Multi-Task Learning: a Large-scale Face Study" [101, 102] η οποία ασχολείται με την ετερογενή μάθηση πολλαπλών εργασιών (Heterogeneous Multi-Task Learning). Πιο συγκεκριμένα, στην συγκεκριμένη εργασία κάνουν χρήση της αντιστοίχισης κατανομής (distribution matching) και της συν-επισημείωση ετικετών (label coannotation) προκειμένου να εκτελεστούν ετερογενείς εργασίες παράλληλα. Επίσης προτείνουν το πρώτο ολιστικό δίκτυο (το FaceBehaviorNet) για την ανάλυση της συμπεριφοράς του προσώπου το οποίο μπορεί να προβλέψει ταυτόχρονα τα 7 βασικά συναισθήματα, 17 μονάδες δράσης και τις συνεχείς τιμές σθένους/διέγερσης. Οι δύο μελέτες περιπτώσεων (case studies) με τις οποίες ασχολούνται είναι ο συναισθηματικός υπολογισμός (affective computing) και η αναγνώριση προσώπου (face recognition). Τέλος, διεξήγαγαν μία εκτεταμένη πειραματική μελέτη στην οποία αξιολόγησαν το FaceBehaviorNet σε 10 βάσεις δεδομένων και παρατήρησαν ότι μετά την εκπαίδευση το νευρωνικό τους υπερτερεί έναντι των δικτύων μεμονωμένης εργασίας (single-task) σε όλες τις εργασίες και σε όλες τις βάσεις δεδομένων.

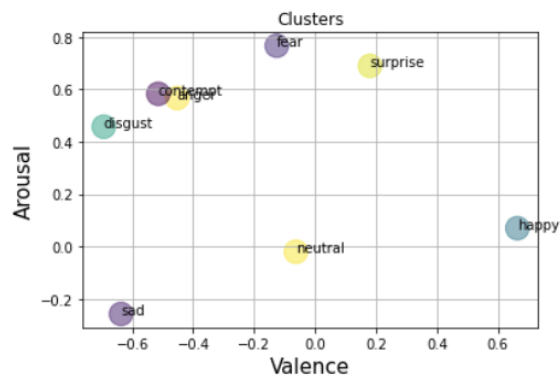
Κεφάλαιο 6

Μεθοδολογία

Στο κεφάλαιο αυτό θα αναλύσουμε την μεθοδολογία που ακολουθήσαμε για την επίλυση του προβλήματος με το οποίο ασχολούμαστε. Στόχος μας ήταν να συνδυάσουμε τις δύο ανεξάρτητες εργασίες (αυτή της ταξινόμησης και αυτή της παλινδρόμησης) δημιουργώντας ένα σύστημα που με δεδομένη την μία μπορεί να εκτελέσει την άλλη. Αυτό το πετύχαμε με την προεπεξεργασία πριν την εκπαίδευση του νευρωνικού, διαδικασία την οποία θα συζητήσουμε πιο αναλυτικά παρακάτω.

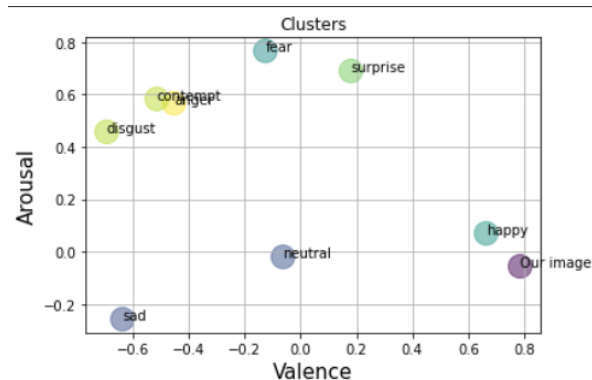
6.1 Εκ νέου επισήμανση με βάση την σχετικότητα διαφορετικών ετικετών

Όπως προαναφέρθηκε, το σύνολο δεδομένων Affectnet που χρησιμοποιήσαμε αποτελείται από εικόνες που έχουν επισημανθεί τόσο όσον αφορά το συναίσθημα όσο και για τις τιμές σθένους και διέγερσης. Επομένως για την πρώτη μέθοδο που υλοποιήσαμε, αρχικά λάβαμε την ετικέτα κάθε εικόνας αναφορικά με το συναίσθημα, η οποία λαμβάνει τιμές από 0 έως 7. Για να διευκολύνουμε τις μετέπειτα πράξεις μετατρέπουμε κάθε ετικέτα σε one hot encoded μορφή, δηλαδή κάθε εικόνα θα έχει 1 στο συναίσθημα το οποίο αποτελεί την ετικέτα του και 0 σε όλα τα υπόλοιπα. Στην συνέχεια, θέλουμε να υπολογίσουμε σημασιολογικά σύμφωνες ομάδες (clusters) με βάση το συναίσθημα. Στην συγκεκριμένη περίπτωση, οι ομάδες αυτές καθορίζονται υπολογίζοντας την μέση τιμή σθένους και διέγερσης ανά συναίσθημα και τα αποτελέσματα που παίρνουμε φαίνονται σχηματικά παρακάτω:



Σχήμα 6.1: Τα clusters που έχουν δημιουργηθεί.

Έπειτα, υπολογίζουμε την ευκλείδια απόσταση κάθε εικόνας από κάθε cluster με βάση τις αντίστοιχες τιμές σθένους και διέγερσης. Έτσι για κάθε εικόνα παίρνουμε 8 ευκλείδιες αποστάσεις. Για να είναι πιο πλήρης η περιγραφή μας παραθέτουμε το παρακάτω παράδειγμα:



Σχήμα 6.2: Τα clusters που έχουν δημιουργηθεί και πού τοποθετείται σε σχέση με τα clusters μία τυχαία εικόνα με σθένος = 0.785714 και διέγερση = -0.055556.

```
[0.8491955280720885,
0.1748405011561023,
1.4369115551722629,
0.9597933698921617,
1.2271946409006502,
1.565818575439517,
1.3860217547600933,
1.4485501798575644]
```

Σχήμα 6.3: Οι αποστάσεις από κάθε cluster (neutral, happy, sad, surprise, fear, disgust, anger, contempt αντίστοιχα).

Από αυτές τις αποστάσεις κάθε εικόνας κρατάμε την μέγιστη και την διαιρούμε με κάθε μία εκ των 8 αποστάσεων. Το αποτέλεσμα που προκύπτει είναι μία νέα λίστα 8 αντικειμένων για κάθε εικόνα, δηλαδή:

```
[1.8438846221839666,
8.95569713587993,
1.0897111724122786,
1.6314121607398118,
1.275933355030256,
1.0,
1.129721499725338,
1.080955701233273]
```

Σχήμα 6.4: Οι νέες αποστάσεις που υπολογίζονται από τον τύπο $\max(\text{distance})/\text{distance}$.

Την νέα λίστα που προέκυψε την περνάμε από την συνάρτηση softmax η οποία μετατρέπει κάθε αντικείμενο της λίστας σε πιθανότητα. Παρατηρούμε ότι το συναίσθημα που είχε την μικρότερη απόσταση από την συγκεκριμένη εικόνα, στο τέλος έχει την μεγαλύτερη πιθανότητα μετά την εφαρμογή της softmax.

```
[0.0008126117694640895,
0.996561348641954,
0.00038225197722363975
0.0006570637237013212,
0.00046049509308150605
0.0003494529196855682,
0.0003978560851138351,
0.0003789197897761381]
```

Σχήμα 6.5: Οι αποστάσεις μετά την εφαρμογή της softmax.

Τέλος, θέλουμε οι τελικές ετικέτες κάθε εικόνας να είναι γραμμικός συνδυασμός των αρχικών ετικετών και των ετικετών μετά την softmax σύμφωνα με τον τύπο:

$$labels_{new} = \lambda \cdot labels_{real} + (1 - \lambda) \cdot labels_{smoothed} \quad (6.1)$$

όπου το λ παίρνει τιμές στο διάστημα (0, 1). Στην συνέχεια εκπαιδεύουμε κατά τα γνωστά το νευρωνικό με τις νέες ετικέτες.

6.2 Εκκαθάριση Δεδομένων (Data Cleaning)

Όπως είναι γνωστό, τα σύνολα δεδομένων που διατίθενται για εκπαίδευση έχουν επισημανθεί είτε χειροκίνητα από ανθρώπους είτε με κάποιον αυτόματο τρόπο, και με τους δύο τρόπους ωστόσο να είναι επιρρεπείς σε σφάλματα λόγω υποκειμενικότητας. Προκειμένου να έχουμε όσο το δυνατόν πιο ακριβείς ετικέτες στα δεδομένα, εφαρμόσαμε μία μέθοδο εκκαθάρισης δεδομένων πριν την χρήση τους. Για κάθε μία εικόνα εξετάσαμε αν οι τιμές σθένους και διέγερσης που της έχουν ανατεθεί συμμορφώνονται με τις θεωρητικά αποδεκτές για το συναίσθημα το οποίο παρουσιάζει, αν δεν συμβαίνει αυτό τότε τις διαγράψαμε από το σύνολο εκπαίδευσης. Παρακάτω παραθέτουμε τα αποτελέσματα που λάβαμε για κάθε κλάση συναισθήματος.

6.2.1 Ουδέτερη έκφραση (Neutral)

Όσον αφορά το ουδέτερο συναίσθημα, πρέπει το σθένος και η διέγερση να λαμβάνουν τιμές στο διάστημα [-0.1,0.1]. Οι ιδανικές τιμές θα ήταν 0 και για τα δύο μιας και το ουδέτερο εξ ορισμού δεν θεωρείται ούτε αρνητικό αλλά ούτε και θετικό συναίσθημα και παράλληλα ούτε ενεργητικό αλλά ούτε και παθητικό. Παρ' όλα αυτά, στις ανθρώπινες εκφράσεις είναι πρακτικά αδύνατο να υπάρξει το πλήρως ουδέτερο συναίσθημα οπότε τυπικά θεωρούνται αποδεκτές οι τιμές μέχρι ± 0.1 . Ορίζοντας αυτούς τους περιορισμούς βρήκαμε ότι από τις 74.874 εικόνες του συνόλου εκπαίδευσης που έχουν ετικέτα ουδέτερο, οι 31.745 έχουν λάθος ετικέτα.

6.2.2 Χαρούμενη έκφραση (Happy)

Αναφορικά με το χαρούμενο συναίσθημα δοκιμάσαμε δύο πιθανά εύρη τιμών. Το πρώτο είναι ότι οι τιμές σθένους και διέγερσης θα πρέπει να είναι αμφότερες θετικές. Αυτή η δοκιμή μας έδειξε ότι από τις 134.415 εικόνες με ετικέτα χαρούμενο, οι 37.317 ήταν λάθος.

Στην συνέχεια δοκιμάσαμε ένα ακόμα πιο αυστηρό κριτήριο, όπου τόσο το σθένος όσο και η διέγερση έχουν τιμές μεγαλύτερες από 0.1. Αυτή την τιμή την βάλουμε με το σκεπτικό ότι εφόσον το ουδέτερο περιέχει τιμές μέχρι το 0.1, για να είναι σίγουρα χαρούμενη έκφραση και όχι ουδέτερη, θα πρέπει να είναι πιο θετική και πιο ενεργητική από το ουδέτερο. Με αυτές τις τιμές βγάλαμε ότι από τις 134.415 εικόνες, οι 75.111 είναι λάθος, δηλαδή αριθμός μεγαλύτερος από την πρώτη δοκιμή όπως και αναμέναμε.

6.2.3 Λυπημένη έκφραση (Sad)

Σχετικά με την λύπη, ο μόνος περιορισμός που μπορούμε να θέσουμε είναι το σθένος να είναι αρνητικό, μιας και σαν συναίσθημα μπορεί να εκφραστεί ως ενεργητική λύπη αλλά και ως παθητική λύπη. Μόνο με αυτόν τον περιορισμό λοιπόν, πήραμε ότι από τις 25.459 εικόνες που έχουν ετικέτα λυπημένο, 5 εικόνες είχαν λάθος ετικέτα.

6.2.4 Έκπληκτη έκφραση (Surprise)

Όσον αφορά το συναίσθημα της έκπληξης, δεν θέσαμε περιορισμούς και επιλέξαμε να κρατήσουμε όλα τα αντίστοιχα δεδομένα αυτούσια. Ο λόγος για κάτι τέτοιο είναι διότι οι απόψεις διίστανται σχετικά με το τι εύρος τιμών σθένους και διέγερσης έχει το συγκεκριμένο συναίσθημα με μερικούς ερευνητές να υποστηρίζουν ότι έχει αρνητικό σθένος και θετική διέγερση [103] και άλλοι ότι έχει θετικό σθένος και θετική/μηδενική διέγερση [104]. Επειδή λοιπόν δεν υπάρχει συμφωνία στις θεωρητικές τιμές αυτού του συναισθήματος, κρατήσαμε όλα τα δείγματα του.

6.2.5 Έκφραση φόβου (Fear)

Για το συναίσθημα του φόβου, γνωρίζουμε με σιγουριά ότι θα πρέπει να έχει αρνητικό σθένος το οποίο ορίσαμε ως περιορισμό. Με αυτόν τον τρόπο, είδαμε ότι από τις 6.378 εικόνες με φοβισμένες εκφράσεις είχαν λάθος ετικέτα οι 29.

6.2.6 Έκφραση αηδίας (Disgust)

Η έκφραση αηδίας, αντίστοιχα με την λύπη και τον φόβο, πρέπει να έχει αρνητικό σθένος. Επομένως με αυτόν τον περιορισμό βρήκαμε 8 λάθος εικόνες από τις 3.803 που ήταν αρχικά.

6.2.7 Θυμωμένη έκφραση (Anger)

Αναφορικά με τον θυμό, δεδομένου ότι είναι ένα αρνητικό συναίσθημα, είναι αυτονόητο ότι έχει αρνητικό σθένος, το οποίο οδήγησε στο να βρεθούν 7 λάθος εικόνες από τις 24.882.

6.2.8 Έκφραση περιφρόνησης (Contempt)

Τέλος, για την περιφρόνηση θέσαμε πάλι τον περιορισμό αρνητικού σθένους το οποίο μας έδειξε ότι από τις 3.750 εικόνες, οι 109 ήταν λάθος.

Όπως βλέπουμε στον παρακάτω πίνακα, οι μεγαλύτεροι αριθμοί λάθος εικόνων αφορούν το ουδέτερο και το χαρούμενο συναίσθημα. Αυτό αφενός εξηγείται από το γεγονός ότι αυτές

Expression	Αριθμός αρχικών εικόνων	Αριθμός λάθος εικόνων	Ποσοστό λάθος εικόνων (επί των αρχικών)
Neutral	74.874	31.745	42.4 %
Happy (1 ^η δοκιμη)	134.415	37.317	27.8 %
Happy (2 ^η δοκιμη)	134.415	75.111	55.9 %
Sad	25.459	5	0.02 %
Fear	6.378	29	0.5 %
Disgust	3.803	8	0.2 %
Anger	24.882	7	0.03 %
Contempt	3.750	109	2.9 %

Πίνακας 6.1: Αριθμός εικόνων πριν την εκκαθάριση δεδομένων και ποσοστά λάθος εικόνων ανά συναίσθημα

οι δύο κλάσεις περιέχουν τις περισσότερες εικόνες συγκριτικά με τις υπόλοιπες αλλά και επειδή τα δύο αυτά συναισθήματα αρκετές φορές έχουν επικαλυπτόμενες τιμές σθένους και διέγερσης μιας και στις οριακές τιμές μπορεί μία ουδέτερη έκφραση να χαρακτηριστεί ως χαρά και το αντίστροφο.

6.3 Εκκαθάριση δεδομένων στο σύνολο δοκιμής

Ακολουθώντας την παραπάνω διαδικασία, για στατιστικούς λόγους ελέγξαμε αν οι ετικέτες του συνόλου δοκιμής ικανοποιούν τους περιορισμούς που έχουν ήδη αναφερθεί. Τα αποτελέσματα που λάβαμε συνοψίζονται στον πίνακα της επόμενης σελίδας.

Λάμβάνοντας υπόψιν τον παρακάτω πίνακα, παρατηρούμε ότι με βάση τα στατιστικά υπάρχουν αρκετά λάθη κυρίως στα συναισθήματα Neutral, Happy όπως αντίστοιχα υπέδειξαν τα στατιστικά που παράξαμε για το σύνολο εκπαίδευσης. Δεδομένου του γεγονότος όμως ότι όλοι οι ερευνητές που έχουν χρησιμοποιήσει το συγκεκριμένο σύνολο δεδομένων δεν έχουν εκτελέσει κάποια εκκαθάριση δεδομένων στο σύνολο δοκιμής, για λόγους ευκολίας στην σύγκριση της απόδοσης του συστήματός μας σε σχέση με άλλα, δεν εκτελέσαμε εκκαθάριση δεδομένων σε αυτή την περίπτωση.

Expression	Αριθμός αρχικών εικόνων	Αριθμός λάθος εικόνων	Ποσοστό λάθος εικόνων (επί των αρχικών)
Neutral	500	151	30.2 %
Happy (1 ^η δοκιμή)	500	109	21.8 %
Happy (2 ^η δοκιμή)	500	287	57.4 %
Sad	500	0	0 %
Fear	500	3	0.6 %
Disgust	500	1	0.02 %
Anger	500	0	0 %
Contempt	500	13	2.6 %

Πίνακας 6.2: Αριθμός εικόνων και ποσοστά λάθος εικόνων ανά συναίσθημα στο σύνολο δοκιμής

Κεφάλαιο 7

Πειραματική Διαδικασία

Στο συγκεκριμένο κεφάλαιο παρουσιάζουμε την πειραματική διαδικασία που ακολουθήθηκε ώστε να γίνει η εκπαίδευση των κατάλληλων μοντέλων. Για την επιτυχή εκπαίδευση των μοντέλων και την τελική επιλογή των υπερπαραμέτρων έγιναν μία σειρά από πειράματα χρησιμοποιώντας διάφορα υπολογιστικά συστήματα. Παρακάτω γίνεται αναλυτική παρουσίαση των ενεργειών και των πειραμάτων που πραγματοποιήθηκαν σε κάθε υπολογιστικό σύστημα και το λογισμικό που χρησιμοποιήθηκε. Ακόμη, αναγράφεται με συγκεκριμένα βήματα η συνολική διαδικασία εκπαίδευσης των μοντέλων και αναλύονται οι επιλογές κάθε παραμέτρου.

7.1 Υπολογιστικά Συστήματα

Το μεγαλύτερο μέρος της εκπαίδευσης των μοντέλων μας έγινε στην πλατφόρμα ανάπτυξης και εκτέλεσης κώδικα Python, Google Colaboratory. Η πλατφόρμα αυτή διαθέτει μία Μονάδα Επεξεργασίας Γραφικών (GPU) Tesla K80 με μνήμη 12GB αν και οι GPU που διαθέτει μεταβάλλονται κατά διαστήματα σύμφωνα με τα επίσημα έγγραφα της Google. [105]

7.2 Διαδικασία Εκπαίδευσης

Στο συγκεκριμένο σημείο θα περιγράψουμε τα διακριτά βήματα που ακολουθήσαμε για την εκπαίδευση των μοντέλων. Αξίζει να σημειωθεί ότι στην συγκεκριμένη εργασία χρησιμοποιήσαμε το μοντέλο Resnet ωστόσο η μεθοδολογία που έχουμε ακολουθήσει λειτουργεί ανεξάρτητα της αρχιτεκτονικής που χρησιμοποιείται.

- *Βήμα 1* : Διάβαση των αρχείων εκπαίδευσης και δοκιμής από το σύνολο δεδομένων Affectnet, τα οποία περιέχουν την διαδρομή (path) κάθε εικόνας και τις αντίστοιχες ετικέτες.
- *Βήμα 2* : Εφαρμογή data cleaning με τον τρόπο που αναφέρθηκε στην Ενότητα 6.2. Αφού υπολογίσαμε τις λάθος επισημασμένες εικόνες τις διαγράψαμε από το σύνολο εκπαίδευσης.
- *Βήμα 3* : Εφαρμογή του label smoothing με τον τρόπο που έχει αναφερθεί στην Ενότητα 6.1. Ορισμός των 8 ομάδων (clusters) υπολογίζοντας την μέση τιμή του σθένους και

της διέγερσης, μία για κάθε έκφραση (Neutral, Happiness, Sadness, Surprise, Fear, Disgust, Anger, Contempt). Υπολογισμός για κάθε εικόνα της ευκλείδειας απόστασης των τιμών σθένους και διέγερσης της από τις 8 ομάδες. Στην συνέχεια, υπολογισμός των νέων τιμών αποστάσεων οι οποίες ορίζονται ως η διαίρεση της μέγιστης απόστασης ανά εικόνα δια την εκάστοτε απόσταση κάθε συναισθήματος και έπειτα πέρασμα των νέων αποστάσεων από την συνάρτηση softmax. Τέλος, ορισμός των νέων ετικέτων με βάση την σχέση

$$labels_{new} = \lambda \cdot labels_{real} + (1 - \lambda) \cdot labels_{smoothed} \quad (7.1)$$

Πειραματιστήκαμε με τις τιμές $\lambda = 0, 0.3, 0.5, 0.7, 0.9$.

- **Βήμα 4 :** Εκ νέου εφαρμογή της μεθόδου data cleaning μιας και μπορεί μερικές εικόνες μετά την μέθοδο label smoothing να έχουν αλλάξει ετικέτα βασικού συναισθήματος επομένως θα πρέπει να υπακούουν στους αντίστοιχους περιορισμούς. Στην συγκεκριμένη περίπτωση της Affectnet δεν βρήκαμε λάθος επισημασμένες εικόνες μετά το label smoothing.
- **Βήμα 5 :** Μετατροπή των εικόνων εισόδου ανά παρτίδα έτσι ώστε να γίνει η τροφοδότηση τους στο δίκτυο. Συγκεκριμένα γίνεται η μετατροπή του μεγέθους των RGB εικόνων σε $3 \times 96 \times 96$ από $3 \times 224 \times 224$ και κανονικοποίηση των τιμών κάθε εικονοστοιχείου.
- **Βήμα 6 :** Φόρτωση των εικόνων του συνόλου δεδομένων ανά παρτίδα (batch) με χρήση generators. Το μέγεθος παρτίδας (batch size) μεταβαλλόταν για το σύνολο εκπαίδευσης, ανάλογα με την εργασία που εκτελούσαμε, ενώ για το σύνολο δοκιμής παρέμενε σταθερό και ίσο με 93.
- **Βήμα 7 :** Ορισμός του μοντέλου και των κατάλληλων υπερπαραμέτρων. Συγκεκριμένα επιλέξαμε το μοντέλο Resnet, ως συνάρτηση ενεργοποίησης την softmax, ως συνάρτηση απωλειών την categorical crossentropy, ως βελτιστοποιητή τον Adam και ως ρυθμό εκμάθησης το 0.0001.
- **Βήμα 8 :** Εκπαίδευση των μοντέλων στο σύνολο δεδομένων εκπαίδευσης για ένα πλήθος εποχών (epoch). Το μοντέλο κάθε εποχή τροφοδοτείται με όλο το σύνολο δεδομένων και προσαρμόζει τις παραμέτρους του. Συνεπώς με κάθε εποχή βελτιώνεται στην πρόβλεψη των σωστών ετικετών. Το πλήθος των εποχών εδώ είναι 5-15.
- **Βήμα 9 :** Αξιολόγηση του μοντέλου σύμφωνα με το σύνολο δοκιμής. Μετά το πέρας κάθε εποχής το μοντέλο πραγματοποιεί προβλέψεις για το σύνολο δοκιμής με βάση τις μετρικές αξιολόγησης που αναλύσαμε στην Ενότητα 2.8 και με αυτό τον τρόπο επιλέγεται το καλύτερο μοντέλο, το οποίο και αποθηκεύεται

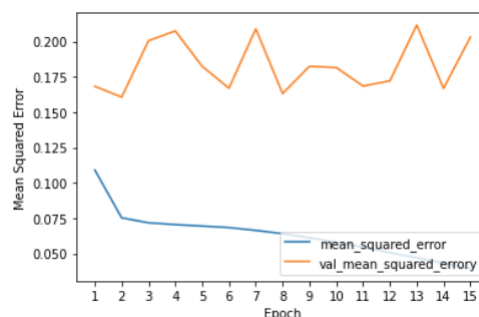
Έχουμε εκπαιδεύσει επίσης την Affectnet και χωρίς την μεθοδολογία του label smoothing (τόσο για την ανίχνευση συναισθημάτων όσο και σθένους/διέγερσης) και του data cleaning προκειμένου να χρησιμεύσουν ως baselines, όπου ακολουθήσαμε τα ίδια βήματα με παραπάνω παραλείποντας τα βήματα 2-4.

Αποτελέσματα και Αξιολόγηση

Στην συγκεκριμένη ενότητα θα αναφέρουμε τις υπερπαραμέτρους που επιλέχθηκαν για την εκπαίδευση κάθε μοντέλου με βάση τις καμπύλες μετρικής εκπαίδευσης και αξιολόγησης για τα καλύτερα μοντέλα κάθε κατηγορίας. Αξίζει να αναφερθεί ότι για κάθε μοντέλο η εκπαίδευση γινόταν για ένα πλήθος 10-20 εποχών από τις οποίες εν τέλει κρατούσαμε τα βάρη που έδιναν το βέλτιστο αποτέλεσμα στην μετρική αξιολόγησης στο σύνολο εκπαίδευσης. Ακόμα, δοκιμάσαμε διαφορετικές τιμές ρυθμών εκμάθησης στο εύρος $10^{-3} - 10^{-6}$ καθώς και batch size και παρακάτω αναφέρουμε τις βέλτιστες τιμές για κάθε μοντέλο. Όσον αφορά τις συναρτήσεις απωλειών, επιλέξαμε σε όλες τις εργασίες παλινδρόμησης να είναι η mean squared error ενώ στις εργασίες ταξινόμησης η categorical crossentropy. Επιπλέον, θα παρουσιάσουμε τα αποτελέσματα των μοντέλων όπως αυτά αξιολογήθηκαν με βάση τις μετρικές αξιολόγησης του κάθε προβλήματος (Mean squared error, Pearson correlation coefficient για παλινδρόμηση, Accuracy, F1 score, Average accuracy για ταξινόμηση).

8.1 Αξιολόγηση Μοντέλων Σθένους-Διέγερσης

Αναφορικά με την ανίχνευση του σθένους και της διέγερσης, χρησιμοποιήσαμε το μοντέλο ResNet-50 σε ολόκληρο το σύνολο δεδομένων Affectnet. Μετά από αρκετές δοκιμές, καταλήξαμε ότι τα βέλτιστα αποτελέσματα σε σχέση με τις επιλεγμένες μετρικές προέκυπταν με ρυθμό εκμάθησης ίσο με 10^{-2} , γραμμική συνάρτηση ενεργοποίησης και batch size ίσο με 203 για το σύνολο εκπαίδευσης και 93 για το σύνολο δοκιμής. Τρέξαμε το μοντέλο για 15 εποχές και βρήκαμε ότι τα βέλτιστα αποτελέσματα λήφθηκαν στην 2η εποχή όπως φαίνεται και στο Σχήμα 8.1.

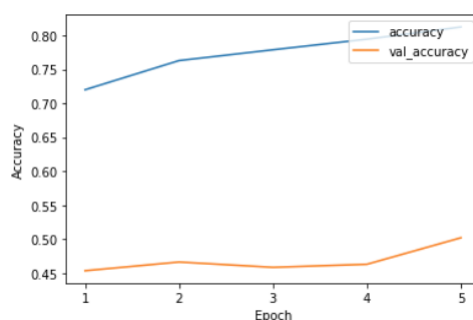


Σχήμα 8.1: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος παλινδρόμησης.

Μοντέλο	Μετρική			Επαύξηση δεδομένων
	Mean Squared Error	Pearson CC valence	Pearson CC arousal	
ResNet-50	0.16	0.57	0.50	-
ResNet-50	0.14	0.60	0.48	✓

Πίνακας 8.1: Αποτελέσματα των μοντέλων παλινδρόμησης σθένους-διέγερσης στο σύνολο αξιολόγησης.

Έπειτα, δοκιμάσαμε να εκτελέσουμε επαύξηση δεδομένων στο συγκεκριμένο μοντέλο με τις ίδιες παραμέτρους. Πιο συγκεκριμένα, εφαρμόσαμε horizontal flip, rotation και brightness range. Παρατηρήσαμε ότι για το συγκεκριμένο σύνολο δεδομένων η μεταβολή της φωτεινότητας δεν συνείσφερε στο να αυξηθεί σημαντικά κάποια μετρική, ενώ η περιστροφή των εικόνων -δεδομένου ότι οι εικόνες στην Affectnet είναι ευθυγραμμισμένες προκειμένου να διευκολυνθεί η σωστή ταξινόμηση- ουσιαστικά αναιρούσε αυτή την ευθυγράμμιση και έβγαζε χειρότερα αποτελέσματα. Γι'αυτό τον λόγο κρατήσαμε μόνο την αναστροφή η οποία μας έβγαλε τα καλύτερα αποτελέσματα στην 5η εποχή όπως φαίνεται στο Σχήμα 8.2.

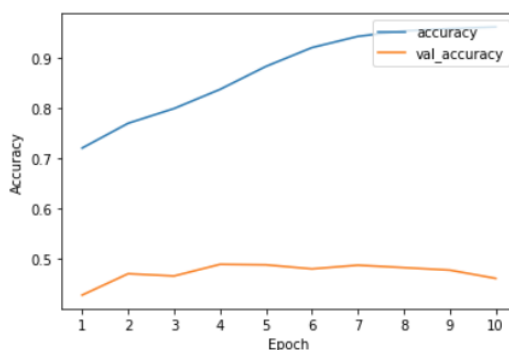


Σχήμα 8.2: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος παλινδρόμησης με εφαρμογή επαύξησης δεδομένων.

8.2 Αξιολόγηση Μοντέλων Βασικών Συναισθημάτων

Αρχικά εκπαιδεύσαμε το μοντέλο ResNet-50 σε ολόκληρο το σύνολο δεδομένων Affectnet προκειμένου να λειτουργήσει ως baseline για τα μετέπειτα πειράματα. Η εκπαίδευση έγινε για ένα σύνολο 10 εποχών, ωστόσο τα καλύτερα αποτελέσματα λήφθηκαν μετά το τέλος της 4ης εποχής, επομένως κρατήσαμε τα συγκεκριμένα βάρη προκειμένου να αξιολογήσουμε το νευρωνικό μας. Αναφορικά με τις υπερπαραμέτρους, ο ρυθμός εκμάθησης τέθηκε 10^{-4} , χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης softmax και επιλέξαμε batch size ίσο με 203 για το σύνολο εκπαίδευσης και 93 για το σύνολο δοκιμής. Στο Σχήμα 8.3 παρουσιάζονται οι καμπύλες της μετρικής ορθότητας κατά την εκπαίδευση και την αξιολόγηση. Βλέπουμε ότι πράγματι η ορθότητα αξιολόγησης φτάνει σε μέγιστο στις 4 εποχές.

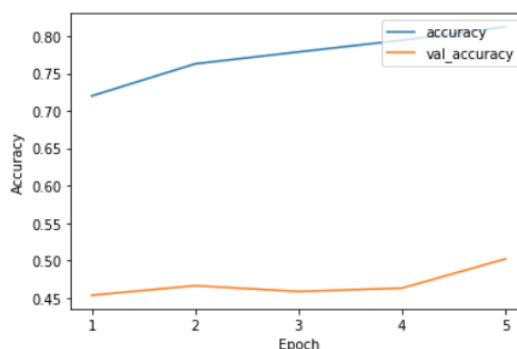
Στην συνέχεια δοκιμάσαμε να εκτελέσουμε επαύξηση δεδομένων στο συγκεκριμένο μοντέλο με τις ίδιες παραμέτρους. Για τους λόγους που αναφέραμε και παραπάνω κρατήσαμε μόνο την αναστροφή η οποία μας έβγαλε τα καλύτερα αποτελέσματα στην 5η εποχή όπως φαίνεται στο Σχήμα 8.4.



Σχήμα 8.3: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης.

Μοντέλο	Μετρική			Επαύξηση δεδομένων
	Accuracy	F1 score	Average accuracy	
ResNet-50	0.49	0.45	0.49	-
ResNet-50	0.50	0.46	0.50	✓

Πίνακας 8.2: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης.



Σχήμα 8.4: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με εφαρμογή επαύξησης δεδομένων.

8.3 Αξιολόγηση Μοντέλων Συνδυασμού Βασικών Συναισθημάτων και Σθένους/Διέγερσης

Στην συνέχεια, εκτελέσαμε multitask learning όπου ανιχνεύονται παράλληλα και τα βασικά συναισθήματα και οι τιμές σθένους/διέγερσης. Παρά το γεγονός ότι η αναγνώριση των βασικών συναισθημάτων γινόταν με ικανοποιητικό βαθμό, η ανίχνευση των τιμών σθένους και διέγερσης είχε χαμηλή απόδοση. Θεωρήσαμε ότι η αιτία αυτού είναι πιθανόν οι μικρός αριθμός εποχών (το τρέξαμε για 15 εποχές) δεδομένου ότι πλέον χρειάζεται περισσότερη εκπαίδευση για να εκτελεί και τις δύο εργασίες καθώς και το μέγεθος του batch size καθώς πιθανόν ήθελε μεγαλύτερο από 203 που το είχαμε για το σύνολο εκπαίδευσης, ωστόσο και για τα δύο είχαμε το πρόβλημα της έλλειψης υπολογιστικών πόρων (τόσο όσον αφορά την GPU όσο και αναφορικά με τον χώρο στην μνήμη). Δοκιμάσαμε τιμές ρυθμού εκμάθησης στο διάστημα 10^{-4} - 10^{-1} και τελικά επιλέξαμε ρυθμό εκμάθησης 10^{-3} που έβγαλε τα καλύτερα αποτελέσματα στις 7 εποχές.

Μοντέλο	Μετρική					Επαύξηση δεδομένων
	Accuracy	F1 score	Avg accuracy	Pearson CC val.	Pearson CC ar.	
ResNet-50	0.49	0.45	0.49	$5.78 \cdot 10^{-11}$	$5 \cdot 10^{-11}$	-
ResNet-50	0.46	0.42	0.46	$6.11 \cdot 10^{-12}$	$9.72 \cdot 10^{-11}$	✓

Πίνακας 8.3: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων και ανίχνευσης σθένους/διέγερσης στο σύνολο αξιολόγησης.

Στην συνέχεια, δοκιμάσαμε επαύξηση δεδομένων χρησιμοποιώντας την αναστροφή η οποία είδαμε ότι επέφερε βελτίωση της απόδοσης αναφορικά με την ανίχνευση σθένους/διέγερσης αλλά χειροτέρευση της αναγνώρισης συναισθήματος (Σχήμα 8.3).

8.4 Αξιολόγηση Μοντέλων εκ νέου Επισήμανσης με βάση την Σχετικότητα Διαφορετικών Ετικετών

Αφού εκπαιδεύσαμε όλα τα παραπάνω νευρωνικά προκειμένου να έχουμε κάποιες τιμές απλών unit tasks προς σύγκριση, στην συνέχεια ασχοληθήκαμε με την μεθοδολογία που προτείναμε σε αυτή την εργασία. Πειραματιστήκαμε με τιμές από $\lambda = 0$ (κρατώντας δηλαδή μόνο τις νέες ετικέτες) έως $\lambda = 0.7$. Δοκιμάσαμε επίσης τις μεθόδους data cleaning και εκ νέου relabelling προκειμένου να δούμε κατά πόσο συνεισφέρουν στην αύξηση των μετρικών.

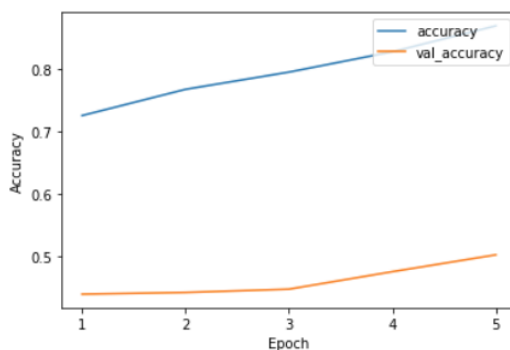
Αρχικά δοκιμάσαμε την δημιουργία clusters απευθείας πάνω στο σύνολο εκπαίδευσης. Οι ομάδες που δημιουργήθηκαν έχουν ως εξής:



Σχήμα 8.5: Τα clusters που έχουν δημιουργηθεί.

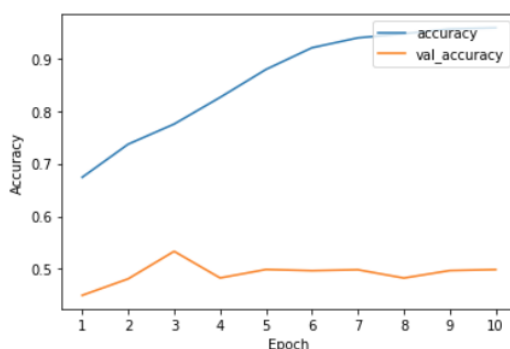
Στα πειράματα που τρέξαμε δοκιμάσαμε τις τιμές λ που αναφέρθηκαν παραπάνω και τιμές ρυθμού εκμάθησης 10^{-3} και 10^{-4} και βρήκαμε ότι σε όλα τα μοντέλα, η τιμή 10^{-4} έδωσε τα καλύτερα αποτελέσματα. Το μοντέλο που έδωσε τα βέλτιστα αποτελέσματα είναι αυτό για $\lambda = 0.7$ με ρυθμό εκμάθησης 10^{-4} , batch size 91 για το σύνολο εκπαίδευσης και 93 για το σύνολο δοκιμής και συνάρτηση ενεργοποίησης την softmax. Τα καλύτερα αποτελέσματα τα έδωσε στην 5η εποχή όπως φαίνεται και στο παρακάτω σχήμα. Οι τιμές που λάβαμε για τα υπόλοιπα λ συνοψίζονται στον Πίνακα στο τέλος της ενότητας.

Έπειρα εφαρμόσαμε την μέθοδο της εκκαθάρισης λάθος επισημασμένων εικόνων προκειμένου να έχει σωστές ετικέτες το σύνολο εκπαίδευσης και άρα να έχουμε καλύτερα αποτελέσματα. Στην παρούσα φάση δοκιμάσαμε και σαν συνάρτηση απωλειών την Kull-



Σχήμα 8.6: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\lambda = 0.7$.

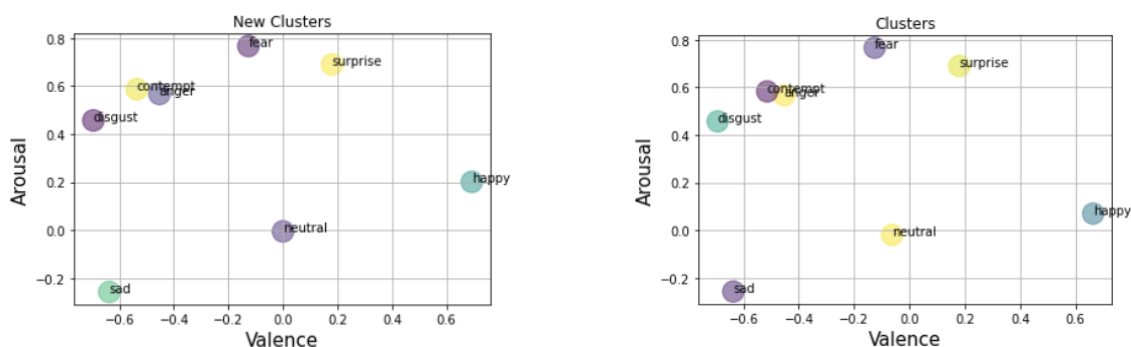
back-Leibler divergence η οποία ωστόσο μας έδωσε χειρότερα αποτελέσματα από την Categorical crossentropy επομένως κρατήσαμε την δεύτερη. Τα βέλτιστα αποτελέσματα μας τα έδωσε το $\lambda = 0.9$ στις 3 εποχές όπως βλέπουμε και στο Σχήμα 8.7.



Σχήμα 8.7: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\lambda = 0.9$ έπειτα από εκκαθάριση δεδομένων.

Τέλος, υπολογίσαμε εκ νέου τα clusters μετά την εκκαθάριση δεδομένων μιας και οι ετικέτες που έχουν μείνει είναι πιο ακριβείς (Σχήμα ;;).

Σε αυτή την περίπτωση, δοκιμάσαμε μόνο τις τιμές $\lambda=0.9$, $\lambda = 0.7$, $\lambda = 0.5$ εφόσον αυτές είχαν βέλτιστη απόδοση στα προηγούμενα στάδια. Τα καλύτερα αποτελέσματα τα πήραμε για $\lambda = 0.7$ και συγκεκριμένα στην 4η εποχή. Η καμπύλη ορθότητας εκπαίδευσης και

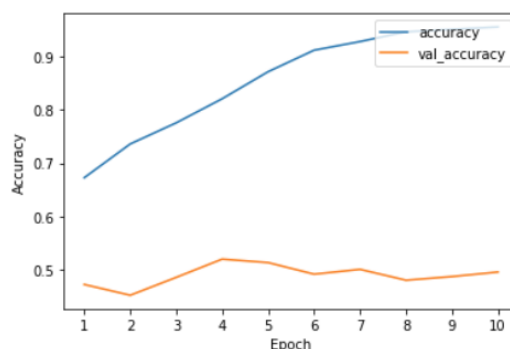


Σχήμα 8.8: Τα νέα clusters που έχουν δημιουργηθεί σε σύγκριση με τα αρχικά. Παρατηρούμε ότι μετά την εκκαθάριση τα νέα clusters έχουν μετατοπιστεί.

Μοντέλο	Μετρική			Data Cleaning	Εκ νέου Data Cleaning
	Accuracy	F1 score	Average accuracy		
ResNet-50,λ=0.9	0.48	0.45	0.48	-	-
ResNet-50,λ=0.7	0.49	0.45	0.49	-	-
ResNet-50,λ=0.5	0.47	0.43	0.46	-	-
ResNet-50,λ=0.3	0.48	0.45	0.48	-	-
ResNet-50,λ=0	0.44	0.40	0.44	-	-
ResNet-50,λ=0.9	0.54	0.52	0.54	✓	-
ResNet-50,λ=0.7	0.52	0.50	0.52	✓	-
ResNet-50,λ=0.5	0.51	0.49	0.51	✓	-
ResNet-50,λ=0.3	0.50	0.47	0.50	✓	-
ResNet-50,λ=0	0.50	0.47	0.50	✓	-
ResNet-50,λ=0.9	0.52	0.49	0.52	✓	✓
ResNet-50,λ=0.7	0.52	0.50	0.52	✓	✓
ResNet-50,λ=0.5	0.51	0.49	0.51	✓	✓

Πίνακας 8.4: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης.

αξιολόγησης φαίνεται στο Σχήμα 8.9.



Σχήμα 8.9: Καμπύλες ορθότητας εκπαίδευσης και αξιολόγησης προβλήματος ταξινόμησης με $\beta = 0.7$ έπειτα από εκκαθάριση δεδομένων και εκ νέου υπολογισμό των clusters.

8.5 Στατιστικά που αφορούν την εκκαθάριση δεδομένων

Όπως έχουμε αναφέρει στο Κεφάλαιο 6, εκτελέσαμε εκκαθάριση δεδομένων στο σύνολο εκπαίδευσης μετά το relabelling προκειμένου να απομακρύνουμε ετικέτες που είχαν λάθη. Με τον τρόπο αυτό, σε μερικές περιπτώσεις άλλαξαν οι ακραίες τιμές (outliers) για κάθε συναίσθημα, με τρόπο που συνοψίζεται στους παρακάτω πίνακες. Οι πίνακες αυτοί αφορούν κάθε εικόνα που έχει ως ετικέτα το συναίσθημα που αναγράφεται ως τίτλος, και αναφέρουν την μέγιστη, ελάχιστη, μέση τιμή κάθε συναισθήματος μετά το relabelling πριν και μετά την εκκαθάριση δεδομένων. Όπως θα παρατηρήσει κανείς, δεν αναφέρουμε τον πίνακα που αφορά το συναίσθημα surprise μιας και όπως έχουμε προαναφέρει δεν κάναμε κάποια εκκαθάριση σε δεδομένα με αυτή την έκφραση εφόσον δεν μπορούσαμε να θέσουμε κάποιους σαφείς περιορισμούς.

Με βάση τα παρακάτω αποτελέσματα, παρατηρούμε ότι στις περισσότερες περιπτώσεις

Neutral	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	1	0.3	0.28	0.20	0.29	0.1	0.14	0.07
max μετά	1	0.002	0.003	0.003	0.003	0.003	0.004	0.003
min πριν	0.7	$6.13 \cdot 10^{-117}$	$7.37 \cdot 10^{-117}$	$5.96 \cdot 10^{-117}$	$5.65 \cdot 10^{-117}$	$5.61 \cdot 10^{-117}$	$6.35 \cdot 10^{-117}$	$5.91 \cdot 10^{-117}$
min μετά	0.97	0	0	0	0	0	0	0
mean πριν	0.92	0.01	0.01	0.006	0.006	0.01	0.01	0.01
mean μετά	0.99	$3.41 \cdot 10^{-5}$	$4.07 \cdot 10^{-5}$	$3.69 \cdot 10^{-5}$	$3.43 \cdot 10^{-5}$	$3.26 \cdot 10^{-5}$	$3.83 \cdot 10^{-5}$	$3.43 \cdot 10^{-5}$

Πίνακας 8.5: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Neutral**.

Happy	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.29	1	0.29	0.3	0.29	0.02	0.18	0.09
max μετά	0.23	1	0.01	0.3	0.04	0.01	0.02	0.02
min πριν	0	0.7	0	0	0	0	0	0
min μετά	$1.57 \cdot 10^{-276}$	0.7	$6.10 \cdot 10^{-277}$	$1.64 \cdot 10^{-276}$	$9.24 \cdot 10^{-277}$	$6.08 \cdot 10^{-277}$	$7.23 \cdot 10^{-277}$	$6.69 \cdot 10^{-277}$
mean πριν	0.007	0.97	0.002	0.005	0.003	0.002	0.002	0.002
mean μετά	0.004	0.98	0.001	0.004	0.001	0.001	0.001	0.001

Πίνακας 8.6: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Happy**.

Sad	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.29	0.26	1	0.06	0.3	0.3	0.3	0.3
max μετά	0.29	0.03	1	0.04	0.3	0.3	0.3	0.3
min πριν	$4.62 \cdot 10^{-314}$	$1.45 \cdot 10^{-314}$	0.7	$1.56 \cdot 10^{-314}$	$1.73 \cdot 10^{-314}$	$3.49 \cdot 10^{-314}$	$2.62 \cdot 10^{-314}$	$2.6 \cdot 10^{-314}$
min μετά	0	0	0.7	0	0	0	0	0
mean πριν	0.02	0.003	0.94	0.003	0.003	0.008	0.004	0.006
mean μετά	0.01	0.003	0.95	0.003	0.003	0.008	0.005	0.006

Πίνακας 8.7: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Sad**.

Fear	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.18	0.03	0.29	0.29	1	0.3	0.29	0.3
max μετά	0.16	0.02	0.29	0.13	1	0.3	0.29	0.29
min πριν	0	0	0	0	0.7	0	0	0
min μετά	0	0	0	0	0.7	0	0	0
mean πριν	0.002	0.0008	0.001	0.006	0.97	0.002	0.004	0.003
mean μετά	0.002	0.0009	0.001	0.005	0.97	0.002	0.004	0.003

Πίνακας 8.8: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Fear**.

Disgust	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.29	0.01	0.29	0.29	0.29	1	0.29	0.3
max μετά	0.26	0.01	0.29	0.05	0.29	1	0.29	0.3
min πριν	$1.1 \cdot 10^{-303}$	$5 \cdot 10^{-304}$	$1.32 \cdot 10^{-303}$	$8.77 \cdot 10^{-304}$	$1.63 \cdot 10^{-303}$	0.7	$3.8 \cdot 10^{-302}$	$1.15 \cdot 10^{-301}$
min μετά	$1.61 \cdot 10^{-204}$	$8.05 \cdot 10^{-205}$	$2.12 \cdot 10^{-204}$	$1.4 \cdot 10^{-204}$	$2.62 \cdot 10^{-204}$	0.7	$6.09 \cdot 10^{-203}$	$2.87 \cdot 10^{-202}$
mean πριν	0.002	0.0006	0.002	0.001	0.003	0.94	0.015	0.02
mean μετά	0.002	0.0006	0.002	0.001	0.003	0.94	0.015	0.02

Πίνακας 8.9: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Disgust**.

Anger	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.29	0.02	0.29	0.29	0.3	0.3	1	0.3
max μετά	0.29	0.02	0.29	0.13	0.3	0.3	1	0.3
min πριν	0	0	0	0	0	0	0.7	0
min μετά	0	0	0	0	0	0	0.7	0
mean πριν	0.008	0.002	0.005	0.006	0.02	0.07	0.8	0.07
mean μετά	0.008	0.002	0.005	0.006	0.02	0.07	0.8	0.06

Πίνακας 8.10: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Anger**.

Contempt	Συναίσθημα							
	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
max πριν	0.29	0.18	0.08	0.3	0.3	0.29	0.3	1
max μετά	0.26	0.02	0.28	0.06	0.3	0.29	0.3	1
min πριν	$4.87 \cdot 10^{-133}$	$2.41 \cdot 10^{-133}$	$4.04 \cdot 10^{-133}$	$5.51 \cdot 10^{-133}$	$1.76 \cdot 10^{-132}$	$3.31 \cdot 10^{-131}$	$1.64 \cdot 10^{-125}$	0.7
min μετά	$8.57 \cdot 10^{-115}$	$4.66 \cdot 10^{-115}$	$7.72 \cdot 10^{-115}$	$1.02 \cdot 10^{-114}$	$3.14 \cdot 10^{-114}$	$7.6 \cdot 10^{-113}$	$9.37 \cdot 10^{-109}$	0.7
mean πριν	0.005	0.001	0.001	0.007	0.01	0.02	0.07	0.87
mean μετά	0.003	0.0009	0.001	0.002	0.01	0.01	0.07	0.89

Πίνακας 8.11: Αποτελέσματα των μοντέλων ταξινόμησης των 7 βασικών συναισθημάτων στο σύνολο αξιολόγησης για το συναίσθημα **Contempt**.

άλλαξαν οι τιμές max, min, mean άλλοτε κατά πολύ άλλοτε λίγο. Όσον αφορά το ουδέτερο συναίσθημα, βλέπουμε ότι μετά την εκκαθάριση μίκρυναν αρκετά οι τιμές των υπόλοιπων συναισθημάτων σε σύγκριση με πριν, το οποίο δείχνει ότι η μέθοδος αυτή ήταν επιτυχημένη αφού πλέον αναδεικνύεται η ουδετερότητα του συγκεκριμένου συναισθήματος και η μη συσχέτιση του με τα υπόλοιπα. Σχετικά με το χαρούμενο συναίσθημα, η δεύτερη μεγαλύτερη τιμή είναι του ουδέτερου συναισθήματος κάτι το οποίο είναι λογικό μιας και όπως είδαμε στο Κεφάλαιο 6 στις οριακές τιμές μπορεί μία ουδέτερη έκφραση να χαρακτηριστεί ως χαρά και το αντίστροφο. Για τα αρνητικά συναισθήματα sad, fear, disgust, anger, contempt βλέπουμε ότι έχουν κοντινές τιμές μεταξύ τους και απόκλιση από τα happy, surprise.

Μέρος **III**

Επίλογος

Επίλογος και Μελλοντικές Επεκτάσεις

Στο Κεφάλαιο αυτό συνοψίζουμε το περιεχόμενο της παρούσας διπλωματικής εργασίας και αναφέρουμε τα γενικά συμπεράσματα που προέκυψαν από την πειραματική μελέτη του θέματος που επιλέξαμε. Στην συνέχεια, προτείνουμε μελλοντικές επεκτάσεις των εργασιών που έχουμε ήδη επιτελέσει οι οποίες θα έχουν ως στόχο την ενδεχόμενη βελτίωση των αποτελεσμάτων.

9.1 Συμπεράσματα Διπλωματικής Εργασίας

Όπως έχουμε προαναφέρει, στόχος της παρούσας διπλωματικής ήταν η προσέγγιση του προβλήματος της αναγνώρισης ανθρωπίνων συναισθημάτων και της ανίχνευσης των τιμών σθένους και διέγερσης με μεθόδους μηχανικής μάθησης χρησιμοποιώντας εικόνες προσώπων. Προκειμένου να διευκολυνθεί η διαδικασία κατανόησης του προβλήματος και υλοποίησης μοντέλων για την επίλυση του, εκτελέσαμε μία εκτενή μελέτη του πεδίου της ανάλυσης συναισθημάτων και των θεωριών γύρω από αυτήν καθώς και του πεδίου της μηχανικής μάθησης δίνοντας την αρμόζουσα προσοχή στο θεωρητικό υπόβαθρο. Στο πλαίσιο της απόκτησης γνώσεων σχετικών με τους τομείς αυτούς, έπαιξε σημαντικό ρόλο και η μελέτη σχετικών ερευνητικών δημοσιεύσεων οι οποίες μας παρείχαν ιδέες που οδήγησαν στην τελική υλοποίηση της μεθόδου μας.

Ο συνδυασμός γνώσεων και για τα δύο πεδία αποτέλεσε την βάση για την εφαρμογή των δικών μας μεθόδων. Πιο συγκεκριμένα, καταφέραμε να αποφασίσουμε ποιο σύνολο δεδομένων θα ήταν το ιδανικό για την δική μας περίπτωση, να σχεδιάσουμε την μέθοδο της εκ νέου επισήμανσης εικόνων με βάση την σχετικότητα των ετικετών σθένους/διέγερσης και βασικών συναισθημάτων καθώς και να αποφασίσουμε ποιές παράμετροι (π.χ. τρόπος υπολογισμού αποστάσεων από κάθε cluster συναισθήματος) ήταν οι καλύτεροι, να επιλέξουμε τα κατάλληλα μοντέλα και τις κατάλληλες υπερπαραμέτρους και τελικά να αποφασίσουμε τις μετρικές αξιολόγησης που θα χρησιμοποιήσουμε, όλα αυτά αξιοποιώντας τις δυνατότητες που μας προσφέρει η γλώσσα προγραμματισμού Python.

Σαν τελικό συμπέρασμα της παρούσας διπλωματικής εργασίας, λαμβάνουμε το γεγονός ότι η αναγνώριση των ανθρωπίνων συναισθημάτων και των τιμών σθένους/διέγερσης είναι μία αρκετά πολύπλοκη διαδικασία που απαιτεί εκτεταμένη μελέτη και πειραματισμούς. Επίσης είδαμε ότι η σύζευξη των δύο αυτών εργασιών -οι οποίες υπό άλλες συνθήκες φαίνονται ασυσχέτιστες- μέσω της εκμετάλλευσης των ετικετών της μίας εργασίας για τον υπολογισμό

της άλλης εργασίας, εγγυάται καλύτερα αποτελέσματα από ότι η εκτέλεση της κάθε εργασίας ανεξάρτητα από την άλλη. Ωστόσο αυτό που δεν θα έπρεπε να ξεχνάμε είναι ότι η διάκριση των ανθρωπίνων συναισθημάτων καθώς και η αναγνώριση του πόσο ενεργητικό/παθητικό και θετικό/αρνητικό είναι ένα συναίσθημα, αποτελεί μερικές φορές μία δύσκολη διαδικασία ακόμα και για τον ίδιο τον άνθρωπο, πόσω μάλλον για τα νευρωνικά δίκτυα.

9.2 Μελλοντικές Επεκτάσεις

Παρά το γεγονός ότι προσπαθήσαμε η παρούσα διπλωματική να είναι όσο το δυνατόν πιο πλήρης, υπάρχει ένας μεγάλος αριθμός επεκτάσεων της προς διαφορετικές κατευθύνσεις.

Αρχικά, το σύνολο δεδομένων παίζει βασικό ρόλο στην επιτυχή εκπαίδευση και την απόδοση των μοντέλων μηχανικής μάθησης. Η επιλογή του συνόλου δεδομένων Affectnet ήταν μία αρκετά καλή επιλογή, ωστόσο περιείχε μερικά λάθη στις ετικέτες όπως έχουμε αναφέρει στο Κεφάλαιο 6 και επίσης θα μπορούσε να έχει περισσότερες εικόνες ειδικά όσον αφορά το σύνολο δοκιμής που περιείχε μόνο 500 ανά συναίσθημα. Μία ίσως ακόμα καλύτερη επιλογή θα ήταν αυτή της Aff-Wild2 [91] η οποία έχει πολύ μεγαλύτερο μέγεθος, επομένως το νευρωνικό θα εκπαιδευόταν και θα δοκιμαζόταν σε μεγαλύτερο αριθμό εικόνων δίνοντας έτσι ακριβέστερα αποτελέσματα.

Εξίσου σημαντικό ρόλο παίζει και η επιλογή του νευρωνικού δικτύου, στην περίπτωση μας του Resnet-50. Δεδομένου ότι η μεθοδολογία μας δουλεύει ανεξάρτητα από την επιλογή του συνελκτικού νευρωνικού δικτύου, ο πειραματισμός με άλλα μοντέλα και η βελτιστοποίηση των υπερπαραμέτρων τους ίσως οδηγούσε σε καλύτερα αποτελέσματα από τα δικά μας.

Επιπρόσθετα, δεδομένου ότι το multitask learning που δοκιμάσαμε είχε χαμηλή απόδοση όσον αφορά την ανίχνευση σθένους και διέγερσης, θα μπορούσαν να γίνουν περισσότεροι πειραματισμοί σε αυτό το κομμάτι προκειμένου να επιτευχθούν καλύτερα αποτελέσματα, κάτι το οποίο δεν υλοποιήσαμε εφόσον δεν ήταν στα πλαίσια της παρούσας διπλωματικής εργασίας. Επιπλέον, θα ήταν ενδιαφέρον να χρησιμοποιηθεί η μεθοδολογία μας κατά την εκπαίδευση ενός νευρωνικού με multitask learning μιας και πιθανότατα θα βελτίωνε την απόδοση.

Μία άλλη επέκταση της μεθοδολογίας μας θα ήταν αντί να αποτελεί ουσιαστικά προεπεξεργασία δεδομένων πριν την εκπαίδευση του νευρωνικού, να ενσωματωθεί το relabelling ως μέρος της εκπαίδευσης του νευρωνικού. Αυτό θα μπορούσε να το πετύχει κανείς συνδυάζοντας την ανίχνευση σθένους/διέγερσης με την αναγνώριση συναισθημάτων μέσω της δημιουργίας μίας εξειδικευμένης συνάρτησης απωλειών. Τέλος, θα μπορούσε κάποιος να εκμεταλλευτεί τις ετικέτες συναισθήματος προκειμένου να βγάλει συμπεράσματα για τις ετικέτες σθένους και διέγερσης, το αντίθετο δηλαδή από αυτό που υλοποιήσαμε εμείς.

Βιβλιογραφία

- [1] Ελένη-Γεωργία Αλεβιζάκου. *Η ΧΡΗΣΗ ΤΩΝ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΤΗΣ ΓΕΩΔΑΙΣΙΑΣ ΜΕ ΕΜΦΑΣΗ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΚΑΤΑΚΟΡΥΦΩΝ ΜΕΤΑΚΙΝΗΣΕΩΝ*. Μεταπτυχιακή διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, 2012.
- [2] *Basics of CNN in Deep Learning*. <https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/>. Ημερομηνία πρόσβασης: 21-06-2022.
- [3] *Training Neural Networks, Part I*. http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture6.pdf. Ημερομηνία πρόσβασης: 21-06-2022.
- [4] *Padding (Machine Learning)*. <https://deeptai.org/machine-learning-glossary-and-terms/padding>. Ημερομηνία πρόσβασης: 21-06-2022.
- [5] *Layers of a Convolutional Neural Network*. <https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network>. Ημερομηνία πρόσβασης: 31-07-2022.
- [6] *Convolutional Neural Networks (CNNs) and Layer Types*. <https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/>. Ημερομηνία πρόσβασης: 31-07-2022.
- [7] Huo Yingge, Imran Ali και Kang Yoon Lee. *Deep Neural Networks on Chip - A Survey*. σελίδες 589–592, 2020.
- [8] Άγγελος Γερούλανος. *Τεχνικές Βαθιάς Μηχανικής Μάθησης για Αναγνώριση Μουσικού Συναισθήματος*. Διπλωματική εργασία, ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ, 2021.
- [9] Xiangyu Ren Shaoqing Sun Jian He, Kaiming Zhang. *Deep Residual Learning for Image Recognition*. *Proceedings of the IEEE Computer Society Conference*, σελίδες 770–778, 2015.
- [10] *ML | Underfitting and Overfitting*. <https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/>. Ημερομηνία πρόσβασης: 11-9-2022.
- [11] *Setting the learning rate of your neural network*. <https://www.jeremyjordan.me/nn-learning-rate/>. Ημερομηνία πρόσβασης: 11-9-2022.
- [12] *Hidden Layer Activation Functions*. <https://kinder-chen.medium.com/hidden-layer-activation-functions-6fd65489ed25>. Ημερομηνία πρόσβασης: 11-9-2022.

- [13] Fazal Ahmed Talukdar Krishna Dharavath και Rabul Hussain Laskar. *Improving Face Recognition Rate with Image Preprocessing*. *Indian Journal of Science and Technology*, 7(8):1170–1175, 2014.
- [14] *How to Improve Image Processing in Facial Recognition Systems*. <https://visionify.ai/how-to-improve-image-processing-in-facial-recognition-systems/>. Ημερομηνία πρόσβασης: 6-9-2022.
- [15] *Understanding Confusion Matrix*. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>. Ημερομηνία πρόσβασης: 5-9-2022.
- [16] *Data Augmentation | How to use Deep Learning when you have Limited Data—Part 2*. <https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/>. Ημερομηνία πρόσβασης:12-9-2022.
- [17] *Random Brightness Image Augmentation - Keras ImageDataGenerator*. <https://www.studytonight.com/post/random-brightness-image-augmentation-keras-imagedatagenerator>. Ημερομηνία πρόσβασης:29-9-2022.
- [18] *Seven Universal Facial Expressions*. <https://www.paulekman.com/resources/universal-facial-expressions/>. Ημερομηνία πρόσβασης: 18-06-2022.
- [19] Yong Man Ro Seung Ho Lee. *Partial Matching of Facial Expression Sequence Using Over-Complete Transition Dictionary for Emotion Recognition*. *IEEE Transactions on Affective Computing*, 7(4):389 – 408, 2015.
- [20] Arvind Bansal Mehdi Ghayoumi. *Unifying Geometric Features and Facial Action Units for Improved Performance of Facial Expression Analysis*. 2016.
- [21] *Real-world Affective Faces Database*. <http://www.whdeng.cn/raf/model1.html#dataset>. Ημερομηνία πρόσβασης: 21-06-2022.
- [22] Tai Do Nhu, S.H. Kim, Hyung Jeong Yang, Guee Sang Lee και Soonja Yeom. *Context-Aware Emotion Recognition in the Wild Using Spatio-Temporal and Temporal-Pyramid Models*. *Sensors*, 21:2344, 2021.
- [23] S. Todorovic J. Kossaifi, G. Tzimiropoulos και M. Pantic. *AFEW-VA database for valence and arousal estimation in-the-wild*. *Image and Vision Computing*, 2017.
- [24] *AffectNet*. <http://mohammadmahoor.com/affectnet/>. Ημερομηνία πρόσβασης: 22-06-2022.
- [25] Behzad Hasani Ali Mollahosseini και Mohammad H. Mahoorn. *AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild*. *IEEE Transactions on Affective Computing*, 2017.
- [26] Miao Yu, Dimitrios Kollias, James Wingate, Niro Siriwardena και Stefanos Kollias. *Machine learning for predictive modelling of ambulance calls*. *Electronics*, 10(4):482, 2021.

- [27] Athanasios Tagaris, Dimitrios Kollias και Andreas Stafylopatis. *Assessment of Parkinson's disease based on deep neural networks. International Conference on Engineering Applications of Neural Networks*, σελίδες 391–403. Springer, 2017.
- [28] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και Stefanos D Kollias. *Transparent Adaptation in Deep Medical Image Diagnosis. TAILOR*, σελίδες 251–267, 2020.
- [29] Dimitrios Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Deep transparent prediction through latent representation analysis. arXiv preprint arXiv:2009.07044*, 2020.
- [30] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris και Stefanos Kollias. *Machine Learning for Neurodegenerative Disorder Diagnosis—Survey of Practices and Launch of Benchmark Dataset. International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018.
- [31] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare. Complex & Intelligent Systems*, 4(2):119–131, 2018.
- [32] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *AI-MIA: COVID-19 Detection & Severity Analysis through Medical Imaging. arXiv preprint arXiv:2206.04732*, 2022.
- [33] Anastasios Arsenos, Dimitrios Kollias και Stefanos Kollias. *A Large Imaging Database and Novel Deep Neural Architecture for Covid-19 Diagnosis. 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, σελίδες 1–5. IEEE, 2022.
- [34] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Miacov19d: Covid-19 detection through 3-d chest ct image analysis. Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537–544, 2021.
- [35] Dimitrios Kollias, Athanasios Tagaris και Andreas Stafylopatis. *On line emotion detection using retrainable deep neural networks. Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, σελίδες 1–8. IEEE, 2016.
- [36] Dimitrios Kollias και Stefanos Zafeiriou. *Training Deep Neural Networks with Different Datasets In-the-wild: The Emotion Recognition Paradigm. 2018 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [37] Dimitris Kollias, George Marandianos, Amaryllis Raouzaίου και Andreas Georgios Stafylopatis. *Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction. 2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, σελίδες 1–6. IEEE, 2015.

- [38] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models. Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, σελίδες 1–8. IEEE, 2017.
- [39] Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker και Rosalind W Picard. *Driver emotion recognition for intelligent vehicles: A survey. ACM Computing Surveys (CSUR)*, 53(3):1–30, 2020.
- [40] Mouhannad Ali, Ahmad Haj Mosa, Fadi Al Machot και Kyandoghene Kyamakya. *Emotion recognition involving physiological and speech signals: A comprehensive review. Recent advances in nonlinear dynamics and synchronization*, σελίδες 287–302, 2018.
- [41] Eihab Bashier Mohammed Bashier Mohssen Mohammed, Muhammad Badrudin Khan. *Machine Learning: Algorithms and Applications*. CRC Press, 2017.
- [42] Kevin Ferguson Max Pumperla. *Deep Learning and the game of Go*. Manning Publications Co., 2019.
- [43] Nishant Shukla. *Machine Learning with TensorFlow*. Manning Publications Co., 2018.
- [44] Στυλιανή Τερζάκη Παπαδοπούλου. *Πολυτροπική Προσέγγιση Αναγνώρισης Συναισθήματος Βασισμένη σε Δίκτυα Βαθιάς Μάθησης*. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, 2021.
- [45] Παναγιώτου Κωνσταντίνος. *Εφαρμογή των "NN για Αναγνώριση και Εντοπισμό Αντικειμένων Ψηλοποίηση στο Ενσωματωμένο Σύστημα Θεισον-ΤΚ1*. Διπλωματική εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2016.
- [46] Antoine Bordes Xavier Glorot και Yoshua Bengio. *Deep sparse rectifier neural networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 2011.
- [47] N. Zhang Y. Gao, O. Beijbom και T. Darrell. *Compact Bilinear Pooling. IEEE Computer Vision and Pattern Recognition*, (2):317–326, 2015.
- [48] K. Kavukcuoglu M. Ranzato Jarrett, K. και Y. LeCun. *What is the best multistage architecture for object recognition? Computer Vision, 2009 IEEE 12th International Conference*, 2009.
- [49] Alex Krizhevsky. *Convolutional deep belief networks on cifar-10. Computer Vision, 2010 IEEE 13th International Conference*, 2010.
- [50] Sergey Ioffe και Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015.

- [51] ΧΡΗΣΤΟΣ ΘΕΟΔΩΡΟΠΟΥΛΟΣ. *Χρήση τεχνικών βαθιάς μάθησης για την αναγνώριση συναισθημάτων μέσα από εκφράσεις του προσώπου*. Διπλωματική εργασία, ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, 2018.
- [52] Τίτα Παγουλάτου . *Βαθιά Μάθηση με Βελτιστοποίηση Υπερ-παραμέτρων με την Μέθοδο Σμήνους Σωματιδίων για Πρόβλεψη Χρήσης Πόρων σε Υπολογιστικές Υποδομές των Άκρων* . Διπλωματική εργασία, ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, 2020.
- [53] *What are Hyperparameters ? and How to tune the Hyperparameters in a Deep Neural Network?* <https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a>. Ημερομηνία πρόσβασης:10-9-2022.
- [54] *Understand the Impact of Learning Rate on Neural Network Performance.* <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/>. Ημερομηνία πρόσβασης:11-9-2022.
- [55] ΕΛΕΝΑ Γ. ΒΕΡΓΟΠΟΥΛΟΥ . *Ανάλυση Συναισθημάτων Σε Εικόνες Με Χρήση Γεννητικών Ανταγωνιστικών Δικτύων* . Διπλωματική εργασία, ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ, 2021.
- [56] *A Quick Guide on Preprocessing Facial Images for Neural Networks using OpenCV in Python.* <https://medium.com/yottabytes/a-quick-guide-on-preprocessing-facial-images-for-neural-networks-using-opencv-in-python-47ee3438abd4>. Ημερομηνία πρόσβασης: 6-9-2022.
- [57] Nawaf Hazim Barnouti. *Improve Face Recognition Rate Using Different Image Pre-Processing Techniques. American Journal of Engineering Research (AJER)*, 5(4):46–53, 2016.
- [58] *Statistical Package for Social Sciences.* https://eclass.uth.gr/modules/document/file.php/PE_U_164/8%20%CE%97%CE%A5%20-%20SPSS%20Lesson%208%202020.pdf. Ημερομηνία πρόσβασης: 6-9-2022.
- [59] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.
- [60] P Ekman. *Facial expression and emotion. American Psychologist*, 48(4):384–392, 1993.
- [61] Βουτυρά, Αναστασία. *Ο ρόλος της έκφρασης συναισθημάτων μέσα στην οικογένεια και της συναισθηματικής νοημοσύνης νεαρών ατόμων στις ακαδημαϊκές τους επιδόσεις*. Διδακτορική Διατριβή, Χαροκόπειο Πανεπιστήμιο, 2009.
- [62] Paul Ekman. *Facial Expressions*. University of California, 1999.
- [63] *About Paul Ekman.* <https://www.paulekman.com/about/paul-ekman/>. Ημερομηνία πρόσβασης: 18-06-2022.

- [64] Γιατί η έκφραση των συναισθημάτων είναι ίδια σε όλο τον κόσμο; - Η θεωρία του Έκμαν. <http://www.viralgr.com/giati-i-ekfrasi-ton-synaisthimaton-einai-idia-se-olo-ton-kosmo-i-theoria-toy-ekman/>. Ημερομηνία πρόσβασης: 18-06-2022.
- [65] Bakeman R. Chen Y. Norris K. Thomas R. Camras, L.A. *Culture, Ethnicity, and Children's Facial Expressions: A Study of European American, Mainland Chinese, Chinese American, and Adopted Chinese Girls. Emotion*, (6):103-114, 2006.
- [66] Caldara R. Schyns P.G. Jack, R.E. *Internal Representations Reveal Cultural Diversity in Expectations of Facial Expressions of Emotion. Journal of Experimental Psychology*, (141):19-25, 2012.
- [67] *Facial Expression Differences Between Chinese and European Americans.* <https://sites.miis.edu/eappicc/2016/04/11/facial-expression-differences-between-chinese-and-european-americans/>. Ημερομηνία πρόσβασης: 18-06-2022.
- [68] T. Kanade Y. I. Tian και J. F. Cohn. *Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97-115, 2001.
- [69] Λώρα (Καλλιόπη-Ειρήνη) Μαλατέστα. *Επικοινωνία Ανθρώπου-Υπολογιστή Βασισμένη σε Ανάλυση και Σύλληψη Οπτικών Πληροφοριών. Διδακτορική Διατριβή, ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ*, 2009.
- [70] W. Wundt. *Grundzuge der Physiologischen Psychologie*. Engelmann, Leipzig, 1η έκδοση, 1993.
- [71] H. Schlosberg. *A scale for the judgment of facial expressions. Journal of Experimental Psychology*, 29:497-510, 1941.
- [72] H. Schlosberg. *Three dimensions of emotion. Psychological review*, 61(2):81-88, 1954.
- [73] J. A. Russell. *A circumplex model of affect. Journal of Personality and Social Psychology*, 39:1161-1178, 1980.
- [74] C.M. Whissell. *The Measurement of Emotions: Theory, Research and Experience: The Measurement of Emotions*. Academic Press, 4η έκδοση, 1989.
- [75] R. Plutchik. *A general psychoevolutionary theory of emotion*. 1η έκδοση, 1980.
- [76] W.J. Lancee E.M. Daly και J. Polivy. *A conical model for the taxonomy of emotional experience. Journal of Personality and Social Psychology*, 45:443-457, 1983.
- [77] Neveux. *Valence and arousal in WHAT and HOW exclamatives: cognitive simplification versus emotive implication. Journal in English Lexicology*, 13, 2019.
- [78] Kate E.Walton David S.Kosson Samantha Fonda Jessica Simonetti Malek Mneimnea, Alice S.Powers. *Emotional valence and arousal effects on memory and hemispheric asymmetries*. 1η έκδοση, 2010.

- [79] Tuerlinckx F. Russell J. A. Barrett L. F. Kuppens, P. *The relation between valence and arousal in subjective experience. Psychological Bulletin*, 139:917–940, 2013.
- [80] Shan Li, Weihong Deng και JunPing Du. *Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 2584–2593. IEEE, 2017.
- [81] Shan Li και Weihong Deng. *Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition. IEEE Transactions on Image Processing*, 28(1):356–370, 2019.
- [82] *AFEW-VA Database for Valence and Arousal Estimation In-The-Wild*. <https://ibug.doc.ic.ac.uk/resources/afew-va-database/>. Ημερομηνία πρόσβασης: 22-06-2022.
- [83] S. Lucey A. Dhall, R. Goecke και T. Gedeon. *Collecting Large, Richly Annotated Facial-Expression Databases from Movies. IEEE MultiMedia*, 19(3):34–41, 2012.
- [84] Kevin Bartlett Philip Trinh S Mohammad Mavadati, Mohammad H Mahoor και Jeffrey F Cohn. *Disfa: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.
- [85] Peng X. Yang J. Meng D. Qiao Y. Wang, K. *Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition. 2019*.
- [86] Dimitrios Kollias, Shiyang Cheng, Evangelos Ververas, Irene Kotsia και Stefanos Zafeiriou. *Deep Neural Network Augmentation: Generating Faces for Affect Analysis. International Journal of Computer Vision*, 128, 2020.
- [87] Dimitrios Kollias, Shiyang Cheng, Maja Pantic και Stefanos Zafeiriou. *Photorealistic Facial Synthesis in the Dimensional Affect Space. European Conference on Computer Vision*, σελίδες 475–491. Springer, 2018.
- [88] Dimitrios Kollias και Stefanos Zafeiriou. *VA-StarGAN: Continuous Affect Generation*, σελίδες 227–238. 2020.
- [89] Dimitrios Kollias και Stefanos Zafeiriou. *Exploiting multi-CNN features in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset*, 2019.
- [90] Dimitrios Kollias και Stefanos Zafeiriou. *A Multi-component CNN-RNN Approach for Dimensional Emotion Recognition in-the-wild. arXiv preprint arXiv:1805.01452*, 2018.
- [91] Dimitrios Kollias, Panagiotis Tzirakis, Mihalis Nicolaou, Athanasios Papaioannou, Guoying Zhao, Björn Schuller, Irene Kotsia και Stefanos Zafeiriou. *Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond. International Journal of Computer Vision*, 127, 2019.

- [92] Dimitrios Kollias, Mihalis A Nicolaou, Irene Kotsia, Guoying Zhao και Stefanos Zafeiriou. *Recognition of affect in the wild using deep neural networks*. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, σελίδες 1972–1979. IEEE, 2017.
- [93] Stefanos Zafeiriou, Dimitrios Kollias, Mihalis A Nicolaou, Athanasios Papaioannou, Guoying Zhao και Irene Kotsia. *Aff-Wild: Valence and Arousal In-The-Wild Challenge*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, σελίδες 34–41, 2017.
- [94] Dimitrios Kollias και Stefanos Zafeiriou. *Expression, Affect, Action Unit Recognition: Aff-Wild2, Multi-Task Learning and ArcFace*. *arXiv preprint arXiv:1910.04855*, 2019.
- [95] Dimitrios Kollias και Stefanos Zafeiriou. *Affect Analysis in-the-wild: Valence-Arousal, Expressions, Action Units and a Unified Framework*. *arXiv preprint arXiv:2103.15792*, 2021.
- [96] Dimitrios Kollias και Stefanos Zafeiriou. *Aff-Wild2: Extending the Aff-Wild Database for Affect Recognition*. *arXiv preprint arXiv:1811.07770*, 2018.
- [97] Dimitrios Kollias, Attila Schulc, Elnar Hajiyev και Stefanos Zafeiriou. *Analysing Affective Behavior in the First ABAW 2020 Competition*. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*, σελίδες 794–800. IEEE Computer Society, 2020.
- [98] Dimitrios Kollias. *Abaw: Valence-arousal estimation, expression recognition, action unit detection & multi-task learning challenges*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2328–2336, 2022.
- [99] Dimitrios Kollias και Stefanos Zafeiriou. *Analysing affective behavior in the second abaw2 competition*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 3652–3660, 2021.
- [100] Dimitrios Kollias. *Abaw: Learning from synthetic data & multi-task learning challenges*. *arXiv preprint arXiv:2207.01138*, 2022.
- [101] Dimitrios Kollias, Viktoriia Sharmanska και Stefanos Zafeiriou. *Distribution Matching for Heterogeneous Multi-Task Learning: a Large-scale Face Study*, 2021.
- [102] Dimitrios Kollias, Viktoriia Sharmanska και Stefanos Zafeiriou. *Face Behavior a la carte: Expressions, Affect and Action Units in a Single Network*. *arXiv preprint arXiv:1910.11111*, 2019.
- [103] M.U.; Khawaja S.G.; Pujari A.N. Dar, M.N.; Akram. *CNN and LSTM-Based Emotion Charting Using Physiological Signals*. *Signal Processing Using Non-invasive Physiological Sensors*, 2020.
- [104] Marko Tkalcic, Andrej Kosir, Jurij Tasic και Matevž Kunaver. *Affective recommender systems: the role of emotions in recommender systems*. σελίδες 9–13, 2011.

- [105] *DColaboratory*. <https://research.google.com/colaboratory/faq.html>. Ημερομηνία πρόσβασης: 13-9-2022.