



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Πρωτόκολλο για δίκαιη αναγνώριση  
συναισθηματικών εκφράσεων προσώπου  
με χρήση βαθιών νευρωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΕΛΕΝΗΣ Α. ΠΑΠΑΔΟΠΟΥΛΟΥ

Επιβλέπων: Στέφανος Κόλλιας  
Καθηγητής

Αθήνα, Σεπτέμβριος 2022

---





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Πρωτόκολλο για δίκαιη αναγνώριση  
συναισθηματικών εκφράσεων προσώπου με  
χρήση βαθιών νευρωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΕΛΕΝΗΣ Α. ΠΑΠΑΔΟΠΟΥΛΟΥ

Επιβλέπων: Στέφανος Κόλλιας  
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28<sup>η</sup> Σεπτεμβρίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Στέφανος Κόλλιας  
Καθηγητής

.....  
Αθανάσιος Βουλόδημος  
Επίκουρος Καθηγητής

.....  
Γιώργος Στάμου  
Καθηγητής

Αθήνα, Σεπτέμβριος 2022





Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.  
Ελένη Παπαδοπούλου, 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)

.....  
Ελένη Παπαδοπούλου

28<sup>η</sup> Σεπτεμβρίου 2022



## Περίληψη

---

Η εργασία της ανάλυσης και αναγνώρισης εκφράσεων προσώπου διαδραματίζει βασικό ρόλο στη μοντελοποίηση συμπεριφοράς, στην αλληλεπίδραση ανθρώπου - μηχανής και στον συναισθηματικό υπολογισμό, με ποικίλες εφαρμογές σε διάφορους τομείς. Λόγω της κοινωνικής φύσης της εργασίας αυτής πρέπει να είναι δίκαιη απέναντι σε όλους τους ανθρώπους ανεξάρτητα από τη φυλή, την ηλικία και το φύλο τους. Επομένως, τα συστήματα που αναπτύσσονται για την υλοποίηση της εργασίας αυτής θα πρέπει να είναι ηθικά και απαλλαγμένα από άδικες προκαταλήψεις.

Η ανάλυση και αναγνώριση εκφράσεων προσώπου αναπτύσσεται με χρήση μηχανικής μάθησης από τρεις ετερογενείς εργασίες: την αναγνώριση βασικών κατηγορικών συναισθημάτων (π.χ. θυμός, φόβος, ευτυχία, λύπη), την εκτίμηση διαστατικών αναπαραστάσεων συναισθημάτων από τις τιμές valence (σθένος) και arousal (διέγερση) και την ανίχνευση ενεργοποιημένων κινήσεων των μυών του προσώπου οι οποίες περιγράφονται ως Action Units. Για την εκπαίδευση, αξιολόγηση και τον έλεγχο των συστημάτων χρησιμοποιούνται βάσεις δεδομένων με έγκυρες επισημάνσεις ως προς τις τρεις εργασίες συναισθηματικού υπολογισμού. Οι βάσεις αυτές παρέχουν μεγάλο αριθμό και μια ποικιλία δειγμάτων και στα δύο φύλα, εθνικότητα, και ηλικίες αλλά δεν είναι ισότιμα και δίκαια διαμορφωμένες στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου.

Στόχος της διπλωματικής εργασίας είναι η δίκαιη αναδιοργάνωση των βάσεων δεδομένων στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου με βάση τη φυλή, την ηλικία, το φύλο και τις επισημάνσεις της εργασίας συναισθηματικού υπολογισμού που αντιμετωπίζουν. Η ανάπτυξη δηλαδή ενός πρωτοκόλλου, το οποίο θα ανακατασκευάζει τις υπάρχουσες βάσεις δεδομένων στα τρία σύνολα με ένα δίκαιο τρόπο.

Για κάθε εργασία αναγνώρισης συναισθηματικών εκφράσεων προσώπου, υλοποιείται ένα μοντέλο το οποίο εκπαιδεύεται και αξιολογείται πριν και μετά την εφαρμογή του δίκαιου χωρισμού στις βάσεις δεδομένων. Καταδεικνύεται ότι η προσέγγιση αυτή επιφέρει κοντινότερα αποτελέσματα απόδοσης στις πλείστες κατηγορίες φυλής, ηλικίας και φύλου. Ταυτόχρονα προσφέρει ομοιομορφία αφού αναδιοργανώνονται οι βάσεις δεδομένων στα τρία σύνολα εκπαίδευσης με ένα συνεχή τρόπο.

## Λέξεις Κλειδιά

Μηχανική Μάθηση, Βαθιά Νευρωνικά Δίκτυα, Εικόνες, Βίντεο, Αναγνώριση Συναισθηματικών Εκφράσεων Προσώπου, Συναισθηματικός Υπολογισμός, Ταξινόμηση, Εκτίμηση, Ανίχνευση, Βασικά Συναισθήματα, διέγερση, σθένος, Action Units, Δικαιοσύνη, Σύνολο Εκπαίδευσης, Σύνολο Επικύρωσης, Σύνολο Ελέγχου, Κατανομή Δεδομένων





# Abstract

---

The task of Affect analysis and recognition, based on a subject's facial expressions, performs a key role in behavioural modelling, human machine interaction and affective computing, with applications in a variety of fields. Due to the social nature of this task it needs to be inclusive towards all people regardless of their race, age and gender. Thus, it should be ethical and free from unjust biases.

Affective computing is developed using machine learning and consists of three heterogeneous tasks; recognition of basic categorical emotions (e.g. anger, fear, happiness, sadness), estimation of dimensional representations of emotions from the values of valence and arousal and detection of active facial muscle movements which are described as Action Units (AUs).

In the affective computing task, there are several databases that provide a large number of samples with accurate annotations of basic expressions, valence-arousal and AUs. These databases provide a large number and a variety of samples in both genders, ethnicities, and ages, but are not equally and fairly distributed in the three sets of training, validation, and test. The aim of this thesis is to fairly reorganize the databases into the three sets of training, validation and test, based on the value of race, age, gender and the affective computing task that they tackle. In other words, the development of a protocol which will reconstruct the existing databases in the three sets in a fair way.

For each facial expression recognition task, an individual model is implemented which is trained and evaluated before and after the application of the fair split on the databases. The approach illustrates closer performance results across most race, age and gender categories. In addition, uniformity is resorted as the databases are reorganized into the three sets in a consistent way

## Keywords

Machine Learning, Deep Neural Networks, Images, Videos, Facial Emotion Recognition, Affect Recognition, Classification, Estimation, Detection, Basic Emotions, Arousal, Valence, Action Units, Fairness, Training Set, Validation Set, Test Set, Distribution of Data



*στην οικογένεια μου*



## Ευχαριστίες

---

Κατ' αρχάς θα ήθελα να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης. Επίσης ευχαριστώ ιδιαίτερα τους συν επιβλέποντες Διδάκτορες, κ.Δημήτρη Κόλλια και κ.Παρασκευή Τζούβελη για την πολύτιμη βοήθεια και καθοδήγησή τους καθ' όλη τη διάρκεια της προετοιμασίας και της εκπόνησης της παρούσας διπλωματικής εργασίας. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Αθήνα, Σεπτέμβριος 2022

*Ελένη Παπαδοπούλου*



# Περιεχόμενα

---

Περίληψη	1
Abstract	3
Ευχαριστίες	7
<b>1 Εισαγωγή</b>	<b>17</b>
1.1 Αντικείμενο της διπλωματικής	17
1.2 Οργάνωση του τόμου	18
<b>I Θεωρητικό Μέρος</b>	<b>19</b>
<b>2 Θεωρία Συναισθηματικού Υπολογισμού</b>	<b>21</b>
2.1 Εργασία 1: Αναγνώριση βασικών συναισθημάτων	21
2.2 Εργασία 2: Εκτίμηση τιμών valence και arousal	21
2.3 Εργασία 3: Ενεργοποίηση τιμών Action Units	22
<b>3 Θεωρία μοντέλων μηχανικής μάθησης</b>	<b>25</b>
3.1 Μηχανική μάθηση	25
3.2 Νευρωνικά δίκτυα	26
3.2.1 Μετρικές αξιολόγησης νευρωνικών δικτύων	27
3.2.2 Υπερπαραμέτροι νευρωνικών δικτύων	29
3.2.3 Συνελικτικά νευρωνικά δίκτυα	32
3.3 Classification task	36
3.4 Regression task	37
<b>II Πρακτικό Μέρος</b>	<b>39</b>
<b>4 Μεθοδολογία</b>	<b>41</b>
4.1 Μοντέλα αναγνώρισης φυλής, ηλικίας και φύλου	41
4.2 Βάσεις δεδομένων	43
4.2.1 Παρουσίαση βάσεων	43
4.2.2 Στατιστικά βάσεων	47
4.3 Δίκαιος Διαχωρισμός	57
4.4 Μοντέλα εργασιών αναγνώρισης έκφρασης	60

<b>5</b>	<b>Σχετικές εργασίες</b>	<b>61</b>
<b>6</b>	<b>Λεπτομέρειες Εκπαίδευσης</b>	<b>63</b>
6.1	Μεθοδολογία βελτιστοποίησης . . . . .	63
6.2	Τελικά δίκτυα και υπερπαραμέτροι . . . . .	63
6.3	Προ-επεξεργασία εικόνων . . . . .	64
6.3.1	Μοντέλα ανίχνευσης φυλής, ηλικίας και φύλου . . . . .	64
6.3.2	Μοντέλα αναγνώρισης συναισθηματικής κατάστασης . . . . .	64
<b>7</b>	<b>Αποτελέσματα</b>	<b>67</b>
7.1	Αποτελέσματα μοντέλων ανίχνευσης φυλής, ηλικίας και φύλου . . . . .	67
7.2	Αποτελέσματα βελτιστοποίησης βάσης RAF-DB . . . . .	68
7.3	Αποτελέσματα δίκαιου διαχωρισμού βάσεων . . . . .	69
7.3.1	Πίνακες αποτελεσμάτων . . . . .	69
7.3.2	Σχολιασμοί . . . . .	75
<b>III</b>	<b>Επίλογος</b>	<b>77</b>
<b>8</b>	<b>Επίλογος</b>	<b>79</b>
8.1	Συμπεράσματα . . . . .	79
8.2	Μελλοντικές Επεκτάσεις . . . . .	80
	<b>Βιβλιογραφία</b>	<b>88</b>



## Κατάλογος Σχημάτων

---

4.1	Σχηματική αναπαράσταση της μεθοδολογίας του δίκαιου διαχωρισμού . . . . .	58
4.2	Σχηματική αναπαράσταση της μεθοδολογίας του δίκαιου διαχωρισμού για τη βάση RAF-DB . . . . .	59



## Κατάλογος Εικόνων

---

2.1	Six basic emotions [1] . . . . .	22
2.2	Most common AUs [2] . . . . .	23
3.1	Τα επίπεδα ενός νευρωνικού δικτύου . . . . .	26
3.2	Οι υπολογισμοί που εκτελεί ένας υπολογιστικός νευρώνας . . . . .	27
3.3	Confusion Matrix for binary classification problem . . . . .	28
3.4	Γραφική παράσταση γραμμικής συνάρτησης ενεργοποίησης . . . . .	30
3.5	Γραφική παράσταση σιγμοειδές συνάρτησης ενεργοποίησης . . . . .	31
3.6	Γραφική παράσταση συνάρτησης ενεργοποίησης tanh . . . . .	31
3.7	Γραφική παράσταση συνάρτησης ενεργοποίησης ReLU . . . . .	31
3.8	Softmax formula . . . . .	32
3.9	The impementation of a Convolutional Layer . . . . .	33
3.10	Resnet Block: skip connection . . . . .	34
3.11	Architecture of Resnet-50 Network . . . . .	34
3.12	Architecture of VGG-16 Network . . . . .	35
3.13	Dense Connection . . . . .	35
3.14	Architecture of DenseNet-121 Network . . . . .	35
3.15	Binary cross entropy loss function. $\hat{y} : prediction, y : label$ . . . . .	37
3.16	Categorical cross entropy loss function. $\hat{y} : prediction, y : label$ . . . . .	37
3.17	Mean absolute error loss function. $y : prediction, x : label$ . . . . .	38
3.18	Mean squared error loss function. $\hat{Y} : prediction, Y : label$ . . . . .	38
3.19	Concordance Correlation Coefficient . . . . .	38
4.1	Εξόδοι του μοντέλου Face Analysis Pytorch με είσοδο τη βάση δεδομένων AffectNet . . . . .	42
4.2	Εξόδοι του μοντέλου Deepface με είσοδο τη βάση δεδομένων AffectNet . . . . .	42
4.3	Εξόδοι του μοντέλου Fairface με είσοδο τη βάση δεδομένων AffectNet . . . . .	42
4.4	Εξόδοι του μοντέλου FaceRec με είσοδο τη βάση δεδομένων AffectNet . . . . .	43
4.5	AffectNet, aligned images with annotations . . . . .	44
4.6	RAF-DB, aligned images with basic emotions annotations . . . . .	44
4.7	RAF-AU, aligned images with AUs annotations . . . . .	45
4.8	EmotioNet, aligned images with AUs annotations . . . . .	45
4.9	GFT, aligned images with AUs annotations . . . . .	46
4.10	AFEW-VA, aligned video frames with valence & arousal annotations . . . . .	46
4.11	DISFA, aligned images with AUs annotations . . . . .	47

4.12	Valence & Arousal Distribution, BEFORE and AFTER split, AffectNet	. 50
4.13	Valence & Arousal Distribution, BEFORE and AFTER split, AFEW-VA	. 51
6.1	Example of face detection and alignment [3]	. . . . . 65
6.2	Face detection and alignment of GFT database (video: 004C, frame: 15)	. 65
6.3	Data augmentation examples: Brightness and Flipping, RAF-DB database	65

## Κατάλογος Πινάκων

---

4.1	Statistics BEFORE and AFTER fair split,RAF-DB, Emotion Recognition	48
4.2	Statistics BEFORE and AFTER fair split, AffectNet, Emotion Recognition	49
4.3	Statistics AFTER fair split, AffectNet, Valence & Arousal Estimation . . .	51
4.4	Statistics BEFORE and AFTER fair split, AFEW-VA, Valence & Arousal Estimation . . . . .	52
4.5	AUs activation BEFORE and AFTER fair split, EmotioNet . . . . .	53
4.6	Statistics BEFORE and AFTER fair split, EmotioNet, AUs Detection . . .	53
4.7	AUs activation BEFORE and AFTER fair split, DISFA . . . . .	54
4.8	Statistics BEFORE and AFTER fair split, DISFA, AUs Detection . . . . .	54
4.9	AUs activation BEFORE and AFTER fair split, GFT . . . . .	55
4.10	Statistics BEFORE and AFTER fair split, GFT, AUs Detection . . . . .	55
4.11	AUs activation BEFORE and AFTER fair split, RAF-AU . . . . .	55
4.12	Statistics BEFORE and AFTER fair split, RAF-AU, AUs Detection . . . .	56
6.1	Statistics BEFORE and AFTER fair split,RAF-DB, Emotion Recognition	64
7.1	Απόδοση των μοντέλων FairFace και FaceRec στην αναγνώριση φυλής, ηλικία και φύλου . . . . .	68
7.2	Optimization for optimizer (Adam, SGD) and learning rate (0.0001 - 0.1) .	68
7.3	Optimization for batch size (16,32,64) . . . . .	68
7.4	Optimization for Architecture (number of layers) and number of units(512, 1024) . . . . .	69
7.5	Optimization for activation function (relu, sigmoid) and DROPOUT layer	69
7.6	Best hyperparameters for RAF-DB, Emotion Recognition . . . . .	69
7.7	BEFORE and AFTER fair split results for RAF-DB database without data augmentation, Emotion Recognition task in all categories . . . . .	70
7.8	BEFORE and AFTER fair split results for RAF-DB database with data augmentation, Emotion Recognition task in all categories . . . . .	70
7.9	BEFORE and AFTER fair split results for AffectNet database without data augmentation, Emotion Recognition task in all categories . . . . .	70
7.10	BEFORE and AFTER fair split results for AffectNet database with data augmentation, Emotion Recognition task in all categories . . . . .	71
7.11	BEFORE and AFTER fair split results for AffectNet database without data augmentation, Valence and Arousal estimation task in all categories .	71

7.12	BEFORE and AFTER fair split results for AffectNet database with data augmentation, Valence and Arousal estimation task in all categories . . . .	71
7.13	BEFORE and AFTER fair split results for AFEW-VA database without data augmentation, Valence and Arousal estimation task in all categories .	72
7.14	BEFORE and AFTER fair split results for AFEW-VA database with data augmentation, Valence and Arousal estimation task in all categories . . . .	72
7.15	BEFORE and AFTER fair split results for EmotioNet database without data augmentation, AUs detection task in all categories . . . . .	72
7.16	BEFORE and AFTER fair split results for EmotioNet database with data augmentation, AUs detection task in all categories . . . . .	73
7.17	BEFORE and AFTER fair split results for DISFA database without data augmentation, AUs detection task in all categories . . . . .	73
7.18	BEFORE and AFTER fair split results for DISFA database with data augmentation, AUs detection task in all categories . . . . .	73
7.19	BEFORE and AFTER fair split results for GFT database without data augmentation, AUs detection task in all categories . . . . .	73
7.20	BEFORE and AFTER fair split results for GFT database with data augmentation, AUs detection task in all categories . . . . .	74
7.21	BEFORE and AFTER fair split results for RAF-AU database without data augmentation, AUs detection task in all categories . . . . .	74
7.22	BEFORE and AFTER fair split results for RAF-AU database with data augmentation, AUs detection task in all categories . . . . .	74

## Κεφάλαιο **1**

### Εισαγωγή

---

**Η** ανάλυση και αναγνώριση συναισθημάτων διαδραματίζει σημαντικό ρόλο σε θέματα μο-  
ντελοποίησης συμπεριφοράς, αλληλεπίδρασης μηχανής και ανθρώπου και συναισθημα-  
τικού υπολογισμού. Εφαρμόζεται σε ποικίλους τομείς όπως τον τομέα της ιατρικής, της υγείας,  
της ηλεκτρονικής μάθησης, της νομικής, του μάρκετινγκ, της ανίχνευσης ψεύδους κ.α. Συγκεκρι-  
μενά κάποια παραδείγματα εφαρμογών αποτελούν την ανίχνευση εκφράσεων του προσώπου  
και προσδιορισμός της κόπωσης ενός οδηγού αυτοκινήτου, η μελέτη των συναισθημάτων των  
μαθητών και προσαρμογή των τεχνικών μάθησης σύμφωνα με τις αντιδράσεις τους, η ανίχνευ-  
ση της συναισθηματικής κατάστασης των υποψήφιων κατά τη διάρκεια μιας συνέντευξης, ο  
εντοπισμός θετικών ή αρνητικών αντιδράσεων του κοινού σε εκδηλώσεις κ.α. Η εργασία αυ-  
τή αναφέρεται στην αναγνώριση ανθρωπίνων προσώπων και γι' αυτό συμπεριλαμβάνεται ένας  
κοινωνικός παράγοντας στη διεκπεραίωσή της. Τα μοντέλα μηχανικής μάθησης που εκπαιδε-  
ύονται σε εφαρμογές της εργασίας αυτής πρέπει να χαρακτηρίζονται από δικαιοσύνη και όχι  
από μεροληψία.

Η δικαιοσύνη στη μηχανική μάθηση αναφέρεται στις προσπάθειες διόρθωσης αλγοριθμικής  
μεροληψίας σε διάφορες εφαρμογές που βασίζονται σε μοντέλα μηχανικής μάθησης. Οι απο-  
φάσεις που λαμβάνονται από υπολογιστές μετά από μια διαδικασία μηχανικής μάθησης μπορεί  
να θεωρηθούν άδικες εάν βασίζονταν σε μεταβλητές που θεωρούνται ευαίσθητες, όπως το  
φύλο, η εθνικότητα, η ηλικία κ.α. Όπως συμβαίνει με πολλές ηθικές έννοιες, οι ορισμοί της  
δικαιοσύνης και της μεροληψίας είναι πάντα αμφιλεγόμενοι. Γενικά, η δικαιοσύνη και η μερο-  
ληψία θεωρούνται συναφείς έννοιες, όταν η διαδικασία λήψης αποφάσεων επηρεάζει τις ζωές  
των ανθρώπων.

#### 1.1 Αντικείμενο της διπλωματικής

Οι πλείστες βάσεις δεδομένων που χρησιμοποιούνται για την εργασία αναγνώρισης έκ-  
φρασης δεν λαμβάνουν υπόψη την ηλικία, την εθνικότητα και το φύλο για την κατανομή  
των δεδομένων στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Ως αποτέλεσμα οι  
αποφάσεις που εξάγουν τα μοντέλα με τις βάσεις αυτές μπορούν να θεωρηθούν άδικες και με-  
ροληπτικές. Ταυτόχρονα οι βάσεις δεδομένων που μελετούνται δεν παρέχουν διαμερισμό και  
στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου κάτι που δυσκολεύει τη διαδικασία εκ-  
μάθησης. Επίσης, κάποιες βάσεις δεδομένων δεν έχουν σωστές κατανομές των επισημάνσεων  
τους στα σύνολα εκπαίδευσης, με αποτέλεσμα τα μοντέλα να μην αποδίδουν τόσο καλά.

Αντικείμενο της διπλωματικής αυτής είναι η προσαρμογή των υπαρχουσών βάσεων δεδομένων στα τρία σύνολα εκπαίδευσης με στόχο να αντιμετωπιστούν τα προβλήματα που προαναφέρθηκαν. Συγκεκριμένα για κάθε βάση δεδομένων που μελετάται λαμβάνονται αυτόματες ή χειροκίνητες επισημάνσεις για τις κατηγορίες φύλο, ηλικία και εθνικότητα. Έπειτα διαμοιράζονται δίκαια οι βάσεις δεδομένων στα τρία σύνολα εκπαίδευσης με γνώμονα τις κατηγορίες αυτές και την εργασία συναισθηματικού υπολογισμού που αντιμετωπίζουν. Για την αξιολόγηση του δίκαιου χωρισμού στις βάσεις δημιουργούνται μοντέλα που επιτελούν κάθε εργασία, τα οποία εκπαιδεύονται και ελέγχονται με τις βάσεις δεδομένων πριν και μετά το διαχωρισμό.

## 1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε 8 κεφάλαια: Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο όσον αφορά την ανάλυση και την αναγνώριση συναισθημάτων από εικόνες. Στο Κεφάλαιο 3 παρουσιάζεται η θεωρία πίσω από τη μηχανική μάθηση και τα νευρωνικά δίκτυα με έμφαση σε ορολογίες που θα χρησιμοποιηθούν κατά την πρακτική υλοποίηση της εργασίας. Στο Κεφάλαιο 4 περιγράφεται αναλυτικά η μεθοδολογία που έλαβε μέρος για τη διεκπεραίωση της προκειμένης εργασίας. Συγκεκριμένα παρουσιάζονται τα μοντέλα που δοκιμάστηκαν για λήψη αυτόματων επισημάνσεων για την ηλικία, τη φυλή και το φύλο. Έπειτα αναλύονται οι βάσεις δεδομένων που χρησιμοποιήθηκαν και η μεθοδολογία του δίκαιου διαχωρισμού που εφαρμόστηκε σε κάθε βάση. Τέλος προβάλλονται τα μοντέλα για κάθε ετερογενή εργασία συναισθηματικού υπολογισμού με τα οποία εκπαιδεύτηκαν και αξιολογήθηκαν οι βάσεις δεδομένων πριν και μετά το δίκαιο διαχωρισμό. Στο Κεφάλαιο 5 παρουσιάζονται οι σχετικές εργασίες που συνδέονται με το αντικείμενο της παρούσας εργασίας. Στο Κεφάλαιο 6 παρουσιάζονται οι λεπτομέρειες της εκπαίδευσης των μοντέλων συναισθηματικού υπολογισμού. Στο Κεφάλαιο 7 προβάλλονται τα αποτελέσματα που παράχθηκαν κατά τη διεκπεραίωση της παρούσας εργασίας. Συγκεκριμένα προβάλλονται τα αποτελέσματα της αξιολόγησης των μοντέλων ανίχνευσης φυλής, ηλικίας και φύλου, τα αποτελέσματα των δοκιμών βελτιστοποίησης και τέλος τα κύρια αποτελέσματα της εργασίας, δηλαδή η απόδοση όλων των εργασιών αναγνώρισης συναισθήματος με τις αντίστοιχες βάσεις δεδομένων, πριν και μετά το δίκαιο διαχωρισμό. Τέλος, στο Κεφάλαιο 8 δίνεται ο επίλογος, τα συμπεράσματα καθώς και οι μελλοντικές επεκτάσεις.



Μέρος **I**

Θεωρητικό Μέρος

---



## Κεφάλαιο **2**

# Θεωρία Συναισθηματικού Υπολογισμού

---

**Σ**το παρόν κεφάλαιο περιγράφεται η θεωρία πίσω από την ανάλυση και αναγνώριση συναισθημάτων, από εικόνες με χρήση μηχανικής μάθησης. Η ανάλυση και αναγνώριση συναισθημάτων διαδραματίζει σημαντικό ρόλο σε θέματα μοντελοποίησης συμπεριφοράς, αλληλεπίδρασης μηχανής και ανθρώπου και συναισθηματικού υπολογισμού. Εφαρμόζεται σε ποικίλους τομείς όπως τον τομέα της ιατρικής, της υγείας, της ηλεκτρονικής μάθησης, της νομικής, του μάρκετινγκ, της ανίχνευσης ψεύδους κ.α. Συγκεκριμένα κάποια παραδείγματα εφαρμογών αποτελούν η ανίχνευση εκφράσεων του προσώπου και προσδιορισμός της κόπωσης ενός οδηγού αυτοκινήτου, η μελέτη των συναισθημάτων των μαθητών και προσαρμογή των τεχνικών μάθησης σύμφωνα με τις αντιδράσεις τους, η ανίχνευση της συναισθηματικής κατάστασης των υποψήφιων κατά τη διάρκεια μιας συνέντευξης, ο εντοπισμός θετικών ή αρνητικών αντιδράσεων του κοινού σε εκδηλώσεις κ.α. Ο συναισθηματικός υπολογισμός αποτελείται από τρεις ετερογενείς εργασίες: την αναγνώριση βασικών συναισθημάτων, την εκτίμηση διαστατικών αναπαραστάσεων συναισθημάτων ως προς τις συνεχείς τιμές διέγερσης (arousal) και σθένους (valence) και την ενεργοποίηση συγκεκριμένων μυών του προσώπου με χρήση των Action Units (AUs). Οι τρεις προαναφερθείσες εργασίες θα αναπτυχθούν στις επόμενες ενότητες. [4]

### 2.1 Εργασία 1: Αναγνώριση βασικών συναισθημάτων

Η αναγνώριση των βασικών κατηγορικών συναισθημάτων είναι ο πιο διαδεδομένος τρόπος της ανάλυσης της συμπεριφοράς του προσώπου. Τα έξι βασικά συναισθήματα, Εικόνα 2.1, (θυμός, αγδία, φόβος, ευτυχία, λύπη, έκπληξη), συν την ουδέτερη κατάσταση, καθορίστηκαν από τον Ekman. Το μοντέλο ανάλυσης έκφρασης προσώπου που βασίζεται σε βασικά συναισθήματα έχει περιορισμούς στην αναπαράσταση της πολυπλοκότητας και της ευαισθησία των καθημερινών συναισθηματικών καταστάσεων. [4]

### 2.2 Εργασία 2: Εκτίμηση τιμών valence και arousal

Η επόμενη εργασία αφορά το διαστατικό μοντέλο κατά το οποίο η έκφραση συναισθηματικής κατάστασης μπορεί να περιγραφεί από συνεχείς διαστάσεις. Με αυτό το μοντέλο μπορούν να αναπαρασταθούν και οι πιο διακριτικές συναισθηματικές καταστάσεις. Οι πιο συχνά χρησιμοποιούμενες διαστάσεις περιλαμβάνουν την τιμή του σθένους (valence), που δείχνει πόσο



Εικόνα 2.1: *Six basic emotions* [1]

θετική ή αρνητική είναι μια συναισθηματική κατάσταση, και την τιμή της διέγερσης (arousal), που δείχνει πόσο ενεργή ή παθητική είναι μια συναισθηματική κατάσταση. Οι δυο αυτές τιμές σχετίζονται άμεσα με συγκεκριμένες λειτουργίες περιοχών του εγκεφάλου. Η πλευρική περιοχή του δεξιού ημισφαιρίου φαίνεται να παίζει ιδιαίτερο ρόλο στη μεσολάβηση της διέγερσης, ενώ οι μετωπιαίες περιοχές φαίνεται να παίζουν ιδιαίτερο ρόλο στην τιμή του σθένους. [4]

### 2.3 Εργασία 3: Ενεργοποίηση τιμών Action Units

Τέλος, η τρίτη εργασία αναγνώρισης έκφρασης συναισθηματικής κατάστασης αφορά τον εντοπισμό Action Units. Το σύστημα Κωδικοποίησης Action Units (FACS - Facial Action Coding System) παρέχει μια ταξινόμηση των κινήσεων των μυών του προσώπου. Στην εικόνα 2.2 απεικονίζονται τα πιο κοινά AUs. Παρατηρείται ότι οποιαδήποτε έκφραση του προσώπου μπορεί να αναπαρασταθεί ως συνδυασμός Action Units. Έτσι αποτελούν μια βάση για την ανάλυση προσώπου η οποία επεκτείνεται σε ένα ευρύ φάσμα εφαρμογών, πέρα από την ταξινόμηση συναισθημάτων. Η έκφραση των AUs είναι συνήθως σύντομη και ασυνείδητη, και έτσι η ανίχνευση τους απαιτεί την ανάλυση μικρών αλλαγών στο πρόσωπο. Επιπλέον, τα AUs δεν εμφανίζονται μεμονωμένα, αλλά χαρακτηρίζουν μια έκφραση προσώπου, επομένως κάποια AUs συνυπάρχουν, ενώ άλλα αλληλοαναιρούνται. [4]

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

Εικόνα 2.2: Most common AUs [2]



## Κεφάλαιο **3**

# Θεωρία μοντέλων μηχανικής μάθησης

---

**Σ**το κεφάλαιο αυτό παρουσιάζονται θεωρητικά τα χαρακτηριστικά των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται για τις τρεις ετερογενείς εργασίες αναγνώρισης συναισθήματος από εικόνες (Κεφάλαιο 2). Αρχικά γίνεται μία σύντομη εισαγωγή στη μηχανική μάθηση, τη λειτουργία των νευρωνικών δικτύων και των χαρακτηριστικών τους. Έπειτα αναλύονται οι δύο διαφορετικοί τρόποι επιβλεπόμενης μάθησης (classification και regression task) οι οποίοι θα χρησιμοποιηθούν για την υλοποίηση των τριών εργασιών.

### 3.1 Μηχανική μάθηση

Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασισμένες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

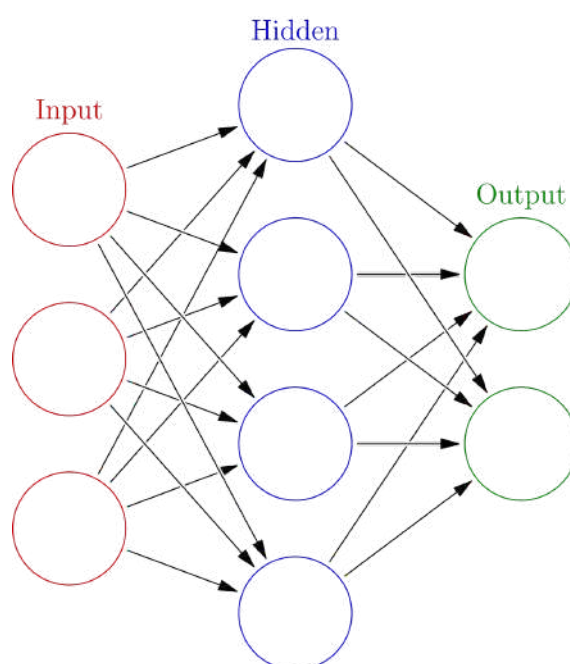
Οι αλγόριθμοι μηχανικής μάθησης ταξινομούνται ευρέως σε τρεις κατηγορίες: επιβλεπόμενη μάθηση (Supervised learning), μη εποβλεπόμενη μάθηση (Unsupervised learning) και ενισχυτική μάθηση (Reinforcement learning). Τα μοντέλα που θα χρησιμοποιηθούν στα πλαίσια της παρούσας εργασίας αφορούν την πρώτη κατηγορία, της επιβλεπόμενης μάθησης. Στην επιβλεπόμενη μάθηση, τα μοντέλα εκπαιδεύονται καθώς επεξεργάζονται δεδομένα εισόδου τα οποία, περιέχουν μια γνωστή είσοδο και ένα επιθυμητό αποτέλεσμα (label).

Οι βάσεις δεδομένων που χρησιμοποιούνται για εκπαίδευση, αξιολόγηση και έλεγχο των μοντέλων μηχανικής μάθησης χωρίζονται σε τρία σύνολα δεδομένων. Το πρώτο σύνολο είναι το σύνολο εκπαίδευσης (training set), το οποίο χρησιμοποιείται για την προσαρμογή του μοντέλου. Εκπαιδεύεται, δηλαδή, το μοντέλο πάνω στα δεδομένα του συνόλου αυτού. Έτσι το σύνολο δεδομένων εκπαίδευσης πρέπει να είναι επαρκώς αντιπροσωπευτικό του προβλήματος. Το δεύτερο σύνολο αποτελεί το σύνολο επικύρωσης (validation set), το οποίο χρησιμοποιείται για προσαρμογή των υπερπαραμέτρων (υποενότητα 3.2.2) του μοντέλου. Με το σύνολο αυτό αξιολογείται η απόδοση του μοντέλου κατά την εκπαίδευση. Το σύνολο ελέγχου (test set) αποτελεί το τρίτο σύνολο και χρησιμοποιείται για την παροχή μιας αμερόληπτης αξιολόγησης ενός τελικού μοντέλου που ταιριάζει στο σύνολο δεδομένων εκπαίδευσης. Το σύνολο ελέγχου πρέπει να είναι καλά επιμελημένο, δηλαδή να περιέχει προσεκτικά δειγματοληπτικά δεδομένα που εκτείνονται στις διάφορες κατηγορίες που θα αντιμετωπίσει το μοντέλο, όταν

χρησιμοποιείται στον πραγματικό κόσμο.

## 3.2 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα, γνωστά και ως τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANN) αποτελούν υποσύνολο της μηχανικής μάθησης. Το όνομα και η δομή τους είναι εμπνευσμένα από τον ανθρώπινο εγκέφαλο, μιμούμενα τον τρόπο που οι βιολογικοί νευρώνες δίνουν σήμα ο ένας στον άλλο. Τα νευρωνικά δίκτυα κατασκευάζονται από τρεις τύπους επιπέδων. Το επίπεδο εισόδου (input layer), τα κρυφά επίπεδα (hidden layers) και το επίπεδο εξόδου (output layer) (Σχήμα 3.1).

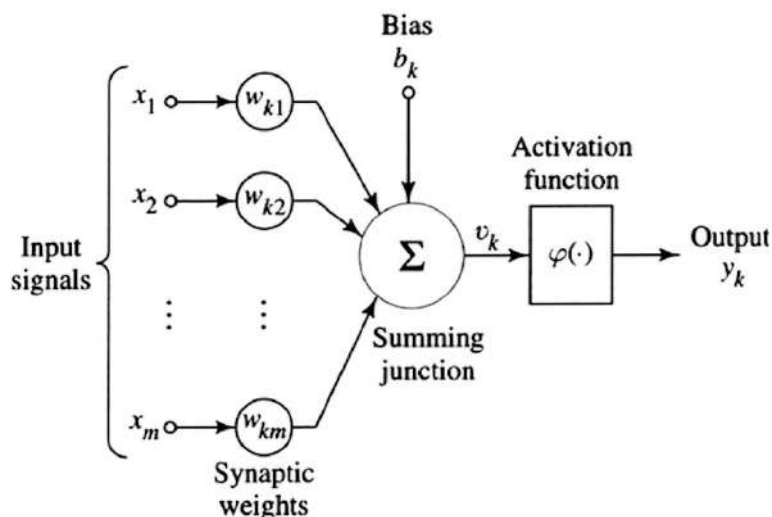


Εικόνα 3.1: Τα επίπεδα ενός νευρωνικού δικτύου

Κάθε επίπεδο αποτελείται από νευρώνες οι οποίοι συνδέονται με κάθε κόμβο από το επόμενο επίπεδο και κάθε σύνδεση έχει μια τιμή βάρους. Το βάρος μπορεί να θεωρηθεί ως επίδραση που έχει αυτός ο κόμβος στον κόμβο του επόμενου επιπέδου. Οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις εισόδους του δικτύου και στους κρυμμένους νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες (νευρώνες στο κρυφό επίπεδο) πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο βάρος, και υπολογίζουν το ολικό άθροισμα των γινομένων, όπως φαίνεται στο σχήμα 3.2. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης (activation function), την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που εξάγει η συνάρτηση είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

Η εκπαίδευση ενός νευρωνικού δικτύου, στην επιβλεπόμενη μάθηση, διεξάγεται συνήθως με τον προσδιορισμό της τιμής σφάλματος, δηλαδή της διαφοράς μεταξύ της τελικής εξόδου του δικτύου (prediction) και της εξόδου στόχου (label). Η συνάρτηση με την οποία θα ληφθεί





Εικόνα 3.2: Οι υπολογισμοί που εκτελεί ένας υπολογιστικός νευρώνας

αυτή η διαφορά ονομάζεται συνάρτηση σφάλματος (loss function) και διαφέρει ανάλογα με το είδος της εργασίας του νευρωνικού δικτύου. Στη συνέχεια, το δίκτυο προσαρμόζει τα βάρη του σύμφωνα με έναν αλγόριθμο βελτιστοποίησης (optimizer), χρησιμοποιώντας αυτήν την τιμή σφάλματος. Οι διαδοχικές προσαρμογές θα αναγκάσουν το νευρωνικό δίκτυο να παράγει έξοδο που είναι όλο και πιο παρόμοια με την έξοδο στόχο. Μετά από έναν επαρκή αριθμό από αυτές τις προσαρμογές, η εκπαίδευση μπορεί να τερματιστεί με βάση ορισμένα κριτήρια. Οι όροι activation function, loss function, optimizer που αναφέρθηκαν θα επεξηγηθούν στις επόμενες υποενότητες.

### 3.2.1 Μετρικές αξιολόγησης νευρωνικών δικτύων

Οι μετρικές αξιολόγησης χρησιμοποιούνται για τη μέτρηση της ποιότητας του μοντέλου μηχανικής μάθησης. Η αξιολόγηση των μοντέλων είναι απαραίτητη αφού διασφαλίζει τη σωστή και βέλτιστη λειτουργία τους. Παρακάτω επισημαίνονται οι βασικότερες μετρικές αξιολόγησης νευρωνικών δικτύων για την εργασία τις ταξινόμησης (classification task).

#### 1. Ακρίβεια ταξινόμησης (Classification Accuracy)

Η ακρίβεια ταξινόμησης είναι ο λόγος του αριθμού των σωστών προβλέψεων προς το συνολικό αριθμό των δειγμάτων εισόδου. Υπολογίζει δηλαδή το ποσοστό των σωστών προβλέψεων που εκτέλεσε το μοντέλο.

#### 2. Confusion matrix (Σχήμα 3.3)

Το Confusion Matrix αναπαρίσταται σε μορφή μήτρας και δίνει μια σύγκριση μεταξύ πραγματικών και προβλεπόμενων τιμών. Είναι ένας πίνακας  $N \times N$ , όπου  $N$  είναι ο αριθμός των κλάσεων ή των εξόδων του μοντέλου. Ο Confusion Matrix χαρακτηρίζεται από τις τιμές True Positive, False Negative, False Positive, True Negative. Η τιμή True Positive δείχνει ότι η πραγματική τιμή (θετική) και η προβλεπόμενη τιμή (θετική) είναι οι ίδιες, η τιμή False Negative δείχνει ότι η πραγματική τιμή (θετική) και η πρόβλεψη του μοντέλου (αρνητική) δεν συμφωνούν, η τιμή False Positive δείχνει ότι η πραγματική

τιμή (αρνητική) και η πρόβλεψη του μοντέλου (θετική) δεν συμφωνούν και τέλος η τιμή True Negative δείχνει ότι η πραγματική τιμή (αρνητική) και η πρόβλεψη του μοντέλου (αρνητική) συμφωνούν.

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

Εικόνα 3.3: *Confusion Matrix for binary classification problem*

### 3. Precision

Η μετρική precision δείχνει τον αριθμό των σωστών θετικών αποτελεσμάτων (True Positives) διαιρεμένος με τον αριθμό των θετικών αποτελεσμάτων που προβλέπονται από το μοντέλο (True Positives + False Positives). Υπολογίζει δηλαδή το ποσοστό των προβλέψεων που ήταν πραγματικά σωστό για κάθε κλάση.

$$Precision = \frac{TP}{TP + FP} \quad (3.1)$$

### 4. Recall

Η μετρική recall δείχνει τον αριθμό των σωστών θετικών αποτελεσμάτων (True Positives) διαιρεμένος με τον αριθμό όλων των δειγμάτων που έπρεπε να έχουν αναγνωρισθεί ως θετικά από το μοντέλο (True Positives + False Negatives). Υπολογίζει δηλαδή το ποσοστό των στοιχείων της κλάσης που ταξινομήθηκε σωστά.

$$Recall = \frac{TP}{TP + FN} \quad (3.2)$$

### 5. F1 score

Το f1 score είναι ο αρμονικός μέσος όρος μεταξύ precision και recall. Όσο η τιμή του f1 score φτάνει τη μονάδα, τόσο καλύτερο το μοντέλο.

$$f1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3.3)$$

## 6. Averaged Accuracy

Η μέση ακρίβεια (Averaged Accuracy) είναι ο μέσος όρος κάθε ακρίβειας (Accuracy) ανά κατηγορία. Η τιμή αυτή μπορεί να εκφραστεί και ως η μέση τιμή για τις τιμές Recall κάθε κατηγορίας.

Για την εργασία της παλινδρόμησης (regression task), χρησιμοποιούνται μετρικές απόδοσης όπως το Mean Absolute Error 1 και το Mean Squared Error 2.

### 3.2.2 Υπερπαραμέτροι νευρωνικών δικτύων

Ένα μοντέλο μηχανικής μάθησης αντιπροσωπεύεται από τις παραμέτρους του (τα βάρη του). Ωστόσο, η διαδικασία εκπαίδευσης ενός μοντέλου περιλαμβάνει την επιλογή των βέλτιστων υπερπαραμέτρων που θα χρησιμοποιηθούν για να βρεθούν οι βέλτιστες παράμετροι, που αντιστοιχίζουν σωστά τα χαρακτηριστικά εισόδου στις ετικέτες. Έτσι οι υπερπαραμέτροι ορίζονται ως οι παράμετροι που επιλέγονται ρητά από το χρήστη για τον έλεγχο της μαθησιακής διαδικασίας. Οι πιο κύριες υπερπαραμέτροι μοντέλων επιβλεπόμενης μηχανικής μάθησης αναλύονται παρακάτω.

#### 1. Αλγόριθμος βελτιστοποίησης (Optimizer)

Ο αλγόριθμος βελτιστοποίησης βρίσκει την τιμή των παραμέτρων (βαρών) που ελαχιστοποιούν το σφάλμα κατά την αντιστοίχιση εισόδων σε εξόδους. Συγκεκριμένα ο optimizer είναι μια συνάρτηση ή ένας αλγόριθμος ο οποίος κατά την εκπαίδευση του μοντέλου βαθιάς μάθησης, τροποποιεί τα βάρη κάθε εποχής (epoch) με στόχο την ελαχιστοποίηση της συνάρτησης απώλειας (loss function) και τη βελτίωση της ακρίβειας. Αυτοί οι αλγόριθμοι επηρεάζουν ευρέως την ακρίβεια και την ταχύτητα εκπαίδευσης του μοντέλου μηχανικής μάθησης. Οι αλγόριθμοι βελτιστοποίησης που χρησιμοποιήθηκαν για την υλοποίηση των μοντέλων υπολογιστικού συναισθηματισμού είναι οι SGD (Stochastic Gradient Descent), Adam (Adaptive Moment Estimation), RMS (Root Mean Square propagation).

#### 2. Ρυθμός εκμάθησης (Learning rate - lr)

Ο ρυθμός εκμάθησης πρόκειται για την υπερπαραμέτρο στους αλγόριθμους βελτιστοποίησης που ελέγχει το ρυθμό που πρέπει να ανανεωθούν οι παράμετροι του μοντέλου σε σχέση με την τιμή του εκτιμώμενου σφάλματος. Λαμβάνει δεκαδικές τιμές στο εύρος 0 μέχρι 1. Η επιλογή του βέλτιστου ρυθμού εκμάθησης είναι κρίσιμη. Εάν ο ρυθμός εκμάθησης είναι πολύ μικρός τότε επιβραδύνεται η διαδικασία εκπαίδευσης και μπορεί να μην επέλθει η σύγκλιση στη βέλτιστη λύση, ενώ εάν το ποσοστό εκμάθησης είναι πολύ μεγάλο το μοντέλο μπορεί να συγκλίνει γρήγορα σε μια μη βέλτιστη λύση.

#### 3. Batch size

Για να βελτιωθεί η ταχύτητα της διαδικασίας εκμάθησης, το σετ εκπαίδευσης χωρίζεται σε διαφορετικά υποσύνολα, τα οποία είναι γνωστά ως batches. Το μέγεθος των υποσυνόλων αυτών ορίζεται ως batch size.

#### 4. Εποχές (Epochs)

Μια εποχή μπορεί να οριστεί ως ο πλήρης κύκλος για την εκπαίδευση του μοντέλου. Η εποχή αντιπροσωπεύει μια επαναληπτική διαδικασία μάθησης. Για τον προσδιορισμό του σωστού αριθμού εποχών, λαμβάνεται υπόψη το σφάλμα επικύρωσης (validation error). Ο αριθμός των εποχών αυξάνεται μέχρι να μειωθεί το σφάλμα επικύρωσης. Εάν δεν υπάρχει βελτίωση για διαδοχικές εποχές, τότε υποδεικνύει τη διακοπή της αύξησης του αριθμού των εποχών.

#### 5. Αριθμός κρυφών νευρώνων (Hidden Units)

Όπως προαναφέρθηκε, οι κρυφοί νευρώνες αποτελούν τις υπολογιστικές μονάδες του νευρωνικού δικτύου. Ο αριθμός των κρυφών νευρώνων καθορίζει πόσοι νευρώνες θα αποτελούν κάθε κρυφό επίπεδο.

#### 6. Αριθμός Επιπέδων (Layers)

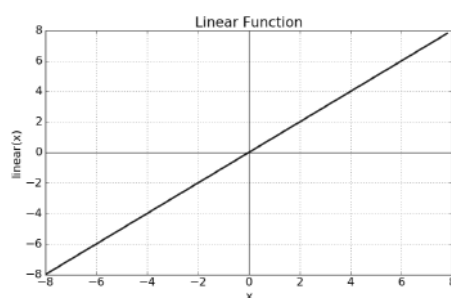
Ο αριθμός των επιπέδων ορίζει πόσο βαθύ θα είναι το νευρωνικό δίκτυο.

#### 7. Συνάρτηση Ενεργοποίησης (Activation Function)

Η συνάρτηση ενεργοποίησης (activation function) ενός κόμβου καθορίζει την έξοδο αυτού του κόμβου με δεδομένη μια είσοδο ή ένα σύνολο εισόδων. Κάποιες από τις βασικότερες activation functions παρουσιάζονται παρακάτω.

(α') Γραμμική συνάρτηση (linear function)

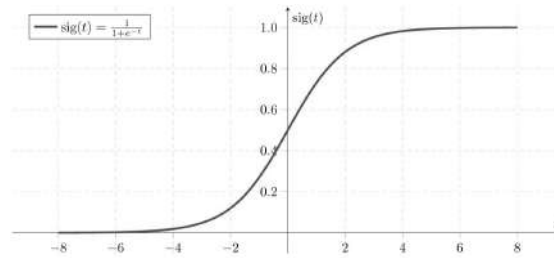
Στη Γραμμική συνάρτηση ενεργοποίησης η έξοδος είναι γραμμική συνάρτηση της εισόδου (σχήμα 3.4).



Εικόνα 3.4: Γραφική παράσταση γραμμικής συνάρτησης ενεργοποίησης

## (β') Σιγμοειδής συνάρτηση (Sigmoid function)

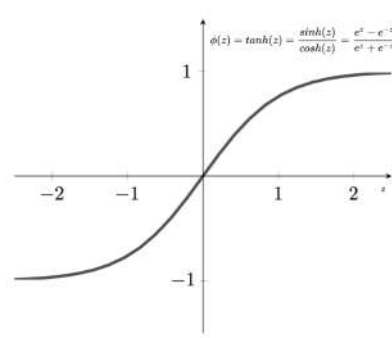
Η Σιγμοειδής συνάρτηση είναι μια μη γραμμική συνάρτηση ενεργοποίησης που μετασχηματίζει τις τιμές εισόδου στο εύρος 0 και 1 (σχήμα 3.5).



Εικόνα 3.5: Γραφική παράσταση σιγμοειδής συνάρτησης ενεργοποίησης

## (γ') Tanh function

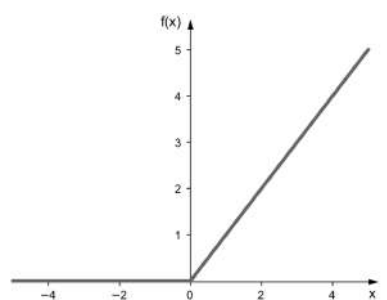
Η Tanh είναι μια μη γραμμική συνάρτηση ενεργοποίησης που μετασχηματίζει τις τιμές εισόδου στο εύρος -1 και 1 (σχήμα 3.6).



Εικόνα 3.6: Γραφική παράσταση συνάρτησης ενεργοποίησης tanh

## (δ') ReLU function

Η συνάρτηση ReLU - Rectified Linear Unit, μια μη γραμμική συνάρτηση κατά την οποία η έξοδος λαμβάνει την τιμή της εισόδου μόνο όταν η αυτή είναι μεγαλύτερη του μηδενός. Σε αντίθετη περίπτωση η έξοδος μηδενίζεται. Η συνάρτηση αυτή δεν ενεργοποιεί όλους τους νευρώνες ταυτόχρονα, αφού οι νευρώνες απενεργοποιούνται όταν η είσοδος είναι μικρότερη του μηδενός (σχήμα 3.7).



Εικόνα 3.7: Γραφική παράσταση συνάρτησης ενεργοποίησης ReLU

## (ε') Softmax function

Η συνάρτηση Softmax μετατρέπει ένα διάνυσμα πραγματικών αριθμών σε κατανομή πιθανότητας. Χρησιμοποιείται συχνά ως η τελευταία συνάρτηση ενεργοποίησης ενός νευρωνικού δικτύου ταξινόμησης (υποενότητα 3.3), για τη μετατροπή της εξόδου σε κατανομή πιθανότητας. Ο τύπος υπολογισμού φαίνεται παρακάτω (3.8)

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

$\sigma$  = softmax

$\vec{z}$  = input vector

$e^{z_i}$  = standard exponential function for input vector

$K$  = number of classes in the multi-class classifier

$e^{z_j}$  = standard exponential function for output vector

$e^{z_j}$  = standard exponential function for output vector

Εικόνα 3.8: *Softmax formula*

### 3.2.3 Συνελικτικά νευρωνικά δίκτυα

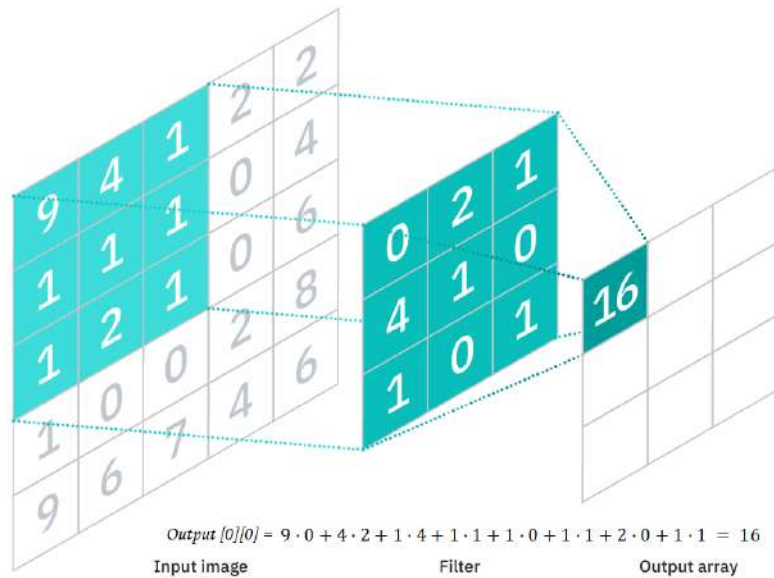
Στη βαθιά μάθηση, ένα συνελικτικό νευρωνικό δίκτυο (Convolutional Neural Networks - CNN) είναι μια εξειδικευμένη κατηγορία νευρωνικών δικτύων. Χρησιμοποιούν τη μαθηματική πράξη της συνέλιξης (convolution) στη θέση του πολλαπλασιασμού γενικού πίνακα σε τουλάχιστον ένα από τα στρώματά τους. Είναι ειδικά σχεδιασμένα για την επεξεργασία δεδομένων pixel και χρησιμοποιούνται στην αναγνώριση και επεξεργασία εικόνων.<sup>1</sup>

Ένα συνελικτικό νευρωνικό δίκτυο αποτελείται από ένα επίπεδο εισόδου, κρυφά επίπεδα και ένα επίπεδο εξόδου. Σε ένα συνελικτικό νευρωνικό δίκτυο, τα κρυφά επίπεδα (μεσαία επίπεδα) περιλαμβάνουν επίπεδα που εκτελούν συνέλιξεις (convolution layers). Συνήθως περιλαμβάνουν ένα επίπεδο που εκτελεί εσωτερικό γινόμενο μεταξύ του πυρήνα συνέλιξης (convolution kernel) με τον πίνακα εισόδου του επιπέδου. Καθώς ο πυρήνας συνέλιξης ολισθαίνει κατά μήκος του πίνακα εισόδου για το επίπεδο, η λειτουργία συνέλιξης δημιουργεί ένα χάρτη χαρακτηριστικών (feature map), ο οποίος συμβάλλει στην είσοδο του επόμενου επιπέδου. Αυτό ακολουθείται από άλλα επίπεδα, όπως επίπεδα pooling, πλήρως συνδεδεμένα επίπεδα. Τα επίπεδα αυτά θα αναλυθούν παρακάτω.

Τα συνελικτικά επίπεδα των CNN είναι τα βασικά δομικά στοιχεία τους, και σε αυτά υλοποιούνται οι περισσότεροι υπολογισμοί. Τα στοιχεία που απαιτούνται για την υλοποίηση των επιπέδων αυτών είναι τα δεδομένα εισόδου, ένα φίλτρο (ή πυρήνας - kernel) και ένας χάρτης χαρακτηριστικών (feature map). Ο πυρήνας είναι μία διδιάστατη διάταξη βαρών, η οποία αντιπροσωπεύει μέρος της εικόνας. Εφαρμόζεται σε μια περιοχή της εικόνας και υπολογίζεται το εσωτερικό γινόμενο μεταξύ των pixels εισόδου και του φίλτρου. Το εσωτερικό γινόμενο τροφοδοτείται στη συνέχεια σε έναν πίνακα εξόδου. Στη συνέχεια, το φίλτρο μετατοπίζεται κατά ένα βήμα, επαναλαμβάνοντας τη διαδικασία μέχρι ο πυρήνας να σαρώσει ολόκληρη την

<sup>1</sup>Στα πλαίσια της παρούσας εργασία χρησιμοποιούνται προ-εκπαιδευμένα συνελικτικά δίκτυα για την υλοποίηση των εργασιών αναγνώρισης συναισθηματικών εκφράσεων προσώπου.

εικόνα. Η τελική έξοδος από τη σειρά εσωτερικών γινομένων από την είσοδο και το φίλτρο είναι γνωστή ως χάρτης χαρακτηριστικών (feature map). Η διαδικασία αυτή απεικονίζεται στο σχήμα 3.9. Ένα άλλο συνελικτικό επίπεδο μπορεί να ακολουθήσει το αρχικό επίπεδο



Εικόνα 3.9: *The implementation of a Convolutional Layer*

συνέλιξης. Η δομή των CNN γίνεται ιεραρχική καθώς αυξάνονται τα συνελικτικά επίπεδα. Τα αρχικά επίπεδα λαμβάνουν χαρακτηριστικά χαμηλού επιπέδου και τα μεταγενέστερα επίπεδα χαρακτηριστικά υψηλού επιπέδου.

Τα επίπεδα pooling πραγματοποιούν μείωση διαστάσεων, αφού μειώνουν τον αριθμό των παραμέτρων της εισόδου. Όπως και το συνελικτικό επίπεδο, τα επίπεδα pooling σαρώνουν με ένα φίλτρο ολόκληρη την είσοδο, χωρίς όμως τη χρήση βαρών. Ο πυρήνας εφαρμόζει μια συνάρτηση συνάνθροισης στις τιμές εισόδου και συμπληρώνει έτσι τον πίνακα εξόδου. Οι δύο κύριοι τύποι είναι το max pooling και το average pooling. Στο πρώτο, καθώς το φίλτρο κινείται κατά μήκος της εισόδου, επιλέγει το pixel με τη μέγιστη τιμή και το στέλνει στη διάταξη εξόδου. Στο δεύτερο καθώς το φίλτρο μετακινείται κατά μήκος της εισόδου, υπολογίζει τη μέση τιμή των pixels και τη στέλνει στη διάταξη εξόδου. Τα επίπεδα αυτά παρόλο που χάνουν πολλές πληροφορίες, βοηθούν στη μείωση της πολυπλοκότητας, στη βελτίωση της αποτελεσματικότητας και περιορίζουν τον κίνδυνο του overfitting.

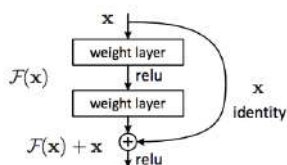
Στα πλήρως συνδεδεμένα επίπεδα (fully connected layers) κάθε κόμβος στο επίπεδο εξόδου συνδέεται απευθείας με έναν κόμβο του προηγούμενου επιπέδου. Αυτό το επίπεδο εκτελεί την ταξινόμηση με βάση τα χαρακτηριστικά που εξάγονται μέσω των προηγούμενων επιπέδων και των διαφορετικών φίλτρων τους. Ενώ τα convolutional και pooling επίπεδα τείνουν να χρησιμοποιούν συναρτήσεις ενεργοποίησης ReLu, τα πλήρως συνδεδεμένα επίπεδα χρησιμοποιούν συνήθως συνάρτηση ενεργοποίησης softmax για την κατάλληλη ταξινόμηση των εισόδων.

### Προ-εκπαιδευμένα συνελικτικά νευρωνικά δίκτυα

Σε αυτή την παράγραφο θα περιγραφούν τρεις αρχιτεκτονικές προ-εκπαιδευμένων συνελικτικών νευρωνικών δικτύων τα οποία θα χρησιμοποιηθούν στα πλαίσια της παρούσας εργασίας.

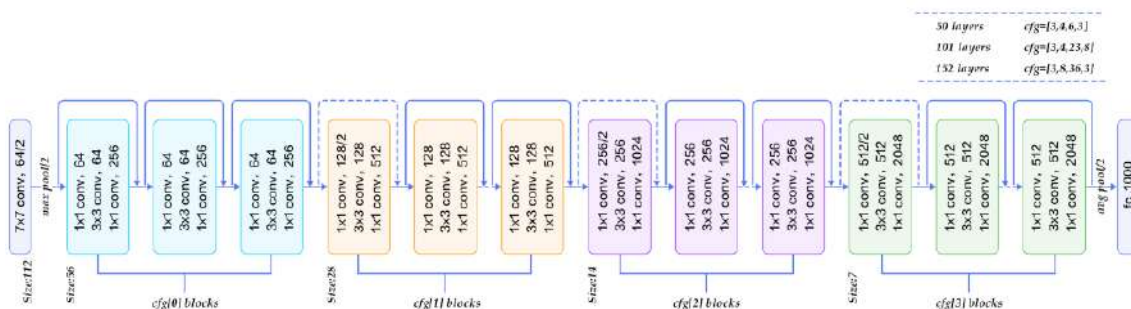
- ResNet [5]

Το ResNet, προτάθηκε το 2015 και εισήγαγε μια νέα αρχιτεκτονική που ονομάζεται Residual Network. Το Residual Network εισάγει μία έννοια που ονομάζεται Residual Blocks, με στόχο να λυθεί το πρόβλημα της εξαφάνισης ή έκρηξης παραγώγων. Το πρόβλημα αυτό συμβαίνει όταν αυξάνονται σημαντικά τα επίπεδα σε ένα βαθύ νευρωνικό δίκτυο και οι τιμές των παραγώγων των βαρών μηδενίζονται ή γίνονται πολύ μεγάλες. Έτσι για τη διαμόρφωση ενός Residual Block, εφαρμόζεται μια τεχνική η οποία ονομάζεται παράκαμψη συνδέσεων (skip connections). Κατά την τεχνική αυτή ενώνονται ενεργοποιήσεις από επόμενα επίπεδα, παρακάμπτοντας ορισμένα επίπεδα ενδιάμεσα, όπως φαίνεται στο σχήμα 3.10. Τα δίκτυα ResNet δημιουργούνται από πολλά Residual Blocks. Στα πλαίσια της παρούσας εργασίας χρησιμοποιείται το ResNet50.



Εικόνα 3.10: Resnet Block: skip connection

Το ResNet-50 είναι ένα συνελικτικό νευρωνικό δίκτυο με βάθος 50 στρωμάτων. Το προεκπαιδευμένο δίκτυο μπορεί να ταξινομήσει τις εικόνες σε 1000 κατηγορίες αντικειμένων. Ως αποτέλεσμα το δίκτυο έχει μάθει πλούσιες αναπαραστάσεις χαρακτηριστικών για ένα ευρύ φάσμα εικόνων. Το δίκτυο έχει μέγεθος εισαγωγής εικόνας 224 επί 224. Αναλυτικά η αρχιτεκτονική του δικτύου αυτού φαίνεται στο σχήμα 3.11.



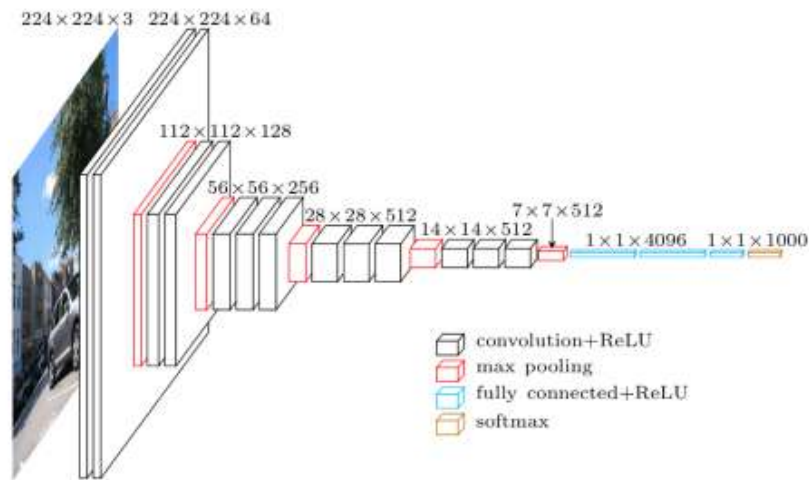
Εικόνα 3.11: Architecture of Resnet-50 Network

- VGG [6]

Το VGG ( Visual Geometry Group ) είναι ένα βαθύ συνελικτικό νευρωνικό δίκτυο με πολλά επίπεδα. Η αρχιτεκτονική VGG είναι η βάση των πρωτοποριακών μοντέλων αναγνώρισης αντικειμένων. Στα πλαίσια της εργασίας χρησιμοποιήθηκε το VGG16, το οποίο υποστηρίζει 16 επίπεδα και μπορεί να ταξινομήσει τις εικόνες σε 1000 κατηγορίες

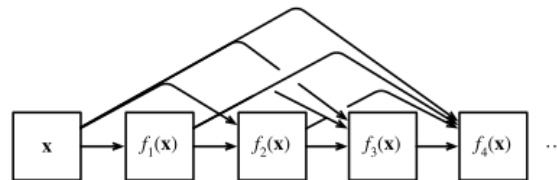


αντικειμένων. Επιπλέον, το μοντέλο έχει μέγεθος εισαγωγής εικόνας 224 επί 224. Αναλυτικά η αρχιτεκτονική του δικτύου αυτού φαίνεται στο σχήμα 3.12.



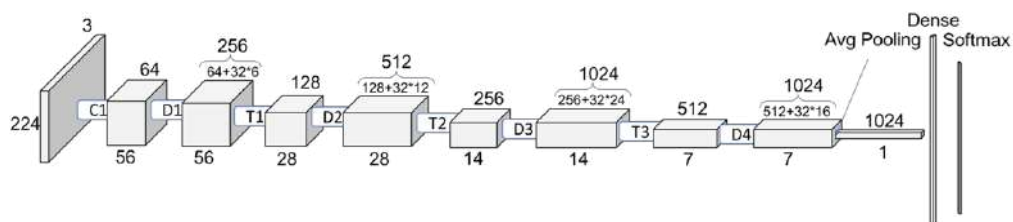
Εικόνα 3.12: *Architecture of VGG-16 Network*

- DenseNet [7] Το DenseNet είναι ένας τύπος συνελικτικού νευρωνικού δικτύου που χρησιμοποιεί πυκνές συνδέσεις (dense connections, σχήμα 3.13), μεταξύ των επιπέδων, μέσω των Dense Blocks. Στα Dense Blocks, συνδέονται όλα τα επίπεδα απευθείας μεταξύ τους. Για κάθε επίπεδο, οι χάρτες χαρακτηριστικών όλων των προηγούμενων επιπέδων αντιμετωπίζονται ως ξεχωριστές εισοδοι, ενώ οι δικόι του χάρτες χαρακτηριστικών μεταβιβάζονται ως εισοδοι σε όλα τα επόμενα επίπεδα. Στα πλαίσια της εργασίας



Εικόνα 3.13: *Dense Connection*

χρησιμοποιήθηκε το DenseNet121, το οποίο έχει τέσσερα Dense Blocks. Επιπλέον, το μοντέλο έχει μέγεθος εισαγωγής εικόνας 224 επί 224. Αναλυτικά η αρχιτεκτονική του δικτύου αυτού φαίνεται στο σχήμα 3.14.



Εικόνα 3.14: *Architecture of DenseNet-121 Network*

### 3.3 Classification task

Η ταξινόμηση (classification task) στη μηχανική μάθηση, πρόκειται για μια εργασία επιβλεπόμενης μηχανικής μάθησης κατά την οποία αναζητείται μια συνάρτηση που κατηγοριοποιεί κάθε στοιχείο του συνόλου εισόδου σε μία κλάση (διαφορετική κατηγορία). Συγκεκριμένα, το μοντέλο μηχανικής μάθησης θα χρησιμοποιήσει το σύνολο δεδομένων εκπαίδευσης και θα υπολογίσει τον καλύτερο τρόπο αντιστοίχισης των δεδομένων εισόδου με τις ετικέτες κλάσεων (διακριτές έξοδοι).

Η απλούστερη κατηγορία ταξινόμησης είναι η δυαδική ταξινόμηση (binary classification) και αναφέρεται στις εργασίες ταξινόμησης στις οποίες κάθε δεδομένο εισόδου λαμβάνει μία ετικέτα η οποία μπορεί να είναι μια από δύο πιθανές κλάσεις. Η δυαδική ταξινόμηση μπορεί να χρησιμοποιηθεί για ποικίλες εφαρμογές, όπως ανίχνευση ανεπιθύμητων μηνυμάτων, ανίχνευση απάτης και ιατρική διάγνωση [8, 9, 10, 11, 12, 13, 14].

Οι εργασίες ταξινόμησης στις οποίες κάθε δεδομένο εισόδου λαμβάνει μία ετικέτα από περισσότερες από δύο πιθανές κλάσεις ανήκουν στην κατηγορία ταξινόμησης πολλαπλών κατηγοριών (multiclass classification). Μια εφαρμογή της εργασίας αυτής θα μπορούσε να είναι η ταξινόμηση εικόνων ζώων σε διαφορετικές κατηγορίες, όπως σκύλοι, γάτες και πουλιά. Επίσης η πρώτη εργασία συναισθηματικού υπολογισμού (ενότητα 2.1), ταξινόμηση εικόνες βασικών συναισθημάτων εμπίπτει στην κατηγορία αυτή [15, 16, 17, 18].

Μια άλλη κατηγορία ταξινόμησης είναι η ταξινόμηση πολλαπλών ετικετών (multi-label classification), η οποία αναφέρεται στις εργασίες ταξινόμησης στις οποίες κάθε δεδομένο εισόδου λαμβάνει περισσότερες από μία ετικέτες από δύο πιθανές κλάσεις. Για παράδειγμα, η πρόβλεψη μιας δεδομένης κατηγορίας ταινίας που μπορεί να ανήκει σε τρόμο, ρομαντισμό, περιπέτεια, δράση ή όλα ταυτόχρονα ανήκει στην κατηγορία αυτή. Επιπρόσθετα στην κατηγορία αυτή ανήκει η τρίτη εργασία συναισθηματικού υπολογισμού (ενότητα 2.3), δηλαδή η ανίχνευση πολλαπλών τιμών AUs.

#### Συναρτήσεις σφάλματος (loss functions)

Για προβλήματα που εμπίπτουν στο είδος της ταξινόμησης, χρησιμοποιούνται συγκεκριμένες συναρτήσεις σφάλματος.

#### 1. Απώλεια δυαδικής διασταυρούμενης εντροπίας (Binary Cross Entropy Loss) <sup>2</sup>

Αυτή είναι η πιο κοινή συνάρτηση σφάλματος που χρησιμοποιείται για δυαδικά προβλήματα ταξινόμησης. Η εντροπία είναι το μέτρο της τυχαιότητας και η διασταυρούμενη εντροπία είναι ένα μέτρο της διαφοράς της τυχαιότητας μεταξύ δύο τυχαίων μεταβλητών. Εάν η απόκλιση της προβλεπόμενης πιθανότητας από την πραγματική ετικέτα αυξηθεί, η απώλεια διασταυρούμενης εντροπίας αυξάνεται (τύπος 3.15).

<sup>2</sup>Αυτή η συνάρτηση σφάλματος χρησιμοποιήθηκε για την τρίτη εργασία του εκφραστικού υπολογισμού, ανίχνευση AUs

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

Εικόνα 3.15: *Binary cross entropy loss function.  $\hat{y}$  : prediction,  $y$  : label*

## 2. Κατηγορική Απώλεια διασταυρούμενης εντροπίας (Categorical Cross Entropy Loss) <sup>3</sup>

Κατηγορική απώλεια διασταυρούμενης εντροπίας είναι ουσιαστικά η απώλεια δυαδικής διασταυρούμενης εντροπίας επεκτεινόμενη σε πολλαπλές κατηγορίες (τύπος 3.16).

$$\text{Loss} = - \sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

Εικόνα 3.16: *Categorical cross entropy loss function.  $\hat{y}$  : prediction,  $y$  : label*

## 3.4 Regression task

Η παλινδρόμηση (regression task) στη μηχανική μάθηση, πρόκειται για μια εργασία επιβλεπόμενης μηχανικής μάθησης κατά την οποία αναζητείται μια συνάρτηση που αντιστοιχεί κάθε στοιχείο του σύνολο εισόδου με μία συνεχή τιμή εξόδου. Σε αυτή την περίπτωση η έξοδος είναι μια πραγματική τιμή η οποία μπορεί να είναι ακέραια ή δεκαδική. Κάποια παραδείγματα (regression task) είναι η πρόβλεψη της ηλικίας κάποιου ατόμου, η πρόβλεψη εάν η τιμή της μετοχής μιας εταιρείας θα αυξηθεί αύριο. Παραδείγματα (regression task) αποτελούν και η πρόβλεψη συνεχών τιμών valence και arousal) που αποτελούν τη δεύτερη εργασία συναισθηματικού υπολογισμού (ενότητα 2.2).<sup>4</sup>

### Συναρτήσεις σφάλματος (loss functions)

Για την παλινδρόμηση χρησιμοποιούνται συγκεκριμένες συναρτήσεις σφάλματος (loss functions). Παρακάτω θα περιγραφούν οι κύριες συναρτήσεις σφάλματος της κατηγορίας αυτής.

#### 1. Μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE)

Το μέσο απόλυτο σφάλμα είναι μια από τις πιο απλές αλλά ισχυρές συναρτήσεις απώλειας που χρησιμοποιούνται για μοντέλα παλινδρόμησης. Όπως υποδηλώνει το όνομα, το MAE παίρνει το μέσο άθροισμα των απόλυτων διαφορών μεταξύ των πραγματικών και των προβλεπόμενων τιμών, όπως φαίνεται παρακάτω (σχήμα 3.17).

<sup>3</sup>Αυτή η συνάρτηση σφάλματος χρησιμοποιήθηκε για την πρώτη εργασία συναισθηματικού υπολογισμού, αναγνώριση βασικών συναισθημάτων

<sup>4</sup>Οι τιμές αυτές έχουν εκπαιδευτεί σε κοινό δίκτυο, βρίσκονται κάτω από την κατηγορία του Multitask regression

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

Εικόνα 3.17: *Mean absolute error loss function.  $y$ : prediction,  $x$ : label*

## 2. Μέσο τετραγωνικό σφάλμα (Mean Squared Error - MAE)

Το μέσο τετραγωνικό σφάλμα είναι ο μέσος όρος των διαφορών στο τετράγωνο μεταξύ των πραγματικών και των προβλεπόμενων τιμών, όπως φαίνεται παρακάτω (σχήμα 3.18).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Εικόνα 3.18: *Mean squared error loss function.  $\hat{Y}$  : prediction,  $Y$  : label*

## 3. Συντελεστής Συσχέτισης Συμφωνίας (Concordance Correlation Coefficient - CCC)

Ο Concordance Correlation Coefficient αναπτύχθηκε ως μέτρο για τη συσχέτιση μεταξύ δύο συνόλων δεδομένων, και ορίζεται όπως φαίνεται παρακάτω (σχήμα 3.19) Για συνάρτηση σφάλματος χρησιμοποιείται η αντίστροφη τιμή,  $1 - \text{CCC}$ , αφού για είναι επιθυμητή η μεγιστοποίηση της συσχέτισης μεταξύ πραγματικών και προβλεπόμενων τιμών.<sup>5</sup> Παρατηρείται ότι ο CCC εξαρτάται από τις τιμές  $\mu_x, \mu_y$  (μέσες τιμές των δύο μεταβλητών),  $\sigma_x^2, \sigma_y^2$  (διακυμάνσεις -variance των δύο μεταβλητών),  $\rho$  (συσχέτιση -correlation των δύο μεταβλητών).

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}$$

Εικόνα 3.19: *Concordance Correlation Coefficient*

<sup>5</sup> Αυτή η συνάρτηση σφάλματος χρησιμοποιήθηκε για τη δεύτερη εργασία συναισθηματικού υπολογισμού, εκτίμηση τιμών valence & arousal

Μέρος **II**

Πρακτικό Μέρος

---



## Κεφάλαιο 4

# Μεθοδολογία

---

Στο κεφάλαιο αυτό παρουσιάζεται αναλυτικά η μεθοδολογία που ακολουθήθηκε στην παρούσα εργασία. Το πρώτο στάδιο είναι η αξιολόγηση τεσσάρων μοντέλων που ανιχνεύουν τη φυλή, την ηλικία και το φύλο. Συγκεκριμένα, δοκιμάστηκαν τα μοντέλα `face_analysis_pytorch`, `deepface`, `FairFace` και `FaceRec`. Το τελευταίο εξάγει προβλέψεις μόνο για την ηλικία και το φύλο. Από τα τέσσερα αυτά μοντέλα επιλέχθηκε το πιο αποτελεσματικό το οποίο εφαρμόστηκε στις βάσεις δεδομένων για λήψη αυτόματων επισημάνσεων για κάθε κατηγορία. Στην εργασία αυτή χρησιμοποιήθηκαν οι ακόλουθες βάσεις δεδομένων `AffectNet`, `RAF-DB`, `AFEW-VA`, `EmotioNet`, `DISFA`, `GFT` και `RAF-AU`. Το επόμενο στάδιο είναι η αναδιάταξη των υπάρχουσών βάσεων δεδομένων με δίκαιο τρόπο στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Έπειτα δημιουργήθηκε ένα μοντέλο για κάθε ετερογενή εργασία αναγνώρισης έκφρασης, χρησιμοποιώντας προ-εκπαιδευμένα βαθιά νευρωνικά δίκτυα με το οποίο εκπαιδεύτηκαν και αξιολογήθηκαν οι βάσεις δεδομένων πριν και μετά το δίκαιο χωρισμό. Στο σύνδεσμο που ακολουθεί επισυνάπτεται ο κώδικας που χρησιμοποιήθηκε στο πρακτικό μέρος. <https://github.com/elenipapadopoulou4/thesis-code.git>

### 4.1 Μοντέλα αναγνώρισης φυλής, ηλικίας και φύλου

Στην ενότητα αυτή παρουσιάζεται η μελέτη που έγινε για επιλογή του κατάλληλου μοντέλου αναγνώρισης φυλής, ηλικίας και φύλου. Τα αποτελέσματα της μελέτης αυτής παρουσιάζονται στο τελευταίο κεφάλαιο 7.1.

Τα μοντέλα που αξιολογήθηκαν για την ανίχνευση φυλής, ηλικίας και φύλου φαίνονται παρακάτω. Το τέταρτο μοντέλο εξάγει προβλέψεις μόνο για δύο κατηγορίες (ηλικία και φύλο).

- Μοντέλο Αναγνώρισης 1: `Face Analysis PyTorch` [19]
- Μοντέλο Αναγνώρισης 2: `deepface` [20][21]
- Μοντέλο Αναγνώρισης 3: `FairFace` [22]
- Μοντέλο Αναγνώρισης 4: `FaceRec` [23]

Το πρώτο μοντέλο, `Face Analysis PyTorch`, παρέχει τρία διαφορετικά μοντέλα τα οποία αναλαμβάνουν την ανίχνευση φυλής, ηλικίας και φύλου. Τα μοντέλα πρόβλεψης φυλής και φύλου χρησιμοποιούν ως συνάρτηση σφάλματος τη σταθμισμένη διασταυρούμενη εντροπία

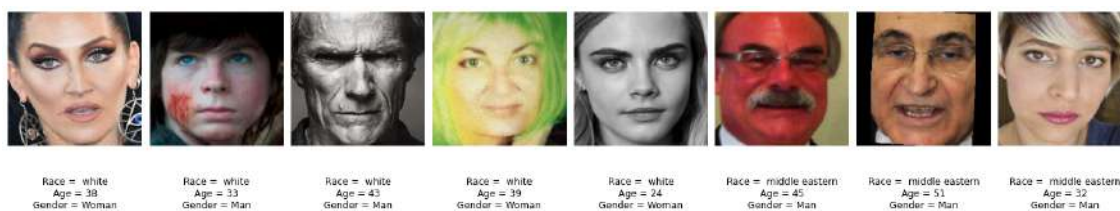
(weighted cross-entropy loss). Το μοντέλο ηλικίας χρησιμοποιεί τη συνάρτηση σφάλματος τακτικής παλινδρόμησης (ordinal regression) με συνάρτηση ενεργοποίησης τη σιγμοειδή αντί του softmax.[24]. Για την ηλικία η έξοδος είναι ακέραιος αριθμός (από 0 μέχρι 100), για τη φυλή μία από τις πέντε κατηγορίες (Caucasian, African-American, Asian, Indian, Άλλο) και τέλος για το φύλο μια από τις δύο κατηγορίες Αρσενικό ή Θηλυκό. Στο μοντέλο αυτό δεν έχει εφαρμοστεί κάποιος ανιχνευτής προσώπου. 4.1

Εικόνα 4.1: Εξόδοι του μοντέλου *Face Analysis Pytorch* με είσοδο τη βάση δεδομένων *AffectNet*



Το δεύτερο μοντέλο, *deepface*, έχει εξόδους για την ηλικία έναν ακέραιο αριθμό, για την εθνικότητα μία από τις έξι κατηγορίες (Asian, White, Middle Eastern, Indian, Latino και Black) και τέλος για το φύλο μια από τις δύο κατηγορίες γυναίκα ή άνδρας. 4.2

Εικόνα 4.2: Εξόδοι του μοντέλου *Deepface* με είσοδο τη βάση δεδομένων *AffectNet*



Το τρίτο μοντέλο, *FairFace*, έχει δύο επιλογές για πρόβλεψη φυλής. Η πρώτη έχει τέσσερις φυλές (White, Black, Asian, Indian) και η δεύτερη επτά φυλές (White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, Latino). Η ηλικία λαμβάνεται ως εννιά διαστήματα (0-2, 3-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+) και το φύλο Αρσενικό ή Θηλυκό. Το μοντέλο αυτό δημιουργήθηκε για να μετριάσει το πρόβλημα της φυλετικής μεροληψίας στα σύνολα δεδομένων. Οι δημιουργοί ανέπτυξαν μια εξισορροπημένη βάση δεδομένων η οποία περιέχει 108.501 εικόνες. Για να εκπαιδευτεί η βάση δημιουργήθηκε το μοντέλο για το οποίο χρησιμοποιήθηκε πανομοιότυπη αρχιτεκτονική με το ResNet-34. Χρησιμοποιήθηκε ο ADAM optimizer με ρυθμό εκμάθησης 0,0001. 4.3

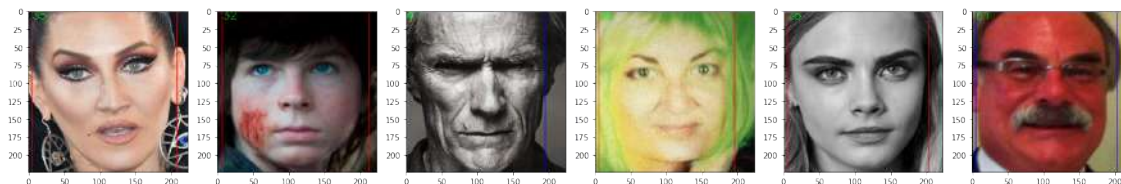
Εικόνα 4.3: Εξόδοι του μοντέλου *Fairface* με είσοδο τη βάση δεδομένων *AffectNet*





Το τέταρτο μοντέλου που δοκιμάστηκε, FaceRec, εξάγει προβλέψεις για ηλικία και φύλο. Για την ηλικία η έξοδος είναι αριθμός και για το φύλο μια από τις δύο κατηγορίες Άρσενικό ή Θηλυκό. Για την εκπαίδευση των μοντέλων αυτών χρησιμοποιήθηκε η συνάρτηση απώλειας Softmax.4.4

Εικόνα 4.4: Εξόδοι του μοντέλου FaceRec με είσοδο τη βάση δεδομένων AffectNet



## 4.2 Βάσεις δεδομένων

Στην ενότητα αυτή παρουσιάζονται οι βάσεις δεδομένων οι οποίες χρησιμοποιήθηκαν στη μελέτη. Επίσης παρουσιάζονται τα στατιστικά στα τρία σύνολα εκπαίδευσης όσον αφορά τις διάφορες ετικέτες, την ηλικία, τη φυλή και το φύλο για τις βάσεις δεδομένων πριν και μετά το δίκαιο χωρισμό

### 4.2.1 Παρουσίαση βάσεων

Έγινε χρήση των παρακάτω βάσεων δεδομένων επειδή παρέχουν μεγάλο αριθμό δειγμάτων με έγκυρες επισημάνσεις για valence-arousal, βασικά συναισθήματα και AUs. Αυτά τα σύνολα δεδομένων περιλαμβάνουν επίσης μια ποικιλία δειγμάτων και στα δύο φύλα, εθνικότητα, και ηλικίες.

Οι βάσεις που θα παρουσιαστούν είναι οι ακόλουθες

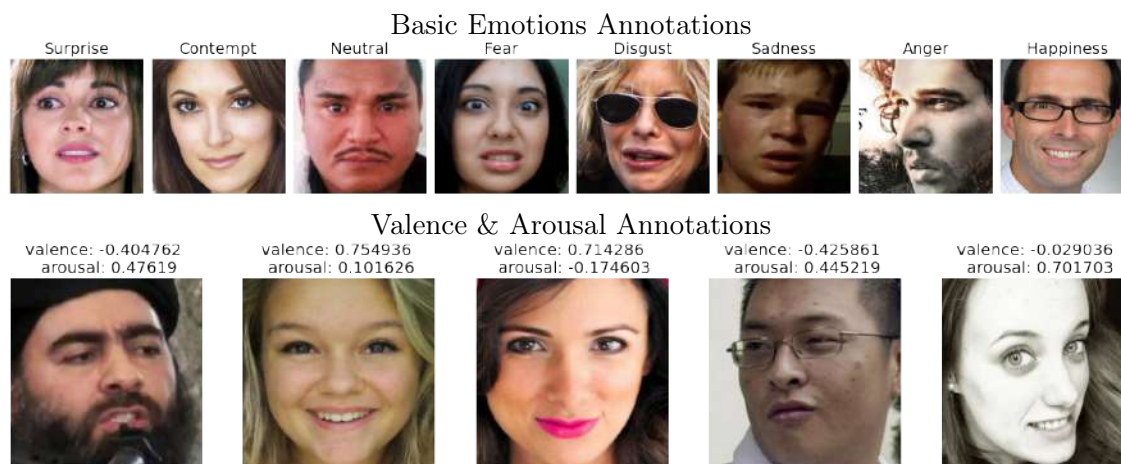
1. AffectNet [25]
2. RAF-DB [26]
3. RAF-AU [27][28]
4. EmotioNet [29]
5. GFT [30]
6. AFEW-VA [31][32]
7. DISFA [33]

#### 4.2.1.1 AffectNet

Η πρώτη βάση δεδομένων που χρησιμοποιήθηκε είναι AffectNet. Αυτή η βάση περιέχει περίπου 1 εκατομμύριο εικόνες προσώπων, 400.000 από τις οποίες σχολιάστηκαν με μη αυτόματο τρόπο (manually) ως προς 7 βασικά συναισθήματα (συν contempt) και ως προς τα valence-arousal που λαμβάνουν τιμές από [-1, 1]. Το σύνολο εκπαίδευσης (training set) αυτής της

βάσης δεδομένων αποτελείται από περίπου 321K εικόνες. Το σύνολο επικύρωσης (validation set) αποτελείται από 5K εικόνες και είναι ισορροπημένο ως προς τα 7 βασικά συναισθήματα (συν την περιφρόνηση - contempt) σε αντίθεση με το σύνολο εκπαίδευσης. Παραδείγματα εικόνων από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.5).

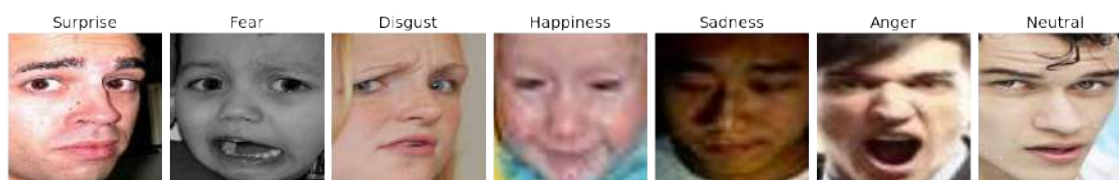
Εικόνα 4.5: *AffectNet, aligned images with annotations*



#### 4.2.1.2 RAF-DB

Η δεύτερη βάση δεδομένων, RAF-DB, περιέχει 12,2K εικόνες εκπαίδευσης (training images) και 3K εικόνες ελέγχου (testing images). Περιλαμβάνει επισημάνσεις για τα 7 βασικά συναισθήματα και για 11 σύνθετα συναισθήματα. Αυτή η βάση δεδομένων παρέχει σχολιασμούς φυλής, ηλικίας και φύλου για κάθε εικόνα. Ειδικότερα, οι σχολιασμοί αποτελούνται από τρεις κατηγορίες φυλών (Καυκάσιοι, Αφροαμερικάνοι, Ασιάτες), πέντε ηλικιακά διαστήματα (0-3, 4-19, 20-39, 40-69, 70+) και τρεις κατηγορίες φύλου (Άντρας, Γυναίκα, Αβέβαιο). Παραδείγματα εικόνων από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.6).

Εικόνα 4.6: *RAF-DB, aligned images with basic emotions annotations*



#### 4.2.1.3 RAF-AU

Η τρίτη βάση δεδομένων, RAF-AU, παρέχει κωδικοποίηση AUs σε εικόνες προσώπων με μεγάλη ποικιλία που έχουν ληφθεί από το Διαδίκτυο. Κατά τη διάρκεια του σχολιασμού, δύο έμπειροι κωδικοποιητές κωδικοποίησαν ανεξάρτητα τις εικόνες προσώπων και επίλυσαν οποιαδήποτε διαφωνία. Αυτή η βάση δεδομένων περιλαμβάνει 4601 εικόνες με επισημάνσεις για 26 διαφορετικά AUs. Τα πειράματα της παρούσας εργασίας έγιναν με τα 13 είδη AUs που

εμφανίζονται με ρυθμό μεγαλύτερο από 8%. Η βάση δεδομένων RAF-AU δεν είχε καθορισμένο διαχωρισμό σε σύνολο εκπαίδευσης, επικύρωσης και ελέγχου. Παραδείγματα εικόνων από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.7).

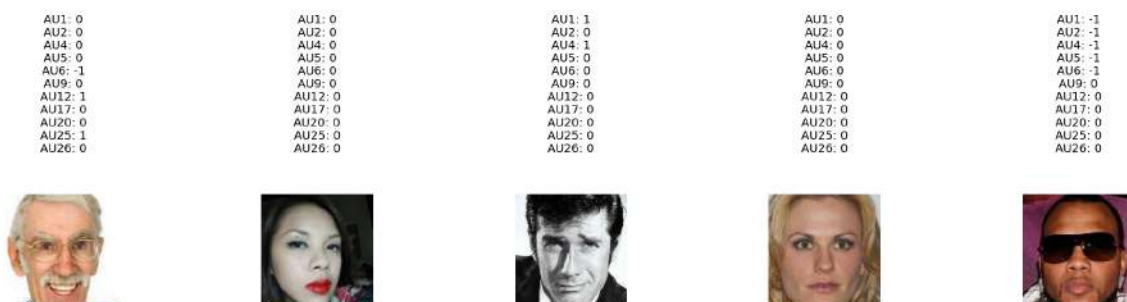
Εικόνα 4.7: *RAF-AU, aligned images with AUs annotations*



#### 4.2.1.4 EmotioNet

Η τέταρτη βάση δεδομένων, EmotioNet, κυκλοφόρησε για το EmotioNet Challenge το 2017. Εμπεριέχει περίπου 1 εκατομμύριο εικόνες από τις οποίες οι 950 χιλιάδες σχολιάστηκαν αυτόματα (automatically) και οι υπόλοιπες 50 χιλιάδες σχολιάστηκαν χειροκίνητα (manually) με 11 AUs (1,2,4,5,6,9,12,17,20,25,26). Από τις εικόνες που σχολιάστηκαν χειροκίνητα οι μισές αποτέλεσαν το σύνολο επικύρωσης (validation set) και οι άλλες μισές το σύνολο ελέγχου (test set). Σε αυτή την εργασία χρησιμοποιήθηκε το σύνολο επικύρωσης για εκπαίδευση δεδομένων και το σύνολο ελέγχου για αξιολόγηση πριν το δίκαιο διαχωρισμό. Παραδείγματα εικόνων από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.8).

Εικόνα 4.8: *EmotioNet, aligned images with AUs annotations*



#### 4.2.1.5 GFT

Η πέμπτη βάση δεδομένων, GFT αποτελείται από 96 βίντεο με 96 πρόσωπα, συνολικού μεγέθους περίπου 130.000 καρέ. Παρέχει επισημάνσεις για 14 AUs (1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 15, 17, 23, 24). Το σετ εκπαίδευσης αποτελείται από 78 πρόσωπα με μέγεθος περίπου 108K καρέ και το σετ δοκιμής με 18 πρόσωπα με μέγεθος περίπου 24,5K καρέ. Αυτή η βάση δεδομένων περιέχει επίσης έγκυρους σχολιασμούς για τη φυλή (Λευκός, Μαύρος, Ισπανόφωνος, Ασιάτης, Άλλος), την ηλικία (ακέραιες τιμές) και το φύλο (άνδρας, γυναίκα). Παραδείγματα εικόνων από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.9).

Εικόνα 4.9: *GFT, aligned images with AUs annotations*



#### 4.2.1.6 AFEW-VA

Η βάση δεδομένων AFEW-VA αποτελείται από 600 βίντεο κλιπ που εξήχθησαν από ταινίες μεγάλου μήκους και έχουν σχολιαστεί ανά καρτέ με τιμές σθένους (valence) και διέγερσης (arousal) που κυμαίνονται από [-1,1]. Αυτή η βάση δεδομένων δεν χωρίζεται σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Παραδείγματα καρτέ βίντεο από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.10).

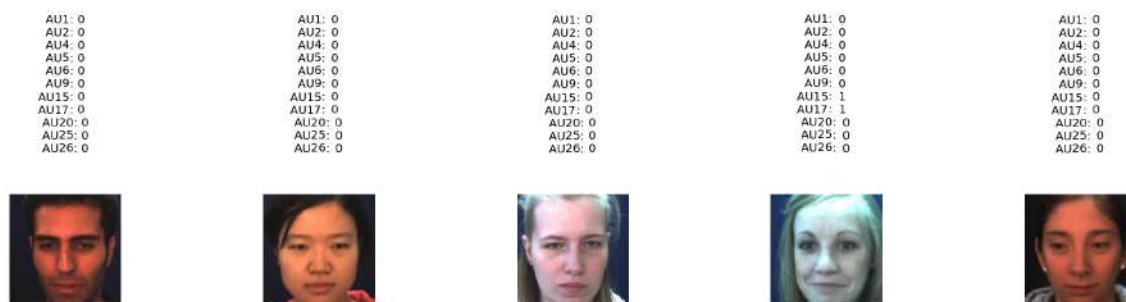
Εικόνα 4.10: *AFEW-VA, aligned video frames with valence & arousal annotations*



#### 4.2.1.7 DISFA

Τέλος, η έβδομη βάση δεδομένων, DISFA, είναι μια εργαστηριακά ελεγχόμενη βάση δεδομένων που αποτελείται από 27 βίντεο (22 στο σετ εκπαίδευσης και 5 στο σετ ελέγχου) το καθένα από τα οποία έχει 4.845 καρτέ, όπου κάθε καρτέ κωδικοποιείται με την ένταση AU σε μια διακριτή κλίμακα έξι σημείων. Οι εντάσεις AU ίσες ή μεγαλύτερες από 2 λαμβάνονται ως εμφανίσεις, ενώ μικρότερες αντιμετωπίζονται ως μη εμφανίσεις. Υπάρχουν συνολικά 12 AU (1,2,4,5,6,9,12,15,17,20,25,26). Στα πειράματα της παρούσας εργασίας χρησιμοποιήθηκε αυτή τη βάση δεδομένων με 11 AU, εξαιρουμένου του AU12 καθώς δεν ενεργοποιήθηκε ποτέ. Παραδείγματα καρτέ βίντεο από τη βάση αυτή φαίνονται παρακάτω (εικόνα 4.11).

Εικόνα 4.11: DISFA, aligned images with AUs annotations



## 4.2.2 Στατιστικά βάσεων

Σ αυτή την υπό-ενότητα παρουσιάζονται τα στατιστικά των βάσεων δεδομένων ως προς τη φυλή, την ηλικία, το φύλο στα τρία σύνολα εκπαίδευσης πριν και μετά το δίκαιο χωρισμό. Συγκεκριμένα, για κάθε βάση δεδομένων και κάθε εργασία αναγνώρισης έκφρασης φαίνονται τα ποσοστά των εικόνων που ανήκουν στην κάθε κατηγορία (φυλή, ηλικία, φύλο) για τα τρία σύνολα εκπαίδευσης. Επίσης, συμπεριλαμβάνονται και τα στατιστικά για τις επισημάνσεις κάθε εργασίας πριν και μετά το δίκαιο χωρισμό. Για την εργασία αναγνώρισης βασικών συναισθημάτων περιλαμβάνονται τα ποσοστά για κάθε κλάση συναισθήματος. Για την εργασία εκτίμησης σθένους και διέγερσης παρουσιάζονται τα δισδιάστατα ιστογράμματα για τα σύνολα εκπαίδευσης πριν και μετά το δίκαιο διαχωρισμό. Κάθε άξονας  $\chi, \psi$  επεκτείνεται από το -1 μέχρι το 1 και αναπαριστά τις τιμές διέγερσης (arousal) και σθένους (valence) αντίστοιχα. Τα ιστογράμματα αυτά δείχνουν το πλήθος των εικόνων που έχουν επισήμανση το αντίστοιχο ζεύγος τιμών valence και arousal. Όπως εξηγεί η επισήμανση στα δεξιά των ιστογραμμάτων, στις ανοιχτόχρωμες περιοχές υπάρχουν περισσότερα δείγματα, ενώ στις σκουρόχρωμες λιγότερα. Για την εργασία εκτίμησης AUs, φαίνονται τα ποσοστά ενεργοποίησης για κάθε AU στα τρία σύνολα εκπαίδευσης, πριν και μετά το δίκαιο χωρισμό.

### 4.2.2.1 Εργασία 1: Αναγνώριση βασικών συναισθημάτων

Η πρώτη εργασία εκφραστικού υπολογισμού είναι η αναγνώριση βασικών κατηγορικών συναισθημάτων. Σε αυτή την παράγραφο παρουσιάζονται τα ποσοστά των εικόνων που ανήκουν σε κάθε κατηγορία (φυλή, ηλικία, φύλο) για τις βάσεις RAF-DB (Πίνακας 4.1) και AffectNet (Πίνακας 4.2). Τα ποσοστά αυτά δείχνουν για κάθε σύνολο (εκπαίδευσης - επικύρωσης - ελέγχου) πόσες εικόνες ανήκουν σε κάθε υποκατηγορία ηλικίας, φυλής και φύλου. Επίσης, στους ίδιους πίνακες φαίνονται και τα ποσοστά των εικόνων σε όλα τα σύνολα που ανήκουν σε κάθε κλάση συναισθήματος.

4.1: Statistics BEFORE and AFTER fair split, RAF-DB, Emotion Recognition

RAF-DB, Emotion Recognition								
BEFORE split								
Labels	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	#images
Train %	0.057	0.058	0.023	0.389	0.206	0.162	0.105	12271
Validation %	0.053	0.052	0.024	0.386	0.222	0.156	0.107	3068
Race	Caucasian	African-American	Asian					
Train %	0.765	0.079	0.156					
Validation %	0.766	0.076	0.157					
Age	0-3	4-19	20-39	40-69	70+			
Train %	0.105	0.177	0.532	0.156	0.030			
Validation %	0.107	0.158	0.542	0.164	0.029			
Gender	Male	Female	Unsure					
Train %	0.404	0.535	0.061					
Validation %	0.407	0.528	0.065					
AFTER split								
Labels	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	#images
Train %	0.056	0.057	0.022	0.391	0.209	0.160	0.105	8327
Validation %	0.055	0.054	0.020	0.398	0.210	0.160	0.104	2206
Test %	0.059	0.060	0.026	0.379	0.207	0.161	0.108	4806
Race	Caucasian	African-American	Asian					
Train %	0.770	0.076	0.154					
Validation %	0.780	0.071	0.149					
Test %	0.751	0.086	0.163					
Age	0-3	4-19	20-39	40-69	70+			
Train %	0.104	0.172	0.538	0.158	0.028			
Validation %	0.099	0.171	0.549	0.156	0.024			
Test %	0.110	0.176	0.520	0.159	0.035			
Gender	Male	Female	Unsure					
Train %	0.404	0.535	0.061					
Validation %	0.404	0.539	0.057					
Test %	0.406	0.528	0.066					

4.2: Statistics BEFORE and AFTER fair split, AffectNet, Emotion Recognition

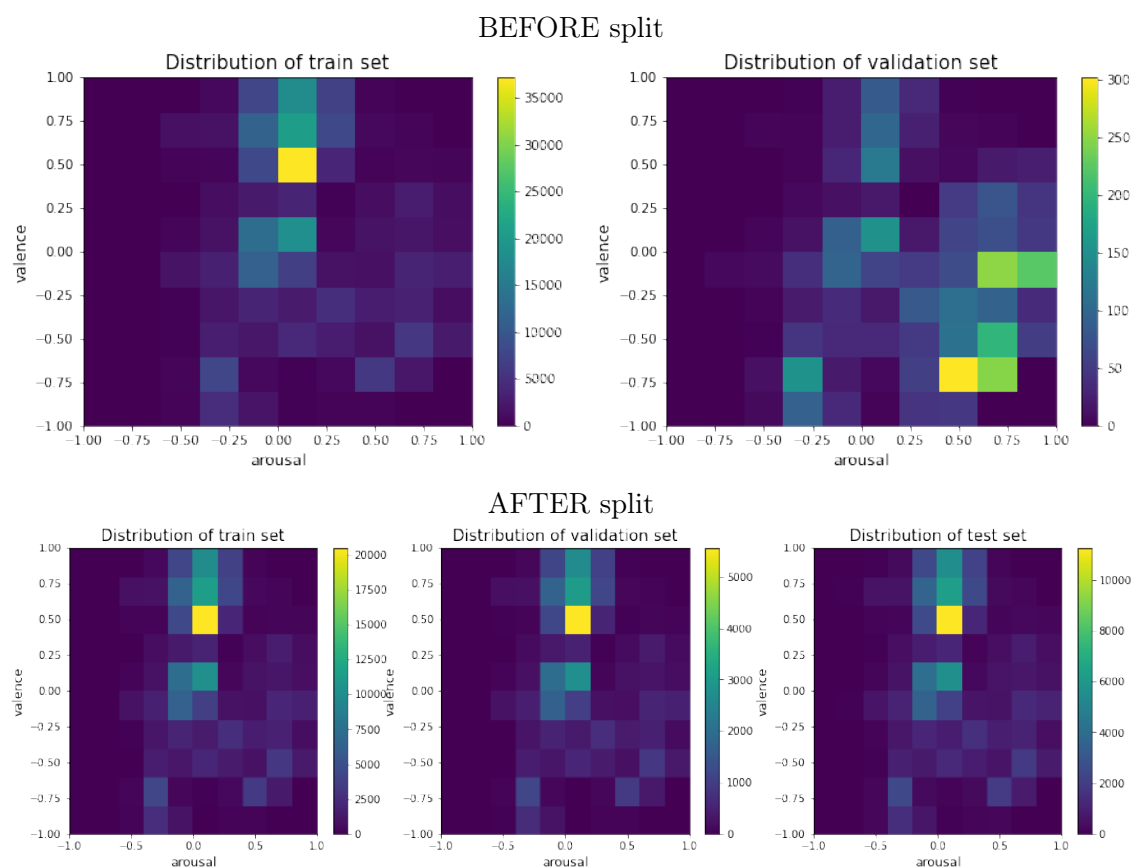
AffectNet, Emotion Recognition									
BEFORE split									
Labels	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	#images
Train %	0.086	0.013	0.013	0.022	0.468	0.260	0.088	0.049	286594
Validation %	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	3986
Race	White	Black	Asian	Indian					
Train %	0.768	0.082	0.089	0.060					
Validation %	0.783	0.081	0.078	0.059					
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.054	0.062	0.052	0.411	0.197	0.099	0.080	0.033	0.012
Validation %	0.072	0.064	0.044	0.367	0.206	0.107	0.091	0.041	0.007
Gender	Male	Female							
Train %	0.496	0.504							
Validation %	0.550	0.450							
AFTER split									
Labels	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	#images
Train %	0.087	0.014	0.015	0.023	0.464	0.259	0.089	0.050	159540
Validation %	0.087	0.014	0.014	0.023	0.465	0.259	0.089	0.050	43330
Test %	0.087	0.015	0.015	0.024	0.461	0.258	0.089	0.050	87710
Race	White	Black	Asian	Indian					
Train %	0.770	0.082	0.089	0.060					
Validation %	0.772	0.081	0.088	0.059					
Test %	0.766	0.083	0.090	0.061					
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.054	0.062	0.052	0.411	0.197	0.099	0.080	0.033	0.011
Validation %	0.054	0.062	0.052	0.413	0.198	0.099	0.080	0.032	0.011
Test %	0.054	0.062	0.053	0.409	0.197	0.099	0.080	0.033	0.012
Gender	Male	Female							
Train %	0.497	0.503							
Validation %	0.497	0.503							
Test %	0.497	0.503							



### 4.2.2.2 Εργασία 2: Εκτίμηση τιμών valence & arousal

Η δεύτερη εργασία εκφραστικού υπολογισμού είναι η εκτίμηση των τιμών σθένους και διέγερσης. Στους πίνακες, παρουσιάζονται τα ποσοστά των εικόνων που ανήκουν σε κάθε κατηγορία (φυλή, ηλικία, φύλο) για τις βάσεις AffectNet (Πίνακας 4.3) και AFEW-VA (Πίνακας 4.4). Τα ποσοστά, όπως και στην προηγούμενη εργασία δείχνουν για κάθε σύνολο (εκπαίδευσης - επικύρωσης - ελέγχου) πόσες εικόνες υπάρχουν σε κάθε υποκατηγορία ηλικίας, φυλής και φύλου. Επίσης επισυνάπτονται και τα διδιάστατα ιστογράμματα που δείχνουν την κατανομή των τιμών valence και arousal πριν και μετά το δίκαιο χωρισμό και για τις δύο βάσεις δεδομένων (AffectNet: Εικόνα 4.12, AFEW-VA: Εικόνα 4.13). Τα ιστογράμματα αυτά δείχνουν το πλήθος των δειγμάτων που έχουν επισήμανση το αντίστοιχο ζεύγος τιμών valence και arousal.

Εικόνα 4.12: *Valence & Arousal Distribution, BEFORE and AFTER split, AffectNet*

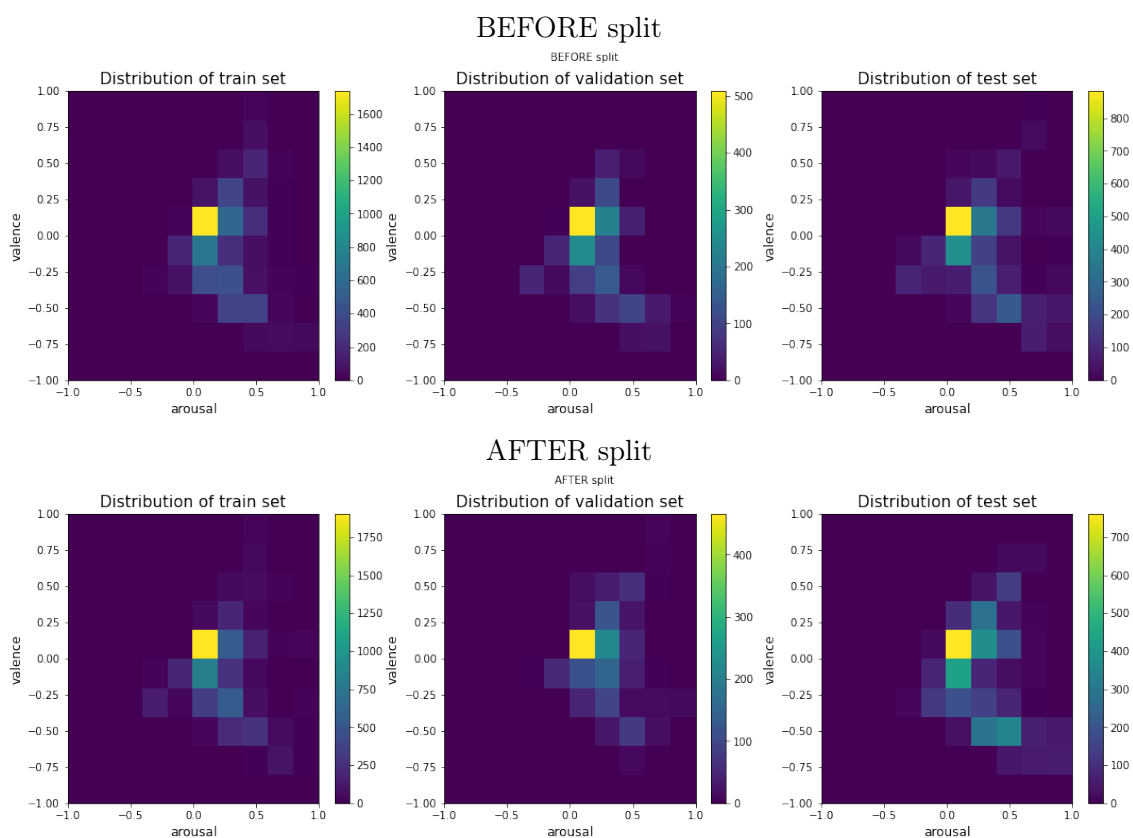




4.3: Statistics AFTER fair split, AffectNet, Valence &amp; Arousal Estimation

AffectNet, Valence&Arousal Estimation									
AFTER split									
Race	White	Black	Asian	Indian	#images				
Train %	0.775	0.080	0.087	0.058	157650				
Validation %	0.790	0.076	0.081	0.053	41633				
Test %	0.748	0.089	0.096	0.067	91297				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.031	0.036	0.030	0.242	0.115	0.057	0.464	0.019	0.006
Validation %	0.052	0.059	0.049	0.423	0.200	0.098	0.079	0.030	0.009
Test %	0.056	0.065	0.055	0.399	0.194	0.100	0.082	0.036	0.014
Gender	Male	Female							
Train %	0.497	0.503							
Validation %	0.495	0.505							
Test %	0.498	0.502							

Εικόνα 4.13: Valence &amp; Arousal Distribution, BEFORE and AFTER split, AFEW-VA



## 4.4: Statistics BEFORE and AFTER fair split, AFEW-VA, Valence &amp; Arousal Estimation

AFEW-VA, Valence & Arousal Estimation									
BEFORE split									
Race	White	Black	Asian	Indian	# images				
Train %	0.892	0.057	0.051	0.000	6483				
Validation %	0.975	0.000	0.025	0.000	1966				
Test %	0.932	0.000	0.048	0.020	3708				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.009	0.012	0.048	0.401	0.294	0.121	0.106	0.000	0.009
Validation %	0.000	0.000	0.000	0.417	0.377	0.096	0.084	0.025	0.000
Test %	0.000	0.000	0.080	0.374	0.296	0.135	0.096	0.019	0.000
Gender	Male	Female							
Train %	0.575	0.425							
Validation %	0.457	0.543							
Test %	0.658	0.342							
AFTER split									
Race	White	Black	Asian	Indian					
Train %	0.950	0.013	0.033	0.004	6345				
Validation %	0.988	0.000	0.012	0.000	1701				
Test %	0.838	0.069	0.080	0.012	4111				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.000	0.000	0.052	0.410	0.318	0.114	0.105	0.000	0.000
Validation %	0.000	0.000	0.044	0.406	0.319	0.120	0.111	0.000	0.000
Test %	0.015	0.019	0.050	0.367	0.287	0.132	0.087	0.029	0.014
Gender	Male	Female							
Train %	0.567	0.433							
Validation %	0.595	0.405							
Test %	0.598	0.402							

### 4.2.2.3 Εργασία 3: Ανίχνευση AUs

Η τρίτη εργασία εκφραστικού υπολογισμού είναι η ανίχνευση των ενεργοποιημένων Action Units. Σε αυτή την παράγραφο παρουσιάζονται τα ποσοστά των εικόνων που ανήκουν σε κάθε κατηγορία (φυλή, ηλικία, φύλο) για τις βάσεις EmotioNet (Πίνακας 4.6), DISFA (Πίνακας 4.8), GFT (Πίνακας 4.10) και RAF-AU (Πίνακας 4.12). Όπως και στις δύο προηγούμενες εργασίες, τα ποσοστά αυτά δείχνουν για κάθε σύνολο (εκπαίδευσης - επικύρωσης - ελέγχου) το πόσες εικόνες αντιστοιχούν σε κάθε υποκατηγορία ηλικίας, φυλής και φύλου. Επίσης επισυνάπτονται και τα ποσοστά των AUs που είναι ενεργοποιημένα στα σύνολα εκπαίδευσης πριν και μετά το δίκαιο διαχωρισμό (EmotioNet: Πίνακας 4.5, DISFA: Πίνακας 4.7, GFT: Πίνακας 4.9, RAF-AU: Πίνακας 4.11). Τα ποσοστά των ενεργοποιημένων AUs δείχνουν για κάθε Action Unit και κάθε σύνολο (εκπαίδευσης - επικύρωσης - ελέγχου), το πλήθος των εικόνων στις οποίες το συγκεκριμένο AU είναι ενεργοποιημένο.

4.5: AUs activation BEFORE and AFTER fair split, EmotioNet

EmotioNet, AUs activation					
AUs	BEFORE		AFTER		
	Train	Validation	Train	Validation	Test
1	0.0632	0.0591	0.0616	0.0834	0.0499
2	0.0298	0.0424	0.0299	0.0591	0.0348
4	0.1189	0.0658	0.1152	0.0723	0.0813
5	0.0377	0.0643	0.0369	0.0620	0.0668
6	0.1962	0.2403	0.2002	0.2717	0.2181
9	0.0213	0.0070	0.0197	0.0166	0.0051
12	0.3822	0.4002	0.3859	0.4039	0.3932
17	0.0215	0.0092	0.0203	0.0157	0.0080
20	0.0055	0.0072	0.0051	0.0083	0.0072
25	0.4939	0.4759	0.4945	0.5030	0.4612
26	0.0866	0.0884	0.0866	0.0942	0.0857

4.6: Statistics BEFORE and AFTER fair split, EmotioNet, AUs Detection

EmotioNet, AUs Detection									
BEFORE split									
Race	White	Black	Asian	Indian	# images				
Train %	0.804	0.080	0.057	0.059	24608				
Validation %	0.794	0.095	0.051	0.060	20320				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.024	0.038	0.043	0.410	0.204	0.127	0.101	0.042	0.011
Validation %	0.013	0.024	0.037	0.439	0.207	0.127	0.103	0.041	0.009
Gender	Male	Female							
Train %	0.524	0.476							
Validation %	0.501	0.499							
AFTER split									
Race	White	Black	Asian	Indian					
Train %	0.800	0.087	0.054	0.059	24674				
Validation %	0.802	0.086	0.054	0.058	6705				
Test %	0.796	0.088	0.056	0.060	13549				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.019	0.032	0.040	0.424	0.206	0.127	0.102	0.042	0.010
Validation %	0.018	0.031	0.040	0.425	0.206	0.127	0.102	0.041	0.010
Test %	0.019	0.032	0.041	0.421	0.205	0.127	0.102	0.042	0.011
Gender	Male	Female							
Train %	0.514	0.486							
Validation %	0.514	0.486							
Test %	0.514	0.486							

## 4.7: AUs activation BEFORE and AFTER fair split, DISFA

DISFA, AUs activation					
AUs	BEFORE		AFTER		
	Train	Test	Train	Validation	Test
1	0.0619	0.0952	0.0668	0.0716	0.0695
2	0.0513	0.0804	0.0530	0.0176	0.0767
4	0.1855	0.2107	0.1822	0.1073	0.2324
5	0.0241	0.0075	0.0289	0.0104	0.0113
6	0.1456	0.1669	0.1713	0.0986	0.1311
9	0.0456	0.0906	0.0532	0.0574	0.0545
15	0.2450	0.2132	0.2393	0.3037	0.2159
17	0.1003	0.0967	0.0860	0.0788	0.1294
20	0.0290	0.0543	0.0370	0.0166	0.0344
25	0.3715	0.2997	0.3223	0.6515	0.3161
26	0.1893	0.2149	0.1876	0.4311	0.1236

## 4.8: Statistics BEFORE and AFTER fair split, DISFA, AUs Detection

DISFA, AUs Detection									
BEFORE split									
Race	White	Black	Asian	Indian	# images				
Train %	0.812	0.047	0.094	0.047	206103				
Validation %	0.800	0.000	0.200	0.000	48450				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.000	0.000	0.047	0.825	0.094	0.034	0.000	0.000	0.000
Validation %	0.000	0.000	0.000	0.800	0.000	0.200	0.000	0.000	0.000
Gender	Male	Female							
Train %	0.551	0.449							
Validation %	0.600	0.400							
AFTER split									
Race	White	Black	Asian	Indian					
Train %	0.931	0.000	0.069	0.000	141041				
Validation %	1.000	0.000	0.000	0.000	29070				
Test %	0.542	0.115	0.229	0.115	84442				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.000	0.000	0.000	0.931	0.000	0.069	0.000	0.000	0.000
Validation %	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
Test %	0.000	0.000	0.115	0.573	0.229	0.083	0.000	0.000	0.000
Gender	Male	Female							
Train %	0.550	0.450							
Validation %	0.667	0.333							
Test %	0.541	0.459							

4.9: AUs activation BEFORE and AFTER fair split, GFT

GFT, AUs activation					
AUs	BEFORE			AFTER	
	Train	Test	Train	Validation	Test
1	0.0370	0.0448	0.0431	0.0182	0.0423
2	0.1338	0.1221	0.1266	0.1233	0.1454
4	0.0367	0.0339	0.0527	0.0197	0.0172
5	0.0240	0.0161	0.0174	0.0070	0.0405
6	0.2836	0.2770	0.3010	0.2971	0.2412
7	0.4551	0.4507	0.4890	0.4467	0.3983
9	0.0140	0.0153	0.0185	0.0040	0.0129
10	0.2463	0.2467	0.2734	0.2443	0.2006
11	0.1375	0.1505	0.1597	0.1276	0.1127
12	0.2950	0.2916	0.3099	0.3135	0.2558
15	0.1063	0.0954	0.1238	0.0942	0.0762
17	0.3015	0.3240	0.2930	0.2745	0.3462
23	0.2488	0.2549	0.2449	0.2392	0.2650
24	0.1426	0.1412	0.1452	0.0802	0.1742

4.10: Statistics BEFORE and AFTER fair split, GFT, AUs Detection

GFT, AUs Detection						
BEFORE split						
Race	White	Black	Hispanic	Asian	Other	# images
Train %	0.895	0.074	0.007	0.024	0.000	108549
Validation %	0.751	0.198	0.000	0.000	0.051	24645
Age	21-22	23-25	26-28			
Train %	0.566	0.354	0.080			
Validation %	0.822	0.088	0.090			
Gender	Male	Female				
Train %	0.579	0.421				
Validation %	0.549	0.451				
AFTER split						
Race	White	Black	Hispanic	Asian	Other	# images
Train %	0.870	0.130	0.000	0.000	0.000	69618
Validation %	1.000	0.000	0.000	0.000	0.000	23685
Test %	0.787	0.096	0.019	0.066	0.031	39891
Age	21-22	23-25	26-28			
Train %	0.547	0.376	0.077			
Validation %	0.597	0.403	0.000			
Test %	0.536	0.349	0.114			
Gender	Male	Female				
Train %	0.563	0.437				
Validation %	0.587	0.413				
Test %	0.583	0.417				

4.11: AUs activation BEFORE and AFTER fair split, RAF-AU

RAF-AU, AUs activation						
AUs	BEFORE			AFTER		
	Train	Validation	Test	Train	Validation	Test
1	0.2410	0.2264	0.2254	0.2410	0.2543	0.2153
2	0.1736	0.1698	0.1740	0.1669	0.1876	0.1781
4	0.3899	0.4025	0.4112	0.4026	0.3643	0.4039
5	0.2144	0.2096	0.2093	0.2426	0.2171	0.1578
6	0.0966	0.1048	0.0977	0.0883	0.0930	0.1164
9	0.1718	0.1509	0.1696	0.1559	0.1364	0.2062
10	0.3058	0.3229	0.2900	0.2677	0.3178	0.3569
12	0.2795	0.2683	0.2695	0.2365	0.2760	0.3422
16	0.1551	0.1363	0.1645	0.1223	0.1659	0.2097
17	0.1196	0.1216	0.1175	0.1300	0.1318	0.0947
25	0.6190	0.5912	0.6116	0.5986	0.6326	0.6318
26	0.2362	0.2327	0.2372	0.2341	0.2481	0.2342
27	0.1773	0.1761	0.1689	0.1547	0.1519	0.2195

4.12: Statistics BEFORE and AFTER fair split, RAF-AU, AUs Detection

RAF-AU, AUs Detection									
BEFORE split									
Race	White	Black	Asian	Indian	# images				
Train %	0.793	0.056	0.099	0.053	2701				
Validation %	0.774	0.061	0.101	0.065	477				
Test %	0.772	0.070	0.101	0.058	1362				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.076	0.126	0.047	0.376	0.234	0.082	0.033	0.014	0.012
Validation %	0.082	0.145	0.040	0.331	0.258	0.090	0.027	0.017	0.010
Test %	0.079	0.130	0.055	0.358	0.230	0.089	0.043	0.007	0.010
Gender	Male	Female							
Train %	0.491	0.509							
Validation %	0.472	0.528							
Test %	0.506	0.494							
AFTER split									
Race	White	Black	Asian	Indian					
Train %	0.790	0.058	0.098	0.053	2469				
Validation %	0.812	0.053	0.090	0.045	645				
Test %	0.762	0.067	0.107	0.065	1426				
Age	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Train %	0.077	0.129	0.047	0.369	0.237	0.084	0.035	0.011	0.011
Validation %	0.074	0.129	0.047	0.381	0.242	0.082	0.031	0.008	0.006
Test %	0.079	0.129	0.053	0.353	0.230	0.087	0.039	0.016	0.014
Gender	Male	Female							
Train %	0.494	0.506							
Validation %	0.491	0.509							
Test %	0.493	0.507							

## 4.3 Δίκαιος Διαχωρισμός

Το επόμενο στάδιο στην εργασία είναι η εφαρμογή του δίκαιου χωρισμού των βάσεων στα τρία σύνολα εκπαίδευσης, με βάση τις ετικέτες, τη φυλή, την ηλικία και το φύλο. Στην ενότητα αυτή, αρχικά, περιγράφεται η γενική μεθοδολογία η οποία χρησιμοποιήθηκε για τον σκοπό αυτό και έπειτα αναπτύσσεται για κάθε βάση ξεχωριστά.

### Γενική μεθοδολογία

Αφού έχουν ληφθεί οι αυτόματες ή χειροκίνητες επισημάνσεις για φυλή, ηλικία και φύλο στα δεδομένα των βάσεων εφαρμόζεται η δίκαια διάσπαση τους στα τρία σύνολα εκπαίδευσης. Θα πρέπει να σημειωθεί ότι για τις βάσεις δεδομένων RAF-DB και GFT χρησιμοποιήθηκαν οι δοσμένοι χειροκίνητοι σχολιασμοί για τη φυλή, την ηλικία και το φύλο ενώ για τις υπόλοιπες βάσεις δεδομένων λήφθηκαν αυτόματα οι τιμές αυτές από το μοντέλο ανίχνευσης. Σε αυτή την περίπτωση για τις βάσεις δεδομένων που είναι οπτικοακουστικές, έγινε έλεγχος και προσαρμογή έτσι ώστε όλα τα καρέ από κάθε βίντεο να συμφωνούν στις τιμές φυλής, ηλικίας και φύλου.

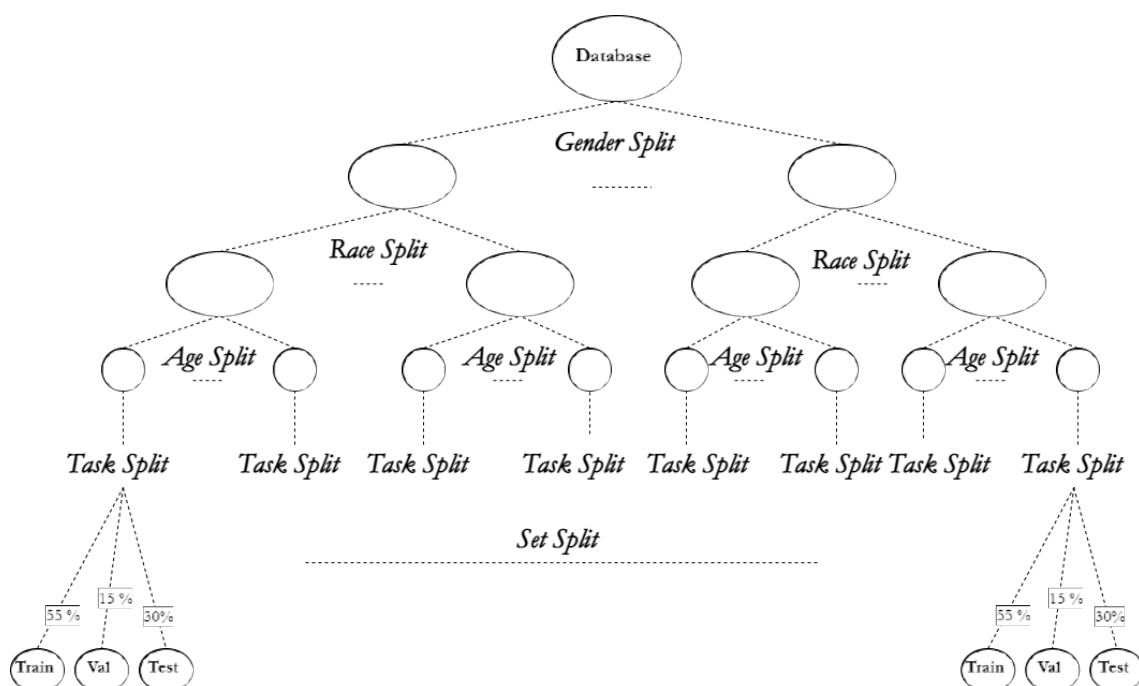
Κάθε εικόνα κάθε βάσης δεδομένων προστίθεται σε μία διαφορετική υποκατηγορία με βάση την τιμή επισημάνσης για φυλή ηλικία και φύλο. Έπειτα, εφαρμόζεται επιπλέον διαχωρισμός στις υποκατηγορίες αυτές με βάση τις ετικέτες που έχουν στη αντίστοιχη εργασία αναγνώρισης έκφρασης (task split). Το task split διαφέρει από τη μια εργασία στην άλλη. Ειδικότερα για την εργασία αναγνώρισης συναισθημάτων, ο διαχωρισμός εκτελέστηκε με διαχωρισμό των εικόνων κάθε υποκατηγορίας στις κατηγορίες συναισθημάτων. Για την εργασία εκτίμησης σθένους και διέγερσης, οι εικόνες κάθε υποκατηγορίας διαχωρίστηκαν σε 100 κατηγορίες οι οποίες σχηματίστηκαν με το διαχωρισμό σε διαστήματα 0.2 των τιμών σθένους και διέγερσης (από την τιμή -1 στην τιμή 1). Για την εργασία ανίχνευσης AUs, το task split παραλείφθηκε, αλλά έγινε έλεγχος έτσι ώστε τα ποσοστά κάθε AU που είναι ενεργοποιημένα να είναι συγκρίσιμα στα τρία σετ (εκπαίδευσης, επικύρωσης, ελέγχου). Στη συνέχεια, δημιουργήθηκαν τα τρία σύνολα λαμβάνοντας από κάθε τελική υποκατηγορία τα ποσοστά 55%, 15%, 30% και τοποθετώντας τα στα σύνολα εκπαίδευσης, επικύρωσης και ελέγχου αντίστοιχα. Η μεθοδολογία δίκαιου διαχωρισμού δεδομένων φαίνεται το Σχήμα 4.1. Κατά τη διάρκεια των διαχωρισμών όταν οι βάσεις δεδομένων αποτελούνταν από εικόνες, κάθε εικόνα λήφθηκε ξεχωριστά, ωστόσο όταν οι βάσεις δεδομένων αποτελούνταν από βίντεο, τα καρέ κάθε βίντεο λήφθηκαν μαζί.

### Μεθοδολογία για κάθε βάση δεδομένων

Ο δίκαιος χωρισμός εφαρμόστηκε σε κάθε μία από τις προαναφερθείσες βάσεις δεδομένων για την εργασία αναγνώρισης έκφρασης που αντιμετωπίζουν.

Στην εργασία αναγνώρισης βασικών συναισθημάτων χρησιμοποιήθηκαν οι βάσεις δεδομένων AffectNet και RAF-DB. Για τη βάση δεδομένων AffectNet λήφθηκαν αυτόματοι σχολιασμοί για τη φυλή (4 κατηγορίες), την ηλικία (9 κατηγορίες) και το φύλο (2 κατηγορίες) από το μοντέλο FairFace, ενώ η βάση δεδομένων RAF-DB περιείχε μη αυτόματους σχολιασμούς με 3 φυλές, 5 διαστήματα ηλικίας και 3 κατηγορίες για το φύλο. Έτσι, στις βάσεις δεδομένων

Σχήμα 4.1: Σχηματική αναπαράσταση της μεθοδολογίας του δίκαιου διαχωρισμού



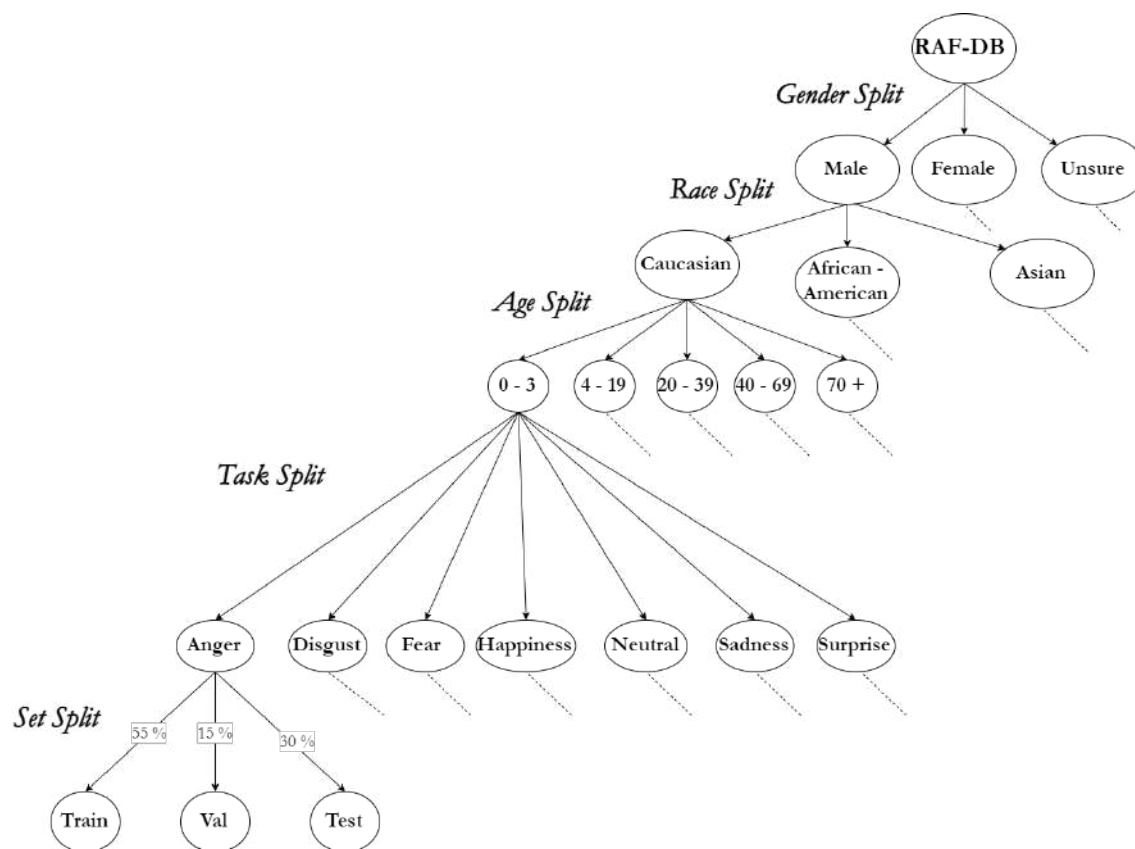
AffectNet και RAF-DB οι εικόνες χωρίστηκαν σε 72 και 45 υποκατηγορίες αντίστοιχα, με βάση τις τιμές φυλής, ηλικίας και φύλου. Στη συνέχεια, οι υποκατηγορίες χωρίστηκαν σε 8 υποκατηγορίες στην πρώτη βάση δεδομένων και σε 7 υποκατηγορίες στη δεύτερη βάση δεδομένων, με βάση την κλάση συναισθήματος. Στη συνέχεια, λήφθηκε το ποσοστό 55%, 15% και 30% από κάθε υποκατηγορία και τοποθετήθηκε στα σύνολα εκπαίδευσης, επικύρωσης και ελέγχου αντίστοιχα. Στο σχήμα 4.2 παρουσιάζεται σχηματική ο δίκαιος χωρισμός για τη βάση RAF-DB.

Στην εργασία εκτίμησης σθένους και διέγερσης χρησιμοποιήθηκε η στατική βάση δεδομένων AffectNet και η οπτικοακουστική βάση δεδομένων AFEW-VA. Και οι δύο αυτές βάσεις δεδομένων έλαβαν αυτόματους σχολιασμούς για τη φυλή, την ηλικία και το φύλο από το μοντέλο FairFace. Οι εικόνες της AffectNet χωρίστηκαν σε 72 υποκατηγορίες και στη συνέχεια σε 100 υποκατηγορίες με βάση τις τιμές σθένους και διέγερσης. Τα τρία σύνολα διαμορφώθηκαν με το χωρισμό των υποκατηγοριών με βάση τα τρία ποσοστά. Τα βίντεο της οπτικοακουστικής βάσης δεδομένων AFEW-VA χωρίστηκαν σε 72 υποκατηγορίες. Από αυτές τις υποκατηγορίες τοποθετήθηκε κάθε βίντεο σε ένα από τα τρία σύνολα εκπαίδευσης χειροκίνητα, λαμβάνοντας υπόψη το επιθυμητό μέγεθος του κάθε συνόλου και την κατανομή του σθένους και της διέγερσης να είναι όσο το δυνατόν πιο όμοια στα τρία σύνολα.

Για την τελευταία εργασία, τον εντοπισμό AUs, χρησιμοποιήθηκαν οι βάσεις δεδομένων EmotioNet, RAF-DB, DISFA και GFT. Για τις δύο πρώτες βάσεις δεδομένων ανακτήθηκαν αυτόματοι σχολιασμοί φυλής, ηλικίας και φύλου και οι εικόνες χωρίστηκαν σε 72 υποκατηγορίες και στη συνέχεια χωρίστηκαν στα τρία σετ με τα προαναφερθέντα ποσοστά. Η βάση δεδομένων DISFA σχολιάστηκε επίσης αυτόματα για τη φυλή, την ηλικία και το φύλο, επομένως τα βίντεο χωρίστηκαν σε 72 κατηγορίες και στη συνέχεια ταξινομήθηκαν χειροκίνητα στα τρία σετ. Τέλος, η βάση δεδομένων GFT περιείχε χειροκίνητους σχολιασμούς για τη φυλή



Σχήμα 4.2: Σχηματική αναπαράσταση της μεθοδολογίας του δίκαιου διαχωρισμού για τη βάση RAF-DB



(5 κατηγορίες), την ηλικία (21-28) και το φύλο (2 κατηγορίες). Για την ηλικία σχηματίστηκαν τρεις κατηγορίες 21-22, 23-25 και 26-28. Τα βίντεο αυτής της βάσης δεδομένων χωρίστηκαν σε 30 κατηγορίες και στη συνέχεια οργανώθηκαν χειροκίνητα στα τρία σετ. Για την εργασία αυτή οι εικόνες (ή τα καρέ) δεν διαχωρίστηκαν επιπλέον με βάση την τιμή των AUs, αλλά έγινε έλεγχος έτσι ώστε το ποσοστό κάθε ενεργοποιημένου AU να είναι συγκρίσιμο στα τρία σετ (εκπαίδευση, επικύρωση, έλεγχος).

## 4.4 Μοντέλα εργασιών αναγνώρισης έκφρασης

Το τελευταίο βήμα της μεθοδολογίας είναι η εκπαίδευση, η αξιολόγηση και ο έλεγχος των βάσεων δεδομένων πριν και μετά το δίκαιο χωρισμό, για κάθε εργασία αναγνώρισης έκφρασης. Αρχικά δημιουργήθηκαν τρία μεμονωμένα δίκτυα που εκτελούσαν κάθε εργασία, προσαρμόζοντας προ-εκπαιδευμένα βαθιά νευρωνικά δίκτυα. Συγκεκριμένα, απορρίφθηκε το πλήρως συνδεδεμένο επίπεδο στην κορυφή των προ-εκπαιδευμένων δικτύων και προστέθηκε ένα νέο πλήρως συνδεδεμένο επίπεδο που υλοποιεί κάθε εργασία. Οι προ-εκπαιδευμένες αρχιτεκτονικές μοντέλων που χρησιμοποιήθηκαν είναι οι ResNet50 [5], VGG16 [6], DenseNet121 [7].

Η εργασία αναγνώρισης συναισθημάτων είναι μια εργασία ταξινόμησης πολλών κλάσεων (Multiclass classification task). Για αυτήν την εργασία προστέθηκε ένα τελικό πλήρως συνδεδεμένο επίπεδο με τόσες εξόδους όσες και ο αριθμός των κατηγορικών συναισθημάτων κάθε βάσης δεδομένων. Για συνάρτηση ενεργοποίησης χρησιμοποιήθηκε η softmax, για να ανατεθούν δεκαδικές πιθανότητες σε κάθε κλάση. Η συνάρτηση απώλειας που χρησιμοποιήθηκε είναι η κατηγορική διασταυρούμενη εντροπία (cross entropy loss) και για τη μέτρηση της απόδοσης το f1 score και η averaged accuracy.

Η εργασία εκτίμησης σθένους και διέγερσης είναι μια εργασία πολλαπλών εργασιών παλινδρόμησης (multitask regression task). Το τελικό πλήρως συνδεδεμένο επίπεδο είναι γραμμικό και αποτελείται από δύο εξόδους που αντιστοιχούν στις τιμές του σθένους και της διέγερσης. Για τη συνάρτηση απώλειας χρησιμοποιήθηκε ο συντελεστή συσχέτισης (Concordance Correlation Coefficient - CCC) και το Mean Squared Error - MSE και για τη μέτρηση της απόδοσης, το CCC.

Η εργασία εκτίμησης Action Units είναι μια δυαδική εργασία ταξινόμησης πολλαπλών ετικετών. Προστέθηκε ένα τελικό πλήρως συνδεδεμένο επίπεδο που έχει τόσες εξόδους όσες είναι ο αριθμός των AU κάθε βάσης δεδομένων. Η συνάρτηση ενεργοποίησης είναι σιγμοειδής, για να προβλέψει την ενεργοποίηση (1) ή την απενεργοποίηση (0) κάθε AU. Η συνάρτηση απώλειας που χρησιμοποιήθηκε είναι η δυαδική διασταυρούμενη εντροπία (cross entropy loss) και η μέτρηση απόδοσης είναι ο μέσος όρος του binary f1 score κάθε AU.

## Κεφάλαιο 5

### Σχετικές εργασίες

---

Στη βιβλιογραφία υπάρχουν πολλές εργασίες που το αντικείμενό τους σχετίζεται με το αντικείμενο της παρούσας εργασίας.

Τα άρθρα [34, 35] παρουσιάζουν μια νέα προσέγγιση για τη σύνθεση συναισθήματος του προσώπου, είτε ως προς τις έξι βασικές εκφράσεις, είτε ως προς το σθένος και τη διέγερση. Λαμβάνεται ως είσοδος μία ουδέτερη διδιάστατη εικόνα ενός ατόμου και μία βασική έκφραση ή ένα ζεύγος τιμών σθένους και διέγερσης και παράγεται μία ρεαλιστική, συνθετική εικόνα με το αντίστοιχο συναίσθημα. Στη συνέχεια, διεξάγονται ποσοτικά πειράματα, στα οποία οι συνθετικές εικόνες χρησιμοποιούνται για επαύξηση δεδομένων στην εκπαίδευση των Deep Neural Networks για την εκτέλεση εργασιών αναγνώρισης εκφράσεων σε όλες τις βάσεις δεδομένων. Σε κάθε περίπτωση παρατηρούνται βελτιωμένες επιδόσεις.

Στις εργασίες [36, 37], παρουσιάζεται μια νέα προσέγγιση βασισμένη στο CNN-RNN, η οποία εκμεταλλεύεται πολλαπλά χαρακτηριστικά του CNN για αναγνώριση διαστατικών συναισθημάτων in-the-wild, χρησιμοποιώντας τη βάση δεδομένων OMG-Emotion. Γίνεται προ-εκπαίδευση με μεγάλες βάσεις δεδομένων με χρήση CNN, από την οποία εξάγονται χαρακτηριστικά χαμηλού, μεσαίου και υψηλού επιπέδου και αξιοποιούνται από υποδίκτυα RNN.

Στο [38], δημιουργείται ένα δίκτυο το VA-StarGAN, το οποίο συνθέτει διαστατικές αναπαραστάσεις προσώπου (τιμές σθένους και διέγερση) και δημιουργεί ρεαλιστικές εικόνες. Συγκεκριμένα προσαρμόζεται το StarGAN, το οποίο δημιουργεί μόνο έναν διακριτό αριθμό εκφράσεων, για συνεχή σύνθεση συναισθημάτων.

Η μελέτη [39] αφορά την επαύξηση δεδομένων σε προβλήματα αυτόματης αναγνώρισης εκφράσεων προσώπου με χρήση Deep Neural Networks. Συγκεκριμένα προτείνεται μία νέα στρατηγική για data augmentation, η οποία βασίζεται στο Mixup, η MixAugment. Σύμφωνα με αυτή τη στρατηγική το δίκτυο εκπαιδεύεται ταυτόχρονα πάνω σε ένα συνδυασμό από εικονικά και πραγματικά παραδείγματα τα οποία συμβάλλουν στη συνολική συνάρτηση απώλειας. Αποδεικνύεται η αποτελεσματικότητα του MixAugment έναντι του Mixup και διαφόρων μεθόδων τελευταίας τεχνολογίας.

Το άρθρο [40] αναφέρεται στα εμπόδια της απόφραξη και στις παραλλαγές στάσης που αλλάζουν σημαντικά την εμφάνιση του προσώπου στην αναγνώριση εκφράσεων. Εξετάζονται τα προβλήματα αυτά και σχολιάζονται πολλά in-the-wild σύνολα δεδομένων με χαρακτηριστικά πόζας και απόφραξης. Ταυτόχρονα προτείνεται ένα νέο Region Attention Network – RAN, το οποίο καταγράφει τη σημασία των περιοχών προσώπου για απόφραξη και παραλλαγή πόζας. Τέλος παρουσιάζεται ένα biased loss το οποίο ενθαρρύνει τα βάρη υψηλής προσοχής στις ποιο

σημαντικές περιοχές.

Η μελέτη [41] αντιμετωπίζει το πρόβλημα των εξαιρετικά ανισόρροπων κατανομών κλάσεων στην αναγνώριση τιμών Action Units με χρήση νευρωνικών δικτύων. Τα θετικά ενεργοποιημένα Action Units είναι πολύ λιγότερα από τα αρνητικά ενεργοποιημένα και αυτή η ανισορροπία προκαλεί πολύ αργό ρυθμό σύγκλισης σφάλματος. Το πρόβλημα αυτό αντιμετωπίζεται με χρήση πολύ βαθιών (10 επίπεδα) αρχιτεκτονικών, για να αυξηθεί η πιθανότητα η εκπαίδευση δικτύου να έχει αποδεκτό ρυθμό σύγκλισης χρησιμοποιώντας εξαιρετικά ανισόρροπα σύνολα δεδομένων.

Στο άρθρο [42] προτείνεται ένα νέο πλαίσιο end-to-end βαθιάς μάθησης για κοινή ανίχνευση AU και ευθυγράμμιση προσώπου, αφού οι δύο αυτές εργασίες έχουν υψηλή συσχέτιση.

Στη μελέτη [43], προτείνεται μια νέα μέθοδος ανίχνευσης AUs, η οποία αντιμετωπίζει τα προβλήματα των διαφορετικών AUs σε ξεχωριστά άτομα και την αδυναμία εμφάνισης σήματος AUs. Συγκεκριμένα σε αυτή τη μέθοδο ανίχνευσης χρησιμοποιούνται τοπικές πληροφορίες και η σχέση μεταξύ μεμονωμένων περιοχών προσώπου.

Στο άρθρο [44] προτείνεται ένα πλαίσιο εκμάθησης βαθιάς μάθησης end-to-end για την ανίχνευση AU μόνο με ετικέτες AU. Συγκεκριμένα τα χαρακτηριστικά πολλαπλής κλίμακας που μοιράζονται από κάθε AU μαθαίνονται πρώτα και μετά μαθαίνονται τα channel & spatial attentions adaptively, για επιλογή και εξαγωγή τοπικών χαρακτηριστικών που σχετίζονται με AUs. Επιπλέον λαμβάνονται οι σχέσεις σε επίπεδο pixel για τα AU για να βελτιώσουν τα spatial attentions και να εξαχθούν σχετικότερα τοπικά χαρακτηριστικά.

Η εργασία [45] ασχολείται με το ετερογενές Multi Task Learning, με συσχέτιση εργασιών μέσω συνεκπαίδευσης με εργασίες οι οποίες περιέχουν λίγους ή μη επικαλυπτόμενους σχολιασμούς. Συγκεκριμένα προτείνεται μια νέα προσέγγιση αντιστοίχισης κατανομής, στην οποία η ανταλλαγή γνώσεων είναι δυνατή μεταξύ των εργασιών, μέσω της αντιστοίχισης των κατανομών των προβλέψεών τους και χτίζεται ένα πλαίσιο για ανάλυση προσώπου, το FaceBehaviorNet το οποίο μαθαίνει από κοινού όλες τις εργασίες συμπεριφοράς προσώπου. Αναπτύσσονται περιπτωσιολογικές μελέτες για: ι) συνεχή εκτίμηση συναισθημάτων, ανίχνευση AUs, αναγνώριση βασικών συναισθημάτων. ιι) ανίχνευση χαρακτηριστικών, αναγνώριση προσώπου.

## Κεφάλαιο 6

# Λεπτομέρειες Εκπαίδευσης

---

Σε αυτό το κεφάλαιο περιγράφονται λεπτομέρειες για την εκπαίδευση των τριών μοντέλων αναγνώρισης έκφρασης σε κάθε βάση δεδομένων. Περιγράφεται η μεθοδολογία για τη βελτιστοποίηση της απόδοσης των μοντέλων, και αναφέρονται τα δίκτυα και οι υπερπαραμέτροι που χρησιμοποιήθηκαν για κάθε βάση δεδομένων και κάθε εργασία. Έπειτα περιγράφεται η προ-επεξεργασία που έγινε στα σύνολα εισόδου, τόσο στην περίπτωση των μοντέλων ανίχνευσης φυλής, ηλικίας και φύλου, όσο και στην περίπτωση των μοντέλων αναγνώρισης συναισθήματος.

### 6.1 Μεθοδολογία βελτιστοποίησης

Για την εύρεση του καλύτερου δικτύου έγιναν πειράματα με διάφορα δίκτυα και υπερπαραμέτρους. Συγκεκριμένα δοκιμάστηκαν διάφορες αρχιτεκτονικές με προσθήκη πλήρως συνδεδεμένων επιπέδων πάνω από το προ-εκπαιδευμένο δίκτυο και έγινε αλλαγή στον αριθμό των νευρώνων κάθε επιπέδου. Επίσης αξιολογήθηκαν διάφοροι optimizers (Adam, SGD, RMS), διάφορες τιμές learning rate ( $10^{-5} - 10^{-2}$ ), διάφορες τιμές batch size (16, 32, 64). Αρχικά υλοποιήθηκε μια απλή αρχιτεκτονική με προσθήκη μόνο του επιπέδου εξόδου πάνω από το προ-εκπαιδευμένο δίκτυο και έγιναν δοκιμές με διάφορα optimizers, learning rates, batch sizes. Αφού λήφθηκαν οι τιμές των υπερπαραμέτρων αυτών που επιφέρουν την καλύτερη απόδοση, εισήχθησαν περισσότερα επίπεδα στην αρχιτεκτονική του δικτύου, με στόχο την περαιτέρω βελτιστοποίηση. Στα πλαίσια αυτής της εργασίας έγιναν δοκιμές μόνο με μία βάση δεδομένων (RAF-DB) για την εργασία αναγνώρισης βασικού συναισθήματος. Στις υπόλοιπες βάσεις δεδομένων εφαρμόστηκε μία αρχιτεκτονική, χωρίς βελτιστοποίηση υπερπαραμέτρων.<sup>1</sup> Τα αποτελέσματα της βελτιστοποίησης της βάσης δεδομένων για το μοντέλο αναγνώρισης βασικών συναισθημάτων παρουσιάζονται στο τελευταίο κεφάλαιο 7.2.

### 6.2 Τελικά δίκτυα και υπερπαραμέτροι

Στον πίνακα φαίνονται η αρχιτεκτονική και οι υπερπαραμέτροι που χρησιμοποιήθηκαν στα μοντέλα αναγνώρισης έκφρασης για κάθε βάση δεδομένων. Για την πρώτη εργασία, πρόβλεψη κατηγορικών συναισθημάτων, χρησιμοποιήθηκαν οι βάσεις δεδομένων RAF-DB και Affect-

<sup>1</sup>Η επίτευξη των καλύτερων δυνατών αποδόσεων στα μοντέλα αναγνώρισης συναισθηματικής κατάστασης δεν είναι ο σκοπός της προκειμένης εργασίας.

Net, για τη δεύτερη εργασία, πρόβλεψη συνεχών τιμών valence και arousal, οι βάσεις AffectNet και AFEW-VA και για την τρίτη εργασία, πρόβλεψη Facial Action Units, οι βάσεις δεδομένων EmotioNet, DISFA, GFT και RAF-AU. Τα προεκπαιδευμένα δίκτυα που χρησιμοποιήθηκαν στην παρούσα εργασία είναι τα Resnet50, VGG16, Densenet121. Η εκπαίδευση πραγματοποιήθηκε σε GPU Tesla K80 12GB και ο χρόνος εκπαίδευσης ήταν περίπου 1.5 ώρα. Όλα τα μοντέλα εκπαιδεύτηκαν για 10 εποχές.

6.1: *Statistics BEFORE and AFTER fair split,RAF-DB, Emotion Recognition*

Architecture and Hyperparameters for all models				
TASK	Dataset	Architecture	Learning rate	batch size
Emotion Recognition	RAF-DB	PR - De(512) - Dr(0.2) - De(512) - De(7)	0.0001	32
Emotion Recognition	AffectNet	PR - De(8)	0.0001	32
Valence & Arousal Estimation	AffectNet	PR - De(512) - De(2)	0.001	128
Valence & Arousal Estimation	RAF-DB	PR - De(2)	0.001	128
AUs Detection	EmotioNet	PR - De(11)	0.0001	32
AUs Detection	DISFA	PR - De(512) - De(11)	0.00001	32
AUs Detection	GFT	PR - De(512) - De(14)	0.0001	32
AUs Detection	RAF-AU	PR - De(512) - De(13)	0.001	32

*Note:* PR symbolizes the pretrained network, De symbolizes Dense layer (#units), Dr symbolizes Dropout layer (Dropout rate)

### 6.3 Προ-επεξεργασία εικόνων

Σε αυτή την ενότητα παρουσιάζεται η προ-επεξεργασία που εφαρμόστηκε στα δεδομένα εισόδου για τη σωστή εκπαίδευση των μοντέλων ανίχνευσης φυλής, ηλικίας και φύλου και των μοντέλων αναγνώρισης συναισθηματικής κατάστασης.

#### 6.3.1 Μοντέλα ανίχνευσης φυλής, ηλικίας και φύλου

Για την εργασία ανίχνευσης φυλής, ηλικίας και φύλου έγιναν δοκιμές τεσσάρων μοντέλων (1.Face Analysis PyTorch, 2.deepface, 3.FairFace, 4.FaceRec). Η προ-επεξεργασία των εικόνων έγινε αυτόματα από όλα τα μοντέλα, εκτός από το πρώτο. Έτσι στις εικόνες που στέλνονται στο πρώτο μοντέλο, face\_analysis\_pytorch, γίνεται κάποια προ επεξεργασία κατά την οποία εφαρμόζεται ο ανιχνευτής προσώπου MTCNN[46]. Το δεύτερο μοντέλο, deepface, συμπεριλαμβάνει την ανίχνευση και ευθυγράμμιση προσώπου και έχει τη δυνατότητα επιλογής κάποιου ανιχνευτή προσώπου (OpenCV, SSD, Dlib, MTCNN, RetinaFace, MediaPipe). Η επιλογή που έγινε είναι ο ανιχνευτής προσώπου RetinaFace ο οποίος εξάγει καλύτερα αποτελέσματα. Στο τρίτο μοντέλο, FairFace, εφαρμόζεται αυτόματα ο ανιχνευτής προσώπου του Dlib που βασίζεται σε CNN. Τέλος, το τέταρτο μοντέλο, FaceRec, περιλαμβάνει ανίχνευση προσώπου και ευθυγράμμιση με χρήση dlib.

#### 6.3.2 Μοντέλα αναγνώρισης συναισθηματικής κατάστασης

Για τα μοντέλα αναγνώρισης συναισθηματικής κατάστασης η προ-επεξεργασία που εφαρμόζεται είναι η ανίχνευση προσώπου (face detection) και η ευθυγράμμιση προσώπου (face

alignment). Αρχικά γίνεται η ανίχνευση του προσώπου από την εικόνα. Συγκεκριμένα εντοπίζεται ένα πλαίσιο οριοθέτησης (bounding box) του προσώπου, από τη θέση τεσσάρων pixels. Έπειτα γίνεται η ευθυγράμμιση προσώπου κατά την οποία εφαρμόζονται κάποιοι μετασχηματισμοί (π.χ. περιστροφή) στην εικόνα έτσι ώστε το πρόσωπο να εμφανίζεται πάντα σε μία συγκεκριμένη θέση (σχήμα 6.1). Η ανίχνευση και η ευθυγράμμιση του προσώπου καθιστά πιο εύκολη τη διαδικασία εκμάθησης και τη γενίκευση από το νευρωνικό δίκτυο. Έτσι οι εικόνες ή τα καρέ των βίντεο της κάθε βάσης δεδομένων εισέρχονται σε κατάλληλο ανιχνευτή και ευθυγραμμιστή προσώπου και έπειτα εισχωρούν στο μοντέλο (σχήμα 6.2). Πριν τροφοδοτηθούν τα δεδομένα στο μοντέλο, προσαρμόστηκαν οι τιμές της εικόνας από 0-255 σε 0-1 για να μειωθεί το εύρος των τιμών των βαρών.

Εικόνα 6.1: *Example of face detection and alignment [3]*



Εικόνα 6.2: *Face detection and alignment of GFT database (video: 004C, frame: 15)*



Τέλος, έγιναν πειράματα με την εκπαίδευση των δικτύων σε κάθε βάση δεδομένων με και χωρίς αύξηση δεδομένων (data augmentation). Η επαύξηση δεδομένων αποτελεί μια τεχνική της ανάλυσης δεδομένων κατά την οποία αυξάνεται ο όγκος δεδομένων με την προσθήκη ελαφρώς τροποποιημένων αντιγράφων ήδη υπάρχοντων δεδομένων. Λόγω του γεγονότος ότι οι βάσεις δεδομένων απεικονίζουν πρόσωπα, εφαρμόστηκαν αυξήσεις αναστροφής και φωτεινότητας στα σύνολα δεδομένων, όπως φαίνεται στις εικόνες 6.3. Για το εύρος της εναλλαγής στην φωτεινότητα επιλέχθηκε το  $[0.2,1]$ , έτσι οι εικόνες θα σκουρύνουν.

Εικόνα 6.3: *Data augmentation examples: Brightness and Flipping, RAF-DB database*







## Κεφάλαιο **7**

### Αποτελέσματα

---

**Σ**το κεφάλαιο αυτό παρουσιάζονται όλα τα αποτελέσματα που παράχθηκαν κατά τη διεκπεραίωση της παρούσας εργασίας. Αρχικά προβάλλονται τα αποτελέσματα της αξιολόγησης των μοντέλων ανίχνευσης της φυλής, της ηλικίας και του φύλου. Έπειτα παρουσιάζονται τα αποτελέσματα των δοκιμών βελτιστοποίησης του μοντέλου αναγνώρισης βασικών συναισθημάτων με χρήση της βάσης δεδομένων RAF-DB. Στην τελευταία ενότητα του κεφαλαίου αυτού παρουσιάζονται τα κύρια αποτελέσματα της εργασίας, δηλαδή η απόδοση όλων των εργασιών αναγνώρισης συναισθήματος με τις αντίστοιχες βάσεις δεδομένων, πριν και μετά το δίκαιο χωρισμό.

#### 7.1 Αποτελέσματα μοντέλων ανίχνευσης φυλής, ηλικίας και φύλου

Για τη δημιουργία αυτόματων επισημάνσεων στις κατηγορίες φυλής, ηλικία και φύλου αξιολογήθηκαν τα μοντέλα που αναλύονται στην ενότητα 4.1. Μετά την εφαρμογή αυτών των μοντέλων ανίχνευσης σε υποσύνολα των βάσεων δεδομένων, παρατηρήθηκε ότι μόνο το τρίτο (FairFace) και το τέταρτο (FaceRec) μοντέλο εξάγουν αξιοπρεπή αποτελέσματα. Από τις βάσεις δεδομένων που χρησιμοποιήθηκαν σε αυτήν την εργασία, οι RAF-DB και GFT παρέχουν έγκυρους χειροκίνητους σχολιασμούς για τις κατηγορίες της φυλής, της ηλικίας και του φύλου. Επομένως, αξιολογήθηκε η απόδοση των δύο προαναφερθέντων μοντέλων με αυτές τις βάσεις δεδομένων χρησιμοποιώντας τις ακόλουθες μετρήσεις: ταξινόμηση ηλικίας, φυλής και φύλου: f1 score, παλινδρόμηση ηλικίας: συντελεστής συσχέτισης (Correlation Coefficient). Προσαρμόστηκαν επίσης οι εξόδοι του μοντέλου και οι ετικέτες της βάσης δεδομένων ώστε να υπάρχει συνέπεια μεταξύ τους. Τα αποτελέσματα του Πίνακα 7.1 φανερώνουν ότι το μοντέλο FairFace έχει καλύτερη απόδοση από το μοντέλο FaceRec σε όλες τις εργασίες. Όπως προαναφέρθηκε στην υποενότητα 4.1, το μοντέλο FairFace έχει δύο εξόδους φυλής, μία με τέσσερις και μία με επτά φυλές. Επιλέχθηκε η έξοδος με τις τέσσερις κατηγορίες φυλής καθώς οι πρόσθετες κατηγορίες της εξόδου με τις επτά κατηγορίες, επιδεινώνουν την απόδοση του μοντέλου. Έτσι, επιλέχθηκε το FairFace ως το μοντέλο από το οποίο θα ληφθούν οι αυτόματοι σχολιασμοί για τις κατηγορίες φυλή, ηλικία και φύλο στις βάσεις δεδομένων.

Μοντέλο	Φυλή		Ηλικία		Φύλο	
	RAF-DB	GFT	RAF-DB	GFT	RAF-DB	GFT
FairFace	0.8 (f1)	0.52 (f1)	0.64 (f1)	0.15 (f1)	0.89 (f1)	0.97 (f1)
FaceRec	/	/	0.55 (f1)	0.106 (CC)	0.87 (f1)	0.97 (f1)

Πίνακας 7.1: Απόδοση των μοντέλων *FairFace* και *FaceRec* στην αναγνώριση φυλής, ηλικία και φύλου

## 7.2 Αποτελέσματα βελτιστοποίησης βάσης RAF-DB

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα της βελτιστοποίησης του μοντέλου αναγνώρισης βασικών συναισθημάτων με τη βάση δεδομένων RAF-DB, η οποία έγινε με τη μεθοδολογία που εξηγήθηκε στην υποενότητα 6.1.

Με βάση τους πίνακες 7.2, 7.3, 7.4, 7.5 την καλύτερη απόδοση του μοντέλου αναγνώρισης βασικών συναισθημάτων με χρήση της βάσης δεδομένων RAF-DB επιφέρουν τα αποτελέσματα που φαίνονται στον πίνακα 7.6.

7.2: *Optimization for optimizer (Adam, SGD) and learning rate (0.0001 - 0.1)*

RAF-DB, Optimizer & lr optimization		
Adam optimizer		
learning rate	Accuracy	f1 score
0.1	0.4381	0.4348
0.001	0.7689	0.768
<b>0.0001</b>	0.7784	0.7803
0.0005	0.7699	0.7718
SGD optimizer		
learning rate	Accuracy	f1 score
0.01	0.7516	0.7522
0.001	0.6496	0.6345
0.0001	0.4726	0.3859

7.3: *Optimization for batch size (16,32,64)*

RAF-DB, batch size optimization		
batch size	Accuracy	f1 score
16	0.7386	0.7391
<b>32</b>	0.7842	0.7838
64	0.7588	0.7628

7.4: *Optimization for Architecture (number of layers) and number of units(512, 1024)*

RAF-DB, architecture and #units optimization		
units	Accuracy	f1 score
512	0.7735	0.7746
1024	0.7601	0.7598
<b>512,512</b>	0.7823	0.7866
512, 1024	0.7621	0.764
512,512,512	0.7383	0.7412

7.5: *Optimization for activation function (relu, sigmoid) and DROPOUT layer*

RAF-DB, activation function optimization		
activation function	Accuracy	f1 score
<b>relu, relu</b>	0.7748	0.7759
relu, sigmoid	0.7451	0.7449
<b>relu, dropout,relu</b>	0.779	0.7828

7.6: *Best hyperparameters for RAF-DB, Emotion Recognition*

RAF-DB,Emotion Recognition	
Optimizer	Adam
Learning Rate	0.0001
Batch size	32
extra layers	Dense Dropout Dense
# units of Dense layers	512
Activation function of Dense layers	ReLU

## 7.3 Αποτελέσματα δίκαιου διαχωρισμού βάσεων

Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα των μοντέλων αναγνώρισης βασικών συναισθημάτων, εκτίμησης των τιμών valence & arousal και ανίχνευσης AUs σε όλες τις βάσεις δεδομένων που μελετήθηκαν για τις τρεις εργασίες, πριν και μετά τη δίκαιη ανακατασκευή τους στα τρία σύνολα (εκπαίδευσης, επικύρωσης, ελέγχου). Φαίνονται τα αποτελέσματα με τη χρήση και χωρίς τη χρήση επαύξησης δεδομένων (data augmentation). Στα μοντέλα που υλοποιήθηκαν για λήψη των αποτελεσμάτων αυτών χρησιμοποιήθηκε το προ-εκπαιδευμένο δίκτυο ResNet50.

### 7.3.1 Πίνακες αποτελεσμάτων

7.7: *BEFORE* and *AFTER* fair split results for RAF-DB database without data augmentation, Emotion Recognition task in all categories

RAF-DB, Emotion Recognition, NO data augmentation											
SPLIT	BEFORE					AFTER					
macro f1 score	0.705					0.648					
averaged accuracy	0.705					0.646					
RACE	Caucasian	African-American			Asian	Caucasian	African-American		Asian		
#images	2351	234			483	3610	415		781		
macro f1 score	0.703	0.694			0.699	0.654	0.671		0.596		
averaged accuracy	0.701	0.696			0.717	0.651	0.688		0.597		
AGE	0-3	4-19	20-39	40-69	70+	0-3	4-19	20-39	40-69	70+	
#images	329	486	1662	502	89	528	844	2501	766	167	
macro f1 score	0.550	0.653	0.711	0.651	0.415	0.604	0.560	0.649	0.644	0.485	
averaged accuracy	0.546	0.654	0.701	0.659	0.408	0.572	0.557	0.645	0.657	0.516	
GENDER	Male		Female			Unsure	Male		Female		Unsure
#images	1249		1620			199	1949		2538		319
macro f1 score	0.687		0.703			0.567	0.642		0.633		0.573
averaged accuracy	0.689		0.695			0.581	0.652		0.622		0.570

7.8: *BEFORE* and *AFTER* fair split results for RAF-DB database with data augmentation, Emotion Recognition task in all categories

RAF-DB, Emotion Recognition, WITH data augmentation											
SPLIT	BEFORE					AFTER					
macro f1 score	0.737					0.669					
averaged accuracy	0.731					0.663					
RACE	Caucasian	African-American			Asian	Caucasian	African-American		Asian		
#images	2351	234			483	3610	415		781		
macro f1 score	0.742	0.757			0.678	0.673	0.643		0.641		
averaged accuracy	0.733	0.802			0.677	0.665	0.664		0.635		
AGE	0-3	4-19	20-39	40-69	70+	0-3	4-19	20-39	40-69	70+	
#images	329	486	1662	502	89	528	844	2501	766	167	
macro f1 score	0.539	0.699	0.737	0.684	0.507	0.611	0.575	0.671	0.665	0.473	
averaged accuracy	0.546	0.694	0.722	0.680	0.519	0.639	0.594	0.658	0.667	0.538	
GENDER	Male		Female			Unsure	Male		Female		Unsure
#images	1249		1620			199	1949		2538		319
macro f1 score	0.715		0.752			0.564	0.650		0.671		0.595
averaged accuracy	0.715		0.732			0.590	0.655		0.655		0.628

7.9: *BEFORE* and *AFTER* fair split results for AffectNet database without data augmentation, Emotion Recognition task in all categories

AffectNet, Emotion Recognition, NO data augmentation																		
SPLIT	BEFORE									AFTER								
macro f1 score	0.475									0.504								
averaged accuracy	0.502									0.491								
RACE	White			Black		Asian		Indian		White			Black		Asian		Indian	
#images	3120			321		311		234		67076			7308		7884		5343	
macro f1 score	0.481			0.436		0.413		0.499		0.507			0.502		0.481		0.493	
averaged accuracy	0.509			0.472		0.442		0.517		0.494			0.484		0.465		0.481	
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
#images	288	257	177	1464	821	427	363	162	27	4742	5465	4645	35861	17219	8658	7050	2910	1061
macro f1 score	0.462	0.536	0.468	0.473	0.464	0.472	0.409	0.367	0.363	0.527	0.528	0.508	0.498	0.514	0.479	0.459	0.412	0.417
averaged accuracy	0.509	0.552	0.488	0.488	0.500	0.506	0.454	0.414	0.385	0.522	0.513	0.487	0.486	0.500	0.459	0.441	0.401	0.399
GENDER	Male				Female					Male					Female			
#images	2192				1794					43480					44131			
macro f1 score	0.463				0.475					0.499					0.506			
averaged accuracy	0.498				0.494					0.483					0.494			

7.10: BEFORE and AFTER fair split results for AffectNet database with data augmentation, Emotion Recognition task in all categories

AffectNet, Emotion Recognition, WITH data augmentation																								
SPLIT	BEFORE									AFTER														
macro f1 score	0.467									0.509														
averaged accuracy	0.496									0.509														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#images	3120			321			311			234			67076			7308			7884			5343		
macro f1 score	0.471			0.451			0.395			0.485			0.509			0.511			0.501			0.505		
averaged accuracy	0.502			0.484			0.423			0.501			0.509			0.502			0.499			0.503		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#images	288	257	177	1464	821	427	363	162	27	4742	5465	4645	35861	17219	8658	7050	2910	1061						
macro f1 score	0.439	0.510	0.432	0.455	0.471	0.470	0.418	0.308	0.441	0.524	0.529	0.497	0.509	0.521	0.485	0.461	0.398	0.440						
averaged accuracy	0.482	0.532	0.443	0.475	0.502	0.500	0.463	0.379	0.438	0.524	0.522	0.493	0.509	0.523	0.481	0.458	0.403	0.440						
GENDER	Male				Female					Male				Female										
#images	2192				1794					43480				44131										
macro f1 score	0.451				0.471					0.499				0.514										
averaged accuracy	0.491				0.491					0.497				0.514										

7.11: BEFORE and AFTER fair split results for AffectNet database without data augmentation, Valence and Arousal estimation task in all categories

AffectNet, Valence & Arousal Estimation , NO data augmentation																								
SPLIT	BEFORE									AFTER														
Valence Correlation	0.485									0.796														
Arousal Correlation	0.444									0.564														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#images	3120			321			311			234			68301			8098			8793			6105		
Valence Correlation	0.475			0.537			0.556			0.432			0.795			0.809			0.792			0.792		
Arousal Correlation	0.443			0.464			0.446			0.417			0.572			0.550			0.523			0.545		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#images	288	257	177	1464	821	427	363	162	27	5116	5905	5049	36404	17732	9097	7441	3243	1310						
Valence Correlation	0.703	0.669	0.419	0.442	0.466	0.432	0.413	0.367	0.434	0.849	0.845	0.794	0.783	0.803	0.790	0.759	0.732	0.721						
Arousal Correlation	0.491	0.531	0.325	0.449	0.478	0.424	0.321	0.343	0.101	0.621	0.635	0.514	0.558	0.576	0.533	0.508	0.496	0.483						
GENDER	Male				Female					Male				Female										
#images	2192				1794					45461				45836										
Valence Correlation	0.446				0.502					0.784				0.795										
Arousal Correlation	0.434				0.451					0.559				0.567										

7.12: BEFORE and AFTER fair split results for AffectNet database with data augmentation, Valence and Arousal estimation task in all categories

AffectNet, Valence & Arousal Estimation , WITH data augmentation																								
SPLIT	BEFORE									AFTER														
Valence Correlation	0.493									0.802														
Arousal Correlation	0.457									0.585														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#images	3120			321			311			234			68301			8098			8793			6105		
Valence Correlation	0.480			0.588			0.569			0.432			0.801			0.814			0.803			0.793		
Arousal Correlation	0.458			0.494			0.420			0.427			0.593			0.573			0.545			0.566		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#images	288	257	177	1464	821	427	363	162	27	5116	5905	5049	36404	17732	9097	7441	3243	1310						
Valence Correlation	0.686	0.683	0.458	0.449	0.487	0.429	0.414	0.356	0.525	0.850	0.850	0.798	0.789	0.806	0.794	0.771	0.734	0.742						
Arousal Correlation	0.510	0.518	0.337	0.451	0.501	0.449	0.363	0.308	0.261	0.629	0.654	0.533	0.577	0.596	0.564	0.533	0.528	0.497						
GENDER	Male				Female					Male				Female										
#images	2192				1794					45461				45836										
Valence Correlation	0.455				0.512					0.789				0.801										
Arousal Correlation	0.452				0.455					0.581				0.586										

7.13: BEFORE and AFTER fair split results for AFEW-VA database without data augmentation, Valence and Arousal estimation task in all categories

AFEW-VA, Valence & Arousal Estimation , NO data augmentation																								
SPLIT	BEFORE									AFTER														
Valence Correlation	0.196									0.021														
Arousal Correlation	0.189									0.124														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#video frames	3456			nan			177			75			3446			285			329			51		
Valence Correlation	0.174			nan			0.343			-0.122			0.033			-0.043			0.002			0.000		
Arousal Correlation	0.183			0.nan			0.241			0.054			0.131			0.415			-0.002			-0.793		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#video frames	nan	nan	297	1385	1099	500	357	70	nan	61	77	205	1509	1181	544	356	119	59						
Valence Correlation	nan	nan	0.431	0.157	0.123	0.249	0.280	-0.029	nan	-0.002	0.002	0.341	0.026	-0.048	0.081	0.133	0.030	-0.027						
Arousal Correlation	nan	nan	0.011	0.179	0.269	0.068	0.227	-0.028	nan	0.421	0.077	-0.022	0.082	0.114	0.085	0.225	-0.505	-0.166						
GENDER	Male				Female					Male				Female										
#video frames	2439				1269					2458				1653										
Valence Correlation	0.214				0.170					0.009				0.046										
Arousal Correlation	0.148				0.238					0.097				0.158										

7.14: BEFORE and AFTER fair split results for AFEW-VA database with data augmentation, Valence and Arousal estimation task in all categories

AFEW-VA, Valence & Arousal Estimation , WITH data augmentation																								
SPLIT	BEFORE									AFTER														
Valence Correlation	0.276									0.112														
Arousal Correlation	0.213									0.182														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#video frames	3456			nan			177			75			3446			285			329			51		
Valence Correlation	0.232			nan			0.555			0.090			0.103			0.047			0.217			0.000		
Arousal Correlation	0.198			nan			0.424			-0.404			0.158			0.432			0.213			0.031		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#video frames	nan	nan	297	1385	1099	500	357	70	nan	61	77	205	1509	1181	544	356	119	59						
Valence Correlation	nan	nan	0.277	0.258	0.109	0.534	0.322	0.083	nan	0.024	0.007	0.291	0.064	0.111	0.211	0.211	0.034	-0.209						
Arousal Correlation	nan	nan	0.033	0.228	0.256	0.141	0.167	0.251	nan	0.034	0.031	0.219	0.253	0.085	0.195	-0.024	0.293	-0.413						
GENDER	Male				Female					Male				Female										
#video frames	2439				1269					2458				1653										
Valence Correlation	0.260				0.298					0.095				0.149										
Arousal Correlation	0.166				0.253					0.148				0.235										

7.15: BEFORE and AFTER fair split results for EmotioNet database without data augmentation, AUs detection task in all categories

EmotioNet,AUs detection , NO data augmentation																								
SPLIT	BEFORE									AFTER														
mean binary fl score	0.236									0.237														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#images	16132			1935			1043			1210			10790			1191			754			814		
mean binary fl score	0.236			0.242			0.231			0.233			0.237			0.239			0.228			0.228		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#images	262	488	754	8912	4206	2581	2093	834	190	262	435	549	5710	2777	1721	1381	571	143						
mean binary fl score	0.197	0.234	0.231	0.236	0.239	0.239	0.236	0.224	0.234	0.199	0.238	0.230	0.235	0.241	0.240	0.237	0.231	0.240						
GENDER	Male				Female					Male				Female										
#images	10188				10132					6958				6591										
mean binary fl score	0.232				0.239					0.233				0.240										

7.16: BEFORE and AFTER fair split results for EmotioNet database with data augmentation, AUs detection task in all categories

EmotioNet,AUs detection , WITH data augmentation																								
SPLIT	BEFORE									AFTER														
mean binary fl score	0.232									0.234														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#images	16132			1935			1043			1210			10790			1191			754			814		
mean binary fl score	0.232			0.232			0.225			0.226			0.235			0.235			0.230			0.228		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#images	262	488	754	8912	4206	2581	2093	834	190	262	435	549	5710	2777	1721	1381	571	143						
mean binary fl score	0.171	0.228	0.227	0.230	0.235	0.237	0.229	0.228	0.225	0.176	0.235	0.227	0.233	0.238	0.238	0.235	0.229	0.228						
GENDER	Male									Female														
#images	10188									10132														
mean binary fl score	0.226									0.235														
GENDER	Male									Female														
#images	6958									6591														
mean binary fl score	0.230									0.237														

7.17: BEFORE and AFTER fair split results for DISFA database without data augmentation, AUs detection task in all categories

DISFA, AUs detection , NO data augmentation																								
SPLIT	BEFORE									AFTER														
mean binary fl score	0.450									0.377														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#video frames	38760			0			9690			0			45766			9690			19296			9690		
mean binary fl score	0.488			0			0.331			0			0.379			0.227			0.309			0.482		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#video frames	0	0	0	38760	0	9690	0	0	0	0	0	9690	48413	19333	7006	0	0	0						
mean binary fl score	0	0	0	0.401	0	0.450	0	0	0	0	0	0.272	0.355	0.433	0.217	0	0	0						
GENDER	Male									Female														
#video frames	29070									19380														
mean binary fl score	0.404									0.441														
GENDER	Male									Female														
#video frames	45719									38723														
mean binary fl score	0.365									0.382														

7.18: BEFORE and AFTER fair split results for DISFA database with data augmentation, AUs detection task in all categories

DISFA, AUs detection , WITH data augmentation																								
SPLIT	BEFORE									AFTER														
mean binary fl score	0.429									0.400														
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian		
#video frames	38760			0			9690			0			45766			9690			19296			9690		
mean binary fl score	0.490			0			0.294			0			0.415			0.267			0.349			0.506		
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+						
#video frames	0	0	0	38760	0	9690	0	0	0	0	0	9690	48413	19333	7006	0	0	0						
mean binary fl score	0	0	0	0.400	0	0.469	0	0	0	0	0	0.385	0.354	0.474	0.345	0	0	0						
GENDER	Male									Female														
#video frames	29070									19380														
mean binary fl score	0.390									0.430														
GENDER	Male									Female														
#video frames	45719									38723														
mean binary fl score	0.418									0.380														

7.19: BEFORE and AFTER fair split results for GFT database without data augmentation, AUs detection task in all categories

GFT, AUs detection , NO data augmentation														
SPLIT	BEFORE										AFTER			
mean binary fl score	0.518										0.467			
RACE	White	Black	Hispanic	Asian	Other	White	Black	Hispanic	Asian	Other				
#video frames	18505	4891	0	0	1249	31407	3830	764	2641	1249				
mean binary fl score	0.488	0.526	0	0	0.412	0.447	0.412	0.247	0.492	0.415				
AGE	21-22		23-25			26-28		21-22		23-25		26-28		
#video frames	20267		2165			2213		26801		6034		7056		
mean binary fl score	0.514		0.458			0.494		0.474		0.459		0.408		
GENDER	Male					Female								
#video frames	13525					11120								
mean binary fl score	0.448					0.552								
GENDER	Male					Female								
#video frames	23265					16626								
mean binary fl score	0.416					0.492								

7.20: BEFORE and AFTER fair split results for GFT database with data augmentation, AUs detection task in all categories

GFT, AUs detection , WITH data augmentation															
SPLIT	BEFORE					AFTER									
mean binary fl score	0.525					0.486									
RACE	White	Black	Hispanic	Asian	Other	White	Black	Hispanic	Asian	Other					
#video frames	18505	4891	0	0	1249	31407	3830	764	2641	1249					
mean binary fl score	0.479	0.528	0	0	0.350	0.466	0.436	0.244	0.535	0.443					
AGE	21-22		23-25			26-28		21-22		23-25		26-28			
#video frames	20267		2165			2213		26801		6034		7056			
mean binary fl score	0.525		0.411			0.468		0.490		0.491		0.433			
GENDER	Male					Female									
#video frames	13525					11120									
mean binary fl score	0.438					0.535									
						0.433					0.514				

7.21: BEFORE and AFTER fair split results for RAF-AU database without data augmentation, AUs detection task in all categories

RAF-AU, AUs detection , NO data augmentation																											
SPLIT	BEFORE									AFTER																	
mean binary fl score	0.665									0.537																	
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian					
#images	1051			95			137			79			1086			96			152			92					
mean binary fl score	0.670			0.671			0.663			0.513			0.541			0.515			0.553			0.428					
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+									
#images	107	177	75	488	313	121	58	10	13	113	184	75	504	328	124	55	23	20									
mean binary fl score	0.558	0.648	0.556	0.674	0.672	0.669	0.720	0.717	0.442	0.404	0.490	0.434	0.546	0.583	0.543	0.491	0.480	0.314									
GENDER	Male									Female																	
#images	689									673																	
mean binary fl score	0.681									0.642																	
										0.546									0.526								

7.22: BEFORE and AFTER fair split results for RAF-AU database with data augmentation, AUs detection task in all categories

RAF-AU, AUs detection , WITH data augmentation																											
SPLIT	BEFORE									AFTER																	
mean binary fl score	0.566									0.596																	
RACE	White			Black			Asian			Indian			White			Black			Asian			Indian					
#images	1051			95			137			79			1086			96			152			92					
mean binary fl score	0.566			0.529			0.618			0.481			0.596			0.570			0.616			0.535					
AGE	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+	0-2	3-9	10-19	20-29	30-39	40-49	50-59	60-69	70+									
#images	107	177	75	488	313	121	58	10	13	113	184	75	504	328	124	55	23	20									
mean binary fl score	0.506	0.523	0.487	0.576	0.587	0.535	0.545	0.361	0.437	0.434	0.504	0.494	0.612	0.635	0.595	0.586	0.597	0.553									
GENDER	Male									Female																	
#images	689									673																	
mean binary fl score	0.568									0.562																	
										0.604									0.585								



### 7.3.2 Σχολιασμοί

Από τα αποτελέσματα της απόδοσης των βάσεων δεδομένων πριν και μετά το δίκαιο χωρισμό για κάθε κατηγορία διατυπώνονται οι ακόλουθοι σχολιασμοί.

1. **Βάση δεδομένων: RAF-DB, Εργασία: Αναγνώριση βασικού συναισθήματος** (Πίνακες 7.7, 7.8)

Σε αυτή την περίπτωση παρατηρείται ότι οι αποδόσεις για κάθε κατηγορία ξεχωριστά είναι κοντινότερες πριν αλλά κυρίως μετά το δίκαιο διαχωρισμό. Για αυτή τη βάση δεδομένων τα δεδομένα στα σύνολα εκπαίδευσης και ελέγχου ήταν, ως επί το πλείστον, ισορροπημένα στις κατηγορίες συναισθημάτων, φύλου, ηλικίας και φυλής πριν το δίκαιο χωρισμό. Γι' αυτό και τα αποτελέσματα πριν το δίκαιο χωρισμό έχουν κοντινές τιμές για τις περισσότερες κατηγορίες. Η συνολική απόδοση του μοντέλου μειώθηκε κατά λίγο μετά το δίκαιο χωρισμό. Με την εισαγωγή επαύξησης δεδομένων παρατηρούνται καλύτερα συνολικά αποτελέσματα τόσο πριν, όσο και μετά το δίκαιο χωρισμό. Οι αποδόσεις για κάθε κατηγορία συνεχίζουν να είναι κοντινές.

2. **Βάση δεδομένων: AffectNet, Εργασία: Αναγνώριση βασικού συναισθήματος** (Πίνακες 7.9, 7.10)

Με τη συγκεκριμένη εργασία και βάση δεδομένων παρατηρείται μια μικρή βελτίωση συνολικών αποτελεσμάτων μετά το δίκαιο χωρισμό, με και χωρίς επαύξηση δεδομένων. Η βάση AffectNet είχε εντελώς διαφορετικές κατανομές δεδομένων στις κατηγορίες συναισθημάτων, φύλου, ηλικίας, φυλής και πολύ λιγότερα δεδομένα στο σύνολο ελέγχου, πριν το δίκαιο χωρισμό. Έτσι τα αποτελέσματα μετά το δίκαιο διαχωρισμό είναι κοντινότερα στις κατηγορίες και πιο αντιπροσωπευτικά για κάθε κατηγορία.

3. **Βάση δεδομένων: AffectNet, Εργασία: Εκτίμηση τιμών valence & arousal** (Πίνακες 7.11, 7.12)

Σε αυτό το μοντέλο παρατηρείται σημαντική βελτίωση της απόδοσης μετά τη δίκαιη αναδιάρθρωση των βάσεων στα τρία σύνολα. Επίσης τα αποτελέσματα για κάθε κατηγορία φύλου, ηλικίας και φυλής έχουν πολύ κοντινές τιμές μετά το δίκαιο διαχωρισμό, σε σύγκριση με πριν. Επίσης, τα δεδομένα του συνόλου ελέγχου μετά την εφαρμογή του δίκαιου χωρισμού είναι περισσότερα, άρα τα αποτελέσματα απόδοσης κάθε κατηγορίας είναι πιο έγκυρα. Μετά την εφαρμογή του data augmentation τα αποτελέσματα φαίνονται ελαφρώς καλύτερα.

4. **Βάση δεδομένων: AFEW-VA, Εργασία: Εκτίμηση τιμών valence & arousal** (Πίνακες 7.13, 7.14)

Σε αυτή την περίπτωση το μοντέλο δεν εξάγει καλά αποτελέσματα. Μετά το δίκαιο χωρισμό της βάσης υπάρχουν δείγματα για κάθε κατηγορία φυλής, ηλικία και φύλου. Όμως τα αποτελέσματα για κάθε κατηγορία δεν είναι αντιπροσωπευτικά, αφού υπάρχει ελλιπής αριθμός δειγμάτων στο σύνολο ελέγχου για κάθε κατηγορία. Μετά την εφαρμογή της επαύξησης δεδομένων τα αποτελέσματα είναι λίγο καλύτερα.

5. **Βάση δεδομένων: EmotioNet, Εργασία: Ανίχνευση τιμών AUs** (Πίνακες 7.15, 7.16)

Σε αυτό το μοντέλο με χρήση της βάσης δεδομένων EmotioNet τα αποτελέσματα είναι σχεδόν τα ίδια για κάθε κατηγορία, με και χωρίς εφαρμογή του data augmentation. Γενικά όμως το μοντέλο αυτό δεν αποδίδει καλά.

6. **Βάση δεδομένων: DISFA, Εργασία: Ανίχνευση τιμών AUs** (Πίνακες 7.17, 7.18)

Η οπτικοακουστική βάση δεδομένων DISFA δεν έχει καθόλου βίντεο από άτομα ηλικιών 0-9 και 50-70+. Πριν το δίκαιο χωρισμό της βάσης, στο σύνολο ελέγχου δεν υπάρχουν δεδομένα από πολλές κατηγορίες. Μετά το δίκαιο χωρισμό, υπάρχουν δεδομένα για όλες τις δυνατές κατηγορίες (όλες εκτός των προαναφερόμενων ηλικιών). Η συνολική απόδοση του μοντέλου μειώνεται μετά το δίκαιο διαχωρισμό, αλλά είναι πιο αντιπροσωπευτική, αφού παρέχει δεδομένα για περισσότερες κατηγορίες ατόμων. Η επαύξηση δεδομένων χειροτερεύει ελαφρώς τα αποτελέσματα πριν το δίκαιο χωρισμό και τα βελτιώνει ελαφρώς μετά το δίκαιο χωρισμό.

7. **Βάση δεδομένων: GFT, Εργασία: Ανίχνευση τιμών AUs** (Πίνακες 7.19, 7.20)

Για την περίπτωση αυτή πριν το δίκαιο χωρισμό της βάσης, στο σύνολο ελέγχου δεν υπάρχουν δεδομένα από κάποιες κατηγορίες φυλής. Μετά το δίκαιο χωρισμό, υπάρχουν δεδομένα για όλες τις κατηγορίες. Η συνολική απόδοση του μοντέλου μειώνεται μετά το δίκαιο διαχωρισμό, αλλά είναι πιο αντιπροσωπευτική, αφού παρέχει δεδομένα για περισσότερες κατηγορίες ατόμων. Η επαύξηση δεδομένων βελτιώνει ελαφρώς τα αποτελέσματα πριν και μετά το δίκαιο χωρισμό.

8. **Βάση δεδομένων: RAF-AU, Εργασία: Ανίχνευση τιμών AUs** (Πίνακες 7.21, 7.22)

Σε αυτό το μοντέλο η συνολική απόδοση είναι χειρότερη μετά την εφαρμογή του δίκαιου χωρισμού, αλλά οι τιμές κάθε κατηγορίας είναι λίγο πιο κοντινές μεταξύ τους. Μετά το data augmentation η συνολική απόδοση πριν το δίκαιο διαχωρισμό είναι χειρότερη από αυτήν μετά το δίκαιο διαχωρισμό.

Μέρος **III**

Επίλογος

---



## Κεφάλαιο 8

### Επίλογος

---

Η εργασία αναγνώρισης συναισθηματικών εκφράσεων αποτελεί ένα σημαντικά αναπτυσσόμενο τομέα της μηχανικής μάθησης, με πλήθος εφαρμογών και ερευνών να τον πλαισιώνουν. Πρόκειται για ένα τομέα άμεσα συνδεδεμένο με τον άνθρωπο, γι' αυτό και επιβάλλεται η σημασία στα χαρακτηριστικά όπως η φυλή, η ηλικία και το φύλο κατά τη δημιουργία μοντέλων. Για το συγκεκριμένο τομέα και την ιδιαιτερότητα που έχει, για καλή απόδοση των μοντέλων που παράγονται πρέπει να λαμβάνονται στα υπόψη και κοινωνικοί παράγοντες, μαζί με την υλοποίηση της εργασίας συναισθηματικής έκφρασης. Οι πλείστες βάσεις δεδομένων που αντιμετωπίζουν τα ετερογενή προβλήματα του τομέα αυτού αποτελούνται από πλήθος δειγμάτων με έγκυρες επισημάνσεις και ικανοποιητικό αριθμό ατόμων διάφορων εθνικοτήτων, φύλων και ηλικιών. Ο διαμοιρασμός, όμως, των πλείστων βάσεων δεδομένων στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου είναι ελαττωματικός αφού δεν παρέχει μια αντιπροσωπευτική κατανομή για πρόσωπα κάθε εθνικότητας, φύλου και ηλικίας. Η αναδιάταξη των δεδομένων στα τρία σύνολα με έναν δίκαιο τρόπο φέρνει τα μοντέλα αναγνώρισης συναισθηματικών εκφράσεων ένα βήμα πιο κοντά στην απελευθέρωσή τους από μεροληπτικές συμπεριφορές και λήψεις άδικων αποφάσεων.

#### 8.1 Συμπεράσματα

Μετά το τέλος της παρούσας μελέτης, εξάγονται κάποια βασικά συμπεράσματα, τα οποία παρουσιάζονται παρακάτω.

- Ομοιομορφία στα σύνολα δεδομένων

Δημιουργούνται και τα τρία σύνολα, εκπαίδευσης, επικύρωσης και ελέγχου με λογικό αριθμό δειγμάτων, μετά το δίκαιο χωρισμό σε όλες τις βάσεις. Όλες οι βάσεις δεδομένων που μελετήθηκαν δεν είχαν διαχωρισμό και στα τρία σύνολα, κάτι το οποίο επιφέρει ο δίκαιος διαχωρισμός. Ταυτόχρονα κάποιες βάσεις είχαν προβληματικά ποσοστά διαχωρισμού (π.χ. η βάση δεδομένων EmotioNet περιέχει περίπου ίδιο αριθμό δεδομένων στο σύνολο εκπαίδευσης και ελέγχου). Επίσης κάποιες βάσεις δεδομένων με πολύ μικρό αριθμό δειγμάτων στο σύνολο ελέγχου (π.χ. AffectNet) επωφελούνται από το δίκαιο διαχωρισμό, αφού τα αποτελέσματα που εξάγονται είναι πιο αντιπροσωπευτικά για κάθε κατηγορία. Η εφαρμογή της δίκαιης ανακατανομής των δεδομένων προσφέρει μία ομοιομορφία αφού τα σύνολα δεδομένων χωρίζονται με συγκεκριμένα ποσοστά και στα τρία σύνολα.

- Κοντινότερα αποτελέσματα για κατηγορίες ηλικίας, εθνικότητας και φύλο  
Τα αποτελέσματα που λαμβάνονται μετά το δίκαιο χωρισμό των βάσεων είναι κοντινότερα για τις πλείστες κατηγορίες ηλικίας, εθνικότητας και φύλου. Το δίκαιο διαχωρισμό καταφέρει να εξισορροπήσει τις κατανομές των δεδομένων στα τρία σύνολα (εκπαίδευσης, επικύρωσης και ελέγχου), όσον αφορά τις κατηγορίες ηλικία, εθνικότητας και φύλο. Έτσι οι αποδόσεις για κάθε υποκατηγορία είναι πιο κοντινές. Αυτό δείχνει ότι μετά το δίκαιο χωρισμό τα μοντέλα γίνονται πιο δίκαια απέναντι σε όλες τις κατηγορίες (ακόμα και σε αυτές με λιγότερα δεδομένα).
- Καλύτερα κατανεμημένα δεδομένα ως προς τις επισημάνσεις εκφραστικού υπολογισμού  
Επίσης, ο δίκαιος διαχωρισμός προσφέρει καλύτερα κατανεμημένα δεδομένα όχι μόνο για ηλικία, εθνικότητα και φύλο αλλά και για τις επισημάνσεις κάθε εργασίας αναγνώρισης συναισθηματικής έκφρασης. Παρατηρείται ότι κάποιες βάσεις δεδομένων φέρουν καλύτερα αποτελέσματα μετά το δίκαιο διαχωρισμό, αφού είναι πιο ισορροπημένες στα τρία σύνολα για την εργασία που υλοποιούν.
- Καλύτερα αποτελέσματα μετά την επαύξηση δεδομένων  
Η επαύξηση δεδομένων (data augmentation) όταν εφαρμόστηκε, επιφέρει καλύτερα αποτελέσματα στις πλείστες βάσεις δεδομένων πριν αλλά κυρίως μετά το δίκαιο χωρισμό.

## 8.2 Μελλοντικές Επεκτάσεις

Η μελέτη αυτή μπορεί να θεωρηθεί η βάση για την ανάπτυξη διάφορων εργασιών. Κάποιες από αυτές προβάλλονται παρακάτω.

- Ένωση κάποιων βάσεων για περισσότερη ποικιλία χαρακτηριστικών  
Μία επέκταση της δουλειάς αυτής η οποία θα επιφέρει περισσότερη ποικιλομορφία στα σύνολα δεδομένων είναι η ένωση κάποιων βάσεων και έπειτα η εφαρμογή του δίκαιου χωρισμού στη νέα βάση δεδομένων. Με αυτό τον τρόπο η νέα βάση δεδομένων που θα δημιουργηθεί θα έχει περισσότερα χαρακτηριστικά ατόμων και έτσι θα εκπροσωπεί περισσότερες ομάδες ατόμων.
- Υλοποίηση πρωτοκόλλου και σε άλλες βάσεις δεδομένων.  
Η επεξεργασία του δίκαιου διαχωρισμού των βάσεων δεδομένων μπορεί να εφαρμοστεί και σε άλλες βάσεις δεδομένων όπως οι Aff-Wild [47, 48, 49, 50], Aff-Wild2 [51, 52, 53, 54, 55, 56, 57, 58], BP4DS [59], BP4D+ [60], DFEW [61], EMOTIC [62, 63].
- Εφαρμογή του χωρισμού εργασίας σε όλες τις εργασίες κατά το δίκαιο χωρισμό  
Η παρούσα εργασία καταφέρει να χωρίσει τα δεδομένα των βάσεων στα τρία σύνολα, λαμβάνοντας στα υπόψη τις κατηγορίες ηλικία, φυλής και φύλου και την εργασία αναγνώρισης συναισθηματικής έκφρασης που αντιμετωπίζουν. Για τις εργασίες αναγνώρισης βασικών συναισθημάτων και εκτίμησης τιμών valence & arousal, προτείνει έναν επιπλέον χωρισμό των βάσεων με βάση την ετικέτα τους (task split). Για την

εργασία ανίχνευσης των τιμών Action Units, δεν εφαρμόζεται κάποιος επιπλέον χωρισμός, απλά γίνεται έλεγχος έτσι ώστε τα ποσοστά των ενεργοποιημένων AUs να είναι πανομοιότυπα στα τρία σύνολα. Έτσι δεν υπάρχει μεγάλη ακρίβεια, όσον αφορά το task split για την εργασία αυτή. Μία επέκταση θα ήταν στις βάσεις δεδομένων που αφορούν AUs, να εφαρμοστεί κάποια μεθοδολογία για task split.

- Ανακατασκευή βάσεων με δεδομένα από όλες τις δυνατές κατηγορίες ατόμων

Η παρούσα εργασία προτείνει ιδέες που αφορούν τη δίκαιη αναγνώριση συναισθηματικών εκφράσεων προσώπου, με αναδιάταξη των βάσεων δεδομένων στα τρία σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Καταφέρνει, σε κάποιες βάσεις δεδομένων, να διαμορφώσει ένα αξιοπρεπές σύνολο που καλύπτει κάθε κατηγορία ατόμων. Αρκετές όμως βάσεις δεδομένων που μελετήθηκαν δεν παρέχουν μεγάλο εύρος στις κατηγορίες ηλικίας, εθνικότητας και φύλου με αποτέλεσμα να μην υπάρχει πληροφορία ή να υπάρχει ελάχιστη πληροφορία (π.χ. στη βάση δεδομένων GFT απεικονίζονται μόνο νέοι, στη βάση δεδομένων DISFA απεικονίζεται μόνο ένα πρόσωπο στη κατηγορία Black και Indian κ.α.). Για αυτές τις περιπτώσεις οι βάσεις δεδομένων είναι μεροληπτικές και άδικες. Για τη μετατροπή των βάσεων δεδομένων στη δίκαιη εκδοχή τους θα πρέπει να υφίσταται επαρκής αριθμός δειγμάτων για όλες τις κατηγορίες ατόμων. Έτσι μια επέκταση της εργασίας αυτή αποτελεί η ανακατασκευή των συγκεκριμένων βάσεων, με προσθήκη και άλλων δεδομένων, έτσι ώστε να παρέχουν ικανοποιητικό αριθμό δειγμάτων για όλες τις κατηγορίες.

- Βελτιστοποίηση των αποτελεσμάτων όλων των μοντέλων μετά τον δίκαιο χωρισμό

Στην υποενότητα 6.1, αναφέρεται η μεθοδολογία που ακολουθείται για βελτιστοποίηση της απόδοσης των μοντέλων για κάθε εργασία αναγνώρισης συναισθηματικών εκφράσεων προσώπου. Η μεθοδολογία αυτή εφαρμόστηκε ενδεικτικά μόνο για τη βάση δεδομένων RAF-DB. Έτσι τα αποτελέσματα για τα μοντέλα που παρουσιάζονται στο τελευταίο κεφάλαιο της εργασίας δεν είναι τα καλύτερα δυνατά. Μια επέκταση της δουλειάς αυτής θα ήταν η βελτιστοποίηση των αποτελεσμάτων μετά τον δίκαιο χωρισμό των βάσεων, για εξαγωγή των βελτιστοποιημένων αποτελεσμάτων της δίκαιης εκπαίδευσης των βάσεων.

- Λήψης αυτόματων επισημάνσεων από ένα μοντέλο με περισσότερες κατηγορίες εθνικότητας

Στην παρούσα εργασία για την εφαρμογή του δίκαιου χωρισμού στις βάσεις δεδομένων χρειάστηκε η λήψη αυτόματων επισημάνσεων για την ηλικία, τη φυλή και το φύλο στις βάσεις δεδομένων που δεν παρείχαν χειροκίνητες επισημάνσεις για τις τιμές αυτές. Για το σκοπό αυτό εφαρμόστηκε στις βάσεις δεδομένων το μοντέλο ανίχνευσης ηλικίας, φυλής και φύλου FairFace. Για τις κατηγορίες της φυλής προτιμήθηκε η επιλογή του μοντέλου που εξάγει τέσσερις κατηγορίες (White, Black, Indian, Asian) αντί επτά (White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, Latino), αφού παρατηρήθηκε ότι οι πρόσθετες κατηγορίες δεν ήταν τόσο αποτελεσματικές. Μια επέκταση θα μπορούσε να είναι η λήψη αυτόματων επισημάνσεων από ένα μοντέλο το οποίο περιέχει περισσότερες κατηγορίες φυλής και είναι αποτελεσματικό. Αυτό βέβαια θα προαπαιτούσε και οι βάσεις δεδομένων να έχουν έναν επαρκή αριθμό δειγμάτων για τις κατηγορίες αυτές.

- Χειροκίνητοι σχολιασμοί για τις κατηγορίες ηλικίας, εθνικότητας και φύλου

Η παρούσα εργασία προέβηκε σε μεθόδους αυτόματου σχολιασμού των βάσεων για τις κατηγορίες ηλικίας, εθνικότητας και φύλου. Οι σχολιασμοί αυτοί είναι επιρρεπείς σε λάθη και λιγότερο ακριβείς. Η εργασία αναγνώρισης συναισθηματικών εκφράσεων προσώπου, έχει συγκεκριμένη ιδιαιτερότητα και ευαισθησία και για τη διεκπεραίωση κάθε εργασίας πρέπει να λαμβάνονται στα υπόψη οι παράγοντες που αφορούν την ηλικία, την εθνικότητα και το φύλο. Έτσι, θα έπρεπε κάθε βάση δεδομένων που αντιμετωπίζει κάποια εργασία του τομέα αυτού να είναι χειροκίνητα σχολιασμένη ως προς την ηλικία, την εθνικότητα και το φύλο.



## Βιβλιογραφία

---

- [1] *Universal Emotions*. <https://www.paulekman.com/universal-emotions/>.
- [2] Ruicong Zhi, Mengyi Liu και Dezheng Zhang. *A comprehensive survey on automatic facial action unit analysis*. *The Visual Computer*, 36(5):1067–1093, 2019.
- [3] Sefik Ilkin Serengil. *Face Alignment for Face Recognition in Python within OpenCV*. <https://sefiks.com/2020/02/23/face-alignment-for-face-recognition-in-python-within-opencv/>, 2020.
- [4] Dimitrios Kollias. *Διδακτορική Διατριβή*.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren και Jian Sun. *Deep Residual Learning for Image Recognition*. *CoRR*, αβζ/1512.03385, 2015.
- [6] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. *arXiv 1409.1556*, 2014.
- [7] Gao Huang, Zhuang Liu και Kilian Q. Weinberger. *Densely Connected Convolutional Networks*. *CoRR*, αβζ/1608.06993, 2016.
- [8] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *AI-MIA: COVID-19 Detection & Severity Analysis through Medical Imaging*. *arXiv preprint arXiv:2206.04732*, 2022.
- [9] 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.
- [10] Miao Yu, Dimitrios Kollias, James Wingate, Niro Siriwardena και Stefanos Kollias. *Machine learning for predictive modelling of ambulance calls*. *Electronics*, 10(4):482, 2021.
- [11] 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.
- [12] 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.

- [13] 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.
- [14] 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.
- [15] Dimitris Kollias, George Marandianos, Amaryllis Raouzaiou 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Greg Tarr Vitor Albiero. *GitHub - vitoralbiero/face\_analysis\_pytorch : FaceanalysisinPyTorch*. 2022.
- [20] Sefik Ilkin Serengil και Alper Ozpinar. *LightFace: A Hybrid Deep Face Recognition Framework*. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, σελίδες 23–27. IEEE, 2020.
- [21] Sefik Ilkin Serengil και Alper Ozpinar. *HyperExtended LightFace: A Facial Attribute Analysis Framework*. *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, σελίδες 1–4. IEEE, 2021.
- [22] Kimmo Karkkainen και Jungseock Joo. *FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation*. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, σελίδες 1548–1558, 2021.
- [23] Andrey Savchenko. *GitHub - HSE-asavchenko/HSE\_FaceRectf : Tensorflow/Kerasmodelsforfacerecognition,ag/genderpredictionfromimages*. 2022.
- [24] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao και Gang Hua. *Ordinal Regression with Multiple Output CNN for Age Estimation*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 4920–4928, 2016.

- [25] Ali Mollahosseini, Behzad Hassani και Mohammad H. Mahoor. *AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild*. *CoRR*, abs/1708.03985, 2017.
- [26] 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.
- [27] Wenjing Yan, Shan Li, Chengtao Que, JiQuan Pei και Weihong Deng. *RAF-AU Database: In-the-Wild Facial Expressions with Subjective Emotion Judgement and Objective AU Annotations*. *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2020.
- [28] Shan Li και Weihong Deng. *Blended Emotion in-the-Wild: Multi-label Facial Expression Recognition Using Crowdsourced Annotations and Deep Locality Feature Learning*. *International Journal of Computer Vision*, 127(6-7):884–906, 2019.
- [29] Fabian Benitez-Quiroz, Ramprakash Srinivasan και Aleix Martinez. *EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild*. 2016.
- [30] Jeffrey Girard, Wen Sheng Chu, László Jeni, Jeffrey Cohn, FernandoDe la Torre και Michael Sayette. *GFT Facial Expression Database*, 2018.
- [31] Jean Kossaiifi, Georgios Tzimiropoulos, Sinisa Todorovic και Maja Pantic. *AFEW-VA database for valence and arousal estimation in-the-wild*. *Image and vision computing*, 65:23–36, 2017.
- [32] Abhinav Dhall, Roland Goecke, Simon Lucey και Tom Gedeon. *Collecting Large, Richly Annotated Facial-Expression Databases from Movies*. *IEEE Multimedia*, 19:34–31, 2012.
- [33] S. Mohammad Mavadati, Mohammad H. Mahoor, Kevin Bartlett, Philip Trinh και Jeffrey F. Cohn. *DISFA: A Spontaneous Facial Action Intensity Database*. *IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.
- [34] Dimitrios Kollias, Shiyang Cheng, Evangelos Ververas, Irene Kotsia και Stefanos Zafeiriou. *Generating faces for affect analysis*. *CoRR*, abs/1811.05027, 2018.
- [35] 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.
- [36] Dimitrios Kollias και Stefanos Zafeiriou. *Exploiting multi-CNN features in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset*. *CoRR*, abs/1910.01417, 2019.
- [37] Dimitrios Kollias και Stefanos Zafeiriou. *A Multi-component CNN-RNN Approach for Dimensional Emotion Recognition in-the-wild*. *arXiv preprint arXiv:1805.01452*, 2018.

- [38] Dimitrios Kollias και Stefanos Zafeiriou. *VA-StarGAN: Continuous Affect Generation*, σελίδες 227–238. 2020.
- [39] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment Mixup: Augmentation Methods for Facial Expression Recognition*. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, σελίδες 2366–2374, 2022.
- [40] Kai Wang, Xiaojiang Peng, Jianfei Yang, Debin Meng και Yu Qiao. *Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition*. *CoRR*, abs/1905.04075, 2019.
- [41] Wan Ding, Dong Yan Huang, Zhuo Chen, Xinguo Yu και Weisi Lin. *Facial action recognition using very deep networks for highly imbalanced class distribution*. σελίδες 1368–1372, 2017.
- [42] Zhiwen Shao, Zhilei Liu, Jianfei Cai και Lizhuang Ma. *Deep Adaptive Attention for Joint Facial Action Unit Detection and Face Alignment*. *CoRR*, abs/1803.05588, 2018.
- [43] Xuesong Niu, Hu Han, Songfan Yang, Yan Huang και Shiguang Shan. *Local Relationship Learning With Person-Specific Shape Regularization for Facial Action Unit Detection*. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, σελίδες 11909–11918, 2019.
- [44] Zhiwen Shao, Zhilei Liu, Jianfei Cai, Yunsheng Wu και Lizhuang Ma. *Facial Action Unit Detection Using Attention and Relation Learning*. *CoRR*, abs/1808.03457, 2018.
- [45] Dimitrios Kollias, Viktoriia Sharmanska και Stefanos Zafeiriou. *Distribution Matching for Heterogeneous Multi-Task Learning: a Large-scale Face Study*. *CoRR*, abs/2105.03790, 2021.
- [46] K. Zhang, Z. Zhang, Z. Li και Y. Qiao. *Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks*. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.
- [47] Dimitrios Kollias, Panagiotis Tzirakis, Mihalis A 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(6-7):907–929, 2019.
- [48] 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.
- [49] 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.

- [50] 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.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] Dimitrios Kollias και Stefanos Zafeiriou. *Expression, Affect, Action Unit Recognition: Aff-Wild2, Multi-Task Learning and ArcFace*. *arXiv preprint arXiv:1910.04855*, 2019.
- [55] Dimitrios Kollias και Stefanos Zafeiriou. *Aff-Wild2: Extending the Aff-Wild Database for Affect Recognition*. *arXiv preprint arXiv:1811.07770*, 2018.
- [56] Dimitrios Kollias και Stefanos Zafeiriou. *A Multi-Task Learning & Generation Framework: Valence-Arousal, Action Units & Primary Expressions*. *arXiv preprint arXiv:1811.07771*, 2018.
- [57] 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.
- [58] Dimitrios Kollias. *Abaw: Learning from synthetic data & multi-task learning challenges*. *arXiv preprint arXiv:2207.01138*, 2022.
- [59] Xing Zhang, Lijun Yin, Jeffrey F Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu και Jeffrey M Girard. *Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database*. *Image and Vision Computing*, 32(10):692–706, 2014.
- [60] Zheng Zhang, Jeff M Girard, Yue Wu, Xing Zhang, Peng Liu, Umut Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang και others. *Multimodal spontaneous emotion corpus for human behavior analysis*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, σελίδες 3438–3446, 2016.
- [61] Xingxun Jiang, Yuan Zong, Wenming Zheng, Chuangao Tang, Wanchuang Xia, Cheng Lu και Jiateng Liu. *DFEW: A Large-Scale Database for Recognizing Dynamic Facial Expressions in the Wild*. *Proceedings of the 28th ACM International Conference on Multimedia*, σελίδες 2881–2889, 2020.

- [62] Ronak Kosti, Jose M Alvarez, Adria Recasens καὶ Agata Lapedriza. *Emotion recognition in context. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [63] Ronak Kosti, Jose Alvarez, Adria Recasens καὶ Agata Lapedriza. *Context based emotion recognition using emotic dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.