



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Διπλωματική Εργασία

Συναισθηματική ευφύια σε συστήματα διεπαφής χρήστη



Φοιτητής: Ουάγκντι Νάντι
ΑΜ: 71347257

Επιβλέπων Καθηγητής

Χρήστος Τρούσσας
Επίκουρος Καθηγητής

ΑΙΓΑΛΕΩ, ΙΟΥΝΙΟΣ 2022



UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF INFORMATICS AND COMPUTER ENGINEERING

Diploma Thesis

Emotional intelligence in user interface systems



Student: Ouagknti Nanti
Registration Number: 71347257

Supervisor

Christos Troussas
Assistant Professor

EGALEO, JUNE 2022

Τίτλος εργασίας
Συναισθηματική ευφυΐα σε συστήματα διεπαφής χρήστη

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

A/a	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
	ΧΡΗΣΤΟΣ ΤΡΟΥΣΣΑΣ	ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ	
	ΑΚΡΙΒΗ ΚΡΟΥΣΚΑ	ΜΕΤΑΔΙΔΑΚΤΟΡΙΚΗ ΕΡΕΥΝΗΤΡΙΑ	
	ΠΑΝΑΓΙΩΤΑ ΤΣΕΛΕΝΤΗ	ΕΔΠ	

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Νάντι Ουάγκντι,
Ιούνιος, 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. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Ουάγκντι Νάντι του Χάσσαν με αριθμό μητρώου 71347257 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Ο Δηλών
Ουάγκντι Νάντι

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου για την για την καθοδήγηση και την υποστήριξη που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Περίληψη

Η συναισθηματική ευφυΐα ορίζεται ως η ικανότητα αντίληψης, χρήσης, κατανόησης, διαχείρισης και χειρισμού των συναισθημάτων. Τα άτομα με υψηλή συναισθηματική ευφυΐα έχουν την ικανότητα να αναγνωρίζουν τα δικά τους συναισθήματα αλλά και των άλλων, να χρησιμοποιούν τη συναισθηματική πληροφορία και να αλληλοεπιδρούν με το ερέθισμα αυτό. Ακόμη, διακρίνουν τα ποικίλα συναισθήματα και προσαρμόζουν τα συναισθήματα τους ώστε να εναρμονιστούν με ένα κοινωνικό περιβάλλον.

Η ραγδαία ανάπτυξη της επιστήμης και συγκεκριμένα του κλάδου της πληροφορικής έχει πετύχει να προσομοιώσει την ανθρώπινη νοημοσύνη σε τέτοιον βαθμό που οι υπολογιστές μπορούν να μιμούνται λειτουργίες του ανθρώπου. Ως ένας σημαντικός κλάδος της πληροφορικής και της τεχνολογίας, η τεχνητή νοημοσύνη δημιουργεί ευφυή συστήματα και προγράμματα υπολογιστών για την εκτέλεση διαφόρων εργασιών που απαιτείται ανθρώπινη νοημοσύνη. Ένα υποπεδίο της τεχνητής νοημοσύνης, η μηχανική μάθηση επικεντρώνεται στην ανάλυση και ερμηνεία προτύπων και δομών σε δεδομένα για να επιτρέψει τη μάθηση, τη λογική και τη λήψη αποφάσεων εκτός της ανθρώπινης παρέμβασης.

Ωστόσο, η ύπαρξη της αλληλεπίδρασης είναι μια απαραίτητη συνθήκη για την επικοινωνία ανθρώπου και υπολογιστή. Μια διεπαφή στην οποία περιέχεται ένα ευφυές σύστημα προϋποθέτει τον εύχρηστο και φιλικό προς τον χρήστη σχεδιασμό της, ώστε να αποδίδεται ως ένα ικανοποιητικό διαδραστικό σύστημα.

Η παρούσα διπλωματική εργασία χρησιμοποιεί μεθόδους μηχανικής μάθησης ώστε να μπορεί να αναγνωρίζει κάποια βασικά συναισθήματα (χαρά, λύπη, θυμό) μέσω κάμερας και να αλληλεπιδρά με τον χρήστη ανάλογα με το συναίσθημα που εντοπίζει εμφανίζοντας το κατάλληλο ενθαρρυντικό μήνυμα σε συνδυασμό με μουσικό ερέθισμα. Η εμφάνιση του μηνύματος και της μουσικής έχει ως σκοπό τη παρότρυνση του χρήστη στην διεκπεραίωση της εργασίας του.

Λέξεις – κλειδιά

Συναισθηματική ευφυΐα, μηχανική μάθηση, βαθιά μάθηση, αλληλεπίδραση ανθρώπου-υπολογιστή

Abstract

Emotional intelligence is usually defined as the ability to perceive, use, understand, manage and manipulate emotions. People with high emotional intelligence can recognize their own feelings and those of others, use emotional information and interact with this stimulus, distinguish between different emotions and adjust their emotions to harmonize with a social environment.

The rapid development of science and specifically of the IT industry has succeeded in simulating human intelligence to such an extent that computers can mimic human functions. As an important branch of information technology, artificial intelligence creates intelligent computer systems and programs to perform various tasks that require human intelligence. A subfield of artificial intelligence, machine learning focuses on analyzing and interpreting patterns and structures in data to enable learning, logic, and decision-making outside of human intervention.

Nevertheless, the existence of interaction is a necessary condition for human-computer communication. An interface containing an intelligent system presupposes its easy-to-use and user-friendly design to be rendered as a satisfactory interactive system.

The present dissertation uses machine learning methods so that it can recognize some basic emotions (happiness, sadness, anger) through a camera and interact with the user depending on the emotion it detects by displaying the appropriate encouraging message combined with a musical stimulus. The display of the message and music is intended to motivate the user to complete his task.

Keywords

Emotional intelligence, machine learning, deep learning, human-computer interaction

Περιεχόμενα

Κατάλογος Πινάκων.....	10
Κατάλογος Εικόνων	11
1 ΚΕΦΑΛΑΙΟ 1^ο : Εισαγωγή	13
1.1 Συναισθηματική ευφυΐα	13
1.1.1 Αναγνώριση συναισθημάτων.....	13
1.1.2 Ανάλυση βασικών συναισθημάτων.....	14
1.2 Τεχνητή Νοημοσύνη.....	15
1.2.1 Μηχανική Μάθηση.....	16
1.2.2 Αλγοριθμικές Τεχνικές.....	16
1.3 Αλληλεπίδραση ανθρώπου - υπολογιστή	17
2 ΚΕΦΑΛΑΙΟ 2^ο : Θεωρητική Ανασκόπηση.....	19
2.1 Αναγνώριση συναισθημάτων μέσω προσώπου	19
2.1.1 Ανίχνευση προσώπου (Face detection).....	21
2.1.2 Εξαγωγή χαρακτηριστικών (Feature extraction)	23
2.1.3 Κατηγοριοποίηση συναισθημάτων (Emotion classification).....	25
2.2 Υπάρχουσες τεχνικές.....	26
2.2.1 Facial Action Coding System (FACS).....	26
2.2.2 Principle Component Analysis (PCA)	27
2.2.3 Νευρωνικά Δίκτυα (Neural Networks)	27
2.2.4 Support Vector Machines (SVM)	28
2.2.5 Hidden Markov Model.....	28
2.2.6 Linear discriminant analysis (LDA).....	28
3 ΚΕΦΑΛΑΙΟ 3^ο : Μεθοδολογία.....	29
3.1 Σύνολο δεδομένων	29
3.1.1 FER-2013	29
3.1.2 Cohn-Kanade (CK).....	31
3.1.3 Τελικό σύνολο δεδομένων	31
3.2 Προεπεξεργασία δεδομένων	32
3.3 Τεχνική αναγνώρισης συναισθήματος.....	33
3.3.1 CNN.....	33
3.3.2 Εκπαίδευση νευρωνικού δικτύου	34
3.4 Αξιολόγηση αποτελεσμάτων	34
3.5 Αξιοποίηση αποτελεσμάτων	36
3.6 Αλληλεπίδραση διεπαφής με τον χρήστη	37
4 ΚΕΦΑΛΑΙΟ 4^ο : Αρχιτεκτονική και Περιγραφή Συστήματος.....	38
4.1 Προεπεξεργασία και διαχείριση δεδομένων.....	38
4.2 Εκπαίδευση νευρωνικού δικτύου	41
4.2.1 Κλάση BaseModel.....	41
4.2.2 Κλάση CNNModel	47
4.2.3 Κώδικας εκπαίδευσης.....	47
4.3 Σχεδιασμός διεπαφής χρήστη (User Interface).....	50
4.3.1 Συνάρτηση dnn_detector	50
4.3.2 Διεπαφή χρήστη	51
5 ΚΕΦΑΛΑΙΟ 5^ο : Ανάλυση Αποτελεσμάτων.....	54
5.1 Dataset FER-2013	54
5.1.1 Εκπαίδευση μοντέλου με έξι συναισθήματα	54

5.1.2	Εκπαίδευση μοντέλου με πέντε συναισθήματα	56
5.1.3	Εκπαίδευση μοντέλου με τέσσερα συναισθήματα.....	57
5.1.4	Εκπαίδευση μοντέλου με τρία συναισθήματα.....	59
5.2	Dataset FER-2013 (filtered).....	61
5.2.1	Εκπαίδευση μοντέλου με πέντε συναισθήματα	61
5.2.2	Εκπαίδευση μοντέλου με τέσσερα συναισθήματα.....	63
5.2.3	Εκπαίδευση μοντέλου με τρία συναισθήματα.....	65
5.3	Dataset CK	66
5.3.1	Εκπαίδευση μοντέλου με πέντε συναισθήματα	66
5.3.2	Εκπαίδευση μοντέλου με τέσσερα συναισθήματα.....	68
5.4	Τελικό dataset.....	70
6	ΚΕΦΑΛΑΙΟ 6^ο : Συμπεράσματα.....	72
6.1	Γενικά συμπεράσματα	72
6.2	Πρακτικές εφαρμογές	72
6.3	Περιορισμοί.....	73
6.4	Μελλοντικές επεκτάσεις	73
	Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές	75

Κατάλογος Πινάκων

Πίνακας 1: Αναγνώριση έκφρασης σε πρόσφατα συστήματα. SVM, MLR, multinomial logistic ridge regression: HMM, Bayesian network, Gaussian mixture model: RegRankBoost, RankBoost με κανονικοποίηση l1 (Rathgeb et al., 2022).....	26
Πίνακας 2: Εκδόσεις βιβλιοθηκών που χρησιμοποιήθηκαν.....	38
Πίνακας 3: Δείγμα αναλυτικής αξιολόγησης του μοντέλου.....	42
Πίνακας 4: Διαχωρισμός των δεδομένων με αναλογία 92%.....	49
Πίνακας 5: Στιγμιότυπο εκπαίδευσης νευρωνικού δικτύου.....	49
Πίνακας 6: Ενεργοποίηση των callback συναρτήσεων.....	50
Πίνακας 7: Αναλυτική αξιολόγηση στο FER-2013 με 6 συναισθήματα.....	54
Πίνακας 8: Αναλυτική αξιολόγηση στο FER-2013 με 5 συναισθήματα.....	56
Πίνακας 9: Αναλυτική αξιολόγηση στο FER-2013 με 4 συναισθήματα.....	58
Πίνακας 10: Αναλυτική αξιολόγηση στο FER-2013 με 3 συναισθήματα.....	60
Πίνακας 11: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 5 συναισθήματα.....	62
Πίνακας 12: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 4 συναισθήματα.....	64
Πίνακας 13: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 3 συναισθήματα.....	65
Πίνακας 14: Αναλυτική αξιολόγηση στο CK με 5 συναισθήματα.....	67
Πίνακας 15: Αναλυτική αξιολόγηση στο CK με 4 συναισθήματα.....	69
Πίνακας 16: Αναλυτική αξιολόγηση στο τελικό dataset.....	70

Κατάλογος Εικόνων

Εικόνα 1: Συστατικά στοιχεία συναισθηματικής νοημοσύνης (Salovey and Mayer, 1990).	13
Εικόνα 2: Τα στάδια της αναγνώρισης συναισθήματος μέσω προσώπου (EDPS, 2021).	19
Εικόνα 3: Παραδείγματα προσώπου (πάνω σειρά) και μη-προσώπου (κάτω σειρά) (Rathgeb et al, 2022)..	22
Εικόνα 4: Εξαγωγή χαρακτηριστικών προσώπου για την ανάλυση της έκφρασης. (a) Μοντέλα πολλαπλών καταστάσεων για εξαγωγή των γεωμετρικών χαρακτηριστικών, (b) Θέσεις για τον υπολογισμό των εμφανισιακών χαρακτηριστικών (Rathgeb et al., 2022).....	23
Εικόνα 5: Παράδειγμα εξαγωγής χαρακτηριστικών προσώπου. (a) Εξαγωγή μονίμων χαρακτηριστικών (μάτια, φρύδια και στόμα). (b) Εξαγωγή προσωρινών χαρακτηριστικών (ρυτίδες) (Rathgeb et al., 2022). ..	24
Εικόνα 6: Παράδειγμα εξαγωγής χαρακτηριστικών. (a) Βίντεο εισόδου (b) Καρέ βίντεο εισόδου (c) Στιγμιότυπο του συστήματος εξαγωγής γεωμετρικών χαρακτηριστικών (d) Επιλεγμένες περιοχές προσώπου για την εξαγωγή εμφανισιακών χαρακτηριστικών (Rathgeb et al., 2022).....	24
Εικόνα 7: Αρχική κατανομή συναισθημάτων του συνόλου δεδομένων FER-2013.	29
Εικόνα 8: Τελική κατανομή συναισθημάτων του συνόλου δεδομένων FER-2013 μετά το φιλτράρισμα.	30
Εικόνα 9: Δείγματα του συνόλου δεδομένων FER-2013 για το κάθε συναίσθημα.....	30
Εικόνα 10: Κατανομή συναισθημάτων του συνόλου δεδομένων CK.....	31
Εικόνα 11: Δείγματα του συνόλου δεδομένων CK για το κάθε συναίσθημα.....	31
Εικόνα 12: Κατανομή του συνόλου δεδομένων (FER-2013 + CK) που χρησιμοποιήθηκε στην μελέτη.....	32
Εικόνα 13: Αναπαράσταση μιας εικόνας ως ένα πλέγμα εικονοστοιχείων.	33
Εικόνα 14: Μια CNN ακολουθία κατηγοριοποίησης χειρόγραφων ψηφίων (Di Guan, 2020).	34
Εικόνα 15: Παράδειγμα confusion matrix.	35
Εικόνα 16: Ιστόγραμμα μιας τυχαίας εικόνας του συνόλου δεδομένων.	40
Εικόνα 17: Ιστόγραμμα μιας τυχαίας εικόνας του συνόλου δεδομένων μετά την κανονικοποίηση.....	41
Εικόνα 18: Δείγμα αναλυτικής αξιολόγησης του μοντέλου.	43
Εικόνα 19: Ιστορικό ορθότητας και απώλειας του μοντέλου.....	44
Εικόνα 20: Τα τρία στρώματα ενός νευρωνικού δικτύου (Ργο, 2017).	44
Εικόνα 21: Αρχιτεκτονική συνελκτικού νευρωνικού δικτύου.....	46
Εικόνα 22: Αναγνώριση συναισθήματος χαράς.....	52
Εικόνα 23: Αναγνώριση συναισθήματος λύπης.	52
Εικόνα 24: Αναγνώριση συναισθήματος θυμού.....	53

Εικόνα 25: Αναγνώριση ουδέτερου συναισθήματος.....	53
Εικόνα 26: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 6 συναισθήματα.	55
Εικόνα 27: Confusion matrix στο FER-2013 με 6 συναισθήματα.....	55
Εικόνα 28: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 5 συναισθήματα.	56
Εικόνα 29: Confusion matrix στο FER-2013 με 5 συναισθήματα.....	57
Εικόνα 30: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 4 συναισθήματα.	58
Εικόνα 31: Confusion matrix στο FER-2013 με 4 συναισθήματα.....	59
Εικόνα 32: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 3 συναισθήματα.	60
Εικόνα 33: Confusion matrix στο FER-2013 με 3 συναισθήματα.....	61
Εικόνα 34: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 5 συναισθήματα.	62
Εικόνα 35: Confusion matrix στο FER-2013 (filtered) με 5 συναισθήματα.	63
Εικόνα 36: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 4 συναισθήματα.	64
Εικόνα 37: Confusion matrix στο FER-2013 (filtered) με 4 συναισθήματα.	65
Εικόνα 38: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 3 συναισθήματα.	66
Εικόνα 39: Confusion matrix στο FER-2013 (filtered) με 3 συναισθήματα.	66
Εικόνα 40: Ιστορικό ορθότητας και απώλειας στο CK με 5 συναισθήματα.	67
Εικόνα 41: Confusion matrix στο CK με 5 συναισθήματα.....	68
Εικόνα 42: Ιστορικό ορθότητας και απώλειας στο CK με 4 συναισθήματα.	69
Εικόνα 43: Confusion matrix στο CK με 4 συναισθήματα.....	70
Εικόνα 44: Ιστορικό ορθότητας και απώλειας στο τελικό dataset.	71
Εικόνα 45: Confusion matrix στο τελικό dataset.	71

1 ΚΕΦΑΛΑΙΟ 1^ο : Εισαγωγή

1.1 Συναισθηματική ευφυΐα

Η συναισθηματική νοημοσύνη είναι ένας τύπος κοινωνικής νοημοσύνης που περιλαμβάνει την ικανότητα να παρακολουθεί κανείς τα δικά του συναισθήματα και τα συναισθήματα των άλλων, να κάνει διακρίσεις μεταξύ τους και να χρησιμοποιεί τις πληροφορίες που καθοδηγούν τη σκέψη και τις πράξεις κάποιου (Salovey and Mayer, 1990). Το εύρος της συναισθηματικής νοημοσύνης ενός ατόμου περιλαμβάνει τη λεκτική και τη μη λεκτική αξιολόγηση και έκφραση συναισθήματος, τη ρύθμιση του συναισθήματος στον εαυτό του και στους άλλους, και την αξιοποίηση του συναισθηματικού περιεχομένου στην επίλυση προβλημάτων.

Σε τι απαντάται όμως ο όρος νοημοσύνη γενικότερα; Σε πρόσφατα άρθρα, ο Scarr (Scarr, 1985) περιέγραψε τη νοημοσύνη ως το συγκεκριμένο των ανθρώπινων αρετών. Χρησιμοποιώντας την κοινωνική ικανότητα ως παράδειγμα (Scarr, 1989), αναφέρει πως οι προϋποθέσεις για την ομαλή αλληλεπίδραση μεταξύ των ατόμων περιλαμβάνουν την εξωστρέφεια, την αυτοπεποίθηση, το χαμηλό άγχος και την κοινωνική διορατικότητα. Ωστόσο, σημείωσε ότι παρόλο που αυτές οι αρετές συσχετίζονται με την ευφυΐα, δεν ταυτίζονται με τον ορισμό της.

Πιο συγκεκριμένα, τα χαρακτηριστικά της προσωπικότητας, όπως η εξωστρέφεια, είναι περισσότερο χαρακτηριστικό συμπεριφοράς, παρά νοημοσύνης, καθώς η νοημοσύνη περιλαμβάνει οργανικές ικανότητες συμπεριφοράς. Η αναγνώριση των συναισθημάτων σε αντίθεση, αποτελεί νοητική ικανότητα, επομένως, θεωρείται χαρακτηριστικό ευφυΐας.



Εικόνα 1: Συστατικά στοιχεία συναισθηματικής νοημοσύνης (Salovey and Mayer, 1990).

1.1.1 Αναγνώριση συναισθημάτων

Καθώς οι υπολογιστές και οι εφαρμογές τους εξελίσσονται και εμπλέκονται όλο και περισσότερο στην καθημερινότητά μας, σε επαγγελματικό, προσωπικό ή κοινωνικό επίπεδο, γίνεται όλο και πιο σημαντικό να μπορούμε να αλληλεπιδρούμε μαζί τους με φυσικό τρόπο, παρόμοιο με τον τρόπο που αλληλεπιδρούμε με άλλους ανθρώπους. Το πιο κρίσιμο χαρακτηριστικό της ανθρώπινης αλληλεπίδρασης που προσδίδει νατουραλισμό στη διαδικασία αυτή είναι η ικανότητά μας να

συμπεραίνουμε τις συναισθηματικές καταστάσεις των άλλων με βάση κρυφά ή/και φανερά σήματα αυτών των συναισθηματικών καταστάσεων. Αυτό μας επιτρέπει να προσαρμόζουμε ανάλογα τις αποκρίσεις και τα πρότυπα συμπεριφοράς μας, διασφαλίζοντας έτσι τη σύγκλιση και τη βελτιστοποίηση της διαδραστικής διαδικασίας.

Οι πιθανές εφαρμογές μιας διεπαφής ικανής να αξιολογεί τις ανθρώπινες συναισθηματικές καταστάσεις είναι πολλές. Μία από τις χρήσεις μιας τέτοιας διεπαφής είναι η ενίσχυση της ανθρώπινης κρίσης για το συναίσθημα σε καταστάσεις όπου απαιτείται αντικειμενικότητα και ακρίβεια. Η ανίχνευση ψεύδους είναι ένα προφανές παράδειγμα τέτοιων καταστάσεων, αν και η βελτίωση της ανθρώπινης απόδοσης θα απαιτούσε ένα πολύ αποτελεσματικό σύστημα αναγνώρισης συναισθημάτων. Ένα άλλο παράδειγμα είναι οι κλινικές μελέτες της σχιζοφρένειας και ιδιαίτερα η διάγνωση του «κενού» συναισθήματος, που μέχρι στιγμής βασίζεται στην υποκειμενική κρίση των ψυχιάτρων για τη συναισθηματικότητα των ασθενών με βάση διάφορες φυσιολογικές ενδείξεις.

Γενικότερα, η αυτόματη ανίχνευση και ταξινόμηση συναισθημάτων μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα ψυχολογικών και νευροφυσιολογικών μελετών της ανθρώπινης συναισθηματικής έκφρασης που μέχρι στιγμής βασίζονται στην αυτοαναφορά των ασθενών για τη συναισθηματική τους κατάσταση, η οποία συχνά αποδεικνύεται προβληματική. Σε ένα επαγγελματικό περιβάλλον, ο εμπλουτισμός μιας τηλεδιάσκεψης με πληροφορίες σε πραγματικό χρόνο σχετικά με τη συναισθηματική κατάσταση των συμμετεχόντων θα μπορούσε να βοηθήσει τους ανθρώπους να διακρίνουν πιο εύκολα τη συναισθηματική τους κατάσταση.

Μια άλλη χρήση ενός τέτοιου συστήματος θα μπορούσε να είναι η ενσωμάτωση του σε μια εφαρμογή αυτόματης διδασκαλίας (Valagkouti, et al., 2022). Το σύστημα αυτό θα μπορούσε να προσαρμόσει διαδραστικά το περιεχόμενο του μαθήματος και την ταχύτητα με την οποία παραδίδεται με βάση το αν ο χρήστης το βρίσκει βαρετό ή θλιβερό, απρόσιτο ή συναρπαστικό. Ομοίως, η ευαισθησία στο συναίσθημα μπορεί να προστεθεί σε αυτόματες υπηρεσίες πελατών και τηλεφωνικά κέντρα για παράδειγμα, για να βοηθήσει στην ανίχνευση της απογοήτευσης και στην αποφυγή περαιτέρω εκνευρισμού, με επιλογές να μεταβιβαστεί η αλληλεπίδραση σε έναν άνθρωπο ή ακόμη και να τερματιστεί εντελώς. Θα μπορούσε επίσης να φανταστεί κανείς ένα αυτοκίνητο που ανταποκρίνεται στα συναισθήματα που μπορεί να ειδοποιήσει τον οδηγό όταν ανιχνεύσει σημάδια άγχους ή θυμού που θα μπορούσαν να βλάψουν τις οδηγικές του ικανότητες.

Τέλος, χάρη στην κοινή βάση της αναγνώρισης ανθρώπινων συναισθημάτων και της συναισθηματικής έκφρασης, η κατανόηση και η ανάπτυξη αυτόματων συστημάτων αναγνώρισης συναισθημάτων μπορεί να βοηθήσει στη δημιουργία προσώπων ή/και φωνών προικισμένων με πειστικά ανθρώπινες συναισθηματικές ιδιότητες. Αυτό μπορεί με τη σειρά του να οδηγήσει σε ένα πλήρως διαδραστικό σύστημα που μπορεί να αντιληφθεί το συναίσθημα.

1.1.2 Ανάλυση βασικών συναισθημάτων

Ακολουθώντας μια μακρά παράδοση που χρονολογείται από τον Ντεκάρτ και τον Δαρβίνο που υποστηρίζει την ύπαρξη ενός μικρού, σταθερού αριθμού διακριτών (βασικών) συναισθημάτων, ο Silvan Tomkins πρότεινε το 1962 (Tomkins, 1962) ότι υπάρχουν εννέα βασικές συναισθηματικές καταστάσεις (δύο θετικές, μία ουδέτερη και έξι αρνητικές), που η καθεμία υποδηλώνεται από μια συγκεκριμένη διαμόρφωση των χαρακτηριστικών του προσώπου. Αυτή η υπόθεση διαγωνίστηκε

από πολλούς ερευνητές που ακολούθησαν (Ekman et al., 1972; Frijda, 1986; Oatley and Johnson-Laird; 1987) με κάθε ερευνητή να παράγει τη δική του λίστα βασικών συναισθημάτων που είναι διαφορετική ως προς τον αριθμό και το είδος των βασικών συναισθημάτων με αυτά στις λίστες των άλλων.

Αυτή η διαφορά είναι τουλάχιστον μπερδεμένη στην προσπάθεια κατανόησης των χαρακτηριστικών των εσωτερικών αναπαραστάσεων των διαφόρων συναισθηματικών καταστάσεων που θεωρούνται πιο κρίσιμες για την ανάπτυξη ενός συστήματος αυτόματης αναγνώρισης συναισθημάτων (Troussas, 2016). Επιπλέον, ενώ θα περίμενε κανείς ένα σύνολο βασικών συναισθημάτων να αναγνωρίζεται με συνέπεια μεταξύ των πολιτισμών - με άλλα λόγια, να είναι καθολικά - τα στοιχεία δείχνουν ότι υπάρχει ελάχιστη καθολικότητα τουλάχιστον στην αναγνώριση των συναισθημάτων από τις εκφράσεις του προσώπου (Russell, 1994) αν και αυτή η άποψη έχει αμφισβητηθεί από τον Ekman (Ekman, 1994).

Τα συναισθήματα μπορούν να διαιρεθούν σε δύο κατηγορίες: τα «πρωτεύοντα» και τα «δευτερεύοντα». Αυτά που ο Damasio (Damasio, 1994) αποκαλεί «πρωτεύοντα» συναισθήματα είναι τα πιο πρωτόγονα συναισθήματα όπως ο φόβος που βασίζεται στην έκπληξη, καθώς και οι έμφυτες αποστροφές και οι έλξεις. Από την άλλη πλευρά, τα «δευτερεύοντα» συναισθήματα είναι πιο λεπτά και περίπλοκα καθώς απαιτούν τη συμμετοχή της γνωστικής επεξεργασίας για να προκύψουν.

1.2 Τεχνητή Νοημοσύνη

Νοημοσύνη συνήθως θεωρείται η ικανότητα συλλογής πληροφορίας και λογικής σχετικά με τη γνώση για την επίλυση σύνθετων προβλημάτων. Στο εγγύς μέλλον οι έξυπνες μηχανές θα αντικαταστήσουν τις ανθρώπινες σε πολλούς τομείς. Η τεχνητή νοημοσύνη είναι η μελέτη και η ανάπτυξη ευφυών μηχανών και λογισμικού που μπορούν να συλλογιστούν, να χειρίζονται και αντιλαμβάνονται έννοιες. Ο John McCarthy επινόησε τον όρο το 1956 ως ένας κλάδος της επιστήμης των υπολογιστών που ασχολείται με το να κάνει τους υπολογιστές να συμπεριφέρονται σαν άνθρωποι (John McCarthy, 1956). Είναι διαφορετική από την ψυχολογία επειδή δίνει έμφαση στον υπολογισμό και διαφέρει επίσης από την επιστήμη των υπολογιστών λόγω της έμφασης στην αντίληψη, τη λογική και τη δράση. Λειτουργεί με τη βοήθεια τεχνητών νευρώνων (τεχνητό νευρωνικό δίκτυο) και επιστημονικών θεωρημάτων. Οι τεχνολογίες της τεχνητής νοημοσύνης έχουν εξελιχθεί σε τέτοιο σημείο ώστε να προσφέρουν πραγματικά πρακτικά οφέλη σε πολλές από τις εφαρμογές τους. Οι κύριοι τομείς της τεχνητής νοημοσύνης είναι τα εξειδικευμένα συστήματα, η επεξεργασία φυσικής γλώσσας, η κατανόηση του λόγου, η ρομποτική, τα αισθητήρια συστήματα, η όραση Υπολογιστή, η Ευφυής Εκπαίδευση με τη βοήθεια υπολογιστή.

Έχει αρκετά πλεονεκτήματα έναντι της πραγματικής νοημοσύνης, καθώς είναι πιο μόνιμη, συνεπής, μπορεί να τεκμηριωθεί και έχει την ικανότητα να πραγματοποιεί ορισμένες διεργασίες πολύ πιο γρήγορα και αποτελεσματικά από τον άνθρωπο. Τα πεδία στα οποία διαχωρίζεται η τεχνητή νοημοσύνη είναι τα εξής:

1. Κατανόηση γλώσσας: η ικανότητα του συστήματος να αντιλαμβάνεται και να ανταποκρίνεται στη φυσική γλώσσα. Χρησιμοποιείται για μετάφραση από τη προφορική γλώσσα σε γραπτή μορφή και για μετάφραση από μια φυσική γλώσσα σε μια άλλη.

2. Συστήματα μάθησης και προσαρμογής: η ικανότητα προσαρμογής συμπεριφοράς που βασίζεται στην εμπειρία και ανάπτυξης γενικών κανόνων που αφορούν στον κόσμο που βασίζονται σε αυτήν.
3. Επίλυση προβλημάτων: η ικανότητα διατύπωσης ενός προβλήματος σε κατάλληλη αναπαράσταση, προγραμματισμού για την επίλυση του και γνώση για το πότε χρειάζονται νέες πληροφορίες και πώς αποκτούνται.
4. Αντίληψη: η ικανότητα ερμηνείας των δεδομένων με τρόπο παρόμοιο με τον τρόπο που οι άνθρωποι χρησιμοποιούν τις αισθήσεις τους για να συσχετιστούν με τον κόσμο γύρω τους.
5. Μοντελοποίηση: η ικανότητα ανάπτυξης μιας εσωτερικής αναπαράστασης και ενός συνόλου κανόνων μετασχηματισμού που μπορούν να χρησιμοποιηθούν για την πρόβλεψη της συμπεριφοράς και της σχέσης μεταξύ κάποιου συνόλου αντικειμένων ή οντοτήτων του πραγματικού κόσμου.
6. Ρομποτική: ο συνδυασμός σχεδόν όλων των παραπάνω ικανοτήτων με τη δυνατότητα κίνησης στο έδαφος και χειρισμού αντικειμένων.
7. Παιχνίδια: η ικανότητα αποδοχής ενός επίσημου συνόλου κανόνων για παιχνίδια όπως για παράδειγμα το σκάκι και η μετάφραση αυτών των κανόνων σε αναπαράσταση ή δομή που επιτρέπει τη χρήση των ικανοτήτων επίλυσης προβλημάτων και μάθησης για την επίτευξη ενός επαρκούς επιπέδου απόδοσης.

1.2.1 Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης και της επιστήμης των υπολογιστών που εστιάζει στη χρήση δεδομένων και αλγορίθμων για τη μίμηση του τρόπου με τον οποίο μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβεια της (Troussas et al., 2018; Virvou et al., 2018).

Είναι ένα σημαντικό συστατικό του αναπτυσσόμενου τομέα της επιστήμης δεδομένων. Μέσω της χρήσης στατιστικών μεθόδων, οι αλγόριθμοι εκπαιδεύονται να κάνουν ταξινομήσεις ή προβλέψεις, αποκαλύπτοντας βασικές γνώσεις στα έργα εξόρυξης δεδομένων. Αυτές οι πληροφορίες στη συνέχεια οδηγούν στη λήψη αποφάσεων εντός των εφαρμογών και των επιχειρήσεων, επηρεάζοντας την ανάπτυξη τους (Kanetaki et al., 2022).

Ένα μεγάλο μέρος της εμπνέεται από σοβαρά προβλήματα από τη βιολογία, την ιατρική και την αστρονομία. Το αυξανόμενο πεδίο της υπολογιστικής βιωσιμότητας (Gomes and Carla, 2009) επιδιώκει να συνδέσει τις προόδους της μηχανικής μάθησης με τις πραγματικές προκλήσεις στο περιβάλλον, την οικονομία και την κοινωνία. Η μηχανική μάθηση έχει επιλύσει αποτελεσματικά τον εντοπισμό ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου (Zdziarski & Jonathan, 2005) και την μηχανική μετάφραση (Koehn et al., 2003), δύο προβλήματα παγκόσμιας σημασίας.

1.2.2 Αλγοριθμικές Τεχνικές

Η μηχανική μάθηση βασίζεται σε διάφορους αλγόριθμους για την επίλυση προβλημάτων, όπως είναι η ταξινόμηση. Δεν υπάρχει ένας ενιαίος τύπος αλγορίθμου που να θεωρείται βέλτιστος για την επίλυση ενός προβλήματος (Troussas, 2018). Το είδος του αλγορίθμου που χρησιμοποιείται εξαρτάται από το είδος του προβλήματος προς επίλυση, τον αριθμό των μεταβλητών, το είδος του μοντέλου που θα ταίριαζε καλύτερα και ούτω καθεξής. Παρακάτω αναφέρονται οι τύποι των αλγοριθμικών τεχνικών που χρησιμοποιούνται στην μηχανική μάθηση (Troussas et al., 2021).

1. Εποπτευόμενη μάθηση: προσφέρεται ένα εκπαιδευτικό σύνολο παραδειγμάτων με κατάλληλους στόχους και βάσει αυτού του συνόλου εκπαίδευσης, οι αλγόριθμοι ανταποκρίνονται σωστά σε όλες τις δυνατές εισόδους.
2. Μη εποπτευόμενη μάθηση: δεν παρέχονται οι στόχοι όπως στην εποπτευόμενη μάθηση. Η μη εποπτευόμενη μάθηση προσπαθεί να ανακαλύψει τις ομοιότητες μεταξύ των δεδομένων εισόδου και με βάση αυτές τις ομοιότητες, ταξινομεί τα δεδομένα.
3. Ημιοπτευόμενη μάθηση: είναι μια κατηγορία εποπτευόμενης μάθησης. Βρίσκεται ανάμεσα στην εποπτευόμενη και την μη εποπτευόμενη μάθηση.
4. Ενισχυτική μάθηση: αυτή η μάθηση ενθαρρύνεται από τη συμπεριφοριστική ψυχολογία. Ο αλγόριθμος ενημερώνεται όταν η απάντηση είναι λάθος, αλλά όχι για το μπορεί να διορθωθεί. Χρειάζεται να εξερευνήσει και να δοκιμάσει διάφορες πιθανότητες μέχρι να βρει τη σωστή απάντηση και δεν προτείνει βελτιώσεις. Διαφέρει από την εποπτευόμενη μάθηση με την έννοια ότι δεν προσφέρονται ακριβή σύνολα εισόδου και εξόδου.
5. Βαθιά μάθηση: αυτός ο κλάδος της μηχανικής μάθησης βασίζεται σε ένα σύνολο αλγορίθμων. Στα δεδομένα, αυτοί οι αλγόριθμοι μοντελοποιούν αφαίρεση υψηλού επιπέδου. Χρησιμοποιεί βαθύ γράφημα με διάφορα στρώματα επεξεργασίας, που αποτελούνται από πολλούς γραμμικούς και μη γραμμικούς μετασχηματισμούς.

1.3 Αλληλεπίδραση ανθρώπου - υπολογιστή

Η αλληλεπίδραση ανθρώπου-υπολογιστή είναι η μελέτη του τρόπου με τον οποίο οι άνθρωποι σχεδιάζουν, υλοποιούν και χρησιμοποιούν διαδραστικά συστήματα υπολογιστών και πώς οι υπολογιστές επηρεάζουν τα άτομα, τους οργανισμούς και την κοινωνία. Αυτό περιλαμβάνει όχι μόνο την ευκολία χρήσης αλλά και νέες τεχνικές αλληλεπίδρασης για την υποστήριξη εργασιών των χρηστών, την παροχή καλύτερης πρόσβασης στις πληροφορίες και τη δημιουργία πιο ισχυρών μορφών επικοινωνίας (Troussas et al., 2021).

Οι χρήστες αναμένουν εξαιρετικά αποτελεσματικές και εύκολη στην εκμάθηση διεπαφές και οι προγραμματιστές συνειδητοποιούν τώρα τον κρίσιμο ρόλο που παίζει η διεπαφή. Έρευνες δείχνουν ότι πάνω από το 50% της προσπάθειας σχεδιασμού και προγραμματισμού σε έργα αφιερώνεται στο τμήμα της διεπαφής χρήστη (Myers and Rosson, 1992). Η διεπαφή ανθρώπου-υπολογιστή είναι κρίσιμη για την επιτυχία των προϊόντων στην αγορά, καθώς και για την ασφάλεια, την χρησιμότητα και την ευχαρίστηση της χρήσης συστημάτων που βασίζονται σε υπολογιστή (Troussas et al., 2021).

Υπάρχουν ουσιαστικά εμπειρικά στοιχεία που ότι η χρήση των διαδικασιών, τεχνικών και εργαλείων που αναπτύχθηκαν από την κοινότητα αλληλεπίδρασης-ανθρώπου υπολογιστή μπορεί να μειώσει δραματικά το κόστος και να αυξήσει την παραγωγικότητα. Για παράδειγμα, μια μελέτη (Karat, 1990) ανέφερε εξοικονόμηση λόγω της χρήσης της μηχανικής ευχρηστίας (Nielsen, 1993) 41.700 δολάρια σε μια μικρή εφαρμογή που χρησιμοποιείται από 23.000 στελέχη μάρκετινγκ και 6.800.000 δολάρια για μια μεγάλη επιχειρηματική εφαρμογή που χρησιμοποιείται από 240.000 υπαλλήλους. Η εξοικονόμηση αποδόθηκε σε μειωμένο χρόνο εργασιών, λιγότερα λάθη, σημαντικά μειωμένη αναστάτωση χρήστη, μειωμένο φόρτο στο προσωπικό υποστήριξης, εξάλειψη της εκπαίδευσης και αποφυγή αλλαγών στο λογισμικό μετά την κυκλοφορία.

Μια άλλη ανάλυση εκτιμά το μέσο όφελος για την εύρεση κάθε προβλήματος χρηστικότητας στα 19.300 δολάρια (Mantei and Teorey 1988).

Μια ανάλυση χρηστικότητα ενός προτεινόμενου σταθμού εργασίας εξοικονόμησε μια τηλεφωνική εταιρεία 2 εκατομμύρια δολάρια ετησίως σε λειτουργικά έξοδα (Gray et al, 1993).

Ένα μαθηματικό μοντέλο που βασίζεται σε 11 μελέτες προτείνει ότι η χρήση λογισμικού που έχει υποβληθεί σε ενδεδειγμένη μηχανική ευχρηστίας θα εξοικονομήσει ένα μικρό έργο 39.000\$, ένα μεσαίο έργο 613.000\$ και ένα μεγάλο έργο 8.200.000\$ (Nielsen and Landauer 1993).

Με την εκτίμηση όλων των δαπανών που σχετίζονται με τη μηχανική χρηστικότητα, μια άλλη μελέτη διαπίστωσε ότι τα οφέλη μπορεί να είναι έως και 5000 φορές το κόστος (Nielsen and Landauer 1993).

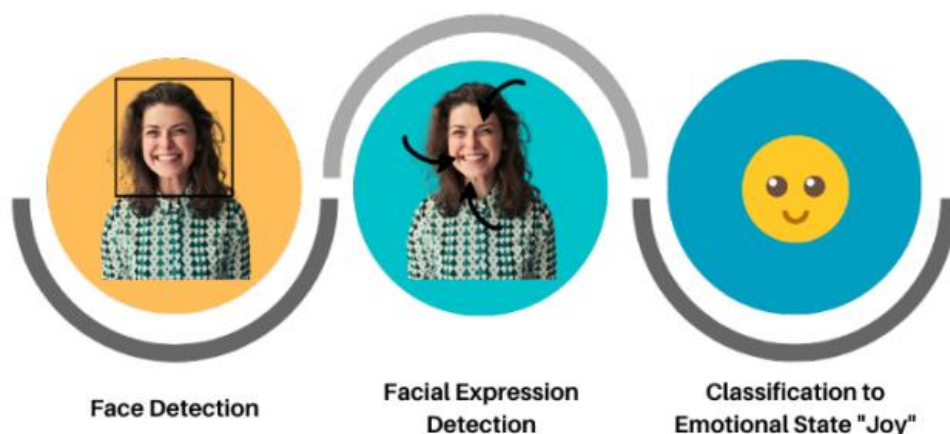
2 ΚΕΦΑΛΑΙΟ 2^ο : Θεωρητική Ανασκόπηση

2.1 Αναγνώριση συναισθημάτων μέσω προσώπου

Η αναγνώριση συναισθημάτων προσώπου είναι μια τεχνολογία που χρησιμοποιείται για την ανάλυση των συναισθημάτων από διαφορετικές πηγές, όπως φωτογραφίες και βίντεο. Ανήκει στην οικογένεια τεχνολογιών που συχνά αναφέρεται ως «συναισθηματικός υπολογισμός», ένα διεπιστημονικό πεδίο έρευνας σχετικά με τις ικανότητες του υπολογιστή να αναγνωρίζει και να ερμηνεύει τα ανθρώπινα συναισθήματα και τις συναισθηματικές καταστάσεις και συχνά βασίζεται σε τεχνολογίες Τεχνητής Νοημοσύνης (Troussas et al., 2019).

Οι εκφράσεις του προσώπου είναι μορφές μη λεκτικής επικοινωνίας, που παρέχουν υποδείξεις για τα ανθρώπινα συναισθήματα. Για δεκαετίες, η αποκωδικοποίηση τέτοιων εκφράσεων συναισθημάτων ήταν ερευνητικό ενδιαφέρον στον τομέα της ψυχολογίας (Ekman and Friesen 2003; Lang and Petter, 1993) αλλά και στο πεδίο της αλληλεπίδρασης με τον ανθρώπινο υπολογιστή (Cowie 2001; Abdat 2011). Πρόσφατα, η υψηλή διάδοση των καμερών και οι τεχνολογικές εξελίξεις στη βιομετρική ανάλυση, τη μηχανική μάθηση και την αναγνώριση προτύπων έπαιξαν εξέχοντα ρόλο στην ανάπτυξη της τεχνολογίας FER (Facial Emotion Recognition).

Η ανάλυση FER περιλαμβάνει τρία βήματα: α) ανίχνευση προσώπου, β) ανίχνευση έκφρασης προσώπου, γ) ταξινόμηση έκφρασης σε συναισθηματική κατάσταση. Βασίζεται στην ανάλυση των θέσεων των σημείων αναφοράς του προσώπου (π.χ. τέλος μύτης, φρύδια). Επιπλέον, σε βίντεο, αναλύονται αλλαγές σε αυτές τις θέσεις, προκειμένου να εντοπιστούν οι συσπάσεις σε μια ομάδα μυών του προσώπου (Ko, 2018). Ανάλογα με τον αλγόριθμο, οι εκφράσεις του προσώπου μπορούν να ταξινομηθούν σε βασικά συναισθήματα (π.χ. θυμός, αηδία, φόβος, χαρά, λύπη και έκπληξη) ή σε σύνθετα συναισθήματα (π.χ. χαρούμενα λυπημένος, χαρούμενος έκπληκτος, χαρούμενος αηδιασμένος, δυστυχώς φοβισμένος, θλιμμένος θυμωμένος, δυστυχώς έκπληκτος) (Du, 2014).



Εικόνα 2: Τα στάδια της αναγνώρισης συναισθήματος μέσω προσώπου (EDPS, 2021).

Σε άλλες περιπτώσεις, οι εκφράσεις του προσώπου θα μπορούσαν να συνδέονται με τη φυσιολογική ή ψυχική κατάσταση του νου (π.χ. κούραση ή πλήξη). Η πηγή των εικόνων ή των βίντεο που χρησιμεύουν ως είσοδος στους αλγόριθμους FER ποικίλλει από κάμερες παρακολούθησης έως κάμερες τοποθετημένες κοντά σε οθόνες διαφημίσεων σε καταστήματα, καθώς και σε μέσα κοινωνικής δικτύωσης και υπηρεσίες ροής ή προσωπικές συσκευές. Το FER μπορεί επίσης να συνδυαστεί με βιομετρική αναγνώριση. Η ακρίβειά του μπορεί να βελτιωθεί με την τεχνολογία που αναλύει διαφορετικούς τύπους πηγών, όπως φωνή, κείμενο, δεδομένα υγείας από αισθητήρες ή μοτίβα ροής αίματος που συνάγονται από την εικόνα.

Οι πιθανές χρήσεις του FER καλύπτουν ένα ευρύ φάσμα εφαρμογών, παραδείγματα των οποίων παρατίθενται παρακάτω σε ομάδες ανάλογα με το πεδίο εφαρμογής τους:

- Παροχή εξατομικευμένων υπηρεσιών

Ανάλυση συναισθημάτων για την εμφάνιση εξατομικευμένων μηνυμάτων σε έξυπνα περιβάλλοντα

Παροχή εξατομικευμένων συστάσεων π.χ. για μία μουσική επιλογή ή ένα πολιτιστικό υλικό

Ανάλυση των εκφράσεων του προσώπου για την πρόβλεψη αντιδράσεων θεατών σε ταινίες

Ανάλυση συμπεριφοράς πελατών και διαφήμιση

- Ανάλυση των συναισθημάτων των πελατών σε μία αγορά, βοηθώντας είτε στην επιλογή του προϊόντος είτε στην διευθέτηση της αγοράς
- Διαφημιστική σήμανση σε έναν σιδηροδρομικό σταθμό χρησιμοποιώντας σύστημα αναγνώρισης και παρακολούθησης προσώπου για σκοπούς μάρκετινγκ

Σύστημα υγείας

- Ανίχνευση αυτισμού ή νευροεκφυλιστικών ασθενειών
- Πρόβλεψη ψυχωτικών διαταραχών ή κατάθλιψης ώστε να ταυτοποιηθούν οι ασθενείς που χρειάζονται βοήθεια
- Πρόληψη αυτοκτονιών
- Ανίχνευση κατάθλιψης σε ηλικιωμένους
- Παρατήρηση των συνθηκών των ασθενών κατά τη διάρκεια της επίσκεψής τους στο νοσοκομείο

Εργασία

- Βοήθεια στην λήψη αποφάσεων για ενδεχόμενες προσλήψεις
- Αναγνώριση αδιάφορων υποψήφιων σε μια συνέντευξη
- Παρακολούθηση της διάθεσης και της προσοχής των εργαζομένων

Εκπαίδευση

- Παρακολούθηση της προσοχής των μαθητών
- Εντοπισμός της συναισθηματικής αντίδρασης των μαθητών σε ένα εκπαιδευτικό πρόγραμμα
- Σχεδίαση συναισθηματικού συστήματος διδασκαλίας
- Ανίχνευση εμπλοκής στη διαδικτυακή μάθηση

Δημόσια ασφάλεια

- Ανίχνευση ψεύδους και έξυπνος έλεγχος συνόρων
- Προγνωστικός έλεγχος δημόσιων χώρων για τον εντοπισμό συναισθημάτων που προκαλούν πιθανή τρομοκρατική απειλή
- Ανάλυση υλικού από σκηνές εγκλήματος ώστε να υποδειχθούν πιθανά κίνητρα σε ένα έγκλημα

Εντοπισμός εγκλήματος

- Εντοπισμός και μείωση ψευδών ασφαλιστικών απαιτήσεων
- Ανάπτυξη στρατηγικών πρόληψης απάτης

Άλλο

- Εντοπισμός κόπωσης ενός οδηγού
- Εντοπισμός πολιτικών συμπεριφορών

2.1.1 Ανίχνευση προσώπου (Face detection)

Η ανίχνευση προσώπου είναι το πρώτο βήμα στην αυτοματοποιημένη αναγνώριση προσώπου. Η αξιοπιστία της επηρεάζει σημαντικά την απόδοση και τη χρηστικότητα ολόκληρου του συστήματος αναγνώρισης προσώπου. Με δεδομένη μια μεμονωμένη εικόνα ή ένα βίντεο, ένας ιδανικός ανιχνευτής προσώπου θα πρέπει να μπορεί να αναγνωρίζει και να εντοπίζει όλα τα παρόντα πρόσωπα ανεξάρτητα από τη θέση, την κλίμακα, τον προσανατολισμό, την ηλικία και την έκφρασή τους (Kontellis et al., 2021). Επιπλέον, η ανίχνευση θα πρέπει να γίνεται ανεξάρτητα από το εξωτερικές συνθήκες φωτισμού και το περιεχόμενο εικόνας και βίντεο.

Η ανίχνευση προσώπου μπορεί να πραγματοποιηθεί με βάση πολλά στοιχεία: χρώμα δέρματος (για πρόσωπα σε έγχρωμες εικόνες και βίντεο), κίνηση (για πρόσωπα σε βίντεο), σχήμα προσώπου/κεφαλιού, εμφάνιση προσώπου ή συνδυασμό αυτών των παραμέτρων. Οι περισσότεροι επιτυχημένοι αλγόριθμοι ανίχνευσης προσώπου βασίζονται στην εμφάνιση χωρίς να χρησιμοποιούν άλλα στοιχεία. Η επεξεργασία γίνεται ως εξής: Μια εικόνα εισόδου σαρώνεται σε όλες τις πιθανές θέσεις και κλίμακες από ένα υποπαράθυρο. Η ανίχνευση προσώπου τίθεται ως ταξινόμηση του μοτίβου στο υποπαράθυρο είτε ως πρόσωπο είτε ως μη-πρόσωπο. Ο ταξινομητής προσώπου/μη

προσώπου μαθαίνεται από παραδείγματα εκπαίδευσης προσώπου και μη με χρήση στατιστικών μεθόδων εκμάθησης.



Εικόνα 3: Παραδείγματα προσώπου (πάνω σειρά) και μη-προσώπου (κάτω σειρά) (Rathgeb et al, 2022).

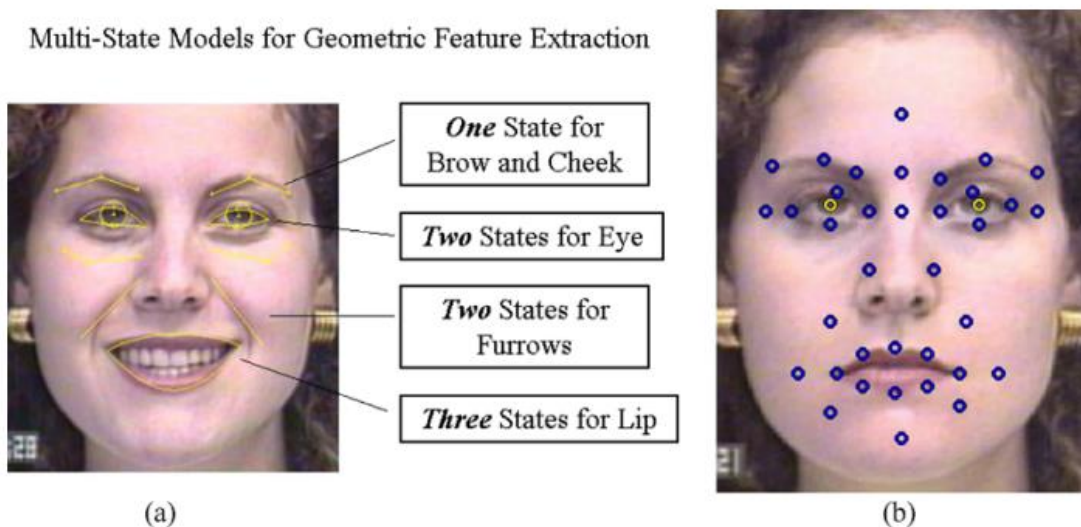
Με τη χρήση μεθόδων που βασίζονται στην εμφάνιση, η ανίχνευση προσώπου αντιμετωπίζεται ως πρόβλημα ταξινόμησης κάθε σαρωμένου υποπαραθύρου ως μία από τις δύο κατηγορίες (δηλαδή, πρόσωπο και μη-πρόσωπο). Οι μέθοδοι που βασίζονται στην εμφάνιση αποφεύγουν δυσκολίες στη μοντελοποίηση τρισδιάστατων δομών προσώπων εξετάζοντας πιθανές εμφανίσεις προσώπων υπό διάφορες συνθήκες. Ένας ταξινομητής προσώπου/μη-προσώπου μπορεί να εκπαιδευτεί πάνω σε ένα σύνολο δεδομένων εκπαίδευσης που αποτελείται από παραδείγματα προσώπων που λαμβάνονται υπό πιθανές συνθήκες, όπως θα φαινόταν επίσης στο στάδιο εκτέλεσης και στα παραδείγματα χωρίς πρόσωπο, όπως φαίνεται και στην Εικόνα 3. Η δημιουργία ενός τέτοιου ταξινομητή είναι δυνατή επειδή τα εικονοστοιχεία σε ένα πρόσωπο συσχετίζονται σε μεγάλο βαθμό, ενώ αυτά σε ένα υποπαραθύρο χωρίς πρόσωπο παρουσιάζουν πολύ λιγότερη κανονικότητα.

Ωστόσο, οι μεγάλες παραλλαγές που προκαλούνται από αλλαγές στην εμφάνιση του προσώπου, τον φωτισμό και την έκφραση καθιστούν την πολλαπλότητα του προσώπου ή τα όρια προσώπου/μη προσώπου αρκετά περίπλοκα (Bichsel and Pentland, 1994). Οι αλλαγές στην όψη του προσώπου (πόζα κεφαλιού) περιπλέκουν περαιτέρω την κατάσταση. Απαιτείται ένας μη γραμμικός ταξινομητής για την αντιμετώπιση της περίπλοκης κατάστασης. Η ταχύτητα είναι επίσης ένα σημαντικό ζήτημα για την απόδοση σε πραγματικό χρόνο.

Έχει γίνει μεγάλη ερευνητική προσπάθεια για την κατασκευή πολύπλοκων αλλά γρήγορων ταξινομητών και έχει επιτευχθεί μεγάλη πρόοδος από τη δεκαετία του 1990. Οι Turk and Pentland (Turk and Pentland, 1991) περιγράφουν ένα σύστημα ανίχνευσης που βασίζεται στον υποχώρο ανάλυσης κύριου συστατικού (PCA) ή στην αναπαράσταση ιδιομορφίας. Ενώ μόνο η πιθανότητα στον υποχώρο PCA λαμβάνεται υπόψη στη βασική μέθοδο PCA, οι Moghaddam και Pentland (Moghaddam and Pentland, 1997) εξετάζουν επίσης την πιθανότητα στον υποχώρο του ορθογώνιου συμπληρώματος, χρησιμοποιώντας αυτό το σύστημα, η πιθανότητα στον χώρο της εικόνας (η ένωση των δύο υποχώρων) μοντελοποιείται ως το γινόμενο των δύο εκτιμήσεων πιθανότητας, οι οποίες παρέχουν μια πιο ακριβή εκτίμηση πιθανότητας για την ανίχνευση. Οι Sung και Poggio (Sung and Poggio, 1998) κατανέμουν πρώτα τον χώρο της εικόνας σε πολλά συμπλέγματα προσώπων και μη προσώπων και στη συνέχεια αποσυνθέτουν περαιτέρω κάθε σύμπλεγμα στους υποχώρους PCA και null. Στη συνέχεια εφαρμόζεται ο εκτιμητής Bayes για να ληφθούν χρήσιμα στατιστικά χαρακτηριστικά. Το σύστημα του Rowley (Rowley et. al, 1998) χρησιμοποιεί νευρωνικά δίκτυα συνδεδεμένα στον αμφιβληστροειδή. Μέσω ενός συρόμενου παραθύρου, η εικόνα εισόδου εξετάζεται αφού περάσει από ένα εκτεταμένο στάδιο προεπεξεργασίας. Ο Osuna (Osuna et. al, 1997) εκπαιδευσε μια μη γραμμική μηχανή διανυσμάτων υποστήριξης (support vector machine) για την ταξινόμηση μοτίβων προσώπων και μη, και ο Yang (Yang et. al, 2000) χρησιμοποίησε την αρχιτεκτονική εκμάθησης SnoW (Sparse Network of Winnows) για ανίχνευση προσώπου.

2.1.2 Εξαγωγή χαρακτηριστικών (Feature extraction)

Το επόμενο βήμα μετά τον εντοπισμό του προσώπου, είναι η εξαγωγή των χαρακτηριστικών του. Μπορούν να εξαχθούν δύο τύποι χαρακτηριστικών: τα γεωμετρικά χαρακτηριστικά και εμφανισιακά χαρακτηριστικά. Τα γεωμετρικά χαρακτηριστικά παρουσιάζουν το σχήμα και τις θέσεις των στοιχείων του προσώπου (συμπεριλαμβανομένου του στόματος, των ματιών, των φρυδιών και της μύτης), ενώ τα χαρακτηριστικά του προσώπου εξάγονται για να σχηματιστεί ένα διάλυμα χαρακτηριστικών που αντιπροσωπεύει τη γεωμετρία του (Krouska et al., 2016).

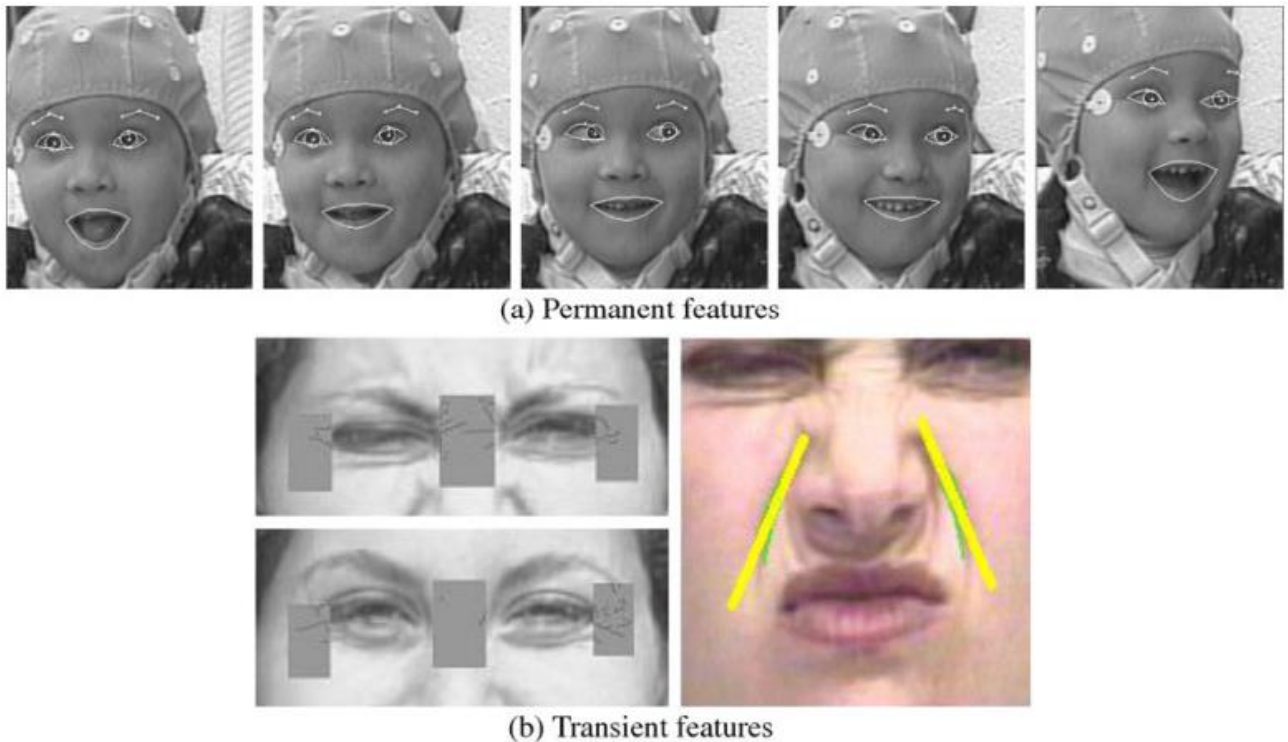


Εικόνα 4: Εξαγωγή χαρακτηριστικών προσώπου για την ανάλυση της έκφρασης. (α) Μοντέλα πολλαπλών καταστάσεων για εξαγωγή των γεωμετρικών χαρακτηριστικών, (β) Θέσεις για τον υπολογισμό των εμφανισιακών χαρακτηριστικών (Rathgeb et al., 2022).

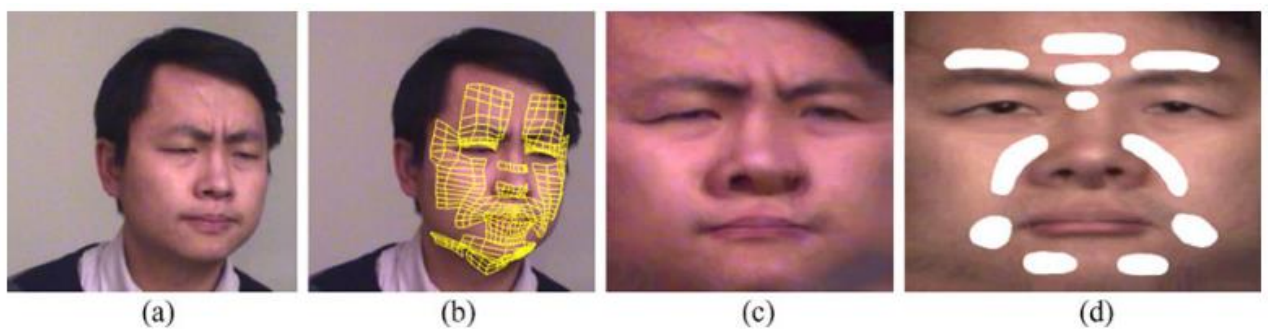
Εξαγωγή γεωμετρικών χαρακτηριστικών

Όπως φαίνεται και στην εικόνα 4, προκειμένου να ανιχνευθούν οι αλλαγές των στοιχείων του προσώπου, αναπτύχθηκαν μοντέλα πολλαπλών καταστάσεων για την εξαγωγή των γεωμετρικών χαρακτηριστικών. Ένα μοντέλο για τα χείλη τριών καταστάσεων περιγράφει την κατάσταση των χειλιών: ανοιχτά, κλειστά, ερμητικά κλειστά. Ένα μοντέλο δύο καταστάσεων χρησιμοποιείται για κάθε ένα από τα μάτια (κλειστό, ανοιχτό). Κάθε φρύδι και μάγουλο έχει ένα μοντέλο μίας μόνο κατάστασης. Ορισμένα εμφανισιακά χαρακτηριστικά, όπως οι ρυτίδες, αναπαρίστανται ρητά χρησιμοποιώντας μόνο δύο καταστάσεις: παρουσία ή απουσία. Δεδομένης μιας ακολουθίας εικόνων, η περιοχή του προσώπου και η κατά προσέγγιση θέση των μεμονωμένων χαρακτηριστικών του ανιχνεύονται αυτόματα στο αρχικό καρέ (Rowley et. al, 1998). Στη συνέχεια, τα περιγράμματα των χαρακτηριστικών και των στοιχείων του προσώπου προσαρμόζονται χειροκίνητα στο αρχικό πλαίσιο. Μετά την αρχικοποίηση, όλες οι αλλαγές χαρακτηριστικών προσώπου εντοπίζονται αυτόματα και παρακολουθούνται στην ακολουθία εικόνων. Το σύστημα ομαδοποιεί 15 παραμέτρους για το επάνω μέρος του προσώπου και 9 παραμέτρους για το κάτω, οι οποίες περιγράφουν το σχήμα, την κίνηση και την κατάσταση των στοιχείων του προσώπου. Για την εξάλειψη των επιπτώσεων της κλίμακας μεταξύ των ακολουθιών εικόνας στο μέγεθος του

προσώπου, όλες οι παράμετροι υπολογίζονται ως αναλογίες των τρεχουσών τιμών τους προς αυτές στο πλαίσιο αναφοράς.



Εικόνα 5: Παράδειγμα εξαγωγής χαρακτηριστικών προσώπου. (a) Εξαγωγή μονίμων χαρακτηριστικών (μάτια, φρύδια και στόμα). (b) Εξαγωγή προσωρινών χαρακτηριστικών (ρυτίδες) (Rathgeb et al., 2022).



Εικόνα 6: Παράδειγμα εξαγωγής χαρακτηριστικών. (a) Βίντεο εισόδου (b) Καρέ βίντεο εισόδου (c) Στιγμιότυπο του συστήματος εξαγωγής γεωμετρικών χαρακτηριστικών (d) Επιλεγμένες περιοχές προσώπου για την εξαγωγή εμφανισιακών χαρακτηριστικών (Rathgeb et al., 2022).

Εξαγωγή εμφανισιακών χαρακτηριστικών

Τα κυματίδια Gabor (Daugmen, J, 1988) χρησιμοποιούνται ευρέως για την εξαγωγή των αλλαγών στα εμφανισιακά χαρακτηριστικά του προσώπου ως ένα σύνολο συντελεστών πολλαπλής κλίμακας και πολλαπλού προσανατολισμού. Το φίλτρο Gabor μπορεί να εφαρμοστεί σε συγκεκριμένες θέσεις σε ένα πρόσωπο ή σε ολόκληρη την εικόνα του προσώπου. Ο Zhang (Zhang et al., 1998) ήταν ο πρώτος που συνέκρινε δύο τύπους χαρακτηριστικών για την αναγνώριση εκφραστικών σημείων, τις γεωμετρικές θέσεις των 34 βασικών σημείων σε ένα πρόσωπο και τους 612 συντελεστές κυματιδίων Gabor που εξήχθησαν από την εικόνα του προσώπου σε αυτά τα 34

βασικά σημεία. Τα ποσοστά αναγνώρισης για έξι εκφράσεις που καθορίζονται από το συναίσθημα (π.χ. χαρά και θυμός) ήταν σημαντικά υψηλότερα για τους συντελεστές κυματιδίων Gabor. Ο Donato (Donato et al., 1999) συνέκρινε διάφορες τεχνικές για την αναγνώριση έξι μονάδων δράσης μονής άνω όψης και έξι μονάδων δράσης κάτω προσώπου. Αυτές οι τεχνικές περιλαμβάνουν οπτική ροή, ανάλυση κύριας συνιστώσας, ανάλυση ανεξάρτητων συνιστωσών, τοπική ανάλυση χαρακτηριστικών και αναπαράσταση κυματιδίων Gabor. Οι καλύτερες επιδόσεις λήφθηκαν χρησιμοποιώντας αναπαράσταση κυματιδίων Gabor και ανάλυση ανεξάρτητων συνιστωσών. Όλα αυτά τα συστήματα χρησιμοποίησαν ένα χειροκίνητο βήμα για να ευθυγραμμίσουν κάθε εικόνα εισόδου με μια τυπική εικόνα προσώπου χρησιμοποιώντας το κέντρο των ματιών και του στόματος.

2.1.3 Κατηγοριοποίηση συναισθημάτων (Emotion classification)

Το τελευταίο βήμα των συστημάτων αυτόματης ανάλυσης εκφράσεων προσώπου είναι η κατηγοριοποίηση της έκφρασης του προσώπου με βάση τα εξαγόμενα χαρακτηριστικά (Giannakas et al., 2021; Krouska et al., 2017). Πολλοί ταξινομητές έχουν εφαρμοστεί στην κατηγοριοποίηση εκφράσεων όπως τα νευρωνικά δίκτυα (Neural Networks), οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines), η διακριτή γραμμική ανάλυση (Linear Discriminant Analysis), η μέθοδος k-πλησιέστερων γειτόνων (K-nearest neighbors, η πολυωνυμική λογιστική παλινδρόμηση (Multinomial Logistic ridge Regression), τα κρυφά μοντέλα Markov (Hidden Markov Models), η τεχνική naïve Bayes και άλλοι (Troussas et al., 2013; Troussas et al., 2017). Ορισμένα συστήματα χρησιμοποιούν μόνο μια ταξινόμηση βασισμένη σε κανόνες που βασίζεται στον ορισμό των ενεργειών του προσώπου. Οι μέθοδοι αναγνώρισης εκφράσεων μπορούν να χωριστούν σε δύο βασικές κατηγορίες, σε μεθόδους που βασίζονται σε καρτέ και σε μεθόδους που βασίζονται σε ακολουθίες. Η μέθοδος αναγνώρισης βάσει καρτέ χρησιμοποιεί μόνο το τρέχον καρτέ με ή χωρίς εικόνα αναφοράς (είναι κυρίως μια εικόνα ουδέτερου προσώπου) για να αναγνωρίσει τις εκφράσεις (Troussas et al., 2015). Η μέθοδος αναγνώρισης βάσει ακολουθίας χρησιμοποιεί τις χρονικές πληροφορίες των ακολουθιών για να αναγνωρίσει τις εκφράσεις για ένα ή περισσότερα πλαίσια (Troussas et al., 2018). Ο Πίνακας 1 συνοψίζει τις μεθόδους αναγνώρισης, τα ποσοστά αναγνώρισης, τα αποτελέσματα αναγνώρισης και τις βάσεις δεδομένων που χρησιμοποιούνται στα πιο πρόσφατα συστήματα. Για τα συστήματα που χρησιμοποίησαν περισσότερους ταξινομητές, έχει επιλεγεί η καλύτερη απόδοση για ανεξάρτητη δοκιμή (Troussas et al., 2016).

Αναγνώριση βάσει καρτέ

Η αναγνώριση βάσει καρτέ δεν χρησιμοποιεί χρονικές πληροφορίες για τις εικόνες εισόδου. Χρησιμοποιεί τις πληροφορίες της τρέχουσας εικόνας εισόδου με/χωρίς καρτέ αναφοράς. Η εικόνα εισόδου μπορεί να είναι μια στατική εικόνα ή ένα πλαίσιο μιας ακολουθίας που αντιμετωπίζεται ανεξάρτητα. Διάφορες μέθοδοι έχουν χρησιμοποιηθεί για την αναγνώριση της έκφρασης του προσώπου, όπως τα νευρωνικά δίκτυα, οι μηχανές διανυσμάτων υποστήριξης, η ανάλυση γραμμικής διάκρισης, το δίκτυο Bayes και οι ταξινομητές βάση κανόνων.

Αναγνώριση βάσει ακολουθίας

Η μέθοδος αναγνώρισης βάσει ακολουθίας χρησιμοποιεί τις χρονικές πληροφορίες των ακολουθιών για να αναγνωρίσει τις εκφράσεις ενός ή περισσότερων πλαισίων. Για να χρησιμοποιηθούν οι χρονικές πληροφορίες, οι τεχνικές όπως το HMM (Hidden Markov Models), τα επαναλαμβανόμενα νευρωνικά δίκτυα RNN (Recurrent Neural Networks) και ο ταξινομητής βάση κανόνων (Rule-

based Classifier) χρησιμοποιήθηκαν στην ανάλυση της έκφρασης του προσώπου. Τα συστήματα του χρησιμοποιούσαν έναν ταξινομητή βασισμένο σε ακολουθία.

Recognition methods	Recognition rate	Recognition outputs	Databases
Neural network (frame)	95.5%	16 single AUs and their combinations	Ekman–Hager Cohn–Kanade
Rule-based (sequence)	100% 57%	Blink, nonblink Brow up, down, and non-motion	Frank–Ekman
SVM + MLR (frame)	91.5%	6 Basic expressions	Cohn–Kanade
Adaboost + SVM (sequence)	80.1%	20 facial actions	Frank–Ekman
BN + HMM (frame & sequence)	73.22% 66.53%	6 Basic expressions 6 Basic expressions	UIUC–Chen
NN + GMM (frame)	71%	6 Basic expressions	Cohn–Kanade
RegRankBoost (frame)	88%	6 Basic expressions	Cohn–Kanade

Πίνακας 1: Αναγνώριση έκφρασης σε πρόσφατα συστήματα. SVM, MLR, multinomial logistic ridge regression: HMM, Bayesian network, Gaussian mixture model: RegRankBoost, RankBoost με κανονικοποίηση l1 (Rathgeb et al., 2022).

2.2 Υπάρχουσες τεχνικές

Μέχρι στιγμής, έχουν γίνει πολυάριθμες έρευνες για την αναγνώριση συναισθήματος μέσω του προσώπου. Οι εκφράσεις του προσώπου παρέχουν τα δομικά στοιχεία με τα οποία μπορεί κανείς να κατανοήσει το συναίσθημα. Προκειμένου να χρησιμοποιηθούν αποτελεσματικά οι εκφράσεις του προσώπου, είναι απαραίτητο να κατανοηθεί πώς να ερμηνεύονται οι εκφράσεις και είναι επίσης σημαντικό να μελετηθεί ποιες τεχνικές έχουν εφαρμοστεί στο παρελθόν (Troussas, 2020).

2.2.1 Facial Action Coding System (FACS)

Το Facial Action Coding System (FACS) είναι ένα σύστημα για τον προσδιορισμό των εκφράσεων του ανθρώπινου προσώπου, που αναπτύχθηκε αρχικά από τους Paul Ekman και Wallace V. Friesen και δημοσιεύτηκε το 1978. Οι Ekman, Friesen και Joseph C. Hager δημοσίευσαν μια σημαντική ενημέρωση για το FACS το 2002. Οι κινήσεις των μεμονωμένων μυών του προσώπου κωδικοποιούνται από το FACS από ελαφρές διαφορετικές στιγμιαίες αλλαγές στην εμφάνιση του προσώπου. Η συστηματική κατηγοριοποίηση της φυσικής έκφρασης των συναισθημάτων έχει καθιερωθεί στα τα συστήματα αυτά. Πρόσφατα, το FACS καθιερώθηκε ως ένα υπολογιστικό αυτοματοποιημένο σύστημα που ανιχνεύει πρόσωπα σε βίντεο, εξάγει τα γεωμετρικά χαρακτηριστικά των προσώπων και στη συνέχεια παράγει χρονικά προφίλ κάθε κίνησης του

προσώπου. Το FACS είναι ένα βασικό σύστημα για τον προσδιορισμό της εξαγωγής χαρακτηριστικών προσώπου.

Οι Fasel και Luetlin πραγματοποίησαν μια εις βάθος μελέτη σε μια προσπάθεια να κατανοήσουν τις πηγές που ελέγχουν τις εκφράσεις. Μεγάλο μέρος της έρευνας έκφρασης μέχρι σήμερα έχει επικεντρωθεί στην κατανόηση του τρόπου με τον οποίο οι υποκείμενοι μύες κινούνται για να δημιουργήσουν εκφράσεις. Για παράδειγμα, μελέτες έχουν δείξει ότι η κίνηση των ματιών, των φρυδιών και του στόματος ενός ατόμου είναι πρωταρχικοί δείκτες σχεδόν όλων των συναισθημάτων.

Πολλές προηγούμενες μελέτες έχουν χρησιμοποιήσει το FACS ως εργαλείο ταξινόμησης. Επιπλέον, προηγούμενες μελέτες έχουν παραδοσιακά υιοθετήσει δύο προσεγγίσεις για την ταξινόμηση των συναισθημάτων σύμφωνα με τους Fasel και Luetlin.

- Προσέγγιση βασισμένη στην κρίση
- Προσέγγιση βασισμένη σε σήματα

Η προσέγγιση βασισμένη στην κρίση αναπτύσσει εκ των προτέρων τις κατηγορίες των συναισθημάτων, όπως τα παραδοσιακά έξι καθολικά συναισθήματα, ενώ η προσέγγιση βασισμένη σε σήματα χρησιμοποιεί ένα σύστημα FACS, που κωδικοποιεί μονάδες ενεργειών προκειμένου να κατηγοριοποιήσει μια έκφραση με βάση τα χαρακτηριστικά της.

Αυτή η προσέγγιση αποδίδει μια συναισθηματική αξία σε ένα πρόσωπο χρησιμοποιώντας έναν συνδυασμό των βασικών μονάδων δράσης που δημιουργούν την έκφραση. Ένα νευρωνικό δίκτυο χρησιμοποιείται για την αναγνώριση μονάδων δράσης από τις συντεταγμένες των χαρακτηριστικών του προσώπου όπως οι γωνίες των χειλιών ή η καμπύλη των φρυδιών.

2.2.2 Principle Component Analysis (PCA)

Για τη μέθοδο αυτοοργανωμένης μάθησης, η ανάλυση βασικών συνιστωσών (PCA) χρησιμοποιείται ευρέως στον τομέα της συμπίεσης δεδομένων και της εξαγωγής χαρακτηριστικών.

Υπάρχουν δύο βασικές προσεγγίσεις για τον υπολογισμό των κύριων στοιχείων: κατά παρτίδες και προσαρμοστικές μέθοδοι. Οι μέθοδοι παρτίδας περιλαμβάνουν τη μέθοδο της αποσύνθεσης ιδιογένειας και τη μέθοδο της αποσύνθεσης μοναδικής τιμής (SVD), ενώ οι προσαρμοστικές μέθοδοι γίνονται κυρίως από νευρωνικά δίκτυα. Ο κύριος στόχος του PCA είναι να εξηγήσει τη δομή διακύμανσης-συνδιακύμανσης των δεδομένων μέσω μερικών γραμμικών συνδυασμών των αρχικών μεταβλητών.

Το κύριο πρόβλημα που υπάρχει χρησιμοποιώντας την τεχνική αυτή, είναι ότι χρησιμοποιεί μόνο τις καθολικές πληροφορίες των εικόνων, καθιστώντας την μη αποτελεσματική για διαφορετικές εκφράσεις του προσώπου.

2.2.3 Νευρωνικά Δίκτυα (Neural Networks)

Τα νευρωνικά δίκτυα είναι ένα άλλο βασικό εργαλείο εξαγωγής χαρακτηριστικών από μια εικόνα. Πολλά προβλήματα αναγνώρισης προτύπων, όπως η αναγνώριση αντικειμένων και η αναγνώριση

χαρακτήρων έχουν αντιμετωπιστεί με επιτυχία από τα νευρωνικά δίκτυα. Αυτά τα συστήματα μπορούν να χρησιμοποιηθούν στην ανίχνευση προσώπου με διαφορετικούς τρόπους.

2.2.4 Support Vector Machines (SVM)

Τα SVM είναι γραμμικοί ταξινομητές που μεγιστοποιούν το περιθώριο μεταξύ του υπερεπίπεδου απόφασης και των παραδειγμάτων στο σύνολο εκπαίδευσης. Έτσι, ένα βέλτιστο υπερεπίπεδο θα πρέπει να ελαχιστοποιεί το σφάλμα ταξινόμησης των αόρατων μοτίβων δοκιμής.

2.2.5 Hidden Markov Model

Το Hidden Markov Model είναι ένα άλλο στατιστικό μοντέλο που έχει χρησιμοποιηθεί για την ανίχνευση προσώπου. Η πρόκληση είναι να δημιουργηθεί ένα σωστό HMM, έτσι ώστε να είναι αξιόπιστη η πιθανότητα εξόδου. Οι καταστάσεις του μοντέλου θα είναι τα χαρακτηριστικά του προσώπου που συχνά ορίζονται ως λωρίδες εικονοστοιχείων. Η πιθανολογική μετάβαση μεταξύ των καταστάσεων είναι συνήθως τα όρια μεταξύ αυτών των λωρίδων εικονοστοιχείων. Όπως και στην περίπτωση των Bayesians, τα HMM χρησιμοποιούνται συνήθως μαζί με άλλες μεθόδους για τη δημιουργία αλγορίθμων ανίχνευσης.

2.2.6 Linear discriminant analysis (LDA)

Αυτή η μέθοδος χρησιμοποιείται στη στατιστική και την αναγνώριση προτύπων για την εύρεση ενός γραμμικού συνδυασμού χαρακτηριστικών. Ο συνδυασμός που προκύπτει μπορεί να χρησιμοποιηθεί ως γραμμικός ταξινομητής ή, πιο συχνά, για μείωση διαστάσεων πριν από μεταγενέστερη ταξινόμηση. Το LDA επιχειρεί ρητά να μοντελοποιήσει τη διαφορά μεταξύ των κατηγοριών δεδομένων. Από την άλλη πλευρά, το PCA δεν λαμβάνει υπόψη καμία διαφορά στην κατηγορία και η ανάλυση παραγόντων δημιουργεί τους συνδυασμούς χαρακτηριστικών με βάση τις διαφορές και όχι τις ομοιότητες.

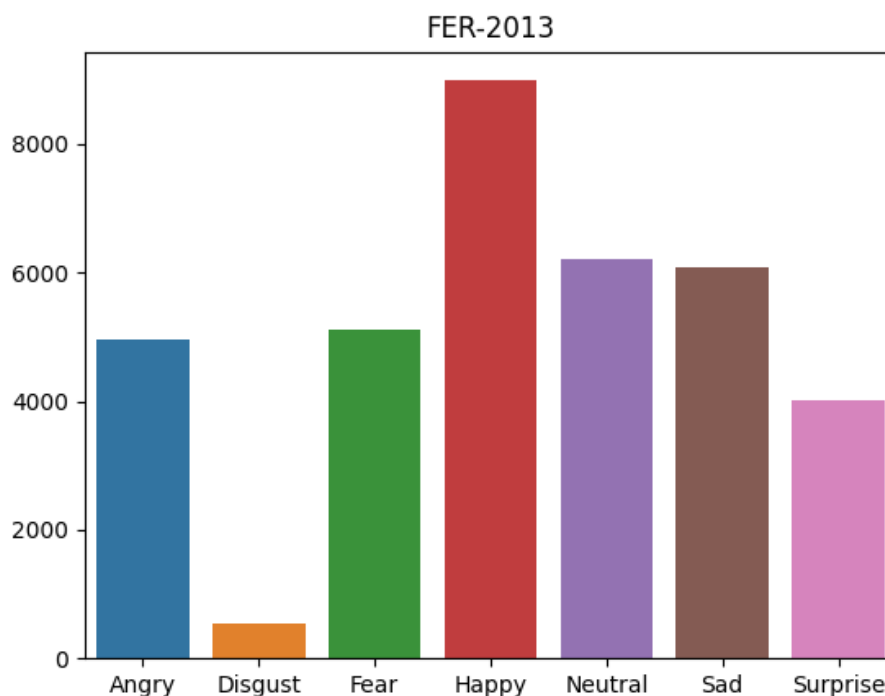
3 ΚΕΦΑΛΑΙΟ 3^ο : Μεθοδολογία

3.1 Σύνολο δεδομένων

Καθώς η μελέτη που πραγματοποιήθηκε είχε ως στόχο την κατηγοριοποίηση των συναισθημάτων βάσει εικόνας, το πρώτο βήμα που υλοποιήθηκε ήταν να γίνει αναζήτηση στο διαδίκτυο για την επιλογή ενός κατάλληλου συνόλου δεδομένων, το οποίο θα τηρούσε όλες τις προδιαγραφές που απαιτούνταν. Μετά από αρκετή αναζήτηση, ξεχώρισαν δύο σύνολα δεδομένων, το Facial Emotion Recognition 2013 (FER-2013) και το Cohn-Kanade (CK). Και τα δύο αυτά σύνολα δεδομένων αποτελούνται από εικόνες 48x48 εικονοστοιχείων σε ασπρόμαυρη μορφή (απουσία χρώματος), οι οποίες παρουσιάζουν διάφορες εκφράσεις προσώπου και αντιστοιχούνται στα εξής συναισθήματα: χαρά, λύπη, θυμός, φόβος, έκπληξη, αποστροφή, ουδέτερο συναίσθημα.

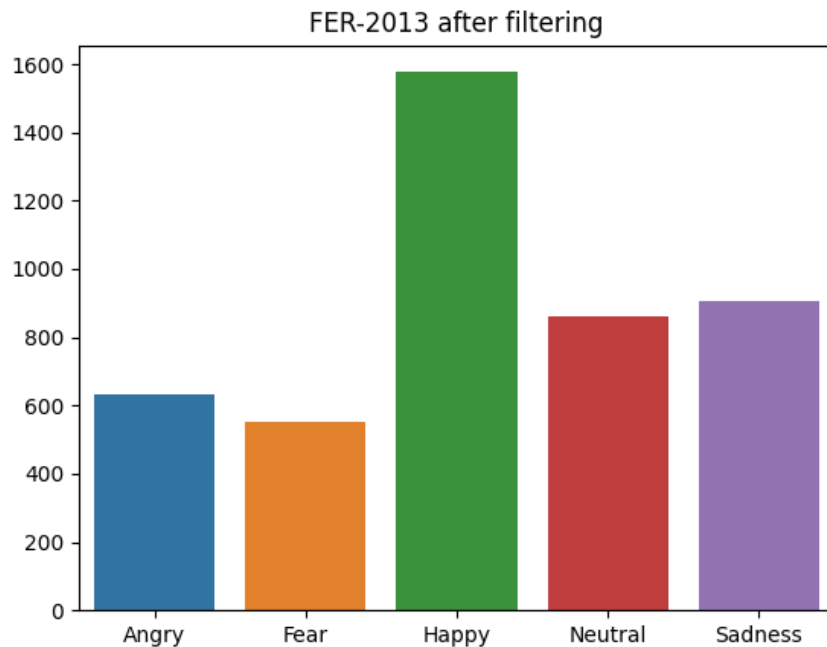
3.1.1 FER-2013

Το σύνολο δεδομένων Facial Emotion Recognition 2013 (FER-2013) δημιουργήθηκε από τους Pierre Luc Carrier και Aaron Courville και έκανε την εμφάνισή του στην δοκιμασία αναγνώρισης εκφράσεων προσώπου του εργαστηρίου ICML 2013. Διαμορφώθηκε χρησιμοποιώντας το API αναζήτησης εικόνων της Google για την ανίχνευση εικόνων που ταιριάζουν με συναισθηματικές λέξεις-κλειδιά. Συνολικά, αποτελείται από 35719 εικόνες προσώπων. Από αυτές, 4953 εικόνες εκφράζουν θυμό, 547 εικόνες εκφράζουν αποστροφή, 4953 εικόνες εκφράζουν φόβο, 8989 εικόνες εκφράζουν την ευτυχία, 6077 εικόνες εκφράζουν θλίψη, 4002 εικόνες εκφράζουν έκπληξη και 6198 εικόνες είναι συναισθηματικά ουδέτερες. Αξίζει να σημειωθεί πως στο συγκεκριμένο σύνολο δεδομένων τα πρόσωπα διαφέρουν πολύ ως προς την ηλικία, την στάση και στις συνθήκες που τραβήχτηκαν οι εικόνες, προσφέροντας έτσι ποικιλομορφία.



Εικόνα 7: Αρχική κατανομή συναισθημάτων του συνόλου δεδομένων FER-2013.

Παρόλα αυτά, επειδή ένα μέρος των εικόνων του προκαλούσαν άστοχες προβλέψεις, εφαρμόστηκε ένα φίλτρο ευθυγράμμισης προσώπου και όσες εικόνες δεν περνούσαν τον έλεγχο, απορρίφθηκαν. Έτσι το τελικό τροποποιημένο FER-2013 κατέληξε να έχει συνολικό αριθμό εικόνων 5751 με κατανομή όπως φαίνεται στην παρακάτω εικόνα.



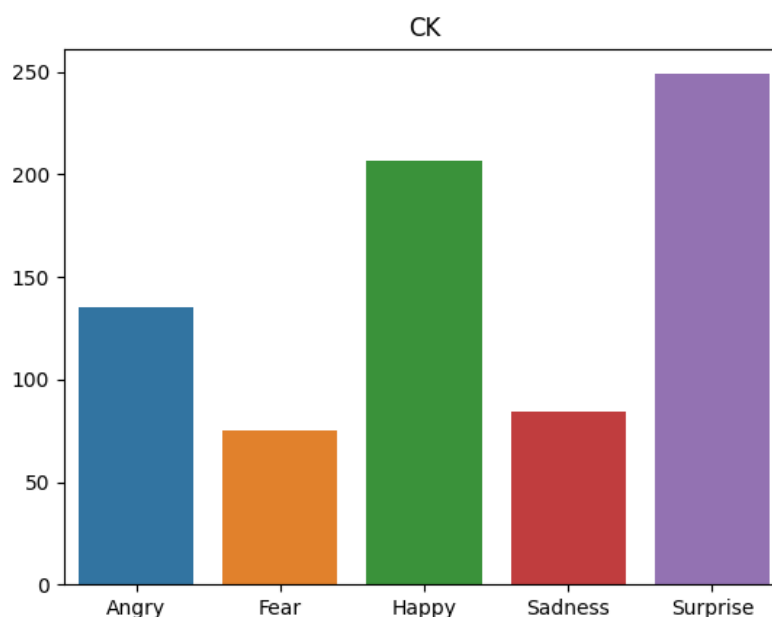
Εικόνα 8: Τελική κατανομή συναισθημάτων του συνόλου δεδομένων FER-2013 μετά το φιλτράρισμα.



Εικόνα 9: Δείγματα του συνόλου δεδομένων FER-2013 για το κάθε συναίσθημα.

3.1.2 Cohn-Kanade (CK)

Το 2000 το σύνολο δεδομένων Cohn-Kanade (CK) κυκλοφόρησε με σκοπό την προώθηση της έρευνας για την ανίχνευση συναισθημάτων μέσω προσώπου. Αποτελεί ένα από τα πιο ευρέως χρησιμοποιούμενα σύνολα δεδομένων για την ανάπτυξη και την αξιολόγηση αλγορίθμων. Για τη δημιουργία του CK χρησιμοποιήθηκαν εικόνες από 100 διαφορετικά άτομα και ηλικία των ατόμων που χρησιμοποιήθηκαν στις εικόνες κυμαίνονταν από 18 έως 30 ετών. Το 65% των ανθρώπων αποτελούνταν από γυναίκες, το 15% από αφροαμερικανούς και το 3% από Ασιάτες ή λατίνους. Στη θέση του ουδέτερου συναισθήματος υπάρχει το συναίσθημα της έκπληξης. Έχει αρκετά πιο μικρό μέγεθος συγκριτικά με το FER-2013, χρησιμοποιώντας μόλις 750 φωτογραφίες, εκ των οποίων 75 εκφράζουν φόβο, 135 θυμό, 207 χαρά, 84 λύπη και 249 έκπληξη.



Εικόνα 10: Κατανομή συναισθημάτων του συνόλου δεδομένων CK.

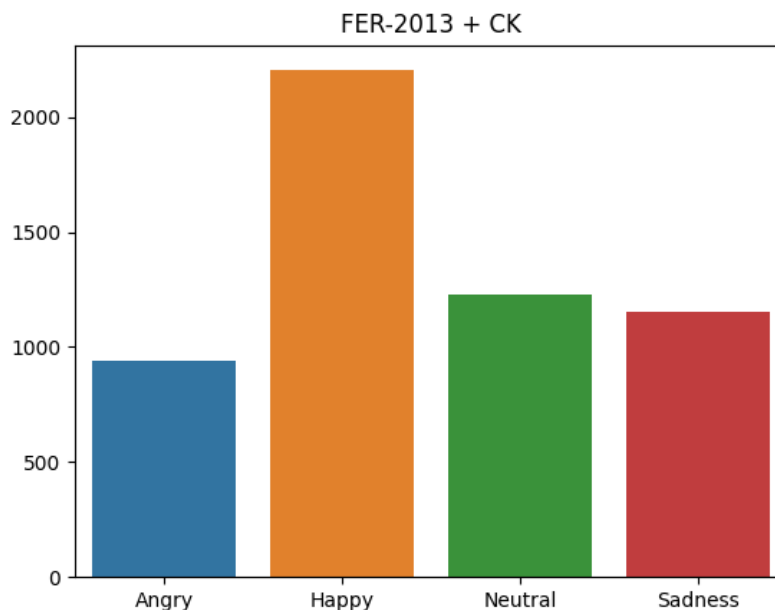


Εικόνα 11: Δείγματα του συνόλου δεδομένων CK για το κάθε συναίσθημα.

3.1.3 Τελικό σύνολο δεδομένων

Για την καλύτερη απόδοση στο μοντέλο μηχανικής μάθησης που χρησιμοποιήθηκε και επειδή και τα δύο σύνολα δεδομένων είχαν κοινά χαρακτηριστικά, προτιμήθηκε από το απορριφθεί κάποιο εκ των δύο να γίνει χρήση ενός κοινού συνόλου το οποίο αποτελούσε μίξη των προαναφερθέντων. Έτσι, προέκυψε ένα μεγαλύτερο σε πλήθος εικόνων σύνολο δεδομένων, πράγμα το οποίο βοήθησε στην βέλτιστη απόδοση του μοντέλου κατά την εκπαίδευση. Λόγω του αρκετά μικρού όγκου των συναισθημάτων του φόβου και της έκπληξης και λόγω της ομοιότητας με άλλες εκφράσεις,

προτιμήθηκε τα δύο αυτά συναισθήματα να μην συμπεριληφθούν στο τελικό σύνολο δεδομένων ώστε να βελτιστοποιηθεί η απόδοση του μοντέλου μάθησης.



Εικόνα 12: Κατανομή του συνόλου δεδομένων (FER-2013 + CK) που χρησιμοποιήθηκε στην μελέτη.

3.2 Προεπεξεργασία δεδομένων

Για να επιτευχθούν τα επιθυμητά αποτελέσματα κατά την εκπαίδευση ενός μοντέλου μηχανικής μάθησης, είναι σημαντικό τα δεδομένα τα οποία χρησιμοποιεί το μοντέλο να έχουν υποβληθεί σε προεπεξεργασία.

Χρησιμοποιώντας ένα σύνολο δεδομένων το οποίο αποτελείται από εικόνες, προκύπτουν κάποια προβλήματα σχετικά την πολυπλοκότητα και την ανακρίβεια τους. Για αυτό λοιπόν κρίνεται απαραίτητο οι εικόνες να καθαριστούν και να επεξεργαστούν στην επιθυμητή μορφή ώστε να μπορεί ο αλγόριθμος μηχανικής μάθησης να τις κατηγοριοποιήσει.

Να σημειωθεί πως στο σύνολο δεδομένων FER-2013 πραγματοποιήθηκε προεπεξεργασία, όπως αναφέρθηκε στην προηγούμενη ενότητα, χρησιμοποιώντας ένα φίλτρο ευθυγράμμισης προσώπου.

Πέραν της προεπεξεργασίας που πραγματοποιήθηκε στο FER-2013, σε όλες τις εικόνες του τελικού συνόλου δεδομένων εφαρμόστηκε κανονικοποίηση, μετατρέποντας όλες τις τιμές των εικονοστοιχείων των εικόνων από το εύρος 0-255 με φυσικούς αριθμούς στο οποίο βρίσκονταν, στο εύρος 0-1 με δεκαδικούς. Αυτό βοήθησε το μοντέλο, διότι το εύρος αυτό ήταν πιο κατανοητό σε αυτό και έτσι βελτιώθηκε ο ρυθμός εκπαίδευσής του.

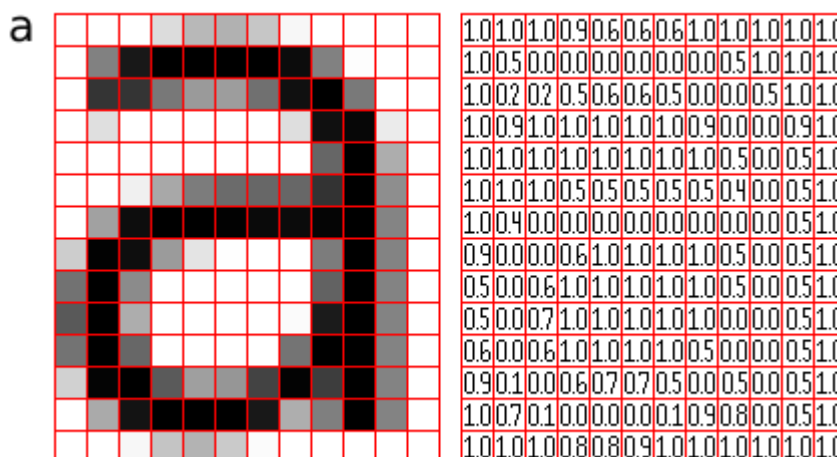
Άλλη μία τεχνική προεπεξεργασίας που εφαρμόστηκε στις εικόνες ήταν η επεξεργασία των εικόνων πραγματοποιώντας μικρές αλλαγές αναφορικά με τον προσανατολισμό τους, το γύρισμα τους είτε στον οριζόντιο είτε στον κάθετο άξονα και στο πόσο εύρος ζουμ. Έτσι, το μοντέλο είχε μεγαλύτερο σύνολο εκμάθησης, βελτιώνοντας την απόδοση του.

3.3 Τεχνική αναγνώρισης συναισθήματος

Το επόμενο βήμα μετά την προεπεξεργασία των εικόνων, ήταν να επιλεγθεί ο κατάλληλος αλγόριθμος κατηγοριοποίησης τους. Από τις διάφορες τεχνικές αναγνώρισης εικόνων που αναφέρθηκαν στο 2ο κεφάλαιο, προτιμήθηκε η χρήση ενός νευρωνικού δικτύου, και πιο συγκεκριμένα ενός συνελκτικού νευρωνικού δικτύου (Convolutional Neural Network – CNN).

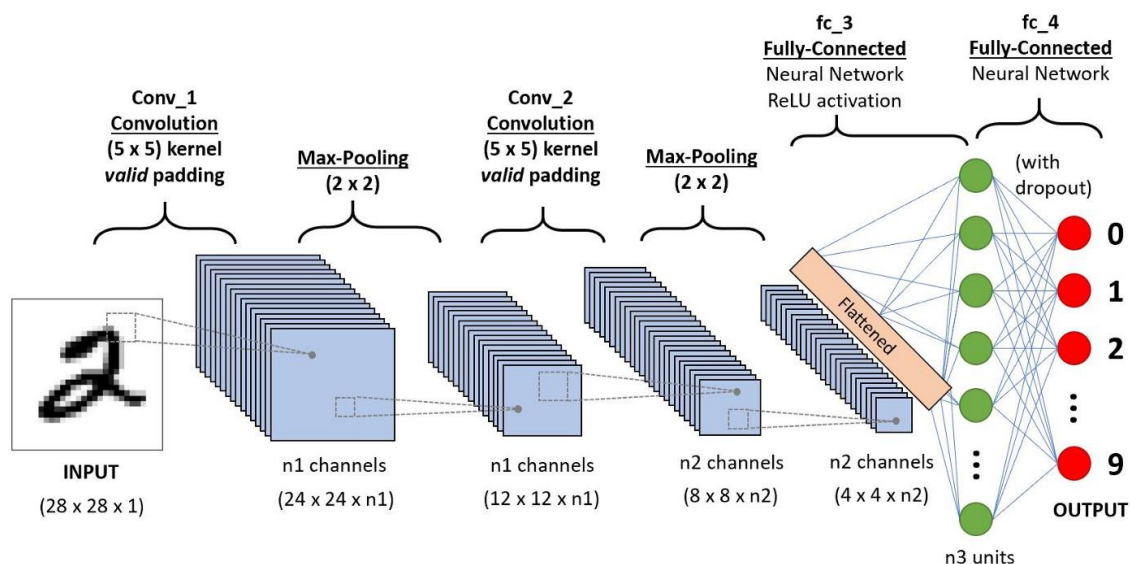
3.3.1 CNN

Το CNN είναι μια κατηγορία νευρωνικών δικτύων που ειδικεύονται στην επεξεργασία δεδομένων που έχουν μια τοπολογία σαν πλέγμα, όπως μια εικόνα. Μια ψηφιακή εικόνα είναι μια δυαδική αναπαράσταση οπτικών δεδομένων. Περιέχει μια σειρά εικονοστοιχείων διατεταγμένα με τρόπο πλέγματος που περιέχει τιμές εικονοστοιχείων για να υποδηλώσει πόσο φωτεινό και τι χρώμα πρέπει να είναι κάθε εικονοστοιχείο.



Εικόνα 13: Αναπαράσταση μιας εικόνας ως ένα πλέγμα εικονοστοιχείων.

Ο ανθρώπινος εγκέφαλος επεξεργάζεται έναν τεράστιο όγκο πληροφοριών τη στιγμή που βλέπει μια εικόνα. Κάθε νευρώνας λειτουργεί στο δικό του δεκτικό πεδίο και συνδέεται με άλλους νευρώνες με τρόπο που καλύπτουν ολόκληρο το οπτικό πεδίο. Όπως κάθε νευρώνας ανταποκρίνεται σε ερεθίσματα μόνο στην περιορισμένη περιοχή του οπτικού πεδίου που ονομάζεται δεκτικό πεδίο στο βιολογικό σύστημα όρασης, κάθε νευρώνας σε ένα CNN επεξεργάζεται δεδομένα μόνο στο δεκτικό του πεδίο επίσης. Τα στρώματα είναι διατεταγμένα με τέτοιο τρόπο ώστε να ανιχνεύουν πρώτα πιο απλά μοτίβα (γραμμές, καμπύλες κ.λπ.) και πιο σύνθετα μοτίβα (πρόσωπα, αντικείμενα κ.λπ.) στη συνέχεια. Χρησιμοποιώντας ένα CNN, μπορεί κανείς να ενεργοποιήσει την όραση σε υπολογιστές.



Εικόνα 14: Μια CNN ακολουθία κατηγοριοποίησης χειρόγραφων ψηφίων (Di Guan, 2020).

3.3.2 Εκπαίδευση νευρωνικού δικτύου

Στο στάδιο αυτό, τα δεδομένα τα οποία είχαν προεπεξεργαστεί με τις μεθόδους που προαναφέρθηκαν χωρίστηκαν σε δύο υποσύνολα, στο σύνολο εκπαίδευσης (training set) και στο σύνολο επικύρωσης (validation set). Το CNN τροφοδοτούνταν με το σύνολο εκπαίδευσης και αξιολογούσε τα δεδομένα του επανειλημμένα για να μάθει περισσότερα για τη συμπεριφορά τους και στη συνέχεια προσαρμοζόταν ώστε να εξυπηρετήσει τον επιδιωκόμενο σκοπό του. Κατά τη διάρκεια της εκπαίδευσης, το σύνολο επικύρωσης εισήγαγε νέα δεδομένα στο CNN που δεν είχε αξιολογήσει πριν. Τα δεδομένα αυτού του συνόλου παρείχαν την πρώτη δοκιμή έναντι δεδομένων που δεν είχαν δοκιμαστεί από το CNN, επιτρέποντας έτσι την αξιολόγηση του κατά τη διάρκεια της εκπαίδευσης του. Η χρήση του συνόλου επικύρωσης δεν κρίνεται απαραίτητη για την εκπαίδευση ενός μοντέλου, αλλά μπορεί να παρέχει χρήσιμες πληροφορίες για την βελτιστοποίηση των παραμέτρων του μοντέλου, οι οποίες επηρεάζουν τον τρόπο που αξιολογεί τα δεδομένα.

3.4 Αξιολόγηση αποτελεσμάτων

Για την αξιολόγηση της απόδοσης του εκπαιδευμένου μοντέλου χρησιμοποιήθηκαν ειδικές μετρικές οι οποίες επηρεάζουν τα συμπεράσματα που προκύπτουν από αυτό. Οι μετρικές που χρησιμοποιήθηκαν για το συγκεκριμένο μοντέλο ήταν οι εξής:

Confusion matrix:

Είναι ένας πίνακας ο οποίος περιγράφει την πλήρη απόδοση του μοντέλου. Θεωρώντας ένα πρόβλημα δυαδικής ταξινόμησης, υπάρχουν δεδομένα τα οποία ανήκουν σε δύο κατηγορίες: ΝΑΙ ή ΟΧΙ. Μετά το πέρασμα των δεδομένων από ένα μοντέλο με 165 δείγματα, προκύπτει το ακόλουθο confusion matrix.

		Predicted: NO	Predicted: YES
n=165	Actual: NO	50	10
	Actual: YES	5	100

Εικόνα 15: Παράδειγμα confusion matrix.

Υπάρχουν τέσσερις σημαντικοί όροι οι οποίοι χρήζουν ανάλυσης:

- Αληθώς θετικά (True Positives): Οι περιπτώσεις στις οποίες η πρόβλεψη ήταν ΝΑΙ και το πραγματικό αποτέλεσμα επίσης ΝΑΙ.
- Αληθώς αρνητικά (True Negatives): Οι περιπτώσεις στις οποίες η πρόβλεψη ήταν ΟΧΙ και το πραγματικό αποτέλεσμα επίσης ΟΧΙ.
- Ψευδώς θετικά (False Positives): Οι περιπτώσεις στις οποίες η πρόβλεψη ήταν ΝΑΙ και το πραγματικό αποτέλεσμα ΟΧΙ.
- Ψευδώς αρνητικά (False Negatives): Οι περιπτώσεις στις οποίες η πρόβλεψη ήταν ΟΧΙ και το πραγματικό αποτέλεσμα ΟΧΙ.

Το confusion matrix θέτει τις βάσεις για τις υπόλοιπες μετρικές.

Ορθότητα (accuracy):

Η ακρίβεια είναι ίσως η πιο απλή μετρική στη χρήση και ορίζεται ως ο αριθμός των σωστών προβλέψεων (True Positives + True Negatives) προς τον αριθμό όλων των προβλέψεων. Ο γενικευμένος της τύπος είναι ο εξής:

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

Ακρίβεια (precision):

Ορίζεται ως ο αριθμός των True Positives προς τον συνολικό αριθμό της θετικής κλάσης (positives). Μια μετρική precision προς το 1 σημαίνει ότι το μοντέλο δεν έχασε κανένα αληθινά θετικό δεδομένο (TP) και είναι σε θέση να ταξινομήσει καλά τις δύο κλάσεις. Αυτό το οποίο δεν μπορεί να προσδιορίσει είναι τον αριθμό των ψευδώς θετικών αποτελεσμάτων. Εάν η μετρική πάρει τιμή μικρότερη του 0.5, σημαίνει ότι το μοντέλο έχει μεγάλο αριθμό ψευδώς θετικών που μπορεί να είναι αποτέλεσμα ανισορροπίας στις κατηγορίες των δεδομένων.

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

Ανάκληση (recall):

Είναι η αναλογία των αληθώς θετικών προς όλων των θετικών. Μια μετρική recall προς το 1 σημαίνει ότι το μοντέλο δεν έχασε κανένα αληθώς θετικό δεδομένο και είναι σε θέση να ταξινομήσει καλά και τις δύο κλάσεις. Δεν μπορεί να προσδιορίσει το σφάλμα στον αριθμό των ψευδώς θετικών. Μια χαμηλή τιμή recall μικρότερη του 0.5, όπως και στην μετρική precision, ενδεχομένως να υποδηλώνει ανισορροπία στις κατηγορίες των δεδομένων.

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

F1-score:

Η μετρική f1-score συνδυάζει τις μετρικές precision και recall, υπολογίζοντας τον αρμονικό μέσο των δύο. Μια υψηλή τιμή της μετρικής συμβολίζει υψηλή τιμή τόσο στην μετρική precision, όσο και στην μετρική recall. Το αντίθετο όμως δεν μπορεί να θεωρηθεί δεδομένο, καθώς ενδέχεται να μία από τις δύο μετρικές χαμηλή, ή και οι δύο. Ο μαθηματικός της τύπος είναι ο εξής:

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$$

Καθώς στο συγκεκριμένο σύνολο δεδομένων υπήρχαν τέσσερις κλάσεις, χρησιμοποιήθηκε η μέθοδος μέσου όρου macro average η οποία μετατρέπει την κατηγοριοποίηση πολλών κλάσεων σε ισάριθμες δυαδικές κλάσεις. Έπειτα υπολογίζει τις μετρικές precision, recall και f1-score ξεχωριστά και τέλος εξάγει τον μέσο όρο τους.

3.5 Αξιοποίηση αποτελεσμάτων

Μετά την ολοκλήρωση της εκπαίδευσης του νευρωνικού δικτύου, δημιουργήθηκε ένα γραφικό περιβάλλον στο οποίο φαινόταν η λειτουργία του μοντέλου. Πιο συγκεκριμένα, επειδή το τελικό σχέδιο ήταν το μοντέλο να εφαρμοστεί πάνω σε βίντεο από μια κάμερα υπολογιστή ώστε να προκύπτουν προβλέψεις σε ζωντανό χρόνο, δημιουργήθηκε ένα απλό παράθυρο το οποίο έδειχνε την εικόνα της κάμερας. Επειδή το μοντέλο χρειαζόταν εικόνες για να κάνει πρόβλεψη, το δείγμα της κάμερας χωρίστηκε σε καρέ και έκανε περίπου 10 προβλέψεις ανά δευτερόλεπτο. Το κάθε καρέ όμως δεν περιείχε απομονωμένο το πρόσωπο, όπως ήταν οι εικόνες του συνόλου δεδομένων. Για αυτό τον λόγο, χρησιμοποιήθηκε ένα προεκπαιδευμένο νευρωνικό δίκτυο εντοπισμού προσώπου, το dnn detector. Μέσω αυτού, γινόταν η απομόνωση του προσώπου και ζωγραφίζονταν στην περιοχή ενδιαφέροντος ένα τετράγωνο ώστε να είναι πιο ευδιάκριτη στον χρήστη το δείγμα που δεχόταν το μοντέλο για την πρόβλεψη του συναίσθηματος. Το αποτέλεσμα της πρόβλεψης

εμφανιζόταν στο ίδιο παράθυρο με αναλυτικό τρόπο, δείχνοντας ποσοστά για το κάθε συναίσθημα που εντόπιζε.

Σκοπός όμως ήταν το πρόγραμμα που συντάχθηκε να είναι αυτοτελές ώστε να μπορεί να μεταφερθεί και προσαρμοστεί σε οποιαδήποτε πλατφόρμα με εύκολο τρόπο. Έτσι λοιπόν, χρησιμοποιήθηκε ένα εργαλείο, το pyinstaller, το οποίο επέτρεπε την μετατροπή όλου του προγράμματος σε ένα εκτελέσιμο αρχείο, έτσι ώστε να μπορεί ο χρήστης να το χρησιμοποιήσει χωρίς να χρειάζεται να έχει εγκατεστημένα τα εργαλεία που χρησιμοποιήθηκαν για την σύνταξη του.

Όταν ο χρήστης έτρεχε το εκτελέσιμο αρχείο, το άνοιγε το παράθυρο σε μια μικρή διάσταση και συνέχιζε τη δουλειά του, αφήνοντας το να κάνει τις προβλέψεις του στο παρασκήνιο. Το πρόγραμμα τερμάτιζε, όταν ο χρήστης πατούσε το πλήκτρο 'Q' ή όταν έκλεινε το παράθυρο.

3.6 Αλληλεπίδραση διεπαφής με τον χρήστη

Πέραν της αλληλεπίδρασης της διεπαφής με τον χρήστη μέσω της πρόβλεψης, η οποία ήταν οπτική αλληλεπίδραση, προστέθηκαν ακόμα δύο τρόποι, ένας βασισμένος στην οπτική και ένας στην ακουστική αλληλεπίδραση.

Οπτική αλληλεπίδραση

Όπως αναφέρθηκε, πέραν της προβολής της αναγνώρισης του συναισθήματος στο παράθυρο του προγράμματος, χρησιμοποιήθηκαν και 3 μηνύματα για κάθε συναίσθημα (εκτός του ουδέτερου που χρησιμοποιήθηκε σαν μεταβατικό στάδιο για τα συναισθήματα) τα οποία είχαν ως σκοπό την ενθάρρυνση του χρήστη να συνεχίσει την δουλειά του. Τα μηνύματα αυτά είχαν ουδέτερο περιεχόμενο και δύναται προσαρμογή τους, ανάλογα με το πεδίο εφαρμογής του προγράμματος. Σε ένα ενδεχόμενο σενάριο κατά το οποίο ο χρήστης συντάσσει κώδικα, το πρόγραμμα εντοπίζοντας ότι είναι λυπημένος, του εμφανίζει ένα μήνυμα που είναι σχετικό με το συναίσθημα και έχει ως σκοπό να τον παροτρύνει να συνεχίσει τη δουλειά του.

Ακουστική αλληλεπίδραση

Τέλος, για να προστεθεί ακόμα ένας τρόπος αλληλεπίδρασης, χρησιμοποιήθηκαν 3 μουσικά κομμάτια, ένα για το κάθε συναίσθημα (εκτός του ουδέτερου) τα οποία επιτελούσαν τον ίδιο σκοπό με τα προαναφερθέντα ενθαρρυντικά μηνύματα. Όπως στα μηνύματα, έτσι και τα μουσικά κομμάτια είχαν σχετικά ουδέτερο χαρακτήρα και μπορούν να αλλαχθούν, ανάλογα με το πεδίο εφαρμογής του προγράμματος.

4 ΚΕΦΑΛΑΙΟ 4^ο : Αρχιτεκτονική και Περιγραφή Συστήματος

Όλα τα αρχεία κώδικα συντάχθηκαν χρησιμοποιώντας τη γλώσσα Python (έκδοση 3.7.4). Η Python είναι αρκετά σταθερή και ευέλικτη για προγράμματα μηχανικής μάθησης και τεχνητής νοημοσύνης. Χρησιμοποιώντας τις κατάλληλες βιβλιοθήκες, οι σύνθετοι αλγόριθμοι που βρίσκονται πίσω από τη μηχανική μάθηση απλοποιούνται σε τέτοιο βαθμό, που καθιστούν αρκετά εύκολη την ανάπτυξη ενός μοντέλου εκπαίδευσης. Οι βιβλιοθήκες και οι εκδόσεις τους που χρησιμοποιήθηκαν στο πρόγραμμα φαίνονται πιο αναλυτικά στον παρακάτω πίνακα.

Library	version
joblib	1.1.0
keras	2.7.0
tensorflow	2.7.0
scikit-learn	1.0.1
opencv-python	4.2.0.34
numpy	1.21.4
dlib	19.19.0
imutils	0.5.4
matplotlib	3.5.0
python-vlc	3.0.12118
seaborn	0.11.2
scikit-plot	0.3.7

Πίνακας 2: Εκδόσεις βιβλιοθηκών που χρησιμοποιήθηκαν.

Για την εκπαίδευση του νευρωνικού δικτύου, χρησιμοποιήθηκαν οι πόροι κάρτας γραφικών (GPU) χρησιμοποιώντας το λογισμικό Cuda της εταιρίας Nvidia. Προτιμήθηκε έναντι της χρήσης των πόρων του επεξεργαστή (CPU) διότι η ικανότητα της να κάνει τις κατάλληλες πράξεις που απαιτούνται κατά την εκπαίδευση επιτρέπει την πιο γρήγορη απόδοση του νευρωνικού δικτύου σε σχέση με την CPU.

4.1 Προεπεξεργασία και διαχείριση δεδομένων

Για την προεπεξεργασία και τη διαχείριση του συνόλου δεδομένων, συντάχθηκε αρχικά ένας κώδικας με όνομα `data_builder.py`. Ο κώδικας αυτός περιείχε δύο κλάσεις, την `DataBuilder` και την `ImagetoArray` οι οποίες ήταν υπεύθυνες για όλη τη διαχείριση του συνόλου δεδομένων, το διάβασμα του από τους φακέλους στο οποίο ήταν χωρισμένο σε κατηγορίες, και την προεπεξεργασία του σε τέτοια μορφή ώστε να μπορεί να την αναγνωρίσει το νευρωνικό δίκτυο.

Η βασική λειτουργία της κλάσης ήταν η συνάρτηση `build_from_directory`, η οποία όπως υποδηλώνει το όνομα, μετέτρεπε το σύνολο δεδομένων από τον βασικό φάκελο σε κατάλληλη

μορφή πίνακα και επέστρεφε τις εικόνες σε αυτήν τη μορφή, τις ετικέτες τους (δηλαδή το συναίσθημα που περιέγραφε η κάθε φωτογραφία) και ένα dictionary με τις ετικέτες σε μορφή κειμένου.

Συνάρτηση build from directory

Το σύνολο δεδομένων ήταν χωρισμένο σε τέσσερις υποφακέλους, έναν για κάθε συναίσθημα που χρησιμοποιήθηκε. Τα ονόματα των φακέλων είχαν το όνομα κάθε συναισθήματος για την ευκολότερη διαχείριση στην προεπεξεργασία. Η συνάρτηση `build_from_directory` για κάθε υποφάκελο του συνόλου δεδομένων διάβαζε τις εικόνες χρησιμοποιώντας εργαλεία της βιβλιοθήκης `opencv` που διαχειρίζεται εικόνες, και με τη χρήση μαθηματικών πράξεων της βιβλιοθήκης `numpy` που έχει πληθώρα συναρτήσεων για τέτοιες χρήσεις, μετέτρεπε τις εικόνες σε διδιάστατο πίνακα. Στο dictionary που είχε δημιουργηθεί για τις ετικέτες των εικόνων, αποθηκεύονταν στην κατάλληλη θέση το συναίσθημα της κάθε εικόνας.

Τέλος, μετά την προσπέλαση όλων των εικόνων, γινόταν μετατροπή των λιστών που διαχειρίζονταν τις εικόνες και τις ετικέτες τους σε μορφή ειδικού πίνακα χρησιμοποιώντας τη βιβλιοθήκη `numpy`, ώστε να μπορεί το νευρωνικό δίκτυο να διαχειριστεί τις εικόνες και να τις αναγνωρίσει. Επειδή το νευρωνικό δίκτυο δεν μπορεί να αναγνωρίσει τις ετικέτες των εικόνων σαν κείμενο, χρησιμοποιήθηκε η συνάρτηση `OneHotEncoder` η οποία μετέτρεπε το κείμενο, σε αριθμούς σε μορφή λίστας.

Για κάθε εικόνα, στη θέση της ετικέτας αντί για το όνομα του συναισθήματος, υπήρχε τώρα μια λίστα με τέσσερις θέσεις, μία για το κάθε συναίσθημα και τοποθετούνταν άσσος στο συναίσθημα που ανήκε η εικόνα και μηδενικά στα υπόλοιπα. Η μετατροπή αυτή πραγματοποιήθηκε, διότι το πρόβλημα της κατηγοριοποίησης των συναισθημάτων ήταν πολλαπλών ετικετών (`multi label`) το οποίο σημαίνει ότι το νευρωνικό δίκτυο μπορούσε να προβλέψει πως μία εικόνα ανήκει σε παραπάνω από ένα συναισθήματα. Όπως αναφέρθηκε παραπάνω, κανονικά σε μια εικόνα μπαίνει άσσος στην θέση της λίστας για το συναίσθημα στο οποίο ανήκει και μηδενικά στα υπόλοιπα. Σε μία πρόβλεψη όμως, μια εικόνα ενδέχεται να προβλεφθεί πως ανήκει σε δύο ή και παραπάνω συναισθήματα. Έτσι, μια πρόβλεψη μπορεί να πάρει τιμές στο εύρος 0-1 για όλα τα συναισθήματα, των οποίων το άθροισμα θα ισούται με τον άσσο.

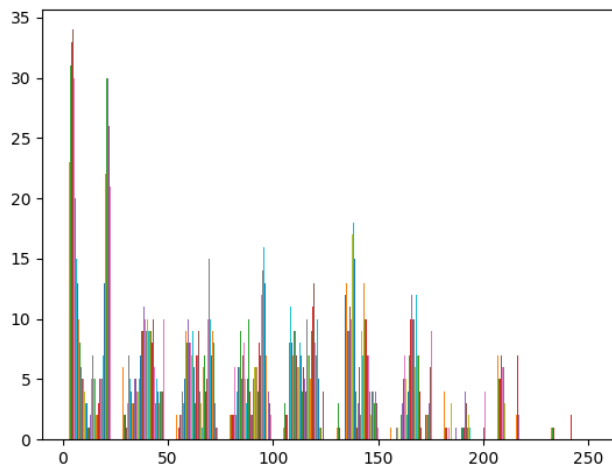
Το τελευταίο κομμάτι προεπεξεργασίας πραγματοποιήθηκε στον κώδικα `trainer.py` ο οποίος ήταν υπεύθυνος για την εκπαίδευση του νευρωνικού δικτύου. Αρχικά, χρησιμοποιήθηκε η συνάρτηση `ImageDataGenerator` η οποία δημιούργησε ένα επιπλέον πακέτο δεδομένων με διαφορές στις εικόνες ως προς την περιστροφή τους, το εύρος ζουμ, το γύρισμα τους στον οριζόντιο άξονα, και την αλλαγή τους στους δύο άξονες. Η τεχνική αυτή εφαρμόστηκε ώστε να μεγαλώσει το σύνολο δεδομένων με τρόπο κατάλληλο ώστε να μην μειωθεί η απόδοση του νευρωνικού δικτύου. Ουσιαστικά, δημιουργήθηκαν αντίγραφα των δεδομένων με μικρές παραμετροποιήσεις οι οποίες βελτίωσαν την απόδοση, προσφέροντας ποικιλομορφία.

Η τελευταία διαδικασία που εκτελέστηκε για την προεπεξεργασία ήταν να κανονικοποιηθεί η είσοδος του νευρωνικού δικτύου. Μετά την κλήση της συνάρτησης `build_from_directory`, η λίστα που περιείχε τις εικόνες στην κατάλληλη μορφή υπέστη κανονικοποίηση, διαιρώντας όλες τις τιμές των εικονοστοιχείων κάθε εικόνας με τον αριθμό 255. Η τιμή αυτή επιλέχθηκε, διότι οι τιμές των εικονοστοιχείων κάθε εικόνας κυμαίνονται στο εύρος 0-255, όπου το 0 ταυτίζεται με το μαύρο και το 255 με το άσπρο (οι ενδιάμεσες τιμές είναι όλες αποχρώσεις του γκρι). Διαιρώντας με το 255, το

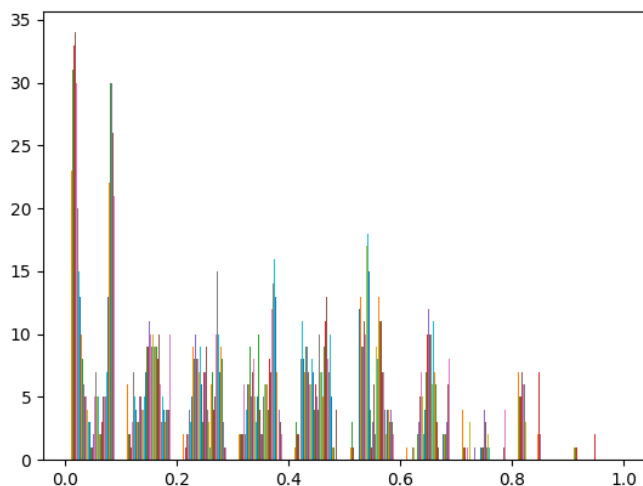
εύρος από 0-255, άλλαξε στο εύρος 0-1 όπου οι ενδιάμεσες τιμές ήταν δεκαδικοί αριθμοί. Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, η κανονικοποίηση πραγματοποιήθηκε διότι συμβάλλει στην μείωση της πολυπλοκότητας των πράξεων και του πλεονασμού των πόρων. Ο μαθηματικός τύπος της κανονικοποίησης είναι ο εξής (όπου x_{min} και x_{max} μπόκαν τα 0 και 255 αντίστοιχα):

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Στις παρακάτω δύο εικόνες φαίνονται τα ιστογράμματα μιας τυχαίας εικόνας από το σύνολο δεδομένων σε κανονική μορφή και έπειτα με κανονικοποίηση. Το ιστόγραμμα είναι η γραφική απεικόνιση της συχνότητας περιοχών τιμών ενός μεγέθους. Πιο συγκεκριμένα, απεικονίζονται οι συχνότητες εμφάνισης των τιμών των εικονοστοιχείων. Στην πρώτη εικόνα φαίνεται πως το εύρος των τιμών κυμαίνεται στο 0-255, ενώ στην δεύτερη εικόνα που έχει υποστεί την κανονικοποίηση το εύρος κυμαίνεται στο 0-1.



Εικόνα 16: Ιστόγραμμα μιας τυχαίας εικόνας του συνόλου δεδομένων.



Εικόνα 17: Ιστόγραμμα μιας τυχαίας εικόνας του συνόλου δεδομένων μετά την κανονικοποίηση.

4.2 Εκπαίδευση νευρωνικού δικτύου

Ο κύριος κώδικας που πραγματοποιούσε την εκπαίδευση του νευρωνικού δικτύου ήταν ο `trainer.py`. Ήταν υπεύθυνος για το διάβασμα των δεδομένων, την προεπεξεργασία τους, την εκπαίδευση του δικτύου και την αποθήκευση του μαζί με τα αποτελέσματα της επίδοσης του.

Ο δεύτερος κώδικας που συντάχθηκε για την εκπαίδευση ήταν ο `models.py`. Περιείχε δύο κλάσεις, την `BaseModel` και την `CNNModel`. Η `BaseModel` που περιέχει όλες τις απαραίτητες συναρτήσεις για την πρόβλεψη, την αξιολόγηση και την αποθήκευση των αποτελεσμάτων του μοντέλου. Περιέχει επίσης την βασική συνάρτηση που κατασκευάζει τη δομή του νευρωνικού δικτύου με όλα τα επίπεδα.

Η κλάση `CNNModel` βασίζεται στην προηγούμενη κλάση και περιέχει την συνάρτηση κατασκευής του νευρωνικού δικτύου η οποία καλεί την συνάρτηση της `BaseModel` και την συνάρτηση της εκπαίδευσης.

Ο κώδικας `models.py` συντάχθηκε με τη δομή που προαναφέρθηκε, διότι με αυτόν τον τρόπο υπάρχει η δυνατότητα να προστεθούν επιπλέον μοντέλα, προσθέτοντας απλά μια κλάση. Έτσι, μελλοντικά αν βρεθεί κάποιο νέο μοντέλο που παράγει καλύτερα αποτελέσματα, δε χρειάζεται να αφαιρεθεί το προηγούμενο μοντέλο.

4.2.1 Κλάση `BaseModel`

Η κλάση `BaseModel` περιέχει τον κατασκευαστή της που δηλώνει μια κενή μεταβλητή με όνομα `model` για την αποθήκευση του μοντέλου και μια `boolean` μεταβλητή με όνομα `trained` για τον

έλεγχο πραγματοποίησης της εκπαίδευσης. Περιέχει επίσης δύο συναρτήσεις `model_builder` και `train` οι οποίες αντικαθίστανται από τις κλάσεις που κληρονομούν την `BaseModel`.

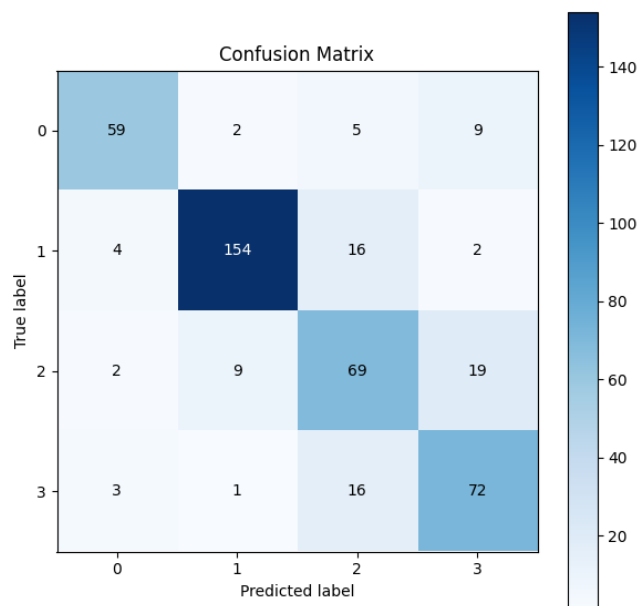
Συνάρτηση evaluate:

Η συνάρτηση αυτή είναι υπεύθυνη για την αξιολόγηση του μοντέλου. Δέχεται δύο παραμέτρους, το `X_test` και το `y_test` που αποτελούν υποσύνολα του αρχικού συνόλου δεδομένων. Το `X_test` είναι οι εικόνες οι οποίες έχουν υποστεί προεπεξεργασία και το `y_test` οι ετικέτες τους. Εκτελώντας μία πρόβλεψη χρησιμοποιώντας το `X_test`, το μοντέλο παράγει τα αποτελέσματα του και έπειτα από κάποιες μετατροπές ώστε να ακολουθεί τη δομή της μεταβλητής `y_test`, υπολογίζεται η ακρίβεια του.

Τέλος, τυπώνεται στο τερματικό η αξιολόγηση του μοντέλου χρησιμοποιώντας τις μετρικές `accuracy`, `precision`, `recall` και `f1-score` για κάθε συναίσθημα ξεχωριστά και έπειτα τους μέσους όρους. Το `support` είναι ο αριθμός των δειγμάτων που χρησιμοποιήθηκαν. Με τη χρήση της βιβλιοθήκης `matplotlib` σχεδιάζεται το `confusion matrix` του μοντέλου και αποθηκεύεται σε αρχείο εικόνας.

test accuracy: 78.0362%				
	precision	recall	f1-score	support
0	0.81	0.76	0.78	131
1	0.92	0.88	0.90	309
2	0.60	0.77	0.67	172
3	0.75	0.62	0.68	162
accuracy			0.78	744
macro avg	0.77	0.76	0.76	744
weighted avg	0.79	0.78	0.78	744
{0: 'Angry', 1: 'Happy', 2: 'Neutral', 3: 'Sadness'}				

Πίνακας 3: Δείγμα αναλυτικής αξιολόγησης του μοντέλου.



Εικόνα 18: Δείγμα αναλυτικής αξιολόγησης του μοντέλου.

Συνάρτηση predict:

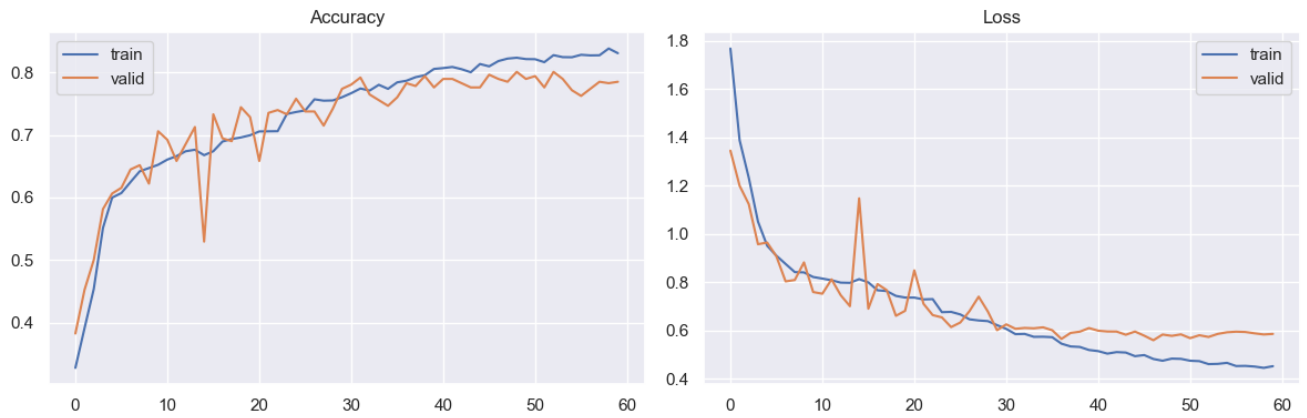
Πραγματοποιεί την πρόβλεψη δοθέντος ενός υποσυνόλου εκπαίδευσης X. Καλείται από τη συνάρτηση αξιολόγησης και από τον κώδικα που υλοποιεί την εφαρμογή ώστε να πραγματοποιήσει πρόβλεψη σε κάθε καρέ από το δείγμα της κάμερας. Επιστρέφει την πρόβλεψη στην κατάλληλη μορφή που απαιτείται από το πρόγραμμα.

Συνάρτηση save_model:

Αποθηκεύει το μοντέλο στην κατάλληλη μορφή ώστε να χρησιμοποιεί από τον κώδικα που υλοποιεί την εφαρμογή.

Συνάρτηση save_training_history:

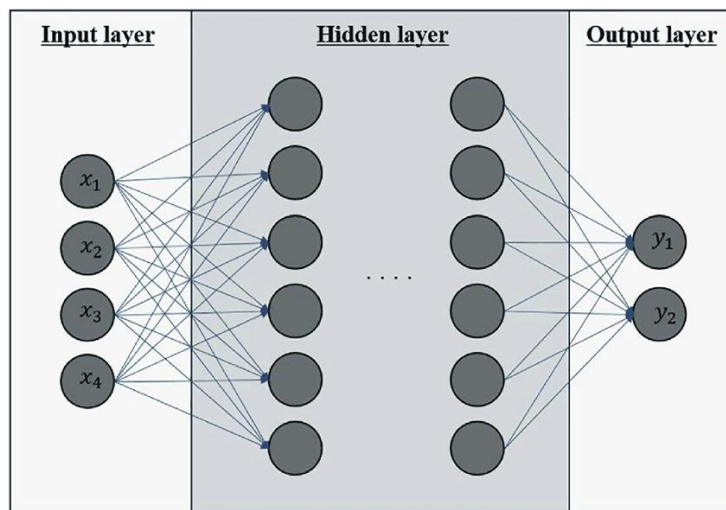
Χρησιμοποιώντας την βιβλιοθήκη matplotlib σχεδιάζει ένα αναλυτικό διάγραμμα της πορείας της εκπαίδευσης του μοντέλου. Στο αριστερό τμήμα αναγράφεται η ορθότητα του μοντέλου για το κάθε στάδιο της εκπαίδευσης. Με μπλε γραμμή σχεδιάζεται η γραμμή πορείας της ορθότητας των δεδομένων εκπαίδευσης, ενώ με πορτοκαλί η γραμμή πορείας της ορθότητας των δεδομένων αξιολόγησης. Στο δεξιό τμήμα σχεδιάζονται οι αντίστοιχες γραμμές πορείας του μοντέλου για την απώλεια του μοντέλου σε κάθε στάδιο.



Εικόνα 19: Ιστορικό ορθότητας και απώλειας του μοντέλου.

Συνάρτηση cnn for raw img:

Είναι υπεύθυνη για την κατασκευή του συνελκτικού νευρωνικού δικτύου. Ένα νευρωνικό δίκτυο αποτελείται από τρία στρώματα (layers), το στρώμα εισόδου (input layer), τα κρυφά στρώματα (hidden layers) και το στρώμα εξόδου (output layer). Οι εισοδοί εισέρχονται στο input layer και κάθε κόμβος παρέχει μια τιμή εξόδου μέσω μιας συνάρτησης ενεργοποίησης. Οι εξοδοί του input layer χρησιμοποιούνται ως εισοδοί στο επόμενο hidden layer.



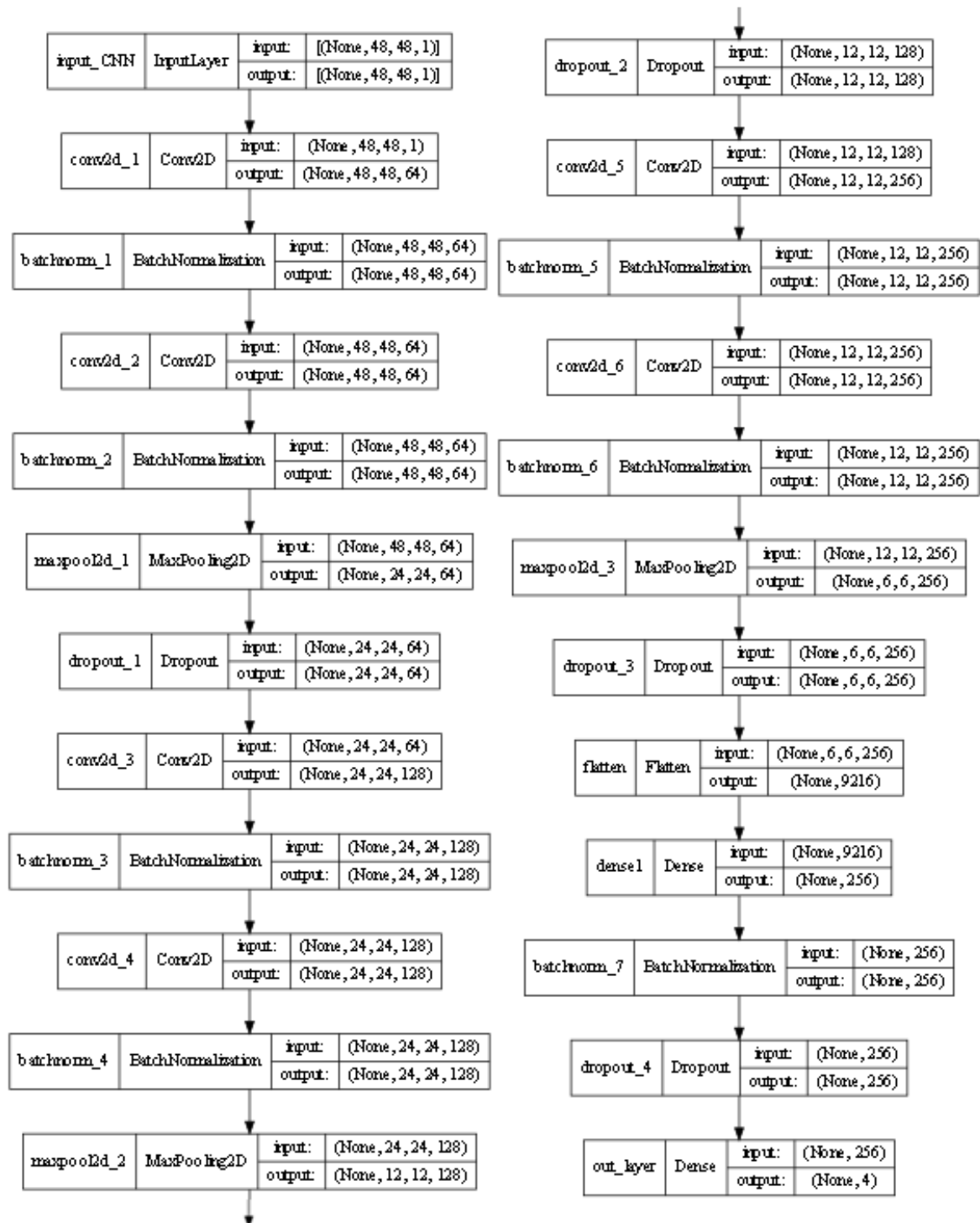
Εικόνα 20: Τα τρία στρώματα ενός νευρωνικού δικτύου (Pygo, 2017).

Στα hidden layers χρησιμοποιήθηκαν έξι διαφορετικών ειδών layers τα οποία περιγράφονται παρακάτω:

- Conv2D layer: Το layer αυτό δημιουργεί έναν συνελκτικό πυρήνα και κάνει συνέλιξης με το input layer. Το αποτέλεσμα της συνέλιξης δίνεται σαν είσοδος στο επόμενο layer το οποίο είναι το batch normalization.
- BatchNormalization layer: εφαρμόζει έναν μετασχηματισμό που διατηρεί τη μέση έξοδο κοντά στο 0 και την τυπική απόκλιση εξόδου κοντά στο 1. Κατά την εκπαίδευση το layer

αυτό κανονικοποιεί την έξοδο του χρησιμοποιώντας τη μέση και την τυπική απόκλιση της τρέχουσας παρτίδας (batch) εισόδων. Η έξοδος του δίνεται σαν είσοδος είτε σε ένα Conv2d layer είτε σε ένα MaxPooling2D layer.

- MaxPooling2D layer: Μειώνει δειγματοληπτικά την είσοδο του κατά μήκος των χωρικών του διαστάσεων (ύψος και πλάτος) λαμβάνοντας τη μέγιστη τιμή σε ένα παράθυρο εισόδου για κάθε κανάλι της εισόδου. Η έξοδος του δίνεται σαν είσοδος σε ένα dropout layer.
- Dropout layer: Ορίζει τυχαία τις μονάδες εισόδου στο 0 με συχνότητα ρυθμού σε κάθε βήμα κατά τη διάρκεια του χρόνου εκπαίδευσης, κάτι που βοηθά στην αποφυγή υπερβολικής προσαρμογής (overfitting). Οι εισοδοί που δεν έχουν οριστεί στο 0 κλιμακώνονται κατά $\frac{1}{1-\text{ρυθμός}}$ έτσι ώστε το άθροισμα όλων των εισόδων να παραμένει αμετάβλητο. Η έξοδος του δίνεται σαν είσοδος είτε σε ένα Conv2d layer είτε σε ένα flatten layer.
- Flatten layer: Χρησιμοποιείται για την εξομάλυνση της εισόδου. Μετατρέπει τις διαστάσεις της εισόδου σε μία. Για παράδειγμα αν οι διαστάσεις της εισόδου είναι [2,2], οι διαστάσεις της εξόδου θα γίνουν [4]. Η έξοδος δίνεται σε ένα dense layer.
- Dense layer: Είναι ένα layer το οποίο συνδέεται βαθιά, πράγμα που σημαίνει ότι κάθε νευρώνας στο dense layer λαμβάνει είσοδο από όλους τους νευρώνες του προηγούμενου layer. Εκτελεί ένα πολλαπλασιασμό πίνακα-διανύσματος. Οι τιμές που χρησιμοποιούνται στον πίνακα είναι στην πραγματικότητα παράμετροι που μπορούν να εκπαιδευτούν και να ενημερωθούν με τη βοήθεια της οπισθοδιάδοσης. Η βασική του χρήση είναι η αλλαγή των διαστάσεων του διανύσματος. Η έξοδος του καταλήγει στο output layer.



Εικόνα 21: Αρχιτεκτονική συνελκτικού νευρωνικού δικτύου.

Το CNN που χρησιμοποιήθηκε αποτελείται από τρία τμήματα σε σειρά που ακολουθούν την εξής δομή: conv2D layer → batchnormalization layer → conv2D layer → batchnormalization layer → maxpooling2D layer → dropout layer. Στο τέλος, υπάρχει ένα dense layer το οποίο καταλήγει στο output layer. Χρησιμοποιώντας την αρχιτεκτονική αυτή το CNN κατάφερε να αποδώσει σε αρκετά ικανοποιητικό βαθμό.

4.2.2 Κλάση CNNModel

Η κλάση αυτή όπως προαναφέρθηκε, αντικαθιστά τις δύο συναρτήσεις `model_builder` και `train`, καθώς κληρονομεί την κλάση `BaseModel`.

Συνάρτηση `model_builder`:

Η συνάρτηση `model_builder` δημιουργεί το CNN μοντέλο της εικόνας 18, προσθέτει άλλο ένα `dense layer` στην έξοδο του και δέχεται σαν είσοδο το CNN. Καλείται στον κώδικα `trainer.py` του μοντέλου ώστε να πραγματοποιηθεί η εκπαίδευση.

Συνάρτηση `train`:

Είναι υπεύθυνη για την εκπαίδευση του CNN. Δέχεται σαν ορίσματα τις εικόνες μαζί με τις ετικέτες τους, τα δεδομένα αξιολόγησης, το μέγεθος παρτίδας (`batch size`), τον βελτιστοποιητή (`optimizer`), τις συναρτήσεις ανάκλησης (`callback functions`) και το σύνολο `train_datagen` που αναφέρθηκε στην προεπεξεργασία των δεδομένων. Καλεί την συνάρτηση `model_builder` ώστε να φτιάξει το CNN και χρησιμοποιώντας σαν μετρική την ορθότητα και σαν συνάρτηση απώλειας την `categorical_crossentropy`, πραγματοποιεί την εκπαίδευση του μοντέλου.

Τα ορίσματα που δέχεται η συνάρτηση `train` θα αναλυθούν στον κώδικα `trainer.py`, διότι παίζουν σπουδαίο ρόλο στην εκπαίδευση του CNN.

4.2.3 Κώδικας εκπαίδευσης

Στο αρχείο `trainer.py` πραγματοποιείται ολόκληρο το κομμάτι της εκπαίδευσης του νευρωνικού δικτύου. Δημιουργούνται οι `callbacks` συναρτήσεις, το `train_data_generator`, φορτώνεται ο `optimizer` και το μοντέλο. Γίνεται επίσης η κανονικοποίηση των δεδομένων εισόδου, χρησιμοποιώντας την συνάρτηση `data_builder` για να διαβάσει από το φάκελο του συνόλου δεδομένων και γίνεται ο διαχωρισμός των δεδομένων σε στα δύο υποσύνολα, το σύνολο εκπαίδευσης και το σύνολο αξιολόγησης. Έπειτα καλείται η συνάρτηση `train` που πραγματοποιεί την εκπαίδευση χρησιμοποιώντας τις κατάλληλες παραμέτρους και αφότου ολοκληρωθεί η εκπαίδευση, αποθηκεύονται σε φωτογραφίες το `classification matrix`, το ιστορικό εκπαίδευσης, η αρχιτεκτονική του μοντέλου και το μοντέλο για μελλοντική χρήση από την εφαρμογή.

Συναρτήσεις `callbacks`:

Οι συναρτήσεις αυτές συμβάλλουν στη βελτίωση του μοντέλου κατά την εκπαίδευση. Στο CNN που δημιουργήθηκε χρησιμοποιήθηκαν δύο ειδών `callback` συναρτήσεις, η συνάρτηση `early_stopping` και η `lr_scheduler`.

Η συνάρτηση `early_stopping` σταματάει πρόωρα την εκπαίδευση ώστε να αποφευχθεί η υπερβολική προσαρμογή πάνω στα δεδομένα (`overfitting`), παρατηρώντας μια μετρική. Συγκεκριμένα, επιλέχθηκε η μετρική ορθότητας των δεδομένων αξιολόγησης (`val_accuracy`) με όριο τον αριθμό 0.005 και αριθμό εποχών (`epochs`) 11. Αυτό σημαίνει ότι εάν η μετρική `val_accuracy` για 11 εποχές δεν εμφανίσει βελτίωση παραπάνω από 0.05%, τότε η

εκπαίδευση θα σταματήσει πρόωρα και θα κρατηθεί το μοντέλο εκείνο που είχε την μεγαλύτερη τιμή στην μετρική αυτή.

Η συνάρτηση `lr_scheduler` έχει ως στόχο τη βελτίωση του μοντέλου κατά την εκπαίδευση. Μειώνει τον ρυθμό εκπαίδευσης, παρατηρώντας μία μετρική. Χρησιμοποιήθηκε πάλι σαν κριτήριο η μετρική `val_accuracy` με όριο εντοπισμού τον αριθμό 0.0001, παράγοντα υπολογισμού του ρυθμού εκπαίδευσης το 0.6, αριθμό εποχών 4 και όριο κατώτατου ρυθμού εκπαίδευσης το $1e-7$. Στο ενδεχόμενο που για 4 εποχές η μετρική `val_accuracy` αυξάνεται με τιμή μικρότερη του 0.01% ή μειώνεται, τότε ο ρυθμός εκπαίδευσης θα μειωθεί κατά 40%, με κατώτατο όριο την τιμή $1e-7$.

train data generator:

Είναι μια γεννήτρια που δημιουργεί νέα δεδομένα εισόδου για το νευρωνικό δίκτυο βασισμένα στα υπάρχοντα του συνόλου δεδομένων. Εφαρμόζει περιστροφή 10 μοιρών, μεταβολές στο πλάτος και το ύψος κατά 10%, μεταβολή διάτμησης 10% και μεταβολή στο ζουμ 10%. Τέλος, εφαρμόζει οριζόντιο γύρισμα στις εικόνες. Η γεννήτρια αυτή δίνεται σαν είσοδος στην συνάρτηση της εκπαίδευσης, μαζί με τις συναρτήσεις `callback`.

Adam optimizer:

Οι `optimizers` είναι αλγόριθμοι που χρησιμοποιούνται για τη μείωση της απώλειας και τη μεγιστοποίηση την αποδοτικότητας του μοντέλου. Χρησιμοποιήθηκε ο `Adam optimizer` ο οποίος αποδίδει καλύτερα στις περισσότερες περιπτώσεις μηχανικής μάθησης. Είναι ένας προσαρμοστικός αλγόριθμος ο οποίος υπολογίζει ξεχωριστούς ρυθμούς εκπαίδευσης για διάφορες παραμέτρους. Η τιμή του ρυθμού εκπαίδευσης ορίστηκε η 0.01.

Μετά τον ορισμό όλων αυτών των συναρτήσεων και των παραμέτρων, φορτώνεται το CNN μοντέλο της εικόνας 18. Ορίζονται τα μονοπάτι του συνόλου δεδομένων που θα χρησιμοποιηθεί και το μονοπάτι όπου θα αποθηκευτεί το εκπαιδευμένο μοντέλο μαζί με τα αποτελέσματα του. Ορίζεται επίσης η λίστα των συναισθημάτων που πρόκειται να επιλεγθούν για το μοντέλο.

Όταν διαβάζονται τα δεδομένα από το σύνολο δεδομένων με τη χρήση της συνάρτησης `data_builder.ImageToArray`, αποθηκεύονται στις αντίστοιχες μεταβλητές οι φωτογραφίες και οι ετικέτες τους. Από τους υποφακέλους που υπάρχουν στον αρχικό φάκελο του συνόλου δεδομένων, επιλέγονται μόνο εκείνοι οι οποίοι ορίστηκαν στη λίστα των συναισθημάτων. Πέραν της προεπεξεργασίας των δεδομένων που πραγματοποιήθηκε μέσω της συνάρτησης, γίνεται και κανονικοποίηση τους διαιρώντας όλες τις τιμές των εικονοστοιχείων με την τιμή 255 ώστε να βρίσκονται στο εύρος [0-1].

Πριν εκτελεστεί η συνάρτηση της εκπαίδευσης, το σύνολο δεδομένων χωρίζεται στα υποσύνολα εκπαίδευσης και αξιολόγησης, δίνοντας στην συνάρτηση `train_test_split` που πραγματοποιεί τον διαχωρισμό ως παραμέτρους τη λίστα των εικόνων και των ετικετών τους και την αναλογία σε ποσοστό που θα διαχωριστούν. Τα δεδομένα ανακατεύονται ώστε να εισαχθούν στο μοντέλο με όσο πιο τυχαίο τρόπο γίνεται.

Οι διαστάσεις των υποσυνόλων εκπαίδευσης και αξιολόγησης φαίνονται στον παρακάτω πίνακα. Στα υποσύνολα `X_train` και `X_val` το πρώτο νούμερο δείχνει τον αριθμό των δειγμάτων, τα και τα άλλα τρία τις διαστάσεις των εικόνων (εικόνες $48*48$ χρησιμοποιώντας μόνο ένα κανάλι

χρώματος). Στα υποσύνολα `y_train` και `y_val` το πρώτο νούμερο δείχνει τον αριθμό των δειγμάτων και το δεύτερο τον αριθμό των συναισθημάτων.

Υποσύνολο	διαστάσεις
<code>X_train:</code>	(5082, 48, 48, 1)
<code>X_val:</code>	(442, 48, 48, 1)
<code>y_train:</code>	(5082, 4)
<code>y_val:</code>	(442, 4)

Πίνακας 4: Διαχωρισμός των δεδομένων με αναλογία 92%.

Συνάρτηση train:

Η συνάρτηση `train` δέχεται σαν ορίσματα τα υποσύνολα εκπαίδευσης και αξιολόγησης, τον `optimizer`, τις συναρτήσεις `callbacks`, τη γεννήτρια δεδομένων, το μέγεθος των παρτίδων (`batch size`) και τον αριθμό των εποχών (`epochs`). Το `batch size` είναι ο αριθμός των δειγμάτων που έχουν επεξεργαστεί πριν ανανεωθεί το νευρωνικό δίκτυο. Το νούμερο αυτό πρέπει να είναι μεγαλύτερο ή ίσο της μονάδας και μικρότερο ή ίσο από το μέγεθος του συνόλου δεδομένων. Ως `epoch` ορίζεται ο αριθμός των περασμάτων ολόκληρου του υποσυνόλου εκπαίδευσης που έχει πραγματοποιήσει το νευρωνικό δίκτυο. Τα σύνολα δεδομένων χωρίζονται σε παρτίδες (`batches`), ιδίως όταν το σύνολο δεδομένων είναι μεγάλου μεγέθους.

Παρακάτω φαίνεται ένα στιγμιότυπο εκπαίδευσης των τριών πρώτων `epochs` με αριθμό `epochs` 100 και `batch size` 16 με 92% διαχωρισμό δεδομένων.

Epoch 1/100 317/317 [=====] - 12s 21ms/step - loss: 1.7215 - accuracy: 0.3266 - val_loss: 1.4917 - val_accuracy: 0.3710 - lr: 0.0100
Epoch 2/100 317/317 [=====] - 5s 17ms/step - loss: 1.3649 - accuracy: 0.4028 - val_loss: 1.4295 - val_accuracy: 0.3100 - lr: 0.0100
Epoch 3/100 317/317 [=====] - 5s 17ms/step - loss: 1.1344 - accuracy: 0.5081 - val_loss: 0.9633 - val_accuracy: 0.5860 - lr: 0.0100

Πίνακας 5: Στιγμιότυπο εκπαίδευσης νευρωνικού δικτύου.

Στον πίνακα 6 φαίνεται η χρήση των `callback` συναρτήσεων, στο 21^ο `epoch` ενεργοποιήθηκε η συνάρτηση `lr_scheduler` και στο 52^ο `epoch` ενεργοποιήθηκε η συνάρτηση `early stopping`, κρατώντας ένα στιγμιότυπο του μοντέλου από το 41^ο `epoch`, στο οποίο το μοντέλο είχε τις καλύτερες τιμές.

```
Epoch 21/100
316/317 [=====>.] - ETA: 0s - loss: 0.6799 - accuracy: 0.7378
Epoch 00021: ReduceLROnPlateau reducing learning rate to 0.00360000031292438.

Epoch 52/100
316/317 [=====>.] - ETA: 0s - loss: 0.4469 - accuracy: 0.8259
Restoring model weights from the end of the best epoch: 41.
317/317 [=====] - 6s 18ms/step - loss: 0.4463 - accuracy: 0.8261 - val_loss: 0.5721 - val_accuracy: 0.7964 - lr: 2.7994e-04
Epoch 00052: early stopping
```

Πίνακας 6: Ενεργοποίηση των callback συναρτήσεων.

Μετά το πέρας της εκπαίδευσης, στο μονοπάτι εξόδου που ορίστηκε πιο πριν, αποθηκεύεται το μοντέλο, το ιστορικό εκπαίδευσης του, το confusion matrix και η αρχιτεκτονική του.

4.3 Σχεδιασμός διεπαφής χρήστη (User Interface)

Ο σχεδιασμός της εφαρμογής πραγματοποιήθηκε μέσω του κώδικα app.py. Μέσω του κώδικα αυτού σχεδιάστηκε όλο το γραφικό περιβάλλον και η λειτουργία της εφαρμογής, καθώς και η συνάρτηση που πραγματοποιεί την πρόβλεψη του συναισθήματος και τον εντοπισμό του προσώπου. Αρχικά δηλώθηκαν τα κομμάτια μουσικής που αντιστοιχούν στο κάθε συναίσθημα και αντίστοιχα ενθαρρυντικά μηνύματα. Έπειτα φορτώθηκε το προεκπαιδευμένο μοντέλο εντοπισμού προσώπου face_detector και ορίστηκε σαν κατώφλι πρόβλεψης το ποσοστό 70%. Ακολούθησε η συνάρτηση dnn_detector η οποία πραγματοποιεί την πρόβλεψη και αποτυπώνει στην οθόνη τα αποτελέσματα αυτής. Τέλος, στο κύριο πρόγραμμα πραγματοποιείται ο βασικός σχεδιασμός της διεπαφής όπου παίρνεται το δείγμα της κάμερας και με τη χρήση της συνάρτησης dnn_detector εκτελείται η πρόβλεψη πάνω στα καρτέ και ανάλογα με το εκάστοτε συναίσθημα, εμφανίζεται ένα τυχαίο ενθαρρυντικό μήνυμα και η μουσική που αντιστοιχεί στο συναίσθημα αυτό.

4.3.1 Συνάρτηση dnn_detector

Η συνάρτηση αυτή είναι υπεύθυνη για τον εντοπισμό του προσώπου ώστε να απομονωθεί το καρτέ, για την πρόβλεψη του συναισθήματος βάσει του συγκεκριμένου καρτέ και για την εμφάνιση των αποτελεσμάτων στην διεπαφή. Δέχεται σαν όρισμα ένα καρτέ, το μετατρέπει σε ασπρόμαυρο ώστε να ακολουθεί τη δομή των εικόνων του συνόλου δεδομένων και μέσω του προεκπαιδευμένου μοντέλου της βιβλιοθήκης opencv απομονώνει το πρόσωπο. Έπειτα υπολογίζοντας τις γωνίες του προσώπου, σχεδιάζει ένα τετράγωνο γύρω από αυτό ώστε να είναι προφανής η περιοχή της πρόβλεψης αλλά και για να δείξει στον χρήστη πως ο εντοπισμός του προσώπου λειτουργεί κανονικά. Στη συνέχεια γίνεται προεπεξεργασία του καρτέ και δίνεται σαν είσοδος στο CNN που παράχθηκε κατά στο κομμάτι της εκπαίδευσης, εμφανίζοντας στην διεπαφή την πρόβλεψη με το μεγαλύτερο ποσοστό. Τέλος, εμφανίζει στην πάνω αριστερά γωνία διεπαφής τα συναίσθημα που χρησιμοποιήθηκαν με τα αποτελέσματα της πρόβλεψης σε ποσοστό και επιστρέφει το όνομα του

συναισθήματος με το υψηλότερο ποσοστό για να αξιοποιηθεί στη λειτουργία του κύριου προγράμματος.

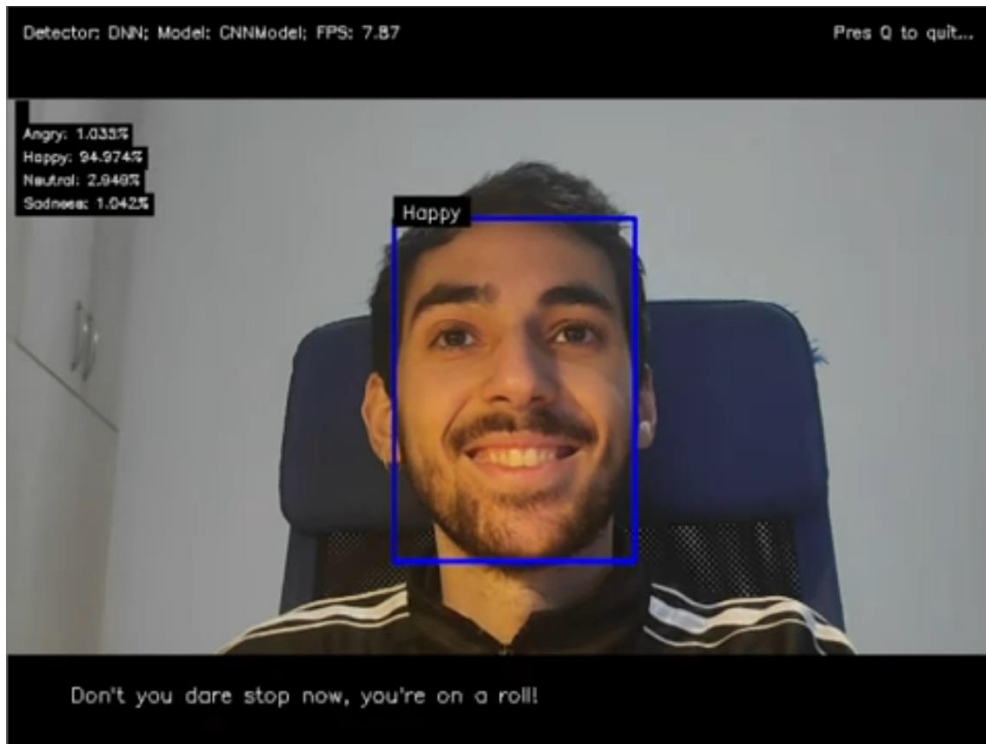
4.3.2 Διεπαφή χρήστη

Το κύριο πρόγραμμα του κώδικα `app.py` δημιουργεί την διεπαφή χρήστη. Ανοίγει την τοπική κάμερα του υπολογιστή και εμφανίζει στην οθόνη ένα παράθυρο διαστάσεων 640*480 εικονοστοιχείων. Για κάθε καρέ που εντοπίζει, καλεί την συνάρτηση `dnn_detector` ώστε να προκύψει η πρόβλεψη. Η έξοδος της συνάρτησης που αποτελεί και την πρόβλεψη, προστίθεται σε μία λίστα που αποθηκεύει όλα τα αποτελέσματα. Αν εντοπιστεί στο τέλος της λίστας το ίδιο αποτέλεσμα 15 φορές, τότε ενεργοποιείται η μουσική του συναισθήματος και εμφανίζεται στο κάτω μέρος του παραθύρου ένα εκ των τριών ενθαρρυντικών μηνυμάτων με τυχαίο τρόπο. Η μέθοδος αυτή προσδίδει μια σταθερότητα στην εναλλαγή της μουσικής και του μηνύματος γιατί το CNN ενδεχομένως να κάνει στιγμιαία άστοχες προβλέψεις. Με αυτόν τον τρόπο, η μουσική και το μήνυμα αλλάζουν όταν για περίπου 2 δευτερόλεπτα η πρόβλεψη παραμένει ίδια. Μετά τον εντοπισμό του συναισθήματος, η λίστα αυτή μηδενίζεται ώστε να συνεχίσει με το επόμενο συναίσθημα. Στην πάνω αριστερή γωνία του παραθύρου αναγράφεται ο αλγόριθμος εντοπισμού προσώπου που χρησιμοποιήθηκε (`dnn_detector`) το μοντέλο (CNN Model) και τα καρέ ανά δευτερόλεπτο (FPS). Το πρόγραμμα τερματίζει όταν ο χρήστης πατήσει το πλήκτρο 'Q'.

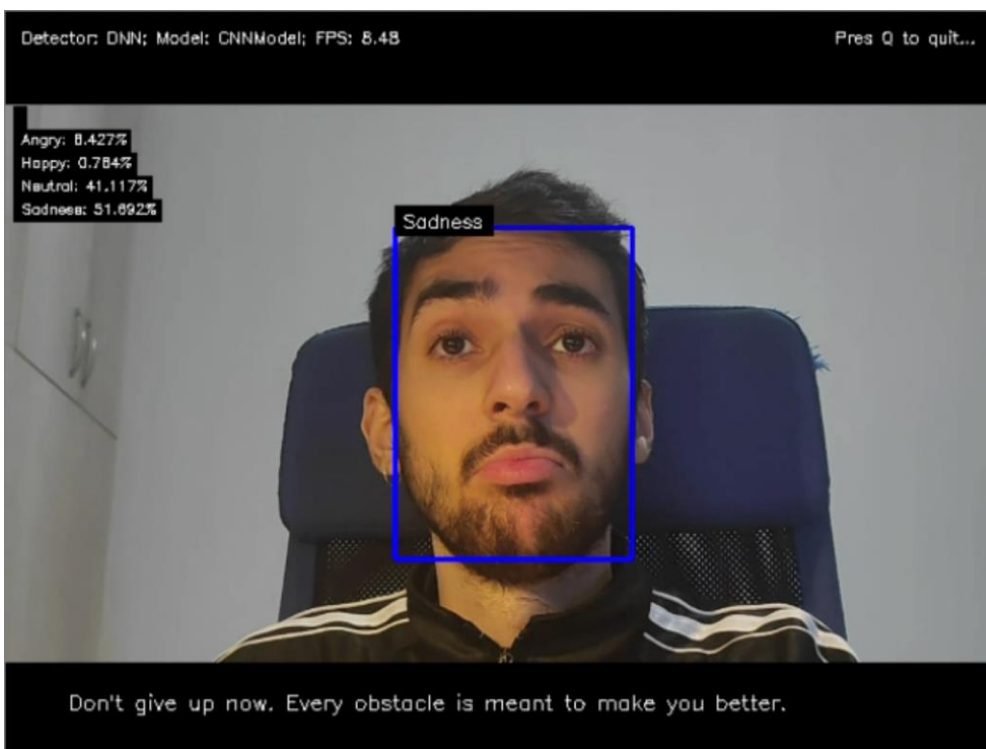
Επειδή όμως η υλοποίηση της εφαρμογής πραγματοποιήθηκε χρησιμοποιώντας τη γλώσσα `python` και τις βιβλιοθήκες της, ο χρήστης για να την θέσει σε λειτουργία θα χρειαζόταν το αρχείο του κώδικα `app.py`, τις συγκεκριμένες εκδόσεις της γλώσσας και των βιβλιοθηκών του αρχείου. Για να αποφευχθεί αυτή η διαδικασία, χρησιμοποιήθηκε το εργαλείο `PyInstaller` το οποίο δημιούργησε ένα εκτελέσιμο αρχείο με όλες τις βιβλιοθήκες ενσωματωμένες. Έτσι, η διεπαφή αυτονομήθηκε και είχε την δυνατότητα να τρέξει σε οποιοδήποτε υπολογιστή χωρίς την εγκατάσταση επιπλέον προγραμμάτων και εργαλείων. Παρακάτω παρουσιάζονται κάποια στιγμιότυπα της διεπαφής για το κάθε συναίσθημα που χρησιμοποιήθηκε.

Στο τελικό πρόγραμμα το οποίο είχε τη βέλτιστη απόδοση, χρησιμοποιήθηκαν μόνο τα συναισθήματα: χαρά, λύπη, θυμός και ουδέτερο συναίσθημα, καθώς είναι τα πιο βασικά και κρίνονται χρήσιμα σε μία εφαρμογή που έχει σκοπό την ενθάρρυνση του χρήστη για την διεκπεραίωση των εργασιών του.

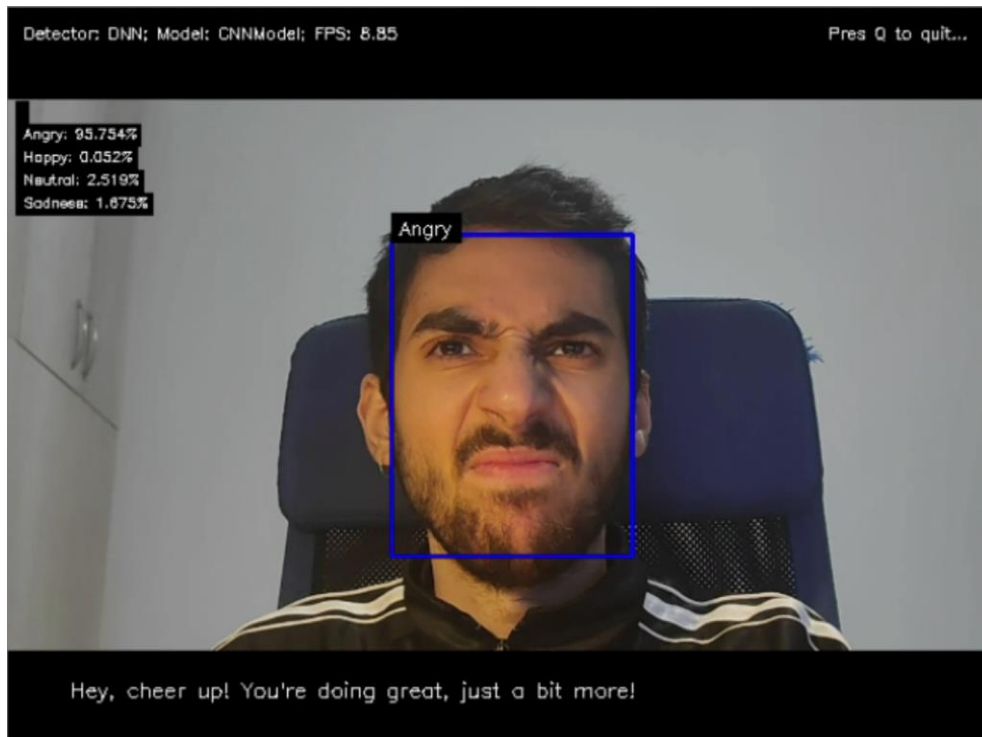
Για αυτό τα στιγμιότυπα που ακολουθούν περιέχουν μόνο τα προαναφερθέντα συναισθήματα.



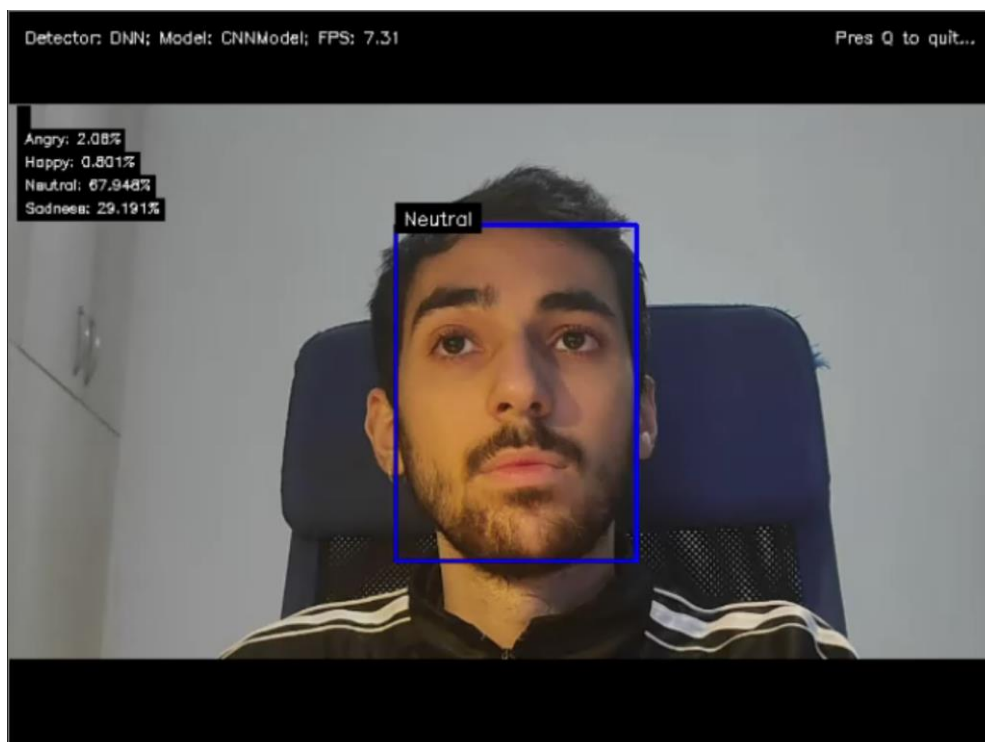
Εικόνα 22: Αναγνώριση συναισθήματος χαράς.



Εικόνα 23: Αναγνώριση συναισθήματος λύπης.



Εικόνα 24: Αναγνώριση συναισθήματος θυμού.



Εικόνα 25: Αναγνώριση ουδέτερου συναισθήματος.

5 ΚΕΦΑΛΑΙΟ 5^ο : Ανάλυση Αποτελεσμάτων

Το CNN που δημιουργήθηκε δοκιμάστηκε σε τέσσερα διαφορετικά σύνολα δεδομένων, στο FER-2013 στο φιλτραρισμένο FER-2013, στο CK και στο τελικό σύνολο δεδομένων το οποίο αποτελούσε την ένωση των δύο τελευταίων. Για κάθε σύνολο δεδομένων πραγματοποιήθηκαν πειράματα με διαφορετικό αριθμό συναισθημάτων. Σε όλες τις δοκιμές το μέγεθος batch size ορίστηκε στο 16, η αναλογία δεδομένων εκπαίδευσης-αξιολόγησης στο 86% και ο αριθμός των epochs στα 100.

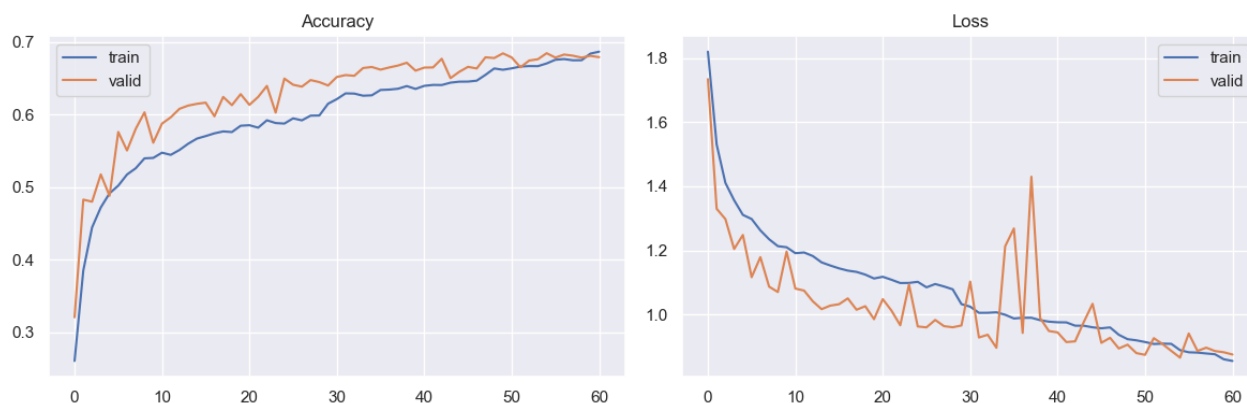
5.1 Dataset FER-2013

5.1.1 Εκπαίδευση μοντέλου με έξι συναισθήματα

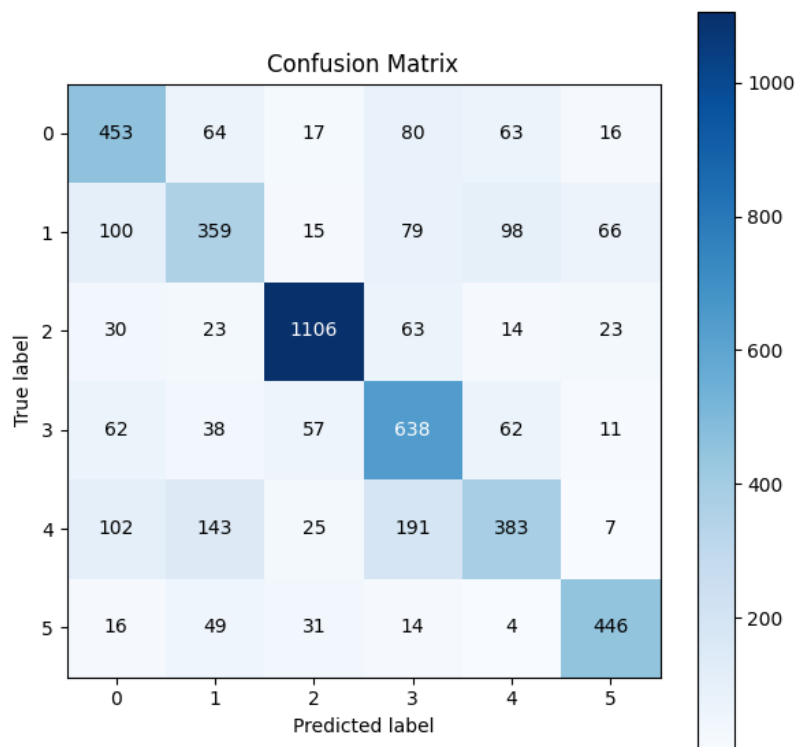
Χρησιμοποιήθηκαν όλα τα συναισθήματα του συνόλου δεδομένων πέραν της αποστροφής: θυμός, φόβος, χαρά, λύπη, έκπληξη και ουδέτερο συναίσθημα. Ο πίνακας αναλυτικής αξιολόγησης μαζί με το ιστορικό ορθότητας και απώλειας και το confusion matrix, παρατίθενται παρακάτω:

test accuracy: 68.4115 %				
	precision	recall	f1-score	support
0	0.59	0.65	0.62	693
1	0.53	0.50	0.52	717
2	0.88	0.88	0.88	1259
3	0.60	0.74	0.66	868
4	0.61	0.45	0.52	851
5	0.78	0.80	0.79	560
accuracy			0.68	4948
macro avg	0.67	0.67	0.66	4948
weighted avg	0.68	0.68	0.68	4948
{0: 'Angry', 1: 'Fear', 2: 'Happy', 3: 'Neutral', 4: 'Sad', 5: 'Surprise'}				

Πίνακας 7: Αναλυτική αξιολόγηση στο FER-2013 με 6 συναισθήματα.



Εικόνα 26: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 6 συναισθήματα.



Εικόνα 27: Confusion matrix στο FER-2013 με 6 συναισθήματα.

Χρησιμοποιώντας 6 συναισθήματα, η ορθότητα του μοντέλου στην αξιολόγηση έφτασε μόλις το 68%. Συνακόλουθα, οι μετρικές f1-score και είχαν ικανοποιητικές τιμές μόνο στο συναίσθημα της χαράς, καθώς είναι εκείνο το οποίο ξεχωρίζει περισσότερο από τα υπόλοιπα. Όπως φαίνεται και από το confusion matrix πραγματοποιήθηκαν αρκετές λάθος προβλέψεις εξαιρώντας το συναίσθημα χαράς και λύπης. Επίσης, παρατηρήθηκε πως το μοντέλο συνέχεε τα συναισθήματα φόβου και θυμού. Ο όγκος του αρχικού FER-2013 και τα συγκεκριμένα συναισθήματα που

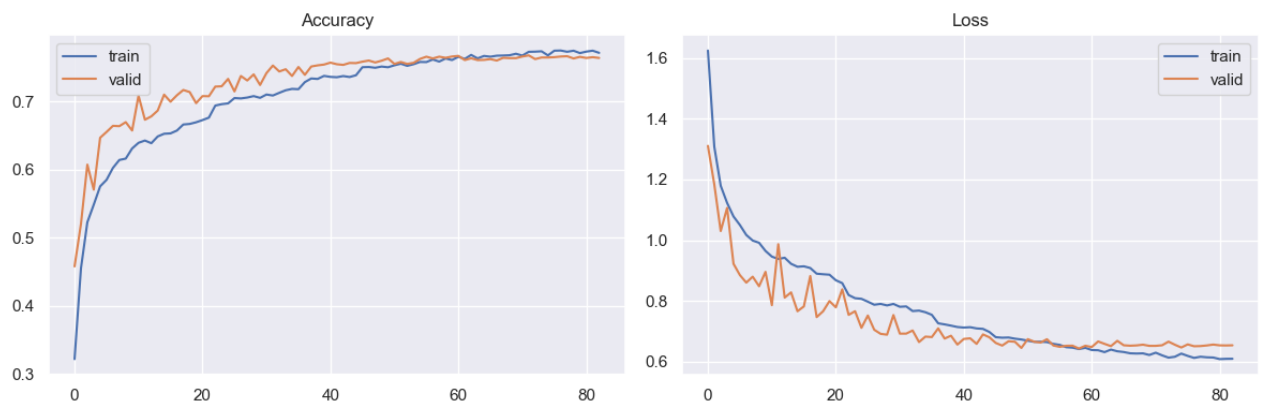
επιλέχθηκαν στην δοκιμή αυτή δείχνουν να μην ωφελούν στην εκπαίδευση του μοντέλου σε ικανοποιητικό βαθμό.

5.1.2 Εκπαίδευση μοντέλου με πέντε συναισθήματα

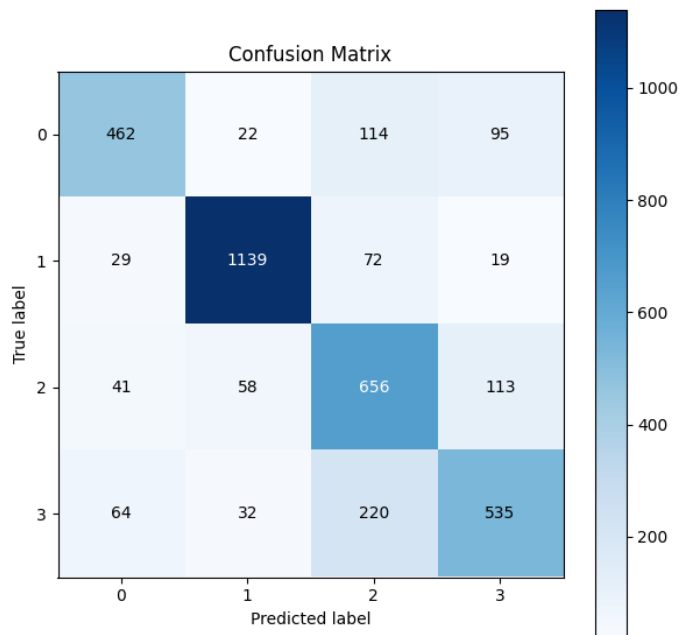
Συνεχίζοντας τις δοκιμές πάνω στο FER-2013 και λαμβάνοντας υπόψιν το γεγονός της ομοιότητας των συναισθημάτων φόβου και θυμού, επιλέχθηκε να μην χρησιμοποιηθεί το πρώτο, καθώς είχε τις χειρότερες επιδόσεις.

test accuracy: 76.814%				
	precision	recall	f1-score	support
0	0.75	0.66	0.70	629
1	0.88	0.88	0.88	1259
2	0.65	0.77	0.70	868
3	0.69	0.63	0.66	851
4	0.86	0.85	0.86	560
accuracy			0.77	4231
macro avg	0.77	0.76	0.76	4231
weighted avg	0.77	0.77	0.77	4231
{0: 'Angry', 1: 'Happy', 2: 'Neutral', 3: 'Sad, 4: Surprise}				

Πίνακας 8: Αναλυτική αξιολόγηση στο FER-2013 με 5 συναισθήματα.



Εικόνα 28: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 5 συναισθήματα.



Εικόνα 29: Confusion matrix στο FER-2013 με 5 συναισθήματα.

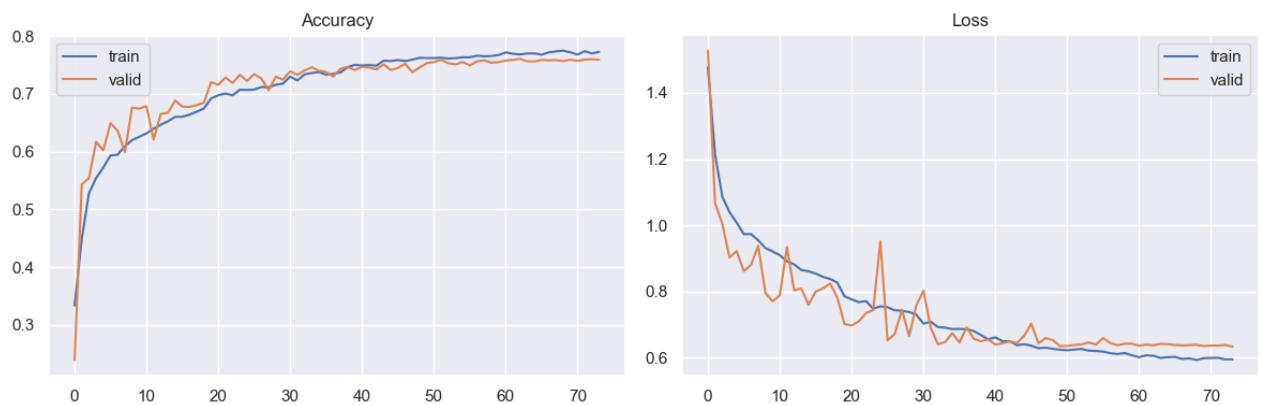
Παρόλου που ο γενικός μέσος όρος ορθότητας στην αξιολόγηση ήταν 77%, παρατηρείται ότι πάλι μόνο το συναίσθημα της χαράς είχε υψηλές επιδόσεις f1-score, σε αντίθεση με τα υπόλοιπα συναισθήματα. Το μοντέλο είχε υψηλή τιμή απώλειας στην αξιολόγηση, πράγμα το οποίο σημαίνει ότι έγιναν αρκετές λάθος προβλέψεις. Τέλος, η υψηλή μετρική του f1-score στο χαρούμενο συναίσθημα παρατηρείται και στο confusion matrix καθώς είναι εκείνο με τις λιγότερες λάθος προβλέψεις. Κατά τη δοκιμή του μοντέλου στην διεπαφή, παρατηρήθηκε ότι το συναίσθημα της έκπληξης δεν εντοπιζόταν καθόλου, παρά την υψηλή του επίδοση.

5.1.3 Εκπαίδευση μοντέλου με τέσσερα συναισθήματα

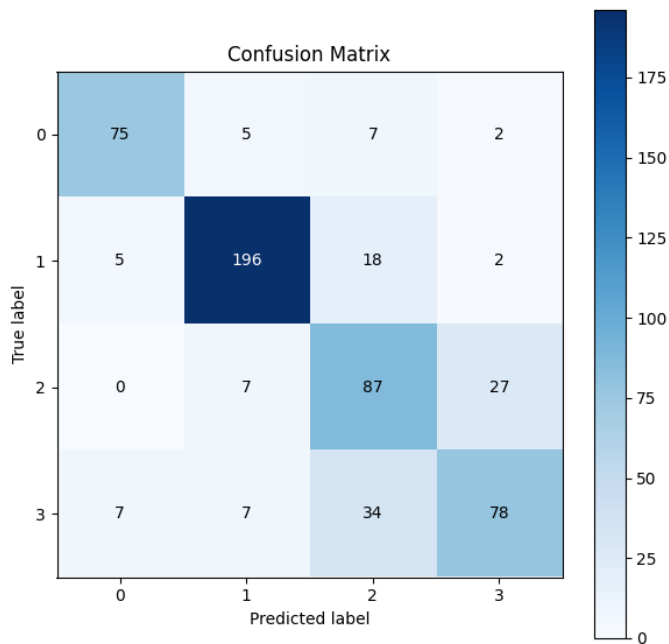
Ακολούθησε ακριβώς η ίδια διαδικασία με την προηγούμενη, αφαιρώντας το συναίσθημα της έκπληξης, λόγω της δυσκολίας του μοντέλου να το εντοπίσει και λόγω της μικρής σημαντικότητας στη λειτουργία της διεπαφής.

test accuracy: 76.0556%				
	precision	recall	f1-score	support
0	0.78	0.67	0.72	693
1	0.91	0.90	0.91	1259
2	0.62	0.76	0.68	868
3	0.70	0.63	0.66	851
accuracy			0.76	3671
macro avg	0.75	0.74	0.74	3671
weighted avg	0.77	0.76	0.76	3671
{0: 'Angry', 1: 'Happy', 2: 'Neutral', 3: 'Sad'}				

Πίνακας 9: Αναλυτική αξιολόγηση στο FER-2013 με 4 συναισθήματα.



Εικόνα 30: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 4 συναισθήματα.



Εικόνα 31: Confusion matrix στο FER-2013 με 4 συναισθήματα.

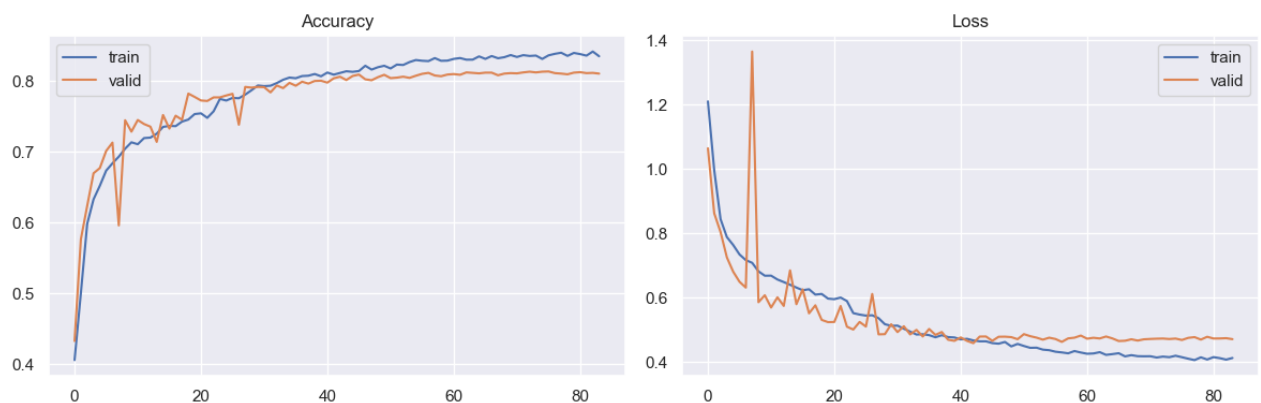
Αφαιρώντας το προβληματικό συναίσθημα, ενώ αναμενόταν μια βελτίωση στην μετρική ορθότητας, η τιμή της παρέμεινε ίδια. Επιπρόσθετα, καμία μετρική από όλα τα συναισθήματα δεν εμφάνισε κάποια αισθητή βελτίωση, το οποίο συμπεραίνεται και από το confusion matrix.

5.1.4 Εκπαίδευση μοντέλου με τρία συναισθήματα

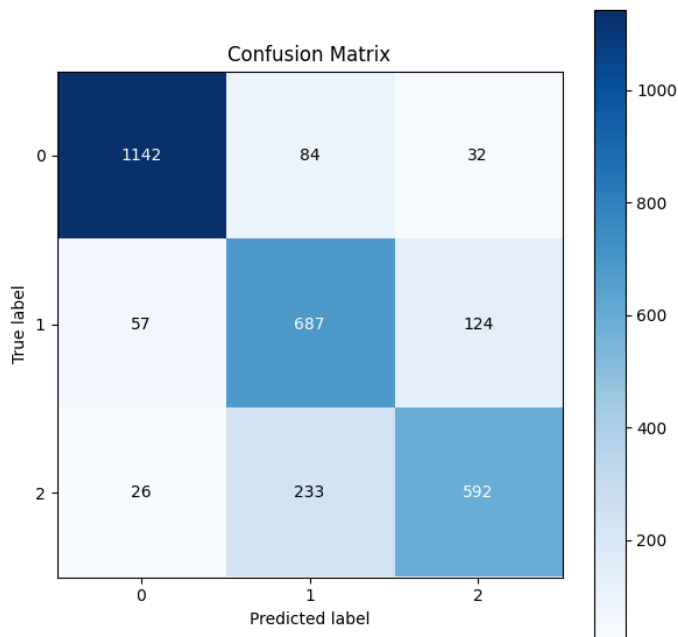
Τέλος, δοκιμάστηκαν μόνο τα τρία βασικά συναισθήματα: χαρά, λύπη και ουδέτερο συναίσθημα για ελεγχθεί η πιθανότητα συμβολής του συναισθήματος του θυμού στα προηγούμενα αποτελέσματα.

test accuracy: 81.3235%				
	precision	recall	f1-score	support
0	0.93	0.91	0.92	1258
1	0.68	0.79	0.73	868
2	0.79	0.70	0.74	851
accuracy			0.81	2977
macro avg	0.80	0.80	0.80	2977
weighted avg	0.82	0.81	0.81	2977
{0: 'Happy', 1: 'Neutral', 2: 'Sad'}				

Πίνακας 10: Αναλυτική αξιολόγηση στο FER-2013 με 3 συναισθήματα.



Εικόνα 32: Ιστορικό ορθότητας και απώλειας στο FER-2013 με 3 συναισθήματα.



Εικόνα 33: Confusion matrix στο FER-2013 με 3 συναισθήματα.

Χρησιμοποιώντας μόλις 3 συναισθήματα, το μοντέλο μπόρεσε να τα ξεχωρίσει σε ικανοποιητικό βαθμό. Παρατηρήθηκε βελτίωση 5 μονάδων στον μέσο όρο ορθότητας και πτώση στις μετρικές του συναισθήματος λύπης. Η απώλεια στα δεδομένα εκπαίδευσης είχε μικρότερο νούμερο, όπως φαίνεται και από το confusion matrix.

Δοκιμάζοντας τα εκάστοτε μοντέλα στην διεπαφή που είχε δημιουργηθεί, παρατηρήθηκε πως όλες οι μετρικές αξιολόγησης απέδωσαν την πραγματική επίδοση τους, αδυνατώντας να εντοπίσουν εύκολα τα συναισθήματα πέραν της χαράς.

5.2 Dataset FER-2013 (filtered)

5.2.1 Εκπαίδευση μοντέλου με πέντε συναισθήματα

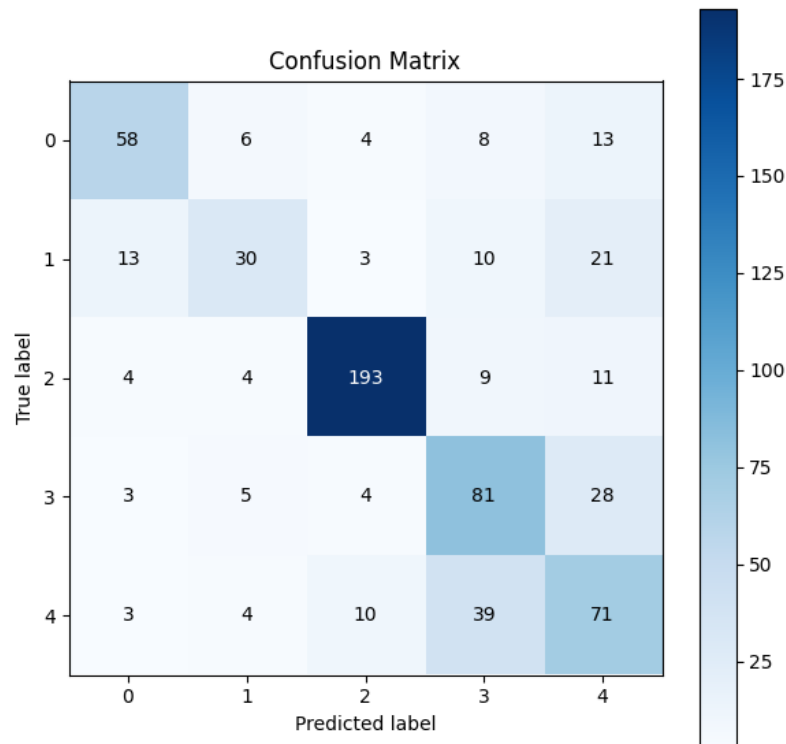
Χρησιμοποιήθηκαν όλα τα συναισθήματα του φιλτραρισμένου συνόλου δεδομένων fer-2013: θυμός, φόβος, χαρά, λύπη και ουδέτερο συναίσθημα. Ο πίνακας αναλυτικής αξιολόγησης μαζί με το ιστορικό ορθότητας και απώλειας και το confusion matrix, παρατίθενται παρακάτω:

test accuracy: 68.189%				
	precision	recall	f1-score	support
0	0.72	0.65	0.68	89
1	0.61	0.39	0.48	77
2	0.90	0.87	0.89	221
3	0.55	0.67	0.60	121
4	0.49	0.56	0.52	127
accuracy			0.68	635
macro avg	0.65	0.63	0.63	635
weighted avg	0.69	0.68	0.68	635
{0: 'Angry', 1: 'Fear', 2: 'Happy', 3: 'Neutral', 4: 'Sadness'}				

Πίνακας 11: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 5 συναισθήματα.



Εικόνα 34: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 5 συναισθήματα.



Εικόνα 35: Confusion matrix στο FER-2013 (filtered) με 5 συναισθήματα.

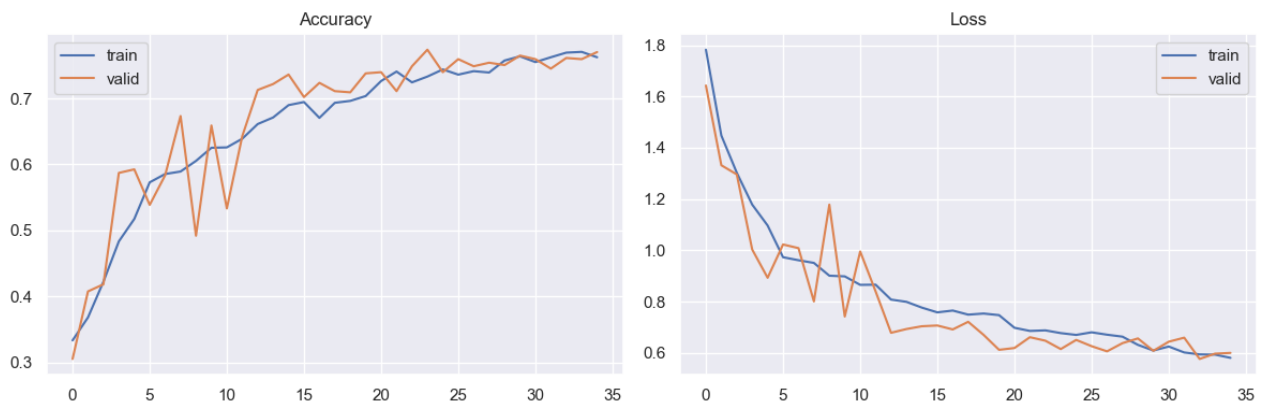
Ο μέσος όρος μετρικής ορθότητας είχε αρκετά χαμηλό νούμερο συγκριτικά με τη δοκιμή στο αρχικό fer-2013 με τον ίδιο αριθμό συναισθημάτων, παρόλο που απορρίφθηκαν αρκετές προβληματικές εικόνες. Εκτός του χαρούμενου συναισθήματος, οι μετρικές f1-score είχαν ιδιαίτερα χαμηλές τιμές.

5.2.2 Εκπαίδευση μοντέλου με τέσσερα συναισθήματα

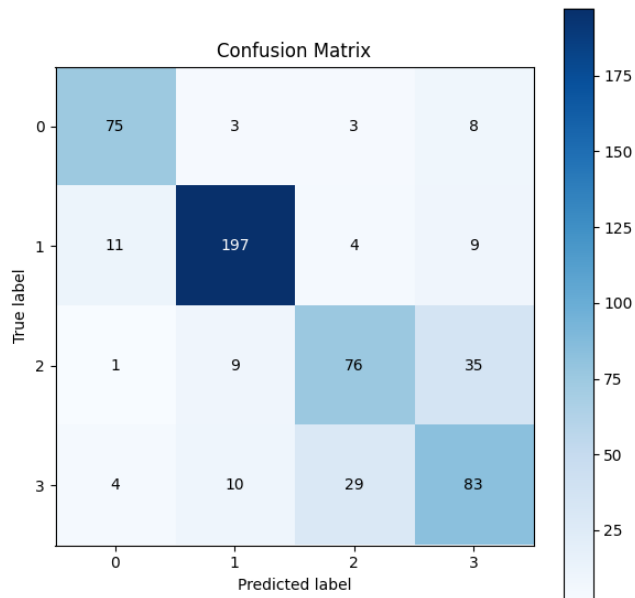
Συνεχίστηκαν οι δοκιμές, αφαιρώντας του συναισθήμα του φόβου, το οποίο είχε τα χειρότερα αποτελέσματα συγκριτικά με τα υπόλοιπα.

test accuracy: 77.3788%				
	precision	recall	f1-score	support
0	0.82	0.84	0.83	89
1	0.90	0.89	0.90	221
2	0.68	0.63	0.65	121
3	0.61	0.66	0.64	126
accuracy			0.77	557
macro avg	0.75	0.76	0.75	557
weighted avg	0.78	0.77	0.77	557
{0: 'Angry', 1: 'Happy', 2: 'Neutral', 3: 'Sadness'}				

Πίνακας 12: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 4 συναισθήματα.



Εικόνα 36: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 4 συναισθήματα.



Εικόνα 37: Confusion matrix στο FER-2013 (filtered) με 4 συναισθήματα.

Σε αυτήν την περίπτωση, παρατηρήθηκε μεγάλη άνοδος σε όλα τα συναισθήματα, ιδίως στο θυμού. Όπως φαίνεται και από το confusion matrix, όλα τα συναισθήματα προβλέφθηκαν σωστά σε μεγάλο βαθμό, εξαιρώντας το συναίσθημα της λύπης που το συνέχευε με το ουδέτερο συναίσθημα.

5.2.3 Εκπαίδευση μοντέλου με τρία συναισθήματα

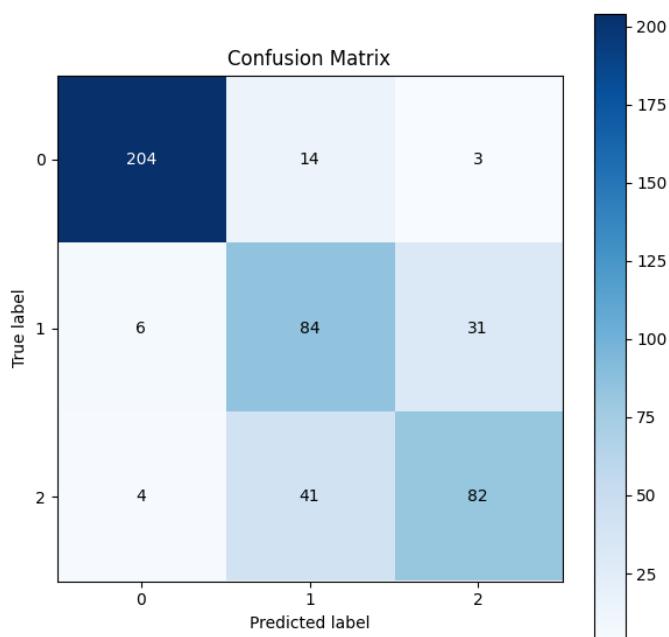
Τελειώνοντας, αφαιρέθηκε το συναίσθημα του θυμού, γιατί ήταν εκείνο με τα λιγότερα δείγματα.

test accuracy: 78.8913%				
	precision	recall	f1-score	support
0	0.95	0.92	0.94	221
1	0.60	0.69	0.65	121
2	0.71	0.65	0.67	127
accuracy			0.79	469
macro avg	0.75	0.75	0.75	469
weighted avg	0.80	0.79	0.79	469
{0: 'Happy', 1: 'Neutral', 2: 'Sadness'}				

Πίνακας 13: Αναλυτική αξιολόγηση στο FER-2013 (filtered) με 3 συναισθήματα.



Εικόνα 38: Ιστορικό ορθότητας και απώλειας στο FER-2013 (filtered) με 3 συναισθήματα.



Εικόνα 39: Confusion matrix στο FER-2013 (filtered) με 3 συναισθήματα.

Τα αποτελέσματα που προέκυψαν ήταν σαφώς καλύτερα από την προηγούμενη περίπτωση, όμως δεν ικανοποιούσαν τις προϋποθέσεις που είχαν τεθεί. Συγκρίνοντας την ίδια δοκιμή με το κανονικό fer-2013, η απόδοση ήταν αρκετά μικρότερη. Σημαντικό ρόλο σε αυτό το αποτέλεσμα έπαιξε ο μικρός όγκος των δεδομένων που προέκυψε μετά την αφαίρεση των προβληματικών εικόνων.

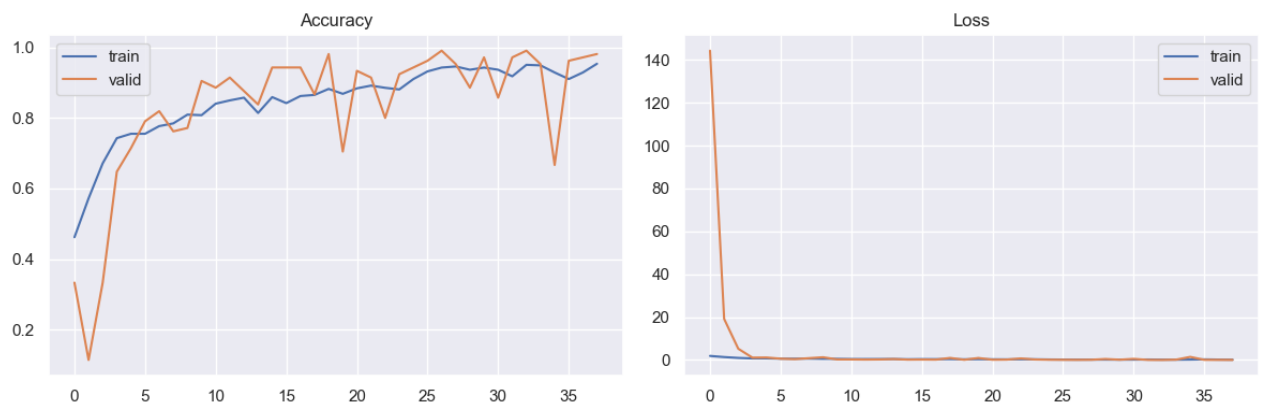
5.3 Dataset CK

5.3.1 Εκπαίδευση μοντέλου με πέντε συναισθήματα

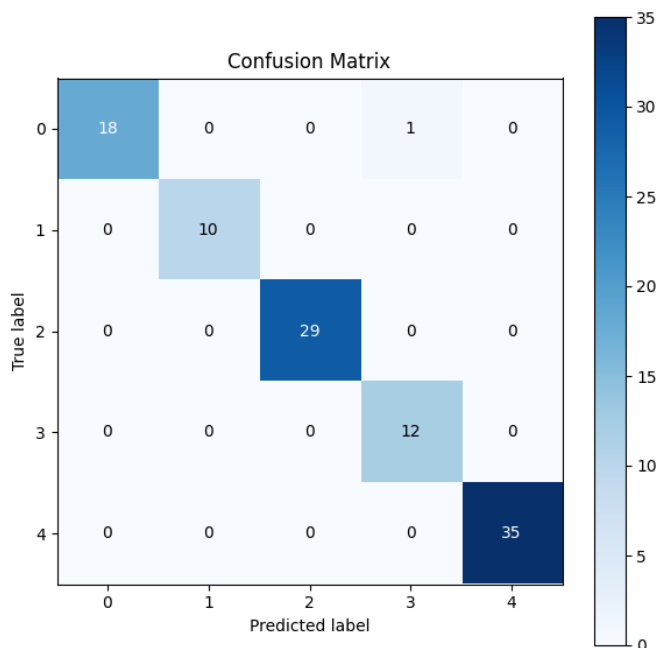
Χρησιμοποιήθηκαν όλα τα συναισθήματα του συνόλου δεδομένων CK: θυμός, φόβος, χαρά, λύπη και έκπληξη. Λόγω του μικρότερου όγκου δεδομένων η εκπαίδευση σε όλες τις δοκιμές πραγματοποιήθηκαν σε λιγότερα epochs.

test accuracy: 99.0476%				
	precision	recall	f1-score	support
0	1.00	0.95	0.97	19
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	29
3	0.92	1.00	0.96	12
4	1.00	1.00	1.00	35
accuracy			0.99	105
macro avg	0.98	0.99	0.99	105
weighted avg	0.99	0.99	0.99	105
{0: 'Angry', 1: 'Fear', 2: 'Happy', 3: 'Sadness', 4: 'Surprise'}				

Πίνακας 14: Αναλυτική αξιολόγηση στο CK με 5 συναισθήματα.



Εικόνα 40: Ιστορικό ορθότητας και απώλειας στο CK με 5 συναισθήματα.



Εικόνα 41: Confusion matrix στο CK με 5 συναισθήματα.

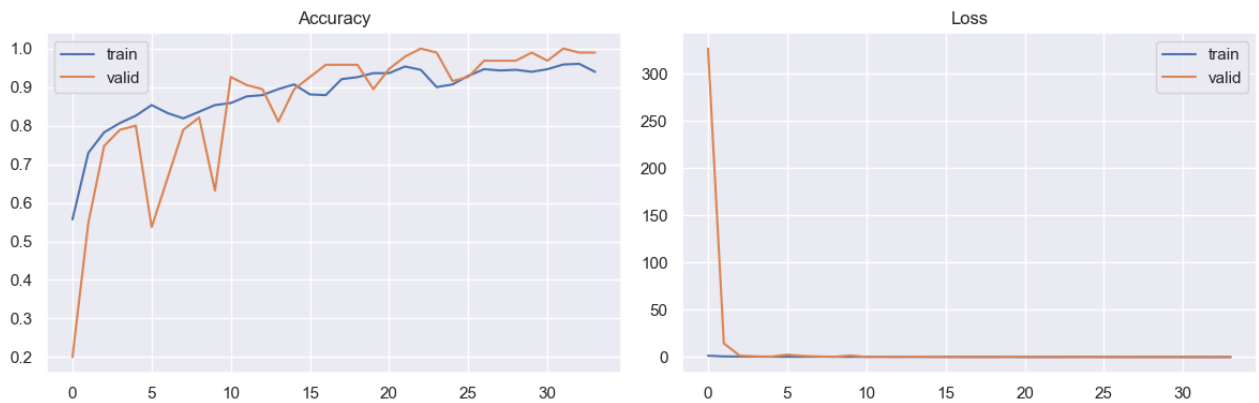
Σε αυτήν την περίπτωση, το μοντέλο εμφάνισε άνογκες τιμές σε όλες τις μετρικές, φτάνοντας το 99% στον μέσο όρο ορθότητας, πραγματοποιώντας μόλις 38 epochs. Η απώλεια έφτασε πολύ κοντά στο μηδέν, το οποίο αποτυπώνεται και από το confusion matrix με μόλις μία λάθος πρόβλεψη. Η απόδοση όμως του μοντέλου δεν μεταφράζεται απαραίτητα και στην πραγματική του επίδοση πάνω στην διεπαφή, καθώς ενδέχεται να έχει πραγματοποιηθεί overfitting λόγω του πολύ μικρού μεγέθους του συνόλου δεδομένων.

5.3.2 Εκπαίδευση μοντέλου με τέσσερα συναισθήματα

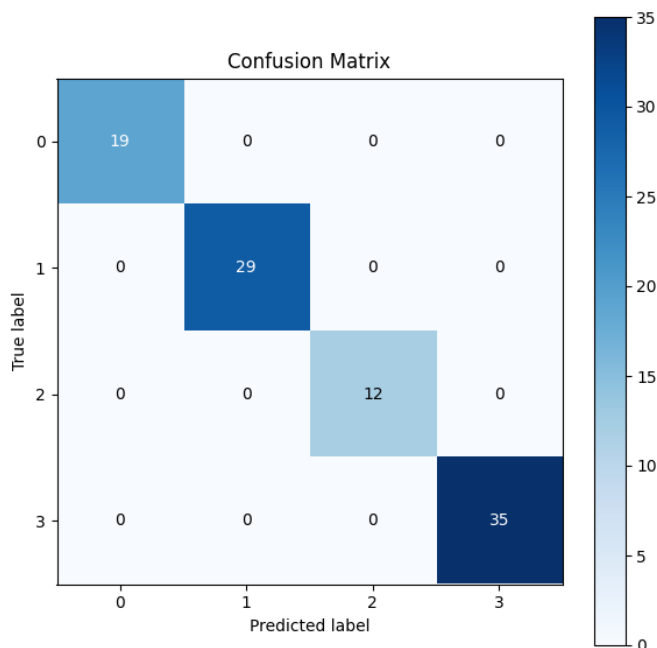
Συνεχίζοντας τις δοκιμές αφαιρέθηκε το συναίσθημα με το μικρότερο αριθμό δειγμάτων το οποίο ήταν το συναίσθημα του φόβου.

test accuracy: 100.0%				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	19
1	1.00	1.00	1.00	29
2	1.00	1.00	1.00	12
3	1.00	1.00	1.00	35
accuracy			1.00	95
macro avg	1.00	1.00	1.00	95
weighted avg	1.00	1.00	1.00	95
{0: 'Angry', 1: 'Happy', 2: 'Sadness', 3: 'Surprise'}				

Πίνακας 15: Αναλυτική αξιολόγηση στο CK με 4 συναισθήματα.



Εικόνα 42: Ιστορικό ορθότητας και απώλειας στο CK με 4 συναισθήματα.



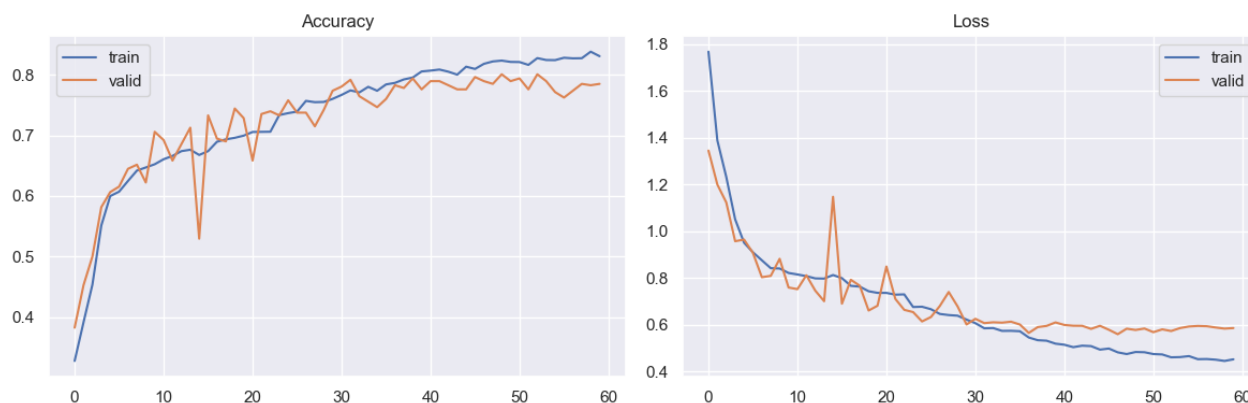
Εικόνα 43: Confusion matrix στο CK με 4 συναισθήματα.

5.4 Τελικό dataset

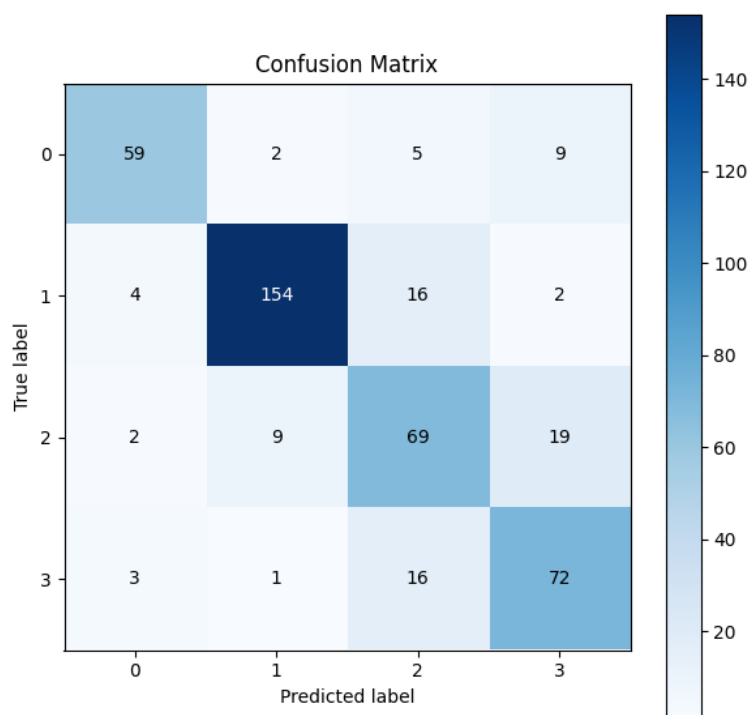
Οι δοκιμές με το τελικό dataset είχαν όλες τον ίδιο αριθμό συναισθημάτων και συγκεκριμένα περιείχαν τα συναισθήματα: θυμός, λύπη, χαρά και ουδέτερο συναίσθημα.

test accuracy: 80.9955%				
	precision	recall	f1-score	support
0	0.94	0.76	0.84	38
1	0.93	0.90	0.91	88
2	0.63	0.86	0.72	49
3	0.76	0.63	0.69	46
accuracy			0.81	221
macro avg	0.81	0.79	0.79	221
weighted avg	0.83	0.81	0.81	221
{0: 'Angry', 1: 'Happy', 2: 'Neutral', 3: 'Sadness'}				

Πίνακας 16: Αναλυτική αξιολόγηση στο τελικό dataset.



Εικόνα 44: Ιστορικό ορθότητας και απώλειας στο τελικό dataset.



Εικόνα 45: Confusion matrix στο τελικό dataset.

Ο συνδυασμός των δύο αυτών συνόλων δεδομένων (FER-2013-filtered και CK) ήταν εκείνος ο οποίος απέδωσε τα βέλτιστα αποτελέσματα στην διεπαφή. Όλες οι τιμές f1-score ήταν από 0.70 και πάνω και ο μέσος όρος ορθότητας του μοντέλου έφτασε το 80%. Παρόλο που μερικές δοκιμές είχαν καλύτερες επιδόσεις στις μετρικές, αυτό δεν αποδιδόταν απαραίτητα και στην εφαρμογή.

6 ΚΕΦΑΛΑΙΟ 6^ο : Συμπεράσματα

6.1 Γενικά συμπεράσματα

Στις δοκιμές εκπαίδευσης του μοντέλου τα σύνολα δεδομένων που χρησιμοποιήθηκαν, παρήγαγαν ένα σύνολο αποτελεσμάτων. Μια αρχική παρατήρηση ήταν πως μερικές από τις μετρικές που προέκυπταν στον έλεγχο του μοντέλου μετά την εκπαίδευση, έφεραν καλύτερα αποτελέσματα σε σχέση με εκείνα της εφαρμογής του μοντέλου στην διεπαφή χρήστη. Αυτό συνέβη κυρίως στα μοντέλα τα οποία είχαν μικρό δείγμα στο εκάστοτε συναίσθημα, με αποτέλεσμα να γίνεται overfitting. Στην ουσία τα μοντέλα αυτά μάθαιναν πολύ καλά τα συγκεκριμένα δείγματα αλλά είχαν δυσκολία να προσαρμοστούν σε άλλα σύνολα δεδομένων. Επίσης παρατηρήθηκε ότι ο αριθμός των συναισθημάτων αύξανε την πολυπλοκότητα της εκπαίδευσης, ιδιαίτερα σε συναισθήματα που μπορούσαν εύκολα να παρερμηνευθούν, όπως τα συναισθήματα φόβου και θυμού.

Πιο συγκεκριμένα, τα καλύτερα αποτελέσματα σημειώθηκαν με τα σύνολα δεδομένων FER-2013 με τρία συναισθήματα και test accuracy 81.32%, FER-2013 (filtered) με τρία συναισθήματα και test accuracy 78.89% και το τελικό σύνολο δεδομένων με τέσσερα συναισθήματα και test accuracy 80.99%. Παρά το γεγονός ότι το FER-2013 με τρία συναισθήματα είχε μεγαλύτερο test accuracy από το τελικό σύνολο δεδομένων, στην διεπαφή χρήστη το τελευταίο απέδιδε καλύτερα λόγω των περισσότερων δειγμάτων στο εκάστοτε συναίσθημα.

Οι υψηλές τιμές των μετρικών των τεσσάρων αυτών συναισθημάτων στο τελικό μοντέλο δεν αποτέλεσε το μοναδικό κριτήριο επιλογής τους. Ο βασικότερος λόγος ήταν το είδος της εφαρμογής, η οποία είχε ως σκοπό την αναπαραγωγή μουσικής και την εμφάνιση ενθαρρυντικών μηνυμάτων για να παροτρύνει τον χρήστη να διεκπεραιώσει τις εργασίες του. Συνεπώς, κρίθηκε ορθότερο να χρησιμοποιηθούν τα τέσσερα βασικά συναισθήματα χαράς, λύπης, θυμού και ουδέτερου.

6.2 Πρακτικές εφαρμογές

Η διεπαφή χρήστη που δημιουργήθηκε απευθύνεται σε ένα ευρύ φάσμα λειτουργιών. Δύναται να ενταχθεί στον τομέα της εκπαίδευσης, στον εργασιακό τομέα, αλλά μπορεί ακόμα να συμβάλλει στον τομέα της ψυχολογίας και της ψυχιατρικής.

Στον τομέα της εκπαίδευσης, η διεπαφή αυτή μπορεί να ενταχθεί σε εκπαιδευτικές πλατφόρμες, βελτιώνοντας την προσοχή των μαθητών, εντοπίζοντας τη συναισθηματική αντίδραση τους σε ένα εκπαιδευτικό πρόγραμμα και ανιχνεύοντας την συμμετοχή τους στην διαδικτυακή μάθηση. Έχει την δυνατότητα να ενθαρρύνει τους μαθητές να συνεχίσουν την προσπάθειά τους να λύσουν μια άσκηση και βοηθάει και τους καθηγητές να αναγνωρίσουν ευκολότερα την ψυχολογική κατάσταση του εκάστοτε μαθητή.

Όσον αφορά στον εργασιακό τομέα, μπορεί να συνεισφέρει στην ομαλότερη λειτουργία του εργασιακού περιβάλλοντος. Η ενθάρρυνση των εργαζομένων να διεκπεραιώσουν τα καθήκοντά τους, βελτιώνοντας τη ψυχολογική τους κατάσταση και παράγοντας έτσι περισσότερο έργο με λιγότερο κόπο, αποτελεί μια πιθανή περίπτωση χρήσης της εφαρμογής. Συνάμα, οι εργοδότες διευκολύνονται γνωρίζοντας πως η διεπαφή αυτή ενθαρρύνει το προσωπικό τους χωρίς τη δική τους παρέμβαση.

Τέλος, η διεπαφή θα μπορούσε να χρησιμοποιηθεί από ψυχολόγους και ψυχιάτρους σε ιδρύματα εντοπίζοντας το συναίσθημα του ασθενή και προλαμβάνοντας εκρήξεις όπως οι κρίσεις πανικού, τα

καταθλιπτικά και σχιζοφρενικά επεισόδια. Πέραν των ιδρυμάτων, εφικτή είναι η εφαρμογή και σε μεμονωμένα άτομα, βοηθώντας στην καταπολέμηση των αρνητικών συναισθημάτων και ενημερώνοντας κοντινά πρόσωπα για την συναισθηματική τους κατάσταση.

Συνεπώς φαίνεται πως οι πρακτικές εφαρμογές της διεπαφής είναι ποικίλες και αποσκοπούν σε κάθε περίπτωση στην πρόληψη αρνητικών καταστάσεων.

6.3 Περιορισμοί

Ωστόσο, παρά το ευρύ πεδίο εφαρμογής της διεπαφής και της μεγάλης δυνατότητας να βελτιώσει σημαντικά πολλές πτυχές της ανθρώπινης ζωής, υπάρχουν ανασταλτικοί παράγοντες που δεν επιτρέπουν την απόλυτη πρόβλεψη του συναισθήματος.

Ένας περιορισμός είναι η απόδοση του μοντέλου η οποία δεν ξεπερνά σε γενικές γραμμές το 85% ακρίβειας, απέχοντας από το επίπεδο ακρίβειας της ανθρώπινης συναισθηματικής ευφυΐας. Η απόδοση αυτή επιδέχεται περιθώρια βελτίωσης, πλησιάζοντας νούμερα κοντά στο 90% με 95%.

Άλλος ένας περιορισμός είναι ο αριθμός των συναισθημάτων που αφορά στην διεπαφή που δημιουργήθηκε, τα οποία είναι τα τέσσερα βασικά συναισθήματα. Επιλέχθηκαν για τον λόγο αυτό αλλά και λόγω του ότι πολλά από τα υπόλοιπα συναισθήματα, όπως για παράδειγμα τα συναισθήματα της έκπληξης και της αποστροφής, δεν εξυπηρετούσαν τον σκοπό της διεπαφής και δεν ωφελούσαν στην έρευνα που πραγματοποιήθηκε.

Η διεπαφή μπορεί να αναφέρεται σε ένα ευρύ πεδίο αλλά σίγουρα θα μπορούσε να επεκταθεί. Διακρίνεται ένας χαρακτήρας εξειδίκευσης ο οποίος από τη μία θα μπορούσε να διευρυνθεί αλλά από την άλλη οδήγησε στην απόδοση που αναφέρθηκε και πρόσφερε λειτουργίες οι οποίες δε θα υπήρχαν χωρίς τον χαρακτήρα αυτόν.

Οπότε παρατηρείται πως παρά το πλήθος των πρακτικών εφαρμογών που έχει η διεπαφή, υπάρχουν σημαντικοί περιορισμοί οι οποίοι πρέπει να ληφθούν υπόψη.

6.4 Μελλοντικές επεκτάσεις

Λόγω των περιορισμών που προαναφέρθηκαν, πρέπει να προταθούν κάποιες μελλοντικές επεκτάσεις που θα βελτιώσουν την διεπαφή και θα την κάνουν να αποδώσει με τον καλύτερο δυνατό τρόπο, ώστε να πλησιάσει όσο το δυνατόν γίνεται την πραγμάτωση του σκοπού της, δηλαδή την μέγιστη βελτιστοποίηση και διευκόλυνση της ανθρώπινης ζωής.

Χρησιμοποιώντας ένα καλύτερο νευρωνικό δίκτυο κατά την εκπαίδευση του μοντέλου, μπορεί να επιτευχθεί ακόμη μεγαλύτερη ακρίβεια στις προβλέψεις. Καθώς η τεχνολογία εξελίσσεται, πραγματοποιούνται μεγαλύτερες έρευνες οι οποίες πρόκειται να οδηγήσουν στην δημιουργία ενός νευρωνικού δικτύου το οποίο καλύπτει το συγκεκριμένο πρόβλημα σε μεγαλύτερο βαθμό σε σχέση με αυτό που χρησιμοποιήθηκε στην έρευνα αυτή.

Η επιλογή ενός καταλληλότερου και πιο σύγχρονου συνόλου δεδομένων το οποίο έχει μεγαλύτερο όγκο δεδομένων και καλύτερη δομή μπορεί επίσης να συνεισφέρει στην μεγαλύτερη απόδοση του μοντέλου και να επιτρέψει επέκταση σε ευρύτερο φάσμα εφαρμογών, χρησιμοποιώντας παραπάνω συναισθήματα.

Τέλος, η προσθήκη λειτουργίας παραμετροποίησης των μηνυμάτων και της μουσικής ώστε να ταιριάζουν καλύτερα ανάλογα με το πεδίο λογισμικού μπορεί να διευρύνει το φάσμα εφαρμογής της διεπαφής. Για παράδειγμα, πιθανή θα ήταν η ύπαρξη μιας ειδικής επιλογής για τον τομέα της

εκπαίδευσης με την οποία θα εμφάνιζε διαφορετικά μηνύματα και θα έπαιζε διαφορετική μουσική από ότι θα εμφανίζονταν σε ένα εργασιακό περιβάλλον.

Φαίνεται λοιπόν ότι με κάποιες αλλαγές θα μπορούσε να επεκταθεί η λειτουργία της διεπαφής διευρύνοντας το φάσμα εφαρμογής της και βελτιώνοντας την ακρίβεια του μοντέλου στις προβλέψεις που πραγματοποιεί.

Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές

1. Abdat, F. (2011). *Human-Computer Interaction Using Emotion Recognition from Facial Expression*. In: 2011 UKSim 5th European Symposium on Computer Modeling and Simulation. IEEE.
2. Bichsel, M., Pentland, A.P. (1994). *Human face recognition and the face image set's topology*. CVGIP, Image Underst. 59, 254–261
3. Cowie, R. (2001). *Emotion recognition in human-computer interaction*. In: IEEE Signal Processing Magazine 18.1, pp. 32–80.
4. Damasio, A.R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam Publishing Group, New York (1994)
5. Daugmen, J., (1988): *Complete discrete 2d Gabor transforms by neural networks for image analysis and compression*. IEEE Trans. Acoust. Speech Signal Process. 36(7), 1169–1179
6. Donato, G., Bartlett, M., Hager, J., Ekman, P., Sejnowski, T. (1999): *Classifying facial actions*. IEEE Trans. Pattern Anal. Mach. Intell. 21(10), 974–989
7. Du, S. (2014). *Compound facial expressions of emotion*. In: Proceedings of the National Academy of Sciences 111.15, E1454–E1462.
8. EDPS:European Data Protection Supervisor. (2021). *EDPS TechDispatch: facial emotion recognition. Issue 1, 2021*. Op.Europa.Eu.
9. Ekman, P., Friesen, W.V., and Ellsworth P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings* Pergamon Press, New York (1972)
10. Ekman, P. (1994). *Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique*. Psychological Bulletin, 115, pp. 268-287

11. Ekman, P. and Wallace, V. F. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.
12. Frijda, N.H. (1986). *The emotions* Cambridge University Press, Cambridge
13. Giannakas, F., Troussas, C., Voyiatzis, I., & Sgouropoulou, C. (2021). A deep learning classification framework for early prediction of team-based academic performance. *Applied Soft Computing*, 106, 107355.
14. Gomes, Carla P. (2009). *Computational sustainability: Computational methods for a sustainable environment, economy, and society*. The Bridge, 39(4):5–13. National Academy of Engineering.
15. Gray, W. D., John, B. E., and Atwood, M. E. (1993). *Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance*. Hum. Comput Interaction 8, 237–309
16. Guan, D. (2020). *Classical Architectures in CNN*. Deep Learning.
17. Kanetaki, Z., Stergiou, C., Bekas, G., Troussas, C., & Sgouropoulou, C. (2022). A Hybrid Machine Learning Model for Grade Prediction in Online Engineering Education. *International Journal of Engineering Pedagogy (iJEP)*, 12(3), 4–24.
18. Karat, C. M. (1990). *Cost-benefit analysis of usability engineering techniques*. In Proceedings of the Human Factors Society 34th Annual Meeting (Orlando, FL, Oct.), Vol. 2, 839–843.
19. Ko, B. C. (2018). *A brief review of facial emotion recognition based on visual information*. In: Sensors 18.2, p. 401.
20. Koehn, Philipp, Och, Franz Josef, and Marcu, Daniel. (2003). *Statistical phrase-based translation*. In Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pp. 48–54, 2003.

21. Kontellis, E., Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Real-Time Face Mask Detector Using Convolutional Neural Networks Amidst COVID-19 Pandemic. *Frontiers in Artificial Intelligence and Applications*.
22. Krouska, A., Troussas, C., & Virvou, M. (2016). The effect of preprocessing techniques on Twitter sentiment analysis. *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*.
23. Krouska, A., Troussas, C., & Virvou, M. (2017). *Comparative Evaluation of Algorithms for Sentiment Analysis over Social Networking Services*. Journal of Universal Computer Science, vol. 23, pp. 755-768
24. Lang, Peter J. (1993). *Looking at pictures: Affective, facial, visceral, and behavioral reactions*. In: Psychophysiology 30.3, pp. 261–273
25. Mantei, M. M. and Teorey, T. J. (1988). *Cost/benefit analysis for incorporating human factors in the software lifecycle*. Commun. ACM31, 4 (April) 428–439.
26. McCarthy J. (1956). *Measures of the value of information*. Proceedings of the National Academy of Sciences of the United States of America, 42(9), 654–655.
27. Moghaddam, B., Pentland, A. (1997). *Probabilistic visual learning for object representation*. IEEE Trans. Pattern Anal. Mach. Intell. 19(7), 696–710
28. Myers, B. A. and Rosson, M. B. (1992). *Survey on user interface programming*. In *Proceedings SIGCHI'92: Human Factors in Computing Systems* (Monterey, CA, May 3–7), 195–202.
29. Nielsen, J. (1993). *Usability Engineering*, Academic Press, Boston
30. Nielsen, J. and Landauer, T. K. (1993). *A mathematical model of the finding of usability problems*. In *Proceedings INTERCHI'93: Human Factors in Computing Systems* (Amsterdam, The Netherlands, April), 206-213

31. Oatley, K., Johnson-Laird, P.N. (1987). *Towards a cognitive theory of emotions*. Cognition and Emotion, 1 pp. 29-50
32. Osuna, E., Freund, R., Girosi, F. (1997): *Training support vector machines: An application to face detection*. In: CVPR, pp. 130–136
33. Pyo, S. (2017). *Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets*. PLOS ONE.
34. Rathgeb, C., Tolosana, R., Vera-Rodriguez, R., & Busch, C. (2022). *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks (Advances in Computer Vision and Pattern Recognition) (1st ed. 2022 ed.)*. Springer.
35. Rowley, H.A., Baluja, S., Kanade, T. (1998): *Neural network-based face detection*. IEEE Trans. Pattern Anal. Mach. Intell. 20(1), 23–28
36. Russell, J.A. (1994) *Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies*. Psychological Bulletin, 115, pp. 102-141
37. Salovey, P., and Mayer, J.D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185-211
38. Scarr, S. (1985). *An author's frame of mind [Review of Frames of Mind by H. Gardner]*. New Ideas in Psychology, 3, 95-100.
39. Scarr, S. (1989). *Protecting general intelligence: Constructs and consequences for interventions*. In R.L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy*. Urbana: University of Illinois Press.
40. Sung, K., Poggio, T. (1998): *Example-based learning for view-based human face detection*. IEEE Trans. Pattern Anal. Mach. Intell. 20(1), 39–51.
41. Tomkins S.S. (1962). *Affect, imagery, consciousness* Springer, New York

42. Troussas, C., Chrysafiadi, K., & Virvou, M. (2018). *Machine Learning and Fuzzy Logic Techniques for Personalized Tutoring of Foreign Languages*. SpringerLink.
43. Troussas, C., Espinosa, K. J., & Virvou, M. (2016). Affect Recognition through Facebook for Effective Group Profiling Towards Personalized Instruction. *Informatics in Education*, 15(1), 147–161.
44. Troussas, C., Krouska, A., Giannakas, F., Sgouropoulou, C., & Voyiatzis, I. (2020). Automated reasoning of learners' cognitive states using classification analysis. *24th Pan-Hellenic Conference on Informatics*.
45. Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Enhancing Human-Computer Interaction in Digital Repositories through a MCDA-Based Recommender System. *Advances in Human-Computer Interaction, 2021*, 1–7.
46. Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Improving Learner-Computer Interaction through Intelligent Learning Material Delivery Using Instructional Design Modeling. *Entropy*, 23(6), 668.
47. Troussas, C., Krouska, A., & Virvou, M. (2016). Evaluation of ensemble-based sentiment classifiers for Twitter data. *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*.
48. Troussas, C., Krouska, A., & Virvou, M. (2018). *Multi-Algorithmic Techniques and a Hybrid Model for Increasing the Efficiency of Recommender Systems*. IEEE In: 30th International Conference on Tools with Artificial Intelligence.
49. Troussas, C., Krouska, A., & Virvou, M. (2018). A Multicriteria Framework for Assessing Sentiment Analysis in Social and Digital Learning: Software Review. *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*.
50. Troussas, C., Krouska, A., Virvou, M. (2019). *Trends on Sentiment Analysis over Social Networks: Pre-processing Ramifications, Stand-Alone Classifiers and Ensemble Averaging*. In: Tsihrintzis, G., Sotiropoulos, D., Jain, L. (eds) *Machine Learning Paradigms*. Intelligent Systems Reference Library, vol 149 . Springer, Cham.

51. Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. *IISA 2013*.
52. Troussas, C., Virvou, M., & Mesaretzidis, S. (2015). Comparative analysis of algorithms for student characteristics classification using a methodological framework. *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*.
53. Troussas, C., & Virvou, M. (2021). *Advances in Social Networking-based Learning: Machine Learning-based User Modelling and Sentiment Analysis (Intelligent Systems Reference Library)* (1st ed. 2020 ed.). Springer.
54. Turk, M.A., Pentland, A.P. (1991). *Eigenfaces for recognition*. *J. Cogn. Neurosci.* 3(1), 71-86
55. Valagkouti, I. A., Troussas, C., Krouska, A., Feidakis, M., & Sgouropoulou, C. (2022). Emotion Recognition in Human–Robot Interaction Using the NAO Robot. *Computers*, 11(5), 72.
56. Virvou, M., Alepis, E., & Troussas, C. (2012). *User Modeling on Communication Characteristics Using Machine Learning in Computer-Supported Collaborative Multiple Language Learning*. 2012 IEEE 24th International Conference on Tools with Artificial Intelligence.
57. Yang, M. H., Roth, D., Ahuja, N. (2000): *A SNoW-based face detector*. In: Proceedings of Neural Information Processing Systems, pp. 855–861
58. Zdziarski, Jonathan A. (2005). *Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification*. No Starch Press, San Francisco
59. Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S., (1998): *Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron*. In: International Workshop on Automatic Face and Gesture Recognition, pp. 454–459