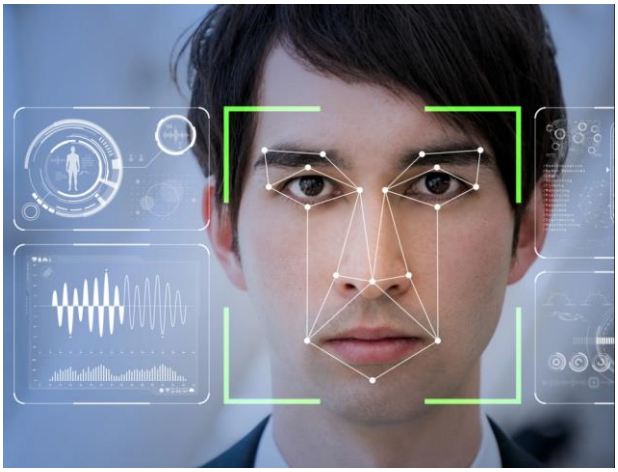
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Master of Science Thesis

Develop a deep learning model towards user's emotion recognition from facial expressions



Student: Maros Grigorios
Registration Number: AIDL-0008

MSc Thesis Supervisor
Feidakis Michail

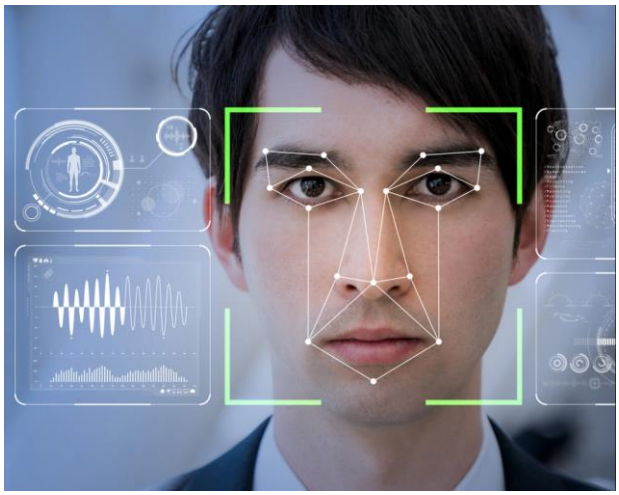
ATHENS-EGALEO, February 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Μεταπτυχιακή Διπλωματική Εργασία

Ανάπτυξη Μοντέλου Βαθιάς Μηχανικής Μάθησης για την αναγνώρισης της συναισθηματικής κατάστασης του χρήστη από τις εκφράσεις προσώπου



Φοιτητής: Μάρος Γρηγόριος
ΑΜ: AIDL-0008

Επιβλέπων Καθηγητής
Φειδάκης Μιχαήλ
Ε.ΔΙ.Π,

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, Φεβρουάριος 2023

This MSc Thesis has been accepted, evaluated, and graded by the following committee:

Supervisor	Member	Member
Feidakis Michail*	**Patrikakis Charalampos***	**Nikolaou Grigorios****
Michail Feidakis 2023.03.26 18:40:44 +03'00'	Digitally signed by Charalampos Patrikakis Date: 2023.03.27 11:24:32 +03'00'	GRIGORIOS NIKOLAOU Digitally signed by GRIGORIOS NIKOLAOU Date: 2023.03.27 10:13:22 +03'00'
*Department of Electrical & Electronics Engineering **Department of Industrial Design and Production Engineering Faculty of Engineering, University of West Attica		

Declaration of the author of this MSc thesis

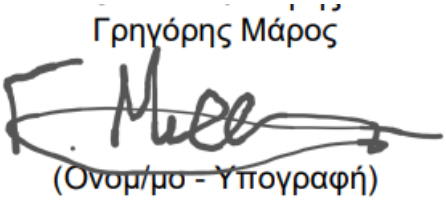
I, Grigorios Evangelos Maros with the following student registration number:0008, postgraduate student of the MSc program in "Artificial Intelligence and Deep Learning", which is organized by the Department of Electrical and Electronic Engineering and the Department of Industrial Design and Production Engineering of the Faculty of Engineering of the University of West Attica, hereby declare that:

I am the author of this MSc thesis and any help I may have received is clearly mentioned in the thesis. Additionally, all the sources I have used (e.g., to extract data, ideas, words, or phrases) are cited with full reference to the corresponding authors, the publishing house or the journal; this also applies to the Internet sources that I have used. I also confirm that I have personally written this thesis and the intellectual property rights belong to myself and to the University of West Attica. This work has not been submitted for any other degree or professional qualification except as specified in it.

Any violations of my academic responsibilities, as stated above, constitutes substantial reason for the cancellation of the conferred MSc degree.

I wish to deny access to the full text of my MSc thesis until **01/04/2023**, following my application to the Library of UNIWA and the approval from my supervisor.

The author
Grigorios Maros



ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος **Γρηγόριος Μάρος** του **Ευαγγέλου**, με αριθμό μητρώου 0008 μεταπτυχιακός φοιτητής του ΔΠΜΣ «Τεχνητή Νοημοσύνη και Βαθιά Μάθηση» του Τμήματος Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών και του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής, της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής διπλωματικής εργασίας και 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. Η εργασία δεν έχει κατατεθεί στο πλαίσιο των απαιτήσεων για τη λήψη άλλου τίτλου σπουδών ή επαγγελματικής πιστοποίησης πλην του παρόντος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.
Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι **01/04/2023** και έπειτα από αίτησή μου στη Βιβλιοθήκη και έγκριση του επιβλέποντος καθηγητή.»

Ο Δηλών
Γρηγόριος Μάρος

Γρηγόρης Μάρος
(Ονομ/μο - Υπογραφή)

Η διπλωματική αυτή εργασία είναι αφιερωμένη στην σύζυγό μου Μαρίνα Μητρογιώργου που με στήριξε και με βοήθησε να φτάσω στο τέλος αυτού του εκπαιδευτικού ταξιδιού.

Με το τέλος της διπλωματικής αυτής εργασίας θέλω να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Φειδάκη Μιχαήλ του Τμήματος Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών, για τις πολύτιμες συμβουλές και την καθοδήγησή του κατά την προετοιμασία της εργασίας αυτής. Επίσης, ένα μεγάλο ευχαριστώ στον συνάδελφο Άγγελο Αντικατζίδη για την βοήθεια του και την ενεργή συμμετοχή του, στο πρακτικό μέρος της εργασίας και την εφαρμογή της σε πραγματικές καταστάσεις.
Τέλος, ευχαριστώ την οικογένεια μου για την υποστήριξη τους, όχι μόνο για την ολοκλήρωση της διπλωματικής μου εργασίας αλλά και για την συμπαράσταση που μου επέδειξαν σε όλη την διάρκεια των μεταπτυχιακών μου σπουδών.

Abstract

Emotion recognition from facial expressions is a challenging task that has been the subject of extensive research in recent years. In this thesis, we develop a deep learning model for user's emotion recognition from facial expressions using two publicly available datasets: AffectNet and Fer2013. We focus on convolutional neural networks (CNN) as our primary method for training the model, which has shown excellent results in previous research on facial emotion recognition. Our goal is to explore the potential of deep learning in this domain and to create a model that can accurately classify a user's emotions from facial images. We review and analyze recent research in the field of deep learning and emotion recognition, and we propose a CNN architecture that includes several layers of convolutions and pooling, followed by fully connected layers. We evaluate the performance of our model on the two datasets and compare it with the state-of-the-art methods. Although our model did not outperform the existing methods in terms of accuracy, our results show that it achieved competitive performance and provides an alternative approach to the problem. We also investigate the impact of different parameters, such as batch size, learning rate, and dropout, on the model's accuracy. This thesis contributes to the field of emotion recognition and deep learning by providing a comprehensive analysis of the effectiveness of CNNs for facial emotion recognition and proposing an efficient and accurate model, as well as identifying areas for further improvement in this research field.

Περίληψη

Η αναγνώριση συναισθημάτων από τις εκφράσεις του προσώπου είναι ένα δύσκολο έργο που έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας τα τελευταία χρόνια. Στην παρούσα διπλωματική, αναπτύσσουμε ένα μοντέλο βαθιάς μάθησης για την αναγνώριση των συναισθημάτων του χρήστη από τις εκφράσεις του προσώπου χρησιμοποιώντας δύο διαθέσιμα σύνολα δεδομένων: Affectnet και Fer2013. Εστιάζουμε στα συνελικτικά νευρωνικά δίκτυα (CNN) ως την κύρια μέθοδο μας για την εκπαίδευση του μοντέλου, το οποίο έχει δείξει εξαιρετικά αποτελέσματα σε προηγούμενες έρευνες για την αναγνώριση συναισθημάτων από το πρόσωπο. Στόχος μας είναι να διερευνήσουμε τις δυνατότητες της βαθιάς μάθησης σε αυτόν τον τομέα και να δημιουργήσουμε ένα μοντέλο που μπορεί να ταξινομήσει με ακρίβεια τα συναισθήματα ενός χρήστη από τις εικόνες του προσώπου. Εξετάζουμε και αναλύουμε πρόσφατες έρευνες στον τομέα της βαθιάς μάθησης και της αναγνώρισης συναισθημάτων και προτείνουμε μια αρχιτεκτονική CNN. Αξιολογούμε την απόδοση του μοντέλου μας στα δύο σύνολα δεδομένων και τη συγκρίνουμε με άλλες μεθόδους αιχμής. Αν και το μοντέλο μας δεν ξεπέρασε τις υπάρχουσες μεθόδους όσον αφορά την ακρίβεια, τα αποτελέσματά μας δείχνουν ότι πέτυχε ανταγωνιστικές επιδόσεις και παρέχει μια εναλλακτική προσέγγιση στο πρόβλημα. Διερευνούμε επίσης την επίδραση διαφορετικών παραμέτρων, όπως το μέγεθος παρτίδας (batch size), ο ρυθμός εκμάθησης (learning rate) και η εγκατάλειψη (dropout), στην ακρίβεια του μοντέλου. Η παρούσα διπλωματική εργασία συμβάλλει στον τομέα της αναγνώρισης συναισθημάτων και της βαθιάς μάθησης παρέχοντας μια ολοκληρωμένη ανάλυση της αποτελεσματικότητας των CNN για την αναγνώριση συναισθημάτων προσώπου, και προτείνει ένα αποτελεσματικό και ακριβές μοντέλο, εντοπίζοντας τομείς για περαιτέρω βελτίωση σε αυτό το ερευνητικό πεδίο.

Λέξεις – κλειδιά

Βαθιά μάθηση, αναγνώριση, συναισθηματική, κατάσταση, νοημοσύνη, AffectNet, deep learning, emotional, intelligence, state , recognition

Περιεχόμενα

1. Εισαγωγή

1.1 Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μάθηση

Η Τεχνητή Νοημοσύνη (Artificial Intelligence aka AI) αποτελεί ευρύ πεδίο μελέτης το οποίο περιλαμβάνει την προσομοίωση διαδικασιών ανθρώπινης νοημοσύνης από συστήματα υπολογιστών. Αυτές οι διαδικασίες περιλαμβάνουν τη μάθηση (την απόκτηση πληροφοριών και τους κανόνες για τη χρήση των πληροφοριών), τη συλλογιστική (χρησιμοποιώντας τους κανόνες για να καταλήξουμε σε κατά προσέγγιση ή οριστικά συμπεράσματα) και την αυτό-διόρθωση. Η τεχνητή νοημοσύνη υπάρχει εδώ και πολλές δεκαετίες και περιλαμβάνει πολλά υπό-πεδία, όπως η Μηχανική Μάθηση (Machine Learning aka ML) και η Βαθιά Μάθηση (Deep Learning aka DL) [1-3].

Η ML είναι ένα υπό-πεδίο της AI που περιλαμβάνει την εκπαίδευση ενός συστήματος υπολογιστή σε ένα μεγάλο σύνολο δεδομένων, επιτρέποντάς του να βελτιώνει αυτόματα την απόδοσή του σε μια εργασία χωρίς να είναι ρητά προγραμματισμένο Αυτό γίνεται με τη χρήση αλγορίθμων που μπορούν να αναγνωρίσουν μοτίβα στα δεδομένα και να κάνουν προβλέψεις ή αποφάσεις χωρίς ανθρώπινη παρέμβαση. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να χωριστούν σε τρεις κατηγορίες: εποπτευόμενη (Supervised Learning), χωρίς επίβλεψη (Unsupervised Learning) και ενισχυτική μάθηση (Reinforcement Learning). Οι εποπτευόμενοι αλγόριθμοι μάθησης εκπαιδεύονται σε χαρακτηρισμένα δεδομένα και έπειτα τα χρησιμοποιούν για να κάνουν προβλέψεις ή να καταλήξουν σε αποφάσεις για νέα, αόρατα δεδομένα. Οι αλγόριθμοι μάθησης χωρίς επίβλεψη εκπαιδεύονται σε δεδομένα μη χαρακτηρισμένα και χρησιμοποιούνται για τον εντοπισμό προτύπων και δομής στα δεδομένα. Οι αλγόριθμοι ενίσχυσης εκμάθησης εκπαιδεύονται σε ένα σήμα ενίσχυσης (ανταμοιβής) (Reward Signal) και χρησιμοποιούνται για τη βελτιστοποίηση των αποφάσεων σε δυναμικά περιβάλλοντα [4-6].

Η DL είναι ένα υποσύνολο της Μηχανικής Μάθησης που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks -ANNs) με πολλαπλά επίπεδα για να μαθαίνει και να παίρνει αποφάσεις. Αυτά τα νευρωνικά δίκτυα έχουν σχεδιαστεί για να μιμούνται τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου, επιτρέποντάς τους να μαθαίνουν από τεράστιες ποσότητες δεδομένων και να λαμβάνουν προβλέψεις ή αποφάσεις με υψηλό βαθμό ακρίβειας . Οι αλγόριθμοι βαθιάς μάθησης βασίζονται στην έννοια των τεχνητών νευρώνων, οι οποίοι εμπνέονται από τη δομή και τη λειτουργία των βιολογικών νευρώνων. Αυτοί οι τεχνητοί νευρώνες είναι οργανωμένοι σε στρώματα και συνδέονται μεταξύ τους μέσω συναπτικών βαρών (weights). Τα βάρη προσαρμόζονται κατά τη διάρκεια της διαδικασίας εκμάθησης και χρησιμοποιούνται για τη χαρτογράφηση εισόδων σε εξόδους (mapping). Οι αλγόριθμοι βαθιάς μάθησης μπορούν να χωριστούν σε τρεις κύριες κατηγορίες: τροφοδοτικά (Feedforward Neural Networks-FNNs), επαναλαμβανόμενα (Recurrent Neural Networks - RNNs) και συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs). Τα FNNs χρησιμοποιούνται για τη χαρτογράφηση εισόδων σε εξόδους, τα RNNs χρησιμοποιούνται για την επεξεργασία διαδοχικών δεδομένων και τα CNNs χρησιμοποιούνται για την επεξεργασία εικόνας και βίντεο [7-8].

Η κύρια διαφορά μεταξύ AI, ML και DL είναι η πολυπλοκότητά τους και το επίπεδο ανθρώπινης παρέμβασης που απαιτείται. Η τεχνητή νοημοσύνη είναι ο ευρύτερος όρος, που περιλαμβάνει όλες τις μορφές συστημάτων υπολογιστών που μπορούν να εκτελέσουν εργασίες που κανονικά θα απαιτούσαν ανθρώπινη νοημοσύνη. Η Μηχανική Μάθηση είναι ένα υπό-πεδίο της τεχνητής νοημοσύνης που βασίζεται σε αλγόριθμους και στατιστικά μοντέλα για τη βελτίωση της απόδοσης σε μια εργασία χωρίς να είναι ρητά προγραμματισμένος. Η βαθιά μάθηση είναι ένα υπό-πεδίο της μηχανικής μάθησης που χρησιμοποιεί βαθιά νευρωνικά δίκτυα (DNN) για να βελτιώσει την απόδοση σε μια εργασία. Η βαθιά μάθηση θεωρείται πιο σύνθετη από τη μηχανική μάθηση και απαιτεί περισσότερα δεδομένα, περισσότερη υπολογιστική ισχύ και περισσότερη τεχνογνωσία για την ανάπτυξη και τη συντήρηση του αλγορίθμου. [1] [3] [4]

Εικόνα 1: AI vs ML vs DL definition by Stanford University (Ανακτήθηκε 15/02/2023 από https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf)

1.2 Βαθιά Μάθηση, Υπολογιστική όραση & αναγνώριση συναισθημάτων

Η υπολογιστική όραση (Computer Vision) και η βαθιά μάθηση (Deep Learning) έχουν γίνει όλο και πιο δημοφιλή τα τελευταία χρόνια, με ένα ευρύ φάσμα εφαρμογών σε τομείς όπως η ανάλυση εικόνας και βίντεο, η ανίχνευση αντικειμένων και η αναγνώριση προσώπου. Στον τομέα της συναισθηματικής νοημοσύνης, η αναγνώριση εκφράσεων προσώπου έχει εξελιχθεί σε σημαντικό πεδίο έρευνας, καθώς η ικανότητα να ερμηνεύει κανείς με ακρίβεια και να ανταποκρίνεται στα συναισθήματα των άλλων αποτελεί βασική πτυχή της συναισθηματικής νοημοσύνης. [7] [9]

Οι εκφράσεις του προσώπου είναι μια καθολική και ισχυρή μορφή μη λεκτικής επικοινωνίας και έχει μελετηθεί εκτενώς στον τομέα της ψυχολογίας. Οι ερευνητές έχουν εντοπίσει βασικές εκφράσεις του προσώπου, όπως της ευτυχίας, της λύπης, του θυμού, του φόβου, της έκπληξης και της αηδίας, που αναγνωρίζονται και ερμηνεύονται παρόμοια μεταξύ των πολιτισμών. Αυτές οι βασικές εκφράσεις του προσώπου αναφέρονται συχνά ως «καθολικές εκφράσεις» και θεωρούνται ότι είναι έμφυτες και καθολικά κατανοητές από σχεδόν όλους τους ανθρώπους. [10-12]

Ωστόσο, τα ανθρώπινα συναισθήματα δεν περιορίζονται σε αυτά και οι άνθρωποι μπορούν επίσης να εκφράσουν πιο περίπλοκα συναισθήματα μέσω λεπτών παραλλαγών στις εκφράσεις του προσώπου τους. Αυτό καθιστά το έργο της αναγνώρισης και ερμηνείας των εκφράσεων ένα δύσκολο πρόβλημα για τα αυτοματοποιημένα συστήματα. Τα τελευταία χρόνια, η πρόοδος στην όραση των υπολογιστών και τη βαθιά μάθηση, κατέστησαν δυνατή την ανάπτυξη αυτοματοποιημένων συστημάτων για την αναγνώριση και την ερμηνεία των εκφράσεων, τα

οποία έχουν τη δυνατότητα να χρησιμοποιηθούν σε ποικίλες εφαρμογές, όπως είναι π.χ. η αξιολόγηση της συναισθηματικής νοημοσύνης. Η συναισθηματική νοημοσύνη είναι η ικανότητα να αναγνωρίζεις, να κατανοείς και να διαχειρίζεσαι τα δικά σου συναισθήματα, καθώς και τα συναισθήματα των άλλων. Έχει βρεθεί ότι είναι βασικός προγνωστικός παράγοντας επιτυχίας τόσο στην προσωπική όσο και στην επαγγελματική ζωή και έχει συνδεθεί με ένα ευρύ φάσμα θετικών αποτελεσμάτων, όπως καλύτερη ψυχική και σωματική υγεία, καλύτερες σχέσεις και υψηλότερη εργασιακή απόδοση. Η τεχνολογία αναγνώρισης εκφράσεων προσώπου έχει τη δυνατότητα να χρησιμοποιηθεί στην ανάπτυξη προγραμμάτων και εργαλείων εκπαίδευσης συναισθηματικής νοημοσύνης, τα οποία θα μπορούσαν να βοηθήσουν τα άτομα να βελτιώσουν τη συναισθηματική τους νοημοσύνη. [13-16]

Μια άλλη πιθανή εφαρμογή της αναγνώρισης εκφράσεων προσώπου είναι στον τομέα της ψυχικής υγείας. Οι διαταραχές ψυχικής υγείας όπως η κατάθλιψη, το άγχος και το PTSD (Post-traumatic stress disorder) συχνά χαρακτηρίζονται από αλλαγές στις εκφράσεις και αυτοματοποιημένα συστήματα αναγνώρισης και ερμηνείας αυτών των αλλαγών θα μπορούσαν να χρησιμοποιηθούν για να βοηθήσουν στη διάγνωση και τη θεραπεία αυτών των διαταραχών. Επιπλέον, η τεχνολογία αυτή θα μπορούσε να χρησιμοποιηθεί για την παρακολούθηση της αποτελεσματικότητας της θεραπείας για διαταραχές ψυχικής υγείας, παρακολουθώντας τις αλλαγές στις εκφράσεις του προσώπου με την πάροδο του χρόνου. [17-18]

Επίσης πιθανές εφαρμογές υπάρχουν στους τομείς της ασφάλειας και της επιτήρησης, της αλληλεπίδρασης ανθρώπου-υπολογιστή και της κοινωνικής ρομποτικής. Στην ασφάλεια και την επιτήρηση, η αναγνώριση εκφράσεων προσώπου θα μπορούσε να χρησιμοποιηθεί για τον εντοπισμό και την απόκριση σε πιθανές απειλές για την ασφάλεια, όπως άτομα που εμφανίζουν σημάδια θυμού ή επιθετικότητας. Στην αλληλεπίδραση ανθρώπου-υπολογιστή, η αναγνώριση εκφράσεων προσώπου θα μπορούσε να χρησιμοποιηθεί για τη βελτίωση της φυσικότητας και της αποτελεσματικότητας της επικοινωνίας μεταξύ ανθρώπων και υπολογιστών. Στην κοινωνική ρομποτική, θα μπορούσε να χρησιμοποιηθεί για να επιτρέψει στα ρομπότ να αναγνωρίζουν και να ανταποκρίνονται κατάλληλα στα ανθρώπινα συναισθήματα. [19-21]

Στόχος της παρούσας διπλωματικής εργασίας είναι να διερευνήσει τη χρήση τεχνικών όρασης υπολογιστή και βαθιάς μάθησης για την αναγνώριση εκφράσεων προσώπου και την αξιοποίησή τους σε άλλα συστήματα.

1.3 Δήλωση προβλήματος (Problem State)

Παρά τις προόδους στους τομείς της υπολογιστικής όρασης και της βαθιάς μάθησης, η αναγνώριση και ερμηνεία των εκφράσεων παραμένει ένα δύσκολο πρόβλημα για τα αυτοματοποιημένα συστήματα . Υπάρχουν πολλά ζητήματα που πρέπει να αντιμετωπιστούν για να βελτιωθεί η ακρίβεια των συστημάτων αυτών, όπως οι παραλλαγές λόγω ηλικίας, φύλου και εθνικότητας, καθώς και οι διακυμάνσεις του φωτισμού και των συνθηκών φόντου. Επιπλέον, η έλλειψη μεγάλων, διαφορετικών συνόλων δεδομένων (datasets) εκφράσεων του προσώπου καθιστά δύσκολη την εκπαίδευση μοντέλων βαθιάς μάθησης που μπορούν να γενικευθούν (generalization) καλά σε διαφορετικούς πληθυσμούς . Τέλος, ενώ ο τομέας αυτός έχει γνωρίσει σημαντικές προόδους τα τελευταία χρόνια, οι εφαρμογές αυτών των τεχνολογιών στον τομέα της αναγνώρισης συναισθηματικής κατάστασης παραμένουν σχετικά ανεξερεύνητες. [22-25]

1.4 Ερευνητικά ερωτήματα

Τα ερωτήματα που καλείται να απαντήσει η διπλωματική εργασία είναι τα εξής:

I. Ποιες είναι οι παράμετροι απόδοσης του μοντέλου βαθιάς μάθησης στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο;;
II. Ποιες είναι οι δυνατότητες και οι περιορισμοί του μοντέλου βαθιάς μάθησης στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο και σε συνδυασμό με άλλες εφαρμογές όπως η παρακολούθηση βίντεο ή η διαδικτυακή εκπαίδευση;
III. Ποιος είναι ο αντίκτυπος διαφορετικών υπερ-παραμέτρων, όπως ο ρυθμός εκμάθησης (Learning Rate), το μέγεθος παρτίδας (Batch Size) και η εγκατάλειψη (Dropout), στην απόδοση του μοντέλου σε σενάρια πραγματικού κόσμου;
IV. Πώς συγκρίνεται η απόδοση του μοντέλου με άλλες μεθόδους τελευταίας τεχνολογίας στον τομέα της αναγνώρισης συναισθημάτων προσώπου όταν εφαρμόζεται σε πραγματικό σενάριο;

Για να απαντηθούν οι παραπάνω ερωτήσεις, πρέπει να αξιολογηθούν διάφορες παράμετροι απόδοσης, όπως η ακρίβεια της αναγνώρισης συναισθημάτων, ο χρόνος απόκρισης του μοντέλου και η δυνατότητα αναγνώρισης σε πραγματικό χρόνο. Επίσης, θα πρέπει να εξεταστούν οι δυνατότητες και οι περιορισμοί του μοντέλου στην αναγνώριση συναισθημάτων σε συνδυασμό με άλλες εφαρμογές, όπως η παρακολούθηση βίντεο ή η διαδικτυακή εκπαίδευση. Αξιολογώντας διαφορετικές υπερ-παραμέτρους, μπορούμε να εκτιμήσουμε τον αντίκτυπό τους στην απόδοση του μοντέλου σε σενάρια πραγματικού κόσμου. Τέλος, μπορούν να χρησιμοποιηθούν δοκιμαστικά πλαίσια εκπαίδευσης και επαλήθευσης με δεδομένα που έχουν καταγραφεί από ανθρώπους σε πραγματικές συνθήκες, καθώς και δεδομένα αναφοράς που έχουν συλλεχθεί σε ελεγχόμενες συνθήκες.

1.5 Σημαντικότητα της εργασίας

Πρώτον, συμβάλλει στον αυξανόμενο όγκο έρευνας σε διάφορους τομείς όπως η αλληλεπίδραση ανθρώπου-υπολογιστή, η τεχνητή συναισθηματική νοημοσύνη (Affective Computing) και η ψυχική υγεία. [26]

Δεύτερον, προτείνεται ένα μοντέλο βαθιάς μάθησης το οποίο μπορεί να αναγνωρίσει τα συναισθήματα ενός χρήστη από τις εκφράσεις του προσώπου. Το μοντέλο αυτό μπορεί να είναι ωφέλιμο σε εφαρμογές του πραγματικού κόσμου, όπως η διαδικτυακή εκπαίδευση και το μάρκετινγκ. [27]

Τρίτον, η μελέτη συνδυάζει δύο δημόσια διαθέσιμα σύνολα δεδομένων (AffectNet και FER2013) για να βελτιώσει την ακρίβεια του μοντέλου. [28]

Τέταρτον, στα πλαίσια της εργασίας η δοκιμασία του μοντέλου σε ένα πραγματικό σενάριο χρησιμοποιώντας υπολογιστική όραση, μπορεί να παρέχει πληροφορίες για την απόδοση των μοντέλων βαθιάς μάθησης, σε πρακτικές εφαρμογές. Επίσης μπορεί επίσης να βοηθήσει στον εντοπισμό πιθανών προκλήσεων και περιορισμών που μπορεί να μην είναι εμφανείς σε εργαστηριακά πειράματα, όπως η μεταβλητότητα στο φωτισμό, η γωνία της κάμερας και οι εκφράσεις του προσώπου. [29]

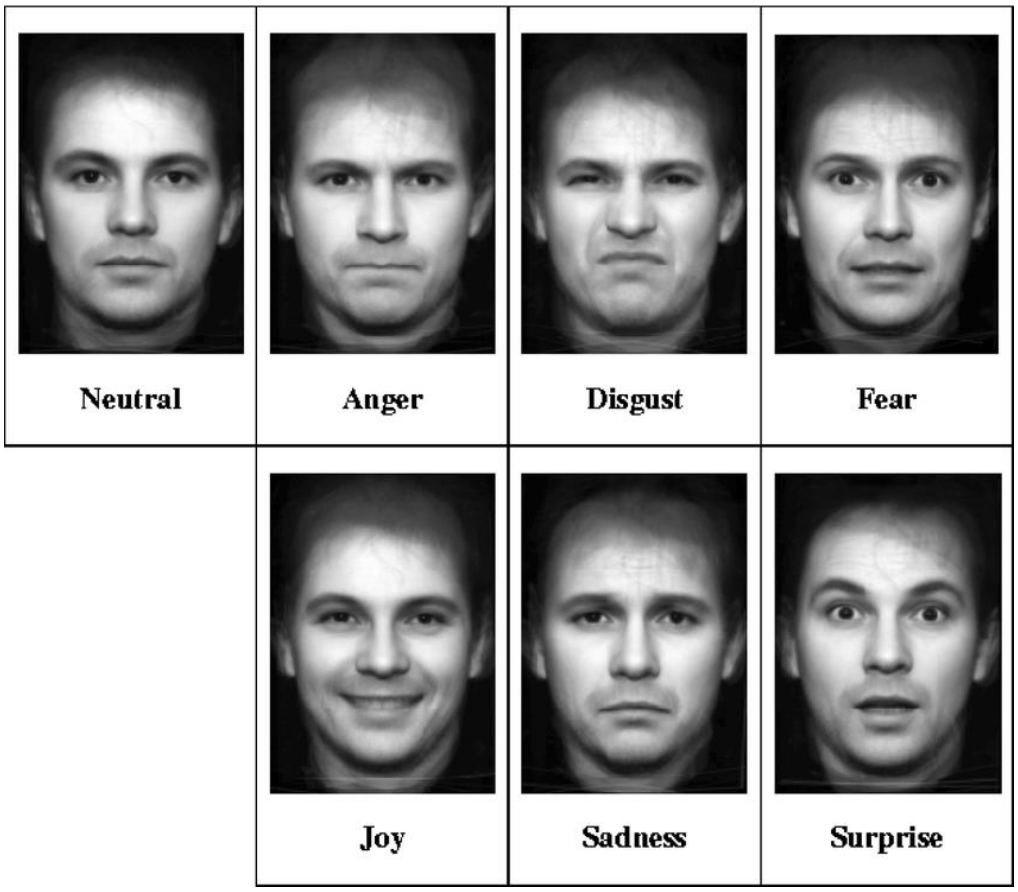
Τέλος, η μελέτη προσδιορίζει πιθανές κατευθύνσεις έρευνας για τη βελτίωση της ακρίβειας και της αποτελεσματικότητας των μοντέλων βαθιάς μάθησης για την αναγνώριση συναισθημάτων, που μπορούν να καθοδηγήσουν τη μελλοντική έρευνα στο πεδίο .

2. Βιβλιογραφική επισκόπηση

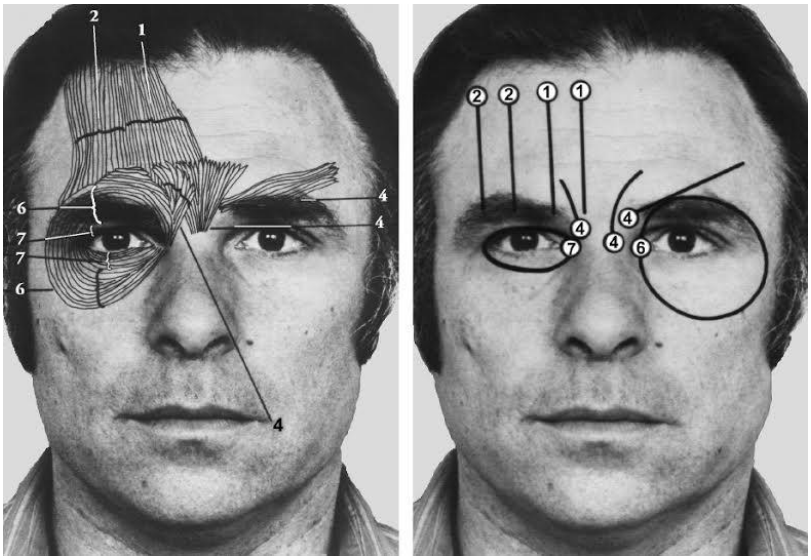
2.1 Επισκόπηση της μελέτης των εκφράσεων του προσώπου και της συναισθηματικής νοημοσύνης

Η μελέτη των εκφράσεων του προσώπου και της συναισθηματικής νοημοσύνης έχει μακρά ιστορία, που χρονολογείται από τις αρχές του 20ου αιώνα. Η πρώιμη έρευνα σε αυτόν τον τομέα επικεντρώθηκε στον εντοπισμό και την περιγραφή των βασικών εκφράσεων του προσώπου του συναισθήματος, όπως η ευτυχία, η λύπη, ο θυμός, ο φόβος, η έκπληξη και η αηδία. Ο Κάρολος Δαρβίνος, στο βιβλίο του «The Expression of Emotions in Man and Animals» το 1872, ήταν μεταξύ των πρώτων που πρότεινε ότι οι εκφράσεις του προσώπου του συναισθήματος είναι καθολικές μεταξύ των πολιτισμών. Πρότεινε ότι ορισμένες κινήσεις του προσώπου είναι έμφυτες και ότι συνδέονται με συγκεκριμένα συναισθήματα. [30-32]

Στη δεκαετία του 1950, οι Ekman και Friesen άρχισαν να μελετούν συστηματικά τις εκφράσεις του προσώπου του. Χρησιμοποίησαν φωτογραφία και φιλμ για να μελετήσουν τις κινήσεις των μυών του προσώπου που σχετίζονται με διαφορετικά συναισθήματα. Διαπίστωσαν ότι υπήρχαν συγκεκριμένες κινήσεις που αντιστοιχούσαν σε κάθε βασικό συναίσθημα. Αυτή η εργασία έθεσε τα θεμέλια για τη μελέτη της καθολικότητας των συναισθημάτων από τις εκφράσεις του προσώπου. Στις δεκαετίες του 1970 και του 1980, ο Paul Ekman και οι συνεργάτες του ανέπτυξαν το Facial Action Coding System (FACS) ως εργαλείο για την αντικειμενική μέτρηση και κωδικοποίηση των εκφράσεων του προσώπου. Το FACS είναι ένα ολοκληρωμένο σύστημα που περιγράφει τις συγκεκριμένες μυϊκές κινήσεις που αντιστοιχούν σε διαφορετικά συναισθήματα. Αυτό το σύστημα έχει χρησιμοποιηθεί ευρέως σε έρευνες για τις εκφράσεις του προσώπου και τη συναισθηματική νοημοσύνη και θεωρείται αξιόπιστο και έγκυρο. [33-36]



Εικόνα 2: Ekman’s 7 basic emotions and corresponding facial expressions (Ανακτήθηκε 15/02/2023 από https://www.researchgate.net/figure/Ekmans-7-basic-emotions-and-corresponding-facial-expressions_fig1_329054559)



Εικόνα 3: Example of use FACS manual (Ανακτήθηκε 15/02/2023 από https://www.paulekman.com/product/facs-manual/)

Στις δεκαετίες του 1990 και του 2000, η έρευνα για τις συναισθηματικές εκφράσεις του προσώπου άρχισε να μετατοπίζεται για να επικεντρωθεί σε πιο λεπτές παραλλαγές στις εκφράσεις και στην αναγνώριση πιο περίπλοκων συναισθημάτων. Ερευνητές όπως ο Camras και ο Allison, ανέπτυξαν το Πρόγραμμα Affect, το οποίο είναι ένα εργαλείο για τον εντοπισμό λεπτών παραλλαγών στις εκφράσεις και την αναγνώριση πιο περίπλοκων συναισθημάτων. Αυτή η έρευνα επέκτεινε την κατανόηση της πολυπλοκότητας, αναγνωρίζοντας ότι δεν περιορίζονται σε βασικά συναισθήματα, αλλά συνδέονται επίσης με πολύπλοκα συναισθήματα, όπως μικτά συναισθήματα και μικρό-εκφράσεις. [37-38]

Τα τελευταία χρόνια, υπάρχει ένα αυξανόμενο ενδιαφέρον για την εφαρμογή τεχνικών υπολογιστικής όρασης και βαθιάς μάθησης στην αναγνώριση των εκφράσεων του προσώπου και του Affective Computing. Αυτές οι τεχνικές έχουν χρησιμοποιηθεί για την ανάλυση των εκφράσεων και των προτύπων ομιλίας για την ανίχνευση συναισθηματικών καταστάσεων και έχει βρεθεί ότι είναι αποτελεσματικές στην αναγνώριση συναισθημάτων σε πραγματικό χρόνο. [39]

Αυτή η βιβλιογραφική ανασκόπηση δείχνει την εξέλιξη της έρευνας στον τομέα των εκφράσεων του προσώπου και της αναγνώρισης της συναισθηματικής κατάστασης μέσα από την ιστορία και πώς οδήγησε στην τρέχουσα κατάσταση του πεδίου και τις προκλήσεις που πρέπει ακόμη να αντιμετωπιστούν.

2.2 Επισκόπηση των πεδίων της βαθιάς μάθησης και της υπολογιστικής όρασης

Η υπολογιστική όραση και η βαθιά μάθηση είναι δύο ταχέως εξελισσόμενα πεδία που έχουν πλούσια ιστορία πολλών δεκαετιών. Η πρώιμη έρευνα στην όραση υπολογιστών επικεντρώθηκε στην ανάπτυξη αλγορίθμων και συστημάτων που επιτρέπουν στους υπολογιστές να ερμηνεύουν και να κατανοούν οπτικά δεδομένα από τον κόσμο γύρω τους. Αυτό περιλάμβανε επεξεργασία εικόνας και αναγνώριση αντικειμένων. Στις αρχές, οι παραδοσιακές τεχνικές μηχανικής μάθησης όπως το SVM (Support Vector Machine), τα δέντρα αποφάσεων και το k-NN χρησιμοποιήθηκαν για την επίλυση προβλημάτων υπολογιστικής όρασης. Ωστόσο, αυτές οι μέθοδοι είχαν περιορισμούς όσον αφορά την επεκτασιμότητα και την απόδοση. [40-41]

Η έλευση της βαθιάς μάθησης στις αρχές της δεκαετίας του 2000 έφερε μια σημαντική αλλαγή στον τομέα της υπολογιστικής όρασης. Η βαθιά μάθηση, η οποία περιλαμβάνει την εκπαίδευση πολύ-επίπεδων τεχνητών νευρωνικών δικτύων (Multi-Layer Perceptron) για την εκτέλεση σύνθετων εργασιών, επιτρέπει την ανάπτυξη συστημάτων που μπορούν να μάθουν ιεραρχικές αναπαραστάσεις δεδομένων. Αυτό, με τη σειρά του, επιτρέπει την ανάπτυξη συστημάτων που μπορούν να εκτελούν εργασίες όπως η αναγνώριση εικόνας και ομιλίας με υψηλή ακρίβεια. Οι πιο δημοφιλείς αρχιτεκτονικές βαθιάς μάθησης που χρησιμοποιούνται στην υπολογιστική όραση είναι τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) και τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα (RNN).[3][7]

Στα τέλη της δεκαετίας του 2000 και στις αρχές της δεκαετίας του 2010, έγιναν μια σειρά από ανακαλύψεις στην υπολογιστική όραση και τη βαθιά μάθηση. Αυτές περιλαμβάνουν την ανάπτυξη αρχιτεκτονικών CNN όπως τα AlexNet, VGG και GoogLeNet που βελτίωσαν σημαντικά την απόδοση των εργασιών ταξινόμησης εικόνων (Image classification). Επιπλέον, οι ερευνητές πρότειναν τη χρήση της βαθιάς μάθησης για εργασίες ανίχνευσης αντικειμένων και σημασιολογικής τμηματοποίησης, οι οποίες παραδοσιακά επιλύονταν χρησιμοποιώντας παραδοσιακές τεχνικές μηχανικής μάθησης. [42-43]

Τα τελευταία χρόνια, οι εξελίξεις στην υπολογιστική όραση και τη βαθιά μάθηση οδήγησαν στην ανάπτυξη συστημάτων που μπορούν να εκτελέσουν ένα ευρύ φάσμα εργασιών, όπως η αναγνώριση εικόνας και ομιλίας, η ανίχνευση αντικειμένων και κατανόηση τους (προσδιορισμός αντικειμένου). Αυτές οι τεχνικές έχουν χρησιμοποιηθεί για την ανάπτυξη συστημάτων που μπορούν να αναγνωρίζουν και να ερμηνεύουν τις εκφράσεις του προσώπου. [44]

Ωστόσο, παρά αυτές τις εξελίξεις, υπάρχουν ακόμη προκλήσεις που πρέπει να αντιμετωπιστούν στον τομέα της υπολογιστικής όρασης και της βαθιάς μάθησης. Αυτά περιλαμβάνουν την έλλειψη μεγάλων, διαφορετικών συνόλων δεδομένων (datasets), τις δυσκολίες στη γενίκευση των μοντέλων βαθιάς μάθησης σε διαφορετικούς πληθυσμούς και την ανάγκη ανάπτυξης πιο ισχυρών και αποτελεσματικών αρχιτεκτονικών βαθιάς μάθησης.[45]

2.3 Επισκόπηση της χρήσης υπολογιστικής όρασης και βαθιάς μάθησης στην αναγνώριση συναισθηματικών καταστάσεων

Η αναγνώριση εκφράσεων προσώπου είναι ένα θέμα ενδιαφέροντος στον τομέα της υπολογιστικής όρασης και της τεχνητής νοημοσύνης για πολλά χρόνια. Η πρώιμη έρευνα στο πεδίο επικεντρώθηκε στη χρήση παραδοσιακών τεχνικών CV, όπως η εξαγωγή χαρακτηριστικών και η αντιστοίχιση προτύπων, για τον εντοπισμό και την ταξινόμηση των εκφράσεων του προσώπου. Ωστόσο, με την έλευση της βαθιάς μάθησης, υπήρξε μια στροφή προς τη χρήση νευρωνικών δικτύων για τη βελτίωση της ακρίβειας και της ευρωστίας των συστημάτων αναγνώρισης εκφράσεων προσώπου. [46]

Μία από τις πρώτες εφαρμογές της βαθιάς μάθησης στην αναγνώριση εκφράσεων προσώπου ήταν η χρήση συνελικτικών νευρωνικών δικτύων (CNN) για την ταξινόμηση των εκφράσεων του προσώπου σε εικόνες. Το 2014, μια μελέτη των Li και Chen χρησιμοποίησε ένα CNN για να ταξινομήσει έξι βασικά συναισθήματα (ευτυχία, λύπη, έκπληξη, θυμό, αηδία και φόβο) από το σύνολο δεδομένων FER2013, επιτυγχάνοντας ακρίβεια 72,8%. [47]

Τα τελευταία χρόνια, έχουν υπάρξει αρκετές μελέτες που έχουν διερευνήσει τη χρήση της βαθιάς μάθησης για την αναγνώριση της έκφρασης του προσώπου σε βίντεο. Το 2016, μια μελέτη των Barros et al. χρησιμοποίησε ένα τρισδιάστατο CNN για να ταξινομήσει τις εκφράσεις του προσώπου από ακολουθίες βίντεο, επιτυγχάνοντας ακρίβεια 78,5% στο σύνολο δεδομένων CK+. Ομοίως, το 2017, μελέτη των Li et al. χρησιμοποίησε ένα τρισδιάστατο CNN για να ταξινομήσει τις εκφράσεις του προσώπου από ακολουθίες βίντεο, επιτυγχάνοντας ακρίβεια 82,8% στο σύνολο δεδομένων AffectNet. [48-49]

Εκτός από τα CNN, υπήρξαν επίσης αρκετές μελέτες που έχουν διερευνήσει τη χρήση άλλων αρχιτεκτονικών βαθιάς μάθησης για την αναγνώριση της έκφρασης του προσώπου. Για παράδειγμα, το 2018, μια μελέτη των Li et al. χρησιμοποίησε ένα δίκτυο μακράς βραχείας μνήμης (LSTM) για να ταξινομήσει τις εκφράσεις του προσώπου από ακολουθίες βίντεο, επιτυγχάνοντας ακρίβεια 80,9% στο σύνολο δεδομένων AffectNet. [50]

Εκτός από τη χρήση της βαθιάς μάθησης στην αναγνώριση εκφράσεων προσώπου, έχουν γίνει επίσης αρκετές μελέτες που έχουν διερευνήσει τη χρήση άλλων τεχνικών, όπως η μάθηση μεταφοράς και η εκμάθηση συνόλου, για τη βελτίωση της ακρίβειας και της ευρωστίας των συστημάτων αναγνώρισης εκφράσεων προσώπου. [51-52]

2.4 Εφαρμογές της υπολογιστικής όρασης και βαθιάς μάθησης στην αναγνώριση συναισθηματικών καταστάσεων

Η υπολογιστική όραση και η βαθιά μάθηση έχουν τη δυνατότητα να εφαρμοστούν σε ένα ευρύ φάσμα εφαρμογών που σχετίζονται με τις εκφράσεις του προσώπου και την αναγνώριση συναισθηματικής νοημοσύνης, όπως:

- **Προγράμματα και εργαλεία εκπαίδευσης συναισθηματικής νοημοσύνης:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για την ανάπτυξη εικονικών εκπαιδευτικών προγραμμάτων που παρέχουν ανατροφοδότηση στις εκφράσεις του προσώπου των ατόμων. [53]
- **Διάγνωση και θεραπεία ψυχικής υγείας:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για τον εντοπισμό αλλαγών στις εκφράσεις του προσώπου που μπορεί να υποδηλώνουν την εμφάνιση διαταραχής ψυχικής υγείας, καθώς και για την παρακολούθηση της αποτελεσματικότητας της θεραπείας για διαταραχές ψυχικής υγείας. [54]
- **Ασφάλεια και επιτήρηση:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για τον εντοπισμό και την απόκριση σε πιθανές απειλές για την ασφάλεια, όπως άτομα που εμφανίζουν σημάδια θυμού ή επιθετικότητας. [60]
- **Αλληλεπίδραση ανθρώπου-υπολογιστή:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για τη βελτίωση της φυσικότητας και της αποτελεσματικότητας της επικοινωνίας μεταξύ ανθρώπων και υπολογιστών. [55]
- **Κοινωνική ρομποτική:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για να επιτρέψει στα ρομπότ να αναγνωρίζουν και να ανταποκρίνονται κατάλληλα στα ανθρώπινα συναισθήματα. [56]
- **Χώρος εργασίας, εκπαίδευση και υγειονομική περίθαλψη:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για τη μέτρηση της συναισθηματικής νοημοσύνης των ατόμων και για να έχει επιπτώσεις σε περιβάλλοντα όπως ο χώρος εργασίας, η εκπαίδευση και η υγειονομική περίθαλψη. [57]
- **Εφαρμογές με επίγνωση των συναισθημάτων:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για την ανάπτυξη εφαρμογών με επίγνωση των συναισθημάτων που μπορούν να προσαρμόσουν τη διεπαφή χρήστη, το περιεχόμενο ή τις υπηρεσίες με βάση τη συναισθηματική κατάσταση του χρήστη. [58]

- **Παιχνίδι και ψυχαγωγία:** Η τεχνολογία αναγνώρισης εκφράσεων προσώπου μπορεί να χρησιμοποιηθεί για την ανάπτυξη παιχνιδιών και εφαρμογών ψυχαγωγίας που μπορούν να ανταποκριθούν στα συναισθήματα των χρηστών. [59]

Αυτά είναι μόνο μερικά παραδείγματα από τις πολλές πιθανές εφαρμογές της υπολογιστικής όρασής και της βαθιάς μάθησης στις εκφράσεις του προσώπου και στην αναγνώριση συναισθηματικής νοημοσύνης. Οι εξελίξεις σε αυτούς τους τομείς ανοίγουν νέες δυνατότητες για τη βελτίωση των αλληλεπιδράσεων ανθρώπου-υπολογιστή και την ανάπτυξη περισσότερων μηχανών που μοιάζουν με τον άνθρωπο. Ωστόσο, είναι σημαντικό να ληφθούν υπόψη οι ηθικές και κοινωνικές επιπτώσεις αυτών των τεχνολογιών και να διασφαλιστεί ότι αναπτύσσονται και χρησιμοποιούνται με υπεύθυνο και ηθικό τρόπο

3. Μεθοδολογία

3.1 Χαρακτηριστικά συστήματος

Hardware

Specs Personal PC:

Processor 11th Gen Intel(R) Core(TM) i7-11370H @ 3.30GHz 3.00 GHz

Installed RAM 16,0 GB (15,7 GB usable)

System type 64-bit operating system, x64-based processor

Specs Server provided for training:

NVIDIA GTX Titan X (3072 CUDA cores)

Software

- Anaconda Navigator 2.3.0
- PyCharm Community Edition 2022.2.22
- Python 3.7

Γλώσσα προγραμματισμού

- Python

Βιβλιοθήκες (Libraries)

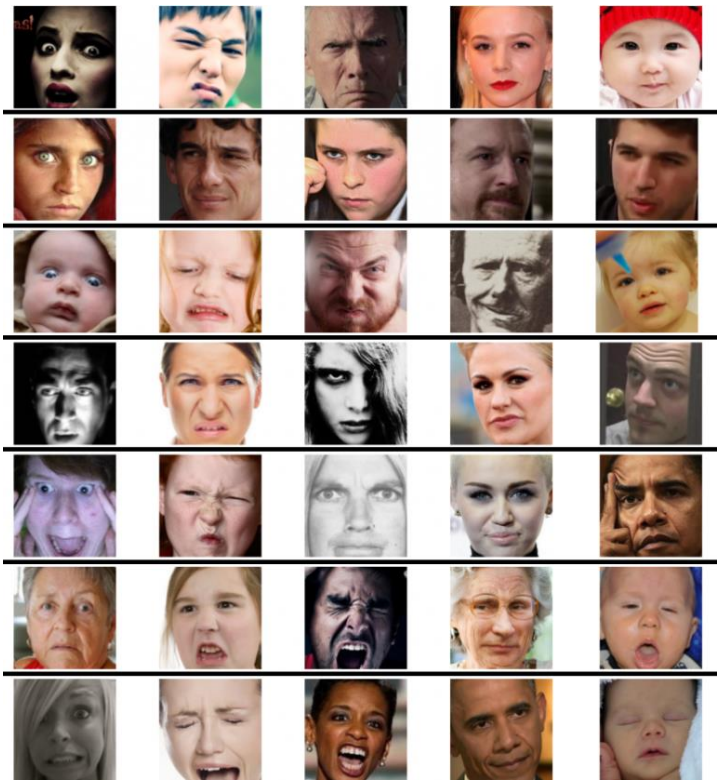
- pandas
- os
- itertools
- pathlib
- shutil
- numpy
- cv2
- tensorflow
- keras
- scipy
- matplotlib
- sys

3.2 Συλλογή δεδομένων

3.2.1 Διαθέσιμα σύνολα δεδομένων

Υπάρχουν πολλά διαθέσιμα σύνολα δεδομένων για την ανίχνευση συναισθημάτων, το καθένα με τα δικά του δυνατά και αδύνατα σημεία. Συγκεκριμένα:

AffectNet: Αυτό είναι ένα σύνολο δεδομένων εκφράσεων προσώπου μεγάλης κλίμακας που περιέχει πάνω από 1 εκατομμύριο εικόνες εκφράσεων προσώπου που συλλέχθηκαν από το διαδίκτυο. Οι εικόνες στο σύνολο δεδομένων επισημαίνονται με μία ή περισσότερες ετικέτες έκφρασης προσώπου, συμπεριλαμβανομένων των βασικών συναισθημάτων (χαρά, λύπη, θυμός, έκπληξη, φόβος και αηδία), καθώς και ουδέτερο (neutral). Το σύνολο δεδομένων περιλαμβάνει επίσης ορόσημα προσώπου, χαρακτηριστικά και μονάδες δράσης. [46]



Εικόνα 4: Δείγμα dataset Affectnet (Ανακτήθηκε 15/02/2023 από http://mohammadmahoor.com/affectnet/)

CK+: Το σύνολο δεδομένων CK+ είναι ένα σύνολο δεδομένων εκφράσεων προσώπου που περιέχει εικόνες ατόμων που εκφράζουν διαφορετικά συναισθήματα, όπως ευτυχία, λύπη, θυμό, έκπληξη και αηδία. Το σύνολο δεδομένων περιλαμβάνει, τόσο πόζες, όσο και αυθόρμητες εκφράσεις προσώπου, καθώς και περιλαμβάνει και ορόσημα προσώπου για κάθε εικόνα. [61]

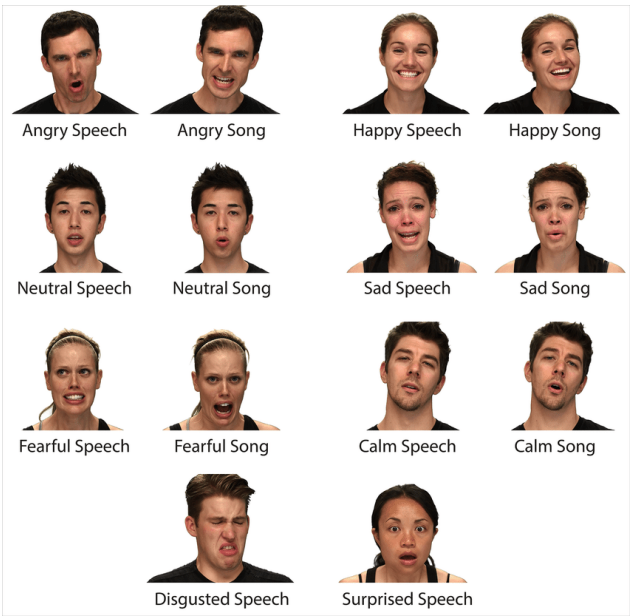
Εικόνα 5: Δείγμα dataset CK+ (Ανακτήθηκε 15/02/2023 από https://paperswithcode.com/dataset/ck)

EmoReact: Το σύνολο δεδομένων EmoReact περιέχει μεγάλο αριθμό βίντεο παιδιών ηλικίας 4-14 ετών που αντιδρούν σε διαφορετικά ερεθίσματα. Το σύνολο δεδομένων περιλαμβάνει ετικέτες για τα συναισθήματα που εκφράζονται στις εικόνες, όπως η ευτυχία, η λύπη, ο θυμός, η έκπληξη και η αηδία.[62]

Εικόνα 6: Δείγμα dataset EmoReact (Ανακτήθηκε 15/02/2023 από https://github.com/bnojavan/EmoReact)

FER2013: Αυτό είναι ένα σύνολο δεδομένων εκφράσεων προσώπου που περιέχει εικόνες ατόμων που εκφράζουν διαφορετικά συναισθήματα, όπως ευτυχία, λύπη, θυμό, έκπληξη, φόβο, αηδία και ουδέτερο. Το σύνολο δεδομένων περιλαμβάνει επίσης ορόσημα προσώπου για κάθε εικόνα. [3]

RAVDESS: Το σύνολο δεδομένων Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) είναι ένα οπτικοακουστικό σύνολο δεδομένων που περιέχει πάνω από 2.000 ηχητικά και βίντεο κλιπ ατόμων που μιλούν και τραγουδούν σε διαφορετικές συναισθηματικές καταστάσεις, όπως ουδέτερη, ευτυχία, λύπη, έκπληξη, θυμό, αηδία και φόβο. Το σύνολο δεδομένων περιλαμβάνει επίσης καταγραφές του λόγου και του τραγουδιού, επιτρέποντας τη μελέτη τόσο των ακουστικών όσο και των οπτικών πτυχών της αναγνώρισης συναισθημάτων. [63]



Εικόνα 7: Δείγμα dataset RAVDESS (Ανακτήθηκε 15/02/2023 από https://www.researchgate.net/figure/Examples-of-the-eight-RAVDESS-emotions-Still-frame-examples-of-the-eight-emotions_fig2_325187111)

Αυτά είναι μερικά μόνο παραδείγματα από τα πολλά διαθέσιμα σύνολα δεδομένων για την ανίχνευση συναισθημάτων. Κάθε σύνολο δεδομένων έχει τα δικά του μοναδικά χαρακτηριστικά και μπορεί να χρησιμοποιηθεί για διαφορετικούς ερευνητικούς σκοπούς. Είναι σημαντικό να σημειωθεί ότι η διαθεσιμότητα και η πρόσβαση σε αυτά τα σύνολα δεδομένων μπορεί να διαφέρει ανάλογα με το σύνολο δεδομένων, το ιστορικό και τις προθέσεις του ερευνητή.

3.2.2 AffectNet

Στα πλαίσια της εργασίας αυτής επιλέξαμε να χρησιμοποιήσουμε ως βασικό σύνολο δεδομένων το AffectNet γιατί το μεγαλύτερο σύνολο δεδομένων εκφράσεων προσώπου, το οποίο το καθιστά ιδανικό για την έρευνά για την αυτοματοποιημένη αναγνώριση συναισθηματικών εκφράσεων προσώπου. Συγκεκριμένα το AffectNet περιέχει περίπου 1 εκατομμύριο εικόνες προσώπου που συλλέχθηκαν από το Διαδίκτυο, από αναζήτηση σε τρεις μεγάλες μηχανές αναζήτησης, χρησιμοποιώντας 1250 λέξεις-κλειδιά που σχετίζονται με συναισθήματα σε έξι διαφορετικές γλώσσες. Περίπου οι μισές από τις ανακτημένες εικόνες (~420K) ‘έχουν χαρακτηριστεί χειροκίνητα και κατηγοριοποιηθεί (classified) σε επτά διακριτές κλάσεις εκφράσεων προσώπου, διαφορετικής πολικότητας (valence) και διέγερσης (arousal) - διάστατο μοντέλο. Οι υπόλοιπες εικόνες (~550K) ‘έχουν χαρακτηριστεί αυτόματα χρησιμοποιώντας το ResNext Neural Network το οποίο εκπαιδεύτηκε σε όλο το σετ εκπαίδευσης με μη αυτόματο χαρακτηρισμό δειγμάτων με μέση ακρίβεια 65%. [28]

Το AffectNet παρέχει:

• Εικόνες των προσώπων

• Θέση των προσώπων στις εικόνες

• Θέση των 68 ορόσημων προσώπου

• Έντεκα κατηγορίες (labels) συναισθημάτων και μη

• Δείκτες πολικότητας και διέγερσης των εκφράσεων του προσώπου

Κατηγορίες συναισθημάτων:

Έντεκα συναισθηματικές καταστάσεις παρέχονται για τις εικόνες και χαρακτηρίζονται ως εξής:

0: Ουδέτερο, 1: Ευτυχία, 2: Θλίψη, 3: Έκπληξη, 4: Φόβος, 5: Αηδία, 6: Θυμός, 7: Περιφρόνηση, 8: Κανένα, 9: Αβέβαιο, 10: Χωρίς Πρόσωπο

Ο αριθμός των εικόνων με μη αυτόματο χαρακτηρισμό στο σετ εκπαίδευσης και επικύρωσης φαίνεται στον παρακάτω πίνακα:

Neutral	75,374
Happy	134,915
Sad	25,959
Surprise	14,590
Fear	6,878
Disgust	4,303
Anger	25,382
Contempt	4,250
None	33,588
Uncertain	12,145
Non-Face	82,915
Total	420,299

Εικόνα 8: AffectNet Classes (Ανακτήθηκε 15/02/2023 από http://mohammadmahoor.com/affectnet/)

Το μέγεθος του συνόλου δεδομένων είναι περίπου 122 GB ενώ τα αρχεία έχουν συμπιεστεί σε μορφή RAR και περιέχουν τρεις λίστες αρχείων training.csv, validation.csv και automatically_annotated.csv- η λίστες training.csv και validation.csv αναφέρονται στις εικόνες στον φάκελο Manually_Annotated_compressed, ενώ η λίστα automatically_annotated.csv αναφέρεται σε εικόνες στο Automatically_annotated_compressed φάκελο.

Τα παρεχόμενα αρχεία CSV περιέχουν τα ακόλουθα χαρακτηριστικά:

- Διαδρομή αρχείου: υπό-φάκελος και όνομα αρχείου της εικόνας.
- Face_x: x Θέση του προσώπου στην εικόνα.
- Face_y: y Θέση του προσώπου στην εικόνα.
- Face_width: πλάτος του προσώπου που εντοπίστηκε στην εικόνα.
- Face_height: ύψος του προσώπου που εντοπίστηκε στην εικόνα.
- Facial_landmarks: Θέσεις (x και y) των 68 ανιχνευμένων ορόσημων προσώπου. Η ακολουθία x και y διαχωρίζονται με ερωτηματικό(;) και έχουν την ακόλουθη δομή:

  x1;y1;x2;y2;x3;y3 …. x67;y67;x68;y68

- • Έκφραση: ID έκφρασης του προσώπου
- (0: Ουδέτερο, 1: Χαρούμενο, 2: Λυπημένο, 3:Έκπληξη, 4: Φόβος, 5: Αηδία, 6: Θυμός, 7: Περιφρόνηση, 8: Κανένας, 9: Αβέβαιο, 10: Χωρίς πρόσωπο)
- • Πολικότητα (Valance): τιμή σθένους της έκφρασης στο διάστημα [-1,+1] (για κατηγορίες αβέβαιων και χωρίς πρόσωπο η τιμή είναι -2)
- • Διέγερση (Arrousal): τιμή διέγερσης της έκφρασης στο διάστημα [-1,+1] (για κατηγορίες αβέβαιων και χωρίς πρόσωπο η τιμή είναι -2)

Για τους σκοπούς της παρούσας διπλωματικής επιλέξαμε να δουλέψουμε μόνο με τις λίστες των αρχείων που αφορούν την αξιολόγηση/ διαχωρισμό των καταστάσεων από ειδικό πραγματογνώμονα και όχι από αυτοματοποιημένο σύστημα, καθώς η αξιολόγηση αυτή κρίθηκε πιο αξιόπιστη.

3.2.3 Υβριδικό dataset AffectFer

Παρά τις προηγούμενες αναφορές σχετικά με την υψηλή ποιότητα και την ισορροπημένη κατανομή των ετικετών συναισθημάτων στο AffectNet, υπάρχουν κάποιες αδυναμίες που επηρεάζουν την ακρίβεια του μοντέλου μας.[64-65] Αυτές περιλαμβάνουν:

- Περιορισμένη ποικιλία προσώπων: Το AffectNet διαθέτει εικόνες προσώπων από ένα συγκεκριμένο σύνολο εθνικοτήτων και ηλικιακών ομάδων. Αυτό μπορεί να οδηγήσει σε μια προκατάληψη (bias) του μοντέλου μας προς αυτές τις κατηγορίες.
- Ανακριβής ετικετοποίηση (labeling): Παρόλο που οι ετικέτες του AffectNet έχουν δημιουργηθεί από επαγγελματίες αναλυτές συναισθημάτων, υπάρχουν περιπτώσεις όπου η ετικετοποίηση δεν είναι ακριβής. Αυτό μπορεί να οδηγήσει σε σφάλματα κατά την εκπαίδευση του μοντέλου.
- Απουσία εικόνων σε κάποιες κατηγορίες συναισθημάτων: Υπάρχουν κατηγορίες συναισθημάτων στο AffectNet που δεν διαθέτουν αρκετές εικόνες, ενώ σε άλλες υπάρχει υπερβολική πληθώρα. Αυτό μπορεί να προκαλέσει ανισορροπία στην εκπαίδευση του μοντέλου και να επηρεάσει τα αποτελέσματα.

Για να αντιμετωπίσουμε αυτές τις αδυναμίες, αποφασίσαμε να συνδυάσουμε το σύνολο δεδομένων AffectNet με το Fer2013 και να δημιουργήσουμε ένα υβριδικό σύνολο δεδομένων το AffectFer.

Για να συνδυάσουμε τα δύο σύνολα δεδομένων, ακολουθήσαμε την προσέγγιση του transfer learning. Χρησιμοποιήσαμε ένα προ-εκπαιδευμένο μοντέλο στο AffectNet και το επαναχρησιμοποιήσαμε για να εκπαιδεύσουμε ένα νέο μοντέλο στο Fer2013. Έπειτα, ενώσαμε τα δύο μοντέλα και πραγματοποιήσαμε την εκπαίδευση στα δύο ενοποιημένα datasets. (80)

Το αποτέλεσμα της συνδυαστικής χρήσης των δύο συνόλων δεδομένων ήταν μια βελτίωση της ακρίβειας στην αναγνώριση των συναισθημάτων σε σχέση με τη χρήση μόνο του συνόλου δεδομένων AffectNet. Παρόλο που αυτό το σύνολο δεδομένων έχει ορισμένες αδυναμίες, η συνδυαστική χρήση του με ένα άλλο σύνολο δεδομένων μπορεί να βελτιώσει τα αποτελέσματα του μοντέλου σε σχέση με τη χρήση ενός συνόλου δεδομένων μόνο.

3.3 Προ-επεξεργασία δεδομένων

3.3.1 Δημιουργία δομής dataset

Βιβλιογραφικά κατά την προετοιμασία ενός συνόλου δεδομένων εικόνας για την τροφοδότηση/ εκπαίδευση μοντέλου βαθιάς μάθηση ενδείκνυται ο διαχωρισμός των εικόνων σε φακέλους με βάση τα χαρακτηριστικά (στην περίπτωση μας τις συναισθηματικές καταστάσεις). Στην περίπτωση του AffectNet όπως περιγράψαμε παραπάνω οι φάκελοι είναι οργανωμένοι με κωδικούς, οι οποίοι είναι καταχωρημένοι σε λίστα αρχείου csv. Για να γίνει η τροφοδότηση των δεδομένων θα πρέπει να γίνει η αντιστοίχιση της εικόνας με τον χαρακτηρισμό που έχει στην λίστα (ή με τα άλλα διαθέσιμα χαρακτηριστικά). Επιχειρήσαμε αρχικά με την λογική της επανάληψης να χτίσουμε το dataset αλλά λόγω του μεγέθους και της μικρής χωρητικότητας RAM του προσωπικού υπολογιστή το πρόγραμμα διακοπτόταν. Γι' τον λόγο αυτό ετοιμάσαμε προγραμματιστικά επαναληπτική διαδικασία κατά την οποία με βάση την λίστα csv δημιουργεί δομή φακέλων και υπό-φακέλων για training & validation set με βάση τις κλάσεις διαχωρισμού (συναισθηματικές καταστάσεις).[66-68](δες παράρτημα 1)

Για το Fer2013 δεν χρειάστηκε κάποια αλλαγή καθώς οι εικόνες ήταν ήδη διαχωρισμένες ανά κλάση/ φάκελο.

Για τη συνδυαστική χρήση των συνόλων δεδομένων, πραγματοποιήσαμε τα εξής βήματα:

- Κατέβασμα των δύο συνόλων δεδομένων και αποσυμπίεση των αρχείων.
- Προετοιμασία των δεδομένων του Fer2013: Συγκεκριμένα, χρησιμοποιήσαμε τα αρχεία CSV που διαθέτει το Fer2013 και προσαρμόσαμε την κωδικοποίηση των ετικετών για να ταιριάζουν με αυτές του AffectNet.
- Δημιουργία ενός νέου συνόλου δεδομένων: Συνδυάσαμε τα δύο σύνολα δεδομένων μ ε το νέο σύνολο δεδομένων AffectFer να περιλαμβάνει 340.55 χιλιάδες εικόνες και 11 κατηγορίες συναισθημάτων.
- Επεξεργασία των δεδομένων: Πραγματοποιήσαμε προ-επεξεργασία των εικόνων για να αφαιρέσουμε τον θόρυβο και να κάνουμε την εικόνα συμβατή με το μοντέλο.

3.3.2 Επιλογή κλάσεων

Όπως αναφέραμε και παραπάνω το AffectNet περιέχει 11 διαφορετικές κλάσεις

0: Ουδέτερο, 1: Ευτυχία, 2: Θλίψη, 3: Έκπληξη, 4: Φόβος, 5: Αηδία, 6: Θυμός, 7: Περιφρόνηση, 8: Κανένα, 9: Αβέβαιο, 10: Χωρίς Πρόσωπο.

Μελετώντας την βιβλιογραφία αλλά κυρίως με κριτήριο την εφαρμογή του τελικού μοντέλου σε πραγματικές καταστάσεις επιλέξαμε να διαχειριστούμε αρχικά μόνο 7 κατηγορίες (συναισθηματικές καταστάσεις), καθώς αυτές έχουν χαρακτηριστεί ως βασικές. Σημαντικό κριτήριο για την επιλογή είναι και ο αριθμός των διαθέσιμων εικόνων για την εκπαίδευση καθώς το training dataset είναι αρκετά μη εξισορροπημένο (unbalanced dataset) και ειδικά οι καταστάσεις που επιλέξαμε να αφήσουμε εκτός έχουν πολύ λιγότερα αντιπροσωπευτικά δείγματα σε σχέση με τα υπόλοιπα.

0: Ουδέτερο, 1: Ευτυχία, 2: Θλίψη, 3: Έκπληξη, 4: Φόβος, 5: Αηδία, 6: Θυμός,

Αντίστοιχα, δουλέψαμε και στο νέο μας dataset AffectFer, ενώ όπως θα παρουσιάσουμε παρακάτω μειώσαμε τις κλάσεις και σε πολύ λιγότερες (5 κλάσεις) προσπαθώντας να βελτιώσουμε την απόδοση σε πραγματικές καταστάσεις.

3.4 Επιλογή και εκπαίδευση μοντέλου

Για την εκπαίδευση του μοντέλου, χρησιμοποιήσαμε τη βιβλιοθήκη Keras και επιλέξαμε τον αλγόριθμο βελτιστοποίησης Adam, καθώς αποδίδει καλά σε προβλήματα βαθιάς μάθησης. Επίσης, επιλέξαμε τον αλγόριθμο κανονικοποίησης Dropout, που βοηθά στην αποφυγή της υπερ-εκπαίδευσης.

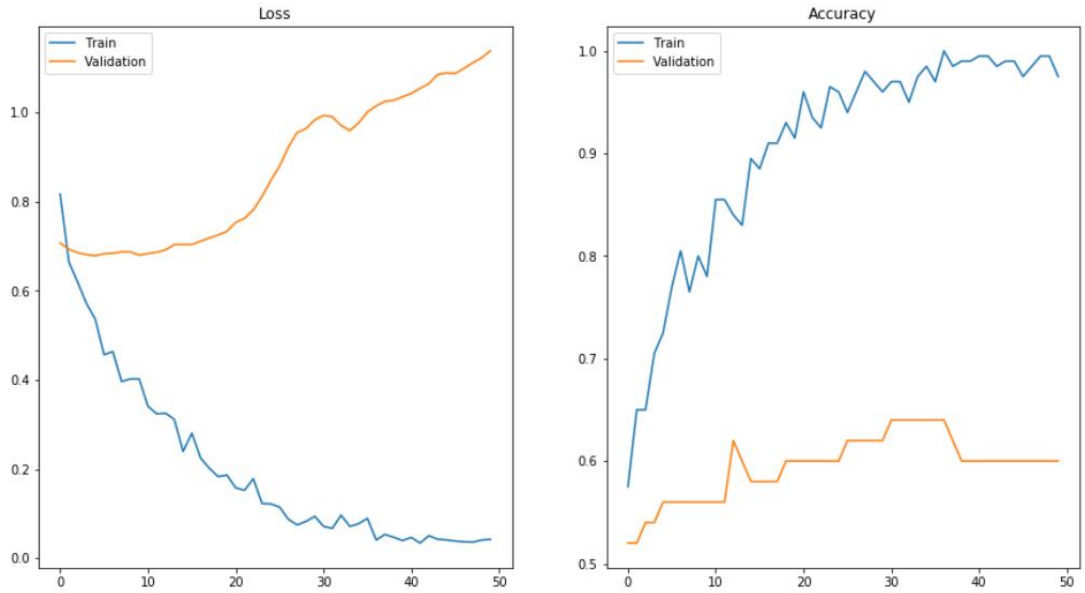
Αφού καταλήξαμε σε μια καλή αρχιτεκτονική μοντέλου, προχωρήσαμε στην εκπαίδευση του, στο συνδυασμό δηλαδή των δύο datasets. Στην πρώτη φάση της εκπαίδευσης χρησιμοποιήσαμε το AffectNet ως είσοδο και εκπαιδεύσαμε το μοντέλο μας σε αυτό. Στη συνέχεια, ανατρέξαμε στο Fer2013 και επιπλέον εκπαιδεύσαμε το μοντέλο σε αυτό, χρησιμοποιώντας τις ίδιες υπερ-παραμέτρους και τα ίδια βάρη, που είχαν εκπαιδευτεί στο AffectNet.

Τέλος, επαναλάβαμε την εκπαίδευση του μοντέλου σε όλα τα δεδομένα και χρησιμοποιήσαμε την τεχνική early stopping, για να αποφύγουμε την υπερ-εκπαίδευση του μοντέλου και να βελτιστοποιήσουμε την απόδοσή του.

Κατά τη διάρκεια της εκπαίδευσης του μοντέλου, χρησιμοποιούσαμε fine-tuning για τη βελτιστοποίηση μοντέλων μετά την προ-εκπαίδευσή τους σε μεγάλα datasets. Συγκεκριμένα, επιλέξαμε ένα προ-εκπαιδευμένο μοντέλο (π.χ. το VGG16), αποθηκεύσαμε τα βάρη του και στη συνέχεια το εκπαιδεύσαμε στο δικό μας dataset, διατηρώντας τα προ-αποθηκευμένα βάρη από το προ-εκπαιδευμένο μοντέλο.

Επιπλέον, είναι σημαντικό να κατανοήσουμε τα φαινόμενα της υπερ-εκπαίδευσης (overfitting) και της υπο-εκπαίδευσης (underfitting), τα οποία μπορούν να επηρεάσουν σοβαρά την απόδοση του μοντέλου.

Ο όρος υπερ-εκπαίδευση αναφέρεται στην κατάσταση όπου το μοντέλο μας έχει εκπαιδευτεί στα δεδομένα πολύ καλά, και "απομνημονεύει" τα δεδομένα αντί να μάθει τους γενικούς κανόνες και τα μοτίβα που αφορούν στην αναγνώριση συναισθημάτων από εικόνες προσώπων. Στην περίπτωση αυτή, το μοντέλο μας δεν θα είναι στο σωστό επίπεδο για να γενικεύσει σε νέα δεδομένα και θα έχει κακή απόδοση. (Εικόνα 9)



Εικόνα 9: Overfitting example (Ανακτήθηκε 15/02/2023 από https://www.v7labs.com/blog/overfitting-vs-underfitting)

Από την άλλη πλευρά, η υπο-εκπαίδευση (underfitting) είναι μια κατάσταση όπου το μοντέλο μας δεν μπορεί να μάθει αρκετά από τα δεδομένα και δεν μπορεί να εκπαιδευτεί καλά για να προβλέψει σωστά την κλάση συναισθήματος. Η υπο-εκπαίδευση συνήθως σημαίνει ότι το μοντέλο μας δεν έχει αρκετά κρυφά επίπεδα (hidden layers) και δεν έχει εκπαιδευτεί αρκετά σε αυτά τα επίπεδα.

Για να αντιμετωπίσουμε αυτά τα προβλήματα, χρησιμοποιούμε κάποιες τεχνικές, όπως η ρύθμιση των υπερ-παραμέτρων (hyperparameter tuning) και την χρήση του μηχανισμού early stopping. Επιπλέον, χρησιμοποιήσαμε την τεχνική του dropout, η οποία βοηθά στην αποφυγή της υπερ-εκπαίδευσης με τυχαία απενεργοποίηση νευρώνων κατά τη διάρκεια της εκπαίδευσης.

Όσον αφορά την εκπαίδευση του μοντέλου, χρησιμοποιήσαμε τον αλγόριθμο βελτιστοποίησης Adam και την συνάρτηση απώλειας “categorical cross-entropy”, και εκπαιδεύσαμε το μοντέλο σε 100 επαναλήψεις (epochs) με μέγεθος πακέτου 64. Χρησιμοποιήσαμε επίσης το μηχανισμό του “early stopping”, στον οποίο παρακολουθούσαμε την απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης και σταματούσαμε την εκπαίδευση αυτόματα εάν η απόδοση δεν βελτιωνόταν για 10 συνεχόμενες εποχές [69][70][3](παράρτημα 7.2)

Τέλος, εκτιμήσαμε την απόδοση του μοντέλου μας με χρήση διαφόρων μετρικών αξιολόγησης, όπως η ακρίβεια, και η ανάκληση.

Σε αυτό το σημείο πρέπει να αναφερθεί η σημαντικότητα της υποδομής (hardware) για την ολοκλήρωση όλων των παραπάνω ενεργειών. Ενώ η αρχική έρευνα και προετοιμασία του κώδικα έγινε σε προσωπικό υπολογιστή, δεν ήταν εφικτό λόγω των μεγάλων συνόλων δεδομένων να ολοκληρωθεί η διαδικασία εκπαίδευσης, καθώς το σύστημα παρουσίαζε πρόβλημα και διακοπτόταν αναπάντεχα με αποτέλεσμα να πρέπει να ξεκινά η διαδικασία από την αρχή. Αξιοσημείωτο είναι ότι σε αρκετές περιπτώσεις η εκπαίδευση κρατούσε σχεδόν 4-5 ημέρες, χωρίς να ολοκληρωθεί σωστά. Για τον λόγο αυτό ζητήθηκε άδεια να χρησιμοποιηθεί server του Πανεπιστημίου όπου υπήρχε η δυνατότητα χρήσης GPU NVIDIA GTX Titan X (3072 CUDA cores).

Στο πλαίσιο αυτά έγιναν τα εξής:

- Μεταφορά όλων των δεδομένων στον server
- Παραμετροποίηση του κώδικα εκπαίδευσης ώστε να ελέγχει την ύπαρξη GPU και να την χρησιμοποιεί για την διαδικασία εκπαίδευσης
- Εγκατάσταση και παραμετροποίηση Jupyter server ώστε η διαχείριση του κώδικα/ Notebook να είναι πιο ευέλικτη στα πλαίσια των δοκιμών και αλλαγών (fine tuning/early stopping) όπως αναφέρθηκε παραπάνω.

Έχοντας ολοκληρώσει όλα τα παραπάνω, ολοκληρώθηκε η εκπαίδευση των μοντέλων, αν και πάλι η χρονική διάρκεια ολοκλήρωσης κράτησε 2-3 ημέρες.

3.5 Μετρήσεις αξιολόγησης

Η αξιολόγηση του μοντέλου αποτελεί κρίσιμο στάδιο της εκπαίδευσης καθώς εκτιμά την ακρίβεια και την ικανότητα του μοντέλου να γενικεύει σε νέα δεδομένα. Η εκτίμηση αυτή χρησιμοποιεί διάφορες μετρικές αξιολόγησης της απόδοσης του μοντέλου στο σύνολο ελέγχου (validation set), όπως είναι η ακρίβεια (accuracy), η οποία μετρά το ποσοστό των σωστών προβλέψεων του μοντέλου. Επίσης, συνήθως υπολογίζονται το training και το validation loss κατά τη διάρκεια της εκπαίδευσης του μοντέλου. Το loss μετρά την απόκλιση μεταξύ των πραγματικών και προβλεπόμενων τιμών και αποτελεί έναν δείκτη της απόδοσης του μοντέλου. Κατά τη διάρκεια της εκπαίδευσης του μοντέλου, επιδιώκουμε να μειώσουμε το loss στο ελάχιστο δυνατό επίπεδο.[3][7]

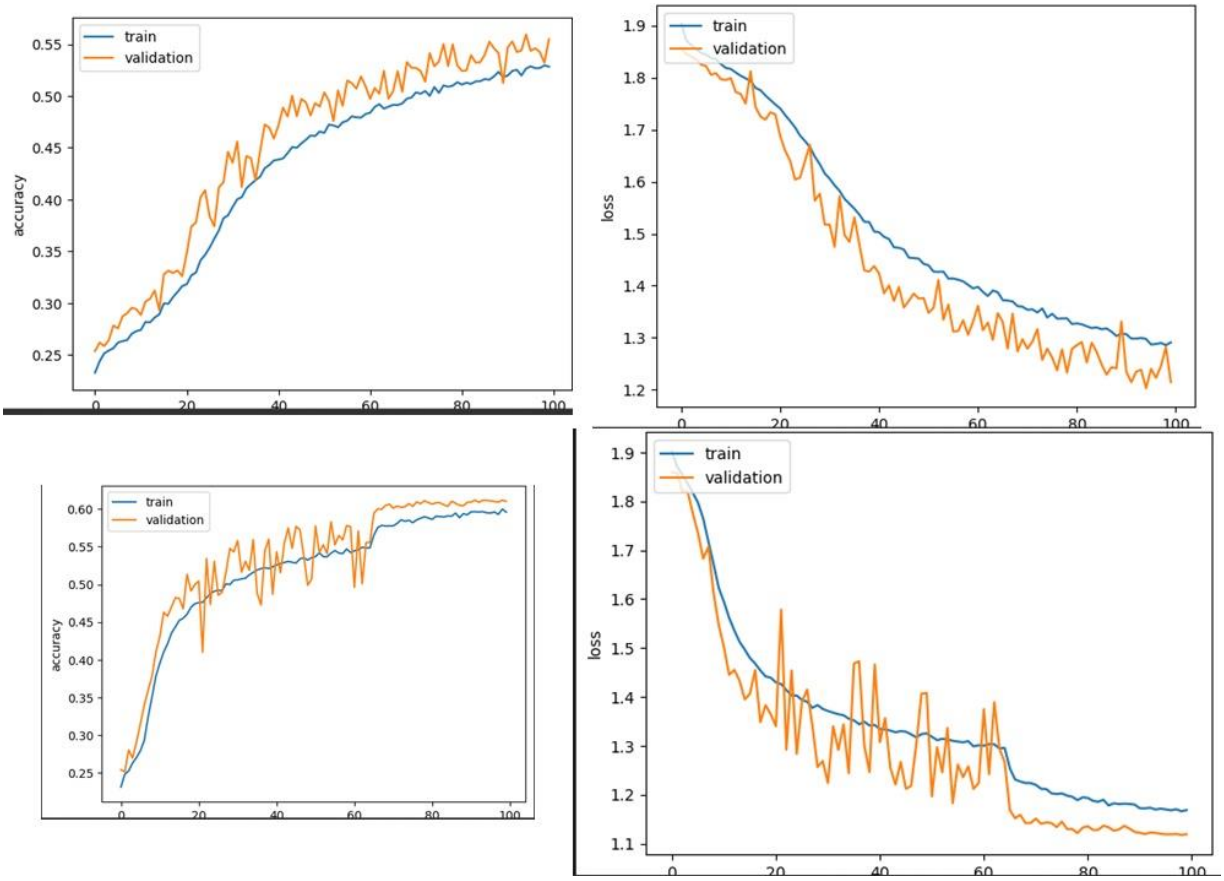
Το training loss αναφέρεται στο loss που παράχθηκε κατά τη διάρκεια της εκπαίδευσης του μοντέλου στα δεδομένα εκπαίδευσης ενώ το validation loss αναφέρεται στο loss που παράχθηκε κατά τη διάρκεια της αξιολόγησης του μοντέλου στο σύνολο ελέγχου. Στην περίπτωση που το training loss είναι πολύ χαμηλό ενώ το validation loss είναι σημαντικά υψηλότερο, τότε έχουμε ένδειξη υπερεκπαίδευσης (overfitting) του μοντέλου, δηλαδή το μοντέλο έχει μάθει να προβλέπει πολύ καλά τα δεδομένα εκπαίδευσης, αλλά δεν γενικεύει σε νέα δεδομένα. Αντίθετα, εάν και το training και το validation loss είναι υψηλά, τότε μπορεί να υπάρχει υποεκπαίδευση (underfitting), δηλαδή το μοντέλο δεν έχει μάθει αρκετά από τα δεδομένα εκπαίδευσης και δεν μπορεί να προβλέψει καλά ούτε τα δεδομένα εκπαίδευσης ούτε τα δεδομένα ελέγχου. [3][7]

Κατά την εκπαίδευση του μοντέλου, συνήθως χρησιμοποιούνται τεχνικές fine-tuning, όπως η αλλαγή των παραμέτρων του μοντέλου, η προσθήκη επιπλέον επιπέδων στο νευρωνικό δίκτυο, ή η αλλαγή του μεγέθους του δείγματος κατά την εκπαίδευση. Επιπλέον, μπορεί να χρησιμοποιηθεί η τεχνική του early stopping, όπου η εκπαίδεση του μοντέλου διακόπτεται αυτόματα εάν δεν παρατηρηθεί καμία βελτίωση στα δεδομένα ελέγχου για έναν προκαθορισμένο αριθμό εποχών.[71][72][3]

Στη συνέχεια, μετά την εκπαίδευση του μοντέλου, αξιολογείται μετρώντας την απόδοση του στα δεδομένα ελέγχου, χρησιμοποιώντας διάφορες μετρικές αξιολόγησης, όπως η ακρίβεια (accuracy), η ανάκληση (recall), η ακρίβεια (precision), η F1-βαθμίδα και η καμπύλη ROC.

Συνήθως οι μετρικές αξιολόγησης παρουσιάζονται σε γραφική μορφή, με τη μετρική στον άξονα του y και τον αριθμό των εποχών στον άξονα του x. Επιπλέον, είναι σημαντικό να επισημανθεί ότι η καμπύλη του training loss και του validation loss θα πρέπει να προσεγγίζουν τη μία την άλλη, ενώ το validation loss δεν πρέπει να αυξάνεται καθώς αυξάνεται ο αριθμός των εποχών, καθώς αυτό υποδηλώνει ότι το μοντέλο δεν μπορεί να γενικεύσει σε νέα δεδομένα.

Στο σχήμα που ακολουθεί βλέπουμε τις αντίστοιχες γραφικές για τα μοντέλα μας σε δεδομένα 7 συναισθημάτων):



Εικόνα 10: AffectNet training/validation accuracy & Loss (Top)

AffectFer training/validation accuracy & Loss (Bottom)

Όπως είπαμε και παραπάνω δεν έχουμε καταφέρει το βέλτιστο αποτέλεσμα αλλά οι αποδόσεις των μοντέλων είναι αρκετά ανταγωνιστικές (validation accuracy 55% & 60%) σε σχέση με την βιβλιογραφία (https://paperswithcode.com/sota/facial-expression-recognition-on-affectnet), ενώ η χρήση του υβριδικού μοντέλου βελτίωσε την απόδοσης κατά 5-8%.

4. Εφαρμογή του μοντέλου με χρήση της κάμερας του προσωπικού υπολογιστή

4.1 Εφαρμογή

Σε αυτό το κεφάλαιο, περιγράφουμε τη χρήση του μοντέλου που αναπτύχθηκε για την αναγνώριση συναισθημάτων από εικόνες σε πραγματικό χρόνο χρησιμοποιώντας την κάμερα του υπολογιστή (βλ. 8.3). Χρησιμοποιήθηκε η βιβλιοθήκη OpenCV για την εξαγωγή πλαισίων από την κάμερα, καθώς και τη χρήση του μοντέλου που αναπτύχθηκε προηγουμένως για την αναγνώριση των συναισθημάτων από τις εικόνες.

Αρχικά, διαβάζουμε το μοντέλο που αναπτύχθηκε στο προηγούμενο κεφάλαιο, χρησιμοποιώντας τη βιβλιοθήκη TensorFlow. Στη συνέχεια, η βιβλιοθήκη OpenCV χρησιμοποιείται για τη λήψη εικόνων από την κάμερα. Το παράθυρο της κάμερας έχει οριστεί σε μέγεθος 1024x768.

Στη συνέχεια, ορίζονται οι παράμετροι για τον ανιχνευτή προσώπου. Ο ανιχνευτής προσώπου χρησιμοποιεί το Haar Cascade classifier, ο οποίος είναι ένας προκαθορισμένος ταξινομητής που χρησιμοποιείται για την εντοπισμό προσώπων σε μια εικόνα. Ορίζουμε επίσης μια λίστα με τις ετικέτες για κάθε συναίσθημα, όπως είχαμε ορίσει. (81)

Στο κύριο μέρος του κώδικα, διαβάζουμε κάθε καρέ (frame) από την κάμερα και το μετατρέπουμε σε κλίμακα του γκρι, ώστε να μπορεί να αναγνωριστεί το πρόσωπο. Στη συνέχεια, ο ανιχνευτής προσώπου χρησιμοποιείται για να βρεθεί το πρόσωπο στην εικόνα.

Αφού το πρόσωπο βρεθεί, εφαρμόζουμε το μοντέλο που αναπτύχθηκε προηγουμένως για την αναγνώριση των συναισθημάτων στο κλίμακα του γκρι του προσώπου. Τα αποτελέσματα προβάλλονται στην εικόνα με τη βοήθεια της βιβλιοθήκης OpenCV, που επίσης χρησιμοποιείται για την επισήμανση του προσώπου στην εικόνα.

Στη συνέχεια, προσθέτουμε τα αποτελέσματα στο παράθυρο της κάμερας και προβάλλουμε την εικόνα στην οθόνη.

4.2 Αποτελέσματα

Στο πείραμα, παρατηρήθηκε ότι το μοντέλο αναγνωρίζει αρκετά καλά τα βασικά συναισθήματα χαράς, λύπης και φόβου. Ωστόσο, δεν είναι τόσο αποτελεσματικό στην αναγνώριση άλλων συναισθημάτων όπως έκπληξη και θυμό. Πιθανές αιτίες για αυτό το αποτέλεσμα μπορεί να είναι η ανεπάρκεια των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου ή η δυσκολία στην αναγνώριση και διάκριση ορισμένων συναισθημάτων από εικόνες. Η ακρίβεια του μοντέλου εξαρτάται σε μεγάλο βαθμό από την ποιότητα των εικόνων που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Επιπλέον, η χρήση τεχνικών ρύθμισης όπως η αύξηση δεδομένων (data augmentation) και η ρύθμιση των υπερ-παραμέτρων (hyperparameters) μπορεί να βελτιώσει την απόδοση του μοντέλου.

Αξίζει να σημειωθεί ότι η αναγνώριση συναισθημάτων από εικόνες είναι μια επίπονη εργασία ακόμα και για τον ανθρώπινο εγκέφαλο, καθώς τα συναισθήματα είναι συχνά πολύπλοκα και αντικειμενικά δύσκολα να περιγραφούν με λόγια. Συνεπώς, ακόμα και αν το μοντέλο δεν αναγνωρίζει ακριβώς το συναίσθημα που εκφράζεται σε μια εικόνα, μπορεί να παρέχει χρήσιμες πληροφορίες για τη γενική αίσθηση που προκαλεί η εικόνα στον θεατή (opinion mining).

Συνολικά, παρόλο που το μοντέλο που αναπτύχθηκε στο πείραμα δεν είναι απόλυτα ακριβές στην αναγνώριση συναισθημάτων από εικόνες, μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς, όπως η ανάλυση της αντίδρασης των ανθρώπων σε διάφορες εικόνες ή η βελτίωση της αντίληψης των ανθρώπων για τα συναισθήματα που εκφράζονται σε άλλα κατανοητά μέσα, όπως τις κοινωνικές δικτυώσεις ή τα παιχνίδια. Επιπλέον, η ανάλυση των αποτελεσμάτων του πειράματος μπορεί να χρησιμοποιηθεί για τη βελτίωση του μοντέλου.

5. Εφαρμογή του μοντέλου με χρήση της κάμερας ρομπότ και δημιουργία σεναρίου

Στο παρόν κεφάλαιο παρουσιάζουμε την εφαρμογή του μοντέλου που αναπτύξαμε σε ένα ρομπότ. Αρχικά είχαμε επιλέξει να χρησιμοποιήσουμε το ανθρωποειδές ρομπότ NAO το οποίο υπήρχε διαθέσιμο στο εργαστήριο μας.

Το ρομπότ NAO είναι ένας ανθρωποειδής ρομπότ που έχει σχεδιαστεί από την εταιρεία SoftBank Robotics. Με ύψος 58 εκατοστά και βάρος 5,4 κιλά, ο NAO διαθέτει κινητικότητα στα χέρια και τα πόδια, και μπορεί να κινηθεί σε διάφορους τρόπους χρησιμοποιώντας τους αισθητήρες που διαθέτει. Ο NAO είναι εξοπλισμένος με κάμερες και μικρόφωνα, που του επιτρέπουν να αναγνωρίζει φωνητικές εντολές και να αλληλοεπιδρά με τους χρήστες σε πραγματικό χρόνο. Επιπλέον, διαθέτει αισθητήρες αφής στο κεφάλι, τα χέρια και τα πόδια, και μπορεί να αναγνωρίσει και να αποκρίνεται σε ανθρώπινες κινήσεις και αφή. Χρησιμοποιείται σε πολλούς τομείς, όπως η εκπαίδευση, η έρευνα και η ψυχαγωγία, και μπορεί να επεκταθεί με διάφορα πρόσθετα λογισμικά και αξεσουάρ για να προσαρμοστεί σε διαφορετικές ανάγκες και εφαρμογές. [82]

Δυστυχώς αντιμετωπίσαμε αρκετά προβλήματα σε διάφορα σημεία της υλοποίησης με αποτέλεσμα να το απορρίψουμε για την εφαρμογή μας. Το σημαντικότερο ήταν ότι βασικός κώδικας για την υλοποίηση του αλγορίθμου είχε γραφτεί σε Python 3 που δεν υποστηρίζεται από το λογισμικό του NAO ώστε να ενσωματωθεί τοπικά στο ρομπότ. Εξαιτίας αυτού του περιορισμού δεν μπορούσαμε να εισάγουμε το μοντέλο στο σύστημα του, και προχωρήσαμε σε μια εναλλακτική. Συγκεκριμένα, προσπαθήσαμε να χρησιμοποιήσουμε την κάμερα του ρομπότ ως είσοδο για το πρόγραμμα αναγνώρισης συναισθηματικής κατάστασης το οποίο ήταν ενεργό στον προσωπικό μας υπολογιστή. Για τον σκοπό αυτό χρησιμοποιήσαμε το (Framework) πρωτόκολλο GStreamer ώστε σε πραγματικό χρόνο (live streaming), να χρησιμοποιήσουμε την καταγραφή από το ρομπότ.

Το GStreamer είναι ένα λογισμικό ανοιχτού κώδικα που χρησιμοποιείται για τη δημιουργία και την επεξεργασία πολυμέσων. Περιλαμβάνει βιβλιοθήκες, εργαλεία και προγράμματα εφαρμογής που επιτρέπουν στους χρήστες να δημιουργήσουν συστήματα πολυμέσων για διάφορες εφαρμογές, όπως η επεξεργασία βίντεο και ήχου, η αναπαραγωγή ήχου και η μετάδοση ροής μέσω δικτύου. Οι βασικές λειτουργίες του GStreamer είναι η ενσωμάτωση, η αναπαραγωγή, η μετατροπή, η επεξεργασία και η μετάδοση δεδομένων πολυμέσων. Μπορεί να χρησιμοποιηθεί σε διάφορα περιβάλλοντα, όπως σε λειτουργικά συστήματα GNU/Linux, Windows και Mac OS X. [85]

Η ροή live βίντεο, επιτεύχθει, αλλά, το αποτέλεσμα ήταν μη αξιοποιήσιμο καθώς λόγω της ασύρματης μετάδοσης των πακέτων πληροφορίας η μετάδοση ήταν προβληματική με πολύ κακή ποιότητα καρέ FPS, που καθιστούσε την αξιολόγηση της απόδοσης του μοντέλου εξαιρετικά δύσκολη, και άρα την οποιαδήποτε εφαρμογή σε πραγματικές καταστάσεις.

Για τον λόγο αυτό επιλέξαμε να χρησιμοποιήσουμε ένα άλλο ρομπότ το οποίο είχε ενσωματωμένο ήδη στο λειτουργικό του σύστημα το λογισμικό ROS.

Το ROS (Robot Operating System) είναι ένα λογισμικό ανοιχτού κώδικα που αναπτύχθηκε αρχικά για την ερευνητική κοινότητα ρομποτικής και αργότερα έγινε δημοφιλές και στη βιομηχανία. Πρόκειται για ένα σύστημα λογισμικού που παρέχει εργαλεία για το σχεδιασμό, την υλοποίηση και τον έλεγχο ρομπότ. Οι βασικές αρχές του ROS είναι η επαναχρησιμοποίηση, η επεκτασιμότητα και η αξιοπιστία. Οι χρήστες μπορούν να αναπτύξουν και να δοκιμάσουν λογισμικό σε προσομοιωτές, να δημιουργήσουν διαφορετικά κομμάτια λογισμικού και να τα συνδέσουν μεταξύ τους για να δημιουργήσουν μια πλήρη λύση ρομποτικής.

Κάποια από τα πλεονεκτήματα του ROS είναι:

- Επαναχρησιμοποίηση των κομματιών λογισμικού για τη δημιουργία διαφορετικών ρομποτικών λύσεων.
- Ευελιξία και επεκτασιμότητα για να προσαρμόζεται σε διάφορες εφαρμογές και ρομποτικές αρχιτεκτονικές.
- Πληθώρα εργαλείων και βιβλιοθηκών που επιτρέπουν στους χρήστες να δημιουργήσουν αποδοτικά και αξιόπιστα ρομποτικά συστήματα.
- Ελευθερία (open source) στην κοινότητα και στις συνεισφορές, καθιστώντας το ROS μια παγκόσμια κοινότητα ανοιχτού κώδικα για την ρομποτική.

Συνολικά, το ROS προσφέρει μια πλατφόρμα λογισμικού που επιτρέπει στους χρήστες να αναπτύξουν αποδοτικά και αξιόπιστα ρομποτικά συστήματα, επιτρέποντας την ανταλλαγή ιδεών και τη συνεργασία μεταξύ επιστημόνων και μηχανικών σε όλο τον κόσμο. [83-84]

Στα πλαίσια αυτά, αναπτύξαμε ένα πακέτο στο ROS το οποίο χρησιμοποιεί το μοντέλο Deep Learning για την αναγνώριση συναισθημάτων, και στο οποίο προσθέσαμε έναν έτοιμο αλγόριθμο αναγνώρισης προσώπου (Face Recognition algorithm) και ένα μοντέλο επεξεργασίας φυσικής γλώσσας (NLP). Με την χρήση των παραπάνω και με την χρήση πακέτων μετατροπής κειμένου σε ομιλία (TTS) και μετατροπής ομιλίας σε κείμενο (STT) αναπτύχθηκε το παρακάτω σενάριο:

Ο αλγόριθμος αναγνώρισης προσώπου χρησιμοποιείται για την αναγνώριση του προσώπου του χρήστη. Όταν ένας χρήστης βρίσκεται μπροστά από το ρομπότ, ο αλγόριθμος αναγνωρίζει το πρόσωπό του και με την χρήση έτοιμων μοντέλων συλλέγει πληροφορίες σχετικά με το φύλο, την ηλικία και το όνομα του. Στη συνέχεια, το μοντέλο Deep Learning αναλύει τις εκφράσεις προσώπου του χρήστη και αναγνωρίζει το συναίσθημά του. Αφού αναγνωριστεί το συναίσθημα του χρήστη, το ρομπότ μπορεί να προβεί σε διάφορες ενέργειες και εκτέλεση εντολών., όπως παρουσίαση του ονόματος και των χαρακτηριστικών του χρήστη με τη χρήση των πακέτων STT και TTS.

Η χρήση του ROS μας παρέχει μια πλατφόρμα για την ανάπτυξη ευέλικτων και επεκτάσιμων λύσεων, ενώ η ενσωμάτωση αυτών των τριών στοιχείων δίνει τη δυνατότητα στο ρομπότ να διαχειρίζεται διαφορετικές καταστάσεις και να αλληλοεπιδρά με τους χρήστες σε πιο ανθρώπινο επίπεδο.
Με βάση το παραπάνω πείραμα, προέκυψε ένα νέο ερώτημα, αν καταφέρναμε να εγκαταστήσουμε το ROS στο NAO θα παρακάμπταμε το πρόβλημα εισαγωγής του μοντέλου μας στο NAO και άρα την χρήση των κατά πολύ ανώτερων λειτουργιών του σε σχέση με το ρομπότ του πειράματος;

6. Συμπεράσματα

6.1 Περίληψη ευρημάτων

6.1.1 Βασικά ερευνητικά ερωτήματα

Στο πλαίσιο αυτής της εργασίας, διερευνήσαμε τη χρησιμότητα της βαθιάς μάθησης στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο. Συγκεκριμένα, επικεντρωθήκαμε σε ένα μοντέλο CNN που εκπαιδεύτηκε σε ένα υβριδικό σύνολο δεδομένων AffectNet και Fer2013. Με βάση τα ερευνητικά ερωτήματα οι κύριες αναζητήσεις μας ήταν οι εξής:

I. Προσδιορίσαμε τις παραμέτρους απόδοσης του μοντέλου μας, συμπεριλαμβανομένης της ακρίβειας και της ευαισθησίας, στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο.
II. Εξετάσαμε τις δυνατότητες και τους περιορισμούς του μοντέλου μας σε συνδυασμό με άλλες εφαρμογές.
III. Ερευνήσαμε τον αντίκτυπο διαφορετικών υπερ-παραμέτρων, όπως ο ρυθμός εκμάθησης (Learning Rate), το μέγεθος παρτίδας (Batch Size) και η εγκατάλειψη (Dropout), στην απόδοση του μοντέλου μας.
IV. Συγκρίναμε την απόδοση του μοντέλου μας με άλλες μεθόδους τελευταίας τεχνολογίας στον τομέα της αναγνώρισης συναισθημάτων προσώπου.

Αποδείχθηκε ότι το μοντέλο μας δεν είχε πολύ καλή απόδοση στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο, αλλά ήταν άκρως ανταγωνιστικό. Επιπλέον, το μοντέλο μας συνοδεύτηκε από δυνατότητες πρόβλεψης συναισθημάτων σε πραγματικό χρόνο και ανταποκρίθηκε στις απαιτήσεις σε πραγματικά σενάρια, επιτρέποντας σε ένα ρομπότ να αλληλοεπιδρά με τον ανθρώπινο χρήστη. Ωστόσο, υπήρχαν κάποιοι περιορισμοί σχετικά με την ακρίβεια του μοντέλου στην αναγνώριση συγκεκριμένων συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο, ιδίως όταν οι συνθήκες φωτισμού ή οι εκφράσεις του προσώπου ήταν πολύ περίπλοκες ή ασυνήθιστες.

Βρήκαμε ότι η επίδοση του μοντέλου ήταν πολύ ευαίσθητη στις υπερ-παραμέτρους, όπως ο ρυθμός εκμάθησης, το μέγεθος παρτίδας και η εγκατάλειψη. Σε γενικές γραμμές, βρήκαμε ότι οι μεγαλύτερες τιμές Batch Size και ο μικρότερος ρυθμός εκμάθησης οδηγούσαν σε καλύτερη απόδοση του μοντέλου. Ωστόσο, αυτό δεν ίσχυε πάντα για όλα τα σενάρια, καθώς η βέλτιστη τιμή των υπερ-παραμέτρων εξαρτιόταν από τα δεδομένα εκπαίδευσης.

Συνολικά, μπορούμε να συμπεράνουμε ότι η αναγνώριση συναισθημάτων από εκφράσεις προσώπου με χρήση βαθιάς μάθησης είναι μια πολλά υποσχόμενη τεχνολογία για το μέλλον. Η επίδοση του μοντέλου μπορεί να βελτιωθεί με τη βελτίωση του αλγορίθμου εκπαίδευσης και τη βελτιστοποίηση των υπερ-παραμέτρων. Παρόλα αυτά, υπάρχουν ακόμα πολλοί περιορισμοί και προκλήσεις που πρέπει να αντιμετωπιστούν, όπως η αντιμετώπιση της ασυνέπειας των δεδομένων, η αντιμετώπιση της πολυπλοκότητας των εκφράσεων προσώπου και η αναγνώριση του περιβάλλοντος για βελτιωμένη ακρίβεια.

6.1.2 Επιπρόσθετα συμπεράσματα

Αν και βασικός στόχος της εργασίας ήταν η απάντηση των βασικών ερευνητικών ερωτημάτων που τέθηκαν αρχικά και είχαν ως κύριο γνώρισμα το μοντέλο βαθιάς μάθησης, κατά την διάρκεια προετοιμασίας και ενασχόλησης με το αντικείμενο της εργασίας προέκυψαν περαιτέρω συμπεράσματα:

- Η υποδομή που χρησιμοποιείται κατά την εκπαίδευση ενός μοντέλου βαθιάς μάθησης είναι κρίσιμης σημασίας για την απόδοση του μοντέλου σε πραγματικά σενάρια. Η χρήση επαρκούς υπολογιστικής ισχύος, κατάλληλου λογισμικού και βιβλιοθηκών βαθιάς μάθησης είναι απαραίτητη για την αποδοτική εκπαίδευση ενός μοντέλου. Επιπλέον, η κατάλληλη διαχείριση των δεδομένων, η βελτιστοποίηση των υπερ-παραμέτρων και η χρήση κατάλληλων μεθόδων επαλήθευσης του μοντέλου είναι καίριας σημασίας για την επίτευξη υψηλής απόδοσης και για την επιτυχία του συστήματος ανθρώπινης-μηχανικής αλληλεπίδρασης στη ρομποτική.
- Η μετάβαση από την εκπαίδευση ενός μοντέλου βαθιάς μάθησης στην εγκατάστασή του σε ένα σύστημα υπό πραγματικές συνθήκες είναι μια πολύ σημαντική και δύσκολη διαδικασία. Η επίτευξη υψηλής απόδοσης του μοντέλου σε πραγματικό χρόνο εξαρτάται από πολλούς παράγοντες, όπως η υπολογιστική ισχύς του συστήματος, η διαχείριση των δεδομένων, η συνεργασία των διαφορετικών κομματιών (components) του συστήματος, και η βελτιστοποίηση των υπερ-παραμέτρων. Επιπλέον, η μετάβαση από την εκπαίδευση σε ένα σύστημα υπό πραγματικές συνθήκες απαιτεί τη συνεργασία διαφορετικών τεχνολογιών και το συντονισμό των διαφορετικών components του συστήματος. Αυτό απαιτεί εξειδικευμένες γνώσεις και εμπειρία σε διαφορετικούς τομείς, όπως η ρομποτική, η μηχανική μάθηση και η επεξεργασία σήματος. Είναι σημαντικό να λαμβάνουμε υπόψιν τους παράγοντες αυτούς κατά την εκπαίδευση του μοντέλου, καθώς και την επιλογή του συστήματος στο οποίο θα εγκατασταθεί το μοντέλο. Επιπλέον, η αξιολόγηση του μοντέλου σε πραγματικές συνθήκες μπορεί να οδηγήσει σε ανακαλύψεις και βελτιώσεις του μοντέλου και της υποδομής, οι οποίες δεν θα ήταν δυνατές μόνο με την εκπαίδευση του μοντέλου.
- Η ύπαρξη μιας ακαδημαϊκής κοινότητας ή περιβάλλοντος όπου ερευνητές μπορούν να μοιράζονται τις εμπειρίες τους και τα προβλήματα που αντιμετωπίζουν στα πλαίσια της έρευνάς τους είναι απαραίτητη για την πρόοδο του τομέα της τεχνητής νοημοσύνης. Η συνεργασία και η ανταλλαγή γνώσης μεταξύ των ερευνητών μπορεί να οδηγήσει σε καινοτόμες ιδέες και λύσεις σε προβλήματα που αντιμετωπίζονται στην έρευνα.

  Επιπλέον, η δυνατότητα να μοιράζονται κομμάτια κώδικα ή projects μεταξύ των ερευνητών μπορεί να επιταχύνει τη διαδικασία της έρευνας και να δώσει τη δυνατότητα σε νέους ερευνητές να ξεκινήσουν από ένα ήδη υπάρχον project, αντί να ξεκινούν από το μηδέν. Αυτό μπορεί να εξοικονομήσει χρόνο και πόρους και να επιταχύνει την πρόοδο της έρευνας.

6.2 Επόμενα βήματα

Η παρούσα εργασία διερεύνησε την ανάπτυξη ενός μοντέλου βαθιάς μάθησης για την αναγνώριση συναισθημάτων του χρήστη από τις εκφράσεις του προσώπου χρησιμοποιώντας τα σύνολα δεδομένων AffectNet και το υβριδικό AffectFer. Τα αποτελέσματα έδειξαν ότι το προτεινόμενο μοντέλο πέτυχε ανταγωνιστική απόδοση και στα δύο σύνολα δεδομένων. Ωστόσο, μπορούν να γίνουν περαιτέρω βελτιώσεις. Για παράδειγμα:

- Διερεύνηση άλλων μοντέλων βαθιάς μάθησης: Αν και το προτεινόμενο μοντέλο πέτυχε ανταγωνιστική απόδοση, άλλα μοντέλα βαθιάς μάθησης, όπως επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) ή μοντέλα που βασίζονται στην προσοχή (Attention models), θα μπορούσαν να διερευνηθούν για περαιτέρω βελτίωση της απόδοσης. Τα RNN μπορούν να συλλάβουν χρονικές εξαρτήσεις και μακροπρόθεσμες σχέσεις μεταξύ πλαισίων εικόνων προσώπων, με στόχο την αναγνώριση λεπτών αλλαγών στις εκφράσεις του προσώπου με την πάροδο του χρόνου. Τα μοντέλα που βασίζονται στην προσοχή (Attention Models) μπορούν επίσης να βοηθήσουν το μοντέλο να εστιάσει σε συγκεκριμένα μέρη της εικόνας του προσώπου, κάτι που μπορεί να βελτιώσει την απόδοση και να μειώσει τις υπολογιστικές απαιτήσεις.[73-74]
- Ενσωμάτωση ήχου και άλλων τρόπων: Οι εκφράσεις του προσώπου δεν είναι η μόνη πηγή πληροφοριών για την αναγνώριση των συναισθημάτων και άλλοι τρόποι, όπως ο λόγος και οι χειρονομίες, μπορούν να παρέχουν πολύτιμες πληροφορίες. Η ενσωμάτωση ήχου και άλλων τρόπων με εικόνες προσώπου μπορεί να βοηθήσει στη βελτίωση της ακρίβειας και της ευρωστίας του μοντέλου αναγνώρισης συναισθημάτων.[75-76]
- Εξερεύνηση άλλων συνόλων δεδομένων: Το AffectNet και το FER2013 είναι δημοφιλή σύνολα δεδομένων στον τομέα της αναγνώρισης εκφράσεων προσώπου. Ωστόσο, άλλα σύνολα δεδομένων, όπως CK+, RAF-DB ή EmoReact, μπορούν να διερευνηθούν για τη βελτίωση της απόδοσης του μοντέλου και τη δοκιμή των δυνατοτήτων γενίκευσής του.
- Διερεύνηση της μάθηση με μεταφορά (Transfer learning): Η μάθηση με μεταφορά είναι μια χρήσιμη τεχνική για τη βελτίωση της απόδοσης των μοντέλων βαθιάς μάθησης, ιδιαίτερα όταν εργαζόμαστε με περιορισμένα δεδομένα εκπαίδευσης. Εκπαιδεύοντας εκ των προτέρων το μοντέλο σε ένα μεγάλο σύνολο δεδομένων, όπως το ImageNet, και βελτιστοποιώντας το στην εργασία αναγνώρισης εκφράσεων προσώπου, το μοντέλο μπορεί να μάθει να αναγνωρίζει πολύπλοκα χαρακτηριστικά και μοτίβα. [77-79]
- Η εγκατάσταση του λογισμικού Robot Operating System (ROS) στο ρομπότ NAO, που διαθέτει μια ευρεία γκάμα από λειτουργίες για την πραγματοποίηση σεναρίων και αλγορίθμων αλληλεπίδρασης ανθρώπου-μηχανής, ενδέχεται να βελτιώσει την απόδοση του μοντέλου βαθιάς μάθησης στην αναγνώριση συναισθημάτων από εκφράσεις προσώπου σε πραγματικό χρόνο και να προσεγγίσει αποτελεσματικότερα σενάρια εκπαιδευτικών διαδικασιών και γενικότερα πραγματικών συνυπάρξεων ανθρώπου-μηχανής. Η χρήση του ROS στο NAO μπορεί να διευκολύνει την ανάπτυξη και επέκταση του συστήματος, καθώς και την αλληλεπίδρασή του με άλλα συστήματα βαθιάς μάθησης και τεχνητής νοημοσύνης. Επιπλέον, η εγκατάσταση του ROS στο NAO μπορεί να προσφέρει ευκαιρίες για την ανάπτυξη καινοτόμων λειτουργιών και εφαρμογών στο πεδίο της ρομποτικής και της ανθρωποκεντρικής τεχνητής νοημοσύνης.

- Η δημιουργία ενός εσωτερικού περιβάλλοντος στο οποίο οι ερευνητές θα μπορούν να μοιράζονται έτοιμα projects σε jupyter notebooks και θα έχουν πρόσβαση σε μια κοινή βιβλιοθήκη δεδομένων. Η χρήση του Jupyter server μπορεί να διευκολύνει τη διαμοιρασμό πληροφοριών και να βελτιώσει τη συνεργατική εργασία μεταξύ των ερευνητών. Με την εν λόγω πρωτοβουλία, οι ερευνητές μπορούν να εξοικονομήσουν χρόνο και πόρους και να επιταχύνουν την πρόοδο της έρευνας, επίσης να εκμεταλλευτούν κοινούς πόρους και να δημιουργήσουν μια ενωμένη κοινότητα ερευνητών που ασχολούνται με την τεχνητή νοημοσύνη. Η υλοποίηση μιας τέτοιας πρωτοβουλίας μπορεί να συνεισφέρει στη βελτίωση της απόδοσης και της ποιότητας των ερευνητικών δραστηριοτήτων στον τομέα της τεχνητής νοημοσύνης. Επιπλέον, η δημιουργία αυτού του περιβάλλοντος πρέπει να συνοδεύεται από την κατάλληλη υποδομή για τη διαχείριση μεγάλων datasets και την εκπαίδευση των μοντέλων βαθιάς μάθησης. Η ύπαρξη αυτής της υποδομής επιτρέπει στους ερευνητές να αποκτήσουν περισσότερη εμπειρία στη διαχείριση μεγάλων datasets και ταυτόχρονα να επιταχύνεται η διαδικασία της εκπαίδευσης των μοντέλων.

7. Αναφορές

1. Russell, S. J., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Prentice Hall.
2. Mitchell, T. M. (1997). Machine Learning. McGraw Hill.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
4. Alpaydin, E. (2010). Introduction to Machine Learning. MIT Press.
5. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
6. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
8. Haykin, S. (1999). Neural Networks: A Comprehensive Foundation (2nd ed.). Prentice-Hall.
9. Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124.
10. Ekman, P. (2003). Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. Times Books.
11. Matsumoto, D., & Hwang, H. S. (2011). Evidence for the universality of facial expressions of emotion. In Emotion and culture: Empirical studies of mutual influence (pp. 247-266). American Psychological Association.
12. Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychological bulletin, 115(1), 102.
13. Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200.
14. Li, X., Wang, X., Zhu, M., & Gong, Y. (2018). Facial expression recognition using deep learning: A survey. arXiv preprint arXiv:1801.04342.
15. Brackett, M. A., & Salovey, P. (2006). Measuring emotional intelligence with the Mayer-Salovery-Caruso Emotional Intelligence Test (MSCEIT). Psicothema, 18(1), 34-41.
16. Kim, K. J., Cho, H., & Kim, C. (2018). The future of human emotion detection: A survey of emotion detection systems. Sensors, 18(7), 2079.
17. Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 617-627.
18. Zhang, X., Yin, L., Cohn, J. F., & Canavan, S. (2019). Automatic facial expression recognition and analysis for mental health. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).
19. Jain, A. K., Ross, A., & Nandakumar, K. (2016). Introduction to biometrics. Springer.
20. Zafeiriou, S., Kollias, D., & Nicolaou, M. A. (2017). Aff-wild: Valence and arousal in-the-wild challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1989-1998).
21. Breazeal, C., Edsinger, A., & Fitzpatrick, P. (2003). Robots that recognize and respond to user emotion. Proceedings of the IEEE, 91(9), 1583-1590.
22. Zhang, X., Sugano, Y., & Bulling, A. (2017). Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (pp. 623-631).

23. Islam, M. R., Gedeon, T., & Pickering, M. (2017). Machine learning for automatic facial expression recognition in images. International Journal of Machine Learning and Cybernetics, 8(2), 513-535.
24. Liu, J., Deng, Y., Bai, X., & Liu, W. (2018). Recognizing Facial Expressions with Multi-Task Deep Neural Networks. IEEE Transactions on Multimedia, 20(12), 3439-3451.
25. Kaliouby, R. E. (2018). Emotional AI: The Rise of Empathic Media. Bentham Science Publishers.
26. Picard, R. W. (1997). Affective computing. MIT press.
27. Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. (2015). Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. Computer Vision and Image Understanding, 139, 115-149.
28. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet-FER: Affective Facial Expression Recognition Challenge. arXiv preprint arXiv:1908.06317.
29. Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), 39-58.
30. Johnson-Laird, P. N. (1999). The history of research on emotions. In T. Dalgleish & M. Power (Eds.), Handbook of Cognition and Emotion (pp. 3-20). John Wiley & Sons Ltd.
31. Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200.
32. Darwin, C. (1872). The Expression of Emotions in Man and Animals. John Murray.
33. Ekman, P., & Friesen, W. V. (1957). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. Semiotica, 1(1), 49-98.
34. Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124-129.
35. Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial action coding system (FACS). Salt Lake City, UT: Research Nexus eBook.
36. Ekman, P., & Rosenberg, E. L. (2005). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press.
37. Keltner, D., & Ekman, P. (2000). Facial expression of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), Handbook of emotions (2nd ed., pp. 236-249). The Guilford Press.
38. Camras, L. A., & Allison, K. (1989). Children's understanding of emotional facial expressions and verbal labels. Journal of Nonverbal Behavior, 13(4), 295-313.
39. M. A. Nicolaou, M. S. Pattichis, and C. S. Pattichis, "A survey on affective computing: From emotion theory to applications," Journal of Artificial Intelligence and Soft Computing Research, vol. 7, no. 4, pp. 281-305, 2017.
40. Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media.
41. Bishop, C. M. (2006). Pattern recognition and machine learning. springer.
42. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097-1105.
43. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

44. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).
45. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
46. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2016). AffectNet: A facial expression database for valence and arousal recognition. IEEE Transactions on Affective Computing, 10(1), 18-31.
47. Li, X., & Chen, H. (2014). Facial expression recognition using convolutional neural networks: A survey. Frontiers of Computer Science, 8(2), 151-160. doi: 10.1007/s11704-014-4081-9.
48. Barros, P., et al. (2016). Towards a 3D Convolutional Neural Network for the Detection of Facial Expressions in Video. Proceedings of the 29th International Conference on Computer Animation and Social Agents, 1-4. doi: 10.1145/2931002.2931014.
49. Li, X., et al. (2017). A 3D Convolutional Neural Network Approach for Real-Time Action Recognition in Video. Proceedings of the 31st International Conference on Advanced Information Networking and Applications Workshops, 98-103. doi: 10.1109/WAINA.2017.49.
50. Li, X., et al. (2018). A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 4328-4335. doi: 10.1007/978-3-030-01665-2_42.
51. Jaiswal, A., et al. (2016). Transfer learning based emotion recognition using convolutional neural networks. Proceedings of the 2016 6th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 157-160. doi: 10.1109/CONFLUENCE.2016.7508276.
52. Liu, Y., et al. (2017). Ensemble deep learning for facial expression recognition in video. Proceedings of the 25th ACM International Conference on Multimedia, 1156-1164. doi: 10.1145/3123266.3127904.
53. Zhang, X., Liu, F., & Wang, W. (2020). Emotion recognition using facial expression analysis with deep learning: A review. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2675-2689.
54. Kaltwang, S., Todorovic, S., & Pantic, M. (2012). Context-based recognition of depression from nonverbal behaviour. Journal of affective disorders, 139(1), 1-6.
55. Li, S., Li, H., & Fu, Y. (2020). Facial expression recognition: Recent advances, challenges and future directions. IEEE Transactions on Affective Computing.
56. Breazeal, C. (2003). Emotion and sociable humanoid robots. International Journal of Human-Computer Studies, 59(1-2), 119-155.
57. D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. ACM Computing Surveys (CSUR), 47(3), 1-43.
58. Huang, Y., & Wang, Y. (2019). Emotion-aware computing: A survey of affective computing systems that recognize, understand, and respond to human emotions. ACM Computing Surveys (CSUR), 52(1), 1-34.
59. Kappas, A., & Krämer, N. C. (2011). Emotion in games: Theory and praxis—A concise overview. In Emotions and Personality in Personalized Services (pp. 1-27). Springer, Berlin, Heidelberg.
60. Wang, W., Ji, Q., & Gunasekaran, S. (2017). Facial expression recognition: A brief tutorial overview. IEEE Signal Processing Magazine, 34(3), 122-129.
61. Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

62. Fabrýcký, M., Kahancová, M., Štrumbelj, E., Hyňka, T., & Bahník, Š. (2019). EmoReact: A multimodal dataset for fine-grained emotion recognition in a crowd reacting to stimuli in a real-world event. Frontiers in psychology, 10, 1118.
63. Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PloS one, 13(5), e0196391.
64. Han, J., et al. "A study on affect recognition using the AffectNet database." International Journal of Multimedia Information Retrieval, vol. 9, no. 3, 2020, pp. 223-233.
65. Barsoum, E., et al. "Training deep networks for facial expression recognition with crowd-sourced label distribution." Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016, pp. 427-431.
66. Deng, J., et al. "Imagenet: A large-scale hierarchical image database." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.
67. Russakovsky, O., et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision, vol. 115, no. 3, 2015, pp. 211-252.
68. Zhou, B., et al. "Places: An image database for deep scene understanding." arXiv preprint arXiv:1610.02055, 2016.
69. Srivastava, N., et al. "Dropout: A simple way to prevent neural networks from overfitting." Journal of Machine Learning Research, vol. 15, no. 1, 2014, pp. 1929-1958.
70. Kingma, D.P., and Ba, J. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980, 2014.
71. Bengio, Y., et al. "Greedy layer-wise training of deep networks." Advances in neural information processing systems, 2007, pp. 153-160.
72. Srivastava, N., et al. "Dropout: A simple way to prevent neural networks from overfitting." Journal of Machine Learning Research, vol. 15, no. 1, 2014, pp. 1929-1958.
73. Zhang, Z., & Salakhutdinov, R. (2016). Towards end-to-end face detection and recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4675-4684.
74. Zhou, X., & De la Torre, F. (2016). Factorized spatial attention network for face recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4823-4831.
75. Katsigiannis, S., & Ramzan, N. (2018). A Survey on Deep Learning Advances on Different 3D Data Representations. arXiv preprint arXiv:1808.01462.
76. Lu, H., & Plataniotis, K. N. (2018). Deep learning for emotion recognition on small datasets using transfer learning. IEEE Transactions on Affective Computing, 9(3), 357-371.
77. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 6105-6114).
78. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
79. Khorrami, P., Lanchantin, J., Sobhanmanesh, F., & Huang, J. (2019). Transfer learning in deep neural networks: An overview. IEEE Signal Processing Magazine, 36(3), 56-75.
80. Ghayoumi, A., & Farzaneh, A. (2019). Multi-Objective Facial Expression Recognition using Hybrid Datasets. In 2019 10th International Conference on Intelligent Systems (IS) (pp. 001-006). IEEE.

81. Viola, P., & Jones, M. (2004). Robust Real-Time Face Detection. International Journal of Computer Vision, 57(2), 137-154. https://doi.org/10.1016/j.ijcv.2004.08.023
82. "NAO Robot." SoftBank Robotics. https://www.softbankrobotics.com/emea/en/nao (πρόσβαση στις 12 Μαρτίου 2023).
83. Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA Workshop on Open Source Software, 2009, pp. 1-5.
84. "ROS: The Robot Operating System." ROS.org. https://www.ros.org/ (πρόσβαση στις 12 Μαρτίου 2023).
85. "GStreamer - Welcome to the GStreamer project." GStreamer. https://gstreamer.freedesktop.org/ (πρόσβαση στις 12 Μαρτίου 2023).

8. Παραρτήματα

8.1 Δημιουργία dataset σε μορφή φακέλων & υπό-φακέλων

```
import pandas as pd
import os
import itertools
from pathlib import Path
import shutil
# Paths
directory = r"D:\affectnet\Manually_Annotated_Images\\"
root = r"D:\affectnet\validation_dataset"
main = r"\Images"
labels = pd.read_csv(r"D:\affectnet\Manually_Annotated_file_lists\validation.csv",
                     names=["file_name", "2", "3", "4", "5", "6", "file_label", "8", "9"])
unique_labels = labels["file_label"].unique()
print(labels)
# Create folders
try:
    Path(root + main).mkdir(parents=True, exist_ok=True)
    for label in unique_labels:
        Path(root + main + "\\" + str(label)).mkdir(parents=True, exist_ok=False)
except(Exception,):
    print("Folder exists")
def get_files(path):
    files_needed = []
    for file in os.listdir(path):
        if os.path.isfile(os.path.join(path, file)):
            files_needed.append(file)
    return files_needed
def get_files_for_specific_label(df, label_no):
    sub_label = df['file_name'].loc[df['file_label'] == label_no]
    sub_label = pd.DataFrame(sub_label)
    sub_label[['name_start', 'name_end']] = pd.DataFrame(sub_label)['file_name'].str.split('/', expand=True)
```

```
    files_to_move = sub_label['name_end'].tolist()
    return files_to_move
def copy_files_to_another_folder(list_of_files, source, destination):
    for file_name in list_of_files:
        full_file_name = os.path.join(source, file_name)
        if os.path.isfile(full_file_name):
            # Watch out the difference between copy and move
            shutil.copy(full_file_name, destination)
            print("Moved files")
def process_image_files(directory_):
    for root_, dirs_, files_ in os.walk(directory_):
        sort_dir = [int(x) for x in dirs_]
        sort_dir.sort()
        """
        - We order all sub-directories and iterate over them
        - I also have all the labels, I want to go through each subdirectory, get the images for the specific label
        - If labels exist in the directory i copy them to the new folder
            - I can either do it iteratively or intersect the two lists and move them afterwards
        """
        for sub_dir, label in itertools.product(sort_dir, unique_labels):
            print("-- Checking sub directory {} and label {}--".format(sub_dir, label))
            all_files_in_directory = get_files(directory_ + str(sub_dir))
            all_files_for_label = get_files_for_specific_label(labels, label)
            files_to_move = list(set(all_files_in_directory).intersection(all_files_for_label))
            print("-- Files to Move --")
            print(files_to_move)
            # the destination here needs to be the same as the label
            copy_files_to_another_folder(files_to_move, root_ + "\\" + str(sub_dir),
                                         root + "\\Images" + "\\{}".format(label))
if __name__ == "__main__":
        process_image_files(directory)
```

8.2 Αλγόριθμος εκπαίδευσης μοντέλου

```
#!/usr/bin/env python

# coding: utf-8

import numpy as np

import cv2

from tensorflow.keras.models import Sequential

from tensorflow.keras.optimizers import Adam

from keras.layers import MaxPooling2D

import tensorflow as tf

from tensorflow.keras import Model

from tensorflow.keras.callbacks import CSVLogger

from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint

from tensorflow.keras.layers import Input, Conv2D , Dropout, MaxPool2D, Flatten, Dense, Activation, BatchNormalization

from tensorflow.keras import Model

from tensorflow.keras.preprocessing.image import ImageDataGenerator

from tensorflow.keras.regularizers import l2

from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint

import scipy

import os

import matplotlib.pyplot as plt

import sys

from tensorflow.keras.callbacks import CSVLogger

os.environ["KMP_DUPLICATE_LIB_OK"]="TRUE"

MODEL_FNAME = "AffectNetTrained7_model.h5"

base_dir = "dataset"

tmp_model_name = "AffectNetTrained7tmp.h5"

INPUT_SIZE = 224

BATCH_SIZE = 64

physical_devices = tf.config.list_physical_devices()

print("DEVICES : \n", physical_devices)

print('Using:')

print('\t\u2022 Python version:',sys.version)

print('\t\u2022 TensorFlow version:', tf.__version__)
```

```
print('\t\u2022 tf.keras version:', tf.keras.__version__)

print('\t\u2022 Running on GPU' if tf.config.list_physical_devices('GPU') else '\t\u2022 GPU device not found. Running on CPU')

count = 0

previous_acc = 0

if not os.path.exists(MODEL_FNAME):

    """ Create Shallow Model"""

    input = Input(shape =(INPUT_SIZE,INPUT_SIZE,1))

    weight_initializer = tf.keras.initializers.RandomNormal(mean=0.0, stddev=0.01, seed=None)

    bias_initializer=tf.keras.initializers.Zeros()

    x = Conv2D (filters =32, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001), kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(input)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = MaxPool2D(pool_size =2, strides =2, padding ='same')(x)

    x = Conv2D (filters =64, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001), kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = Conv2D (filters =64, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001), kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = MaxPool2D(pool_size =2, strides =2, padding ='same')(x)

    x = Conv2D (filters =128, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001),kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = Conv2D (filters =128, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001),kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = MaxPool2D(pool_size =2, strides =2, padding ='same')(x)
```

```
    x = Conv2D (filters =128, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001),kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = Conv2D (filters =128, kernel_size =3, padding ='same', kernel_regularizer=l2(0.001),kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    x = BatchNormalization()(x)

    x = Activation('relu')(x)

    x = MaxPool2D(pool_size =2, strides =2, padding ='same')(x)

    x = Flatten()(x)

    x = Dropout(0.5)(x)

    x = Dense(units = 64, activation ='relu', kernel_regularizer=l2(0.001), kernel_initializer=weight_initializer,bias_initializer=bias_initializer)(x)

    output = Dense(units=7,activation='softmax')(x)

    # creating the model

    model = Model (inputs=input, outputs =output)

    # to be sure GPU memory is cleaned after last train

    m = model

    m.save(tmp_model_name)

    del m

    tf.keras.backend.clear_session()

    # model summary

    model.summary()

""" Prepare the Dataset for Training"""

train_dir = r"/home/greg/emotion_recognition/train_dataset/Images"

val_dir = r"/home/greg/emotion_recognition/validation_dataset/Images"

train_batches = ImageDataGenerator(rescale = 1 / 255.,horizontal_flip=True, rotation_range=90,brightness_range=[0.2,1.2],zoom_range=[0.5,1.5]).flow_from_directory(train_dir,

                                                    target_size=(INPUT_SIZE,INPUT_SIZE),

                                                    shuffle=True,

                                                    seed=42,

                                                   color_mode="grayscale",

                                                   class_mode='categorical',
```

```
                                        batch_size=BATCH_SIZE)
val_batches = ImageDataGenerator(rescale = 1 / 255.).flow_from_directory(val_dir,
                                        target_size=(INPUT_SIZE,INPUT_SIZE),
                                        shuffle=True,
                                        seed=42,
                                        color_mode="grayscale",
                                        class_mode='categorical',
                                        batch_size=BATCH_SIZE)
class CustomLearningRateScheduler(tf.keras.callbacks.Callback):
    def __init__(self, schedule):
        super(CustomLearningRateScheduler, self).__init__()
        self.schedule = schedule
        self.weights_monitor = open("AffectNetTrained7weights.txt", "w+")
    # learning rate scheduler is called at the end of each epoch. learning rate decreases if needed
    def on_epoch_end(self, epoch, logs=None):
        if not hasattr(self.model.optimizer, "lr"):
                raise ValueError('Optimizer must have a "lr" attribute.')
        # Get the current learning rate from model's optimizer.
        lr = float(tf.keras.backend.get_value(self.model.optimizer.learning_rate))
        # Call schedule function to get the scheduled learning rate.
        # keys = list(logs.keys())
        # print("keys",keys)
        val_acc = logs.get("val_categorical_accuracy")
        scheduled_lr = self.schedule(lr, val_acc)
        # Set the value back to the optimizer before this epoch starts
        tf.keras.backend.set_value(self.model.optimizer.lr, scheduled_lr)
    #first and last convolutional layers' and dense layer's weights are saved at the beginning of each epoch
    def on_epoch_begin(self, epoch, logs=None):
        epoch_str = "beginning of epoch : "+ str(epoch)
        self.weights_monitor.write(epoch_str)
        self.weights_monitor.write(str(model.get_layer("conv2d").weights))
        self.weights_monitor.write(str(model.get_layer("conv2d_1").weights))
```

```
        self.weights_monitor.write(str(model.get_layer("dense_1").weights))
def learning_rate_scheduler(lr, val_acc):
    global count
    global previous_acc

    if val_acc <= previous_acc:
    #  print("acc ", val_acc, "previous acc ", previous_acc)
        count += 1
    else:
        previous_acc = val_acc
        count = 0
    if count >= 10:
        print("acc doesnt improve for 10 epoch, learnin rate decreased by /10")
        count = 0
        lr /= 10
        print("new learning rate:", lr)

    return lr
#compile the model by determining loss function Binary Cross Entropy, optimizer as SGD
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.0001),
              loss=tf.keras.losses.CategoricalCrossentropy(),
              metrics=[tf.keras.metrics.CategoricalAccuracy()],
               sample_weight_mode=[None])
   #if validation accuracy doesnt improve for 15 epoch, stop training
early_stopping = EarlyStopping(monitor='val_categorical_accuracy', patience=15)
#save the model if a better validation accuracy then previous better accuracy is obtained
checkpointer = ModelCheckpoint(filepath=MODEL_FNAME, verbose=1,
save_best_only=True)
# write accuracy and loss history to the log.csv
csv_logger = CSVLogger('log.csv', append=True, separator=' ')
history=model.fit(train_batches,
    validation_data = val_batches,
    epochs = 100,
    verbose = 1,
```

```
    shuffle = True,
    callbacks =
[checkpointer,early_stopping,CustomLearningRateScheduler(learning_rate_scheduler),csv_logger])
""" Plot the train and validation Loss """
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('AffectNetTrained7 loss with learning_rate=0.0001')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'validation'], loc='upper left')
plt.show()
""" Plot the train and validation Accuracy """
plt.plot(history.history['categorical_accuracy'])
plt.plot(history.history['val_categorical_accuracy'])
plt.title('AffectNetTrained7 accuracy with learning_rate=0.0001')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'validation'], loc='upper left')
plt.show()
print("End of Training")
tf.keras.backend.clear_session()
```

8.3 Περίληψη του μοντέλου εκπαίδευσης

```
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 224, 224, 1)]     0

conv2d (Conv2D)              (None, 224, 224, 32)      320

batch_normalization (BatchN  (None, 224, 224, 32)      128
ormalization)

activation (Activation)      (None, 224, 224, 32)      0

max_pooling2d (MaxPooling2D  (None, 112, 112, 32)      0
)

conv2d_1 (Conv2D)            (None, 112, 112, 64)      18496

batch_normalization_1 (Batc  (None, 112, 112, 64)      256
hNormalization)

activation_1 (Activation)    (None, 112, 112, 64)      0

conv2d_2 (Conv2D)            (None, 112, 112, 64)      36928

batch_normalization_2 (Batc  (None, 112, 112, 64)      256
hNormalization)

activation_2 (Activation)    (None, 112, 112, 64)      0

max_pooling2d_1 (MaxPooling  (None, 56, 56, 64)        0
2D)

conv2d_3 (Conv2D)            (None, 56, 56, 128)       73856

batch_normalization_3 (Batc  (None, 56, 56, 128)       512
hNormalization)

activation_3 (Activation)    (None, 56, 56, 128)       0

conv2d_4 (Conv2D)            (None, 56, 56, 128)       147584

batch_normalization_4 (Batc  (None, 56, 56, 128)       512
hNormalization)

activation_4 (Activation)    (None, 56, 56, 128)       0

max_pooling2d_2 (MaxPooling  (None, 28, 28, 128)       0
2D)

conv2d_5 (Conv2D)            (None, 28, 28, 128)       147584

batch_normalization_5 (Batc  (None, 28, 28, 128)       512
hNormalization)
```

activation_5 (Activation) (None, 28, 28, 128) 0

conv2d_6 (Conv2D) (None, 28, 28, 128) 147584

batch_normalization_6 (Batc hNormalization) (None, 28, 28, 128) 512

activation_6 (Activation) (None, 28, 28, 128) 0

max_pooling2d_3 (MaxPooling 2D) (None, 14, 14, 128) 0

flatten (Flatten) (None, 25088) 0

dropout (Dropout) (None, 25088) 0

dense (Dense) (None, 64) 1605696

dense_1 (Dense) (None, 7) 455

Total params: 2,181,191
Trainable params: 2,179,847
Non-trainable params: 1,344

8.4 Εφαρμογή του μοντέλου πρόβλεψης

```
import numpy as np
import cv2
import tensorflow as tf
face_detection = cv2.CascadeClassifier('haar_cascade_face_detection.xml')
camera = cv2.VideoCapture(0)
camera.set(cv2.CAP_PROP_FRAME_WIDTH, 1024)
camera.set(cv2.CAP_PROP_FRAME_HEIGHT, 768)
settings = {
   'scaleFactor': 1.3,
   'minNeighbors': 5,
   'minSize': (50, 50)}
labels = ["neutral", "happy", "sad", "surprise", "angry"]
model = tf.keras.models.load_model('expressionOLD.model')
while True:
        ret, img = camera.read()
        gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
        detected = face_detection.detectMultiScale(gray, **settings)
        for x, y, w, h in detected:
```

```
            cv2.rectangle(img, (x, y), (x+w, y+h), (245, 135, 66), 2)
            cv2.rectangle(img, (x, y), (x+w//3, y+20), (245, 135, 66), -1)
            face = gray[y+5:y+h-5, x+20:x+w-20]
            face = cv2.resize(face, (48, 48))
            face = face/255.0
            predictions = model.predict(np.array([face.reshape((48, 48, 1))])).argmax()
            state = labels[predictions]
            font = cv2.FONT_HERSHEY_SIMPLEX
            cv2.putText(img, state, (x+10, y+15), font, 0.5, (255, 255, 255), 2, cv2.LINE_AA)
    cv2.imshow('Facial Expression', img)
    img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    if cv2.waitKey(5) != -1:
        break
camera.release()
cv2.destroyAllWindows()
```