



Πανεπιστήμιο Δυτικής Αττικής
Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Διπλωματική εργασία

Αυτόματη Ανίχνευση, Ανάλυση και Αναγνώριση Συναισθημάτων

Στυλιανός Γεωργαράς
71345354

Επιβλέπων:

Χρήστος Τρούσσας
Επίκουρος Καθηγητής

Αιγάλεω - Αθήνα, Οκτώβριος, 2022

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Γεωργαράς Στυλιανός του Αποστόλου, με αριθμό μητρώου 71345354 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι: «Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

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

Ο Δηλών

Στυλιανός Γεωργαράς



Η διπλωματική εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τρι-μελή επιτροπή:

Όνοματεπώνυμο	Βαθμίδα/Ιδιότητα	Υπογραφή
Παναγιώτα Τσελέντη	Ε.ΔΙ.Π.	
Χρήστος Τρούσσας	Επίκουρος Καθηγητής	
Γεώργιος Μπαρδής	Επίκουρος Καθηγητής	

Ευχαριστίες

Η παρούσα διπλωματική εργασία αποτελεί έργο όλων όσων συνέβαλλαν με οποιοδήποτε τρόπο. Αρχικά, θα ήθελα να ευχαριστήσω τον κ.Χρήστο Τρούσσα για την καθοδήγηση, τον χρόνο που μου παρείχε, καθ' όλη την διάρκεια της εργασίας. Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου για την υποστήριξη όλα τα χρόνια της ακαδημαϊκής μου πορείας. Τέλος, θα ήθελα να ευχαριστήσω την Δήμητρα για την πολύτιμη βοήθεια της, ψυχολογική και έμπρακτη, κατά την διάρκεια της εκπόνησης της διπλωματικής αυτής.

Ημερομηνία 11/10/2022

Στυλιανός Γεωργαράς

Περίληψη

Στην παρούσα διπλωματική ερευνήθηκαν οι τεχνικές μηχανικής και βαθιάς μάθησης για την αναγνώριση συναισθημάτων στη περιοχή του προσώπου μέσω νευρωνικών δικτύων. Ο σκοπός της εργασίας ήταν η αυτόματη ανίχνευση των επτά αποδεκτών συναισθημάτων, όπως χαρά, λύπη, έκπληξη, ουδετερότητα, αηδία, φόβος και θυμός. Το εγχείρημα αυτό είχε στόχο την χρήση της πληροφορίας του συναισθήματος και την κατασκευή ειδικής εφαρμογής που ενσωματώνει το εκπαιδευμένο μοντέλο.

Η τροφοδότησή του έγινε με το dataset FER-2013 και για την εκπαίδευση επιλέχθηκε η αρχιτεκτονική συνελκτικών δικτύων. Χρησιμοποιήθηκαν τελευταίες εκδόσεις από γνωστές βιβλιοθήκες που ειδικεύονται σε θέματα βαθιάς μάθησης, όπως Tensorflow & Keras. Η εκπαίδευσή του, λόγω τον όγκο δεδομένων και της υψηλής απαιτούμενης υπολογιστικής ισχύς, έγινε σε μια υπηρεσία νέφους που παρέχει υπολογιστικούς πόρους κάρτας γραφικών ειδική για το σκοπό αυτό. Το εργαλείο αυτό παρέχεται από την Google και καλείται Colab. Η υλοποίηση εφαρμογής που έχει την λειτουργία της ανίχνευσης συναισθημάτων μέσω κάμερας ή μέσω φωτογραφιών αναπτύχθηκε στην βιβλιοθήκη Streamlit.

Η ακρίβεια του μοντέλου εξετάστηκε από γνωστές μετρήσεις μοντέλων μηχανικής μάθησης. Η αξιολόγηση από τις συναρτήσεις του Tensorflow με βάση τα δεδομένα ελέγχου είναι της τάξης 68%. Τέλος, έγινε σύγκριση του μοντέλου με αυτό της DeepFace, αφού ακολουθούν το ίδιο πρότυπο συναισθημάτων.

Λέξεις Κλειδιά: Βαθία Μάθηση, Αναγνώριση Συναισθήματος, Νευρωνικά Δίκτυα, Όραση Υπολογιστών

Abstract

In the present diploma thesis, we will explore various machine learning and deep learning techniques used in facial recognition of emotions through neural networks. The purpose of this thesis was the automatic recognition of the seven universal emotions; happiness, sadness, surprise, contempt, disgust, fear and anger. The work presented aimed at using the information behind emotions and creating a special application that implements the trained model.

The model was loaded with the FER-2013 dataset and for its training, architecture of convolutional neural networks was chosen. Furthermore, the latest editions of well-known libraries specialized in matters of deep learning, such as TensorFlow & Keras were used. Due to the large amount of data and the high demand in the computational power needed, the training of the model took place in a cloud service that provides graphics card (GPU) computational resources. This tool is provided by Google and is called Colab. The materialization of the application recognizing emotions through a webcam or photos was developed in the Streamlit framework.

The accuracy of the model was examined by well-known metrics of machine learning models. The evaluation, according to Tensorflow's functions, based on validation data reached 68%. Finally, our model was compared to that of DeepFace, since they adopt the same pattern of emotions.

Keywords: Deep Learning, Emotion Recognition, Neural Networks, Computer Vision

Πίνακας περιεχομένων

Ευχαριστίες	iii
Περίληψη	iv
Abstract	v
Πίνακας περιεχομένων	viii
Πίνακας σχημάτων	xi
Πίνακας πινάκων	xii
1 Εισαγωγή	1
1.1 Περιγραφή του προβλήματος	1
1.2 Σκοπός της εργασίας	2
1.3 Ανασκόπηση βιβλιογραφίας	4
1.4 Διάρθρωση της εργασίας	5
2 Θεωρητικό Υπόβαθρο	6
2.1 Τεχνητή Νοημοσύνη - Artificial Intelligence	6
2.2 Μηχανική Μάθηση - Machine Learning	7
2.2.1 Τύποι Μηχανικής Μάθησης	7
2.3 Βαθιά Μάθηση - Deep Learning	9
2.4 Βιολογικός Νευρώνας - Biological Neuron	11
2.5 Τεχνητός Νευρώνας - Artificial Neuron	11
2.5.1 Τύποι Συνάρτηση Ενεργοποίησης	12
2.6 Τεχνητά Νευρωνικά Δίκτυα - Artificial Neural Networks (ANN)	15

2.6.1	Υπερπροσαρμογή - Overfitting	16
2.7	Συνελκτικά Νευρωνικά Δίκτυα - Convolutional Neural Networks	17
2.8	Αναγνώριση Συναισθήματος	19
2.8.1	Το πρόβλημα της Αναγνώρισης Συναισθήματος	19
2.8.2	Μοντέλα Συναισθημάτων	19
2.8.3	Εφαρμογές στην Αναγνώριση Συναισθήματος	20
3	Μεθοδολογία Έρευνας	22
3.1	Πορεία Έρευνας & Υλοποίησης	22
3.2	Περιορισμοί & Λύσεις Υλοποίησης	23
3.3	Στόχοι της Έρευνας	24
4	Αρχιτεκτονική	25
4.1	Dataset	25
4.2	Αρχιτεκτονική Μοντέλου	28
4.3	Λειτουργικότητα	29
4.3.1	Λογισμικό	29
4.3.2	Διαδικτυακή εφαρμογή	30
5	Σχεδιασμός & Υλοποίηση λογισμικού	32
5.1	Περιβάλλοντα ανάπτυξης και Βιβλιοθήκες	32
5.1.1	Tensorflow - Keras	32
5.1.2	OpenCV	33
5.1.3	Google Colab	33
5.1.4	NumPy	33
5.1.5	Matplotlib	34
5.1.6	Streamlit	34
5.2	Υλοποίηση Μοντέλου	34
5.3	Υλοποίηση Εφαρμογής	40
5.3.1	Γενικό Μενού	40
5.3.2	Λειτουργία - Ανάλυση εικόνας	40
5.3.3	Λειτουργία - Ανίχνευση συναισθήματος με βίντεο	41

6	Παραδείγματα Χρήσης (Use Cases)	44
6.1	Ανάλυση εικόνας	44
6.1.1	Παράδειγμα λογισμικού	44
6.2	Ανίχνευση συναισθήματος σε βίντεο	47
6.3	Περίπτωση Διαδικτυακής Εφαρμογής	50
6.4	Προβλήματα	53
7	Αξιολόγηση	55
7.1	Μετρήσεις εκπαιδευμένου μοντέλου	57
7.2	Σύγκριση	59
7.2.1	Περίπτωση φόβου	59
7.2.2	Περίπτωση αηδίας	60
7.2.3	Περίπτωση χαράς	61
7.2.4	Περίπτωση λύπης	62
7.2.5	Περίπτωση ουδέτερος	63
7.2.6	Περίπτωση έκπληξης	64
7.2.7	Περίπτωση θυμού	65
8	Συμπεράσματα & Μελλοντικές Επεκτάσεις	67
8.1	Συμπεράσματα	67
8.2	Περιορισμοί	68
8.3	Μελλοντικές Επεκτάσεις	69
	Βιβλιογραφικές Αναφορές	71

Πίνακας σχημάτων

Εικόνα 1.	Μάθηση με επίβλεψη[19]	8
Εικόνα 2.	Ενισχυτική Μάθηση[19]	9
Εικόνα 3.	Βαθιά Μάθηση[21]	9
Εικόνα 4.	Deep Learning vs Machine Learning Techniques[23]	10
Εικόνα 5.	Αναπαράσταση Βιολογικού Νευρώνα[25]	11
Εικόνα 6.	Αναπαράσταση Τεχνητού Νευρώνα[26]	12
Εικόνα 7.	Συνάρτηση Κατωφλίου	13
Εικόνα 8.	Σιγμοειδής Συνάρτηση	14
Εικόνα 9.	Συνάρτηση ReLU	15
Εικόνα 10.	Απλό ΤΝΔ χωρίς bias [22]	16
Εικόνα 11.	Παράδειγμα Overfitting [28]	17
Εικόνα 12.	Μια από τις πρώτες υλοποιήσεις CNN μοντέλων [22]	18
Εικόνα 13.	Δεδομένα Εκπαίδευσης	26
Εικόνα 14.	Δεδομένα Ελέγχου	26
Εικόνα 15.	Αρχιτεκτονική Μοντέλου	28
Εικόνα 16.	Διάγραμμα ροής για την 1η λειτουργία	29
Εικόνα 17.	Διάγραμμα ροής για την 2η λειτουργία	30
Εικόνα 18.	Αρχικοποίηση παραμέτρων	34
Εικόνα 19.	Πρώτο επίπεδο νευρωνικού συνελκτικού δικτύου	35
Εικόνα 20.	Δεύτερο επίπεδο νευρωνικού συνελκτικού δικτύου	36
Εικόνα 21.	Τρίτο επίπεδο νευρωνικού συνελκτικού δικτύου	36
Εικόνα 22.	Τέταρτο επίπεδο νευρωνικού συνελκτικού δικτύου	36
Εικόνα 23.	Πέμπτο επίπεδο νευρωνικού συνελκτικού δικτύου	37

Εικόνα 24.	Μετατροπή των δεδομένων σε μονοδιάστατο πίνακα	37
Εικόνα 25.	Πλήρης σύνδεση των νευρωνικών συνελκτικών δικτύων	38
Εικόνα 26.	Πλήρης σύνδεση των νευρωνικών συνελκτικών δικτύων	38
Εικόνα 27.	Επίπεδο Κατηγοριοποίησης	38
Εικόνα 28.	Ορισμός παραμέτρων εκπαίδευσης	39
Εικόνα 29.	Εκπαίδευση του μοντέλου	40
Εικόνα 30.	Μενού εφαρμογής	40
Εικόνα 31.	Φόρτωση εικόνας στην εφαρμογή και παραγωγή αποτελεσμάτων	41
Εικόνα 32.	Κατασκευή αντικειμένου για την υποστήριξη βίντεο ζωντανής ροής	42
Εικόνα 33.	Callback συνάρτηση για την διαχείριση βίντεο	43
Εικόνα 34.	Χαρούμενος - Παράδειγμα 1ο	45
Εικόνα 35.	Γραφική παράσταση για την εικόνα 34	45
Εικόνα 36.	Χαρούμενος - Παράδειγμα 2ο	46
Εικόνα 37.	Γραφική παράσταση για την εικόνα 36	46
Εικόνα 38.	Χαρούμενος - Παράδειγμα 3ο	47
Εικόνα 39.	Γραφική παράσταση για την εικόνα 38	47
Εικόνα 40.	Λειτουργία βίντεο για την περίπτωση του ενθουσιασμού	48
Εικόνα 41.	Περίπτωση που κάποιος δεν κάνει κάποια έκφραση	49
Εικόνα 42.	Πιθανά μηνύματα στο χρήστη - Περίπτωση 1η	49
Εικόνα 43.	Πιθανά μηνύματα στο χρήστη - Περίπτωση 2η	49
Εικόνα 44.	Λειτουργίες διαδικτυακής εφαρμογής	50
Εικόνα 45.	Βήματα για την επιλογή εικόνας	51
Εικόνα 46.	Ανάλυση εικόνας και εμφάνιση των συναισθημάτων	51
Εικόνα 47.	Επιλογή νέας εικόνας για νέα ανάλυση	52
Εικόνα 48.	Αρχική σελίδα στην λειτουργία της ανάλυσης βίντεο	52
Εικόνα 49.	Έλεγχος εφαρμογής για βίντεο - Περίπτωση 1η	53
Εικόνα 50.	Έλεγχος εφαρμογής για βίντεο - Περίπτωση 2η	53
Εικόνα 51.	Στιγμιότυπο από την αξιολόγηση του μοντέλου	55
Εικόνα 52.	Ακρίβεια του μοντέλου	56
Εικόνα 53.	Απώλεια του μοντέλου	56

Εικόνα 54.	Πίνακας σύγκρισης - Confusion Matrix	58
Εικόνα 55.	Περίπτωση Φόβου- Μοντέλου DeepFace	59
Εικόνα 56.	Περίπτωση Φόβου- Μοντέλου της εργασίας	60
Εικόνα 57.	Περίπτωση Αηδίας- Μοντέλου DeepFace	61
Εικόνα 58.	Περίπτωση Αηδίας- Μοντέλου της εργασίας	61
Εικόνα 59.	Περίπτωση Χαράς - Μοντέλου DeepFace	62
Εικόνα 60.	Περίπτωση Χαράς - Μοντέλου της εργασίας	62
Εικόνα 61.	Περίπτωση Λύπης - Μοντέλου DeepFace	63
Εικόνα 62.	Περίπτωση Λύπης - Μοντέλου της εργασίας	63
Εικόνα 63.	Περίπτωση Ουδέτερος - Μοντέλου DeepFace	64
Εικόνα 64.	Περίπτωση Ουδέτερος - Μοντέλου της εργασίας	64
Εικόνα 65.	Περίπτωση Έκπληξης - Μοντέλου DeepFace	65
Εικόνα 66.	Περίπτωση Έκπληξης - Μοντέλου της εργασίας	65
Εικόνα 67.	Περίπτωση Θυμού - Μοντέλου DeepFace	66
Εικόνα 68.	Περίπτωση Θυμού - Μοντέλου της εργασίας	66

Πίνακας πινάκων

Πίνακας 1.	Πίνακας Ομάδων & Ακρίβειας μοντέλων[15]	4
Πίνακας 2.	Αναλυτικός Πίνακας με των εικόνων του dataset	27
Πίνακας 3.	Σχέση συναισθήματος με ανάλογη κίνηση [42]	27
Πίνακας 4.	Πίνακας Ταξινόμησης - Classification report	58

Κεφάλαιο 1

Εισαγωγή

Στον παρόν κεφάλαιο παρουσιάζεται η εισαγωγή στο θέμα της διπλωματικής, γίνεται η ανασκόπηση βιβλιογραφίας και ορίζεται ο σκοπός της παρούσας εργασίας.

1.1 Περιγραφή του προβλήματος

Στην εποχή μας, οι εφαρμογές της Τεχνητής Νοημοσύνης συμβάλλουν στην επίλυση προβλημάτων της καθημερινότητας μας όλο και περισσότερο. Οι δυνατότητες των νευρωνικών δικτύων σε συνδυασμό με την εποχή του Big Data¹ μας έδωσαν νέα εργαλεία για την κατασκευή μοντέλων. Τα συγκεκριμένα μοντέλα είναι ικανά να επιλύουν πολύπλοκα και δύσκολα προβλήματα και να φτάνουν τις ανθρώπινες επιδόσεις στην αναγνώριση αντικειμένων, κάτι που πριν από λίγα χρόνια θεωρείτο ανέφικτο.[1]

Η ανάπτυξη της τεχνολογίας και νέων μέσων επικοινωνίας γέννησε και νέες ανάγκες. Οι μελέτες που έχουν στόχο την αναγνώριση συγκεκριμένων συναισθημάτων στο νέο αυτό ψηφιακό κόσμο είναι πολυάριθμες. [2], [3] Οι έρευνες παγκοσμίως στρέφονται σε αυτό τον υποτομέα της επιστήμης των υπολογιστών και προσπαθούν να ανιχνεύσουν συναισθήματα σε οποιαδήποτε μορφή.

Οι άνθρωποι εύκολα μπορούν να βγάλουν κάποιο συμπέρασμα για το πώς νιώθει κάποιος παρατηρώντας το πρόσωπό του και τις αντιδράσεις του. Ωστόσο, για τους υπολογιστές είναι μια πολύπλοκη και δύσκολη διαδικασία, καθώς καλούνται να επεξεργαστούν τεράστιο όγκο εικονοστοιχείων και να παρατηρήσουν μοτίβα που σχηματίζονται ανάμεσά

¹μεγάλος όγκος δεδομένων με μεγάλη ποικιλία, εύκολα προσβάσιμος

τους. Εταιρείες και οργανισμοί στρέφονται σε τέτοιες λύσεις αφού προσφέρουν μια διαφορετική μορφή ανατροφοδότησης για την κατασκευή προϊόντων στην οποία δεν είχαμε πρόσβαση μέχρι τώρα. Άλλες εταιρείες στρέφονται σε μελέτες για να χρησιμοποιήσουν αυτές τις τεχνολογίες για να αυτοματοποιήσουν διαδικασίες αλληλεπίδρασης με ρομπότ. [4] Επομένως, η κατασκευή μοντέλων που εξάγουν πληροφορίες για το συναίσθημα από ένα πρόσωπο έχουν προσελκύσει πλήθος ερευνητών και επιστημόνων την τελευταία δεκαετία. Επίσης, υπάρχουν οι ερευνητές που μελετούν το συναίσθημα σε άλλες μορφές, όπως αυτή του γραπτού κειμένου που παρουσιάζεται καθημερινά σε αναρτήσεις των μέσων κοινωνικής δικτυωτής. [5], [6], [7]

Η γλώσσα του σώματος και κυρίως οι εκφράσεις προσώπου είναι μια πολύ σημαντική πληροφορία για οποιαδήποτε μορφή διαπροσωπικής επικοινωνίας.[8] Τα συναισθήματα που εκφράζονται στο πρόσωπο έχουν κατηγοριοποιηθεί σε επτά, αποδεκτά παγκοσμίως συναισθήματα. [9] Αυτά περιγράφονται ως εξής: Χαρά, Λύπη, Ενθουσιασμός, Αηδία, Φόβος, Θυμός και Αδιαφορία.

1.2 Σκοπός της εργασίας

Η εργασία εστιάζει στην αυτόματη ανίχνευση και αναγνώριση συναισθήματος σε πραγματικό χρόνο κάνοντας χρήση των τεχνικών βαθιάς μάθησης. Οι τεχνικές παρουσιάζουν τεράστια άνοδο λόγω των υψηλών αποδόσεων σε σχέση με τους παραδοσιακούς αλγόριθμους μηχανικής μάθησης. [10] [11] [12] Στόχος είναι η κατασκευή ενός μοντέλου που αυτόματα ταξινομεί τα χαρακτηριστικά του προσώπου με βάση κάποιο παγκοσμίως αποδεκτό συναίσθημα. Πιο συγκεκριμένα, το μοντέλο, που θα κληθεί να αντιμετωπίσει το πρόβλημα, κατασκευάζεται από ένα εξειδικευμένο είδος νευρωνικού δικτύου, το συνελκτικό νευρωνικό δίκτυο. Το συγκεκριμένο είδος αρχιτεκτονικής έχει μεγάλη άνοδο την τελευταία δεκαετία και είναι κατάλληλο για δεδομένα, όπως οι εικόνες. Η ανάπτυξη ενός τέτοιου μοντέλου απαιτεί μεγάλο όγκο δεδομένων για την τροφοδότησή του, χρόνο για την εκπαίδευση και την επανεκπαίδευση, παραμετροποιώντας τα δεδομένα και βελτιστοποιώντας την απόδοση του τελικού μοντέλου. Στη διάρκεια της επανεκπαίδευσης, συχνά, επανεξετάστηκε η ποιότητα του αρχικού όγκου δεδομένων και όπου ήταν δυνατό έγιναν βελτιστοποιήσεις για την μεγαλύτερη ακρίβεια του τελικού μοντέλου. Στόχος της εργα-

σίας είναι να προσεγγίσουμε την ακρίβεια του τελικού μοντέλου κοντά στο 70%, καθώς όπως θα αναλύσουμε και στην ανασκόπηση της βιβλιογραφίας, εκεί κυμαίνονται και οι υπόλοιπες υλοποιήσεις.

Έπειτα από την εκπαίδευση του, σειρά έχει ο έλεγχος του μοντέλου και η αξιολόγηση της απόδοσής του σε πραγματικές συνθήκες. Η ακρίβειά του εξαρτάται από την επιτυχία του μοντέλου να αναγνωρίσει ένα συναίσθημα και να το ταξινομήσει ανάλογα. Η κατηγοριοποίηση αυτή θα γίνεται σε πραγματικό χρόνο καθώς στόχος αυτού του εγχειρήματος είναι να μπορούμε αυτή την πληροφορία να την χρησιμοποιήσουμε σε κάποια εφαρμογή στην καθημερινότητα μας. Η δομή μιας τέτοιας εφαρμογής θα παρουσιαστεί στα τελευταία κεφάλαια για την παρουσίαση και την εφαρμογή του μοντέλου που θα υλοποιηθεί.

Το λογισμικό που αναπτύχθηκε αρχικά, αλλά και η διαδικτυακή εφαρμογή στην συνέχεια, δέχονται από την κάμερα κάθε καρτέ (frame). Ένα ήδη εκπαιδευμένο μοντέλο είναι υπεύθυνο για την αναγνώριση του προσώπου και την αποστολή του στο μοντέλο αναγνώρισης συναισθήματος. Το τελευταίο εξετάζει τις εκφράσεις προσώπου σε πραγματικό χρόνο, με σκοπό την κατηγοριοποίηση αυτών των εκφράσεων σε κάποιο παγκοσμίως αποδεκτό συναίσθημα. Η εργασία έχει στόχο την εξαγωγή αυτής της πληροφορίας, έτσι ώστε άλλες υλοποιήσεις να μπορούν να βασιστούν στην υπάρχουσα για την κατασκευή νέων εφαρμογών. Οι μελλοντικές αυτές εφαρμογές, θα στοχεύουν στην λύση γενικών προβλημάτων στην καθημερινή ζωή, όπως για παράδειγμα στην αποτροπή των εκρήξεων θυμού μέσα σε όχημα εν κινήσει. Αυτό θα έχει ως αποτέλεσμα την καλύτερη διαχείριση θυμού από τον οδηγό αλλά και την μείωση δυστυχημάτων που προκαλούνται από τέτοιους παράγοντες.

Η εφαρμογή θα παρέχει την δυνατότητα ανάλυσης μιας εικόνας που απεικονίζει κάποιο πρόσωπο. Το μοντέλο θα εξάγει την πληροφορία των συναισθημάτων και με την βοήθεια της εφαρμογής θα παρέχει στον χρήστη την πληροφορία αυτή με μορφή γραφημάτων. Το γράφημα θα αποτελείται από τα επτά βασικά συναισθήματα, όπου το καθένα θα λαμβάνει μια τιμή επί τοις εκατό, ανάλογα με την εικόνα που του δόθηκε. Το άθροισμα των ποσοστών αυτών κάθε φορά είναι 100%. Επομένως, στην εικόνα μπορεί να αναγνωριστούν ένα ή περισσότερα κύρια συναισθήματα.

1.3 Ανασκόπηση βιβλιογραφίας

Αρχικά, πραγματοποιήθηκε έρευνα για να βρεθούν παρόμοιες μελέτες και να κατευθυνθεί η εργασία προς ένα συγκεκριμένο dataset. Το FER-2013 είναι αρκετά δημοφιλές με πολλές έρευνες κάθε χρόνο να το χρησιμοποιούν για μελέτη και επιλέχτηκε αυτό για την εργασία. Επομένως, πραγματοποιήθηκε έρευνα στις υλοποιήσεις του διαγωνισμού του συγκεκριμένου dataset [13]. Ωστόσο, τα περισσότερα παραγόμενα μοντέλα δεν ικανοποιούν βασικές προϋποθέσεις καθώς είναι υπερεκπαιδευμένα (overfitted). Παρατηρούμε όμως, ότι και τα μη-υπερεκπαιδευμένα μοντέλα καθώς και οι νικητές του διαγωνισμού αυτού πετυχαίνουν ως και 71.16%, με τους υπόλοιπους διαγωνιζόμενους να κυμαίνονται στα $65 \pm 5\%$. Έπειτα, έγινε έρευνα σε άρθρα και άλλες επιστημονικές πηγές για να εξάγουμε πιο συγκεκριμένα συμπεράσματα για την μέγιστη ακρίβεια του τελικού μοντέλου, βασιζόμενοι σε αυτά τα δεδομένα. Η μεταπτυχιακή διατριβή, που πραγματεύεται παρόμοιο θέμα με την παρούσα εργασία [14], πετυχαίνει 66.55% μέγιστη ακρίβεια. Στην συνέχεια, κατευθύνθηκε η έρευνα στις πηγές που αναφέρονται στην διατριβή και συγκεκριμένα σε μία επιστημονική αναφορά . [15]

Πίνακας 1. Πίνακας Ομάδων & Ακρίβειας μοντέλων[15]

Ομάδες	Ακρίβεια
Yichuan Tang	71.162%
Yingbo Zhou, Chetan Ramaiah	69.267%
Maxim Milakov	68.821%
Radu Ionescu, Marius Popescu, Cristian Grozea	67.484%

Στην αναφορά παρουσιάζονται αποτελέσματα από διαφορετικές ομάδες στο πρόβλημα της ανίχνευσης συναισθήματος και τα αποτελέσματα για την ακρίβεια των μοντέλων που υλοποιήθηκαν βρίσκονται στον πίνακα 1. Τέλος, άλλη μια επιστημονική εργασία που θα λάβουμε υπόψη μας είναι μια υλοποίηση που φτάνει ως και 73.28%. [16]

1.4 Διάρθρωση της εργασίας

Εδώ γίνεται σύνοψη του περιεχομένου του κάθε κεφαλαίου.

- **Κεφάλαιο 2 - Θεωρητικό Υπόβαθρο:** Παρουσιάζονται οι απαραίτητες έννοιες της μηχανικής μάθησης, της βαθιάς μάθησης και ποιο είναι το πρόβλημα της αναγνώρισης συναισθήματος.
- **Κεφάλαιο 3 - Μεθοδολογία Έρευνας:** Στο κεφάλαιο αυτό, αναλύεται η μεθοδολογία που ακολουθήθηκε για την επίλυση του βασικού προβλήματος της εργασίας αλλά και των επιμέρους που προέκυψαν κατά την διαδικασία.
- **Κεφάλαιο 4 - Αρχιτεκτονική:** Στο παρόν κεφάλαιο γίνεται λόγος για την αρχιτεκτονική του μοντέλου που κατασκευάστηκε, την παρουσίαση του dataset με το οποίο τροφοδοτήθηκε το μοντέλο και τα κατάλληλα διαγράμματα ροής για την κατανόηση της λειτουργίας του προγράμματος.
- **Κεφάλαιο 5 - Σχεδιασμός & Υλοποίηση λογισμικού:** Στο εν λόγω κεφάλαιο γίνεται επίδειξη του κώδικα που υλοποιήθηκε και της παρουσίαση των λειτουργιών του.
- **Κεφάλαιο 6 - Παραδείγματα Χρήσης - Use Cases:** Στο κεφάλαιο αυτό παρουσιάζονται μερικές πιθανές λειτουργίες χρήσης, όπου θα μπορούσε να χρησιμοποιηθεί το μοντέλο.
- **Κεφάλαιο 7 - Αξιολόγηση - Συζήτηση:** Γίνεται η αξιολόγηση του μοντέλου βάσει κάποιων μετρικών και ο σχολιασμός των αποτελεσμάτων αυτών.
- **Κεφάλαιο 8- Συμπεράσματα & Μελλοντική Εργασία:** Τέλος, στο κεφάλαιο αυτό παρουσιάζονται τα συμπεράσματα της εργασίας, οι περιορισμοί που προέκυψαν και δίνονται προτάσεις για μελλοντική ανάπτυξη της.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Τεχνητή Νοημοσύνη - Artificial Intelligence

Η Τεχνητή Νοημοσύνη (TN) είναι ο κλάδος της πληροφορικής ο οποίος ασχολείται με την «αυτοματοποίηση των δραστηριοτήτων που συσχετίζουμε με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση...»(Bellman, 1978) [17]. Η επιστήμη της TN άρχισε να αναπτύσσεται στα τέλη του Β΄ Παγκοσμίου Πολέμου με την μεγάλη συνεισφορά του Alan Turing, ο οποίος πρότεινε την γνωστή δοκιμασία «Turing Test». Με λίγα λόγια η δοκιμασία Turing αναφέρεται σε μία διαδικασία πεπερασμένων γραπτών ερωτοαπαντήσεων που γίνεται μεταξύ μίας μη-νοήμων οντότητας (υπολογιστής/μηχανή) και μίας νοήμων (άνθρωπος). Η μηχανή που θα καταφέρει να δίνει απαντήσεις τις οποίες θα μπορούσε να δώσει ένας άνθρωπος και να ξεγελάσει τον εξεταστή περνάει την δοκιμασία.

Η τεχνητή νοημοσύνη αποτελεί κομβικό σημείο ανάμεσα σε πολλές επιστήμες εκ των οποίων είναι ακόμη και η ψυχολογία, η φιλοσοφία, η νευροεπιστήμη, η γλωσσολογία, η επιστήμη μηχανικών και ως επί το πλείστον η πληροφορική. Ο συνδυασμός όλων αυτών των επιστημών αποσκοπεί στην σύνθεση μίας ευφυούς συμπεριφοράς, η οποία θα παίρνει στοιχεία συλλογιστικής, μάθησης και προσαρμογής στο ανάλογο περιβάλλον και θα εφαρμόζεται σε μηχανές ή υπολογιστές ειδικής κατασκευής. Η τεχνητή νοημοσύνη χωρίζεται στην *συμβολική* τεχνητή νοημοσύνη και στην *υπολογιστική* τεχνητή νοημοσύνη. Η πρώτη στοχεύει στην εξομοίωση της ανθρώπινης νοημοσύνης με αλγοριθμικό τρόπο και συστήματα που στηρίζονται σε κανόνες, λογική και βάσεις γνώσης, όπως τα έμπειρα συ-

στήματα (expert systems). Στην δεύτερη κατηγορία, στην οποία γίνεται προσπάθεια ανα- παραγωγής βιολογικών διεργασιών, όπως αυτή της ανθρώπινης ευφυΐας, χρησιμοποιούνται στοιχειώδη αριθμητικά μοντέλα. Συστήματα που μπορούν να προκύπτουν από αυτά είναι τα τεχνητά νευρωτικά δίκτυα (artificial neural network).

2.2 Μηχανική Μάθηση - Machine Learning

Μηχανική μάθηση καλείται το πεδίο στο οποίο μελετάται η ανάπτυξη αλγορίθμων, που έχουν την ικανότητα να μαθαίνουν από τα δεδομένα παράγοντας προβλέψεις σχετικά με αυτά. [18].

2.2.1 Τύποι Μηχανικής Μάθησης

Η μηχανική μάθηση χωρίζεται σε τρεις βασικές επιμέρους κατηγορίες που είναι οι εξής:

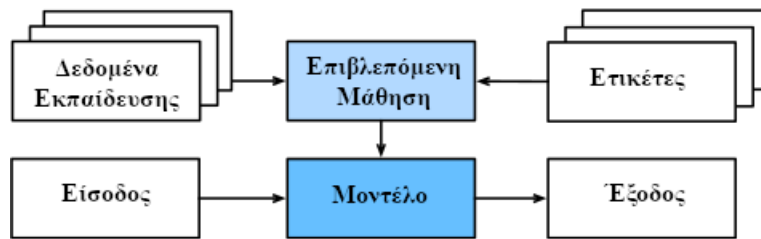
- Μάθηση με επίβλεψη (Supervised learning)
- Μάθηση χωρίς επίβλεψη (Un-supervised learning)
- Ενισχυτική Μάθηση με επίβλεψη (Reinforcement learning)

Θεωρείται ότι ο συνδυασμός της μάθησης με επίβλεψη και με χωρίς επίβλεψη δημιουργεί μια τέταρτη κατηγορία Μάθηση με ημι-επίβλεψη (Semi-Supervised learning).

Μάθηση με επίβλεψη

Στην περίπτωση αυτή ο αλγόριθμος που εκπαιδεύεται έχει σκοπό να παράγει τις κατάλληλες ετικέτες (labels) για τις εισόδους που δέχεται. Ένα ζεύγος εισόδου-ετικέτας καλείται παράδειγμα. [19] Επομένως, στόχος σε αυτήν την κατηγορία είναι να κατασκευαστεί ένα μοντέλο το οποίο θα είναι ικανό να δώσει σωστή ετικέτα στην εκάστοτε είσοδο. Ένα τέτοιο μοντέλο θα μπορεί να χρησιμοποιηθεί ως ανιχνευτής ανεπιθύμητων email (spam filter). Το μοντέλο θα αναθέτει σε ένα μήνυμα την ετικέτα «Spam» ή «Not Spam». Σε αυτήν την κατηγορία σημαντικό ρόλο έχει ο επιβλέπων του μοντέλου, αφού αυτός εί-

να υπεύθυνος για να παρέχει στο μοντέλο το κατάλληλο dataset,¹ που θα αποτελείται από παραδείγματα με ετικέτες.



Εικόνα 1. Μάθηση με επίβλεψη[19]

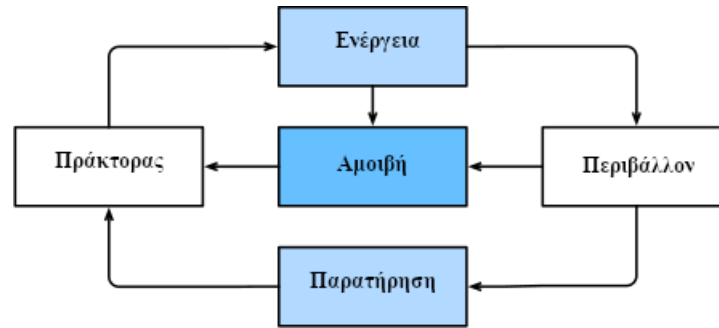
Μάθηση χωρίς επίβλεψη

Σε αυτή τη μορφή μάθησης ο αλγόριθμος μαθαίνει από δεδομένα, χωρίς να του παρέχονται δεδομένα με ετικέτες. Στόχος του μοντέλου, που θα αναπτυχθεί είναι να βρίσκεται σε θέση ο αλγόριθμος να βρει κάποια τάση στα δεδομένα στα οποία δέχεται, χωρίς την παρουσία του ανθρώπου να το επιβλέπει. Συχνά προβλήματα που καλείται η συγκεκριμένη μορφή μάθησης είναι η ομαδοποίηση δεδομένων (clustering) φτιάχνοντας ομάδες με κοινά χαρακτηριστικά. Άλλα παραδείγματα χρήσης είναι η εύρεση ποικιλίας στα δεδομένα, η απομάκρυνση θορύβου από μία κατανομή δεδομένων.[20] Ένας αρκετά δημοφιλής αλγόριθμος μάθησης χωρίς επίβλεψη είναι αυτός των K-μέσων (K-means), που είναι ένας αλγόριθμος ομαδοποίησης και χρησιμοποιείται κατά κόρον στον κλάδο της εξόρυξης δεδομένων.

Ενισχυτική Μάθηση

Η ενισχυτική μάθηση δεν βασίζεται σε κάποιο συγκεκριμένο σύνολο δεδομένων αλλά δρα μέσω της αλληλεπίδρασης με το περιβάλλον. Η διαδικασία εκμάθησης γίνεται με την προϋπόθεση ότι υπάρχει κάποια μορφή θετικής ή αρνητικής ανατροφοδότησης του μοντέλου. Ο εκάστοτε πράκτορας (agent) μαθαίνει από την διαδικασία της δοκιμής (επιλογής) και σφάλματος. [20] Ο αυτόματος αυτός, πράκτορας εκτελεί κάποια ενέργεια και δέχεται κάποια «επιβράβευση» εάν πετύχει το σκοπό του, ενώ λαμβάνει κάποια «παρατήρηση» εάν δεν έκανε την σωστή επιλογή. [19] Στο διάγραμμα ροής 2 παρουσιάζεται η διαδικασία που ακολουθεί ένας τέτοιος αλγόριθμος.

¹ σύνολο δεδομένων με μία ή περισσότερες εγγραφές



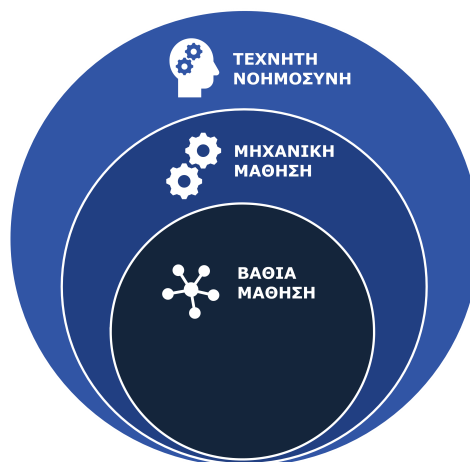
Εικόνα 2. Ενισχυτική Μάθηση[19]

Μάθηση με ημι-επίβλεψη

Η μάθηση με ημι-επίβλεψη είναι ο συνδυασμός των δύο παραπάνω τεχνικών (μάθησης με επίβλεψη και χωρίς). Η χρήση της γίνεται για ένα σύνολο δεδομένων που δεν είναι πλήρως ως προς τις ετικέτες. Επομένως, ο αλγόριθμος έχει το ρόλο να ισοσταθμίσει τα δεδομένα που δεν έχουν χαρακτηριστεί από την αρχή, όπως γίνεται στην μάθηση χωρίς επίβλεψη. Στην συνέχεια, γίνεται η διαδικασία και συνδέει τα δεδομένα εισόδου με την έξοδο, όπως γίνεται στην μάθηση με επίβλεψη.

2.3 Βαθιά Μάθηση - Deep Learning

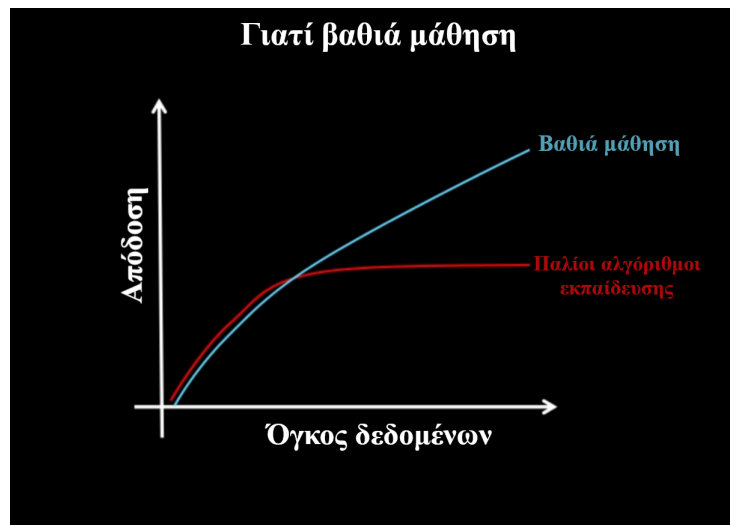
Η Βαθιά Μάθηση είναι υποπεδίο της Μηχανικής Μάθησης και ο όρος «βαθιά» προέκυψε, όχι επειδή ο αλγόριθμος μαθαίνει σε μεγαλύτερο «βάθος», αλλά επειδή υπάρχουν πολλά επίπεδα τεχνητών νευρωνικών δικτύων.



Εικόνα 3. Βαθιά Μάθηση[21]

Οι τεχνικές αυτές δεν ανακαλύφθηκαν τώρα, ήταν γνωστές, αλλά άρχισαν να χρησιμοποιούνται τις τελευταίες δεκαετίες λόγω της μεγάλης διαθεσιμότητας σε δεδομένα και την αύξηση της υπολογιστικής ισχύος.[1] Κάποιοι από τους πρώτους αλγόριθμους μηχανικής μάθησης, που αναγνωρίζουμε σήμερα, επρόκειτο να γίνουν υπολογιστικά μοντέλα για βιολογική μάθηση, δηλαδή μοντέλα για το πώς η μάθηση λειτουργεί ή για το τι συμβαίνει στον εγκέφαλο κατά την διαδικασία της. Αυτό είχε ως αποτέλεσμα, ένα από τα ονόματα, που έχει πάρει η Βαθιά Μάθηση, να είναι Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANNs). Τα μοντέλα Βαθιάς Μάθησης είναι μηχανικά συστήματα, εμπνευσμένα από τον βιολογικό εγκέφαλο (είτε πρόκειται για ανθρώπινο, είτε για κάποιον ενός άλλου ζώου).[20]

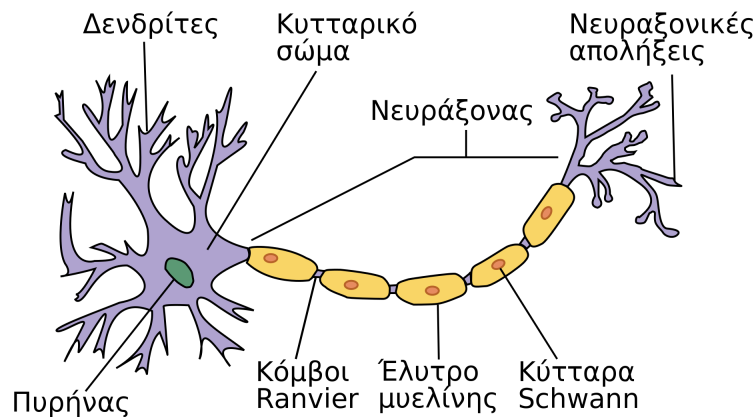
Η αποτελεσματικότητά τους σε δύσκολα προβλήματα μέχρι τώρα τις καθιστά πολύ ανώτερες από τις παραδοσιακές μεθόδους μηχανικής μάθησης. Η αναγνώριση αντικειμένων, για παράδειγμα, θεωρείτο μία δύσκολη δουλειά για μία μηχανή, αφού ο μηχανισμός της όρασης ενός ανθρώπου είναι αρκετά περίπλοκος. Ωστόσο, τα τελευταία χρόνια με την ανάπτυξη ανάλογων μοντέλων βαθιάς μάθησης, σε συνδυασμό με την μηχανική όραση, οι μηχανές ξεπέρασαν την ανθρώπινη απόδοση σε συγκεκριμένες εργασίες. Στο πεδίο της αναγνώρισης εικόνας, οι ειδικοί της υπολογιστικής όρασης πριν από είκοσι χρόνια, δεν θεωρούσαν ότι τέτοια μοντέλα είναι δυνατόν να κατασκευαστούν. [22]



Εικόνα 4. Deep Learning vs Machine Learning Techniques[23]

2.4 Βιολογικός Νευρώνας - Biological Neuron

Ο εγκέφαλος και το υπόλοιπο νευρικό σύστημα αποτελούνται από δισεκατομμύρια νευρώνες. Αν και οι νευρώνες διαφέρουν σε κάποια σημεία μεταξύ τους, ο κάθε νευρώνας έχει τέσσερα βασικά μέρη: (1) το κυτταρικό σώμα, (2) αρκετούς δενδρίτες, (3) έναν ή περισσότερους νευράξονες και (4) τις τελικές απολήξεις. Όταν ένα νευρώνας διεγείρεται κατάλληλα στο κυτταρικό του σώμα ή μέσω των δενδριτών του, μία νευρική ώση, δηλαδή μία αλλαγή στο ηλεκτρικό δυναμικό του κυττάρου, μετακινείται κατά μήκος του νευράξονα μέχρι τις τελικές του απολήξεις.[24]



Εικόνα 5. Αναπαράσταση Βιολογικού Νευρώνα[25]

2.5 Τεχνητός Νευρώνας - Artificial Neuron

Ένας νευρώνας ουσιαστικά είναι μία μονάδα επεξεργασίας και το βασικό δομικό στοιχείο ενός νευρωνικού δικτύου. Η εικόνα 6 είναι μια μοντελοποίηση ενός τεχνητού νευρώνα και η βάση για μια οικογένεια νευρωνικών δικτύων. Το υπολογιστικό μοντέλο ενός νευρώνα έχει τρία βασικά στοιχεία: τις συνάψεις, έναν αθροιστή και μία συνάρτηση ενεργοποίησης.

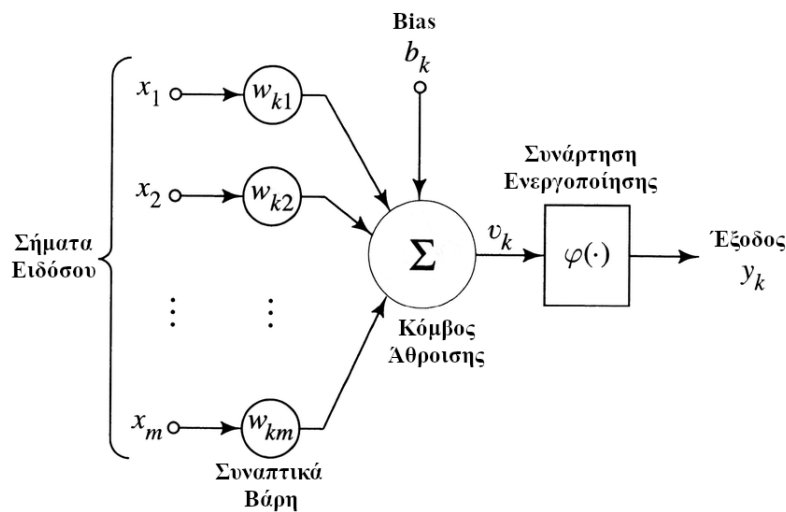
Η σύνναψη χαρακτηρίζεται από ένα βάρος (weight) w_{kj} , το οποίο θα πολλαπλασιάσει το σήμα x_j στην είσοδο της σύνναψης j . Η κάθε σύνναψη καταλήγει στον νευρώνα k . Αντίθετα, με τον βιολογικό νευρώνα, οι τιμές που μπορεί να πάρουν τα βάρη ενός τεχνητού μπορεί να είναι και αρνητικές. Όλες οι συνάψεις καταλήγουν σε έναν αθροιστή (adder), ο οποίος αθροίζει τα σήματα της εισόδου, που είναι σταθμισμένα από τα συναπτικά βάρη του νευ-

ρώνα. Τέλος, η συνάρτηση ενεργοποίησης $\varphi(v)$ (activation function) είναι υπεύθυνη για την πυροδότηση του νευρώνα. [26]

Οι μαθηματικοί τύποι που περιγράφουν το μοντέλο της εικόνας 6 είναι οι εξής:

$$u_k = \sum_{j=1}^m w_{kj}x_j \quad (2.1)$$

$$y_k = \varphi(u_k + b_k) \quad (2.2)$$



Εικόνα 6. Αναπαράσταση Τεχνητού Νευρώνα[26]

2.5.1 Τύποι Συνάρτηση Ενεργοποίησης

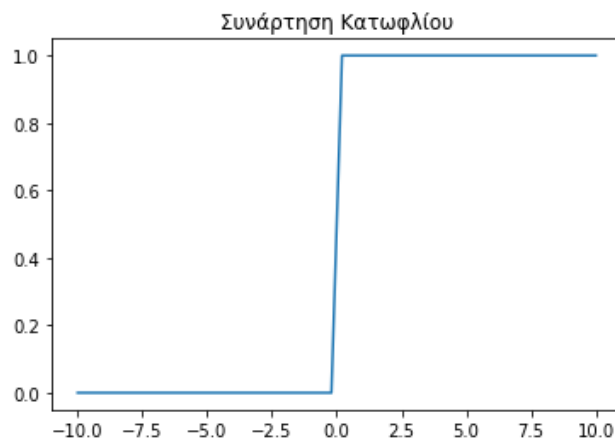
Οι συναρτήσεις ενεργοποίησης, συμβολίζονται ως $\varphi(u)$, περιορίζουν το πλάτος στην έξοδο μετά τον αθροιστή και συχνά καλούνται και συναρτήσεις περιορισμού. Αναλυτικότερα, είναι μαθηματικές εξισώσεις που καθορίζουν την έξοδο ενός νευρωνικού δικτύου. Δέχονται το σήμα εξόδου και είναι υπεύθυνες για την κανονικοποίηση του στο διάστημα $[0,1]$ ή $[-1,1]$ [26]. Παρακάτω θα αναλύσουμε τις βασικές συναρτήσεις ενεργοποίησης, που χρησιμοποιούνται κατά κόρον στα νευρωνικά δίκτυα.

Συνάρτηση Κατωφλίου - Threshold Function

Αρχικά, θα περιγράψουμε ένα μοντέλο τέτοιας συνάρτησης, εις αναγνώριση του πρωτοποριακού έργου των McCulloch και Pitts(1943). Παρουσιάστηκε ως μοντέλο McCulloch-

Pitts με την έξοδο του ενός νευρώνα να λαμβάνει τιμή 1, εάν το τοπικό πεδίο του συγκεκριμένου νευρώνα είναι μη αρνητικό και 0 σε κάθε άλλη περίπτωση. Σήμερα, στους κλάδους της μηχανικής αναφέρεται ως συνάρτηση κατωφλίου ή αλλιώς Heaviside συνάρτηση και ο μαθηματικός τύπος, με τον οποίο εκφράζεται, είναι ο εξής: [26]

$$\varphi(v) = \begin{cases} 1 & \text{εάν } u \geq 0 \\ 0 & \text{εάν } u < 0 \end{cases} \quad (2.3)$$

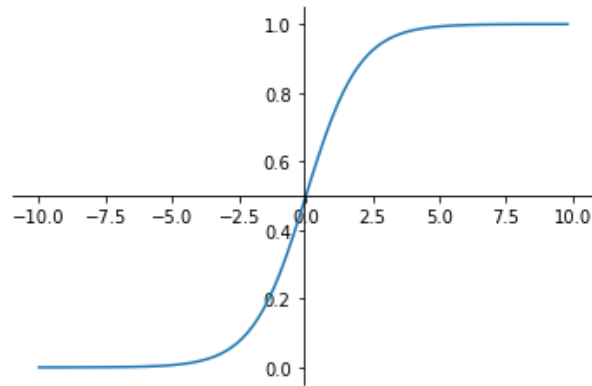


Εικόνα 7. Συνάρτηση Κατωφλίου

Σιγμοειδής Συνάρτηση - Sigmoid Function

Η Σιγμοειδής Συνάρτηση, της οποίας όπως παρατηρούμε στην εικόνα 8 έχει σχήμα «S», είναι πλέον μία κοινή μορφή συνάρτησης ενεργοποίησης που χρησιμοποιείται στην κατασκευή νευρωνικών δικτύων. Ο μαθηματικός της τύπος είναι ο 2.4

$$\varphi(x) = \frac{1}{1 + e^{-x}} \quad (2.4)$$



Εικόνα 8. Σιγμοειδής Συνάρτηση

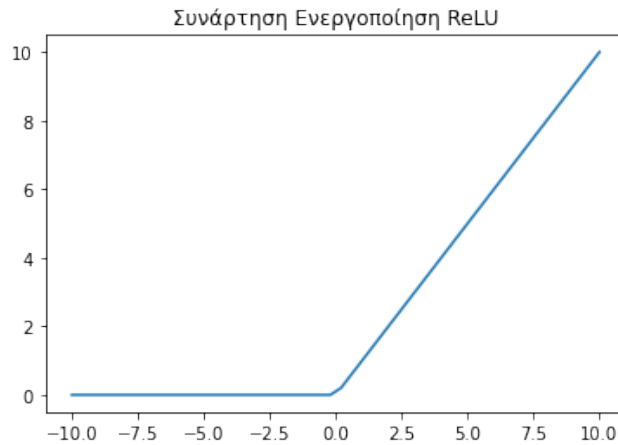
Rectified Linear Unit (ReLU) Function

Η συνάρτηση ReLU είναι μια μη γραμμική συνάρτηση, που είναι πολύ δημοφιλής στα νευρωνικά δίκτυα. Πιο συγκεκριμένα, χρησιμοποιείται για την κατασκευή πολυεπίπεδων νευρωνικών δικτύων και στα δίκτυα νευρώνων βαθιάς μάθησης. Η συνάρτηση περιγράφεται από τον παρακάτω μαθηματικό τύπο. Στην εξίσωση 2.5 το x αναπαριστά την είσοδο.

$$\varphi(x) = \max(x, 0) \quad (2.5)$$

Σύμφωνα με τον τύπο η έξοδος της συνάρτησης αυτής είναι η μέγιστη τιμή ανάμεσα στο 0 και στην τιμή της εισόδου. Η έξοδος δηλαδή, είναι ίση με το 0 όταν η είσοδος είναι αρνητική. Αντιθέτως, όταν η τιμή στην είσοδο είναι θετική τότε η τιμή εξόδου της συνάρτησης είναι 1.[19] Άρα μπορούμε να ξαναγράψουμε την 2.6 ως εξής:

$$\varphi(x) = \begin{cases} 0 & \text{αν } x < 0 \\ x & \text{αν } x \geq 0 \end{cases} \quad (2.6)$$

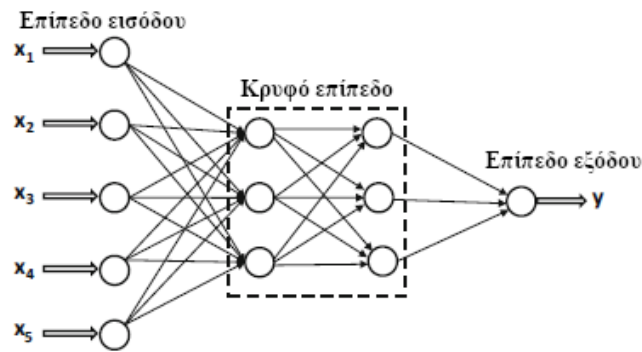


Εικόνα 9. Συνάρτηση ReLU

2.6 Τεχνητά Νευρωνικά Δίκτυα - Artificial Neural Networks (ANN)

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι δομημένα σε επίπεδα (layers), τα οποία είναι γνωστά και ως στρώματα, ενώ τα επίπεδα που μεσολαβούν ανάμεσά τους ονομάζονται κρυφά επίπεδα (hidden layers) και η ύπαρξή τους δεν είναι αναγκαίο στοιχείο για τα ΤΝΔ. Τα πρώτα από τα επίπεδα, που αναφέρθηκαν, αποτελούνται από κάποιες μονάδες (units) ή κόμβους (nodes) οι οποίες συνδέονται μεταξύ τους με τέτοιο τρόπο, ώστε να μπορούν να αναπτύσσουν συνδέσμους και με άλλες μονάδες του δικού τους ή άλλων επιπέδων. Αυτή η διαδικασία ξεκινά με την επίδραση της μίας μονάδας σε άλλες μέσω της διέγερσης ή αναστολής της ενεργοποίησής τους. Προκειμένου να συμβεί κάτι τέτοιο, η μονάδα δέχεται το σταθμισμένο άθροισμα όλων των εισόδων, λειτουργία που επιτυγχάνεται μέσω των συνδέσμων που τερματίζουν σε αυτήν, και φτιάχνει μία μοναδική έξοδο μέσω της συνάρτησης μετάβασης, σε περίπτωση που το άθροισμα υπερβαίνει μία συνάρτηση κατωφλίου. Το επίπεδο εισόδου (input layer) δέχεται τις εισόδους και είναι υπεύθυνο για την τροφοδότηση του όλου του δικτύου. Το συγκεκριμένο επίπεδο εισόδου έχει την ικανότητα επικοινωνίας με ένα ή παραπάνω κρυμμένα επίπεδα, τα οποία με τη σειρά τους είναι συνδεδεμένα με το επίπεδο εξόδου (output layer). Από το επίπεδο εξόδου γίνεται η εξαγωγή της απάντησης.[27]

Πιο συγκεκριμένα, όσον αφορά τη σύνδεση των μονάδων μεταξύ τους, αυτή χωρίζεται σε δύο βασικές κατηγορίες τεχνητών νευρωνικών δικτύων. Πρόκειται για την πρό-



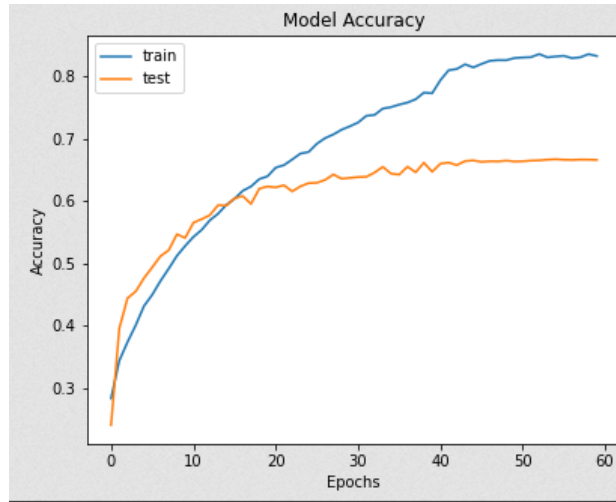
Εικόνα 10. Απλό ΤΝΔ χωρίς bias [22]

σθια τροφοδότηση (feed forward) και την οπίσθια τροφοδότηση (feed backward). Στην περίπτωση των νευρωνικών δικτύων πρόσθιας τροφοδότησης η δομή και οργάνωση των μονάδων γίνεται σε διαφορετικά επίπεδα. Με αυτόν τον τρόπο, οι μονάδες που βρίσκονται στο ένα επίπεδο μπορούν να τροφοδοτήσουν τις μονάδες του επόμενου και ούτω καθεξής. Έτσι, η τροφοδότηση φτάνει κάποια στιγμή και στις μονάδες του τελευταίου επιπέδου και αποκλείεται το ενδεχόμενο εξόδου μονάδας ενός επιπέδου που να γίνεται ταυτόχρονα είσοδος για κάποια άλλη μονάδα προηγούμενου επιπέδου. [27]

2.6.1 Υπερπροσαρμογή - Overfitting

Το πρόβλημα της υπερπροσαρμογής (Overfitting) ή υπερεκπαίδευσης (Overtraining) αναφέρεται στη διαδικασία κατά την οποία το μοντέλο που προκύπτει έχει «φορτωθεί» με ένα συγκεκριμένο dataset εκπαίδευσης και ανταποκρίνεται αποδοτικά μόνο σε αυτό. Η εκπαίδευση του μοντέλου με πολλά δεδομένα δεν εγγυάται, ότι θα παρέχει καλή απόδοση στη πρόβλεψη σε άγνωστα δεδομένα ελέγχου, ακόμη και αν το μοντέλο προβλέπει τέλεια τους στόχους των δεδομένων εκπαίδευσης [22].

Στην εικόνα 11 γίνεται αναπαράσταση ενός μοντέλου, που όπως παρατηρούμε, στην δέκατη πέμπτη περίπου εποχή συμβαίνει το φαινόμενο της υπερεκπαίδευσης. Αυτό που παρατηρούμε είναι ότι η ακρίβεια για τα δεδομένα εκπαίδευσης (μπλε χρώμα), που παρουσιάζονται, αποκλίνουν με τα δεδομένα ελέγχου (πορτοκαλί χρώμα) μετά την εικοστή εποχή. Όπως μπορούμε να παρατηρήσουμε, καθώς φτάνουμε στο τέλος της εκπαίδευσης, (εξηκοστή εποχή) τα δύο γραφήματα αποκλίνουν κατά ένα μεγάλο ποσοστό της τάξης του 20%.



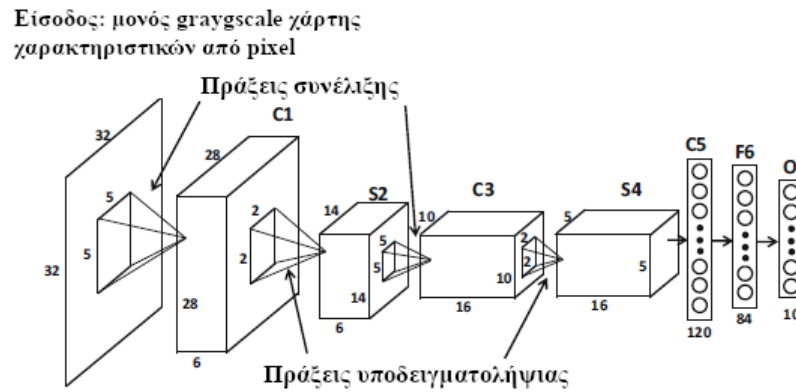
Εικόνα 11. Παράδειγμα Overfitting [28]

2.7 Συνελκτικὰ Νευρωνικά Δίκτυα - Convolutional Neural Networks

Τα συνελκτικὰ δίκτυα (convolutional networks LeCun, 1989), γνωστά και ως συνελκτικὰ νευρωνικά δίκτυα ή CNN (convolutional neural networks), αποτελούν έναν εξειδικευμένο τύπο νευρωνικού δικτύου, οποίος εφαρμόζεται ευρέως σε δεδομένα εικόνας[29] ή άλλα δεδομένα με την μορφή πλέγματος [20]. Τέτοια παραδείγματα περιλαμβάνουν δεδομένα χρονοσειρών, τα οποία μπορούν να θεωρηθούν ως ένα πλέγμα 1D, το οποίο λαμβάνει δείγματα σε τακτά χρονικά διαστήματα, και τα δεδομένα εικόνας, τα οποία μπορούν να θεωρηθούν ως ένα πλέγμα 2D από pixels. Τα συνελκτικὰ δίκτυα έχουν πετύχει εξαιρετικά σε πρακτικές εφαρμογές για το λόγο ότι είναι ιδιαίτερα κατάλληλα για την ταξινόμηση προτύπων[26]. Το όνομα «συνελκτικό νευρωνικό δίκτυο» υποδηλώνει ότι το δίκτυο χρησιμοποιεί μια μαθηματική λειτουργία, που είναι ευρέως γνωστή στον τομέα της ψηφιακής επεξεργασίας σήματος, ονόματι «συνέλιξη» (convolution). Η συνέλιξη είναι ένα εξειδικευμένο είδος γραμμικής λειτουργίας, και ο μαθηματικός τύπος της περιγράφεται στον παρακάτω μαθηματικό τύπο 2.7.

$$y[n] = (x * h)[n] = \sum_k x[k]h[n - k] \quad (2.7)$$

Ένα συνελκτικό δίκτυο είναι χρήσιμο, γιατί η λειτουργία του έχει ως σκοπό την αναγνώριση προτύπων, αλλά ταυτόχρονα έχει υψηλό βαθμό μη-ευαισθησίας στη μετατόπιση,



Εικόνα 12. Μια από τις πρώτες υλοποιήσεις CNN μοντέλων [22]

στην στρέβλωση, στην κλιμάκωση, όπως και σε άλλες μορφές παραμορφώσεις [26]. Αναλυτικότερα, ο κάθε νευρώνας δέχεται τα δεδομένα έχοντας σκοπό να εξάγει κάποιο χαρακτηριστικό. Το κάθε χαρακτηριστικό, αφού εξαχθεί, χάνει τη σημαντικότητα σε επίπεδο τοπολογίας, ωστόσο η θέση του διατηρείται (προσεγγιστικά) με την θέση των άλλων χαρακτηριστικών. Το κάθε επίπεδο αποτελείται από πολλαπλούς χάρτες χαρακτηριστικών (feature map) με το κάθε έναν από αυτούς να είναι στη μορφή ενός επιπέδου.

Στην ορολογία του συνελκτικού δικτύου για το πρώτο επίπεδο το όρισμα, που δέχεται, αναφέρεται ως η είσοδος (input), ενώ για δεύτερο ως ο πυρήνας (kernel). Η έξοδος (output) κάποιες φορές αναφέρεται ως χάρτης χαρακτηριστικών (feature map).

Στις εφαρμογές μηχανικής μάθησης, η είσοδος είναι συνήθως μία πολυδιάστατη διάταξη δεδομένων και ο πυρήνας είναι συνήθως μία πολυδιάστατη διάταξη παραμέτρων, οι οποίες προσαρμόζονται από τον αλγόριθμο εκμάθησης. Γίνεται αναφορά σε αυτές τις πολυδιάστατες διατάξεις ως τανυστές (tensors). Επειδή κάθε στοιχείο της εισόδου και του πυρήνα (βάσης) πρέπει ρητά να αποθηκευτεί χωριστά, συνήθως γίνεται η υπόθεση ότι αυτές οι συναρτήσεις είναι μηδενικές παντού εκτός του πεπερασμένου συνόλου σημείων, για τα οποία οι τιμές αποθηκεύονται. Αυτό σημαίνει, πως στην πράξη μπορεί να εφαρμοστεί το άπειρο άθροισμα ως ένα άθροισμα σε έναν πεπερασμένο αριθμό στοιχείων διάταξης.

[20]

2.8 Αναγνώριση Συναισθήματος

Το πρόβλημα της αναγνώρισης συναισθήματος τα τελευταία χρόνια έχει γνωρίσει μεγάλη ανάπτυξη, προκαλώντας το ενδιαφέρον πολλών κλάδων. Οι ερευνητές προσπαθούν μέσα από την τόνο της φωνής, από αυτά που γράφει ένας χρήστης στα μέσα κοινωνικής δικτύωσης αλλά και από τις εκφράσεις του προσώπου του να προσδιορίσουν το συναίσθημα που νιώθει. Η παρούσα εργασία ερευνά την τελευταία προσέγγιση και πώς θα μπορούσε αυτή η πληροφορία να χρησιμοποιηθεί σε μια εφαρμογή ευρείας χρήσης.

2.8.1 Το πρόβλημα της Αναγνώρισης Συναισθήματος

Αρχικά, για να προσδιορίσουμε το συναίσθημα του ατόμου θα πρέπει να λαμβάνουμε ως είσοδο το πρόσωπο του ατόμου. Η αναγνώριση του προσώπου σε πρώτο στάδιο είναι απαραίτητη και στη συνέχεια αφού έχουμε το πρόσωπο με όλα τα χαρακτηριστικά γίνεται η εξαγωγή των χαρακτηριστικών ξεχωριστά.

Στον τομέα της έρευνας της αναγνώρισης συναισθημάτων προσώπου κυριαρχούν δύο τύποι μεθόδων: οι ολιστικές μέθοδοι και οι αναλυτικές ή τυπικές μέθοδοι. [30] Οι ολιστικές μέθοδοι που προσπαθούν να μοντελοποιήσουν τις παραμορφώσεις του ανθρώπινου προσώπου σε παγκόσμιο επίπεδο, κωδικοποιούν ολόκληρο το πρόσωπο ως σύνολο. Από την άλλη, οι αναλυτικές μέθοδοι παρατηρούν και μετρούν τοπικές ή διακριτές παραμορφώσεις του ανθρώπινου προσώπου, όπως είναι τα μάτια, τα φρύδια, η μύτη, το στόμα κ.λπ. και τις γεωμετρικές τους σχέσεις, προκειμένου να δημιουργήσουν περιγραφικά και εκφραστικά μοντέλα. [31]

2.8.2 Μοντέλα Συναισθημάτων

Ο τρόπος, με τον οποίο τα συναισθήματα αναπαρίστανται, είναι μία βασική πτυχή του συστήματος αναγνώρισης συναισθημάτων. Ένα πολύ δημοφιλές μοντέλο κατηγοριοποίησης, είναι το μοντέλο συναισθημάτων Ekman [32], το οποίο προσδιορίζει έξι βασικά ανθρώπινα συναισθήματα: τον θυμό, την αηδία, τον φόβο, την ευτυχία, την λύπη και την έκπληξη. Το μοντέλο συναισθημάτων Ekman έχει χρησιμοποιηθεί σε πολλές ερευνητικές μελέτες και σε ποικίλα συστήματα που χρησιμοποιούνται για την αναγνώριση της συναισθηματικής κατάστασης από το κείμενο και τις εκφράσεις του προσώπου. Ένα άλλο

δημοφιλές μοντέλο είναι το μοντέλο OCC (Ortony/ Clore/ Collins)[33], το οποίο προσδιορίζει 22 κατηγορίες συναισθημάτων, βασισμένο σε συναισθηματικές αντιδράσεις σε καταστάσεις, και έχει κυρίως σχεδιαστεί για να μοντελοποιεί γενικά τα ανθρώπινα συναισθήματα. Το μοντέλο συναισθημάτων του Plutchik[34], είναι το μοντέλο των διαστάσεων, το οποίο προσφέρει μία ολοκληρωμένη θεωρία, βασισμένη σε εξελικτικές αρχές, και ορίζει οκτώ βασικά διπολικά συναισθήματα. Αυτά τα οκτώ συναισθήματα είναι οργανωμένα σε τέσσερα διπολικά σύνολα: χαρά εναντίον λύπης, θυμός εναντίον φόβου, εμπιστοσύνη εναντίον αηδίας και έκπληξη εναντίον προσμονής.[31]

2.8.3 Εφαρμογές στην Αναγνώριση Συναισθήματος

Ο τομέας της αναγνώρισης προσώπου και πιο συγκεκριμένα του συναισθήματος δεν έχει γνωρίσει τυχαία μεγάλη ανάπτυξη, αφού είναι πολύ σημαντική πληροφορία. Οι ερευνητές, τους οποίους έχει προσεγγίσει είναι από ποικίλους κλάδους, όπως αυτός της ιατρικής, της εκπαίδευσης, της αυτοκινητοβιομηχανίας και της βιομηχανίας ηλεκτρονικών παιχνιδιών.

Αναλυτικότερα, ο τομέας της ιατρικής εκμεταλλεύεται αυτήν την τεχνολογία για την ιατρική περίθαλψη, με λογισμικό αναγνώρισης τεχνητής νοημοσύνης, το οποίο βοηθά στην αξιολόγηση ασθενών και στην απόφαση παροχής ιατρικής βοήθειας σε αυτούς. Οι έρευνες, που γίνονται, αφορούν την κατανόηση ασθενών που έχουν διαγνωστεί με κάποια ψυχική διαταραχή ή ασθένεια [35], [36], [37]. Επίσης, άλλες έρευνες έχουν σκοπό να βοηθήσουν παιδιά, που βρίσκονται στο φάσμα του αυτισμού [38], να κατανοήσουν τα δικά τους συναισθήματα και των άλλων.

Στην βιομηχανία των βιντεοπαιχνιδιών χρησιμοποιούνται για την ανάλυση του συναισθήματος. Τα βιντεοπαιχνίδια σχεδιάζονται έχοντας έναν συγκεκριμένο στόχο για το κοινό στο νου και επιθυμούν να προκαλέσουν μία συγκεκριμένη συμπεριφορά και ένα σύνολο συναισθημάτων στους χρήστες. Κατά το στάδιο της δοκιμής, οι χρήστες καλούνται να παίξουν το παιχνίδι για μία δεδομένη περίοδο και η ανατροφοδότηση που δίνουν, τα σχόλιά τους, λαμβάνονται υπόψιν για την κατασκευή του τελικού προϊόντος[39].

Η αυτοκινητοβιομηχανία εφαρμόζει, επίσης, τεχνολογία αναγνώρισης συναισθημάτων καθώς οι κατασκευαστές αυτοκινήτων σε όλο τον κόσμο εστιάζουν όλο και περισσότερο στο να κάνουν τα αυτοκίνητα πιο προσωπικά και πιο ασφαλή κατά την οδήγηση.

Τα έξυπνα αυτοκίνητα, τα οποία μπορούν να ανιχνεύσουν συναισθήματα του προσώπου, μπορούν να ειδοποιούν τον οδηγό, όταν αυτός βρίσκεται σε κίνδυνο και να βοηθήσουν στη μείωση των τροχαίων ατυχημάτων. Ένα τέτοιο αυτοκίνητο θα μπορούσε να παρέχει στο οδηγό συμβουλές χαλάρωσης σε μία έντονη συναισθηματική κατάσταση ή ακόμα και να τον αποτρέπει να οδηγήσει την συγκεκριμένη στιγμή[40].

Κεφάλαιο 3

Μεθοδολογία Έρευνας

Στο παρόν κεφάλαιο παρουσιάζεται η πορεία που ακολουθήθηκε για την κατασκευή του τελικού μοντέλου και του συνεργαζόμενου προγράμματος. Επίσης, αναφέρονται οι στόχοι που αναμένεται να επιτευχθούν για την λειτουργικότητα των παραπάνω δύο υλοποιήσεων.

3.1 Πορεία Έρευνας & Υλοποίησης

Αρχικά, μελετήθηκε η βιβλιογραφία ανάλογων επιστημονικών άρθρων και βιβλίων που πραγματεύονται θέματα βαθιάς μάθησης σε προβλήματα αναγνώρισης αντικειμένων. Με βάση αυτά, ο κατάλληλος τύπος αρχιτεκτονικής του νευρωνικού δικτύου που είναι αποδοτικός σε δεδομένα εικόνας, είναι αυτός του Συνελικτικού Νευρωνικού Δικτύου (CNN). Στην συνέχεια, έγινε έρευνα για τους διάφορους τύπους τέτοιων δικτύων και συγκεκριμένα στην ανάλυση του συναισθήματος. Τα δεδομένα που επιλέχθηκαν αποτελούν ένα αρκετά δημοφιλές dataset[41], όπως θα αναλύσουμε στο κεφάλαιο 4, που χρησιμοποιείται για μελέτη σε πολλά άρθρα κάθε χρόνο.

Κατασκευάστηκαν μοντέλα τοπικά σε προσωπικό υπολογιστή για να ελεγχθεί ο χρόνος και η ποσότητα υπολογιστικής ισχύος που θα χρειαζόταν το μοντέλο. Συγχρόνως, γινόταν παραμετροποίηση του αλγόριθμου εκπαίδευσης και προσπάθειες βελτιστοποίησης των δεδομένων τροφοδότησης του μοντέλου. Τα πρώτα δοκιμαστικά μοντέλα που υλοποιήθηκαν θα μπορούσαν να χαρακτηριστούν ελαφρώς υπερπροσαρμοσμένα (overfitted) στα δεδομένα εκπαίδευσης. Επομένως, έγιναν βελτιστοποιήσεις τόσο στα δεδομένα εκπαίδευ-

σης, όσο και στην αρχιτεκτονική του αλγορίθμου που χρησιμοποιήθηκε. Αναλυτικότερα, για τον αλγόριθμο δοκιμάστηκαν υλοποιήσεις, διαφορετικών επιπέδων με διαφορετική πολυπλοκότητα και διαφορετικών πυρήνων στα επίπεδα του δικτύου.

Η εγκυρότητα του μοντέλου ελέγχθηκε αρχικά τροφοδοτώντας το, με μία εικόνα προσώπου και τυπώνοντας στην κονσόλα το κυρίαρχο συναίσθημα. Σε επόμενο στάδιο υλοποιήθηκε πρόγραμμα, που κάνει έλεγχο του μοντέλου μέσω ζωντανής ροής βίντεο από μια απλή κάμερα υπολογιστή.

Στην συνέχεια, η ανάλυση της εικόνας και η καταγραφή του συναισθήματος σε πραγματικό χρόνο ενσωματώθηκαν σε μια διαδικτυακή εφαρμογή. Σκοπός αυτής της εφαρμογής θα είναι να παρέχει στο χρήστη κάποια πληροφορία για τις εικόνες ή βίντεο τα οποία την τροφοδοτεί. Ιδανικά θα προσπαθήσουμε να επιτύχουμε κάποια αλληλεπίδραση της εφαρμογής με τον χρήστη, για να μπορεί αυτή να χρησιμοποιηθεί σε καθημερινές καταστάσεις.

Το μοντέλο, στο τέλος, ενσωματώθηκε σε μια διαδικτυακή εφαρμογή, η οποία με τις κατάλληλες αλλαγές εμπεριέχει και τις δύο προαναφερθείσες λειτουργίες. Η εφαρμογή υλοποιήθηκε και τρέχει τοπικά στον browser¹ της επιθυμίας του χρήστη.

3.2 Περιορισμοί & Λύσεις Υλοποίησης

Η πολυπλοκότητα του αλγορίθμου εκπαίδευσης μεγάλωνε καθώς προσπαθούσαμε να αυξήσουμε την ακρίβεια του μοντέλου. Επίσης, τα δεδομένα εκπαίδευσης αποτελούνται από χιλιάδες εγγραφές. Επομένως, ο χρόνος της υλοποίησης του μοντέλου και η επεξεργαστική ισχύς, που απαιτούνταν κάθε φορά, είχαν αυξηθεί αρκετά. Προς επίλυση αυτού του προβλήματος χρησιμοποιήθηκε μια υπηρεσία νέφους από την Google που μας παρέχει επεξεργαστική ισχύ κάρτας γραφικών(GPU), κατάλληλη για τέτοιους αλγόριθμους μηχανικής μάθησης. Αφού κατασκευάστηκε το μοντέλο, ο έλεγχός του έγινε πρώτα με εικόνες προσώπων και έπειτα με τροφοδότηση ζωντανής ροής βίντεο. Προφανώς, επειδή το μοντέλο αυτό εκπαιδεύτηκε και κατασκευάστηκε για να ανιχνεύει το συναίσθημα στο πρόσωπο αλλά όχι το ίδιο το πρόσωπο, έγινε χρήση ενός ήδη εκπαιδευμένου μοντέλου για την αναγνώριση προσώπου. Επίσης, το μοντέλο, όπως θα παρατηρήσουμε και έπειτα

¹φυλλομετρητή

από κάποιο αριθμό εποχών, δύσκολα αποκτά μεγαλύτερη ακρίβεια.

Η υλοποίηση της διαδικτυακής εφαρμογής υλοποιήθηκε θέτοντας κάποιους περιορισμούς στην χρήση και στις λειτουργίες της. Αυτό οφείλεται στις περιορισμένες δυνατότητες των βιβλιοθηκών, οι οποίες ακόμα βρίσκονται σε στάδιο ανάπτυξης και δεν παρέχουν μεγάλη ευελιξία.

3.3 Στόχοι της Έρευνας

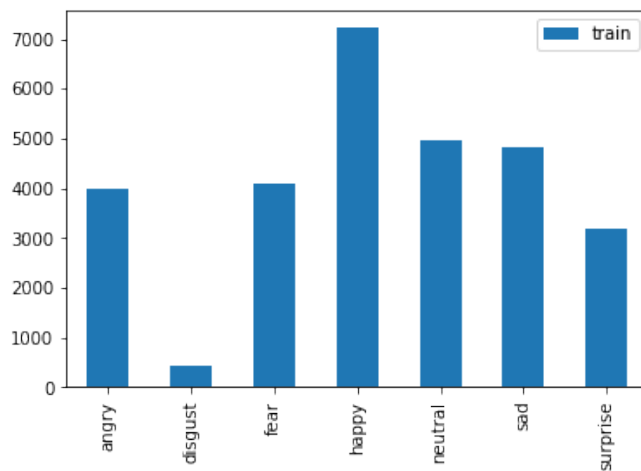
Αρχικό στόχο της έρευνας αποτελεί η κατανόηση των μεθόδων και εννοιών της βαθιάς μάθησης στο πεδίο της αναγνώρισης αντικειμένων και συγκεκριμένα στην ανάλυση συναισθήματος από κάποιο αυτοματοποιημένο λογισμικό. Στο λογισμικό αυτό θέλουμε να πετύχουμε ακρίβεια της τάξης 70% και να δοθούν λύσεις σε παρόμοια προβλήματα αναγνώρισης συναισθήματος. Τέλος, σκοπός αυτού του εγχειρήματος είναι να εκμεταλλευτούμε αυτήν τη σημαντική πληροφορία του συναισθήματος και να εξετάσουμε αν αυτή πληροφορία μπορεί αξιοποιηθεί σε κάποια εφαρμογή στην καθημερινή μας ζωή.

Κεφάλαιο 4

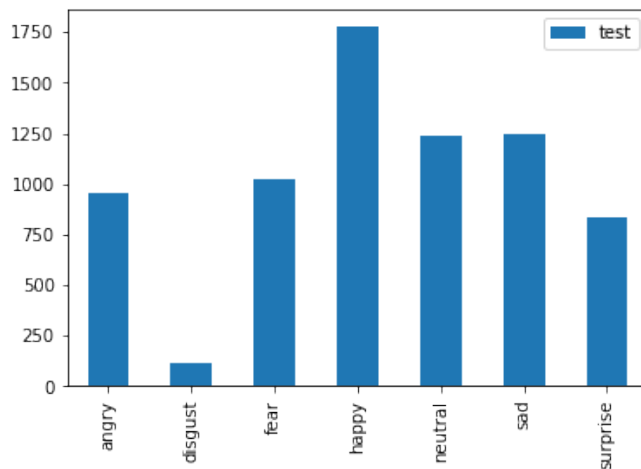
Αρχιτεκτονική

4.1 Dataset

Για την εκπαίδευση του μοντέλου της παρούσας εργασίας χρησιμοποιήθηκε το FER-13[41]. Το παρόν dataset αποτελείται από 35.887 φωτογραφίες, που απεικονίζουν μόνο το πρόσωπο και είναι όλες κεντραρισμένες με το πρόσωπο να καταλαμβάνει τον ίδιο χώρο σε κάθε φωτογραφία. Τα παραδείγματα φωτογραφιών έχουν μέγεθος 48x48 pixels και είναι ασπρόμαυρες (grayscale). Τα πρόσωπα που περιέχει το dataset περιγράφουν ένα από τα επτά συναισθήματα (θυμός, αηδία, φόβος, χαρά, λύπη, έκπληξη, ουδέτερο). Το κάθε dataset χωρίζεται πάντα σε ένα μέρος που είναι υπεύθυνο για την εκπαίδευση του μοντέλου και ένα για την εγκυρότητα του μοντέλου, που προέκυψε από την διαδικασία της εκπαίδευσης. Συνήθως τα δεδομένα έλεγχου - εκπαίδευσης χαρακτηρίζονται από μία συγκεκριμένη αναλογία της τάξης 20% - 80% αντίστοιχα. Επομένως, έχουμε 28,709 εικόνες για την εκπαίδευση του μοντέλου και 7,178 εικόνες για τον έλεγχο του. Στους πίνακες γραφημάτων 13 & 14 παρουσιάζεται η αναλογία των εικόνων που έχουμε για το κάθε συναίσθημα. Παρατηρείται, ότι στο dataset τα παραδείγματα δεν είναι ομοιόμορφα κατανεμημένα. Για παράδειγμα, στα δεδομένα εκπαίδευσης υπάρχουν 7215 εικόνες, που απεικονίζουν την χαρά αλλά μόλις 436, που δείχνουν το συναίσθημα της αηδίας.



Εικόνα 13. Δεδομένα Εκπαίδευσης



Εικόνα 14. Δεδομένα Ελέγχου

Παρακάτω γίνεται η περιγραφή των συναισθημάτων σε σχέση με τις ανάλογες κινήσεις του προσώπου. Όπως παρατηρούμε, τα μάτια, καθώς και τα υπόλοιπα χαρακτηριστικά γύρω από το μάτι, έχουν σημαντικό ρόλο στον καθορισμό του συναισθήματος.

Πίνακας 2. Αναλυτικός Πίνακας με των εικόνων του dataset

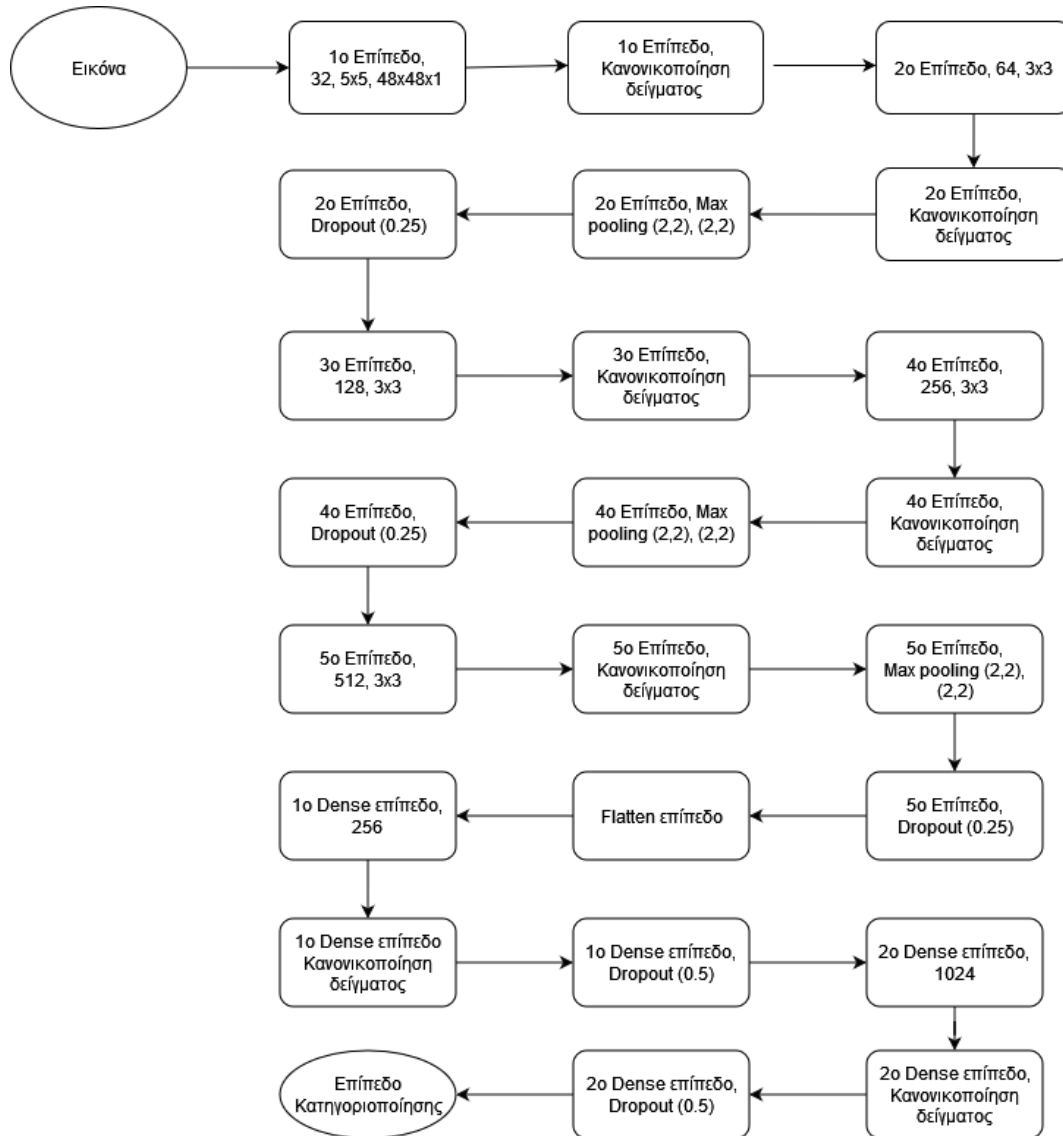
set	θυμός	αηδία	φόβος	χαρά	λύπη	έκπληξη	ουδέτερο
train	3995	436	4097	7215	4965	4830	3171
test	958	111	1024	1774	1233	1247	831

Πίνακας 3. Σχέση συναισθήματος με ανάλογη κίνηση [42]

Συναίσθημα	Κίνηση μελών του προσώπου
Χαρά	Άνοιγμα ματιών, άνοιγμα στόματος, τραβηγμένη γωνία των χειλιών, ανασηκωμένα μάγουλα
Λύπη	Κατεβασμένο εξωτερικό φρύδι, ανασηκωμένα εσωτερικά φρύδια, κλειστά μάτια, κατεβασμένη γωνία των χειλιών
Έκπληξη	Ανασηκωμένα φρύδια, ανοιχτά μάτια, πεσμένο σαγόι
Θυμός	Κατεβασμένα φρύδια, ανοιχτά μάτια, σφιγμένα χείλη
Φόβος	Κατεβασμένο εξωτερικό φρύδι, σηκωμένο εσωτερικό φρύδι, ανοιχτό στόμα
Αηδία	Γωνιακό πιεστικό χειλιών, πιεστήριο κάτω χειλιών, κατεβασμένα φρύδια, ζαρωμένη μύτη

4.2 Αρχιτεκτονική Μοντέλου

Το μοντέλο που κατασκευάστηκε και εκπαιδεύτηκε με το παραπάνω dataset αποτελείται από πέντε επίπεδα συνέλιξης. Ανάμεσα σε αυτά υπάρχουν άλλα αντίστοιχα επίπεδα υποδειγματοληψίας και κανονικοποίησης του δείγματος. Τέλος, υπάρχουν το επίπεδο flatten και δύο πλήρως συνδεδεμένα επίπεδα (dense layer).



Εικόνα 15. Αρχιτεκτονική Μοντέλου

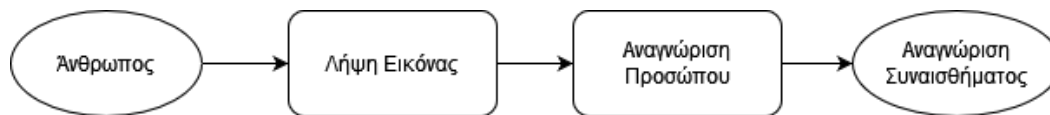
4.3 Λειτουργικότητα

4.3.1 Λογισμικό

Η λειτουργικότητα του λογισμικού που υλοποιήθηκε χωρίζεται σε δύο μέρη:

1. Αναγνώριση Συναισθήματος από εικόνα
2. Αναγνώριση Συναισθήματος σε ζωντανή ροή εικόνας

Η πρώτη περίπτωση, όπως θα δούμε και στο κεφάλαιο 6, αναφέρεται σε εικόνες που απεικονίζουν έναν άνθρωπο με το πρόσωπο να είναι στραμμένο προς την κάμερα. Αυτό που έχει σημασία είναι τα βασικά χαρακτηριστικά του προσώπου (μάτια, στόμα, μύτη) να είναι ευδιάκριτα, προκειμένου να αναγνωρίζεται το πρόσωπο. Ο φάκελος με τις εικόνες φορτώνεται σε ένα υποπρόγραμμα και αυτό δημιουργεί ένα άλλο φάκελο, όπου εκεί αποθηκεύεται μόνο το πρόσωπο του ανθρώπου στην φωτογραφία. Στην συνέχεια, φορτώνεται ο νέος φάκελος στο μοντέλο αναγνώρισης συναισθήματος και παρουσιάζεται η αρχική εικόνα που δέχτηκε το πρόγραμμα και μια γραφική παράσταση σχετική με το ποια συναισθήματα ανιχνεύτηκαν.



Εικόνα 16. Διάγραμμα ροής για την 1η λειτουργία

Στην δεύτερη περίπτωση, το μοντέλο φορτώνεται και ταυτόχρονα υπάρχει ένα πρόγραμμα, το οποίο δέχεται ως είσοδο το βίντεο μέσω της κάμερας του υπολογιστή. Κάθε καρέ από το βίντεο περνάει από αναγνώριση προσώπου και σχηματίζεται ένα σχήμα τετραγώνου ως ένδειξη ότι το πρόσωπο αναγνωρίστηκε. Στην οθόνη πάνω από την ένδειξη αυτή παρουσιάζεται το κυρίαρχο συναίσθημα, το οποίο υπερίσχυσε έναντι των άλλων συναισθημάτων, μαζί με το ποσοστό του.

Στην συνέχεια, υλοποιήθηκε ένα υποπρόγραμμα που τρέχει με την μορφή νήματος (thread). Ο ρόλος του συγκεκριμένου είναι να εξάγει την πληροφορία από την κάμερα για κάθε πέντε δευτερόλεπτα και να παρέχει κατάλληλα μηνύματα. Αναλυτικότερα, για κάθε καρέ που περνάει από την αναγνώριση του συναισθήματος γίνεται η αποθήκευση του συναισθήματος σε μια λίστα. Έπειτα, όταν η λίστα αυτή φτάσει να έχει πέντε εγγραφές,



Εικόνα 17. Διάγραμμα ροής για την 2η λειτουργία

καλείται μια άλλη συνάρτηση. Αυτή με την σειρά της υπολογίζει, εάν υπάρχει κάποια εγγραφή που εμφανίζεται πάνω από τρεις φορές.

Επομένως, με αυτό τον τρόπο το πρόγραμμα, στην ουσία, λαμβάνει υπόψη του τα προηγούμενα πέντε καρέ, που δέχθηκε, και βρίσκει αυτό που υπερτερεί. Η παραγωγή ανάλογων μηνυμάτων στο χρήστη γίνεται από το νήμα, αφού παίζει το ρόλο του χρονομέτρου και για κάθε πέντε δευτερόλεπτα καλείται η συνάρτηση που τυπώνει το μήνυμα στο χρήστη.

Τέλος, οι δύο παραπάνω λειτουργίες ενσωματώθηκαν σε μια δικτυακή εφαρμογή, τροποποιώντας κατάλληλα το λογισμικό για την ομαλή λειτουργία της εφαρμογής.

4.3.2 Διαδικτυακή εφαρμογή

Στη διαδικτυακή εφαρμογή έχουν ενσωματωθεί οι δύο παραπάνω λειτουργίες. Στην αρχική σελίδα της εφαρμογής ο χρήστης πληροφορείται σχετικά με τις λειτουργίες της. Μπορεί να πλοηγηθεί από ένα μενού, το οποίο βρίσκεται στα αριστερά, και να επιλέξει ανάμεσα στις δύο λειτουργίες και σε μια σελίδα που παρέχει πληροφορίες για την εφαρμογή.

Η ανάλυση εικόνας γίνεται, αφού ο χρήστης επιλέξει φάκελο και την εικόνα που θέλει να αναλύσει. Μετά την επιλογή, αυτή εμφανίζεται με μια ένδειξη που δείχνει, εάν αναγνωρίστηκε κάποιο πρόσωπο. Ταυτόχρονα γίνεται η επεξεργασία της εικόνας και δεξιά της θα εμφανιστεί μια γραφική παράσταση που περιγράφει τα συναισθήματα που υπάρχουν στο πρόσωπο της φωτογραφίας.

Στην περίπτωση που ο χρήστης επιλέξει να γίνει αναγνώριση συναισθήματος μέσω βίντεο, φορτώνουν οι ανάλογες επιλογές. Το πρώτο που θα πρέπει να κάνει είναι να επιλέξει

μία από τις διαθέσιμες κάμερες που βρίσκονται στον υπολογιστή. Έπειτα, μπορεί να εκκινήσει την κάμερα και το πρόγραμμα θα αρχίσει να φορτώνει. Τα συναισθήματα που θα αναγνωριστούν θα εμφανιστούν, όπως στο λογισμικό που υλοποιήθηκε. Τέλος, το νήμα θα παράγει στην κονσόλα τα κατάλληλα μηνύματα κάθε πέντε δευτερόλεπτα, ανάλογα με τα συναισθήματα που ανιχνεύονται.

Κεφάλαιο 5

Σχεδιασμός & Υλοποίηση λογισμικού

Η ανάπτυξη του μοντέλου, η διαχείριση των δεδομένων φόρτωσής του αλλά και ο έλεγχος του υλοποιήθηκαν σε γλώσσα Python 3.

5.1 Περιβάλλοντα ανάπτυξης και Βιβλιοθήκες

5.1.1 Tensorflow - Keras

Η εκπαίδευση μεγάλων μοντέλων και η τροφοδότησή τους με μεγάλο όγκο δεδομένων οδήγησε στην υλοποίηση της βιβλιοθήκης Tensorflow από την Google. [43] Ο σκοπός ήταν να υπάρχει μια βιβλιοθήκη που αξιοποιεί κατάλληλα τους υπολογιστικούς πόρους και παρέχει πολλές δυνατότητες. Το Tensorflow είναι μια δωρεάν βιβλιοθήκη ανοικτού κώδικα που εξειδικεύεται σε θέματα τεχνητής νοημοσύνης και μηχανικής μάθησης. Επίσης, μέσω του API¹ του, επικοινωνεί με μια άλλη βιβλιοθήκη το Keras. Το Keras είναι φιλικό προς τους αρχάριους στα θέματα της μηχανικής μάθησης παρέχοντας, όμως, όλες τις δυνατότητες για την κατασκευή των νευρωνικών δικτύων.[44] Οι δύο αυτές συνεργαζόμενες βιβλιοθήκες βοηθούν στην κατασκευή του μοντέλου, την φόρτωση δεδομένων σε αυτό και στη συνέχεια στην εκπαίδευσή του. Στην συγκεκριμένη εργασία χρησιμοποιήσαμε μία από τις τελευταίες εκδόσεις της βιβλιοθήκης (2.8.0).

¹Application Programming Interface

5.1.2 OpenCV

Η OpenCV είναι και αυτή μια βιβλιοθήκη ανοιχτού κώδικα και ειδικεύεται σε θέματα μηχανικής όρασης ή αλλιώς υπολογιστικής όρασης (computer vision).[45] Χρησιμοποιείται κατά κόρον σε θέματα ρομποτικής, επαυξημένης πραγματικότητας, αυτόματης ανίχνευσης αντικειμένων, χειρισμού μη επανδρωμένων σκαφών (drones) και άλλα. Λόγω της χρήσης της σε θέματα μηχανικής μάθησης παρέχει εργαλεία και αλγορίθμους για στατιστική ανάλυση που βοηθούν στην κατασκευή, στην υλοποίηση και στον έλεγχο των παραπάνω κλάδων. Επιπλέον, ο στόχος της, προφανώς, είναι να παρέχει τα κατάλληλα εργαλεία στον προγραμματιστή για την ανάγνωση εικόνων ή ζωντανής ροής βίντεο από κάμερα και να τα τροφοδοτεί στο μοντέλο.

5.1.3 Google Colab

Το Google Colab ή απλά Colab είναι ένα προϊόν της Google Research.[46] Το Colab είναι μια δωρεάν πλατφόρμα, που βασίζεται στις αρχές της υπολογιστικής νέφους προσφέροντας υπολογιστή ισχύ και μια υπηρεσία Jupyter Notebook. Στο Notebook παρέχεται η δυνατότητα υλοποίησης προγραμμάτων σε Python και η χρησιμοποίηση υπολογιστικής ισχύος από πολύ ισχυρές κάρτες γραφικών. Επίσης, εκτός από αυτές τις κάρτες γραφικών υπάρχει και η δυνατότητα ενός τύπου κάρτας, που αναφέρεται ως TPU², η οποία είναι εξειδικευμένη για τη χρησιμοποίηση της βιβλιοθήκης Tensorflow και για τους υπολογισμούς στα νευρωνικά δίκτυα. Τα εργαλεία αυτά, μας δίνουν την δυνατότητα μιας ασφαλούς κατασκευής μοντέλου για την εργασία μας αλλά, παρέχοντας μας ισχυρή υπολογιστική ισχύ, μας διευκολύνουν στην υλοποίηση του μοντέλου που θα παραχθεί.

5.1.4 NumPy

Το NumPy είναι μια απαραίτητη βιβλιοθήκη για την παρούσα εργασία καθώς είναι κατάλληλη για τον υπολογισμό μαθηματικών πράξεων. Πιο συγκεκριμένα, ειδικεύεται στη διαχείριση και στον υπολογισμό αριθμητικών πολυδιάστατων πινάκων.[47] Ένας τέτοιος πίνακας μπορεί να θεωρηθεί μία εικόνα, με την οποία τροφοδοτείται το μοντέλο εκπαίδευσης. Με τη χρήση αυτής της βιβλιοθήκης προ-επεξεργαζόμαστε τα δεδομένα μας με

²Tensor Processing Unit

τις κατάλληλες μαθηματικές συναρτήσεις και έπειτα τα εισάγουμε στο νευρωνικό δίκτυο.

5.1.5 Matplotlib

Η Matplotlib είναι μια βιβλιοθήκη που παρέχει δυνατότητες για παρουσίαση δεδομένων με την δημιουργία διαγραμμάτων. Στη συγκεκριμένη εργασία χρησιμοποιήθηκε, προκειμένου να συγκριθεί η ακρίβεια του μοντέλου στα δεδομένα εκπαίδευσης με τα δεδομένα ελέγχου.

5.1.6 Streamlit

Το Streamlit είναι μια ανοιχτού κώδικα βιβλιοθήκη κατάλληλη για την δημιουργία διαδικτυακών εφαρμογών. Προσφέρει μεγάλη ευκολία στην υλοποίηση διαδικτυακών εφαρμογών, που ενσωματώνουν τη λειτουργία εκπαιδευμένων μοντέλων μηχανικής μάθησης. Επίσης, παρέχει πολλαπλά εργαλεία για την παρουσίαση δεδομένων και την παραμετροποίησή τους σε πραγματικό χρόνο από τον τελικό χρήστη.

5.2 Υλοποίηση Μοντέλου

Η ανάπτυξη του συγκεκριμένου δικτύου έγινε σε Tensorflow και Keras. Σε ένα πρώτο στάδιο γίνεται ο προσδιορισμός του αριθμού των συναισθημάτων και η αρχικοποίηση τους. Δηλώνονται, επίσης, οι 200 εποχές που θα περάσει το μοντέλο.

```
num_classes = 7
epochs = 200
model = tf.keras.models.Sequential()
```

Εικόνα 18. Αρχικοποίηση παραμέτρων

Στην εικόνα 19 παρουσιάζεται το πρώτο επίπεδο. Η κατασκευή του πρώτου επιπέδου του νευρωνικού δικτύου, αποτελείται από ένα επίπεδο συνέλιξης και ένα επίπεδο κανονικοποίησης. Το πρώτο είναι ένα διδιάστατο (2D) επίπεδο συνέλιξης και περιλαμβάνει 32 φίλτρα με πυρήνα διαστάσεων 5x5. Ο αριθμός των φίλτρων δείχνει πόσοι πυρήνες θα χρησιμοποιηθούν για να υλοποιήσουν την πράξη της συνέλιξης με την εικόνα. Τις περισσότερες φορές, με βάση την αρχιτεκτονική του δικτύου, στα πρώτα επίπεδα των

συνελκτικών νευρωνικών δικτύων συναντώνται λιγότερα φίλτρα σε αριθμό από ότι στα βαθύτερα επίπεδα. Έπειτα, καλείται η «ReLU»³ ως συνάρτηση ενεργοποίησης. Η παράμετρος `padding` αναφέρεται στο συμπλήρωμα της εικόνας εισόδου περιμετρικά με κάποιο αριθμό εικονοστοιχείων. Ακόμη, στο πρώτο επίπεδο του δικτύου και μόνο σε αυτό, θα πρέπει να δηλωθεί η παράμετρος `input_space`, η οποία είναι απαραίτητη για να δηλωθούν οι διαστάσεις των δεδομένων της εισόδου. Επομένως, η κάθε εικόνα έχει συγκεκριμένες διαστάσεις, άρα δηλώνεται στην προκειμένη περίπτωση `48x48x1`. Τα 48 δηλώνουν το ύψος και το πλάτος της εικόνα και το 1 δηλώνει ότι έχει ένα κανάλι χρώματος, δηλαδή είναι ασπρόμαυρη.

Στην συνέχεια του πρώτου επιπέδου βρίσκεται ένα επίπεδο κανονικοποίησης παρτίδας (Batch Normalization Layer). Το επίπεδο αυτό χρησιμοποιείται ανάμεσα στα κρυφά επίπεδα για να αποτρέψει την ανισορροπία στα δεδομένα και να τα υποβαθμίσει σε μια χαμηλότερη κλίμακα. Επίσης, το επίπεδο αυτό είναι γνωστό για την μείωση χρόνου της εκπαίδευσης του μοντέλου. [22] Επίσης, το επίπεδο αυτό χρησιμοποιήθηκε γιατί αποτρέπει την υπερπροσαρμοστικότητα του μοντέλου στα δεδομένα εκπαίδευσης. [19]

```
#1st CNN layer
model.add(Conv2D(32,(5,5), activation='relu',
                padding="same",
                input_shape=(48,48,1)))
model.add(BatchNormalization())
```

Εικόνα 19. Πρώτο επίπεδο νευρωνικού συνελκτικού δικτύου

Το δεύτερο επίπεδο (Εικόνα 20) συνέλιξης του νευρωνικού δικτύου αποτελείται από 64 φίλτρα και πυρήνες διαστάσεων `3x3`. Το πρώτο 2D max pooling επίπεδο έχει μέγεθος `2x2` και βηματισμό (strides) επίσης `2x2`. Το max pooling επίπεδο ή αλλιώς επίπεδο μέγιστης ομαδοποίησης είναι υπεύθυνο για την υποδειγματοληψία στην εικόνα. Αυτή πραγματοποιείται μέσω της κατασκευής ενός «πλέγματος», το οποίο έχει μέγεθος `2x2` και λαμβάνει την μεγαλύτερη κάθε φορά τιμή των χαρακτηριστικών που περιέχει η εικόνα (feature map). Αυτό το «πλέγμα» ολισθαίνει κατά `2x2` κάθε φορά έως ότου να διατρέξει ολόκληρη την εικόνα και να εξάγει τις μέγιστες τιμές. Σε αυτό το επίπεδο η μέθοδος dropout (διάλειψη) αφαιρεί ένα ποσοστό των νευρώνων που ανέρχεται στο 25% κάθε φορά. Η

³Rectified Linear Unit

αφαίρεση γίνεται τυχαία και χρησιμοποιούμε αυτήν την παράμετρο για να αποτρέψουμε το μοντέλο από το overfitting⁴.

```
#2nd CNN layer
model.add(Conv2D(64,(3,3), activation='relu',padding="same"))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2), strides=(2, 2)))
model.add(Dropout(0.25))
```

Εικόνα 20. Δεύτερο επίπεδο νευρωνικού συνελκτικού δικτύου

Το τρίτο (Εικόνα 21) περιέχει αριθμό φίλτρων που ανέρχεται στα 128 και πυρήνες διαστάσεων 3x3. Παρόμοιες είναι και οι υπόλοιπες παράμετροι.

```
#3rd CNN layer
model.add(Conv2D(128,(3,3), activation='relu',padding="same"))
model.add(BatchNormalization())
```

Εικόνα 21. Τρίτο επίπεδο νευρωνικού συνελκτικού δικτύου

Το τέταρτο επίπεδο (Εικόνα 22) συνέλιξης ακολουθείται από ένα επίπεδο υποδειγματοληψίας. Η πράξη της συνέλιξης γίνεται με αριθμό φίλτρων που ανέρχεται στα 256 και πυρήνες διαστάσεων 3x3. Παρατηρούμε, ότι σημειώνεται μία σημαντική διαφορά ανάμεσα στα επίπεδα, καθώς ο αριθμός των φίλτρων μεγαλώνει. Ωστόσο ακόμη κι εδώ γίνεται χρήση της συνάρτησης ενεργοποίησης «relu», όπως και σε όλα τα συνελκτικά επίπεδα. Το max pooling επίπεδο έχει διαστάσεις 2x2 όπως και στο προηγούμενο, ο βηματισμός έχει μέγεθος 2x2, ενώ με το dropout αφαιρεί το 25% της εισόδου.

```
#4th CNN layer
model.add(Conv2D(256,(3,3), activation='relu',padding="same"))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2), strides=(2, 2)))
model.add(Dropout(0.25))
```

Εικόνα 22. Τέταρτο επίπεδο νευρωνικού συνελκτικού δικτύου

Το πέμπτο επίπεδο (Εικόνα 23) απαρτίζεται από ένα 2D επίπεδο συνέλιξης, ένα κανονικοποίησης και ένα υποδειγματοληψίας. Το συγκεκριμένο επίπεδο συνέλιξης περιλαμ-

⁴υπερεκπαίδευση

βάνει 512 φίλτρα και πυρήνες διαστάσεων 3x3, κάνει χρήση της συνάρτησης ενεργοποίησης «relu» και το max pooling επίπεδο έχει μέγεθος 2x2 και βηματισμό 2x2, όπως και στις προηγούμενες περιπτώσεις που αναφέρθηκαν. Πάλι εδώ το dropout αφαιρεί το ένα τέταρτο της εισόδου.

```
#5th CNN layer
model.add(Conv2D(512,(3,3), activation='relu',padding="same"))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2,2), strides=(2, 2)))
model.add(Dropout(0.25))
```

Εικόνα 23. Πέμπτο επίπεδο νευρωνικού συνελκτικού δικτύου

Με το επίπεδο Flatten (Εικόνα 24) , τα δεδομένα του δικτύου αυτομάτως μετατρέπονται σε μονοδιάστατο πίνακα. Με αυτόν τον τρόπο γίνεται εφικτή η μεταφορά τους στο επόμενο επίπεδο του νευρωνικού συνελκτικού δικτύου. Είναι μία διαδικασία απαραίτητη, η οποία πρέπει να γίνει προκειμένου να μπορούν τα δεδομένα να εισέλθουν στο πλήρως συνδεδεμένο δίκτυο.

```
model.add(Flatten())
```

Εικόνα 24. Μετατροπή των δεδομένων σε μονοδιάστατο πίνακα

Όπως φαίνεται και στις παρακάτω εικόνες, τα δύο επίπεδα είναι πλήρως συνδεδεμένα. Χρησιμοποιώντας τη συνάρτηση Dense, γίνεται αντιληπτό ότι η σύνδεση των νευρώνων είναι πυκνή. Αυτό σημαίνει πως όλοι οι νευρώνες του πρώτου επιπέδου Dense δέχονται σαν την είσοδο, την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου (flatten layer). Στην συνέχεια, με την ίδια διαδικασία το δεύτερο επίπεδο Dense, στον κάθε νευρώνα του δέχεται όλες τις εξόδους από το πρώτο Dense επίπεδο. Συγκεκριμένα οι αριθμοί 256 και 1024 δείχνουν τον αριθμό των νευρώνων που υπάρχουν στο πρώτο και δεύτερο επίπεδο αντίστοιχα. Χρησιμοποιείται ξανά η συνάρτηση ενεργοποίησης «relu» και το dropout αφαιρεί τυχαία ένα ποσοστό νευρώνων. Όμως, αυτή την φορά και στα δύο επίπεδα αφαιρείται το 50%.

```
#Fully connected 1st layer
model.add(Dense(256,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))
```

Εικόνα 25. Πλήρης σύνδεση των νευρωνικών συνελκτικών δικτύων

```
# Fully connected layer 2nd layer
model.add(Dense(1024,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))
```

Εικόνα 26. Πλήρης σύνδεση των νευρωνικών συνελκτικών δικτύων

Τέλος, το επίπεδο κατηγοριοποίησης (classification layer) περιλαμβάνει τόσους νευρώνες όσες και οι κατηγορίες των συναισθημάτων. Ως συνάρτηση ενεργοποίησης ορίζεται η «softmax», η οποία αποτελεί μία γενίκευση της λογιστικής συνάρτησης.

```
model.add(Dense(num_classes,activation="softmax"))
```

Εικόνα 27. Επίπεδο Κατηγοριοποίησης

Ακολουθεί ο ορισμός των παραμέτρων της εκπαίδευσης του μοντέλου, όπως φαίνεται και στην [28](#).

- Βελτιστοποιητής (Optimizer): Ορίζουμε ως βελτιστοποιητή τον αλγόριθμο Adam. Αυτός ο αλγόριθμος είναι κατάλληλος, γιατί είναι από τους πιο αποδοτικούς αλγόριθμους σε μεγάλο όγκο δεδομένων.
- Ρυθμός εκμάθησης (Learning Rate): Ο ρυθμός αυτός αναφέρεται στην διαδικασία εκπαίδευσης του μοντέλου. Συχνά ο βηματισμός αυτός σημειώνεται ως υπερ-παράμετρος, αφού καθορίζει τις υπόλοιπες παραμέτρους που μαθαίνει το μοντέλο. Ο ρυθμός αυτός πρέπει να είναι κατάλληλος, αφού για πολύ γρήγορο βηματισμό το μοντέλο εμφανίζει το φαινόμενο του overfitting.
- Συνάρτηση απώλειας (Loss function): Ως συνάρτηση απώλειας ορίζουμε την *categorical_crossentropy*. Είναι μια συνάρτηση που έχει τον ρόλο την μέτρηση της διαφοράς μεταξύ δύο κατανομών πιθανοτήτων.
- Μετρική (Metrics): Αναφερόμαστε στον προσδιορισμό της μέτρησης της ακρίβειας (accuracy).

```
model.compile(optimizer=Adam(  
    learning_rate=0.0001),  
    loss='categorical_crossentropy',  
    metrics=['accuracy'])
```

Εικόνα 28. Ορισμός παραμέτρων εκπαίδευσης

Στην εικόνα 29 παρουσιάζεται η εντολή για την εκκίνηση της εκπαίδευσης του μοντέλου.

- x: Το «x» αντιπροσωπεύει προφανώς τα δεδομένα, με τα οποία τροφοδοτούμε το μοντέλο.
- Εποχές (Epochs): Οι εποχές που θα διανύσει το μοντέλο, τις οποίες έχουμε δηλώσει αρχικά.
- Δεδομένα ελέγχου (Validation Data): Τα δεδομένα ελέγχου σε κάθε εποχή είναι υπεύθυνα για την εκτίμηση της απόκλισης από την ακρίβεια του μοντέλου.

```

history = model.fit(x = train_generator,
                    epochs = epochs,
                    validation_data = validation_generator
                    )

```

Εικόνα 29. Εκπαίδευση του μοντέλου

5.3 Υλοποίηση Εφαρμογής

5.3.1 Γενικό Μενού

Η ανάπτυξη της εφαρμογής έγινε με την βοήθεια της βιβλιοθήκης Streamlit, καθώς είναι κατάλληλη για την ενσωμάτωση μοντέλων μηχανικής μάθησης και την παρουσίασή τους. Αρχικά, δημιουργήθηκε η διεπαφή χρήστη (User Interface), που αποτελείται από ένα αναπτυσσόμενο (drop-down) μενού. Όπως φαίνεται στην εικόνα 30, ο χρήστης μπορεί να επιλέξει ανάμεσα σε τέσσερις ενέργειες.

```

st.title("Αυτόματη Ανίχνευση, Ανάλυση και Αναγνώριση Συναισθημάτων")
activities = ["Αρχική",
             "Ανάλυση συναισθήματος σε εικόνα",
             "Ανίχνευση προσώπου μέσω κάμερας",
             "About"]
choice = st.sidebar.selectbox("Επιλογή Ενέργειας", activities)
st.sidebar.markdown(" " " ")
if choice == "Αρχική":...
elif choice == "Ανίχνευση προσώπου μέσω κάμερας":...
elif choice == "Ανάλυση συναισθήματος σε εικόνα":...
elif choice == "About":...

```

Εικόνα 30. Μενού εφαρμογής

5.3.2 Λειτουργία - Ανάλυση εικόνας

Η ανάλυση μια φωτογραφίας που περιέχει κάποιο πρόσωπο γίνεται ανεβάζοντας μια εικόνα από έναν φάκελο της επιλογής του χρήστη. Στην συνέχεια, καλείται η ανάλογη συνάρτηση και η εικόνα φορτώνεται στο μοντέλο. Στην εικόνα 31 παρουσιάζονται οι απαραίτητες εντολές για την φόρτωση της εικόνας. Στην συνέχεια, η εικόνα εισέρχεται στην κατάλληλη συνάρτηση (*image_classification*) για να πραγματοποιηθεί η ανίχνευση των συναισθημάτων. Στην πρώτη στήλη (*coll*) παρατηρούμε ότι η συνάρτηση αυτή επι-

```
if st.checkbox('Επιλογή φακέλου'):
    folder_path = st.text_input('Ο φάκελος (π.χ C:\\Users\\User\\Pictures\\)', 'C:\\')
    try:
        image_file = file_selector(folder_path=folder_path)
        st.write('Επέλεξε το αρχείο `%s`' % image_file)
        col1, col2 = st.columns(2, gap='small')
        with col1:
            domi_emotion, bar_emotions = image_classification(image_file)
            print(domi_emotion)
        with col2:
            st.pyplot(bar_emotions)
    except:
        print("Not found")
```

Εικόνα 31. Φόρτωση εικόνας στην εφαρμογή και παραγωγή αποτελεσμάτων

στρέφει το κυρίαρχο συναίσθημα και την γραφική παράσταση με τα εικονιζόμενα συναισθήματα. Έπειτα, η επιστρεφόμενη γραφική παράσταση, όπως βλέπουμε, τυπώνεται στην δεύτερη στήλη (*col2*).

5.3.3 Λειτουργία - Ανίχνευση συναισθήματος με βίντεο

Η ανίχνευση του προσώπου μέσω κάμερας και η ανάλυση συναισθήματος σε εικόνα είναι, στην ουσία, το λογισμικό που υλοποιήθηκε αρχικά για να ελεγχθεί η εγκυρότητα του μοντέλου. Οι απαραίτητες αλλαγές έγιναν για να συνεργαστούν οι υλοποιημένες συναρτήσεις με την βιβλιοθήκη του Streamlit.

Η ανάλυση μιας εικόνας υλοποιήθηκε χωρίς πολλές αλλαγές στις υπάρχουσες συναρτήσεις. Η μόνη διαφοροποίηση υπάρχει στον τρόπο, με τον οποίο τυπώνεται στην εφαρμογή η γραφική παράσταση με τα συναισθήματα.

Αντιθέτως, στην ανάπτυξη της λειτουργίας για την ανίχνευση προσώπου μέσω κάμερας χρειάστηκαν κάποιες αλλαγές. Η βιβλιοθήκη δεν υποστηρίζει ακόμα βίντεο ζωντανής ροής, για αυτό όμως έχει υλοποιηθεί μία βοηθητική βιβλιοθήκη⁵ που παίζει αυτόν το ρόλο. Στην εικόνα 32 παρουσιάζεται το αντικείμενο που κατασκευάζουμε για να υποστηρίξουμε την λειτουργία βίντεο. Οι παράμετροι του αντικειμένου είναι οι εξής:

- Key (Κλειδί): Το key υποδηλώνει ένα απλό όνομα του αντικειμένου και χρησιμοποιείται κυρίως για διαχείριση του αντικειμένου από άλλες συναρτήσεις της βιβλιοθήκης.

⁵Streamlit-webrtc[48]

- **Mode (Κατάσταση Λειτουργίας):** Είναι η κατάσταση λειτουργίας που ακολουθεί το αντικείμενο. Στην προκειμένη περίπτωση, όπως φαίνεται και στην εικόνα 32 το αντικείμενο και δέχεται και στέλνει εικόνες.
- **Rtc_configuration :** Εδώ ορίζεται ο RTC⁶ του διακομιστή (server) στον οποίο ανατίθεται η κάλυψη απαιτούμενων πόρων. Στην προκειμένη περίπτωση γίνεται σε έναν server της Google που δίνεται δωρεάν.
- **Media_stream_constraints (Περιορισμοί ζωντανής ροής):** Στην προαιρετική αυτή παράμετρο, ορίζουμε τυχόν περιορισμούς που θέλουμε να δώσουμε στην ζωντανή ροή βίντεο. Ορίζουμε ως «True» για το βίντεο, αλλά ως «False» για το κανάλι του ήχου, καθώς δεν μας ενδιαφέρει ο ήχος στον έλεγχο του μοντέλου.
- **Video_processor_factory (Συνάρτηση επεξεργασίας βίντεο):** Η σημαντικότερη παράμετρος, αφού εδώ γίνεται ο ορισμός της συνάρτησης με την οποία θα λειτουργήσει το αντικείμενο του stream⁷ μας.
- **Async_processing (Ασύγχρονη επεξεργασία):** Εδώ δηλώνεται εάν η επεξεργασία εικόνας γίνεται με ασύγχρονο τρόπο. Αφού ορίζουμε ως «True», δηλαδή «αληθής», και έτσι δηλώνουμε ότι θέλουμε ασύγχρονη επεξεργασία.

```
webrtc_streamer(key="example", mode=WebRtcMode.SENDRECV, rtc_configuration=RTC_CONFIGURATION,
media_stream_constraints={"video": True, "audio": False},
video_processor_factory=FaceEmotion,
async_processing=True)
```

Εικόνα 32. Κατασκευή αντικειμένου για την υποστήριξη βίντεο ζωντανής ροής

Στην εικόνα 33 παρουσιάζεται η συνάρτηση, με την οποία φορτώνεται το αντικείμενο που προαναφέραμε για την λειτουργία της ζωντανής ροής βίντεο. Εδώ ορίζεται με ποιον τρόπο γίνεται η εξεργασία του βίντεο, η αναγνώριση του συναισθήματος και η παραγωγή γραφικών ενδείξεων στην οθόνη. Αναλυτικότερα, η callback συνάρτηση «FaceEmotion», είναι μια ειδική συνάρτηση η οποία δέχεται μια άλλη συνάρτηση ως είσοδο.

⁶(Real-time communication)

⁷ζωντανή ροή

Ακολουθεί η συνάρτηση «Face_Emotion» που είναι υπεύθυνη για την ανίχνευση του συναισθήματος και τον σχεδιασμό κατάλληλων γραφικών στην οθόνη. Περιέχει όλες τις λειτουργίες που υλοποιήθηκαν στην απλή μορφή λογισμικού για τον έλεγχο μοντέλου.

Η «recv» είναι μια ειδική συνάρτηση, η οποία καλεί την Face_Emotion. Τα αποτελέσματα της τελευταίας αποτελούνται από το καρέ εικόνας την εκάστοτε χρονική στιγμή και το συναίσθημα που την χαρακτηρίζει. Αυτά επιστρέφουν στην recv και αυτή με την σειρά της, επιστρέφει ένα αντικείμενο *av.VideoFrame*, όπως απαιτείται από τον κατασκευαστή της βιβλιοθήκης.

```
class FaceEmotion(VideoProcessorBase):  
  
    def Face_Emotion(self, frame):  
  
    def recv(self, frame):  
        img, result = self.Face_Emotion(frame)  
  
        if result == "":  
            return av.VideoFrame.from_ndarray(img, format="bgr24")  
        else:  
            self.current_emotion = result[0]  
            return av.VideoFrame.from_ndarray(img, format="bgr24")
```

Εικόνα 33. Callback συνάρτηση για την διαχείριση βίντεο

Κεφάλαιο 6

Παραδείγματα Χρήσης (Use Cases)

Στο παρόν κεφάλαιο ακολουθούν πιθανές χρήσεις του μοντέλου και γίνεται η παρουσίαση της ακρίβειάς του. Οι δύο βασικές λειτουργίες που υλοποιήθηκαν είναι οι εξής:

- Ανάλυση συναισθήματος από εικόνα
- Ανίχνευση συναισθήματος σε βίντεο

6.1 Ανάλυση εικόνας

Αρχικά, υλοποιήθηκε κατάλληλο λογισμικό για την περίπτωση που κάποιος θελήσει να φορτώσει μια εικόνα στο μοντέλο και να πάρει πληροφορίες για τα συναισθήματα που απεικονίζονται σε αυτήν. Ο έλεγχος έγινε τραβώντας αρκετές φωτογραφίες για κάθε ένα από τα επτά συναίσθημα που αναφέραμε στο Κεφάλαιο 4. Πιο συγκεκριμένα, οι φάκελοι περιέχουν δέκα φωτογραφίες οι οποίες απεικονίζουν ένα συναίσθημα.

6.1.1 Παράδειγμα λογισμικού

Παρουσιάζονται ενδεικτικές φωτογραφίες που τροφοδοτήθηκαν στο μοντέλο. Οι αντίστοιχες γραφικές παραστάσεις ακολουθούν μαζί με τον σχολιασμό των αποτελεσμάτων.

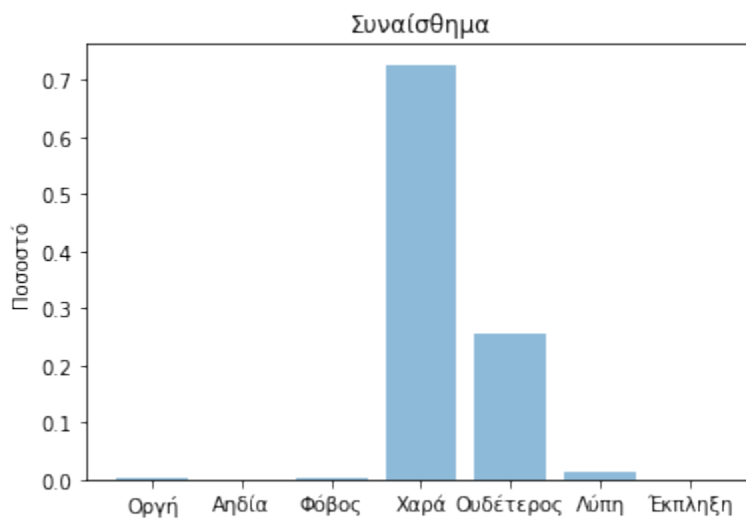
Παράδειγμα για χαρούμενες εικόνες

Οι εικόνες 34 και 36 δείχνουν προφανώς χαρούμενα συναισθήματα. Όπως παρατηρούμε στις αντίστοιχες γραφικές παραστάσεις (Εικόνες 35 και 37) το συναίσθημα που

κυρίως ανιχνεύεται είναι η χαρά. Στην πρώτη εικόνα (35) το ποσοστό χαράς είναι σχεδόν στο 70% με βάση την μέτρηση και το υπόλοιπο διανέμεται κυρίως στο «Ουδέτερος». Αυτό θεωρείται λογικό αφού η εικόνα δεν δείχνει έντονα το συναίσθημα της χαράς, απλώς ένα αμυδρό χαμόγελο. Σε αντίθεση με την δεύτερη εικόνα που στην γραφική παράσταση (37) εμφανίζεται εξ ολοκλήρου το συναίσθημα της χαράς.



Εικόνα 34. Χαρούμενος - Παράδειγμα 1ο

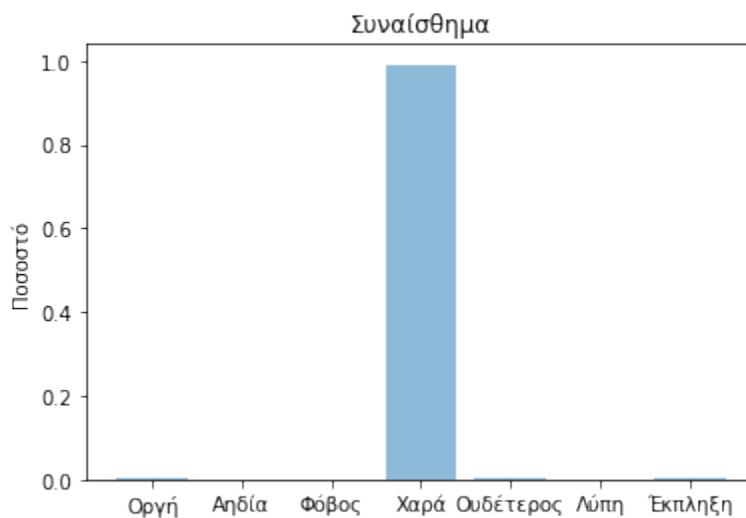


Εικόνα 35. Γραφική παράσταση για την εικόνα 34

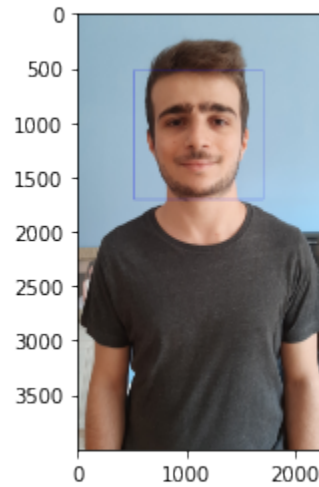
Οι υπόλοιπες περιπτώσεις για τις εικόνες που απεικονίζουν χαρά πετυχαίνουν το σκοπό τους στην πλειοψηφία τους. Ωστόσο, υπάρχουν και περιπτώσεις όπου η εικόνα, ενώ παρουσιάζει χαρούμενο συναίσθημα με συγκεκριμένα χαρακτηριστικά, όπως κλειστό στόμα σε συνδυασμό με ένα αμυδρό χαμόγελο, χαρακτηρίζει το εικονιζόμενο πρόσωπο αρκετά «Ουδέτερος» (Εικόνα 39). Το αποτέλεσμα αυτό είναι αναμενόμενο, αφού και ένας άνθρωπος θα μπορούσε να χαρακτηρίσει ένα ισχνό χαμόγελο και χαρούμενο και ουδέτερο.



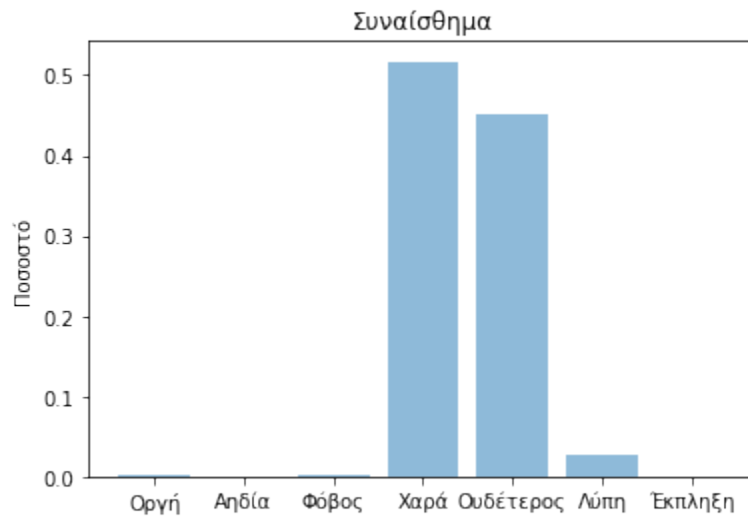
Εικόνα 36. Χαρούμενος - Παράδειγμα 2ο



Εικόνα 37. Γραφική παράσταση για την εικόνα 36



Εικόνα 38. Χαρούμενος - Παράδειγμα 3ο



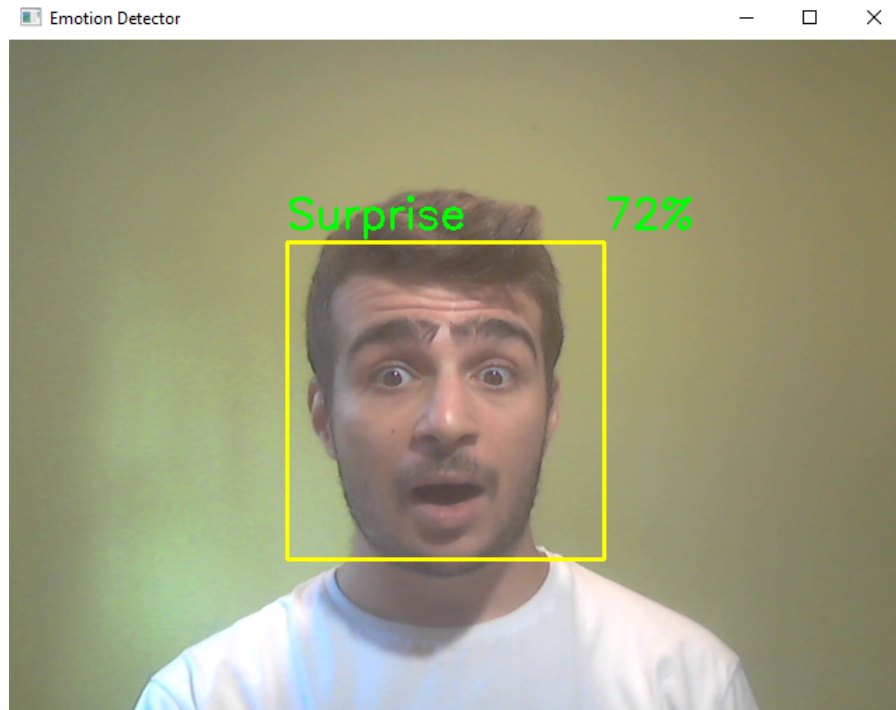
Εικόνα 39. Γραφική παράσταση για την εικόνα 38

6.2 Ανίχνευση συναισθήματος σε βίντεο

Έπειτα από την κατασκευή της παραπάνω λειτουργίας ο επόμενος στόχος ήταν να υλοποιηθεί το λογισμικό, το οποίο θα δέχεται σαν είσοδο κάθε καρέ από μία κάμερα και θα δείχνει το ισχυρότερο συναίσθημα. Θα παρουσιαστούν οι λειτουργίες και τα χαρακτηριστικά του λογισμικού που κατασκευάστηκε.

Στιγμιότυπα οθόνης

Τα παρακάτω στιγμιότυπα οθόνης (40, 41) παρουσιάζουν την χρήση του προγράμματος παίρνοντας είσοδο από μια απλή κάμερα υπολογιστή. Στην οθόνη του χρήστη εμφανίζεται το βίντεο σε πραγματικό χρόνο, καθώς και το συναίσθημα που υπερισχύει επί των άλλων. Το ποσοστό που αναγράφεται δεξιά επάνω αναφέρεται στη τιμή που έχει το ισχυρότερο συναίσθημα την συγκεκριμένη χρονική στιγμή.

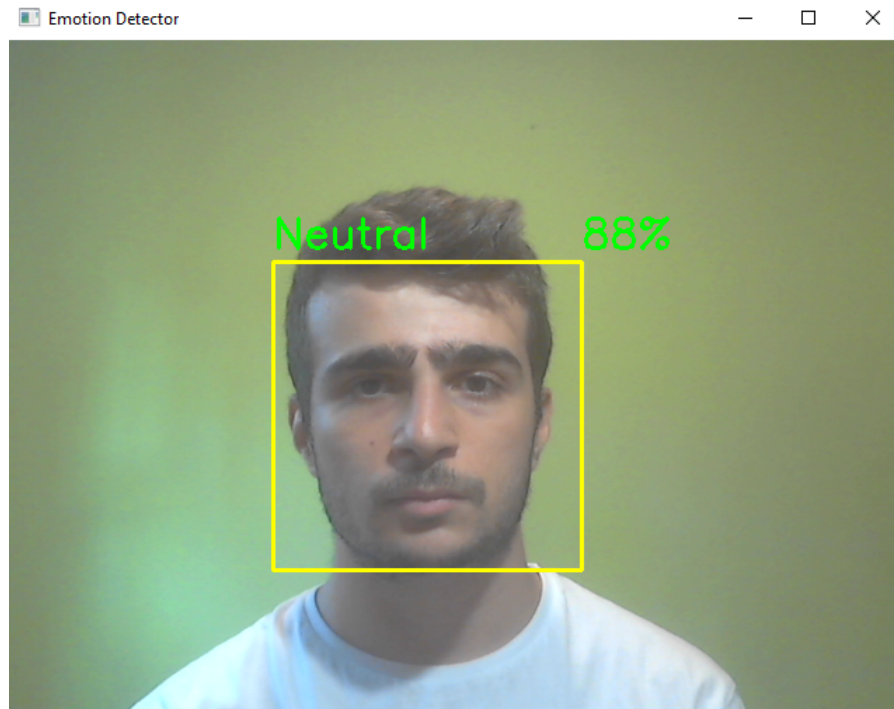


Εικόνα 40. Λειτουργία βίντεο για την περίπτωση του ενθουσιασμού

Αξιοποιώντας την πληροφορία από την κάμερα

Το πρόγραμμα που υλοποιήθηκε ταυτόχρονα με την παρουσίαση του ποσοστού και του συναισθήματος, αποθηκεύει τα προηγούμενα συναισθήματα που επικράτησαν. Αναλυτικότερα, γίνεται η αποθήκευση των πέντε προηγούμενων συναισθημάτων και εάν υπάρχει πλειοψηφία σε αυτά τα πέντε τότε μπορούμε να εξάγουμε κάποια συμπεράσματα για την συγκεκριμένη στιγμή. Αυτή η μέθοδος χρησιμοποιήθηκε για να αποφευχθεί η παραγωγή του συμπεράσματος μόνο από ένα καρέ, δηλαδή κάτι λιγότερο από ένα δευτερόλεπτο.

Στην εικόνα 42 και 43 παρουσιάζονται τα πιθανά μηνύματα που μπορούν να εμφανι-



Εικόνα 41. Περίπτωση που κάποιος δεν κάνει κάποια έκφραση

στούν στην κονσόλα του χρήστη κατά την διάρκεια της λειτουργίας του προγράμματος. Τα μηνύματα αυτά εμφανίζονται στην κονσόλα κάθε πέντε δευτερόλεπτα. Σε περίπτωση απουσίας του προσώπου εμφανίζεται το κατάλληλο μήνυμα.

```
Φαίνεσαι λυπημένος γιατί δεν βάζεις λίγο μουσική;  
Τι κάνεις;  
Σε βλέπω ενθουσιασμένο έγινε κάτι καλό;
```

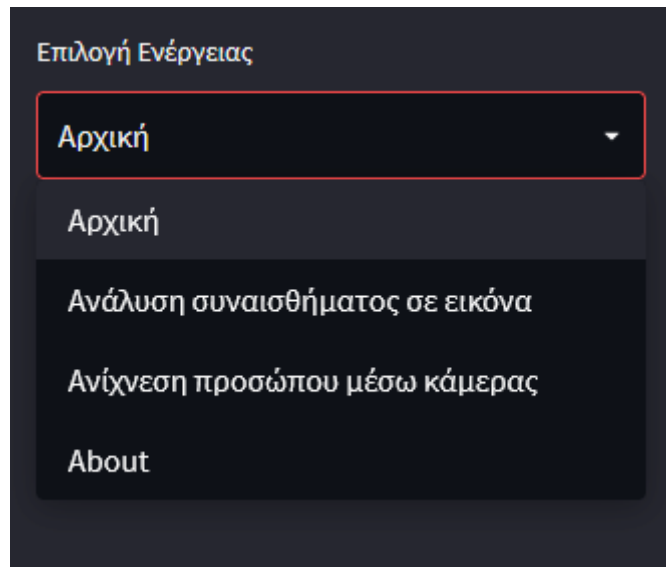
Εικόνα 42. Πιθανά μηνύματα στο χρήστη - Περίπτωση 1η

```
Είσαι χαρούμενος  
Τι κάνεις;  
Είδες κάτι αηδιαστικό;  
Μη φοβάσαι
```

Εικόνα 43. Πιθανά μηνύματα στο χρήση - Περίπτωση 2η

6.3 Περίπτωση Διαδικτυακής Εφαρμογής

Στα πλαίσια της εργασίας δημιουργήθηκε διαδικτυακή εφαρμογή που εμπεριέχει τις δύο παραπάνω λειτουργίες, όπως φαίνεται στην εικόνα 44. Ο χρήστης επιλέγει την επιθυμητή ενέργεια και φορτώνει η ανάλογη σελίδα.

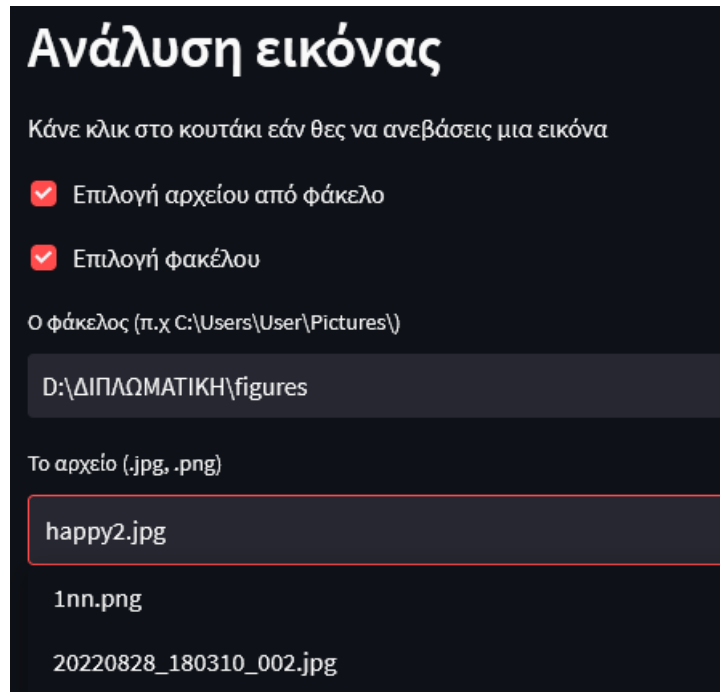


Εικόνα 44. Λειτουργίες διαδικτυακής εφαρμογής

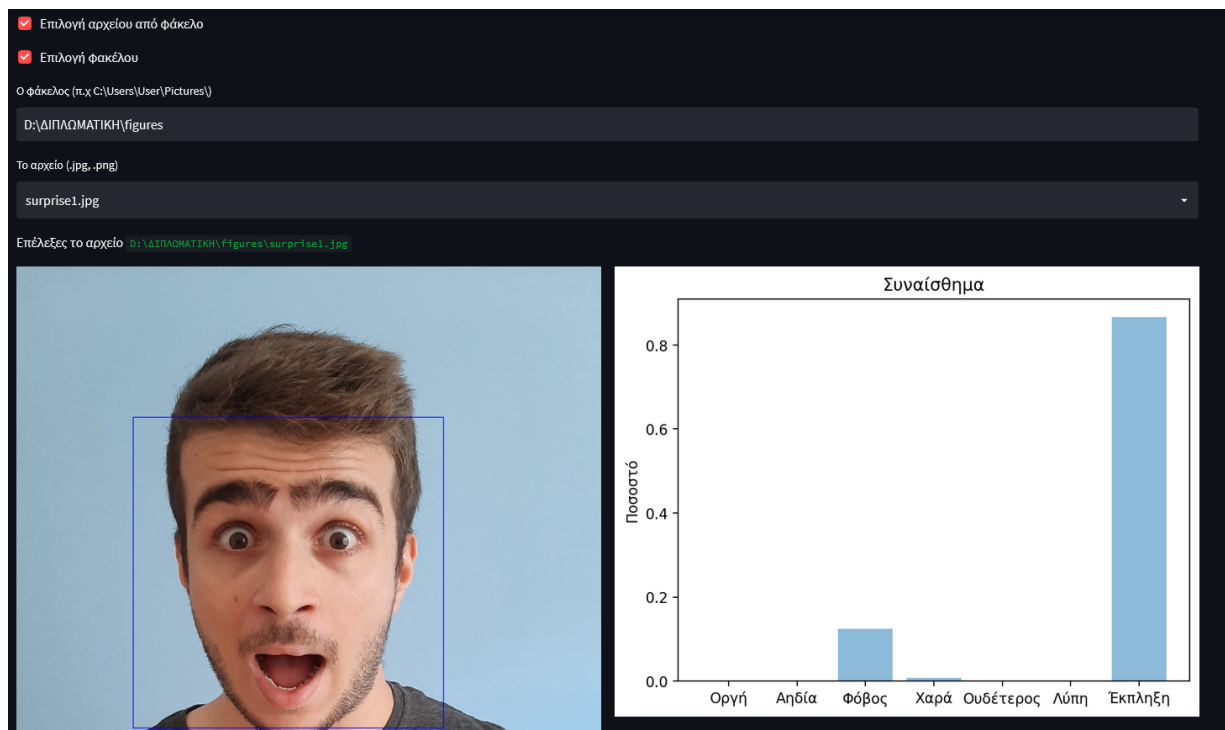
Διαδικτυακή Εφαρμογή - Ανάλυση εικόνας

Αφού ο χρήστης επιλέξει «Ανάλυση συναισθήματος σε εικόνα», φορτώνει η σελίδα όπως παρουσιάζεται στην εικόνα 45. Έπειτα, θα πρέπει να τσεκάρει τα κουτάκια έτσι ώστε να προχωρήσει η διαδικασία και να είναι σε θέση να επιλέξει το φάκελο από τον οποίο θέλει να ανεβάσει μία φωτογραφία. Όταν επιλεγθεί ο φάκελος, εμφανίζεται, όπως βλέπουμε στην εικόνα, 45 μια λίστα με τις διαθέσιμες εικόνες που περιέχει ο φάκελος. Σημειώνεται, επίσης, ότι η εικόνα θα πρέπει να είναι αρχείο μορφής είτε «jpg» είτε «png».

Η διαδικασία ολοκληρώνεται με την επιθυμητή εικόνα από τον χρήστη να εμφανίζεται στα αριστερά και στα δεξιά να υπάρχει η ανάλογη γραφική παράσταση (Εικόνα 46). Η λειτουργία έχει δυναμικό χαρακτήρα. Αυτό σημαίνει, ότι οποιαδήποτε στιγμή ο χρήστης μπορεί να επιλέξει μια άλλη φωτογραφία από τον ίδιο φάκελο. Την επόμενη στιγμή θα φορτωθεί η νέα επιλεγμένη εικόνα και την ίδια στιγμή αλλάζει και η γραφική παράσταση(Εικόνα 47).



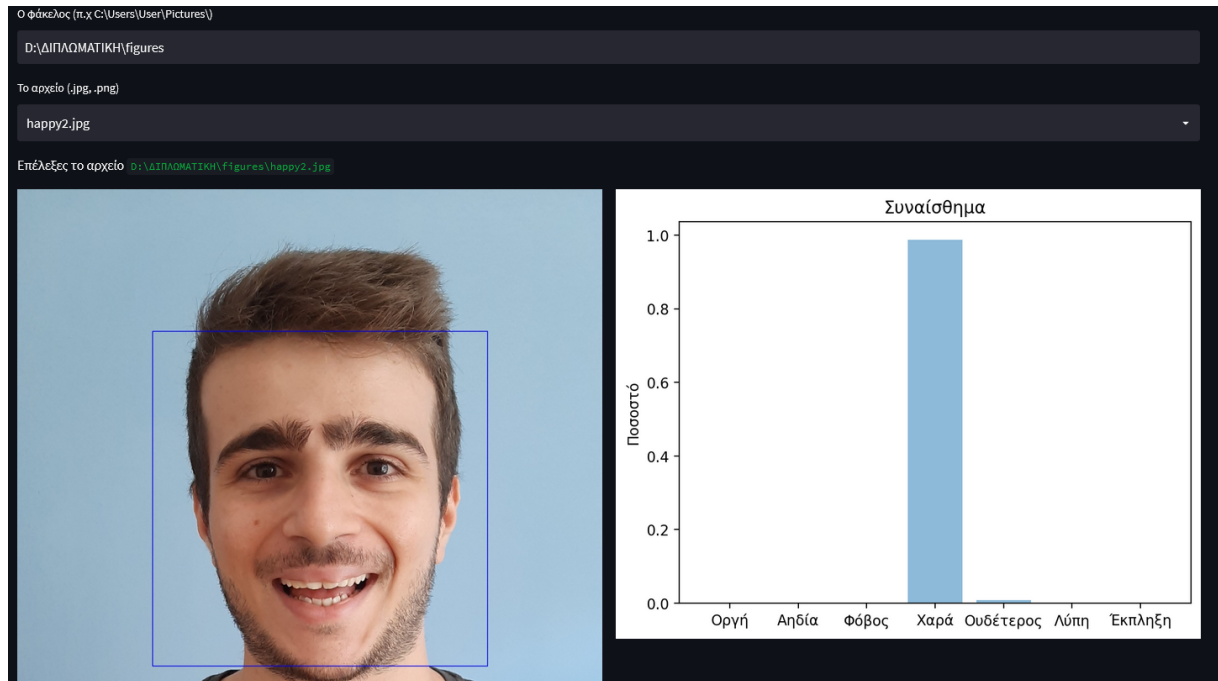
Εικόνα 45. Βήματα για την επιλογή εικόνας



Εικόνα 46. Ανάλυση εικόνας και εμφάνιση των συναισθημάτων

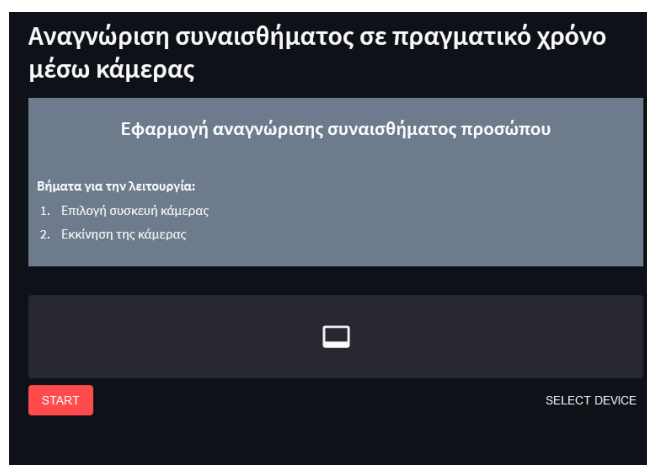
Διαδικτυακή Εφαρμογή - Ανάλυση βίντεο

Ο χρήστης ενημερώνεται για τα βήματα που πρέπει να κάνει για να εκκινήσει τη λειτουργία του βίντεο, όπως βλέπουμε και στην εικόνα 48. Αρχικά, πρέπει να επιλέξει την συ-

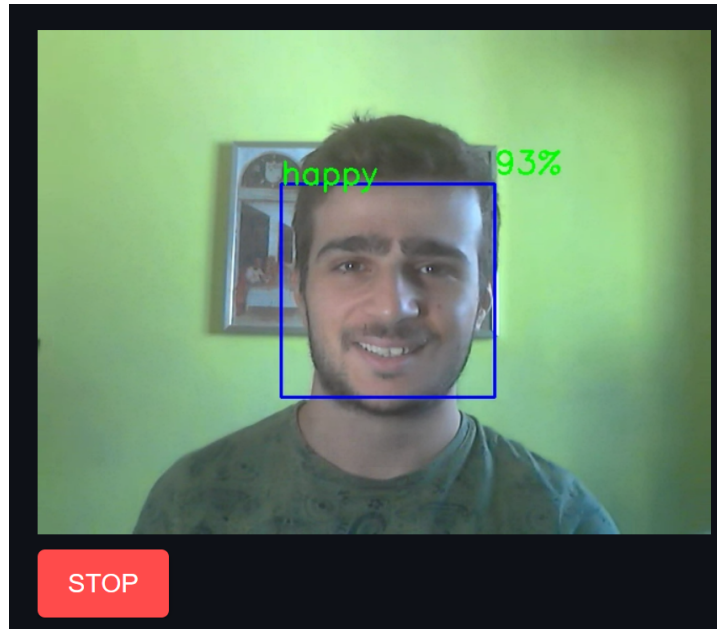


Εικόνα 47. Επιλογή νέας εικόνας για νέα ανάλυση

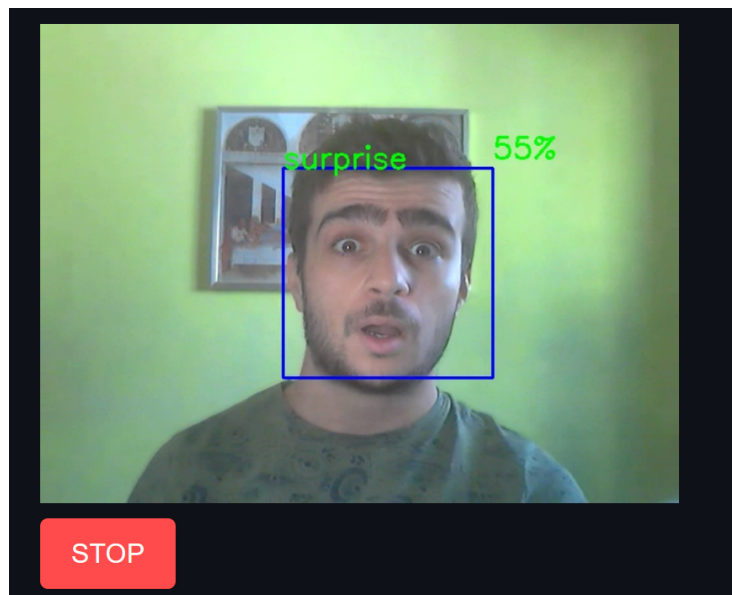
σκευή της επιλογής του και στη συνέχεια να πατήσει το κουμπί εκκίνησης (Start). Έπειτα, φορτώνεται το μοντέλο για την ανίχνευση προσώπου και το μοντέλο για την ανίχνευση συναισθήματος. Όταν ανοίξει η κάμερα, ο χρήστης θα πρέπει να βρίσκεται μπροστά από την κάμερα με μια απόσταση γύρω στο μισό με ένα μέτρο και τα χαρακτηριστικά του προσώπου του να είναι ευδιάκριτα. Στις παρακάτω εικόνες (49, 50) παρουσιάζεται η πιθανή λειτουργία του προγράμματος.



Εικόνα 48. Αρχική σελίδα στην λειτουργία της ανάλυσης βίντεο



Εικόνα 49. Έλεγχος εφαρμογής για βίντεο - Περίπτωση 1η



Εικόνα 50. Έλεγχος εφαρμογής για βίντεο - Περίπτωση 2η

6.4 Προβλήματα

Η ανάπτυξη των αρχικών λογισμικών που υλοποιήθηκαν δεν παρουσίασε κάποιο σοβαρό πρόβλημα. Ωστόσο, η ενσωμάτωση όλων των λειτουργιών στην διαδικτυακή εφαρμογή παρουσίασε αρκετά εμπόδια. Όπως προαναφέραμε η βιβλιοθήκη Streamlit δεν διαθέτει ακόμα από μόνη της την χρήση κάποιου stream για βίντεο και ήχο, για αυτό χρησιμο-

ποιήθηκε η βοηθητική βιβλιοθήκη Stremalit-webrtc. Επειδή, όμως, και οι δύο βρίσκονται σε αρχικά στάδια ανάπτυξης, έχουν κάποιους περιορισμούς όσο αφορά τη χρήση συναρτήσεων και μεταβλητών.

Ο αρχικός στόχος ήταν να ενσωματώσουμε το νήμα (thread), που έχει το ρόλο του χρονομέτρου, στην εφαρμογή για την τύπωση μηνυμάτων δυναμικά. Όμως, επειδή οι συναρτήσεις των βιβλιοθηκών χρησιμοποιούν και αυτές νήματα, θα πρέπει να κάνει σωστή διαχείριση νημάτων αλλιώς το πρόγραμμα καταρρέει. Μια πιθανή λύση θα ήταν να θέσουμε κάποιες καθολικές μεταβλητές (global variables) και με αυτόν τον τρόπο να υλοποιήσουμε μια λειτουργία χρονομέτρου. Παρόλα αυτά, ούτε αυτό επιτρέπεται από τις συγκεκριμένες βιβλιοθήκες ακόμα, γιατί δημιουργούνται άλλες ανεπιθύμητες ενέργειες στο υπόλοιπο πρόγραμμα.

Πιθανά προβλήματα παρατηρούνται όταν τα χαρακτηριστικά του προσώπου δεν είναι ευδιάκριτα. Συνήθως τέτοια φαινόμενα δημιουργούνται όταν η ποιότητα της κάμερας δεν είναι καλή και όταν ο φωτισμός δεν είναι κατάλληλος για την ανάδειξη των κύριων χαρακτηριστικών (μάτια, στόμα, ζυγωματικά) του προσώπου.

Τέλος, το εκπαιδευμένο μοντέλο, θα μπορούσαμε να πούμε ότι έχει προδιάθεση για μερικά συναισθήματα. Για παράδειγμα, το συναίσθημα του ενθουσιασμού είναι παρόμοιο με αυτό του φόβου όσο αφορά την έκφραση του προσώπου. Όμως, οι φωτογραφίες της έκπληξης είναι κατά εφτακόσιες, περισσότερες από αυτές που χαρακτηρίζουν το φόβο. Επομένως, το μοντέλο σε τέτοιες περιπτώσεις ανιχνεύει ευκολότερα το αίσθημα της έκπληξης σε περιπτώσεις που κάποιος άνθρωπος θα το χαρακτήριζε ως φόβο. Το συναίσθημα της αηδίας, παρ' όλο που έχει πολύ λίγα παραδείγματα στο dataset για συγκεκριμένες εκφράσεις και συνδυασμό χαρακτηριστικών, ανιχνεύεται μόνο στην περίπτωση που εμφανίζονται αυτά τα χαρακτηριστικά.

Κεφάλαιο 7

Αξιολόγηση

Στο παρόν κεφάλαιο θα γίνει λόγος για την αξιολόγηση του μοντέλου που εκπαιδεύτηκε. Επίσης, θα παρουσιαστούν τα αποτελέσματα από συγκεκριμένες μετρήσεις που αυτό εξετάστηκε. Θα ακολουθήσει η σύγκριση του με άλλα μοντέλα που ειδικεύονται στην αναγνώριση συναισθήματος.

Η βιβλιοθήκη Tensorflow παρέχει μια δική της συνάρτησης αξιολόγησης του μοντέλου που μας δίνει την ακρίβεια του μοντέλου για τα δεδομένα εκπαίδευσης και την ακρίβεια για τα δεδομένα ελέγχου. Όπως φαίνεται και στην εικόνα 51, η ακρίβεια του μοντέλου για τα δεδομένα εκπαίδευσης είναι 71.80% και τα δεδομένα ελέγχου είναι 68.23%. Παρατηρούμε ότι το μοντέλο δεν έχει υποστεί υπερπροσαρμογή στα δεδομένα εκπαίδευσης, κάτι αρκετά θετικό.

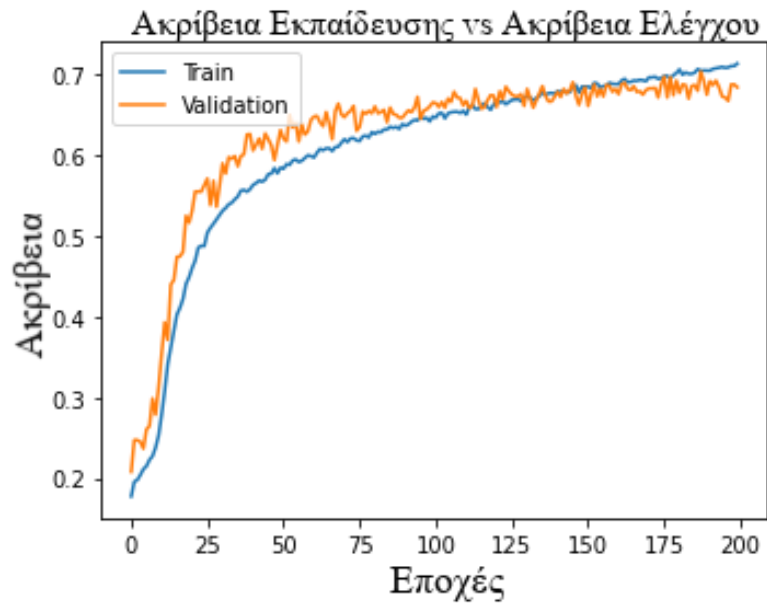
```
train_loss, train_acc = model.evaluate(train_generator)
test_loss, test_acc = model.evaluate(validation_generator)
print("final train accuracy = {:.2f} , validation accuracy = {:.2f}".format(train_acc*100, test_acc*100))

718/718 [=====] - 349s 486ms/step - loss: 0.9818 - accuracy: 0.7180
45/45 [=====] - 21s 457ms/step - loss: 1.1238 - accuracy: 0.6823
final train accuracy = 71.80 , validation accuracy = 68.23
```

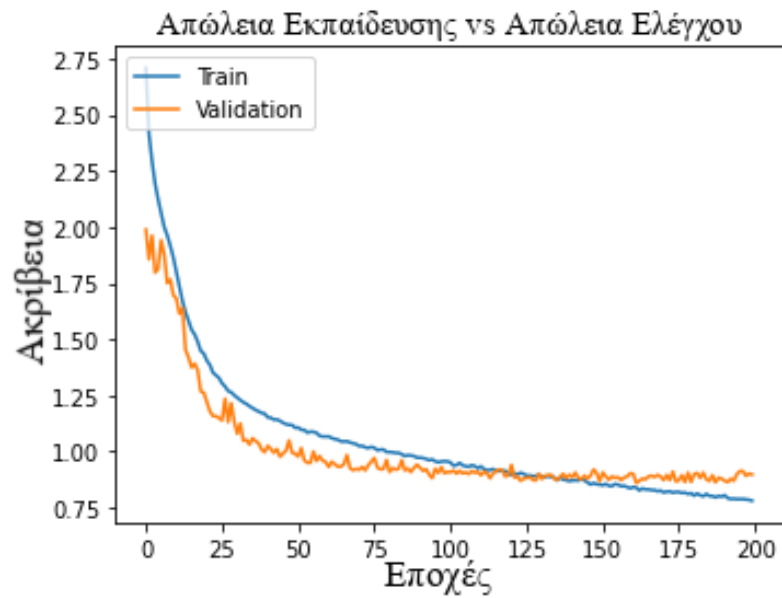
Εικόνα 51. Στιγμιότυπο από την αξιολόγηση του μοντέλου

Στις εικόνες 52 και 53 παρουσιάζονται οι γραφικές αναπαραστάσεις από την εκπαίδευση του μοντέλου. Πιο συγκεκριμένα, το γράφημα 52 είναι η ακρίβεια του μοντέλου για τα δεδομένα εκπαίδευσης με τα δεδομένα ελέγχου. Ενώ το γράφημα 53 είναι η γραφική αναπαράσταση της απώλειας του μοντέλου για τα δεδομένα εκπαίδευσης και ελέγχου. Με πορτοκαλί χρώμα σχεδιάζονται τα δεδομένα ελέγχου και με μπλε χρώμα τα δεδομένα

εκπαίδευσης.



Εικόνα 52. Ακρίβεια του μοντέλου



Εικόνα 53. Απώλεια του μοντέλου

7.1 Μετρήσεις εκπαιδευμένου μοντέλου

Οι μετρήσεις είναι μια μέθοδος για να εξετάσουμε την αποτελεσματικότητα του μοντέλου που υλοποιήθηκε. Θα εξεταστούν συγκεκριμένες μέθοδοι μετρήσεων όπως η Recall, η Precision και η F1.

Η μέτρηση Recall (ανάκληση) αναφέρεται στην διαίρεση των True positive (αληθινών θετικών) αποτελεσμάτων προς το άθροισμα των True positive (αληθινών θετικών) με των False negative (ψευδών αρνητικών) αποτελεσμάτων.

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (7.1)$$

Η μέτρηση Precision (ακρίβεια) αναφέρεται στην διαίρεση των True positive (αληθινών θετικών) αποτελεσμάτων προς το άθροισμα των True positive (αληθινών θετικών) με των False positive (ψευδών θετικών) αποτελεσμάτων.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (7.2)$$

Η μέτρηση F1 χρησιμοποιεί τις δύο προηγούμενες και αναφέρεται όπως στην εξίσωση 7.3. Η μέτρηση αυτή παρουσιάζει υψηλά αποτελέσματα, εάν και οι άλλες μετρήσεις είναι υψηλές. Εάν μια από τις δύο είναι χαμηλή και η άλλη υψηλή, τότε η μέτρηση F1-score θα πάρει μεσαία τιμή. Επομένως, στην περίπτωση που το μοντέλο στις προηγούμενες δύο μετρήσεις λάβει χαμηλή τιμή, τότε θα είναι χαμηλή και η μέτρηση της F1.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7.3)$$

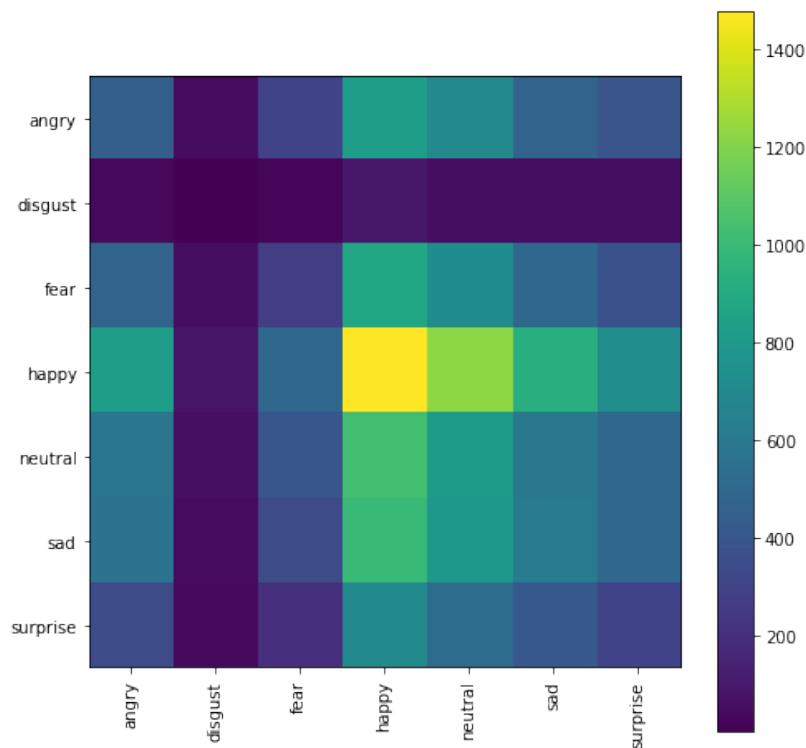
Ακολουθεί ο πίνακας ταξινόμησης (4), όπου συγκρίνονται οι παραπάνω μετρήσεις (Precision, Recall, F1-score) με άξονα το κάθε συναίσθημα. Επίσης, υπάρχει στον πίνακα η στήλη «support», η οποία μας δείχνει τον αριθμό των περιπτώσεων για την εκάστοτε κλάση στο συγκεκριμένο dataset. Το μη ισοσταθμισμένο support στα δεδομένα εκπαίδευσης επηρεάζει την απόδοση των μετρήσεων. Έτσι, παρατηρώντας τον πίνακα διακρίνουμε, ότι τα αποτελέσματα είναι πολύ χαμηλά, ειδικά στο συναίσθημα της αγδίας που είναι σχεδόν μηδενική η τιμή.

Στην εικόνα 54 παρατίθεται ο πίνακας σύγκρισης για τα συναισθήματα. Στον πίνακα γίνεται η αναπαράσταση των περιπτώσεων που είναι αληθώς αρνητικά & θετικά και ψευ-

Πίνακας 4. Πίνακας Ταξινόμησης - Classification report

Score Συναίσθημα	Precision	Recall	F1-score	Support
Θυμός	0.14	0.14	0.14	3196
Αηδία	0.01	0.01	0.01	349
Φόβος	0.13	0.08	0.10	3278
Χαρά	0.24	0.26	0.26	5772
Ουδέτερος	0.17	0.20	0.18	3972
Λύπη	0.17	0.16	0.16	3864
Έκπληξη	0.11	0.12	0.11	2537

δώς αρνητικά & θετικά. Παρατηρώντας τα χρώματα του πίνακα αλλά και όπως ήταν αναμενόμενο, λόγω των πολλών περιπτώσεων, στο συναίσθημα της χαράς έχουμε καλές προβλέψεις. Αντιθέτως, στο συναίσθημα της αηδίας τα χρώματα είναι πολύ σκούρα και αυτό δηλώνει την αποτυχία του μοντέλου να προβλέπει σωστά το αίσθημα αυτό.

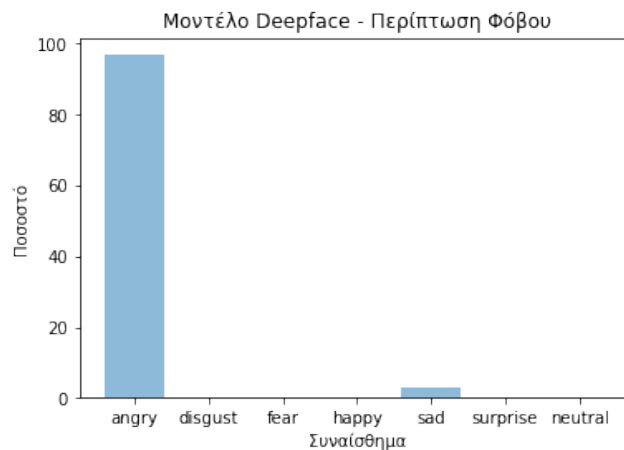


Εικόνα 54. Πίνακας σύγκρισης - Confusion Matrix

7.2 Σύγκριση

7.2.1 Περίπτωση φόβου

Στον παρακάτω πίνακα 55 παρατηρείται ότι, ενώ το πρόσωπο του χρήστη είχε έκφραση φόβου, το μοντέλο εντόπισε μεγαλύτερο ποσοστό θυμού και ένα πολύ μικρό ποσοστό λύπης. Σε αυτήν, λοιπόν, την πρώτη απόπειρα ανίχνευσης του συναισθήματος έχει γίνει χρήση του μοντέλου DeepFace. Στην δεύτερη περίπτωση, πίνακας 56, και έχοντας την ίδια περίπτωση έκφρασης φόβου, φαίνεται ότι εντοπίστηκε μεγαλύτερο ποσοστό στο συναίσθημα της έκπληξης, πολύ μικρά ποσοστά στα συναισθήματα του θυμού, της ουδετερότητας και της λύπης, ενώ σε αυτό του φόβου η μπάρα υπερβαίνει κατά λίγο τα τελευταία που αναφέρθηκαν. Στην δεύτερη περίπτωση έγινε χρήση του μοντέλου εργασίας και παρ' όλο που εντοπίστηκαν κι άλλα, το ζητούμενο συναίσθημα δεν παραλήφθηκε. Η διαφορά αυτή μεταξύ των αποτελεσμάτων των δύο μοντέλων, ίσως, οφείλεται στην ασάφεια της έκφρασης του προσώπου.



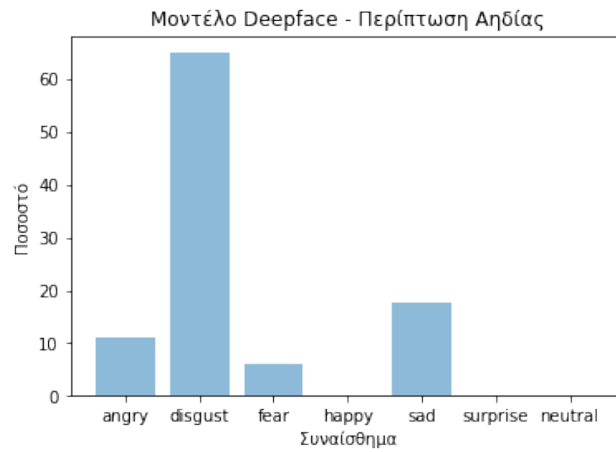
Εικόνα 55. Περίπτωση Φόβου- Μοντέλου DeepFace



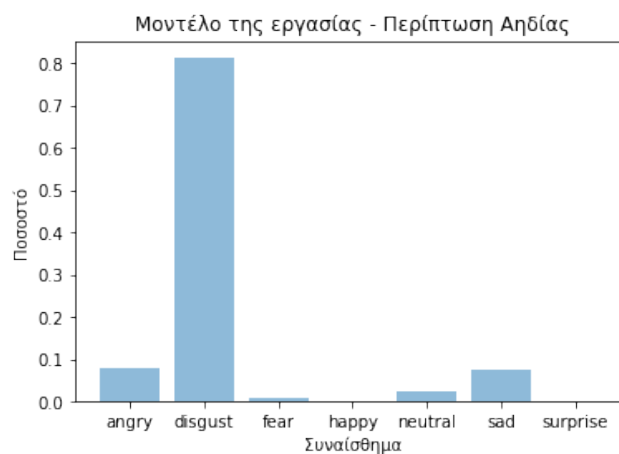
Εικόνα 56. Περίπτωση Φόβου- Μοντέλου της εργασίας

7.2.2 Περίπτωση αηδίας

Στο παράδειγμα περίπτωσης του συναισθήματος αηδίας, παρατηρείται, ότι στην πρώτη προσπάθεια 57 εντοπίστηκε από το μοντέλο ένα πολύ μεγάλο ποσοστό αηδίας 60%, πράγμα που ήταν αναμενόμενο λόγω της έκφρασης του προσώπου του χρήστη, ένα μικρό ποσοστό θυμού 10%, ένα λίγο υψηλότερο ποσοστό λύπης 20% και ένα πολύ μικρό ποσοστό φόβου 5% περίπου. Όπως και στην περίπτωση φόβου, η πρώτη προσπάθεια έγινε με τη χρήση του μοντέλου DeepFace, ενώ η δεύτερη με τη χρήση του μοντέλου εργασίας. Στην δεύτερη προσπάθεια 58, φαίνεται, ότι κι εδώ το μοντέλο έχει ανιχνεύσει σωστά το ζητούμενο συναίσθημα (της αηδίας), ωστόσο εδώ το ποσοστό ανέρχεται στο 80% πράγμα που καθιστά το μοντέλο πιο αξιόπιστο. Τα υπόλοιπα συναισθήματα που ανιχνεύτηκαν (θυμός, φόβος, ουδετερότητα, λύπη) εμφανίζουν ποσοστά μικρότερα από 10% .



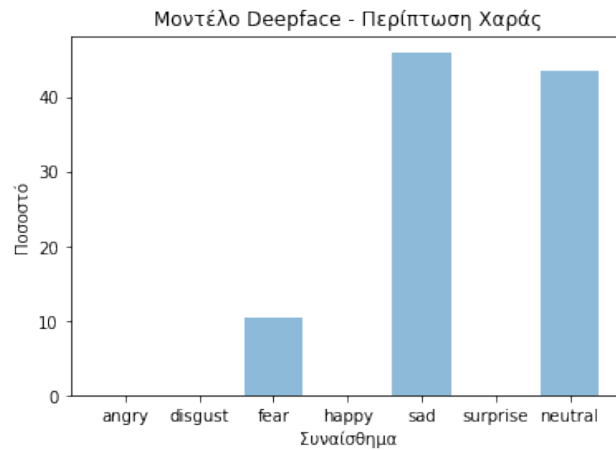
Εικόνα 57. Περίπτωση Αηδίας- Μοντέλου DeepFace



Εικόνα 58. Περίπτωση Αηδίας- Μοντέλου της εργασίας

7.2.3 Περίπτωση χαράς

Στην πρώτη απόπειρα ανίχνευσης του συναισθήματος της χαράς [59](#), μέσω της χρήσης του μοντέλου DeepFace, διαπιστώνεται, ότι το μοντέλο αδυνατεί να εντοπίσει το ζητούμενο συναίσθημα και αντί αυτού εντοπίζει μεγάλο ποσοστό του συναισθήματος της λύπης 50%, επίσης μεγάλο ποσοστό ουδετερότητας πάνω από 40% και ένα ποσοστό 10% φόβου. Αντίθετα, με τη χρήση του μοντέλου εργασίας, κατά την προσπάθεια [60](#), γίνεται σωστά η ανίχνευση μόνο του συναισθήματος της χαράς με ποσοστό 100%.



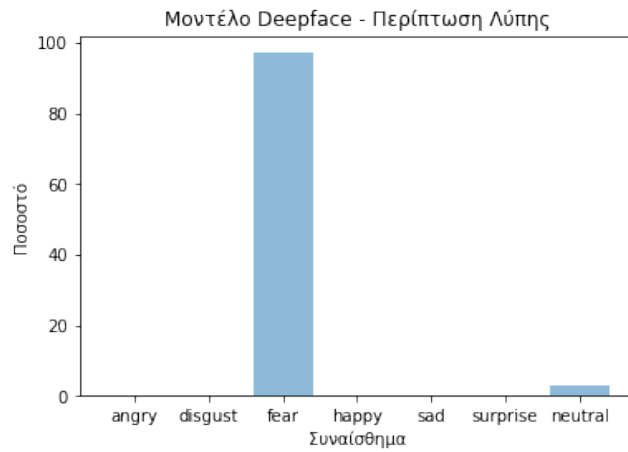
Εικόνα 59. Περίπτωση Χαράς - Μοντέλου DeepFace



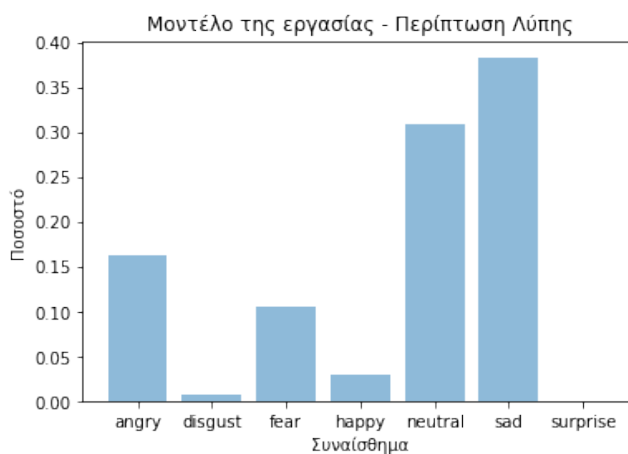
Εικόνα 60. Περίπτωση Χαράς - Μοντέλου της εργασίας

7.2.4 Περίπτωση λύπης

Στην περίπτωση ανίχνευσης του συναισθήματος της λύπης το μοντέλο DeepFace αναγνώρισε το ζητούμενο συναίσθημα ως σχεδόν εξ ολοκλήρου φόβο. Το διάγραμμα 61, που εξήχθη από το συγκεκριμένο μοντέλο, εμφανίζει ποσοστό φόβου σχεδόν 95% και ένα πολύ μικρό ποσοστό ουδετερότητας κάπου 5%. Αντίθετα, σε αυτήν την περίπτωση, το μοντέλο εργασίας 62, ενώ παρουσιάζει υψηλό ποσοστό στο συναίσθημα της λύπης 40%, εξακολουθεί να είναι ανακριβές σε σχέση με το μοντέλο DeepFace καθώς εμφανίζει πληθώρα άλλων συναισθημάτων σε μεγάλα ποσοστά.



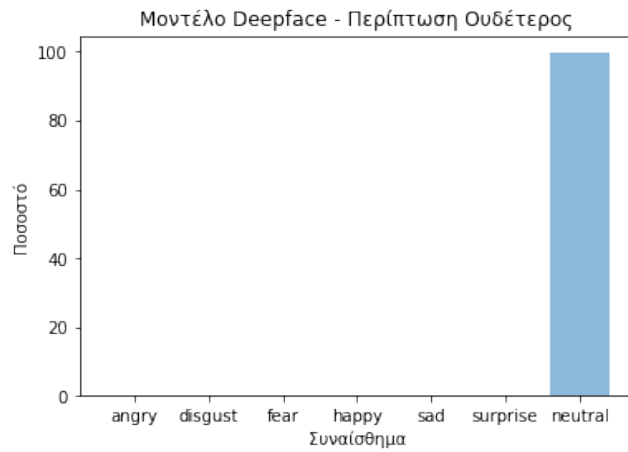
Εικόνα 61. Περίπτωση Λύπης - Μοντέλου DeepFace



Εικόνα 62. Περίπτωση Λύπης - Μοντέλου της εργασίας

7.2.5 Περίπτωση ουδέτερος

Στην προσπάθεια της περίπτωση ουδέτερος, το μοντέλο DeepFace, φαίνεται να είναι πιο ακριβές από το μοντέλο της εργασίας. Στο πρώτο διάγραμμα 63, το συναίσθημα ουδέτερος εμφανίζεται σε ποσοστό 100%, χωρίς κανένα άλλο συναίσθημα. Στο δεύτερο διάγραμμα 64, το μοντέλο, παρ' όλο που εντοπίζει σωστά το ζητούμενο συναίσθημα σε ποσοστό 80%, εμφανίζει και ένα μικρό δείγμα από το συναίσθημα του θυμού, του φόβου, της λύπης και ένα ποσοστό πάνω από 10% για το συναίσθημα της χαράς. Με αυτά τα δεδομένα μπορούμε να συμπεράνουμε ότι το μοντέλο DeepFace λειτούργησε σωστότερα.



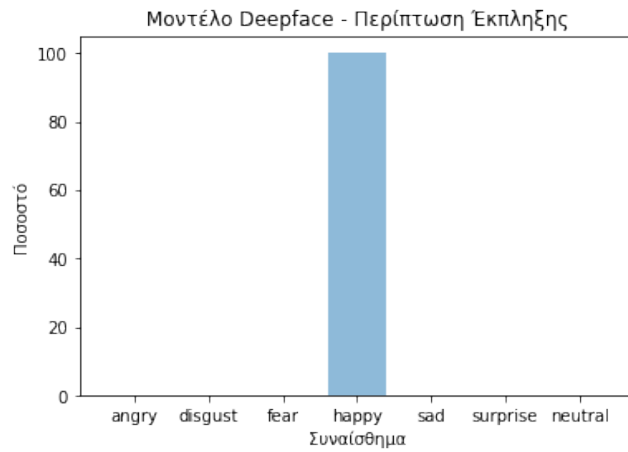
Εικόνα 63. Περίπτωση Ουδέτερος - Μοντέλου DeepFace



Εικόνα 64. Περίπτωση Ουδέτερος - Μοντέλου της εργασίας

7.2.6 Περίπτωση έκπληξης

Στην περίπτωση του συναισθήματος της έκπληξης, το μοντέλο DeepFace ανίχνευσε με χαμηλότερη ακρίβεια το ζητούμενο συναίσθημα καθώς παρατηρείται στο διάγραμμα. Το συναίσθημα, λοιπόν, της ουδετερότητας δεν εμφανίζεται καθόλου στο πρώτο διάγραμμα 65, ενώ βλέπουμε το συναίσθημα της χαράς σε πολύ μεγάλο ποσοστό 100%. Αντίθετα, με τη χρήση του μοντέλου της εργασίας, στο δεύτερο διάγραμμα 66, παρατηρείται ποσοστό έκπληξης 60% και ένα επίσης μεγάλο ποσοστό φόβου 40%. Παρ' όλα αυτά, το μοντέλο εργασίας θεωρείται πιο αξιόπιστο, αφού κατάφερε να εντοπίσει το βασικότερο συναίσθημα.



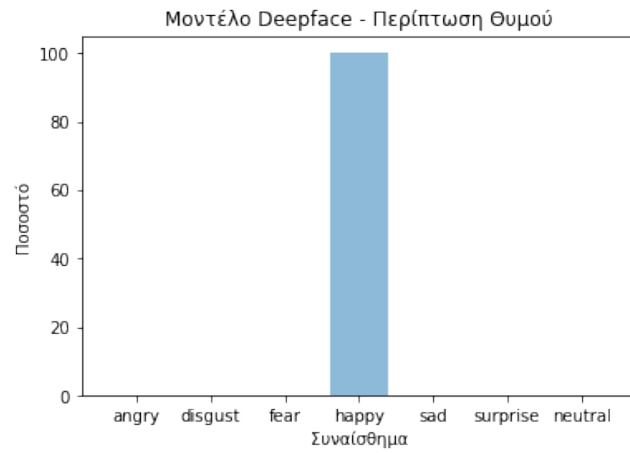
Εικόνα 65. Περίπτωση Έκπληξης - Μοντέλου DeepFace



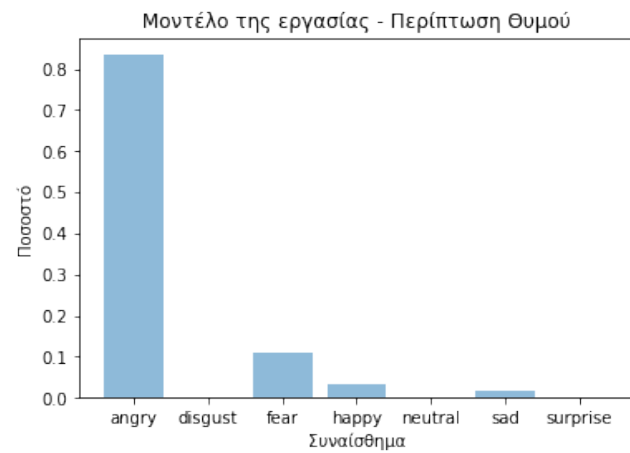
Εικόνα 66. Περίπτωση Έκπληξης - Μοντέλου της εργασίας

7.2.7 Περίπτωση θυμού

Στην περίπτωση αντίχενωσης του συναισθήματος του θυμού, το πρώτο διάγραμμα, 67, που εξήχθη από το μοντέλο DeepFace, δεν εμφανίζει καθόλου το ζητούμενο συναίσθημα και παρουσιάζει ένα μεγάλο ποσοστό χαράς 100%. Αντίθετα, στην περίπτωση του μοντέλου της εργασίας 68, το διάγραμμα που εξήχθη παρουσιάζει ποσοστό θυμού 80%, ποσοστό φόβου 10%, ποσοστό χαράς λιγότερο από 10% και ένα ποσοστό λύπης λίγο πάνω από το 0. Συμπερασματικά, σε αυτήν την περίπτωση συναισθήματος, το μοντέλο της εργασίας λειτούργησε αποδοτικότερα και με περισσότερη ακρίβεια από ότι το μοντέλο DeepFace.



Εικόνα 67. Περίπτωση Θυμού - Μοντέλου DeepFace



Εικόνα 68. Περίπτωση Θυμού - Μοντέλου της εργασίας

Κεφάλαιο 8

Συμπεράσματα & Μελλοντικές

Επεκτάσεις

8.1 Συμπεράσματα

Η συγκεκριμένη εργασία εστιάζει στην κατανόηση και ανάλυση των αρχών σχεδίασης νευρωνικών δικτύων και ως εκ τούτου και συνελκτικών νευρωνικών δικτύων. Η σχεδίαση των τελευταίων, διαδραματίζει πολύ σημαντικό ρόλο στην απόδοση των εκπαιδευόμενων μοντέλων όταν πρόκειται για μεγάλα datasets εικόνων. Η κατανόηση εννοιών που συνθέτουν ένα συνελκτικό νευρωνικό δίκτυο ήταν απαραίτητη για την σωστή εκπαίδευση του μοντέλου και βελτίωση του επαναλαμβάνοντας την διαδικασία εκπαίδευσης. Η υλοποίηση κατάλληλου λογισμικού, που αναπτύχθηκε στο πλαίσιο της εγκυρότητας του μοντέλου, μας παρείχε ανατροφοδότηση για την συμπεριφορά του σε πραγματικές καταστάσεις.

Στόχος της παρούσας διπλωματικής εργασίας, ωστόσο, είναι η εστίαση στη χρήση τέτοιου είδους μοντέλων σε καθημερινές εφαρμογές. Η εφαρμογή που αναπτύχθηκε είναι ένα καλό παράδειγμα χρήσης του λογισμικού που χρησιμοποιεί το εκπαιδευμένο μοντέλο. Τα συναισθήματα που μπορούν να ανιχνευτούν είναι επτά: η χαρά, ο φόβος, η λύπη, η αγάπη, η έκπληξη, ο θυμός και η ουδετερότητα. Η ανίχνευση συναισθήματος του προσώπου ήταν αρκετά επιτυχής σε πραγματικό χρόνο μέσω κάμερας χωρίς να υπάρχουν καθυστερήσεις στην παρουσίαση των αποτελεσμάτων. Επίσης, η εφαρμογή προσφέρει την λειτουργία της μεταφόρτωσης μιας φωτογραφίας από τον υπολογιστή, παρέχοντας έτσι μια

πιο αναλυτική εικόνα των συναισθημάτων που αναγνωρίζονται. Η εφαρμογή παραμένει απλή στην χρήση της, παρέχοντας μενού για πλοήγηση του χρήστη στις λειτουργίες της.

8.2 Περιορισμοί

Κατά τη διάρκεια υλοποίησης του προγράμματος, σημειώθηκαν κάποιοι περιορισμοί και δυσκολίες, οι οποίες καταγράφονται προσεκτικά παρακάτω.

Ο βασικότερος περιορισμός που δυσχέρανε την ομαλή εκπόνηση του προγράμματος ήταν η ποιότητα της εικόνας, η οποία αρκετές φορές ήταν κακή. Ωστόσο, το φαινόμενο αυτό οφείλεται σε αρκετούς εσωτερικούς και εξωτερικούς παράγοντας. Πολύ σημαντικό ρόλο στην ποιότητα της ζητούμενης εικόνας παίζει και η ποιότητα της κάμερας. Εάν η κάμερα είναι χαμηλής ποιότητας, τότε είναι αναμενόμενη η χαμηλή απόδοση μιας ευκρινούς εικόνας. Επιπροσθέτως, άλλα προβλήματα που είναι πιθανό να εμφανιστούν είναι ο κακός ή μέτριος φωτισμός, η απόσταση του ατόμου (προσώπου) από την κάμερα καθώς και διάφορα αξεσουάρ που ενδεχομένως αυτό να φοράει. Τέτοια παραδείγματα είναι τα γυαλιά, ακουστικά κ.λπ. Όλα αυτά μπορούν να συμβάλλουν αρνητικά, καθώς παρεμποδίζουν την σωστή λειτουργία του προγράμματος. Επομένως, γίνεται εύκολα αντιληπτό πως κατά τη διάρκεια αυτής της διαδικασίας θα πρέπει να υπάρχει ευκρίνεια και να μην παρεμβάλλονται εξωτερικοί παράγοντες.

Επίσης, μία σημαντική παρατήρηση αποτελεί και το γεγονός ότι γίνεται χρήση διαφορετικού μοντέλου για την αναγνώριση του προσώπου από εκείνου που χρησιμοποιείται για την αναγνώριση των συναισθημάτων. Ακόμη, πρέπει να αναφερθεί ότι η παρούσα έρευνα δεν έχει δοκιμαστεί σε φυσικές συνθήκες, ενώ εξαιρετικός παράγοντας επιρροής των εξαγόμενων αποτελεσμάτων αποτελεί και η γωνία από την οποία εμφανίζεται το πρόσωπο στην κάμερα. Εάν η γωνία δε συμβάλλει ευνοϊκά στην εμφάνιση του ατόμου, τότε είναι πολύ πιθανό να αλλοιωθούν πολλά από τα χαρακτηριστικά του προσώπου του. Κάτι που πρέπει να ληφθεί υπόψη, σε αυτήν την περίπτωση, είναι το σημείο από το οποίο περνάει το φως. Πιο συγκεκριμένα, το πρόσωπο θα πρέπει να φωτίζεται από πολλές, αν όχι από όλες, τις πλευρές, έτσι ώστε να μην είναι δυνατή η δημιουργία σκιών. Αντίθετα, το υπερβολικό φως μπορεί κι αυτό με τη σειρά του να προκαλέσει αντίστοιχα προβλήματα. Θα πρέπει να σημειωθεί και το ότι, ιδανικά, μπροστά στην κάμερα θα πρέπει να εμφανίζε-

ται ένα άτομο, προκειμένου τα αποτελέσματα να είναι εγκυρότερα. Ωστόσο, έγινε δοκιμή και με δύο πρόσωπα και παρουσιάστηκαν τα ανάλογα συναισθήματα.

Άλλος ένας παράγοντας που παίζει πολύ σημαντικό ρόλο στην απόδοση του μοντέλου είναι τα δεδομένα εκπαίδευσης που τροφοδοτούνται αρχικά στο μοντέλο. Αναλυτικότερα, το συγκεκριμένο dataset που χρησιμοποιήθηκε, κοιτώντας τις εικόνες που χαρακτηρίζουν το συναίσθημα της έκπληξης για παράδειγμα δεν εκφράζουν αυστηρά έκπληξη. Το ίδιο ισχύει και για τις υπόλοιπες περιπτώσεις. Πραγματοποιήθηκαν τεχνικές για βελτιστοποίηση του dataset αλλά στην εκπαίδευση η ακρίβεια του μοντέλου δεν μπορεί να ξεπεράσει συγκεκριμένες τιμές και φτάνει τιμές της τάξης 70%.

8.3 Μελλοντικές Επεκτάσεις

Στην ενότητα αυτή θα παρουσιαστούν συνοπτικά μελλοντικές επεκτάσεις που αφορούν την βελτιστοποίηση του μοντέλου και την χρήση του.

Αρχικά, για την βελτιστοποίηση του μοντέλου θα μπορούσε κάποιος να βασιστεί σε δύο παράγοντες. Τα δεδομένα εκπαίδευσης θα μπορούσαν να επεξεργαστούν με τέτοιο τρόπο ώστε να παρέχουν μια πιο ομαλή κατανομή στις εικόνες με τα συναισθήματα. Επίσης, όσο αφορά το dataset θα μπορούσε να χρησιμοποιηθεί διαφορετικό και να γίνει επανεκπαίδευση με τον ίδιο αλγόριθμο που παρουσιάστηκε.

Ένας άλλος τρόπος διερεύνησης για την βελτίωση του μοντέλου είναι η εκπαίδευση του με διαφορετικές αρχιτεκτονικές συνελκτικών δικτύων. Πιο αναλυτικά, η χρήση αρχιτεκτονικών με μεγαλύτερη πολυπλοκότητα, δηλαδή περισσότερων κρυφών επιπέδων και επιπέδων δειγματοληψίας ενδέχεται να παρουσιάσουν καλύτερα αποτελέσματα στο συγκεκριμένο πρόβλημα.

Η εφαρμογή που παρουσιάστηκε θα μπορούσε να χρησιμοποιηθεί για την πρόληψη αυτοκινητιστικών ατυχημάτων. Δεν είναι λίγες οι φορές που έχουν συμβεί ατυχήματα, τα οποία οφείλονται σε διάφορες εκρήξεις συναισθημάτων των οδηγών. Επομένως, η κατασκευή ενός μοντέλου που βρίσκεται συνεχώς διαθέσιμο ως εφαρμογή στο διαδίκτυο θα μπορούσε να μειώσει το ποσοστό των ατυχημάτων που συμβαίνουν παγκοσμίως. Πιο συγκεκριμένα, το πρόγραμμα θα είναι υπεύθυνο για την παρακολούθηση και την αναγνώριση των συναισθημάτων του οδηγού ενός οχήματος καθ' όλη τη διάρκεια της διαδρομής.

Σε περίπτωση καταγραφής έντονων συναισθημάτων, το πρόγραμμα στέλνει αυτομάτως κάποιο μήνυμα με το οποίο προειδοποιεί τον οδηγό ή τον βοηθά για να ηρεμήσει η συναισθηματική του φόρτιση.

Η υλοποίηση αυτού του εγχειρήματος θα μπορούσε να πραγματοποιηθεί χρησιμοποιώντας τις δυνατότητες της υπολογιστικής νέφους. Το Tensorflow Serving είναι ένα εργαλείο της ομώνυμης βιβλιοθήκης που συνδυάζει τις δυνατότητες και τα χαρακτηριστικά νέφους παρέχοντας εύκολη χρήση του μοντέλου σε οποιοδήποτε απλό χρήστη.

Βιβλιογραφικές Αναφορές

- [1] C.C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018, σ. VII. ISBN: 9783319944630. URL: <https://books.google.gr/books?id=achqDwAAQBAJ>.
- [2] Christos Troussas, Akrivi Krouska και Maria Virvou. «A multicriteria framework for assessing sentiment analysis in social and digital learning: software review». Στο: *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE. 2018, σσ. 1–7.
- [3] Christos Troussas, Akrivi Krouska και Maria Virvou. «A multicriteria framework for assessing sentiment analysis in social and digital learning: Software». Στο: ()
- [4] Iro Athina Valagkouti, Christos Troussas, Akrivi Krouska, Michalis Feidakis και Cleo Sgouropoulou. «Emotion Recognition in Human–Robot Interaction Using the NAO Robot». Στο: *Computers* 11.5 (2022), σ. 72.
- [5] Christos Troussas, Akrivi Krouska και Maria Virvou. «Evaluation of ensemble-based sentiment classifiers for Twitter data». Στο: *2016 7th International Conference on Information, Intelligence, Systems Applications (IISA)*. 2016, σσ. 1–6. DOI: [10.1109/IISA.2016.7785380](https://doi.org/10.1109/IISA.2016.7785380).
- [6] Christos Troussas, Maria Virvou, Kurt Junshean Espinosa, Kevin Llaguno και Jaime Caro. «Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning». Στο: *IISA 2013*. 2013, σσ. 1–6. DOI: [10.1109/IISA.2013.6623713](https://doi.org/10.1109/IISA.2013.6623713).
- [7] Christos Troussas, Akrivi Krouska και Maria Virvou. «Trends on sentiment analysis over social networks: pre-processing ramifications, stand-alone classifiers and ensemble averaging». Στο: *Machine Learning Paradigms*. Springer, 2019, σσ. 161–186.

- [8] Nalini K Ratha, Vishal M Patel και Rama Chellappa. *Deep Learning-Based Face Analytics*. Springer, 2021.
- [9] Paul Ekman και Wallace V. Friesen. «Measuring facial movement». Στο: *Environmental psychology and nonverbal behavior* 1 (1976), σσ. 56–75.
- [10] Akrivi Krouska, Christos Troussas και Maria Virvou. «Comparative evaluation of algorithms for sentiment analysis over social networking services.» Στο: *J. Univers. Comput. Sci.* 23.8 (2017), σσ. 755–768.
- [11] Akrivi Krouska, Christos Troussas και Maria Virvou. «The effect of preprocessing techniques on Twitter sentiment analysis». Στο: *2016 7th international conference on information, intelligence, systems & applications (IISA)*. IEEE. 2016, σσ. 1–5.
- [12] Akrivi Krouska, Christos Troussas και Maria Virvou. «Deep learning for twitter sentiment analysis: the effect of pre-trained word embedding». Στο: *Machine learning paradigms*. Springer, 2020, σσ. 111–124.
- [13] *Challenges in Representation Learning: Facial Expression Recognition Challenge*. Challenge FER13. URL: <https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/overview>.
- [14] Αλέξανδρος Ζάγκος. *Ανίχνευση εκφράσεων προσώπου με χρήση Βαθιών Συνελικτικών Νευρωνικών Δικτύων*. 2019.
- [15] Ian J. Goodfellow κ.ά. «Challenges in Representation Learning: A report on three machine learning contests». Στο: (2013). DOI: [10 . 48550 / ARXIV . 1307 . 0414](https://doi.org/10.48550/ARXIV.1307.0414). URL: <https://arxiv.org/abs/1307.0414>.
- [16] Yousif Khairuddin και Zhuofa Chen. *Facial Emotion Recognition: State of the Art Performance on FER2013*. 2021. DOI: [10 . 48550 / ARXIV . 2105 . 03588](https://doi.org/10.48550/ARXIV.2105.03588). URL: <https://arxiv.org/abs/2105.03588>.
- [17] Peter Norvig Stuart J. Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2003.
- [18] William L. Hosch. *Machine learning*. 2003. URL: <https://www.britannica.com/technology/machine-learning>.

- [19] Aston Zhang, Zachary C. Lipton, Mu Li και Alexander J. Smola. «Dive into Deep Learning». Στο: *CoRR* abs/2106.11342 (2021). arXiv: 2106.11342. URL: <https://arxiv.org/abs/2106.11342>.
- [20] Ian Goodfellow, Yoshua Bengio και Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [21] Deep Learning sub Field of AI and ML. URL: <https://blog.oursky.com/wp-content/uploads/2020/05/ai-vs-ml-vs-dl-e1588741387626.png>.
- [22] C.C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018. ISBN: 9783319944630. URL: <https://books.google.gr/books?id=achqDwAAQBAJ>.
- [23] Deep Learning Performance. URL: <https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/04/04085751/dl3.png>.
- [24] Ann M Kring, Gerald C Davison, John M Neale και Sheri L Johnson. *Abnormal psychology*. John Wiley & Sons Inc, 2007.
- [25] *Wikipedia*. Biological Neuron. URL: https://upload.wikimedia.org/wikipedia/commons/thumb/0/01/Neuron_el.svg/1920px-Neuron_el.svg.png.
- [26] Simon Haykin. *Νευρωνικά Δίκτυα και Μηχανική Μάθηση*. Αθήνα: Παπασωτηρίου, 2015. ISBN: 9789607182647.
- [27] Georgouli Aikaterini. *Τεχνητή νοημοσύνη*. Αθήνα: Kallipos, 2015. ISBN: 9789606030314. URL: <http://hdl.handle.net/11419/3381>.
- [28] *Overfitting*. URL: <https://www.kaggle.com/code/anantgupt/facial-emotion-recognition-91-accuracy-train-set>.
- [29] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. <https://link.springer.com/book/9780387310732>. Springer New York, NY, 2016. ISBN: 9781493938438.
- [30] Maja Pantic και Leon J. M. Rothkrantz. «Automatic analysis of facial expressions: The state of the art». Στο: *IEEE Transactions on pattern analysis and machine intelligence* 22.12 (2000), σσ. 1424–1445.

- [31] Panagiotis Giannopoulos, Isidoros Perikos και Ioannis Hatzilygeroudis. «Deep learning approaches for facial emotion recognition: A case study on FER-2013». Στο: *Advances in hybridization of intelligent methods*. Springer, 2018, σσ. 1–16.
- [32] Paul Ekman. «Basic emotions». Στο: *Handbook of cognition and emotion* 98.45-60 (1999).
- [33] Gerald L Clore και Andrew Ortony. «Psychological construction in the OCC model of emotion». Στο: *Emotion Review* 5.4 (2013).
- [34] Robert Ed Plutchik και Hope R Conte. *Circumplex models of personality and emotions*. American Psychological Association, 1997.
- [35] Almila Erol, Levent Mete, Ipek Sonmez και Elif Keles Unal. «Facial emotion recognition in patients with schizophrenia and their siblings». Στο: *Nordic Journal of Psychiatry* 64.1 (2010), σσ. 63–67.
- [36] Henrik Kessler, Markus Schwarze, Suzanne Filipic, Harald C Traue και Joern von Wietersheim. «Alexithymia and facial emotion recognition in patients with eating disorders». Στο: *International Journal of Eating Disorders* 39.3 (2006), σσ. 245–251.
- [37] Birgit Derntl, Eva-Maria Seidel, Ilse Kryspin-Exner, Alexander Hasmann και Matthias Dobmeier. «Facial emotion recognition in patients with bipolar I and bipolar II disorder». Στο: *British Journal of Clinical Psychology* 48.4 (2009), σσ. 363–375.
- [38] Madeline B Harms, Alex Martin και Gregory L Wallace. «Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies». Στο: *Neuropsychology review* 20.3 (2010), σσ. 290–322.
- [39] Ruth L Diaz, Ulric Wong, David C Hodgins, Carina G Chiu και Vina M Goghari. «Violent video game players and non-players differ on facial emotion recognition». Στο: *Aggressive behavior* 42.1 (2016), σσ. 16–28.
- [40] Christos D Katsis, Nikolaos Katertsidis, George Ganiatsas και Dimitrios I Fotiadis. «Toward emotion recognition in car-racing drivers: A biosignal processing approach». Στο: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38.3 (2008), σσ. 502–512.

- [41] *Dataset FER - 2013*. Dataset FER13. URL: <https://www.kaggle.com/datasets/msambare/fer2013>.
- [42] Renuka S Deshmukh, Vandana Jagtap και Shilpa Paygude. «Facial emotion recognition system through machine learning approach». Στο: *2017 international conference on intelligent computing and control systems (iciccs)*. IEEE. 2017, σσ. 272–277.
- [43] *Tensorflow*. URL: <https://github.com/tensorflow/tensorflow>.
- [44] *Keras*. URL: <https://github.com/keras-team/keras>.
- [45] *OpenCV*. URL: <https://github.com/opencv/opencv>.
- [46] *Google Colab*. URL: <https://research.google.com/colaboratory/faq.html>.
- [47] *NumPy*. URL: <https://github.com/numpy/numpy>.
- [48] *Streamlit-WebrRTC*. URL: <https://github.com/whitphx/streamlit-webrtc>.