



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**ΑΝΑΠΤΥΞΗ ΕΦΑΡΜΟΓΗΣ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΔΕΡΜΑΤΙΚΩΝ
ΒΛΑΒΩΝ ΜΕ ΧΡΗΣΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Διπλωματική Εργασία

Αθανάσιος Πριμικήρης

Επιβλέπουσα: Χαρίκλεια Τσαλαπάτα

Φεβρουάριος, 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

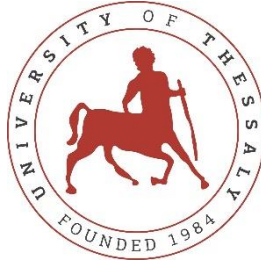
**ΑΝΑΠΤΥΞΗ ΕΦΑΡΜΟΓΗΣ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΔΕΡΜΑΤΙΚΩΝ
ΒΛΑΒΩΝ ΜΕ ΧΡΗΣΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

Διπλωματική Εργασία

Αθανάσιος Πριμικήρης

Επιβλέπουσα: Χαρίκλεια Τσαλαπάτα

Φεβρουάριος, 2024



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

**APP DEVELOPMENT FOR SKIN LESION CLASSIFICATION USING
MACHINE LEARNING**

Diploma Thesis

Athanasios Primikyris

Supervisor: Charikleia Tsalapata

February, 2024

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπουσα

Χαρίκλεια Τσαλαπάτα

Ε.ΔΙ.Π., Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

Αλέξανδρος Χροναίος

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

Γεώργιος Θάνος

Ε.ΔΙ.Π., Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο Δηλών

Πριμικήρης Αθανάσιος

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

The Declarant

Primikyris Athanasios

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω την επιβλέπουσα καθηγήτρια μου κ. Τσαλαπάτα για το χρόνο που αφιέρωσε και τη βοήθεια που μου προσέφερε κατά τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας. Επιπλέον, θα ήθελα να ευχαριστήσω όλους τους καθηγητές του τμήματος για τις γνώσεις και την καθοδήγηση που μου παρείχαν όλα αυτά τα χρόνια των σπουδών μου. Τέλος, ευχαριστώ τους φίλους μου και την οικογένειά μου για τη βοήθεια και την υποστήριξη τους στις δύσκολες στιγμές μου.

Περίληψη

Το αντικείμενο της παρούσας διπλωματικής εργασίας αφορά την ανάπτυξη μοντέλων μηχανικής μάθησης για την κατηγοριοποίηση εικόνων δερματικών βλαβών που μπορεί να προέρχονται είτε από ψηφιακό δερματοσκόπιο, είτε από κάμερα smartphone, στην αντίστοιχη δερματική ασθένεια. Επιπλέον, αναπτύσσεται εφαρμογή σε Android που κάνει χρήση αυτών των μοντέλων με σκοπό την έγκυρη και γρήγορη διάγνωση δερματικών ασθενειών από το χρήστη. Αρχικά, παρουσιάζονται οι βάσεις δεδομένων που χρησιμοποιήθηκαν για την ανάπτυξη των μοντέλων. Στη συνέχεια, περιγράφονται αναλυτικά οι αρχιτεκτονικές των νευρωνικών δικτύων που χρησιμοποιήθηκαν για την αναγνώριση και ταξινόμηση των εικόνων. Επιπρόσθετα, αξιοποιούνται μεταδεδομένα που αφορούν την εκάστοτε δερματική πληγή, αυξάνοντας τις επιδόσεις των μοντέλων. Τέλος, αφού παρουσιαστούν τα αποτελέσματα για κάθε βάση, γίνεται περιγραφή της εφαρμογής Android και του τρόπου ανάπτυξής της.

Abstract

The subject area of the particular thesis is the development of machine learning models for the classification of skin lesion images that may originate from either digital dermoscopy, or from a smartphone camera, to the corresponding skin disease. Moreover, an Android application is developed that utilizes these models, aiming at the valid and rapid diagnosis of skin diseases by the user. Initially, the databases that were used for the development of the models are presented. Then, the architectures of the neural networks that were used are described in detail for the identification and classification of the images. In addition, metadata that are related to each skin lesion are used in order to increase the models' performance. Finally, after the presentation of the results for each database, a description of the Android application and its development process is provided.

Πίνακας περιεχομένων

<i>Ευχαριστίες.....</i>	<i>ix</i>
<i>Περίληψη.....</i>	<i>x</i>
<i>Abstract</i>	<i>xii</i>
<i>Κατάλογος εικόνων</i>	<i>xv</i>
<i>Κατάλογος σχημάτων</i>	<i>xvi</i>
<i>Κατάλογος πινάκων</i>	<i>xvii</i>
<i>Συντομογραφίες.....</i>	<i>xviii</i>
Κεφάλαιο 1 Εισαγωγή	1
1.1 Εισαγωγικές πληροφορίες	1
1.2 Αντικείμενο της διπλωματικής	4
1.3 Σχετικές εργασίες.....	5
1.4 Οργάνωση της διπλωματικής εργασίας.....	6
Κεφάλαιο 2 Βάσεις δεδομένων	7
2.1 ISIC 2019.....	7
2.2 PAD-UFES-20.....	11
Κεφάλαιο 3 Μέθοδοι αξιολόγησης και αρχιτεκτονικές νευρωνικών δικτύων	14
3.1 Μετρικές αξιολόγησης	14
3.2 Νευρωνικά δίκτυα	18
3.2.1 Resnet34	18
3.2.2 MobileNet	21
3.2.3 GoogleNet	25
3.2.4 EfficientNet	27
3.2.5 MobileNet + Self attention layers	29
3.2.6 Συγχώνευση μοντέλων	31
3.3 Μεταφορά μάθησης (Transfer learning)	31
Κεφάλαιο 4 Εκπαίδευση μοντέλων και αποτελέσματα	33
4.1 Επαύξηση δεδομένων	33
4.2 Εξισορρόπηση δεδομένων	34

4.3 Προσδιορισμός υπερ-παραμέτρων	35
4.4 Χρήση μεταδεδομένων	36
4.5 Επεξεργασία και κωδικοποίηση πεδίων	37
4.6 Αρχιτεκτονική νευρωνικού δικτύου με μεταδεδομένα.....	39
4.7 Εκπαίδευση και αποτελέσματα στη βάση PAD-UFES-20.....	40
4.7.1 Διαχωρισμός των δεδομένων σε δεδομένα εκπαίδευσης / αξιολόγησης	40
4.7.2 Εκπαίδευση και αποτελέσματα μόνο με εικόνες	41
4.7.3 Ανάλυση των μεταδεδομένων της βάσης	42
4.7.4 Εκπαίδευση και αποτελέσματα με χρήση μεταδεδομένων	44
4.7.5 Επιλογή μοντέλου για χρήση στην εφαρμογή.....	45
4.8 Εκπαίδευση και αποτελέσματα στη βάση ISIC-2019.....	46
4.8.1 Διαχωρισμός των δεδομένων σε εκπαίδευσης / αξιολόγησης	46
4.8.2 Εκπαίδευση και αποτελέσματα μόνο με εικόνες	47
4.8.3 Ανάλυση των μεταδεδομένων της βάσης	48
4.8.4 Εκπαίδευση και αποτελέσματα με χρήση μεταδεδομένων	49
4.8.5 Αποτελέσματα στο σύνολο ελέγχου (test set)	51
Κεφάλαιο 5 Ανάπτυξη εφαρμογής.....	52
5.1 Παρουσίαση της δομής της εφαρμογής.....	52
5.2 Παρουσίαση του τρόπου ανάπτυξης της εφαρμογής.....	55
5.2.1 Ανάπτυξη κύριας οθόνης (MainScreen)	55
5.2.2 Ανάπτυξη οθόνης μεταδεδομένων (GetMetadataScreen).....	56
5.2.3 Ανάπτυξη της οθόνης αποτελεσμάτων (ResultsScreen).....	57
5.2.4 Ανάπτυξη νευρωνικού δικτύου στην εφαρμογή	57
5.2.5 Ανίχνευση μη έγκυρων εικόνων	58
Κεφάλαιο 6 Συμπεράσματα και προτάσεις	60
6.1 Συμπεράσματα	60
6.2 Μελλοντικές επεκτάσεις	61
Βιβλιογραφία.....	62

Κατάλογος εικόνων

Εικόνα 2.1 Κατανομή των τιμών του πεδίου age_approx για τη βάση ISIC	9
Εικόνα 2.2 Κατανομή των τιμών του πεδίου anatom_site_general για τη βάση ISIC.....	9
Εικόνα 2.3 Κατανομή των τιμών του πεδίου gender για τη βάση ISIC.....	10
Εικόνα 2.4 Κατανομή των τιμών του πεδίου category για τη βάση ISIC	10
Εικόνα 2.5 Κατανομή των τιμών του πεδίου diagnostic για τη βάση PAD-UFES-20	13
Εικόνα 4.1 Το confusion matrix για το MobileNet + Sa χωρίς εικόνες	42
Εικόνα 4.2 Confusion Matrix για το MobileNet + SA με τη χρήση μεταδεδομένων.....	45
Εικόνα 4.3 Confusion Matrix για το Ensemble Model χωρίς τη χρήση μεταδεδομένων ...	48
Εικόνα 4.4 Confusion Matrix για το Ensemble Model με τη χρήση μεταδεδομένων.....	50
Εικόνα 5.1 Η αρχική οθόνη της εφαρμογής.....	53
Εικόνα 5.2 Η οθόνη με τα πεδία των μεταδεδομένων	54
Εικόνα 5.3 Η οθόνη με τα αποτελέσματα	55

Κατάλογος σχημάτων

Σχήμα 3.1 Η δομή ενός υπολειμματικού μπλοκ. Σχήμα από [41]	19
Σχήμα 3.2 Η αρχιτεκτονική του Resnet34. Σχήμα από [41]	20
Σχήμα 3.3 Η αρχιτεκτονική μιας depthwise separable συνέλιξης. Σχήμα από [43]	22
Σχήμα 3.4 Η δομή του inverted residual block. Σχήμα από [44].....	23
Σχήμα 3.5 Η δομή του squeeze-excitation μπλοκ. Σχήμα από [46]	24
Σχήμα 3.6 Το inception block. Εικόνα από [47]	25
Σχήμα 3.7 Διαφορά ανάμεσα σε MBConv και Fused-MBConv. Σχήμα από [49].....	28
Σχήμα 3.8 Το self attention block. Σχήμα από [52]	29
Σχήμα 4.1 Η αρχιτεκτονική του metaModel	39
Σχήμα 5.1 Η ροή λειτουργίας της εφαρμογής	52

Κατάλογος πινάκων

Πίνακας 3.1 Η αρχιτεκτονική του MobileNet v3 Large. Πίνακας από [45]	25
Πίνακας 3.2 Η αρχιτεκτονική του GoogleNet. Πίνακας από [47]	26
Πίνακας 3.3 Η αρχιτεκτονική του EfficientNet B0. Πίνακας από [48]	27
Πίνακας 3.4 Η αρχιτεκτονική του EfficientNet v2 Small. Πίνακας από [49]	29
Πίνακας 4.1 Αποτελέσματα πειραμάτων στη βάση PAD μόνο με εικόνες.....	41
Πίνακας 4.2 Συσχετίσεις του πεδίου diagnostic με τα υπόλοιπα πεδία της βάσης PAD ...	43
Πίνακας 4.3 Αποτελέσματα πειραμάτων στη βάση PAD με τη χρήση μεταδεδομένων	44
Πίνακας 4.4 Αριθμός παραμέτρων των μοντέλων.....	46
Πίνακας 4.5 Αποτελέσματα πειραμάτων στη βάση ISIC χωρίς τη χρήση μεταδεδομένων	47
Πίνακας 4.6 Συσχετίσεις του πεδίου category με τα υπόλοιπα πεδία της βάσης ISIC	49
Πίνακας 4.7 Αποτελέσματα πειραμάτων στη βάση ISIC με τη χρήση μεταδεδομένων.....	49
Πίνακας 4.8 Αποτελέσματα Ensemble Model στο σύνολο ελέγχου.....	51

Συντομογραφίες

AK	Actinic Keratosis
BCC	Basal Cell Carcinoma
BKL	Benign keratosis
DF	Dermatofibroma
MEL	Melanoma
ReLu	Rectified Linear Unit
SCC	Squamous Cell Carcinoma
SA	Self Attention
SEK	Seborrheic Keratosis
NV	Nevus
VASC	Vascular lesion

Κεφάλαιο 1 Εισαγωγή

Η εξέλιξη της επιστήμης των υπολογιστών και ιδιαίτερα της μηχανικής μάθησης έχει οδηγήσει σε επαναστατικές μεθόδους που αφορούν την επιστήμη της ιατρικής, βοηθώντας στην πρόληψη, διάγνωση και θεραπεία διάφορων παθήσεων. Κάποιες από τις πιο χαρακτηριστικές εφαρμογές της μηχανικής μάθησης στην ιατρική σχετίζονται με τη χρήση εφαρμογών ιατρικής απεικόνισης όπως οι ακτινογραφίες (X-rays) [1], η μαγνητική πυρηνική απεικόνιση (MRI) [2] και η υπολογιστική τομογραφία (CT scan) [3], συμβάλλοντας στη βελτίωση και αύξηση της ταχύτητας και ακρίβειας των διαγνώσεων. Μία ακόμα χαρακτηριστική εφαρμογή και συμβολή της μηχανικής μάθησης στην ιατρική εντοπίζεται στον τομέα της ρομποτικής ιατρικής, όπου χρησιμοποιούνται ρομπότ σε επεμβάσεις με εξαιρετική ακρίβεια, ελαττώνοντας τις βλάβες στους ιστούς και μειώνοντας τον χρόνο ανάρρωσης και αποκατάστασης του ασθενούς [4]. Επιπρόσθετα, η μηχανική μάθηση εφαρμόζεται στον τομέα των εξατομικευμένων θεραπειών λαμβάνοντας υπόψη το ιατρικό προφίλ του κάθε ασθενούς, όσον αφορά προηγούμενες ιατρικές παθήσεις και τη γενετική του προδιάθεση, παρέχοντας θεραπείες κατάλληλες και προσαρμοσμένες στις ιατρικές του ανάγκες [5]. Τέλος, η μηχανική μάθηση μπορεί να παίζει καθοριστικό ρόλο στην πρόληψη ασθενειών αναλύοντας τις ιατρικές ενδείξεις του εκάστοτε ασθενούς και κάνοντας προβλέψεις για πιθανές παθήσεις, αποτελώντας με αυτό τον τρόπο ένα εξαιρετικά χρήσιμο εργαλείο για την ιατρική.

1.1 Εισαγωγικές πληροφορίες

Σε αυτή την ενότητα θα γίνει αναφορά και περιγραφή κάποιων βασικών εννοιών και ορολογιών αναφορικά με τη δερματολογία, που είναι απαραίτητες για την κατανόηση του αντικειμένου και του σκοπού της παρούσας εργασίας.

Δερματοσκόπηση:

Η δερματοσκόπηση [6], είναι μία μη επεμβατική μέθοδος διάγνωσης, η οποία χρησιμοποιείται από τους δερματολόγους για την εξέταση ύποπτων δερματικών βλαβών με μεγαλύτερη λεπτομέρεια. Η τεχνική πραγματοποιείται με τη χρήση ενός ειδικού

μικροσκοπίου (δερματοσκόπιο), το οποίο περιλαμβάνει έναν μεγεθυντικό φακό και μία πηγή φωτός. Επιπλέον, τα περισσότερα συστήματα δερματοσκόπησης κάνουν χρήση ψηφιακών μεθόδων για την αποθήκευση και περαιτέρω ανάλυση των εικόνων.

Η εξέταση της δερματοσκόπησης δίνει τη δυνατότητα στους ειδικούς να μελετήσουν την εκάστοτε βλάβη με μεγαλύτερη λεπτομέρεια, ανιχνεύοντας χαρακτηριστικά και δομές του δέρματος που δεν είναι ορατές με γυμνό μάτι. Με αυτό το τρόπο ενισχύεται η ακρίβεια των διαγνώσεων και επομένως προωθείται η πρόληψη των δερματικών ασθενειών.

Στη συνέχεια, γίνεται περιγραφή όλων των δερματικών βλαβών που θα εξεταστούν στην εργασία:

Ακτινική κεράτωση (AK):

Η ακτινική ή ηλιακή κεράτωση [7] οφείλεται στην παρατεταμένη έκθεση του δέρματος στην ηλιακή ακτινοβολία. Εξωτερικά η πληγή μοιάζει με μια επίπεδη τραχιά λεπιδώδη πλάκα και το χρώμα της μπορεί να είναι κόκκινο, καφέ ή στο χρώμα του δέρματος. Η βλάβη αυτή αν δεν αντιμετωπιστεί εγκαίρως, έχει τη δυνατότητα να εξελιχθεί σε καρκίνο του δέρματος.

Βασικοκυτταρικό καρκίνωμα (BCC):

Το βασικοκυτταρικό καρκίνωμα [8] αποτελεί τη συχνότερη μορφή καρκίνου που μπορεί να εμφανιστεί στο δέρμα και αναπτύσσεται κυρίως σε σημεία που έχουν εκτεθεί παρατεταμένα στον ήλιο, κυρίως στο πρόσωπο, το κεφάλι και τα χέρια. Η βλάβη αυτή εμφανίζεται συνήθως ως ένας βραδέως αναπτυσσόμενος όγκος που σπάνια παρουσιάζει μεταστάσεις, ωστόσο μπορεί να γίνει εξαιρετικά καταστροφικός για το συγκεκριμένο σημείο του σώματος. Η κύρια προσέγγιση για την αντιμετώπιση του BCC περιλαμβάνει συνήθως χειρουργικές επεμβάσεις όπως η εκτομή.

Καλοήθης κεράτωση (BKL) ή Σμηγματορροϊκή κεράτωση (SEK):

Η Σμηγματορροϊκή κεράτωση [9] είναι συχνός καλοήθης όγκος του δέρματος. Οι βλάβες αυτές προκύπτουν από τον πολλαπλασιασμό των επιδερμικών κυττάρων και παρουσιάζονται τυπικά σε άτομα μεγαλύτερης ηλικίας. Δεν ενέχουν κανέναν κίνδυνο εξέλιξης σε κακοήθεια, ωστόσο η κλινική εικόνα της βλάβης πολλές φορές συγχέεται με

άλλους κακοήθεις όγκους του δέρματος. Οι όγκοι αυτοί αφαιρούνται κυρίως όταν υπάρχει αμφιβολία για τη διάγνωση ή για αισθητικούς λόγους.

Δερματοϊνίωμα (DF):

Το δερματοϊνίωμα [10] είναι μια κοινή καλοήθης βλάβη που πολλές φορές μοιάζει με σπίλο (ελιά). Οι βλάβες αυτές είναι συνήθως μικρές σε μέγεθος, και χωρίς κάποιο σύμπτωμα. Εμφανίζονται πιο συχνά στις γυναίκες, παρά στους άνδρες. Τα δερματοϊνιώματα δεν απαιτούν κάποια θεραπεία και κυρίως αφαιρούνται για αισθητικούς λόγους.

Μελάνωμα (MEL):

Το μελάνωμα [11] θεωρείται η πιο επικίνδυνη μορφή καρκίνου του δέρματος, καθώς ευθύνεται για τους περισσότερους θανάτους που σχετίζονται με ασθένειες του δέρματος. Χαρακτηριστικά αυτής της βλάβης είναι το ασύμμετρο σχήμα, η ανώμαλη και γρήγορη εξάπλωση και η αλλαγή χρώματος. Το μελάνωμα στα αρχικά στάδια συνήθως αντιμετωπίζεται επιτυχώς με χειρουργική επέμβαση, ωστόσο, εάν υπάρξει μετάσταση τα ποσοστά επιβίωσης μειώνονται. Συνεπώς, είναι σαφές ότι η έγκαιρη διάγνωση είναι πολύ σημαντική ώστε να διασφαλιστεί ότι ο ασθενής θα λάβει την καλύτερη δυνατή θεραπεία. Σε περίπτωση μετάστασης, η κύρια προσέγγιση θεραπείας είναι η χρήση φαρμακευτικών ουσιών και η εφαρμογή χημειοθεραπειών. Τα τελευταία χρόνια η ανακάλυψη πρωτοποριακών θεραπειών έχε βελτιώσει σημαντικά την αντιμετώπιση του μεταστατικού μελανώματος, ωστόσο, η ανάγκη για τη βελτίωση αυτών των θεραπειών παραμένει επιτακτική.

Ακανθοκυτταρικό καρκίνωμα (SCC):

Το ακανθοκυτταρικό καρκίνωμα (SCC) [12] είναι η δεύτερη πιο συχνή δερματική κακοήθεια μετά το βασικοκυτταρικό καρκίνωμα. Η βασική αιτία ανάπτυξης της βλάβης είναι η χρόνια έκθεση στην ηλιακή ακτινοβολία. Εξωτερικά, η πληγή εκδηλώνεται συνήθως ως μία παχιά λεπιδώδης πληγή η οποία δύναται να αιμορραγήσει. Σε περίπτωση που δεν αντιμετωπιστεί εγκαίρως, το SCC μπορεί να εξαπλωθεί (μετάσταση) και σε άλλα μέρη του σώματος, δημιουργώντας σοβαρά προβλήματα υγείας. Η δερματοσκόπηση και η χειρουργική εκτομή αποτελούν τις κύριες προσεγγίσεις για τη διάγνωση και τη θεραπεία του SCC.

Μελανοκυτταρικοί σπίλοι(NV):

Ως μελανοκυτταρικοί σπίλοι [13] ορίζονται οι καλοήθεις πολλαπλασιασμοί μελανοκυττάρων (κυττάρων που παράγουν μελανίνη). Πρόκειται για τους πιο συχνούς καλοήθεις όγκους δέρματος και είναι γνωστοί με τον όρο «ελιά».

Αγγειακές δερματικές βλάβες (VASC):

Οι αγγειακές δερματικές βλάβες [14] περιλαμβάνουν τα αιμαγγειώματα, τις φλεβικές λίμνες και τις ευρυαγγείες. Σχηματίζονται λόγω της διάταξης μιας ομάδας αγγείων που βρίσκονται πολύ κοντά στην επιφάνεια του δέρματος. Παρόλο που μπορούν να εμφανιστούν οπουδήποτε στο σώμα, οι συγκεκριμένες βλάβες είναι πιο συχνές στο πρόσωπο και στα πόδια. Η θεραπεία που πρέπει να εφαρμοστεί ποικίλει αναλόγως τα χαρακτηριστικά κάθε βλάβης, ωστόσο συνήθως οι αγγειακές βλάβες δεν συνιστούν άμεσο κίνδυνο.

1.2 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία επικεντρώνεται στην πρόληψη και διάγνωση δερματικών ασθενειών με χρήση τεχνικών μηχανικής μάθησης, βασισμένων στην ανάλυση και επεξεργασία δεδομένων εικόνας. Συγκεκριμένα, έχουν αναπτυχθεί μοντέλα νευρωνικών δικτύων για την αναγνώριση δερματικών ασθενειών από εικόνα που προέρχεται από δερματοσκοπική λήψη, αλλά και εικόνας απευθείας από κάμερα.

Για την εκπαίδευση των μοντέλων που αφορούν τις δερματοσκοπικές εικόνες αξιοποιήθηκε η βάση δεδομένων ISIC 2019 [15], ενώ για την ανάπτυξη των μοντέλων που αφορούν τις εικόνες που προέρχονται από κάμερες κινητού, η PAD-UFES-20 [16]. Για το πρόβλημα της αναγνώρισης και κατηγοριοποίησης των εικόνων εξετάστηκαν διάφορες αρχιτεκτονικές νευρωνικών δικτύων, καθώς επίσης και συνδυασμοί των συγκεκριμένων δικτύων (ensemble learning).

Επιπλέον, πέραν της εικόνας, αξιοποιούνται κι άλλα χαρακτηριστικά της δερματικής βλάβης, που αφορούν λεπτομέρειες για τη συμπεριφορά της (κνησμό, αιμορραγία, ανομοιόμορφη εξάπλωση, αλλαγή χρώματος), καθώς επίσης και λεπτομέρειες που αφορούν τον εκάστοτε ασθενή (φύλο, ηλικία).

Τέλος, έχει αναπτυχθεί εφαρμογή σε android, η οποία προσφέρει τη δυνατότητα στον χρήστη να φωτογραφήσει ή να ελέγξει μια προϋπάρχουσα εικόνα δερματικής βλάβης, ώστε να του παρέχει μια πρώτη γρήγορη και έγκυρη διάγνωση, κάνοντας χρήση ενός νευρωνικού δικτύου που αναπτύχθηκε για τη βάση δεδομένων PAD-UFES-20.

1.3 Σχετικές εργασίες

Η εξέλιξη των αλγορίθμων τεχνητής νοημοσύνης και βαθιάς μάθησης έχει προσφέρει πολλά εργαλεία σε ερευνητές ώστε να αναπτύξουν έξυπνα συστήματα για την αναγνώριση και κατηγοριοποίηση δερματικών βλαβών. Μία από τις πρώτες προσεγγίσεις που πετυχαίνει υψηλές αποδόσεις στην κατηγοριοποίηση των βλαβών εξετάζει αρχιτεκτονικές προεκπαιδευμένων συνελκτικών δικτύων που χρησιμοποιούνται είτε απευθείας για ταξινόμηση [17], [18], [19], είτε για εξαγωγή χαρακτηριστικών [20], [21], [22]. Άλλες προσεγγίσεις περιλαμβάνουν τη χρήση συγχωνευμένων συνελκτικών δικτύων (Ensemble Models) για τη κατηγοριοποίηση των εικόνων, παρουσιάζοντας αύξηση στην απόδοση [23], [24]. Επιπλέον, έχουν μελετηθεί τεχνικές ενσωμάτωσης Soft Attention νευρωνικών επιπέδων πάνω σε προεκπαιδευμένα συνελκτικά νευρωνικά δίκτυα, επιτυγχάνοντας αυξημένη απόδοση στη ταξινόμηση [25]. Μία άλλη ενδιαφέρουσα προσέγγιση που έχει αποφέρει εξαιρετικά αποτελέσματα χρησιμοποιώντας πολύ μικρό αριθμό παραμέτρων, είναι η ανάπτυξη νευρωνικών δικτύων με βασικό στοιχείο τα Capsule Networks [26], [27]. Επιπρόσθετα, υψηλές επιδόσεις έχουν παρατηρηθεί από συνδυασμούς συνελκτικών νευρωνικών δικτύων με Transformers, κάνοντας χρήση αλγορίθμων εκπαίδευσης αυτό-επίβλεψης (self-supervised) [28].

Άλλες έρευνες έχουν εστιάσει στην προεπεξεργασία των εικόνων και την επαύξηση των δεδομένων χρησιμοποιώντας GAN νευρωνικά δίκτυα, πετυχαίνοντας αύξηση στην επίδοση των μοντέλων ταξινόμησης [29]. Επιπλέον, μελετήθηκε η επίδραση των διάφορων μετασχηματισμών που μπορούν να εφαρμοστούν πάνω στις εικόνες κατά τη διαδικασία της εκπαίδευσης, καταδεικνύοντας έτσι τη σημαντικότητα του σταδίου της προεπεξεργασίας [30].

Επιπρόσθετα, πολλές μελέτες πραγματεύονται το πρόβλημα της χρήσης μεταδεδομένων στα μοντέλα. Τα μεταδεδομένα είναι δεδομένα που περιλαμβάνουν επιπλέον πληροφορίες για την δερματική πληγή που απεικονίζεται στην εικόνα. Προσεγγίσεις για

την αξιοποίηση αυτών των δεδομένων περιλαμβάνουν την χρήση μηχανισμών συσχέτισης, όπως η εφαρμογή νευρωνικών επιπέδων αυτοσυσχέτισης (self attention) και αλληλοσυσχέτισης (cross-attention) για τον συνδυασμό των πληροφοριών εικόνας και δεδομένων [31], [32], αλλά και την εφαρμογή μονοδιάστατων συνελίξεων πάνω στα μεταδεδομένα [33]. Επιπρόσθετα, μελετώνται τεχνικές συνδυασμού των δύο παραπάνω μεθόδων [34]. Σημαντική συνεισφορά επίσης, είχε η μελέτη της χρήσης γραμμικών επιπέδων πάνω σε συνελικτικά δίκτυα, με σκοπό να εξαχθούν χαρακτηριστικά της εικόνας που έπειτα θα συνενωθούν με τα δεδομένα [35].

1.4 Οργάνωση της διπλωματικής εργασίας

Η δομή της υπόλοιπης εργασίας έχει ως εξής:

- Κεφάλαιο 2: Γίνεται περιγραφή και ανάλυση των βάσεων δεδομένων που χρησιμοποιήθηκαν.
- Κεφάλαιο 3: Γίνεται αναλυτική παρουσίαση των αρχιτεκτονικών των νευρωνικών δικτύων που χρησιμοποιήθηκαν, καθώς επίσης και των μετρικών αξιολόγησης που εφαρμόστηκαν .
- Κεφάλαιο 4: Γίνεται παρουσίαση του τρόπου εκπαίδευσης των νευρωνικών δικτύων για κάθε βάση, καθώς επίσης και σχολιασμός των αποτελεσμάτων και εξαγωγή συμπερασμάτων από αυτά.
- Κεφάλαιο 5: Γίνεται αναλυτική παρουσίαση του τρόπου δημιουργίας, αλλά και χρήσης της εφαρμογής που αναπτύχθηκε.
- Κεφάλαιο 6: Παρουσιάζεται η σύνοψη της διπλωματικής εργασίας και γίνεται αναφορά σε πιθανές κατευθύνσεις έρευνας για την περεταίρω βελτίωση της αξιοπιστίας και ακρίβειας των νευρωνικών δικτύων και της εφαρμογής.

Κεφάλαιο 2 Βάσεις δεδομένων

2.1 ISIC 2019

Το ISIC (International Skin Imaging Collaboration) είναι ένας οργανισμός ερευνητών, με σκοπό τη συλλογή εικόνων από δερματικές πληγές και την ανάλυση τους, χρησιμοποιώντας τεχνικές επεξεργασίας εικόνας και όρασης υπολογιστών.

Η ανάπτυξη αξιόπιστων υπολογιστικών εργαλείων για την κατανόηση των δερματολογικών εικόνων μπορεί να παίξει σημαντικό ρόλο στη γρήγορη και έγκυρη διάγνωση δερματικών ασθενειών, καθώς επίσης και στη μείωση αχρείαστων βιοψιών, επιτυγχάνοντας με αυτό το τρόπο την αύξηση της αποδοτικότητας του συστήματος υγείας όσον αφορά την πρόληψη και διάγνωση.

Το αρχείο του ISIC αποτελεί τη μεγαλύτερη δημοσίως διαθέσιμη συλλογή δερματικών εικόνων, καθώς περιλαμβάνει πάνω από 76.108 δερματοσκοπικές εικόνες, που συλλέχθηκαν από διάφορες συνεργαζόμενες κλινικές παγκοσμίως. Οι εικόνες έχουν υποστεί προεπεξεργασία και κατηγοριοποίηση από ειδικούς δερματολόγους.

Το ISIC διοργανώνει διάφορους διαγωνισμούς με θέμα τη ταξινόμηση και τμηματοποίηση (segmentation) εικόνων δερματικών βλαβών. Οι διαγωνισμοί αυτοί δημιουργούν προϋποθέσεις για τη συνεργασία και ανταλλαγή πληροφορίας από ερευνητές παγκοσμίως. Συνεπώς, με αυτό το τρόπο προάγεται η παραγωγή έρευνας και επομένως, η ανάπτυξη σύγχρονων τεχνικών και μεθόδων με εφαρμογές στο κλάδο της δερματολογίας. Ο πρώτος διαγωνισμός πραγματοποιήθηκε το 2016 και έκτοτε γίνονται ετησίως μέχρι και το 2020.

Η συγκεκριμένη διπλωματική εργασία επικεντρώνεται στο διαγωνισμό του 2019 [15]. Στόχος του διαγωνισμού είναι η ανάπτυξη μοντέλων για τη σωστή ταξινόμηση εικόνων δερματικών πληγών που έχουν ληφθεί με δερματοσκόπιο.

Οι κατηγορίες στις οποίες μπορεί να ανήκει μία εικόνα είναι οκτώ(8) και παρατίθενται παρακάτω:

- Actinic Keratosis (Ακτινική Κεράτωση), συντομογραφία: AK.

- Basal Cell Carcinoma (Βασιλοκυτταρικό Καρκίνωμα), συντομογραφία: BCC.
- Benign keratosis (Καλοήθης Κεράτωση), συντομογραφία: BKL.
- Dermatofibroma (Δερματοΐνωμα), συντομογραφία: DF.
- Melanocytic Nevi (Μελανοκυτταρικοί Σπίλοι), συντομογραφία: NV.
- Melanoma (μελάνωμα), συντομογραφία: MEL.
- Squamous Cell Carcinoma (Ακανθοκυτταρικό Καρκίνωμα), συντομογραφία: SCC.
- Vascular lesions (Αγγειακές Βλάβες), συντομογραφία: VASC.

Η βάση δεδομένων που χρησιμοποιήθηκε αποτελείται από τρία υποσύνολα βάσεων:

HAM10000 [36], BCN20000 [37] και MSK Dataset [38].

Συνολικά, το σύνολο δεδομένων περιλαμβάνει 25.331 εικόνες. Στη βάση, εκτός από τη κατηγορία στην οποία αντιστοιχεί κάθε εικόνα (πεδίο: category), προσφέρονται επιπλέον δεδομένα (μεταδεδομένα) που παρέχουν πληροφορίες για τη δερματική πληγή και τον ασθενή.

Συγκεκριμένα, περιλαμβάνονται:

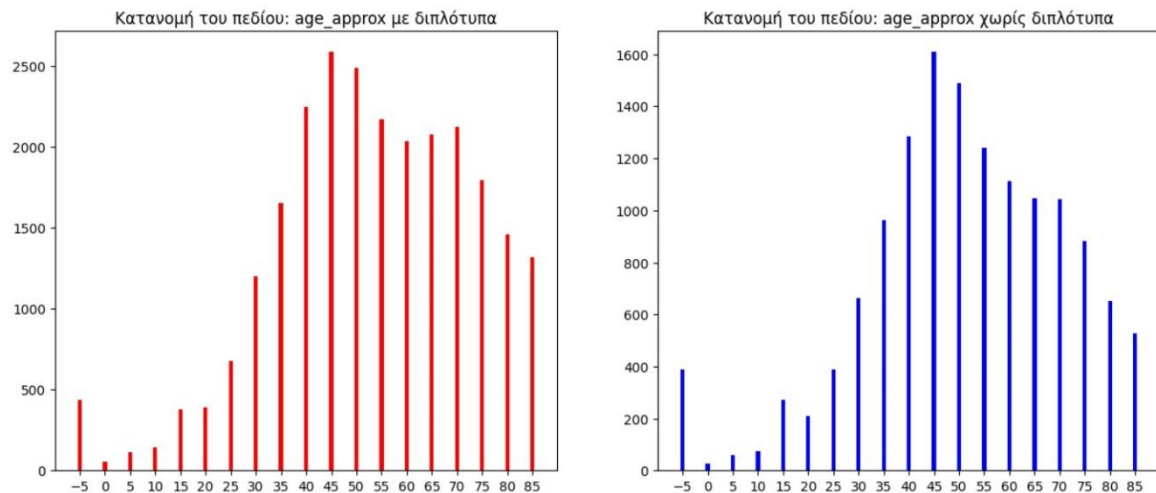
- Πεδίο age_approx: Η ηλικία του ασθενούς.
- Πεδίο anatom_site_general: Σημείο του σώματος στο οποίο βρίσκεται η πληγή.
- Πεδίο gender: Το φύλο του ασθενούς.

Επιπλέον, είναι πολύ σημαντικό να αναφερθεί ότι κάποιες εικόνες από τη βάση απεικονίζουν την ίδια δερματική πληγή. Οι εικόνες αυτές μπορεί να έχουν ληφθεί από διαφορετική γωνία, διαφορετικό φωτισμό, είτε σε διαφορετική χρονική στιγμή. Συνολικά στη βάση ο αριθμός των ξεχωριστών πληγών είναι 13.931.

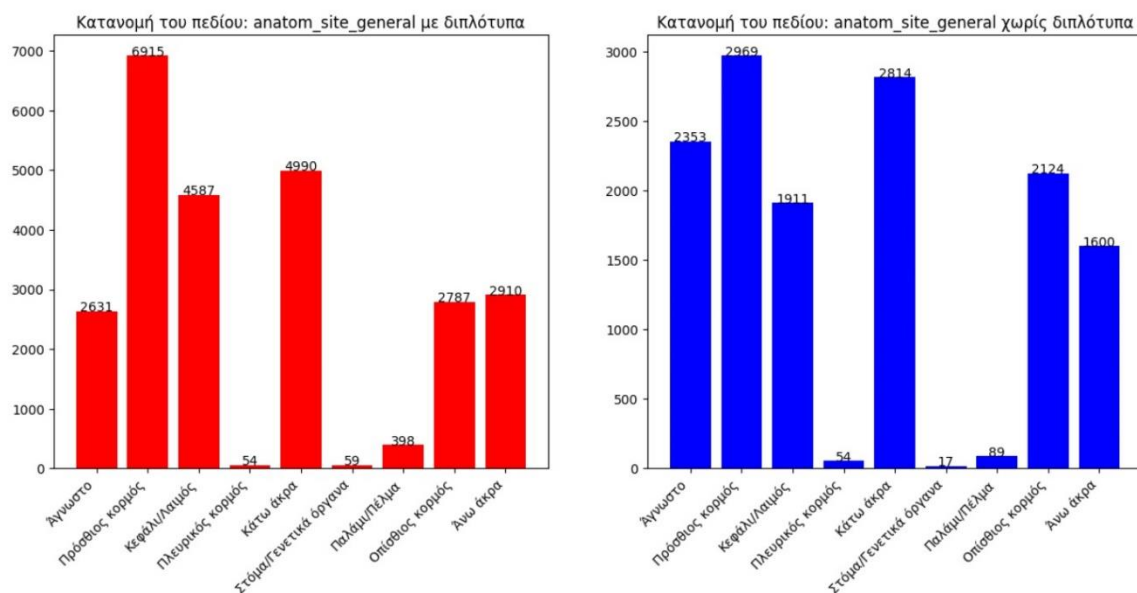
Στη συνέχεια παρουσιάζεται μια συνοπτική διερεύνηση πάνω στη δομή των δεδομένων. Επειδή υπάρχουν πολλές εικόνες που αντιστοιχούν στην ίδια πληγή (και επομένως τα μεταδεδομένα θα είναι τα ίδια), η ανάλυση θα γίνει με και χωρίς διπλότυπα δεδομένα.

Στις ακόλουθες τέσσερις εικόνες παρουσιάζονται οι κατανομές των τιμών για τα μεταδεδομένα, αλλά και για το πεδίο της κατηγορίας. Αυτές οι αναπαραστάσεις θα μας βοηθήσουν να λάβουμε μια καλύτερη εικόνα της δομής της βάσης. Στα αριστερά (κόκκινες τιμές) παρουσιάζεται η κατανομή του εκάστοτε πεδίου με διπλότυπα, ενώ στα δεξιά (μπλε

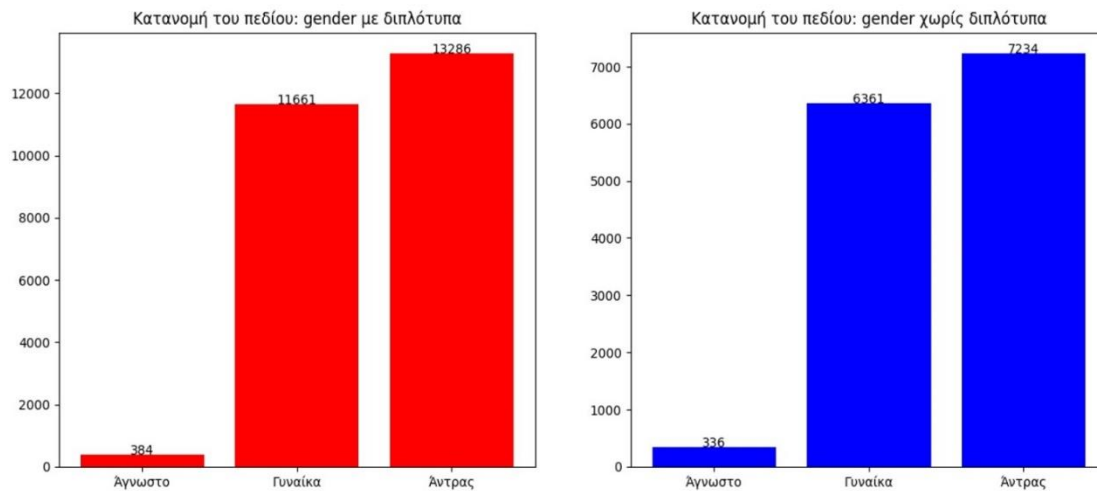
τιμές) παρουσιάζεται η κατανομή του πεδίου χωρίς διπλότυπα. Σημειώνεται πως η τιμή -5 στο πεδίο age_approx αντιστοιχεί σε άγνωστες τιμές.



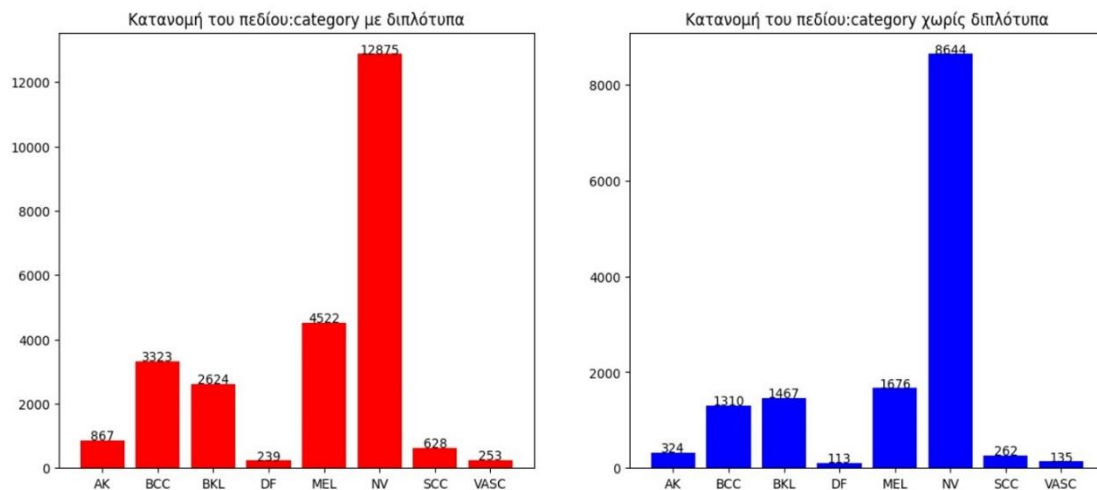
Εικόνα 2.1 Κατανομή των τιμών του πεδίου age_approx για τη βάση ISIC



Εικόνα 2.2 Κατανομή των τιμών του πεδίου anatom_site_general για τη βάση ISIC



Εικόνα 2.3 Κατανομή των τιμών του πεδίου gender για τη βάση ISIC



Εικόνα 2.4 Κατανομή των τιμών του πεδίου category για τη βάση ISIC

Παρατηρείται ότι η κατανομή των κατηγοριών των δεδομένων μας είναι imbalanced, καθώς κάποιες κατηγορίες εμφανίζονται πολύ συχνότερα από κάποιες άλλες. Αυτό μπορεί να προκαλέσει πρόβλημα, καθώς οι κατηγορίες με τις περισσότερες εικόνες θα παρουσιάζονται πολύ συχνότερα κατά τη διάρκεια της εκπαίδευσης, με αποτέλεσμα τα νευρωνικά δίκτυα να μην μπορούν να μοντελοποιήσουν σωστά τις κατηγορίες με τα λιγότερα δεδομένα. Επομένως, είναι σημαντικό να εφαρμοστούν τεχνικές εξισορρόπησης (balancing) για την αντιμετώπιση αυτού του προβλήματος (oversampling, undersampling, weighted loss training) [39].

2.2 PAD-UFES-20

Τα μοντέλα που έχουν αναπτυχθεί πάνω σε βάσεις με δερματοσκοπικές εικόνες δεν αποτελούν πάντα ένα εύχρηστο εργαλείο ανάλυσης δερματικών βλαβών, καθώς για να χρησιμοποιηθούν θα πρέπει να γίνει λήψη μιας δερματοσκοπικής εικόνας της πληγής, κάτι που προϋποθέτει κατάλληλο εξοπλισμό. Ωστόσο, αυτός ο εξοπλισμός δεν μπορεί να είναι εύκολα διαθέσιμος σε ανθρώπους που δεν είναι ειδικοί. Το ιδανικό θα ήταν να αναπτυχθούν μοντέλα που θα λειτουργούν ικανοποιητικά με εικόνες που έχουν ληφθεί απευθείας από κάμερα, χωρίς να απαιτείται ειδικός εξοπλισμός ή κάποια προεπεξεργασία, με στόχο την εύκολη και άμεση χρήση τους από όλο τον κόσμο.

Η PAD-UFES-20 [16] είναι μία βάση δεδομένων που αποτελείται από εικόνες που έχουν ληφθεί από κάμερα smartphone. Περιλαμβάνει 2298 εικόνες οι οποίες διαφέρουν μεταξύ τους ως προς την ανάλυση, και τις ρυθμίσεις φωτισμού, καθώς συλλέχθηκαν από διαφορετικές συσκευές. Αυτή η ποικιλία μεταξύ των εικόνων είναι επιθυμητή καθώς αντιπροσωπεύει την ποικιλομορφία των εικόνων που θα κληθούν να ταξινομήσουν τα μοντέλα μας κατά τη χρήση τους.

Οι έξι διαφορετικές κατηγορίες προς ταξινόμηση που παρέχονται από το πεδίο diagnostic είναι οι εξής:

- Actinic Keratosis (Ακτινική Κεράτωση), συντομογραφία: ACK.
- Basal Cell Carcinoma (Βασικοκυτταρικό Καρκίνωμα), συντομογραφία: BCC.
- Melanoma (μελάνωμα), συντομογραφία: MEL.
- Melanocytic Nevi (Μελανοκυτταρικοί Σπίλοι), συντομογραφία: NV.
- Squamous Cell Carcinoma (Ακανθοκυτταρικό Καρκίνωμα), συντομογραφία: SCC.
- Seborrheic Keratosis (Σμηγματορροϊκή Κεράτωση), συντομογραφία: SEK

Σημειώνεται ότι, όπως και στη βάση ISIC, υπάρχουν εικόνες που αντιστοιχούν στην ίδια δερματική βλάβη.

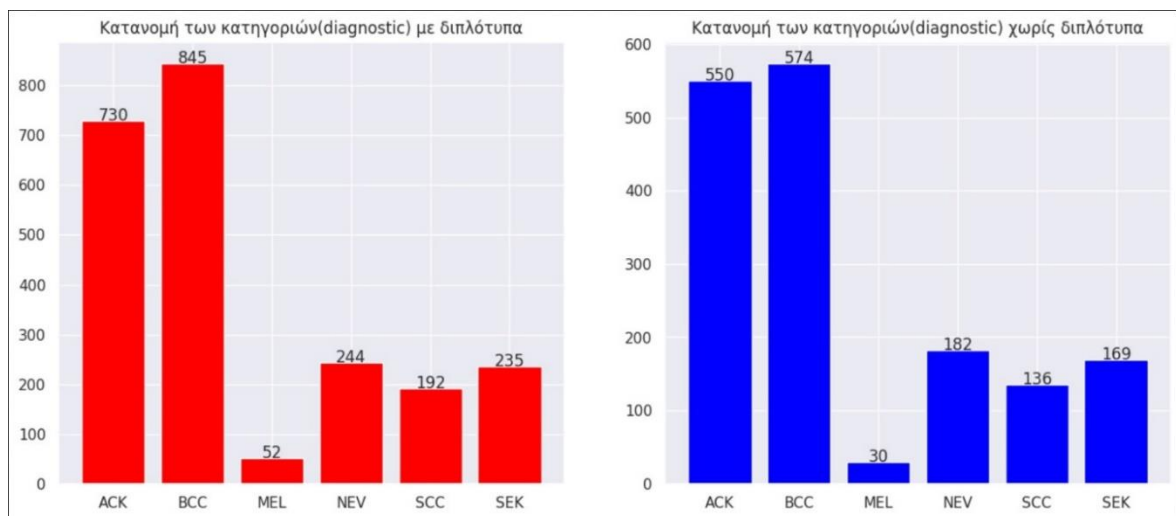
Η βάση δεδομένων PAD-UFES-20 περιέχει περισσότερες πληροφορίες (μεταδεδομένα) για κάθε εικόνα συγκριτικά με τη βάση ISIC. Συγκεκριμένα, εμπεριέχονται 21 πεδία για κάθε

εικόνα, τα οποία παρατίθενται παρακάτω αναλυτικά με το όνομα για κάθε χαρακτηριστικό όπως παρέχεται από τους δημιουργούς της βάσης:

- Age: Ένας integer που υποδεικνύει την ηλικία του ασθενούς.
- Background_father: Ένα string που υποδεικνύει τη χώρα καταγωγής του πατέρα του ασθενούς.
- Background_mother: Ένα string που υποδεικνύει τη χώρα καταγωγής της μητέρας του ασθενούς.
- Bleed: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη αιμορραγεί.
- Cancer_history: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής έχει ιστορικό καρκίνου.
- Changed: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη άλλαξε πρόσφατα.
- Diagnostic: Ένα string που υποδεικνύει τη διάγνωση της δερματικής βλάβης.
- Diameter_1: Ένας float που αναπαριστά το μήκος της οριζόντιας διαμέτρου της δερματικής βλάβης.
- Diameter_2: Ένας float που αναπαριστά το μήκος της κατακόρυφης διαμέτρου της δερματικής βλάβης.
- Drink: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής καταναλώνει αλκοόλ.
- Elevation: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη έχει διογκωθεί.
- Fitzpatrick: Ένας ακέραιος που υποδεικνύει το χρώμα δέρματος του ασθενούς.
- Gender: Ένα string που υποδεικνύει το φύλο του ασθενούς.
- Grew: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη μεγάλωσε.
- Has_piped_water: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής έχει πρόσβαση σε πόσιμο νερό.
- Has_sewage_system: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής έχει αποχετευτικό σύστημα στο σπίτι.
- Hurt: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη προκαλεί πόνο.
- Itch: Μια Boolean τιμή που υποδεικνύει αν η δερματική βλάβη προκαλεί κνησμό.
- Pesticide: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής εκτίθεται σε φυτοφάρμακα.

- Region: Ένα string που υποδεικνύει την περιοχή του σώματος που βρίσκεται η δερματική βλάβη.
- Skin_cancer_history: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής έχει ιστορικό καρκίνου του δέρματος.
- Smoke: Μια Boolean τιμή που υποδεικνύει αν ο ασθενής καπνίζει.

Επειδή τα πεδία σε αυτή τη βάση είναι πολλά, παρουσιάζεται μόνο η κατανομή των κατηγοριών (πεδίο: diagnostic) με και χωρίς διπλότυπα, στην Εικόνα 2.5.



Εικόνα 2.5 Κατανομή των τιμών του πεδίου diagnostic για τη βάση PAD-UFES-20

Παρατηρείται ότι και σε αυτή τη βάση τα δεδομένα είναι imbalanced και επομένως θα πρέπει να χρησιμοποιηθούν τεχνικές εξισορρόπησης.

Κεφάλαιο 3 Μέθοδοι αξιολόγησης και αρχιτεκτονικές νευρωνικών δικτύων

3.1 Μετρικές αξιολόγησης

Στους αλγορίθμους μηχανικής μάθησης είναι σημαντικό να μπορούμε να αξιολογούμε σωστά την ικανότητα του μοντέλου και να έχουμε μια καλή εικόνα της επίδοσης που θα έχει όταν τελικά τεθεί σε λειτουργία στον πραγματικό κόσμο. Επιπλέον, θα πρέπει να είμαστε σε θέση να προσδιορίσουμε ποιες από τις διαφορετικές αρχιτεκτονικές και υπερ-παραμέτρους που έχουμε δοκιμάσει αποδίδουν καλύτερα, ώστε να επιλέξουμε τον καλύτερο συνδυασμό για το τελικό μοντέλο που θα χρησιμοποιηθεί.

Γι' αυτούς τους λόγους, είναι σημαντικό να εφαρμόσουμε τα κατάλληλα metrics (μετρικές αξιολόγησης μοντέλου). Σε περίπτωση που χρησιμοποιηθούν ακατάλληλα metrics, θα καταλήξουμε σε παραπλανητικά αποτελέσματα όσον αφορά την αποτελεσματικότητα του μοντέλου και συνεπώς σε απρόσμενη λειτουργία της εφαρμογής.

Στην παρούσα εργασία αντιμετωπίζεται ένα πρόβλημα κατηγοριοποίησης εισόδου με πολλές κλάσεις (multiclass classification problem) πάνω σε μη ισορροπημένα δεδομένα (imbalanced datasets). Επομένως, αποφασίστηκε να χρησιμοποιηθούν τα εξής metrics: Sensitivity, Precision, AUC, F1-Score, Accuracy [40].

Πριν την ανάλυση καθεμιάς από τις μετρικές θα εισαχθούν κάποιες βασικές έννοιες που απαιτούνται για την περιγραφή τους. Συγκεκριμένα, εισάγονται τέσσερις βασικές ποσότητες που αφορούν κάθε κλάση ξεχωριστά:

1. Σωστά Θετικά (True Positives): Δεδομένα που ταξινομήθηκαν σωστά στην κλάση.
2. Λάθος Θετικά (False Positives): Δεδομένα που ταξινομήθηκαν στην κλάση, χωρίς ωστόσο να ανήκουν σε αυτή.
3. Σωστά Αρνητικά (True Negatives): Δεδομένα που ορθώς δεν ταξινομήθηκαν στη κλάση.
4. Λάθος Αρνητικά (False Negatives): Δεδομένα που ανήκουν στην κλάση, αλλά λανθασμένα δεν ταξινομήθηκαν σε αυτή.

Στη συνέχεια, γίνεται περιγραφή των μετρικών.

Precision:

Το precision για μία κλάση ορίζεται μαθηματικά ως ο λόγος του αριθμού των εισόδων που ταξινομήθηκαν σωστά στην κλάση προς τον συνολικό αριθμό των εισόδων που ταξινομήθηκαν στην κλάση:

$$\text{Precision κλάσης} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Διαισθητικά, το συγκεκριμένο metric καταδεικνύει την ικανότητα του μοντέλου να κάνει έγκυρες προβλέψεις.

Το precision σε πρόβλημα κατηγοριοποίησης πολλών κλάσεων μπορεί να υπολογιστεί ως ο μέσος όρος του precision κάθε κλάσης, δηλαδή:

$$\text{Average Precision} = \frac{\sum_{\text{για όλες τις κλάσεις}} \text{precision κλάσης}}{\text{πλήθος κλάσεων}}$$

Στη δική μας περίπτωση ωστόσο, επειδή τα δεδομένα είναι imbalanced, θα έχει ενδιαφέρον να υπολογίσουμε το precision λαμβάνοντας υπόψιν και την εκπροσώπηση κάθε κλάσης στη βάση. Συγκεκριμένα, θα υπολογίσουμε το precision δίνοντας έμφαση σε κάθε κλάση, ανάλογη με τον αριθμό των δεδομένων που περιέχει στη βάση. Το metric που προκύπτει από αυτή τη διαφοροποίηση ονομάζεται weighted average precision. Για τον υπολογισμό του, ανατίθεται αρχικά ένα βάρος σε κάθε κλάση, το οποίο υπολογίζεται ως εξής:

$$\text{Βάρος κλάσης} = \frac{\text{αριθμός δεδομένων που ανήκουν στην κλάση}}{\text{συνολικός αριθμός δεδομένων}}$$

και τελικά το weighted average precision υπολογίζεται ως:

$$\sum_{\text{για όλες τις κλάσεις}} \text{βάρος κλάσης} * \text{precision κλάσης}$$

Στα πειράματα θα γίνεται παρουσίαση των αποτελεσμάτων και των δύο μεθόδων.

Sensitivity:

Το sensitivity (ή recall) για μία κλάση ορίζεται μαθηματικά ως ο αριθμός των εισόδων που ταξινομήθηκαν σωστά στην κλάση προς τον συνολικό αριθμό εισόδων που ανήκουν στην κλάση:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Διαισθητικά, το συγκεκριμένο metric μας δείχνει την ικανότητα του μοντέλου να ταξινομεί τα δεδομένα σωστά στην κλάση που ανήκουν.

Συνολικά, υπολογίζεται το average sensitivity, αλλά και το weighted average sensitivity, ακολουθώντας την ίδια διαδικασία που περιεγράφηκε στον υπολογισμό του precision.

F1-Score

Το F1 score για κάθε κλάση υπολογίζεται ως η αρμονική μέση τιμή του sensitivity και precision, δηλαδή:

$$F1 = 2 * \frac{\text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

Το F1-score, συνδυάζει το precision με το sensitivity σε ένα metric, δίνοντας μία πιο γενική εικόνα της επίδοσης του μοντέλου. Ένα υψηλό F1-score δείχνει ότι το μοντέλο έχει πετύχει μια καλή ισορροπία ανάμεσα στα precision και sensitivity.

Στα πειράματα παρουσιάζονται τα αποτελέσματα για το average και weighted average F1-Score.

AUC (Area Under the Curve ROC) – Εμβαδόν κάτω από τη καμπύλη ROC

Σημαντικό στοιχείο για την μέτρηση του AUC είναι η έννοια του κατωφλιού πιθανότητας (probability threshold). Αναλυτικά, όταν ένα μοντέλο κατηγοριοποίησης δέχεται μία είσοδο, παράγει πιθανότητες για όλες τις κλάσεις, με τις πιθανότητες να αθροίζουν στο 1. Ωστόσο, για να επιλέξουμε την πρόβλεψη του μοντέλου πολλές φορές ορίζουμε ένα κατώφλι πιθανότητας, δηλαδή μια ελάχιστη τιμή που θα πρέπει να ξεπερνά η πιθανότητα

της κλάσης, ώστε να επιλεγθεί ως η πρόβλεψη του μοντέλου. Διαφορετικές τιμές του κατωφλίου μπορούν να επηρεάσουν σημαντικά τις τιμές του precision και sensitivity.

Υψηλό κατώφλι μπορεί να οδηγήσει σε λιγότερες σωστές ταξινομήσεις για μια κλάση, μειώνοντας το sensitivity, αλλά αυξάνοντας το precision. Ένα υψηλό κατώφλι είναι χρήσιμο όταν θα θέλαμε να έχουμε όσο το δυνατόν λιγότερες λανθασμένες ταξινομήσεις στην κλάση. Αντίθετα, ένα χαμηλό κατώφλι, θα αυξήσει τον αριθμό των εισόδων που ταξινομήθηκαν σωστά στην κλάση, ωστόσο, θα αυξήσει και τον αριθμό των εισόδων που ταξινομήθηκαν λανθασμένα στην ίδια κλάση. Συνεπώς, θα προκύψει μία αύξηση στο sensitivity και μείωση στο precision. Το χαμηλό κατώφλι είναι χρήσιμο όταν θα θέλαμε το μοντέλο μας να ταξινομεί οπωσδήποτε σωστά εισόδους που ανήκουν στην κλάση.

Το AUC είναι άλλο ένα δημοφιλές metric για την αξιολόγηση ενός μοντέλου σε προβλήματα κατηγοριοποίησης. Λειτουργεί αναλύοντας την ικανότητα του μοντέλου να διαχωρίζει τα δεδομένα που ανήκουν σε μία κλάση με αυτά που δεν ανήκουν. Συγκεκριμένα, για μία κλάση υπολογίζονται δύο λόγοι:

1. Ο ρυθμός σωστών ταξινομήσεων (True Positive Rate) στην κλάση, που ορίζεται και υπολογίζεται ακριβώς όπως το sensitivity και
2. ο ρυθμός λανθασμένων ταξινομήσεων στην κλάση (False Positive Rate), που ορίζεται ως ο παρακάτω λόγος:

$$\frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

Στη συνέχεια, δημιουργείται η καμπύλη ROC για την κλάση, σχεδιάζοντας τους δύο παραπάνω λόγους για διαφορετικά κατώφλια πιθανότητας. Τέλος, η τιμή του AUC υπολογίζεται ως το εμβαδόν κάτω από την καμπύλη ROC και κυμαίνεται από 0 έως 1.

Διαισθητικά, το AUC καταδεικνύει πόσο ικανό είναι το μοντέλο, δεδομένου ενός τυχαίου δείγματος, να αναθέσει υψηλότερη πιθανότητα στη θετική κλάση από τι στην αρνητική.

Στα πειράματα θα παρουσιάζεται μόνο το average AUC, που είναι ο μέσος όρος του AUC όλων των κλάσεων.

Accuracy

Το accuracy είναι ένα πολύ δημοφιλές metric για την αξιολόγηση σε προβλήματα ταξινόμησης. Ορίζεται ως ο λόγος των δεδομένων που ταξινομήθηκαν σωστά προς τον συνολικό αριθμό των δεδομένων. Το συγκεκριμένο metric δεν είναι κατάλληλο για να αξιολογούμε την επίδοση σε imbalanced δεδομένα, ωστόσο, παρατίθεται ως μια ένδειξη για την συνολική επίδοση του εκάστοτε μοντέλου.

3.2 Νευρωνικά δίκτυα

Σε αυτή την ενότητα θα παρουσιαστούν οι αρχιτεκτονικές των νευρωνικών δικτύων που επιλέχθηκαν για να εκπαιδευτούν πάνω στις βάσεις.

3.2.1 Resnet34

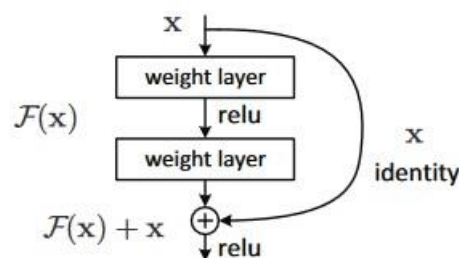
Τα ResNet (Residual Neural Networks) [41] είναι μοντέλα νευρωνικών δικτύων που έχουν παρουσιάσει εξαιρετικές επιδόσεις σε μία σειρά από προβλήματα πάνω στην αναγνώριση εικόνων, αλλά και σε δημοφιλείς βάσεις που χρησιμοποιούνται για την αξιολόγηση τέτοιων μοντέλων (Imagenet). Το κλειδί για την επίδοσή τους βασίζεται στον τρόπο που διαχειρίζονται το πρόβλημα του vanishing gradient [42].

Το πρόβλημα του vanishing gradient εμφανίζεται κατά την διάρκεια της εκπαίδευσης ενός νευρωνικού δικτύου, όταν αυτό είναι πολύ βαθύ (δηλαδή αποτελείται από πάρα πολλά διαδοχικά επίπεδα), έχοντας σαν αποτέλεσμα, οι μερικές παράγωγοι ως προς το σφάλμα, που υπολογίζονται στο τελευταίο επίπεδο, να τείνουν στο μηδέν όσο αυτές διαδίδονται στα πρώτα επίπεδα του δικτύου. Συνέπεια αυτού, είναι τα βάρη των πρώτων επιπέδων να μην εκπαιδεύονται καθόλου.

Η μέθοδος που χρησιμοποιούν τα ResNet για να λύσουν το πρόβλημα του vanishing gradient αλλά και για τη γενική βελτίωση της επίδοσης, είναι η εφαρμογή των παραλειπόμενων συνδέσεων στο δίκτυο (skip connections). Οι παραλειπόμενες συνδέσεις επιτρέπουν στην είσοδο ενός επιπέδου να συνυπολογίζεται αναλλοίωτη στον τελικό υπολογισμό της εξόδου, παραλείποντας ενδιάμεσα επίπεδα του δικτύου. Αυτό επιτρέπει στο νευρωνικό δίκτυο να αυξήσει κατά πολύ τον αριθμό των παραμέτρων του, διατηρώντας όμως τη δυνατότητα να εκπαιδεύει τις παραμέτρους που βρίσκονται στα

πρώτα επίπεδα του, καθώς αυτές είναι άμεσα συνδεδεμένες με την έξοδο μέσω των παραλειπόμενων συνδέσεων.

Αναλυτικά, τα Resnet αποτελούνται από συνεχόμενα υπολειμματικά μπλοκ (residual blocks), τα οποία περιέχουν τις παραλειπόμενες συνδέσεις. Η δομή ενός υπολειμματικού μπλοκ παρουσιάζεται στο σχήμα 3.1. Η είσοδος διέρχεται μέσα από δύο συνελκτικά επίπεδα με batch normalization και στη συνέχεια το αποτέλεσμα προστίθεται στην αρχική είσοδο.



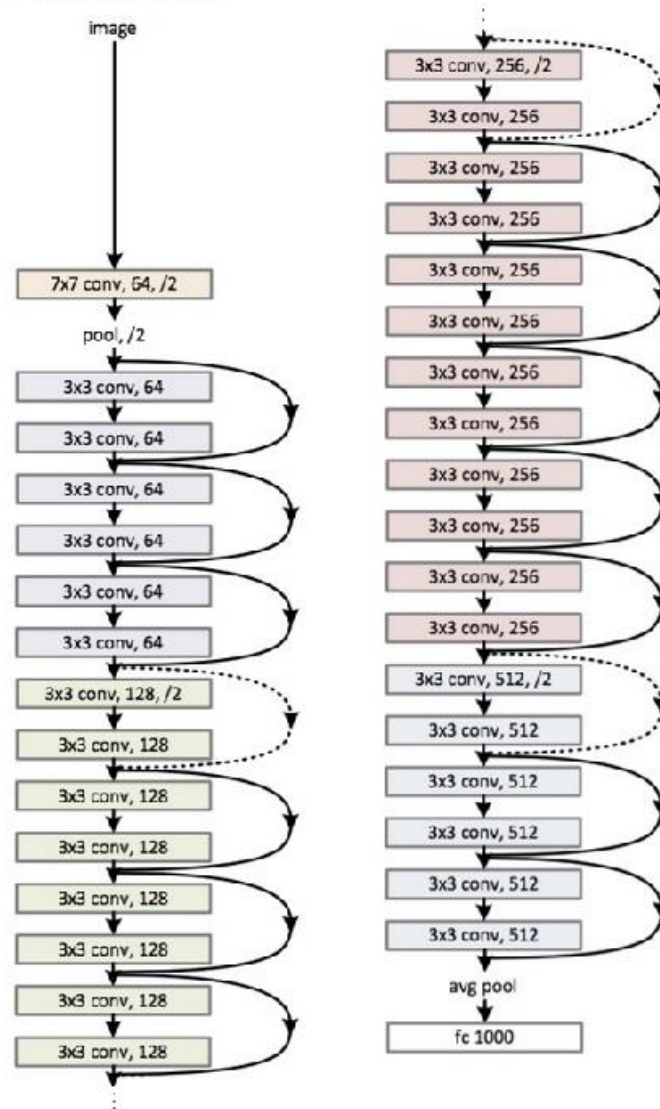
Σχήμα 3.1 Η δομή ενός υπολειμματικού μπλοκ. Σχήμα από [41]

Διαισθητικά, μπορούμε επιπλέον να δούμε αυτή τη λειτουργία ως μία συγχώνευση διαφορετικών νευρωνικών δικτύων όπου το καθένα μοντελοποιεί διαφορετικά χαρακτηριστικά και τελικά προσφέρει την πληροφορία του στην αρχική είσοδο.

Υπάρχουν πολλές διαφορετικές αρχιτεκτονικές Resnet, που διαφέρουν μεταξύ τους κυρίως στο βάθος του δικτύου. Σε αυτή την εργασία θα χρησιμοποιηθεί ένα Resnet34, καθώς παρέχει αρκετά καλή ισορροπία μεταξύ επίδοσης και πολυπλοκότητας. Μοντέλα Resnet περισσότερων παραμέτρων (όπως το Resnet50) δεν δοκιμάζονται, λόγω περιορισμών στους διαθέσιμους πόρους.

Το Resnet-34, όπως υποδηλώνεται και στο όνομά του, περιέχει 34 επίπεδα. Μια αναλυτική εικόνα της δομής και των παραμέτρων του Resnet-34 παρουσιάζεται στο σχήμα 3.2.

34-layer residual



Σχήμα 3.2 Η αρχιτεκτονική του Resnet34. Σχήμα από [41]

Συγκεκριμένα, περιέχει ένα αρχικό συνελκτικό επίπεδο που δέχεται ως είσοδο την εικόνα και παράγει μια έξοδο με 64 κανάλια. Έπειτα, ακολουθούν τα υπολειμματικά μπλοκ.

Τα υπολειμματικά μπλοκ στο Resnet-34 χωρίζονται σε 4 ομάδες:

- Η πρώτη ομάδα περιέχει τρία διαδοχικά υπολειμματικά μπλοκ όπου τα συνελκτικά επίπεδα δέχονται και παράγουν 64 κανάλια.
- Η δεύτερη ομάδα περιέχει τέσσερα διαδοχικά υπολειμματικά μπλοκ, όπου τα συνελκτικά επίπεδα δέχονται και παράγουν 128 κανάλια.
- Η τρίτη ομάδα περιέχει έξι διαδοχικά υπολειμματικά μπλοκ, με τα συνελκτικά επίπεδα να έχουν 256 κανάλια.

- Η τέταρτη ομάδα περιέχει άλλα τρία υπολειμματικά μπλοκ, με τα συνελικτικά επίπεδα να έχουν 512 κανάλια.

Τέλος, υπάρχει ένα γραμμικό επίπεδο, το οποίο παράγει τις τελικές προβλέψεις για κάθε κλάση.

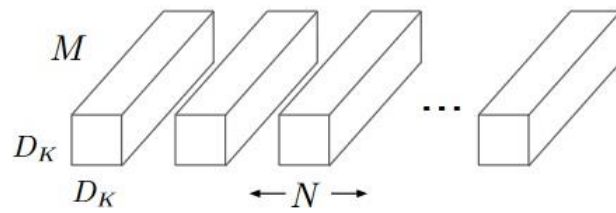
3.2.2 MobileNet

Τα MobileNet [43] είναι μοντέλα νευρωνικών δικτύων που έχουν πετύχει υψηλές επιδόσεις σε μία σειρά από προβλήματα στην όραση υπολογιστών, διατηρώντας πολύ χαμηλό υπολογιστικό κόστος. Τα συγκεκριμένα νευρωνικά δίκτυα περιλαμβάνουν σχετικά μικρό αριθμό παραμέτρων και είναι σχεδιασμένα ώστε να εκτελούν όσο το δυνατόν λιγότερες αριθμητικές πράξεις διατηρώντας ωστόσο, ικανοποιητική απόδοση σε θέματα που αφορούν την αναγνώριση εικόνας. Συνεπώς, είναι μοντέλα που χρησιμοποιούνται κυρίως σε συσκευές smartphone, σε ενσωματωμένα συστήματα και σε άλλες υπολογιστικές μονάδες με μειωμένη υπολογιστική ισχύ. Κύριο χαρακτηριστικό στην αρχιτεκτονική τους είναι η χρήση των depthwise separable συνελίξεων.

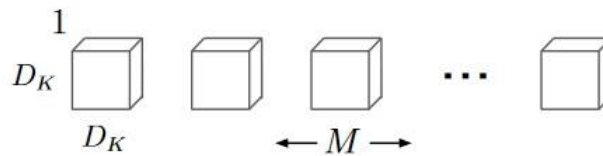
Οι depthwise separable συνελίξεις χωρίζονται σε δύο ξεχωριστές διαδοχικές συνελίξεις: depthwise συνέλιξη και pointwise συνέλιξη. Η depthwise συνέλιξη χρησιμοποιεί ένα φίλτρο για κάθε κανάλι της εισόδου ξεχωριστά, σε αντίθεση με την παραδοσιακή συνέλιξη όπου το φίλτρο εφαρμόζεται σε όλα τα κανάλια της εισόδου. Αυτή η λειτουργία μοντελοποιεί τις χωρικές εξαρτήσεις των pixel για κάθε κανάλι χωριστά και παράγει μια έξοδο με τον ίδιο αριθμό καναλιών της εισόδου.

Στη συνέχεια, στην έξοδο της depthwise εφαρμόζεται η pointwise συνέλιξη. Η συγκεκριμένη συνέλιξη στοχεύει να συνδυάσει τη πληροφορία μεταξύ των καναλιών της εισόδου. Για το σκοπό αυτό, χρησιμοποιούνται φίλτρα $1 \times 1 (\# \text{καναλιών})$ εκτελώντας κανονική συνέλιξη, μοντελοποιώντας έτσι την αλληλεπίδραση των διαφορετικών χαρακτηριστικών ανάμεσα στα κανάλια.

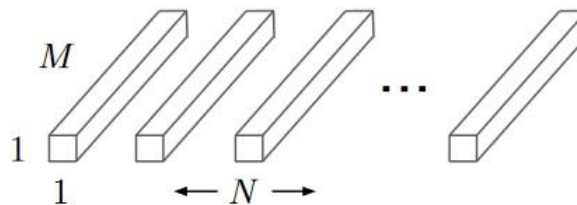
Αυτή η δομή της συνέλιξης έχει ως αποτέλεσμα την δραστική μείωση του αριθμού των υπολογισμών που απαιτούνται σε σχέση με την παραδοσιακή συνέλιξη. Η αρχιτεκτονική μιας depthwise separable συνέλιξης φαίνεται στο σχήμα 3.3. Στο 3.3(a) απεικονίζονται φίλτρα παραδοσιακής συνέλιξης τα οποία αντικαθίστανται από μία depthwise συνέλιξη 3.3(b) και μία pointwise συνέλιξη 3.3(c).



(a) Standard Convolution Filters



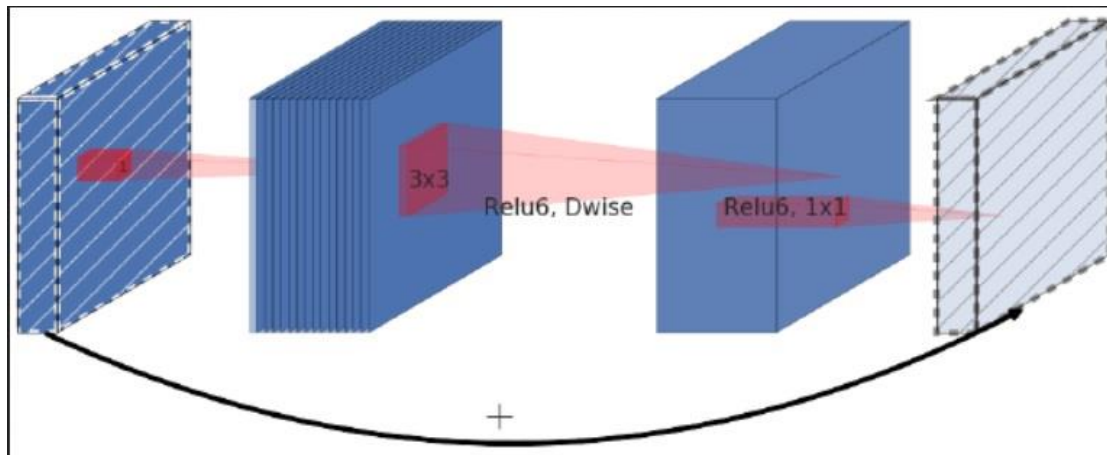
(b) Depthwise Convolutional Filters



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Σχήμα 3.3 Η αρχιτεκτονική μιας depthwise separable συνέλιξης. Σχήμα από [43]

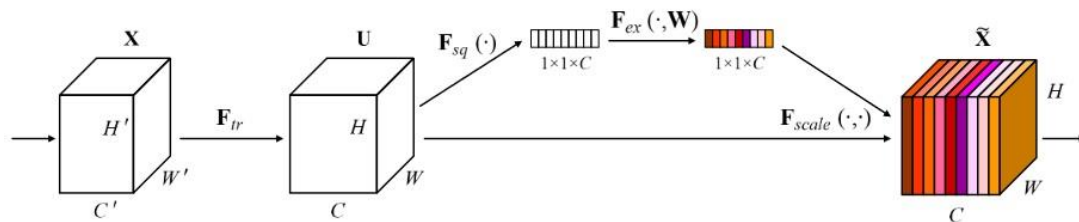
Έπειτα από την αρχική επιτυχία του MobileNet, έχουν προστεθεί περισσότερες προχωρημένες τεχνικές στην αρχιτεκτονική του αρχικού μοντέλου, όπως η χρήση skip connections και επαύξηση καναλιών, καταλήγοντας σε μπλοκ νευρωνικών επιπέδων που ονομάζονται inverted residuals. Αυτά τα μπλοκ είναι το δομικό στοιχείο στα μοντέλα MobileNet v2 [44] και MobileNet v3 [45].



Σχήμα 3.4 Η δομή του inverted residual block. Σχήμα από [44]

Ένα inverted residual block (σχήμα 3.4) αποτελείται από τρεις διαδοχικές συνελίξεις. Η πρώτη συνέλιξη είναι pointwise (διαστάσεις φίλτρου: $1 \times 1 \times \text{\#καναλιών}$) και έχει σκοπό την επαύξηση του αριθμού των καναλιών της εισόδου με στόχο την αναπαράσταση των χαρακτηριστικών σε υψηλότερες διαστάσεις. Στη συνέχεια, εφαρμόζεται με τη σειρά μια depthwise και μία pointwise συνέλιξη με μείωση του αριθμού των καναλιών. Οι τελευταίες δύο συνελίξεις ουσιαστικά συντελούν μία depthwise separable συνέλιξη. Επιπλέον, στο αποτέλεσμα των τριών συνελίξεων, προστίθεται η αρχική είσοδος που δέχτηκε το inverted residual block, εφαρμόζοντας μια παραλειπόμενη σύνδεση (skip connection). Έμφαση δίνεται επίσης στις συναρτήσεις ενεργοποίησης που χρησιμοποιούνται σε κάθε επίπεδο. Συγκεκριμένα, στα δύο πρώτα επίπεδα του μπλοκ (pointwise για επαύξηση των καναλιών και depthwise συνέλιξη) εφαρμόζεται είτε ReLU είτε Hard swish, ενώ στην τελευταία συνέλιξη (pointwise με μείωση του αριθμού των καναλιών) δεν χρησιμοποιείται καμία συνάρτηση ενεργοποίησης.

Επιπρόσθετα, τα inverted residual block του μοντέλου MobileNet v3 περιλαμβάνουν ένα squeeze-excitation [46] νευρωνικό επίπεδο. Τα επίπεδα αυτά προστίθενται στο block, ακριβώς πριν από την τελευταία pointwise συνέλιξη. Η αρχιτεκτονική του φαίνεται στο σχήμα 3.5.



Σχήμα 3.5 Η δομή του squeeze-excitation μπλοκ. Σχήμα από [46]

Το επίπεδο αυτό χρησιμοποιείται με σκοπό να μοντελοποιήσει τις αλληλεξαρτήσεις των καναλιών της εισόδου, ώστε να παραχθούν σημαντικές πληροφορίες που αφορούν την συσχέτισή τους, με όσο το δυνατόν μικρότερο υπολογιστικό κόστος. Αρχικά, χρησιμοποιείται global average pooling σε κάθε κανάλι ξεχωριστά, το οποίο παράγει για έξοδο, μία τιμή για το κάθενα (φάση squeeze). Η τιμή αυτή που ανατίθεται σε κάθε κανάλι καταγράφει μία αναπαράσταση για αυτό. Στη συνέχεια, εφαρμόζονται δύο συνεχόμενα γραμμικά επίπεδα, το πρώτο με συνάρτηση ενεργοποίησης ReLU, ενώ το δεύτερο με sigmoid. Τα γραμμικά επίπεδα έχουν σκοπό να επικοινωνήσουν την πληροφορία κάθε καναλιού με τα υπόλοιπα, ώστε να καταγραφούν αλληλεξαρτήσεις ανάμεσά τους (φάση excitation). Έτσι, το διάνυσμα μεγέθους $C \times 1 \times 1$ που προέκυψε, περιέχει σε κάθε του θέση, μία τιμή που καταδεικνύει την σημαντικότητα του συγκεκριμένου καναλιού. Στο τελικό βήμα, κάθε στοιχείο αυτού του διανύσματος πολλαπλασιάζει τις τιμές του αντίστοιχου καναλιού της αρχικής εισόδου. Αυτό έχει ως αποτέλεσμα, οι τιμές κάθε καναλιού να ρυθμίζονται ανάλογα με τη σημαντικότητά τους, αναδεικνύοντας με αυτό τον τρόπο τη σημαντική πληροφορία της εισόδου.

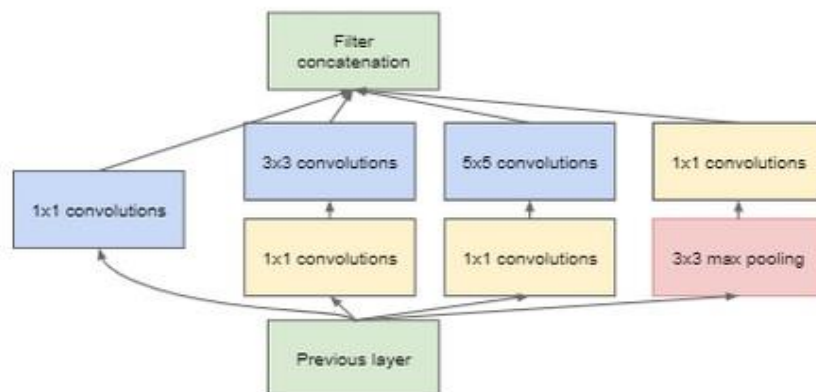
Σε αυτή την εργασία, θα γίνει χρήση του MobileNet v3 Large. Η αναλυτική αρχιτεκτονική του φαίνεται στον Πίνακα 3.1.

Πίνακας 3.1 Η αρχιτεκτονική του MobileNet v3 Large. Πίνακας από [45]

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3x3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3x3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5x5	72	40	✓	RE	2
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 3x3	240	80	-	HS	2
$14^2 \times 80$	bneck, 3x3	200	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	480	112	✓	HS	1
$14^2 \times 112$	bneck, 3x3	672	112	✓	HS	1
$14^2 \times 112$	bneck, 5x5	672	160	✓	HS	2
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	conv2d, 1x1	-	960	-	HS	1
$7^2 \times 960$	pool, 7x7	-	-	-	-	1
$1^2 \times 960$	conv2d 1x1, NBN	-	1280	-	HS	1
$1^2 \times 1280$	conv2d 1x1, NBN	-	k	-	-	1

3.2.3 GoogleNet

Το Google Net [47] είναι μια αρχιτεκτονική συνελκτικού νευρωνικού δικτύου που παρουσιάστηκε το 2014 από ερευνητές της Google. Κύριο προτέρημα της αρχιτεκτονικής είναι η χρήση του inception μπλοκ που παρουσιάζεται στο σχήμα 3.6.



Σχήμα 3.6 Το inception block. Εικόνα από [47]

Ένα inception μπλοκ περιέχει πολλαπλά συνελικτικά μπλοκ που χρησιμοποιούν διαφορετικά μεγέθη φίλτρων και επιπέδων συγκέντρωσης (pooling). Τα συνελικτικά αυτά μπλοκ εφαρμόζονται πάνω στην ίδια είσοδο ξεχωριστά. Στη συνέχεια, οι έξοδοι από κάθε μπλοκ συνενώνονται (concatenate) δημιουργώντας μία μοναδική έξοδο. Το βασικό πλεονέκτημα αυτής της δομής είναι ότι επιτρέπει σε ένα inception μπλοκ να συνδυάζει τα οφέλη διαφορετικών μεγεθών συνελικτικών φίλτρων.

Επιπρόσθετα, ένα ακόμη σημαντικό στοιχείο του Google Net είναι η χρήση των βοηθητικών ταξινομητών (auxiliary classifiers). Οι βοηθητικοί ταξινομητές είναι μπλοκ νευρωνικών δικτύων που περιέχουν συνελικτικά και γραμμικά επίπεδα. Τοποθετούνται σε ενδιάμεσα επίπεδα του Google Net για να κάνουν προβλέψεις. Το σφάλμα αυτών των ταξινομητών, συνυπολογίζεται στο σφάλμα των εξόδων που παράγει το Google Net στο τελευταίο του επίπεδο. Στόχος των βοηθητικών ταξινομητών είναι κυρίως να αντιμετωπίσουν το vanishing gradient πρόβλημα, αλλά και να αποτρέπουν το συνολικό νευρωνικό δίκτυο από να κάνει overfit.

Η συνολική δομή του GoogleNet παρουσιάζεται στον πίνακα 3.2.

Πίνακας 3.2 Η αρχιτεκτονική του GoogleNet. Πίνακας από [47]

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

3.2.4 EfficientNet

Τα EfficientNet [48] είναι μοντέλα νευρωνικών δικτύων που παρουσιάζουν εξαιρετικές επιδόσεις σε προβλήματα κατηγοριοποίησης εικόνας. Σε αυτό τον τύπο μοντέλου, οι ερευνητές επικεντρώνονται στην εύρεση ενός μεθοδικού τρόπου για να καθορίσουν τη σχέση ανάμεσα στο βάθος (αριθμός επιπέδων δικτύου), το πλάτος (αριθμός καναλιών) και την ανάλυση(διαστάσεις εικόνας εισόδου) του νευρωνικού δικτύου. Συγκεκριμένα, οι ερευνητές παρέχουν μαθηματικά εργαλεία για τον προσδιορισμό του τρόπου κλιμάκωσης (scaling) ενός συνελκτικού νευρωνικού δικτύου κάτω από συγκεκριμένους περιορισμούς υπολογιστικών πόρων, επιφέροντας πολύ σημαντικά αποτελέσματα ως προς την εύρεση μιας ισορροπίας ανάμεσα στην ταχύτητα εκπαίδευσης του μοντέλου, τον αριθμό των παραμέτρων και την επίδοσή του.

Το αρχικό μοντέλο που χρησιμοποιείται για να γίνουν τα πειράματα είναι το EfficientNet B0 και η αρχιτεκτονική του φαίνεται στον πίνακα 3.3.

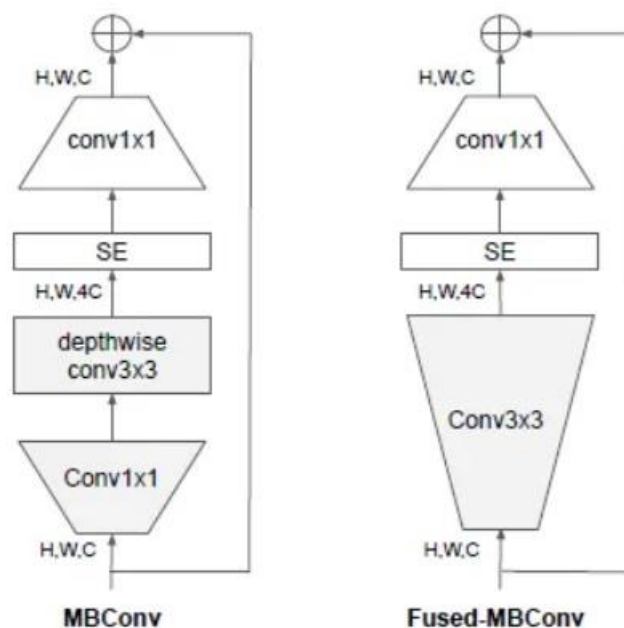
Πίνακας 3.3 Η αρχιτεκτονική του EfficientNet B0. Πίνακας από [48]

Stage i	Operator $\mathcal{F}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Όπως παρατηρούμε, κύριο χαρακτηριστικό της δομής είναι το MBConv, το οποίο είναι βασισμένο στο inverted residual block με squeeze-excitation που περιεγράφηκε στην υποενότητα 3.2.2.

Στη συνέχεια, κλιμακώνοντας αυτό το βασικό μοντέλο κάτω από συγκεκριμένους μαθηματικούς περιορισμούς που θα πρέπει να τηρούνται όσον αφορά τις σχέσεις μεταξύ βάρους, πλάτους, ανάλυσης, προκύπτουν μεγαλύτερα μοντέλα, από το B1 μέχρι και το B7.

Έπειτα από την αρχική επιτυχία των EfficientNet, οι ερευνητές προχώρησαν στην υλοποίηση των EfficientNet v2 [49]. Μεταξύ των αλλαγών που πραγματοποιήθηκαν στην αρχιτεκτονική των αρχικών μοντέλων, είναι η αντικατάσταση στα πρώτα επίπεδα των MBConv με τα Fused-MBConv. Τα Fused-MBConv είναι μια παραλλαγή των MBConv η οποία αντικαθιστά τις πρώτες δύο συνελίξεις (pointwise και depthwise) με μία κανονική συνέλιξη, όπως φαίνεται και στο σχήμα 3.7.



Σχήμα 3.7 Διαφορά ανάμεσα σε MBConv και Fused-MBConv. Σχήμα από [49]

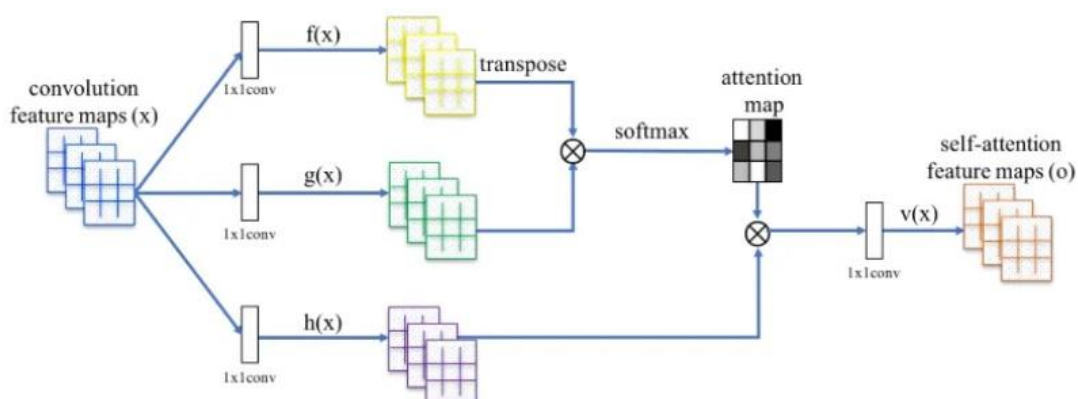
Σε αυτή τη διπλωματική, επιλέχθηκε να χρησιμοποιηθεί το EfficientNet v2 small, καθώς έχει παρουσιάσει εξαιρετικές επιδόσεις με σχετικά μικρό αριθμό παραμέτρων και χρόνο εκπαίδευσης. Η αρχιτεκτονική του φαίνεται στον πίνακα 3.4.

Πίνακας 3.4 Η αρχιτεκτονική του EfficientNet v2 Small. Πίνακας από [49]

Stage	Operator	Stride	#Channels	#Layers
0	Conv3x3	2	24	1
1	Fused-MBConv1, k3x3	1	24	2
2	Fused-MBConv4, k3x3	2	48	4
3	Fused-MBConv4, k3x3	2	64	4
4	MBConv4, k3x3, SE0.25	2	128	6
5	MBConv6, k3x3, SE0.25	1	160	9
6	MBConv6, k3x3, SE0.25	2	256	15
7	Conv 1x1 & Pooling & FC	-	1280	1

3.2.5 MobileNet + Self attention layers

Η χρήση τεχνικών που εφαρμόζουν self attention (αυτό-συσχέτιση) έχει γίνει ιδιαίτερα δημοφιλής στο χώρο της βαθιάς μάθησης λόγω των εξαιρετικών αποτελεσμάτων που παρουσιάζουν σε προβλήματα που αφορούν natural language processing [50], [51]. Επομένως, είναι λογικό, η χρήση τους να επεκταθεί και σε προβλήματα της όρασης υπολογιστών. Στην εργασία Self-Attention Generative Adversarial Networks [52] οι ερευνητές χρησιμοποίησαν επίπεδα self-attention για την αύξηση της απόδοσης των GAN μοντέλων τους, με μεγάλη επιτυχία. Μια αναλυτική παρουσίαση του συγκεκριμένου νευρωνικού επιπέδου παρουσιάζεται στο σχήμα 3.8.



Σχήμα 3.8 To self attention block. Σχήμα από [52]

Αναλυτικά, το self attention επίπεδο δουλεύει ως εξής:

Έστω $C \times W \times H$, οι διαστάσεις της εισόδου, με C να αναπαριστά τον αριθμό των καναλιών της εισόδου. Αρχικά, στην είσοδο του επιπέδου εφαρμόζονται τρεις ξεχωριστές pointwise συνελίξεις. Ο αριθμός των καναλιών που παράγονται στην έξοδο των δύο συνελίξεων (έστω f και g) ισούται με $C^* = C/8$, ενώ για τη τρίτη συνέλιξη (έστω h) ο αριθμός καναλιών παραμένει ίδιος. Στη συνέχεια, τα f και g μετασχηματίζονται στη μορφή $C^* \times N$, όπου $N = W \times H$ και εκτελείται ο ακόλουθος πολλαπλασιασμός πινάκων :

$$\text{attention}_{j,i} = f(x_i)^T g(x_j)$$

Ο πίνακας που παράγεται έχει διαστάσεις $N \times N$ και ονομάζεται attention map (πίνακας αυτό-προσοχής), και διαισθητικά αυτό που μας δείχνει είναι η σημαντικότητα του pixel j σε σχέση με το pixel i της αρχικής εισόδου. Κάθε στήλη του πίνακα κανονικοποιείται, έτσι ώστε το άθροισμα των στοιχείων της να είναι 1 (softmax). Τέλος, πολλαπλασιάζουμε το h με το attention map που προέκυψε και προσθέτουμε το αποτέλεσμα αυτής της πράξης στην αρχική είσοδο που δόθηκε στο επίπεδο.

Κάθε pixel της εξόδου που λαμβάνουμε, περικλείει πληροφορία για τη συσχέτιση και αλληλεπίδραση του με τα υπόλοιπα pixel της εικόνας. Αυτή η διαδικασία επιτρέπει σε pixel της εικόνας που βρίσκονται σε διαφορετικά σημεία της, να επικοινωνήσουν και να ανταλλάξουν πληροφορία, σε αντίθεση με τη διαδικασία της συνέλιξης, όπου το φίλτρο εφαρμόζεται τοπικά πάντα.

Αντλώντας έμπνευση από αυτή την εργασία, θα χρησιμοποιηθεί στην παρούσα διπλωματική εργασία αυτό το νευρωνικό επίπεδο για πειράματα πάνω σε ένα προ-εκπαιδευμένο MobileNet v3 Large. Είναι σημαντικό να σημειωθεί ότι τα συγκεκριμένα επίπεδα μπορούν να αυξήσουν κατά πολύ το κόστος σε υπολογισμούς και μνήμη όταν εφαρμόζονται σε εικόνες με μεγάλες διαστάσεις. Για να αποφευχθεί αυτό το πρόβλημα, τα επίπεδα αυτά έχουν προστεθεί στο τέλος του $4^{ου}, 7^{ου}, 10^{ου}, 13^{ου}, 15^{ου}$ inverted residual block του μοντέλου, όπου οι διαστάσεις της εικόνας έχουν μειωθεί από τις συνελίξεις των προηγούμενων επιπέδων. Το νευρωνικό δίκτυο που προκύπτει θα αναφέρεται ως MobileNet + SA.

3.2.6 Συγχώνευση μοντέλων

Η συγχώνευση μοντέλων (ensemble learning) [72] είναι μια τεχνική μηχανικής μάθησης στην οποία, πολλά μοντέλα συνδυάζονται είτε κατά τη διάρκεια της εκπαίδευσής τους, είτε έπειτα, με σκοπό την αύξηση της επίδοσης του συνολικού μοντέλου, καθώς επίσης και την αποφυγή της υπερ-μοντελοποίησης. Η τεχνική αυτή βασίζεται στην παρατήρηση ότι, διαφορετικά μοντέλα και αρχιτεκτονικές παρουσιάζουν διαφορετικά πλεονεκτήματα και επιδόσεις πάνω στις μετρικές αξιολόγησης. Επομένως, όταν συνδυαστούν τα επιμέρους μοντέλα, θα αξιοποιηθούν καλύτερα τα πλεονεκτήματα του καθενός με αποτέλεσμα το συνολικό μοντέλο να γίνει πιο αξιόπιστο στις προβλέψεις του.

Στη συγκεκριμένη εργασία η τεχνική της συγχώνευσης μοντέλων εφαρμόζεται μετά την εκπαίδευση των επιμέρους μοντέλων, κάνοντας χρήση της τεχνικής Μέσου Όρου. Συγκεκριμένα, σε αυτή τη μέθοδο, οι τελικές προβλέψεις για κάθε κλάση του συγχωνευμένου μοντέλου προκύπτουν λαμβάνοντας τον μέσο όρο των προβλέψεων των επιμέρους μοντέλων για κάθε κλάση.

Τα καλύτερα αποτελέσματα σε κάθε περίπτωση λαμβάνονται όταν αξιοποιούνται όλα τα μοντέλα. Το συγχωνευμένο μοντέλο θα αναφέρεται στα αποτελέσματα των πειραμάτων ως Ensemble.

3.3 Μεταφορά μάθησης (Transfer learning)

Για την επίτευξη της μέγιστης δυνατής επίδοσης, εφαρμόζεται η μέθοδος της μεταφοράς μάθησης σε όλα τα μοντέλα που εξετάζονται.

Η μεταφορά μάθησης είναι μια τεχνική που χρησιμοποιείται για την αύξηση της αποδοτικότητας ενός μοντέλου [53]. Η μέθοδος αυτή χρησιμοποιεί ήδη υπάρχοντα μοντέλα που έχουν εκπαιδευτεί σε συναφή προβλήματα και τα ξανά-εκπαιδεύει στο εκάστοτε πρόβλημα. Είναι ιδιαίτερα χρήσιμη όταν τα δεδομένα που έχουμε να χρησιμοποιήσουμε είναι λίγα. Η ιδέα αυτή, μπορεί και αποδίδει πολύ ικανοποιητικά, καθώς συχνά υπάρχουν κοινά πρότυπα και χαρακτηριστικά μεταξύ βάσεων δεδομένων και προβλημάτων. Έτσι, όταν ένα μοντέλο έχει εκπαιδευτεί σε μια προϋπάρχουσα βάση δεδομένων και έχει μοντελοποιήσει τα χαρακτηριστικά της, είναι πολύ εύκολο να εφαρμοστεί και σε ένα συναφές πρόβλημα με ικανοποιητική ακρίβεια.

Συγκεκριμένα, στην παρούσα εργασία όπου αντιμετωπίζεται ένα πρόβλημα ταξινόμησης εικόνας, χρησιμοποιούνται μοντέλα που έχουν εκπαιδευτεί στη βάση δεδομένων IMAGENET [54]. Η βάση δεδομένων IMAGENET περιέχει περισσότερες από 14 εκατομμύρια εικόνες που ανήκουν σε 1000 κατηγορίες. Το τεράστιο μέγεθος αυτής της βάσης επιτρέπει στα μοντέλα που έχουν εκπαιδευτεί σε αυτή να μοντελοποιούν χαρακτηριστικά και πρότυπα που συναντώνται σε εικόνες με εξαιρετική ακρίβεια.

Είναι σημαντικό να σημειωθεί πως όταν χρησιμοποιείται ένα προ-εκπαιδευμένο μοντέλο σε μία διαφορετική βάση δεδομένων, απαιτείται να αλλάξει η αρχιτεκτονική του δικτύου στα τελευταία επίπεδα που αφορούν την ταξινόμηση, ώστε ο αριθμός των κλάσεων στην έξοδο να συμπίπτει με τον αριθμό των κλάσεων που έχουμε να ταξινομήσουμε. Συγκεκριμένα, για τις δύο βάσεις που χρησιμοποιούμε, αλλάζουμε το τελευταίο γραμμικό επίπεδο όλων των μοντέλων, ώστε να συμπίπτει με τον αριθμό των κλάσεων για κάθε βάση (οκτώ για την ISIC και έξι για την PAD).

Όσον αφορά τη διαδικασία εκπαίδευσης των μοντέλων, στην μεταφορά μάθησης υπάρχουν δύο διαφορετικές προσεγγίσεις. Η πρώτη είναι να εκπαιδευτούν μόνο τα γραμμικά επίπεδα που προστέθηκαν στο τέλος του προ-εκπαιδευμένου μοντέλου, ενώ τα υπόλοιπα επίπεδα να είναι «παγωμένα»(frozen) κατά τη διάρκεια της εκπαίδευσης. Αυτό σημαίνει, πως δεν θα μεταβάλλονται οι τιμές τους, αλλά θα συμμετέχουν στην εκπαίδευση μόνο για να εξάγουν χαρακτηριστικά που θα βοηθήσουν τα γραμμικά επίπεδα να ταξινομήσουν σωστά τις εισόδους. Αυτή η προσέγγιση πλεονεκτεί στη ταχύτητα της εκπαίδευσης, καθώς μόνο λίγες παράμετροι πραγματικά ανανεώνονται, ωστόσο είναι πιθανό να οδηγήσει σε μείωση της επίδοσης του μοντέλου. Η δεύτερη προσέγγιση είναι να ξανά-εκπαιδευτεί ολόκληρο το μοντέλο, αφού προστεθούν τα γραμμικά νευρωνικά επίπεδα. Σε αυτή τη προσέγγιση, ο χρόνος εκπαίδευσης είναι σαφώς μεγαλύτερος, ωστόσο, είναι ο κατάλληλος τρόπος για την επίτευξη της υψηλότερης δυνατής απόδοσης. Στη συγκεκριμένη εργασία, εφαρμόζεται για όλα τα μοντέλα η δεύτερη προσέγγιση.

Κεφάλαιο 4 Εκπαίδευση μοντέλων και αποτελέσματα

Σε αυτό το κεφάλαιο, θα γίνει αναλυτική περιγραφή των μεθόδων και των τεχνικών που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων. Πολλές τεχνικές είναι κοινές και για τις δύο βάσεις, επομένως αυτές θα αναφερθούν στις ίδιες ενότητες, ενώ θα αφιερωθούν δύο ενότητες για τις βάσεις όπου θα αναλυθούν οι λεπτομέρειες και τα αποτελέσματα για καθεμιά.

Η εκπαίδευση των νευρωνικών δικτύων έγινε με χρήση της pytorch (έκδοση 2.0.1) [55] στη γλώσσα python 3.10.12 [56]. Η GPU που χρησιμοποιήθηκε είναι η T4 GPU. Όλη η ανάπτυξη έγινε στο Google Colab.

4.1 Επαύξηση δεδομένων

Ένα νευρωνικό δίκτυο αξιολογείται από το πόσο καλά μπορεί να αποδίδει σε δεδομένα στα οποία δεν έχει εκπαιδευτεί. Η ικανότητα ενός μοντέλου να δίνει σωστά αποτελέσματα σε δεδομένα εκτός του συνόλου εκπαίδευσης ονομάζεται γενίκευση (generalization) [57]. Ωστόσο, κατά την εκπαίδευση νευρωνικών δικτύων, συχνά παρατηρείται το φαινόμενο της υπερ-μοντελοποίησης (over-fit) [58]. Ένα νευρωνικό δίκτυο οδηγείται σε υπερ-μοντελοποίηση όταν αρχίζει να μοντελοποιεί τον θόρυβο και τα τυχαία χαρακτηριστικά που υπάρχουν στο σύνολο εκπαίδευσης. Έτσι, το μοντέλο γίνεται πολύ συγκεκριμένο σε πληροφορία που δεν εμπεριέχεται στη συνολική δομή των δεδομένων, με αποτέλεσμα να παρατηρείται πτώση της ικανότητας του σε καινούρια δεδομένα.

Ένας από τους δημοφιλέστερους τρόπους που χρησιμοποιούνται για να αντιμετωπιστεί αυτό το πρόβλημα είναι η χρήση της τεχνικής επαύξησης δεδομένων (data augmentation). Η τεχνική αυτή εφαρμόζει μετασχηματισμούς πάνω στα δεδομένα εισόδου, με αποτέλεσμα τη δημιουργία νέων δεδομένων με διαφορετική μορφή από τα αρχικά. Στην περίπτωση, που τα δεδομένα είναι εικόνες, οι μετασχηματισμοί που εφαρμόζονται αφορούν την αλλαγή φωτισμού και χρώματος, περιστροφή εικόνας, τυχαία περικοπή εικόνας κ.α.

Οι μετασχηματισμοί εφαρμόζονται πάνω σε RGB εικόνες. Στην συγκεκριμένη εργασία, εφαρμόστηκαν οι εξής μετασχηματισμοί από τη βιβλιοθήκη transforms της pytorch [59]:

- Περικοπή εικόνας με τυχαίο μέγεθος στο διάστημα: (0.4 , 1).
- Οριζόντια αναστροφή εικόνας με πιθανότητα 50%.
- Κατακόρυφη αναστροφή εικόνας με πιθανότητα 50%.
- Αλλαγή φωτισμού, αντίθεσης, κορεσμού και απόχρωσης με τυχαίο μέγεθος στα διαστήματα: (0.7 , 1.3) , (0.7 , 1.3) , (0.7 , 1.3) , (-0.1 , 0.1) αντίστοιχα.
- Κανονικοποίηση κάθε καναλιού της εικόνας με μέσες τιμές (0.485, 0.456, 0.406) και τυπική απόκλιση (0.229, 0.224, 0.225) αντίστοιχα για κάθε κανάλι.

Είναι σημαντικό να σημειωθεί ότι η επαύξηση δεδομένων γίνεται την ώρα της εκπαίδευσης των μοντέλων (on the fly). Επομένως, σε κάθε epoch θα παρουσιάζονται στο νευρωνικό δίκτυο διαφορετικά μετασχηματισμένες εικόνες.

4.2 Εξισορρόπηση δεδομένων

Όπως περιεγράφηκε στο Κεφάλαιο 2, τα δεδομένα και στις δύο βάσεις είναι imbalanced ως προς την κατηγορία ταξινόμησης. Για να αντιμετωπιστεί αυτό το πρόβλημα επιλέχθηκε να χρησιμοποιηθεί η τεχνική της υπερ-δειγματοληψίας (oversampling).

Η ιδέα πίσω από τη τεχνική αυτή είναι να παρουσιάζονται στο νευρωνικό δίκτυο, κατά τη διάρκεια της εκπαίδευσης, παραδείγματα από όλες τις κλάσεις με την ίδια συχνότητα, έτσι ώστε αυτό να μπορεί αμερόληπτα να μοντελοποιήσει τα χαρακτηριστικά κάθε κατηγορίας. Συγκεκριμένα, αυτή η μέθοδος αναθέτει μία τιμή (βάρος) σε κάθε κλάση. Το βάρος αυτό αντιπροσωπεύει την εκπροσώπηση της κλάσης στη βάση δεδομένων και ισούται με το λόγο:

$$\frac{1}{\text{συνολικός αριθμός εικόνων που ανήκουν στην κλάση}}.$$

Είναι προφανές ότι όσο μικρότερη εκπροσώπηση έχει μία κλάση στη βάση, τόσο μεγαλύτερο βάρος θα της ανατεθεί. Στη συνέχεια, κάθε είσοδος(εικόνα) λαμβάνει το βάρος που αντιστοιχεί στην κλάση στην οποία ανήκει και έπειτα, γίνεται κανονικοποίηση

όλων των βαρών των εισόδων, έτσι ώστε να αθροίζουν στο 1. Αυτό ουσιαστικά έχει ως αποτέλεσμα την ανάθεση μιας πιθανότητας επιλογής σε κάθε εικόνα. Εικόνες που ανήκουν στις πιο σπάνιες κλάσεις θα έχουν μεγαλύτερη πιθανότητα επιλογής. Αυτή η διαδικασία δημιουργεί σε ένα batch κατά μέσο όρο ίδιο αριθμό εισόδων από κάθε κλάση, κάνοντας εφικτή τη μάθηση από το νευρωνικό δίκτυο, ακόμα και των κλάσεων που περιέχουν πολύ λίγες εικόνες.

Στην pytorch αυτό υλοποιείται με χρήση της `weightRandomSampler` [60].

4.3 Προσδιορισμός υπερ-παραμέτρων

Σε αυτή την ενότητα, προσδιορίζονται οι υπερ-παραμέτροι που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων και για τις δύο βάσεις. Οι υπερ-παραμέτροι προέκυψαν από πειραματισμό και παρατήρηση του αντίκτυπου των διαφορετικών τιμών τους στην επίδοση των μοντέλων.

Αρχικά, για τη συνάρτηση σφάλματος (loss function) επιλέχθηκε η cross entropy loss. Η συγκεκριμένη συνάρτηση σφάλματος χρησιμοποιείται συχνά στα προβλήματα ταξινόμησης πολλών κλάσεων για να υπολογίσει το μέγεθος του σφάλματος ανάμεσα στις προβλέψεις του μοντέλου και τις πραγματικές τιμές.

Για τον αλγόριθμο βελτιστοποίησης (optimization) χρησιμοποιήθηκε ο Adam [61]. Ο συγκεκριμένος αλγόριθμος παρουσιάζει εξαιρετικά αποτελέσματα ως προς τη ταχύτητα και ακρίβεια σύγκλισης. Το βασικό του προτέρημα είναι ότι συνδυάζει πληροφορία του ιστορικού των τιμών της κλίσεως για κάθε παράμετρο, προσαρμόζοντας έτσι το ρυθμό εκμάθησης κάθε παραμέτρου με πιο αποδοτικό τρόπο.

Ο ρυθμός εκμάθησης (learning rate), τέθηκε στη τιμή 0.001. Πάνω στον ρυθμό εκμάθησης χρησιμοποιήθηκε δρομολογητής (scheduler), ο οποίος πολλαπλασιάζει με 0.2 τη τιμή του ρυθμού εκμάθησης κάθε 5 epochs. Στόχος του δρομολογητή είναι σταδιακά να μειώνει το μέγεθος των βημάτων που θα κάνει ο αλγόριθμος βελτιστοποίησης, με σκοπό να μπορεί να συγκλίνει στο ελάχιστο αντί να ταλαντώνεται γύρω του.

Τέλος, ο αριθμός των epoch τέθηκε στο 30. Ωστόσο, κατά τη διάρκεια της εκπαίδευσης, συγκρίνεται σε κάθε epoch η επίδοση του δικτύου με την προηγούμενη καλύτερη επίδοση, και αποθηκεύεται η τρέχουσα έκδοση του δικτύου μόνο αν ξεπερνά στη

σύγκριση την προηγούμενη καλύτερη έκδοση. Επομένως, δεν επιλέγεται πάντα το μοντέλο του 30^{ου} epoch. Αυτή η μέθοδος ονομάζεται πρόωρη-παύση(early-stopping) και είναι άλλος ένας τρόπος που χρησιμοποιείται για να αποφευχθεί η υπερ-μοντελοποίηση.

4.4 Χρήση μεταδεδομένων

Η χρήση των μεταδεδομένων μπορεί να παίξει σημαντικό ρόλο στην αύξηση της ακρίβειας των μοντέλων. Διαισθητικά, είναι λογικό να περιμένουμε κάτι τέτοιο, καθώς μία εξέταση από δερματολόγο περιλαμβάνει, εκτός από την εξέταση της δερματικής βλάβης βάσει της εικόνας της, και ερωτήσεις του ειδικού για την συμπεριφορά της πληγής (πότε εμφανίστηκε, αν έχει αλλάξει, αν αιμορραγεί).

Ωστόσο, είναι σημαντικό να δοθεί προσοχή στο ποιά πεδία θα χρησιμοποιηθούν στην εκπαίδευση του νευρωνικού δικτύου. Συγκεκριμένα, θα πρέπει να ελεγχθεί, ποια πεδία προσφέρουν ουσιαστική πληροφορία στην εκπαίδευση και ποια όχι [62]. Αν ένα πεδίο αποδειχθεί ότι έχει υψηλή συσχέτιση με το πεδίο της διάγνωσης(δηλαδή των κλάσεων στις οποίες ανήκουν οι εικόνες) θα χρησιμοποιείται στην εκπαίδευση του νευρωνικού δικτύου, καθώς θεωρείται ότι προσφέρει χρήσιμη πληροφορία. Αντίθετα, ένα πεδίο που έχει πολύ χαμηλή συσχέτιση με το πεδίο της διάγνωσης μπορεί να παραληφθεί, καθώς η πληροφορία που θα προσφέρει είναι ασύνδετη με το πρόβλημα και επομένως, η χρήση του θα εισάγει θόρυβο στην εκπαίδευση του μοντέλου. Ωστόσο, θα ήταν σημαντικό να διερευνηθούν και διαφορετικοί συνδυασμοί αυτών των δεδομένων, καθώς είναι πιθανό, πεδία που έχουν χαμηλή συσχέτιση, εν τέλει να προσφέρουν πληροφορία όταν συνδυαστούν με άλλα πεδία.

Αρχικά, βασικό σημείο είναι να προσδιοριστεί το είδος κάθε πεδίου. Στα δεδομένα μας ένα πεδίο μπορεί να είναι είτε κατηγορικό(categorical nominal), είτε αριθμητικό(numerical). Τα κατηγορικά πεδία λαμβάνουν τιμές που είναι διακριτές και δεν έχουν κάποια φυσική ή ποσοτική σημασία. Αντίθετα, τα αριθμητικά πεδία λαμβάνουν τιμές από μία συνεχή κατανομή (μία τιμή μπορεί να είναι ένας οποιοσδήποτε πραγματικός αριθμός) και έχουν ποσοτική σημασία.

Ο καθορισμός της συσχέτισης θα γίνει υπολογίζοντας το μέγεθος επίδρασης (effect size) [63] που έχει κάθε μεταβλητή με το πεδίο των κλάσεων. Για την εύρεση της συσχέτισης

μεταξύ κατηγορικών μεταβλητών χρησιμοποιείται ο αλγόριθμος Cramer's V [64], ενώ για την εύρεση της συσχέτισης μεταξύ κατηγορικής και αριθμητικής μεταβλητής γίνεται υπολογισμός του Eta Squared (η^2) από τον αλγόριθμο one-way ANOVA [65]. Είναι σημαντικό να αναφερθεί ότι ο αλγόριθμος one-way ANOVA προϋποθέτει κανονική κατανομή των τιμών της αριθμητικής μεταβλητής, κάτι το οποίο δεν ισχύει απαραίτητα. Σε περίπτωση που η μεταβλητή δεν ακολουθεί κανονική κατανομή, χρησιμοποιείται ο αλγόριθμος Kruskal-Wallis [66] που είναι η μη παραμετρική εναλλακτική της μεθόδου ANOVA. Για να καθοριστεί αν μία μεταβλητή ακολουθεί κανονική κατανομή εφαρμόζεται ο αλγόριθμος Shapiro-Wilk [67].

Η τιμή που παράγει ο αλγόριθμος Cramer's V και η τιμή Eta Squared βρίσκονται στο διάστημα [0,1] όπου το 0 σημαίνει καθόλου συσχέτιση, ενώ το 1 πλήρη ταύτιση. Η ερμηνεία των αποτελεσμάτων βάσει της τιμής για τον αλγόριθμο του Cramer's V είναι η εξής [68]:

- Cramer's V τιμή ≤ 0.2 : Σημαίνει αδύναμη συσχέτιση μεταξύ των δύο μεταβλητών.
- $0.2 < \text{Cramer's V τιμή} \leq 0.6$: Σημαίνει μέτρια συσχέτιση μεταξύ των μεταβλητών.
- $0.6 < \text{Cramer's V τιμή}$: Σημαίνει ισχυρή συσχέτιση μεταξύ των μεταβλητών.

ενώ, αντίστοιχα για το Eta Squared ισχύει [69]:

- $0.01 \leq \eta^2 < 0.06$ σημαίνει αδύναμη συσχέτιση.
- $0.06 \leq \eta^2 < 0.14$ σημαίνει μέτρια συσχέτιση.
- $0.14 \leq \eta^2$ σημαίνει ισχυρή συσχέτιση.

Σε επόμενες ενότητες θα γίνει ανάλυση συσχέτισης για κάθε βάση ξεχωριστά.

4.5 Επεξεργασία και κωδικοποίηση πεδίων

Για να δοθούν τα πεδία που θα επιλέξουμε στο νευρωνικό δίκτυο, είναι σημαντικό πρώτα να υποστούν μια προεπεξεργασία. Στα κατηγορικά πεδία, απαιτείται να γίνει αυτό, καθώς θα πρέπει να τους ανατεθούν αριθμητικές τιμές, ώστε να μπορούν να εισαχθούν στο νευρωνικό δίκτυο. Για τα αριθμητικά πεδία, παρόλο που έχουν ήδη αριθμητικές τιμές, χρειάζεται να γίνει προεπεξεργασία, καθώς έχει παρατηρηθεί πως, όταν έχουν μέσο όρο(μ) ίσο με 0 και τυπική απόκλιση(σ) ίση με 1, ο αλγόριθμος εκπαίδευσης συγκλίνει πιο

γρήγορα και επιτρέπει στο νευρωνικό δίκτυο να τα μοντελοποιήσει πιο εύκολα [70]. Επιπλέον, η προεπεξεργασία θέτει τις τιμές ενός αριθμητικού πεδίου σε σχετικά μικρές τιμές γύρω από το μηδέν. Αυτό είναι απαραίτητο, καθώς είναι πολύ πιθανό να υπάρχουν πεδία που λαμβάνουν εξ ορισμού μεγάλες τιμές, και άρα να έχουν μεγαλύτερη επίδραση στην εκπαίδευση των μοντέλων από τα άλλα πεδία, κάτι το οποίο είναι ανεπιθύμητο.

Τα κατηγορικά πεδία κωδικοποιούνται χρησιμοποιώντας την τεχνική one-hot encoding. Σε αυτή την τεχνική, ανατίθεται σε κάθε τιμή ένα διάνυσμα με μήκος που ισούται με το πλήθος των διαφορετικών τιμών που λαμβάνει το εκάστοτε πεδίο. Κάθε θέση(index) του διανύσματος αντιστοιχεί σε μία συγκεκριμένη τιμή. Το διάνυσμα περιέχει σε όλες τις θέσεις του 0, εκτός από μία θέση που περιέχει τη τιμή 1. Η θέση που περιέχει 1 αντιστοιχεί στη συγκεκριμένη τιμή του πεδίου που λαμβάνει η εκάστοτε είσοδος. Για παράδειγμα, το πεδίο hurt λαμβάνει δύο διαφορετικές τιμές: False, True. Αρχικά, το διάνυσμα που αντιστοιχεί στο πεδίο θα έχει μήκος δύο. Αν υποθέσουμε ότι στο False αντιστοιχεί η θέση(index) 0 του διανύσματος και στο True η θέση 1, τότε μία είσοδος με hurt == False, θα λαμβάνει το διάνυσμα [1,0] , ενώ μία είσοδος με hurt == True, θα λαμβάνει το διάνυσμα [0,1].

Σημειώνεται πως στα δεδομένα και των δύο βάσεων υπάρχουν πεδία που δεν περιέχουν τιμή σε κάποιες εισόδους (nan). Για τα κατηγορικά πεδία, οι τιμές αυτές συμπληρώνονται πιθανοτικά με μία από τις υπόλοιπες έγκυρες τιμές του πεδίου. Η πιθανότητα επιλογής μίας έγκυρης τιμής είναι ανάλογη της συχνότητας εμφάνισής της. Έτσι, τιμές που περιέχονται περισσότερο σε ένα πεδίο, θα έχουν μεγαλύτερη πιθανότητα να ανατεθούν σε μία άγνωστη τιμή. Με αυτό τον τρόπο διατηρείται η κατανομή των τιμών στο εκάστοτε πεδίο. Περαιτέρω τεχνικές εφαρμόστηκαν, όπως συμπλήρωση των τιμών με το mode (συχνότερη τιμή) καθώς και πρόβλεψη αυτών των τιμών με χρήση RandomForest, ωστόσο καμία από τις μεθόδους δεν φάνηκε να πλεονεκτεί των υπολοίπων.

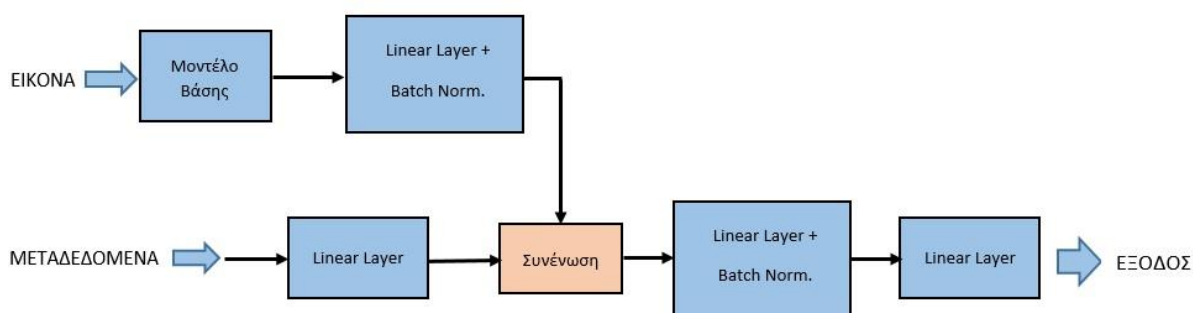
Για τα αριθμητικά πεδία, εφαρμόζεται η τεχνική του standard scale. Σε αυτή τη τεχνική, αρχικά υπολογίζεται για το πεδίο ο μέσος όρος των τιμών του(μ), καθώς επίσης και η τυπική απόκλιση(σ). Έπειτα, κάθε τιμή του πεδίου (x), λαμβάνει τη νέα του τιμή από τον εξής τύπο:

$$z = \frac{x - \mu}{\sigma}$$

Για την εφαρμογή του standard scale χρησιμοποιήθηκε η κλάση StandardScaler της scikit [71]. Οι είσοδοι των αριθμητικών πεδίων που δεν έχουν κάποια έγκυρη τιμή, λαμβάνουν τη τιμή 0.

4.6 Αρχιτεκτονική νευρωνικού δικτύου με μεταδεδομένα

Στο Κεφάλαιο 3 παρουσιάστηκαν μοντέλα νευρωνικών δικτύων που δέχονται για είσοδο μόνο εικόνες. Σε αυτή την ενότητα θα παρουσιαστεί η αρχιτεκτονική νευρωνικού δικτύου που θα δέχεται ως είσοδο την εικόνα και τα μεταδεδομένα της. Η αρχιτεκτονική που υιοθετείται ώστε να συνδυαστεί η πληροφορία της εικόνας με την πληροφορία των μεταδεδομένων φαίνεται στο σχήμα 4.1 και θα αναφέρεται ως metaModel.



Σχήμα 4.1 Η αρχιτεκτονική του metaModel

Συγκεκριμένα, χρησιμοποιείται ένα μοντέλο που εκπαιδεύτηκε μόνο στις εικόνες (μοντέλο βάσης), το οποίο δέχεται για είσοδο μία εικόνα και παράγει ένα διάνυσμα που αντιπροσωπεύει τα χαρακτηριστικά της. Το διάνυσμα που λαμβάνεται είναι η έξοδος του τελευταίου συνελκτικού επιπέδου, ακριβώς πριν τα γραμμικά επίπεδα που χρησιμοποιούνται για τη ταξινόμηση. Στη συνέχεια, το διάνυσμα που περιέχει τα χαρακτηριστικά της εικόνας διέρχεται από ένα γραμμικό επίπεδο διαστάσεων: (μέγεθος διανύσματος χαρακτηριστικών, 256) με batch normalization και έπειτα συνενώνεται (concatenate) με το διάνυσμα των μεταδεδομένων το οποίο επίσης έχει διέλθει από ένα γραμμικό επίπεδο διαστάσεων: (27, 64).

Το διάνυσμα που προκύπτει διέρχεται από δύο συνεχόμενα γραμμικά επίπεδα διαστάσεων: (320, 64) και (64, αριθμός κλάσεων), με το τελευταίο να κάνει τις τελικές προβλέψεις.

Τέλος, είναι σημαντικό να καθοριστεί, ποια κομμάτια του μοντέλου θα εκπαιδευτούν. Μία επιλογή θα ήταν να χρησιμοποιήσουμε τα μοντέλα βάσης απλώς για να εξάγουν τα χαρακτηριστικά των εικόνων χωρίς να συμμετέχουν στην εκπαίδευση (ως feature extractors) και να αφήσουμε μόνο τα γραμμικά επίπεδα να ανανεώνουν τις τιμές τους. Η δεύτερη επιλογή είναι να εκπαιδεύσουμε το συνολικό μοντέλο από την αρχή με το κόστος μιας πιο χρονοβόρας εκπαίδευσης. Και σε αυτή την περίπτωση, επιλέχθηκε η δεύτερη προσέγγιση.

4.7 Εκπαίδευση και αποτελέσματα στη βάση PAD-UFES-20

4.7.1 Διαχωρισμός των δεδομένων σε δεδομένα εκπαίδευσης / αξιολόγησης

Λόγω του μικρού αριθμού εικόνων που υπάρχουν στη βάση PAD-UFES-20 (μόνο 2298), είναι σημαντικό να υπάρξει ειδική μέριμνα στον διαχωρισμό των δεδομένων. Ο κλασικός τρόπος διαχωρισμού, όπου ένα μεγάλο ποσοστό των δεδομένων χρησιμοποιείται για εκπαίδευση και το υπόλοιπο για αξιολόγηση, δεν είναι ο πιο κατάλληλος για βάσεις με μικρό μέγεθος, καθώς είναι πιθανό, το σύνολο αξιολόγησης να μην είναι αντιπροσωπευτικό των συνολικών δεδομένων, με αποτέλεσμα να εξαχθούν λανθασμένα συμπεράσματα ως προς την επίδοση και ικανότητα του μοντέλου. Για το λόγο αυτό, αποφασίστηκε σε αυτή τη βάση να χρησιμοποιηθεί διαχωρισμός 5-fold cross validation.

Τα δεδομένα, χωρίστηκαν σε πέντε διαφορετικά folds με όσο γίνεται ίδιο πλήθος δεδομένων μεταξύ τους, ενώ δόθηκε προσοχή ώστε όλες οι εικόνες που αντιστοιχούν στην ίδια πληγή να κατανέμονται στο ίδιο fold. Στη συνέχεια, όλα τα μοντέλα εκπαιδεύτηκαν πέντε φορές, όπου κάθε φορά ένα διαφορετικό fold χρησιμοποιείται σαν σύνολο αξιολόγησης και τα υπόλοιπα τέσσερα ως σύνολα εκπαίδευσης. Τέλος, παίρνουμε τη γενική εικόνα της ικανότητας ενός μοντέλου υπολογίζοντας τον μέσο όρο (mean) και την τυπική απόκλιση (std) των πέντε διαφορετικών πειραμάτων για κάθε μετρική αξιολόγησης. Αυτή η μέθοδος, έχει το πλεονέκτημα ότι επιτρέπει στα μοντέλα να εκπαιδευτούν και να αξιολογηθούν σε όλα τα δεδομένα εισόδου, δημιουργώντας μια πιο αντιπροσωπευτική εικόνα για την επίδοσή τους.

4.7.2 Εκπαίδευση και αποτελέσματα μόνο με εικόνες

Εκπαιδεύουμε τα μοντέλα στα πέντε folds μόνο πάνω στις εικόνες, χωρίς τη χρήση των μεταδεδομένων. Οι παράμετροι της εκπαίδευσης αναφέρθηκαν στην Ενότητα 4.3. Ο μέσος όρος και η τυπική απόκλιση των αποτελεσμάτων σε μορφή επί τοις εκατό(%) για όλα τα μοντέλα φαίνονται στον Πίνακα 4.1.

Πίνακας 4.1 Αποτελέσματα πειραμάτων στη βάση PAD μόνο με εικόνες

	Accuracy	Average			Weighted			AUC
		Precision	Sensitivity	F-Score	Precision	Sensitivity	F-Score	
MobileNet	72.44 ± 1.58	67.23 ± 3.43	64.21 ± 3.52	64.86 ± 3.48	72.56 ± 1.77	72.44 ± 1.58	72.28 ± 1.63	90.81 ± 0.99
GoogleNet	68.75 ± 0.55	64.56 ± 1.50	64.41 ± 4.17	63.82 ± 2.68	69.92 ± 0.97	68.75 ± 0.55	69.07 ± 0.49	90.31 ± 0.82
Resnet34	64.00 ± 1.50	60.99 ± 3.26	62.75 ± 2.45	60.29 ± 1.41	69.36 ± 0.72	64.00 ± 1.50	65.58 ± 1.18	89.62 ± 0.62
EfficientNet	72.30 ± 1.92	68.15 ± 4.10	66.40 ± 2.65	66.07 ± 1.66	73.64 ± 1.61	72.30 ± 1.92	72.66 ± 1.70	91.77 ± 0.58
MobileNet + SA	71.83 ± 1.30	68.88 ± 3.97	65.42 ± 3.35	66.11 ± 3.04	72.36 ± 1.28	71.83 ± 1.30	71.83 ± 1.24	91.45 ± 0.80
Ensemble	75.70 ± 1.21	73.41 ± 2.70	69.41 ± 1.94	70.60 ± 1.38	75.94 ± 1.12	75.70 ± 1.21	75.63 ± 1.15	92.02 ± 1.10

Όπως παρατηρείται, η συγχώνευση μοντέλων παρουσιάζει την καλύτερη επίδοση σε όλα τα metrics. Από τα ξεχωριστά μοντέλα, φαίνεται ότι το EfficientNet και τα MobileNets έχουν μοντελοποιήσει καλύτερα τις εικόνες, καθώς πέτυχαν υψηλότερες επιδόσεις σε σχέση με τα υπόλοιπα μοντέλα, ενώ μεταξύ τους παρουσιάζουν παρόμοια αποτελέσματα. Τέλος, τα GoogleNet και Resnet34 έδωσαν χειρότερα αποτελέσματα, με το Resnet34 να έχει εμφανώς τις χαμηλότερες επιδόσεις σε όλα τα metrics.

Στην Εικόνα 4.1 παρατίθεται ενδεικτικά το confusion matrix για το MobileNet + SA με κανονικοποιημένες τις τιμές κάθε γραμμής ως προς το άθροισμα τους. Σε κάθε κελί (i, j) η τιμή ερμηνεύεται ως το ποσοστό των στοιχείων που ανήκουν στην κλάση που αντιστοιχεί

στη γραμμή i , και ταξινομήθηκαν στην κλάση που αντιστοιχεί στην στήλη j . Αυτό θα μας επιτρέψει να δούμε το ποσοστό που έχει ταξινομηθεί σωστά για κάθε κλάση, καθώς και ποιες κλάσεις συγχέονται πιο συχνά. Οι τελικές τιμές προέκυψαν λαμβάνοντας τον μέσο όρο των ποσοστών για κάθε fold. Γενικά, θα θέλαμε οι τιμές της διαγωνίου να είναι όσο υψηλότερες γίνεται.



Εικόνα 4.1 Το confusion matrix για το MobileNet + Sa χωρίς εικόνες

Παρατηρούμε ότι, το νευρωνικό δίκτυο είναι αποτελεσματικό για τις κλάσεις ACK, BCC, NEV ενώ αντίθετα, για την κλάση SCC υπάρχει μειωμένη επίδοση. Μεγάλο πλήθος εικόνων που ανήκουν στην SCC, έχουν ταξινομηθεί λανθασμένα στην κλάση BCC. Αυτό έχει λογική εξήγηση καθώς, οι δύο κλάσεις έχουν κοινή εικόνα και ως εκ τούτου είναι λογικό να συγχέονται. Το πρόβλημα των λανθασμένων ταξινομήσεων γίνεται ιδιαίτερως σοβαρό όταν εικόνες που απεικονίζουν καρκίνο του δέρματος, ταξινομούνται σε διαφορετική κλάση. Τέτοια παραδείγματα είναι η σύγχυση της κλάσης MEL με τις κλάσεις NEV και SEK, καθώς επίσης και η σύγχυση των κλάσεων BCC και SCC με την κλάση ACK.

4.7.3 Ανάλυση των μεταδεδομένων της βάσης

Από τα πεδία που υπάρχουν στην βάση PAD-UFES-20, αριθμητικά είναι μόνο τα πεδία age, diameter_1, diameter_2, ενώ όλα τα υπόλοιπα είναι κατηγορικά. Κάνοντας χρήση των

αλγορίθμων που αναφέρθηκαν στην Ενότητα 4.4, καταλήγουμε στις συσχετίσεις του πεδίου διάγνωσης με τα υπόλοιπα πεδία, που παρουσιάζονται στον Πίνακα 4.2. Για τα αριθμητικά πεδία χρησιμοποιήθηκε ο αλγόριθμος Kruskal-Wallis, καθώς δεν ακολουθούσαν κανονική κατανομή.

Πίνακας 4.2 Συσχετίσεις του πεδίου diagnostic με τα υπόλοιπα πεδία της βάσης PAD

Πεδίο	Τύπος πεδίου	Συσχέτιση με diagnostic	Μέγεθος Συσχέτισης
age	Αριθμητικό	0.2075	Ισχυρή
background_father	Κατηγορικό	0.073	Ασθενής
background_mother	Κατηγορικό	0.04	Ασθενής
bleed	Κατηγορικό	0.4990	Μέτρια
cancer_history	Κατηγορικό	0.0	Ασθενής
changed	Κατηγορικό	0.3856	Μέτρια
diameter_1	Αριθμητικό	0.027	Ασθενής
diameter_2	Αριθμητικό	0.030	Ασθενής
drink	Κατηγορικό	0.0	Ασθενής
elevation	Κατηγορικό	0.5203	Μέτρια
fitspatrick	Κατηγορικό	0.1203	Ασθενής
gender	Κατηγορικό	0.069	Ασθενής
grew	Κατηγορικό	0.5007	Μέτρια
has_piped_water	Κατηγορικό	0.070	Ασθενής
has_sewage_system	Κατηγορικό	0.0753	Ασθενής
hurt	Κατηγορικό	0.3923	Μέτρια
itch	Κατηγορικό	0.5266	Μέτρια
pesticide	Κατηγορικό	0.1330	Ασθενής
region	Κατηγορικό	0.2837	Μέτρια
skin_cancer_history	Κατηγορικό	0.059	Ασθενής
smoke	Κατηγορικό	0.1279	Ασθενής

Με αυτά τα στοιχεία, θα πρέπει να καθοριστεί ποια πεδία θα χρησιμοποιηθούν στην εκπαίδευση του μοντέλου. Επίσης πρέπει να ληφθεί υπόψιν ότι τα μοντέλα που θα αναπτυχθούν σε αυτή τη βάση θα χρησιμοποιηθούν στην εφαρμογή και επομένως, ο χρήστης θα κληθεί να θέσει τιμές σε αυτά τα πεδία. Είναι σημαντικό ωστόσο, η εφαρμογή

να είναι εύκολη στη χρήση της, που σημαίνει ότι τα πεδία δεν πρέπει να απαιτούν πληροφορίες που δύσκολα ανακτώνται από το χρήστη. Για το λόγο αυτό, τα πεδία `diameter_1` και `diameter_2` που αφορούν το μέγεθος της δερματικής βλάβης, παραλείπονται. Επιπλέον, τα πεδία `pesticide`, `fitpatrick`, `background_father`, `background_mother` παραλείπονται, καθώς αφορούν πληροφορίες στις οποίες είναι πιθανό να μην έχει πρόσβαση ο χρήστης. Επιπρόσθετα, χρησιμοποιώντας μοντέλα RandomForest ελέγχθηκαν διαφορετικοί συνδυασμοί πεδίων, χωρίς τις εικόνες, έτσι ώστε να μελετηθούν οι αλληλεπιδράσεις μεταξύ τους και να καθοριστεί ο συνδυασμός πεδίων που θα αποφέρει την υψηλότερη επίδοση. Τελικά, καταλήγουμε στα πεδία: `age`, `bleed`, `changed`, `elevation`, `grew`, `hurt`, `itch`, `region`.

4.7.4 Εκπαίδευση και αποτελέσματα με χρήση μεταδεδομένων

Τα μοντέλα που αναπτύχθηκαν χωρίς τη χρήση μεταδεδομένων, χρησιμοποιήθηκαν τώρα ως μοντέλα βάσης για τα μοντέλα που θα κάνουν χρήση των μεταδεδομένων, όπως περιεγράφηκε στην Ενότητα 4.6. Τα συνολικά μοντέλα (`metaModels`) εκπαιδεύτηκαν ξανά πάνω στα ίδια `folds` με τις ίδιες παραμέτρους εκπαίδευσης που προσδιορίστηκαν στην ενότητα 4.3. Τα αποτελέσματα φαίνονται στον Πίνακα 4.3

Πίνακας 4.3 Αποτελέσματα πειραμάτων στη βάση PAD με τη χρήση μεταδεδομένων

	Accuracy	Average			Weighted			AUC
		Precision	Sensitivity	F-Score	Precision	Sensitivity	F-Score	
MobileNet	77.74 ± 1.47	73.26 ± 3.08	70.15 ± 3.95	70.91 ± 3.65	77.73 ± 0.73	77.74 ± 1.47	77.49 ± 1.02	92.47 ± 1.40
GoogleNet	74.28 ± 0.87	70.98 ± 3.74	70.23 ± 3.84	70.28 ± 3.63	75.44 ± 1.30	74.28 ± 0.87	74.73 ± 1.08	92.79 ± 0.67
Resnet34	72.72 ± 1.03	69.17 ± 1.95	71.05 ± 3.56	68.74 ± 2.29	77.08 ± 1.23	72.72 ± 1.50	74.13 ± 0.81	93.14 ± 0.88
EfficientNet	77.71 ± 1.44	74.71 ± 1.56	72.92 ± 2.15	73.25 ± 2.16	78.36 ± 0.73	77.71 ± 1.44	77.81 ± 0.98	93.70 ± 1.01
MobileNet + SA	77.92 ± 1.96	75.28 ± 2.81	70.83 ± 2.88	72.23 ± 2.35	77.50 ± 1.99	77.92 ± 1.96	77.45 ± 1.90	92.62 ± 0.82
Ensemble	79.80 ± 1.24	76.49 ± 2.26	73.28 ± 4.02	74.05 ± 3.31	79.43 ± 1.53	79.80 ± 1.24	79.41 ± 1.51	94.07 ± 0.91

Παρατηρείται πως η χρήση των μεταδεδομένων έχει αυξήσει τις επιδόσεις των νευρωνικών δικτύων σε όλες τις μετρικές αξιολόγησης. Η συγχώνευση μοντέλων πετυχαίνει τις καλύτερες επιδόσεις γενικά, ενώ από τα ξεχωριστά μοντέλα το EfficientNet παρουσιάζει καλύτερα αποτελέσματα όσον αφορά τα average Sensitivity και F-Score. Στην Εικόνα 4.2 παρατίθεται το confusion matrix για το MobileNet + SA, προ-επεξεργασμένο με τον ίδιο τρόπο που έγινε στην υποενότητα 4.7.2.



Εικόνα 4.2 Confusion Matrix για το MobileNet + SA με τη χρήση μεταδεδομένων

Όπως παρατηρείται, όλες οι τιμές τις διαγωνίου αυξήθηκαν, ωστόσο οι κλάσεις SCC και MEL συνεχίζουν να έχουν χαμηλές επιδόσεις. Δοκιμάστηκαν τεχνικές για να αυξηθεί η απόδοση των δικτύων πάνω σε αυτές τις δύο κλάσεις, όπως η χρήση weighted συνάρτηση σφάλματος, ωστόσο αυτό είχε αρνητική επίδραση στις υπόλοιπες κλάσεις και εν τέλει απορρίφθηκε.

4.7.5 Επιλογή μοντέλου για χρήση στην εφαρμογή

Όπως εξηγείται στην υποενότητα 5.2.4, το μοντέλο που θα επιλεγεί να χρησιμοποιηθεί, θα ενσωματωθεί στην εφαρμογή ως Module. Επομένως, είναι σημαντικό να λάβουμε υπόψιν, εκτός από τα αποτελέσματα των μοντέλων που αναπτύχθηκαν, και το μέγεθός

τους. Στον Πίνακα 4.4 παρουσιάζεται ο αριθμός των παραμέτρων για κάθε metaModel με το αντίστοιχο μοντέλο - βάσης .

Πίνακας 4.4 Αριθμός παραμέτρων των μοντέλων

Μοντέλο βάσης	Αριθμός παραμέτρων
MobileNet	3.241.334
GoogleNet	5.885.670
Resnet34	21.439.366
EfficientNet	20.528.790
MobileNet + SA	3.323.989
Ensemble	54.419.149

Από τα αποτελέσματα παρατηρήθηκε ότι η συγχώνευση μοντέλων έχει πετύχει τις υψηλότερες επιδόσεις, ωστόσο ο αριθμός των παραμέτρων του είναι ιδιαίτερα υψηλός. Αντίθετα, το MobileNet και MobileNet + SA έχουν πολύ μικρό αριθμό παραμέτρων σε σχέση με τα υπόλοιπα μοντέλα. Επιπλέον, και τα δύο μοντέλα αποδίδουν αρκετά ικανοποιητικά όπως είδαμε στον Πίνακα 4.3. Ωστόσο, το MobileNet + SA παρουσιάζει ελαφρώς πιο βελτιωμένα αποτελέσματα από το MobileNet, επομένως είναι αυτό που θα επιλεγεί για να χρησιμοποιηθεί στην εφαρμογή.

4.8 Εκπαίδευση και αποτελέσματα στη βάση ISIC-2019

4.8.1 Διαχωρισμός των δεδομένων σε εκπαίδευσης / αξιολόγησης

Για τη βάση δεδομένων ISIC χρησιμοποιήθηκε διαμερισμός 80%/10%/10% των δεδομένων για τα σύνολα εκπαίδευσης(training), αξιολόγησης(validation) και ελέγχου(testing) αντίστοιχα. Επιπλέον, δόθηκε προσοχή ώστε οι εικόνες που απεικονίζουν την ίδια πληγή, να ταξινομηθούν στο ίδιο σύνολο. Ο συγκεκριμένος διαμερισμός επιλέχθηκε έναντι του cross validation, κυρίως λόγω του γεγονότος ότι το πλήθος των μοντέλων και των εικόνων για αυτή τη βάση καθιστά αδύνατη από πλευράς πόρων τη χρήση cross validation στη συγκεκριμένη εργασία. Επιπλέον, το πλήθος των δεδομένων είναι αρκετά μεγάλο και άρα τα σύνολα αξιολόγησης και ελέγχου μπορούν να περιέχουν μια αντιπροσωπευτική εικόνα της βάσης.

4.8.2 Εκπαίδευση και αποτελέσματα μόνο με εικόνες

Τα μοντέλα εκπαιδεύτηκαν πάνω στο σύνολο εκπαίδευσης με τις ίδιες παραμέτρους που αναφέρθηκαν στην ενότητα 4.3. Τα αποτελέσματα όλων των μοντέλων πάνω στο σύνολο αξιολόγησης (validation set) παρατίθενται στον Πίνακα 4.5. Οι αριθμοί είναι σε μορφή επί τοις εκατό(%).

Πίνακας 4.5 Αποτελέσματα πειραμάτων στη βάση ISIC χωρίς τη χρήση μεταδεδομένων

	Accuracy	Average			Weighted			AUC
		Precision	Sensitivity	F-Score	Precision	Sensitivity	F-Score	
MobileNet	73.80	57.82	59.35	57.98	75.34	73.80	74.31	93.82
GoogleNet	72.98	60.38	61.29	60.23	74.47	72.98	73.55	93.51
Resnet34	70.08	57.64	61.51	58.63	72.93	70.08	71.20	92.56
EfficientNet	75.93	63.05	59.04	59.73	77.01	75.93	76.17	94.16
MobileNet + SA	73.96	61.84	60.25	60.55	75.48	73.96	74.49	92.64
Ensemble	78.00	65.87	64.69	64.79	78.11	78.00	77.93	94.76

Παρατηρούμε ότι η συγχώνευση μοντέλων παρουσιάζει την καλύτερη επίδοση σε όλες τις μετρικές ενώ, τα ξεχωριστά μοντέλα φαίνεται να έχουν όλα παρόμοια απόδοση. Στην Εικόνα 4.3 παρατίθεται το confusion matrix για το συγχωνευμένο μοντέλο με κανονικοποιημένες τις τιμές κάθε γραμμής ως προς το άθροισμα τους. Αυτό θα μας επιτρέψει να δούμε το ποσοστό που έχει ταξινομηθεί σωστά για κάθε κλάση, καθώς και ποιες κλάσεις συγχέονται πιο συχνά. Σημειώνεται πως αυτό το confusion matrix δεν έχει κάποιο νόημα ως προς τις στήλες, αλλά μόνο όταν εξετάζεται κατά γραμμές.



Εικόνα 4.3 Confusion Matrix για το Ensemble Model χωρίς τη χρήση μεταδεδομένων

Παρατηρούμε ότι, το νευρωνικό δίκτυο δεν είναι αποτελεσματικό για τις κλάσεις AK, SCC, ενώ στις κλάσεις BKL, DF, MEL αποδίδει μέτρια. Οι περισσότερες εικόνες που ανήκουν στην SCC, έχουν ταξινομηθεί λανθασμένα στις κλάσεις BCC και AK. Αυτό το φαινόμενο παρατηρήθηκε και για τη βάση PAD και είναι αναμενόμενο καθώς οι τρεις κλάσεις έχουν πολλές φορές παρόμοια εικόνα. Επιπλέον, παρατηρούμε ότι η κλάση AK παρουσιάζει πολύ κακά αποτελέσματα, καθώς μεγάλο ποσοστό των εικόνων της ταξινομήθηκε λανθασμένα στις κλάσεις BCC, BKL και MEL.

4.8.3 Ανάλυση των μεταδεδομένων της βάσης

Από τα πεδία που υπάρχουν στην βάση ISIC-2019, κατηγορικά είναι τα πεδία `anatom_site_general` και `gender`, ενώ αριθμητικό είναι το πεδίο `age_approx`.

Κάνοντας χρήση των αλγορίθμων που αναφέρθηκαν στην Ενότητα 4.4, καταλήγουμε στον ακόλουθο πίνακα συσχετίσεων. Για το πεδίο `age_approx` χρησιμοποιήθηκε ο αλγόριθμος Kruskal - Wallis, καθώς δεν ακολουθούσε κανονική κατανομή.

Πίνακας 4.6 Συσχετίσεις του πεδίου category με τα υπόλοιπα πεδία της βάσης ISIC

Πεδίο	Τύπος πεδίου	Συσχέτιση με category	Μέγεθος Συσχέτισης
age_approx	Αριθμητικό	0.2859	Ισχυρή
anatom_site_general	Κατηγορικό	0.1726	Ασθενής
gender	Κατηγορικό	0.091	Ασθενής

Παρατηρούμε ότι, πεδίο age_approx έχει σχετικά ισχυρή συσχέτιση με το πεδίο category, το πεδίο anatom_site_general έχει ασθενή συσχέτιση, ενώ το πεδίο gender δεν έχει καθόλου.

Για την εκπαίδευση των μοντέλων με χρήση μεταδεδομένων χρησιμοποιήθηκαν τα πεδία age_approx και anatom_site_general, ενώ δοκιμάστηκε και η εισαγωγή του πεδίου gender, χωρίς ωστόσο να έχει επίδραση στα αποτελέσματα.

4.8.4 Εκπαίδευση και αποτελέσματα με χρήση μεταδεδομένων

Τα μοντέλα που εκπαιδεύτηκαν στην προηγούμενη υπό-ενότητα χρησιμοποιήθηκαν ως μοντέλα βάσης για τα μοντέλα που κάνουν χρήση μεταδεδομένων, όπως περιεγράφηκε στην ενότητα 4.6. Χρησιμοποιώντας τις ίδιες παραμέτρους εκπαίδευσης, λαμβάνουμε τα αποτελέσματα που φαίνονται στον Πίνακα 4.7.

Πίνακας 4.7 Αποτελέσματα πειραμάτων στη βάση ISIC με τη χρήση μεταδεδομένων

	Accuracy	Average			Weighted			AUC
		Precision	Sensitivity	F-Score	Precision	Sensitivity	F-Score	
MobileNet	76.19	62.30	60.88	61.32	76.21	76.19	76.16	93.96
GoogleNet	73.49	58.35	58.78	58.39	74.74	73.49	73.96	93.21
Resnet34	72.04	59.98	59.46	59.01	74.97	72.04	73.14	93.62
EfficientNet	77.38	65.11	63.13	63.25	77.63	77.38	77.30	94.51
MobileNet + SA	75.41	61.55	61.44	61.10	75.52	75.41	75.39	94.40
Ensemble	78.40	67.00	63.02	64.32	78.77	78.40	78.41	95.03

Σε αυτή τη βάση, η χρήση των μετα-δεδομένων έχει αυξήσει ελάχιστα τις επιδόσεις των νευρωνικών, εκτός του GoogleNet και του συγχωνευμένου μοντέλου. Αυτό εξηγείται από το γεγονός ότι τα νευρωνικά δίκτυα είχαν ελάχιστη αύξηση της απόδοσής τους και επομένως το συγχωνευμένο μοντέλο δεν είχε επιπλέον πληροφορία που μπορούσε να μοντελοποιήσει. Στην Εικόνα 4.4 παρατίθεται το confusion matrix για το συγχωνευμένο μοντέλο, προεπεξεργασμένο με τον ίδιο τρόπο που έγινε στην ενότητα 4.8.3.



Εικόνα 4.4 Confusion Matrix για το Ensemble Model με τη χρήση μεταδεδομένων

Παρατηρούμε ότι παρόλο που αυξήθηκε ελάχιστα η κλάση AK, μειώθηκε η επίδοση στην κλάση SCC με κύριο χαρακτηριστικό την αύξηση της σύγχυσης μεταξύ των εικόνων της με την κλάση BCC. Τέλος, σημαντική πτώση παρουσιάστηκε στη κλάση VASC, όπου πολλές εικόνες της ταξινομούνται λανθασμένα στην κλάση MEL.

4.8.5 Αποτελέσματα στο σύνολο ελέγχου (test set)

Σύμφωνα με τα αποτελέσματα στο σύνολο αξιολόγησης παρατηρούμε ότι λαμβάνουμε την καλύτερη επίδοση όταν χρησιμοποιείται η συγχώνευση των μοντέλων. Για να λάβουμε μια αμερόληπτη εκτίμηση της ικανότητας του μοντέλου σε καινούρια δεδομένα, αξιολογούμε το συγχωνευμένο μοντέλο στο σύνολο ελέγχου. Τα αποτελέσματα φαίνονται στον Πίνακα 4.8.

Πίνακας 4.8 Αποτελέσματα Ensemble Model στο σύνολο ελέγχου

Ensemble Model	Accuracy	Average			Weighted			AUC
		Precision	Sensitivity	F-Score	Precision	Sensitivity	F-Score	
Εικόνες	73.50	65.30	63.26	63.12	76.17	73.50	74.33	91.19

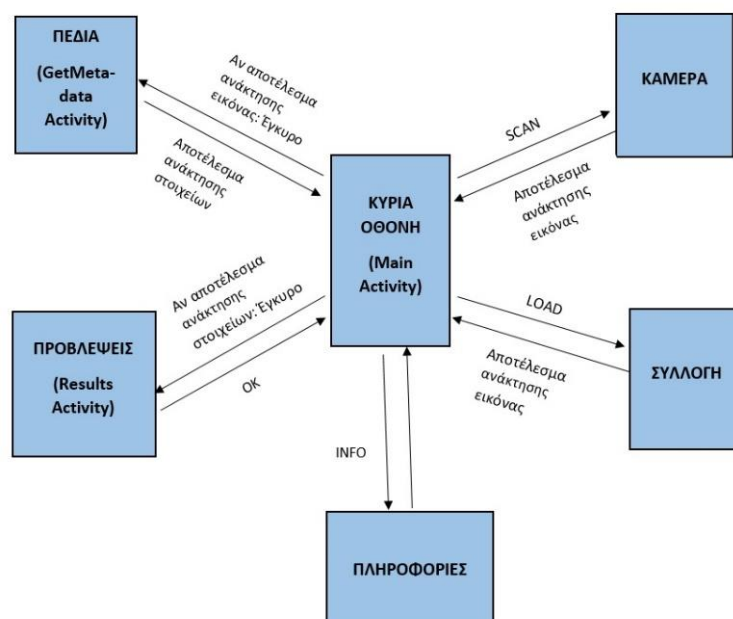
Κεφάλαιο 5 Ανάπτυξη εφαρμογής

Σε αυτό το κεφάλαιο, θα γίνει αναλυτική παρουσίαση του τρόπου ανάπτυξης και χρήσης της εφαρμογής. Η εφαρμογή αναπτύχθηκε στο Android Studio [73] και στη γλώσσα προγραμματισμού Java.

Σκοπός της εφαρμογής είναι να λάβει μία εικόνα και κάνοντας χρήση του νευρωνικού δικτύου που αναπτύχθηκε για τη βάση PAD, να εξαγει προβλέψεις για το ποια δερματική πάθηση μπορεί να απεικονίζεται στη φωτογραφία. Η εικόνα μπορεί να ληφθεί κατά τη διάρκεια χρήσης της εφαρμογής από την κάμερα, είτε να γίνει φόρτωση μιας προϋπάρχουσας φωτογραφίας που είναι αποθηκευμένη στο κινητό. Γενικές οδηγίες που μπορούν να δοθούν στο χρήστη, ώστε η εφαρμογή να κάνει όσο το δυνατόν πιο έγκυρες προβλέψεις είναι πρωτίστως, η δερματική πληγή να απεικονίζεται ολόκληρη στην εικόνα. Επιπλέον, είναι σημαντικό να υπάρχει καλός φωτισμός, ώστε να διακρίνονται καθαρά όλα τα χαρακτηριστικά του δέρματος.

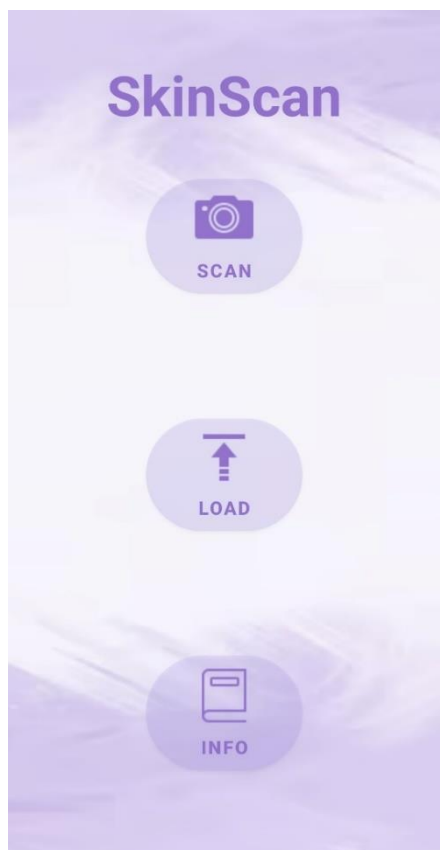
5.1 Παρουσίαση της δομής της εφαρμογής

Μια γενική εικόνα της δομής της εφαρμογής φαίνεται στο σχήμα 5.1.



Σχήμα 5.1 Η ροή λειτουργίας της εφαρμογής

Πιο αναλυτικά, η εφαρμογή περιέχει τέσσερις οθόνες. Η πρώτη είναι η αρχική οθόνη (MainScreen, Εικόνα 5.1) που φορτώνει στην εκκίνηση της εφαρμογής.



Εικόνα 5.1 Η αρχική οθόνη της εφαρμογής

Περιέχει τρία κουμπιά των οποίων η λειτουργία παρουσιάζεται στη συνέχεια.

SCAN: Ξεκινά τη διαδικασία λήψης φωτογραφίας. Συγκεκριμένα ανοίγει την κάμερα, ώστε ο χρήστης να βγάλει μία φωτογραφία της δερματικής βλάβης.

LOAD: Ξεκινά τη διαδικασία φόρτωσης μιας φωτογραφίας που είναι αποθηκευμένη στο κινητό. Συγκεκριμένα ανοίγει την εφαρμογή Συλλογή ή Φωτογραφίες του κινητού, ώστε ο χρήστης να επιλέξει τη φωτογραφία που επιθυμεί να ελέγξει.

INFO: Φορτώνει μία καινούρια οθόνη η οποία περιλαμβάνει πληροφορίες για την εφαρμογή και το τρόπο χρήσης της.

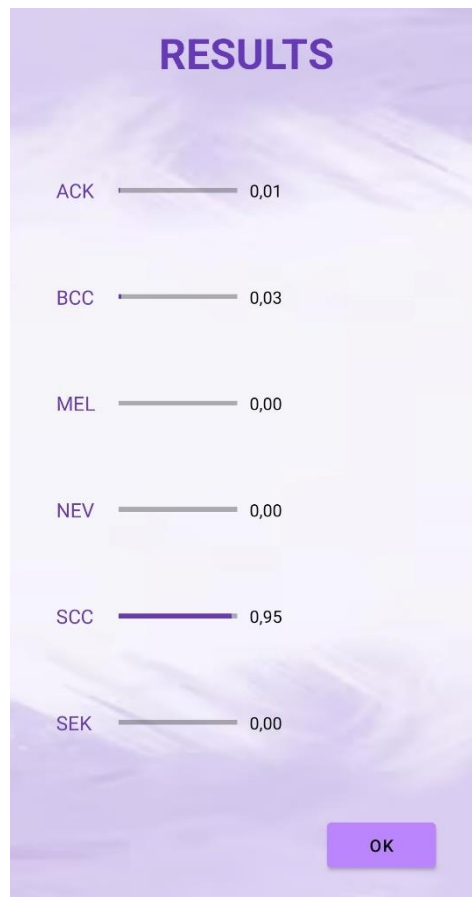
Αφού ληφθεί η φωτογραφία, είτε μέσω της κάμερας (SCAN), είτε μέσω φόρτωσης από τη Συλλογή (LOAD), εμφανίζεται η δεύτερη οθόνη (GetMetadataScreen, Εικόνα 5.2).

The image shows a mobile application screen titled "Data". It contains several input fields and toggle switches. On the left side, there are five toggle switches labeled "Bleed", "Itch", "Grew", "Hurt", and "Changed", all of which are currently turned on (indicated by a purple circle). On the right side, there are two input fields: "Age" with the value "20" and "Region" with the value "ARM" and a dropdown arrow. At the bottom right, there is a purple button labeled "OK".

Εικόνα 5.2 Η οθόνη με τα πεδία των μεταδεδομένων

Ο σκοπός της δεύτερης οθόνης είναι να λάβει από το χρήστη τα απαραίτητα μεταδεδομένα που αφορούν την πληγή. Τα μεταδεδομένα που ζητούνται είναι αυτά που αναφέρθηκαν στην ενότητα 4.7.3. Καθένα από αυτά δέχεται την ίδια επεξεργασία που δέχτηκε και κατά τη διάρκεια εκπαίδευσης.

Όταν ο χρήστης εισάγει τις τιμές στα πεδία, μπορεί να πατήσει το κουμπί OK. Στη συνέχεια, τα μεταδεδομένα και η εικόνα που λήφθηκε σε προηγούμενο στάδιο, εισάγονται στο νευρωνικό δίκτυο το οποίο παράγει τις προβλέψεις. Για την παρουσίαση των προβλέψεων χρησιμοποιείται μία τρίτη οθόνη (ResultsScreen, Εικόνα 5.3).



Εικόνα 5.3 Η οθόνη με τα αποτελέσματα

Σε αυτή την οθόνη ο χρήστης μπορεί να δει τις προβλέψεις του νευρωνικού δικτύου για καθεμιά από τις έξι κλάσεις. Τέλος, ο χρήστης μπορεί να πατήσει το κουμπί OK για να γυρίσει στην κύρια οθόνη.

5.2 Παρουσίαση του τρόπου ανάπτυξης της εφαρμογής

5.2.1 Ανάπτυξη κύριας οθόνης (MainScreen)

Η κύρια οθόνη υλοποιείται από το Activity: MainActivity. Περιλαμβάνει τρία κουμπιά που υλοποιούνται με την κλάση Button [74]. Ένα Button είναι ένα στοιχείο αλληλοεπίδρασης, το οποίο ο χρήστης μπορεί να πατήσει για να εκτελεστεί μια συγκεκριμένη λειτουργία. Συγκεκριμένα, το scan button ξεκινά τη διαδικασία σκαναρίσματος ενεργοποιώντας την κάμερα. Κατά τη διαδικασία αυτή, η εφαρμογή ζητά άδεια από το χρήστη για το άνοιγμα της κάμερας. Εφόσον ο χρήστης παραχωρήσει άδεια, μπορεί στη συνέχεια να κάνει τη

λήψη της φωτογραφίας. Εάν δεν παραχωρήσει άδεια, η εφαρμογή επιστρέφει στην αρχική οθόνη, εμφανίζοντας ένα μήνυμα (Toast) σφάλματος.

Όσον αφορά το load button, αφού λάβει την απαραίτητη άδεια, ανοίγει την εφαρμογή Συλλογή ή Φωτογραφίες του κινητού, έτσι ώστε ο χρήστης να επιλέξει την εικόνα που θέλει να ελέγξει.

Τέλος, το info button, φορτώνει μία καινούρια οθόνη, η οποία περιλαμβάνει ένα μεγάλο κείμενο με λεπτομέρειες που αφορούν την εφαρμογή.

5.2.2 Ανάπτυξη οθόνης μεταδεδομένων (GetMetadataScreen)

Η οθόνη GetMetadataScreen υλοποιείται από το Activity: GetMetadata. Περιλαμβάνει οκτώ πεδία εισόδου με τα οποία ο χρήστης μπορεί να εισάγει τα δεδομένα του. Για τα πεδία bleed, itch, elevation, grew, expanded, hurt, καθώς λαμβάνουν binary τιμές (True ή False), χρησιμοποιείται για το καθένα ένα Switch [75] για να ληφθεί η τιμή τους. Το πεδίο region μπορεί να πάρει πολλαπλές τιμές, επομένως για αυτό το πεδίο χρησιμοποιείται ένα Spinner [76].

Συγκεκριμένα, οι τιμές που μπορεί να πάρει το πεδίο region είναι:

- ABDOMEN (Κουλιακή χώρα)
- ARM (Μπράτσο)
- BACK (Πλάτη)
- CHEST (Στήθος)
- EAR (Αυτί)
- FACE (Πρόσωπο)
- FOOT (Πόδι)
- FOREARM (Πήχης)
- HAND (Χέρι)
- LIP (Χείλη)
- NECK (Λαιμός)
- NOSE (Μύτη)
- SCALP (Κρανίο)
- THIGH (Μηρός)

Τέλος, το πεδίο age δέχεται ως είσοδο έναν αριθμό και επομένως χρησιμοποιείται ένα EditText [77] για να ληφθεί η τιμή του. Αφού ληφθεί η τιμή, μετασχηματίζεται με βάση τον Standard Scaler που εφαρμόστηκε κατά τη διάρκεια της εκπαίδευσης με μέσο όρο: 60,46 και τυπική απόκλιση: 15,89.

Ο χρήστης θα πρέπει να εισάγει για αυτό το πεδίο έναν αριθμό μεταξύ του 0 και του 100. Σε περίπτωση που εισαχθεί ένας αριθμός έξω από αυτό το διάστημα, το πεδίο θα λάβει την τιμή 0. Τέλος, το κουμπί OK, χρησιμοποιείται από το χρήστη, ώστε να επιβεβαιώσει τα δεδομένα που εισήγαγε. Σε αυτό το σημείο το GetMetadataActivity καταστρέφεται, και τα δεδομένα μεταφέρονται πίσω στο MainActivity για να εισαχθούν μαζί με την εικόνα στο νευρωνικό δίκτυο.

5.2.3 Ανάπτυξη της οθόνης αποτελεσμάτων (ResultsScreen)

Η οθόνη αποτελεσμάτων υλοποιείται από το Activity: Results. Περιλαμβάνει τις τελικές προβλέψεις του μοντέλου για κάθε κατηγορία. Οι κατηγορίες είναι αυτές που περιεγράφηκαν στην ενότητα 2.2.

Για την παρουσίαση των προβλέψεων, χρησιμοποιείται για κάθε κατηγορία ένα ProgressBar [78], μαζί με ένα TextView [79] που περιλαμβάνει το ποσοστό. Τέλος, το κουμπί (Button) OK καταστρέφει το Results Activity και επαναφέρει το MainActivity στο προσκήνιο.

5.2.4 Ανάπτυξη νευρωνικού δικτύου στην εφαρμογή

Για τη χρήση ενός νευρωνικού δικτύου σε μία εφαρμογή υπάρχουν δύο επιλογές. Η πρώτη είναι να αναπτυχθεί το μοντέλο σε έναν server με τον οποίο θα επικοινωνεί η εφαρμογή μέσω internet για τη μεταφορά των δεδομένων. Η δεύτερη επιλογή περιλαμβάνει την ενσωμάτωση του νευρωνικού δικτύου στην εφαρμογή ως asset και τη χρήση του απευθείας ως μεταβλητή. Σε κάθε μία από τις προσεγγίσεις υπάρχουν τα αντίστοιχα πλεονεκτήματα και μειονεκτήματα.

Αναφορικά με την προσέγγιση της ανάπτυξης του μοντέλου σε έναν server, εντοπίζονται δύο βασικά πλεονεκτήματα. Το πρώτο πλεονέκτημα σχετίζεται με την ελαχιστοποίηση του υπολογιστικού κόστους της εφαρμογής, καθώς η επεξεργασία δεδομένων τελείται από τον server και όχι από την εφαρμογή. Επιπλέον, οποιεσδήποτε αλλαγές χρειαστεί να

υπάρξουν στο μοντέλο μπορούν να γίνουν απευθείας στον server, χωρίς την παρέμβαση του χρήστη ή της εφαρμογής.

Ωστόσο, ένα από τα μειονεκτήματα αυτής της προσέγγισης εντοπίζεται στην μειωμένη ταχύτητα που μπορεί να παρουσιάζει η εφαρμογή, καθώς τα δεδομένα πρέπει να μεταφέρονται μέσω internet για την επεξεργασία τους.

Από την άλλη πλευρά, η προσέγγιση της ενσωμάτωσης του νευρωνικού δικτύου στην εφαρμογή πλεονεκτεί στη ταχύτητα χρήσης του, καθώς οι προβλέψεις γίνονται τοπικά στην εφαρμογή. Επιπλέον, δεν προϋποθέτει σύνδεση στο internet διασφαλίζοντας έτσι την ομαλή χρήση της εφαρμογής χωρίς πιθανές διακοπές που μπορεί να προκύψουν. Τέλος, εξασφαλίζεται ότι τα δεδομένα του χρήστη θα παραμείνουν ιδιωτικά.

Στα αρνητικά αυτής της προσέγγισης συγκαταλέγονται η μεγάλη χρήση υπολογιστικών πόρων από την εφαρμογή εξαιτίας των λειτουργιών του νευρωνικού δικτύου οι οποίες επιβαρύνουν τη λειτουργία του κινητού. Επιπλέον, παρουσιάζεται δυσκολία στην ενημέρωση και αναβάθμιση της εφαρμογής, αφού για οποιαδήποτε αλλαγή στο μοντέλο, ο χρήστης θα πρέπει να ενημερώσει ή να επαναεγκαταστήσει την εφαρμογή.

Στην εργασία αυτή, το νευρωνικό δίκτυο που επιλέχθηκε να χρησιμοποιηθεί ενσωματώθηκε στην εφαρμογή σύμφωνα με την τελευταία προσέγγιση που αναφέρθηκε, καθώς δεν ήταν εφικτή η χρήση ενός δεσμευμένου server για αυτό το σκοπό.

5.2.5 Ανίχνευση μη έγκυρων εικόνων

Είναι σημαντικό η εφαρμογή να μπορεί να ανιχνεύει εικόνες που δεν είναι έγκυρες, δηλαδή εικόνες που δεν απεικονίζουν μία δερματική πληγή ή η δερματική πληγή δεν φαίνεται καθαρά. Με αυτό το τρόπο εξασφαλίζεται ότι ο χρήστης χρησιμοποιεί σωστά την εφαρμογή και λαμβάνει όσο το δυνατόν πιο ακριβή αποτελέσματα.

Για το σκοπό αυτό ενσωματώθηκε στην εφαρμογή ένα ακόμη νευρωνικό δίκτυο (MobileNet v2 Small) που αναγνωρίζει αν η εικόνα που εισήγαγε ο χρήστης αναπαριστά μία δερματική πληγή. Το μοντέλο δέχεται την εικόνα ακριβώς μετά το scan ή load και αν αποφανθεί ότι η είσοδος είναι μη έγκυρη, εμφανίζει ένα μήνυμα (Toast) στο χρήστη χωρίς να διακόπτει τη ροή της εφαρμογής. Το μοντέλο αυτό εκπαιδεύτηκε χρησιμοποιώντας τις

εικόνες της βάσης PAD ως μία κατηγορία (οι έγκυρες εικόνες) ενώ για τις άκυρες, χρησιμοποιήθηκαν 2298 random εικόνες από τη βάση Open Images [80].

Κεφάλαιο 6 Συμπεράσματα και προτάσεις

6.1 Συμπεράσματα

Σε αυτή τη διπλωματική εργασία αναπτύχθηκαν μοντέλα νευρωνικών δικτύων με σκοπό την κατηγοριοποίηση εικόνων δερματικών πληγών. Συγκεκριμένα, αναπτύχθηκαν μοντέλα για δύο διαφορετικές βάσεις ξεχωριστά. Η μία βάση αφορά δερματοσκοπικές εικόνες, ενώ η δεύτερη αφορά εικόνες που έχουν ληφθεί από κάμερα smartphone. Εξετάστηκαν διάφορες αρχιτεκτονικές, ωστόσο αυτό που αποδίδει καλύτερα φαίνεται να είναι η τεχνική της συγχώνευσης πολλαπλών νευρωνικών δικτύων. Επιπλέον, φάνηκε ότι η εισαγωγή των μεταδεδομένων έπαιξε καθοριστικό ρόλο, αφού όταν χρησιμοποιήθηκαν στην εκπαίδευση των μοντέλων, παρατηρήθηκε αύξηση στις μετρικές αξιολόγησης από τα περισσότερα μοντέλα. Ωστόσο, τα αποτελέσματα και η ικανότητα των νευρωνικών δικτύων παραμένει σχετικά αναξιόπιστη για περίπλοκες εικόνες, ενώ συχνά υπάρχει σύγχυση μεταξύ πολλών κατηγοριών. Το πρόβλημα της σύγχυσης γίνεται ιδιαίτερα σοβαρό όταν δεδομένα που πρέπει να ταξινομηθούν σε μία κλάση που αφορά καρκίνο του δέρματος (όπως για παράδειγμα η SCC), καταλήγουν να ταξινομούνται σε κάποια άλλη κατηγορία που αφορά μία πιο ήπια δερματική πάθηση (όπως η ACK). Είναι σημαντικό κάθε εικόνα που αφορά καρκίνο του δέρματος να ταξινομείται σωστά στην κατηγορία που ανήκει. Για αυτό τον σκοπό, χρησιμοποιήθηκαν τεχνικές κατά τη διάρκεια της εκπαίδευσης, όπου δίνεται περισσότερο έμφαση σε κάποιες κατηγορίες από κάποιες άλλες, εφαρμόζοντας *weighted loss function*. Ωστόσο τα αποτελέσματα για τις υπόλοιπες κλάσεις ήταν πολύ αρνητικά και εν τέλει αυτή η μέθοδος απορρίφθηκε.

Είναι προφανές ότι για τη βελτίωση της αξιοπιστίας αυτών των μοντέλων, πέρα από την εξέταση διάφορων μεθόδων και αρχιτεκτονικών νευρωνικών δικτύων, είναι σημαντικό να συλλεχθούν παραπάνω εικόνες για κάθε κατηγορία, αλλά και πιο κρίσιμα μεταδεδομένα. Παρατηρήθηκε ότι τα μοντέλα της βάσης PAD-UFES-20 παρουσίασαν μεγαλύτερη αύξηση στην επίδοσή τους, σε σχέση με αυτά που αναπτύχθηκαν για τη βάση ISIC-2019 όταν χρησιμοποιήθηκαν τα μεταδεδομένα. Αυτό πιθανώς συνέβη γιατί η βάση PAD περιείχε πιο περιγραφικά δεδομένα αναφορικά με την πληγή.

Τέλος, η εφαρμογή που αναπτύχθηκε είναι ικανή να ταξινομεί εικόνες που έχουν παρόμοια μορφή με αυτές της βάσης PAD, ωστόσο για να γίνει σωστή ταξινόμηση απαιτείται η λήψη της φωτογραφίας να γίνει κάτω από καλό φωτισμό και σε κοντινή απόσταση από την πληγή.

Στο ακόλουθο link παρατίθεται ο κώδικας: <https://github.com/primi9/Thesis>

6.2 Μελλοντικές επεκτάσεις

Όπως αναφέρθηκε στην προηγούμενη ενότητα, για την επίτευξη υψηλότερης επίδοσης των μοντέλων απαιτούνται περισσότερα δεδομένα. Για το σκοπό αυτό, στο μέλλον θα μπορούσε να αναπτυχθεί σελίδα στο internet όπου θα μπορούν ειδικοί δερματολόγοι να ανεβάζουν εικόνες και δεδομένα μαζί με τη διάγνωση, με σκοπό την ανάπτυξη ενός μεγάλου συνόλου δεδομένων.

Τέλος, είναι σημαντικό να διερευνηθούν περισσότερες αρχιτεκτονικές νευρωνικών δικτύων, όπως για παράδειγμα τα CapsuleNets, και οι Visual Transformers (ViT).

Βιβλιογραφία

- [1] Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. P. C. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, 145, 511–518. <https://doi.org/10.1016/j.measurement.2019.05.076>
- [2] Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Network: <https://arxiv.org/abs/1705.03820>
- [3] Zhou, J., Hu, B., Feng, W., Zhang, Z., Fu, X., Shao, H., Wang, H., Jin, L., Ai, S., & Ji, Y. (2023). An ensemble deep learning model for risk stratification of invasive lung adenocarcinoma using thin-slice CT. *Npj Digital Medicine*, 6(1), 1–12. <https://doi.org/10.1038/s41746-023-00866-z>
- [4] Ma, R., Vanstrum, E. B., Lee, R., Chen, J., & Hung, A. J. (2020). Machine learning in the optimization of robotics in the operative field. *Current opinion in urology*, 30(6), 808–816. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7735438/>
- [5] Vegter M. W. (2018). Towards precision medicine; a new biomedical cosmology. *Medicine, health care, and philosophy*, 21(4), 443–456. <https://doi.org/10.1007/s11019-018-9828-z>
- [6] Dermoscopy: <https://www.ncbi.nlm.nih.gov/books/NBK537131/>
- [7] Actinic Keratosis: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6939186>
- [8] Basal-Cell Carcinoma: <https://www.ncbi.nlm.nih.gov/books/NBK482439/>
- [9] BKL: <https://www.ncbi.nlm.nih.gov/books/NBK545285/>
- [10] Dermatofibroma: <https://www.ncbi.nlm.nih.gov/books/NBK470538/>
- [11] Melanoma: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4346328/>
- [12] Squamous Cell Carcinoma: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7319751/>
- [13] Nevus: <https://www.ncbi.nlm.nih.gov/books/NBK559270/>
- [14] VASC: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5615389/>

- [15] ISIC Challenge 2019 Dataset: <https://challenge.isic-archive.com/data/#2019>
- [16] PAD-UFES-20 Dataset:
<https://www.sciencedirect.com/science/article/pii/S235234092031115X>
- [17] Steppan, J., & Hanke, S. (n.d.). Analysis of skin lesion images with deep learning.
<https://arxiv.org/pdf/2101.03814v1.pdf>
- [18] Kassem, M. A., Hosny, K. M., & Fouad, M. M. (2020). Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer learning. IEEE Access, 1–1. <https://doi.org/10.1109/access.2020.3003890>
- [19] Panda, S., Sunil Tiwari, A., & Ranjan Prusty, M. (2020). Comparative study on different Deep Learning models for Skin Lesion Classification using transfer learning approach. International Journal of Scientific and Research Publications (IJSRP), 11(1), 219–229. <https://doi.org/10.29322/ijsrp.11.01.2021.p10923>
- [20] Menegola, A., Tavares, J., Fornaciali, M., Li, L., Avila, S., & Valle, E. (n.d.). RECOD Titans at ISIC Challenge 2017. <https://arxiv.org/pdf/1703.04819.pdf>
- [21] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Elling, I. (2019, May 1). Skin Lesion Classification Using Hybrid Deep Neural Networks. IEEE Xplore. <https://doi.org/10.1109/ICASSP.2019.8683352>
- [22] Lee, Y., Jung, S.-H., & Won, H.-H. (n.d.). ISIC 2018:SKIN LESION ANALYSIS TOWARDS MELANOMA DETECTION 1 WonDerM: Skin Lesion Classification with Fine-tuned Neural Networks. <https://arxiv.org/pdf/1808.03426.pdf>
- [23] Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin Lesion Classification Using Ensembles of Multi-Resolution EfficientNets with Meta Data. MethodsX, 100864. <https://doi.org/10.1016/j.mex.2020.100864>
- [24] Codella, N. C. F., Nguyen, Q.-B. ., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development, 61(4/5), 5:1–5:15. <https://doi.org/10.1147/jrd.2017.2708299>

- [25] Datta, S., Abuzar, M., Sargur, S., Srihari, N., & Gao, M. (n.d.). Soft-Attention Improves Skin Cancer Classification Performance. <https://arxiv.org/pdf/2105.03358v3.pdf>
- [26] Sabour, S., Frosst, N., & Hinton, G. (n.d.). Dynamic Routing Between Capsules. <https://arxiv.org/pdf/1710.09829.pdf>
- [27] Lan, Z., Cai, S., He, X., & Wen, X. (2022). FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer. IEEE Access, 10, 76261–76267. <https://doi.org/10.1109/ACCESS.2022.3181225>
- [28] Singh, P., Sizikova, E., & Cirrone, J. (n.d.). CASS: Cross Architectural Self-Supervision for Medical Image Analysis. <https://arxiv.org/pdf/2206.04170v6.pdf>
- [29] Bisla, D., Choromanska, A., Berman, R., Stein, J., Polsky, D., & Perelman, R. (n.d.). Towards Automated Melanoma Detection with Deep Learning: Data Purification and Augmentation. <https://arxiv.org/pdf/1902.06061.pdf>
- [30] Perez, F., Vasconcelos, C., Avila, S., & Valle, E. (n.d.). Data Augmentation for Skin Lesion Analysis. <https://arxiv.org/pdf/1809.01442.pdf>
- [31] Ou, C., Zhou, S., Yang, R., Jiang, W., He, H., Gan, W., Chen, W., Qin, X., Luo, W., Pi, X., & Li, J. (2022). A deep learning based multimodal fusion model for skin lesion diagnosis using smartphone collected clinical images and metadata. <https://doi.org/10.3389/fsurg.2022.1029991>
- [32] Cai, G., Zhu, Y., Wu, Y., Jiang, X., Ye, J., & Yang, D. (2022). A multimodal transformer to fuse images and metadata for skin disease classification. <https://doi.org/10.1007/s00371-022-02492-4>
- [33] Li, W., Zhuang, J., Wang, R., Zhang, J., & Zheng, W.-S. (n.d.). FUSING METADATA AND DERMOSCOPY IMAGES FOR SKIN DISEASE DIAGNOSIS. https://www.isee-ai.cn/~wangruixuan/files/ISBI2020_2.pdf
- [34] Yin, W., Huang, J., Chen, J., & Ji, Y. (2022). A study on skin tumor classification based on dense convolutional networks with fused metadata. <https://doi.org/10.3389/fonc.2022.989894>

- [35] Pacheco, A. G. C., & Krohling, R. A. (2020). The impact of patient clinical information on automated skin cancer detection. *Computers in Biology and Medicine*, 116, 103545. <https://doi.org/10.1016/j.compbiomed.2019.103545>
- [36] HAM10000 Dataset: (c) by ViDIR Group, Department of Dermatology, Medical University of Vienna; <https://doi.org/10.1038/sdata.2018.161>
- [37] BCN_20000 Dataset: (c) Department of Dermatology, Hospital Clínic de Barcelona
- [38] MSK Dataset: (c) Anonymous; <https://arxiv.org/abs/1710.05006>; <https://arxiv.org/abs/1902.03368>
- [39] Handle imbalanced datasets: <https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>
- [40] Calculate Precision, Recall, F1, and more for Deep Learning Models: <https://machinelearningmastery.com/how-to-calculate-precision-recall-f1-and-more-for-deep-learning-models/>
- [41] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. <https://arxiv.org/pdf/1512.03385.pdf>
- [42] Vanishing Gradient Problem: <https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484>
- [43] Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., & Andreetto, M. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. <https://arxiv.org/pdf/1704.04861.pdf>
- [44] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2019). MobileNetV2: Inverted Residuals and Linear Bottlenecks. <https://arxiv.org/pdf/1801.04381.pdf>
- [45] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q., Adam, H., Ai, G., & Brain, G. (n.d.). Searching for MobileNetV3. <https://arxiv.org/pdf/1905.02244v5.pdf>
- [46] Hu, J. (n.d.). Squeeze-and-Excitation Networks. <https://arxiv.org/pdf/1709.01507.pdf>

- [47] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going deeper with convolutions. <https://arxiv.org/pdf/1409.4842.pdf>
- [48] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. <https://arxiv.org/pdf/1905.11946.pdf>
- [49] Tan, M., & Le, Q. (n.d.). EfficientNetV2: Smaller Models and Faster Training. <https://arxiv.org/pdf/2104.00298.pdf>
- [50] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, & Polosukhin, I. (2017). Attention Is All You Need. <https://arxiv.org/pdf/1706.03762.pdf>
- [51] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (n.d.). AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE. <https://arxiv.org/pdf/2010.11929v2.pdf>
- [52] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (n.d.). Self-Attention Generative Adversarial Networks. <https://arxiv.org/pdf/1805.08318.pdf>
- [53] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, 109(1), 43–76. <https://doi.org/10.1109/jproc.2020.3004555>
- [54] Image Net Database: <https://www.image-net.org/>
- [55] Pytorch: <https://pytorch.org/get-started/pytorch-2.0/>
- [56] Python: <https://www.python.org/downloads/release/python-31012/>
- [57] Generalization in Machine Learning: <https://mathanraisharma.medium.com/generalization-in-machine-learning-for-better-performance-51bed74a3820>
- [58] Overfitting in Machine Learning: <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>
- [59] Pytorch Image Transformations: <https://pytorch.org/vision/stable/transforms.html>

- [60] Weighted Random Sampler: <https://medium.com/@zergtant/improving-control-and-reproducibility-of-pytorch-dataloader-with-sampler-instead-of-shuffle-7f795490256e>
- [61] Kingma, D., & Lei Ba, J. (2017). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. <https://arxiv.org/pdf/1412.6980.pdf>
- [62] Feature Selection: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>
- [63] Effect Size Measures: <https://www.analyticsvidhya.com/blog/2022/08/statistical-effect-size-and-python-implementation/>
- [64] Cramer's V Algorithm: <https://www.statstest.com/cramers-v-2/>
- [65] One-way Anova: <https://libguides.library.kent.edu/spss/onewayanova>
- [66] Kruskal-Wallis Algorithm: <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php>
- [67] Shapiro-Wilk Algorithm: <https://math.mit.edu/~rmd/465/shapiro.pdf>
- [68] Cramer's V Interpretation: <https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramers-v>
- [69] Eta Squared Interpretation: https://rpkgs.datanovia.com/rstatix/reference/kruskal_effsize.html
- [70] de Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. The choice of scaling technique matters for classification performance. Applied Soft Computing, 133, 109924. <https://arxiv.org/pdf/2212.12343.pdf>
- [71] Scikit Learn – Standard Scaler: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- [72] Ensemble Learning: <https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/>
- [73] Android Studio Version: <https://developer.android.com/studio/releases/past-releases/as-chipmunk-release-notes>

[74]Android Studio Button:

<https://developer.android.com/reference/android/widget/Button>

[75] Android Studio Switch:

<https://developer.android.com/reference/android/widget/Switch>

[76] Android Studio Spinner:

<https://developer.android.com/develop/ui/views/components/spinner>

[77] Android Studio EditText:

<https://developer.android.com/reference/android/widget/EditText>

[78] Android Studio Progress Bar:

<https://developer.android.com/reference/android/widget/ProgressBar>

[79] Android Studio TextView:

<https://developer.android.com/reference/android/widget/TextView>

[80] Open Image: https://storage.googleapis.com/openimages/web/download_v7.html