



Εθνικό Μετσόβιο Πολυτεχνείο

ΔΠΜΣ 'Επιστήμη Δεδομένων και Μηχανική Μάθηση'

## Ανάπτυξη Android Εφαρμογής Ταξινόμησης Μελανωμάτων



Εργασία στο μάθημα 'Ανάλυση Βιο-Δεδομένων'

### Στοιχεία Μελών Ομάδας

Μπριλλάκης Βασίλειος - 03400104

Συρράφος Δημήτριος - 03400116

Χατσιαγιάν Μελκόν - 03400114

Κετσέας Σταύρος - 03400084

Αθήνα, Ιούνιος 2021

# Πίνακας Περιεχομένων

<b>Κατάλογος Εικόνων</b>	<b>3</b>
<b>Κατάλογος Πινάκων</b>	<b>4</b>
<b>Περίληψη</b>	<b>5</b>
<b>1. Εισαγωγή</b>	<b>6</b>
1.1. Περιγραφή αντικειμένου	6
1.2. Κλινική συνεισφορά	7
1.3. Δομή αναφοράς	7
<b>2. Βιβλιογραφική επισκόπηση</b>	<b>9</b>
<b>3. Δεδομένα</b>	<b>15</b>
3.1. Παρουσίαση και Ανάλυση Δεδομένων	15
3.2. Επαύξηση δεδομένων	17
<b>4. Μεθοδολογία</b>	<b>19</b>
4.1. Δομή Android Εφαρμογής	19
4.2. Ανάπτυξη μοντέλου	20
<b>5. Αποτελέσματα - Σχολιασμός</b>	<b>24</b>
5.1. Μετρικές αξιολόγησης	24
5.2. Αποτελέσματα	25
5.3. Επεξήγηση αποτελεσμάτων - LIME	27
<b>6. Συμπεράσματα - Μελλοντικές επεκτάσεις</b>	<b>30</b>
6.1. Συμπεράσματα	30
6.2. Μελλοντικές επεκτάσεις	31
6.3. Τεχνικές λεπτομέρειες	32
6.4. Συνεισφορά Μελών Ομάδας	32
<b>Βιβλιογραφία</b>	<b>33</b>

# Κατάλογος Εικόνων

<b>Εικόνα 2.1:</b> Αρχιτεκτονική EfficientNet	9
<b>Εικόνα 2.2:</b> Ανισορροπία δεδομένων στο HAM10000 dataset	10
<b>Εικόνα 2.3:</b> Soft Attention block και End to End αρχιτεκτονική Inception ResNet v2 + SA	12
<b>Εικόνα 2.4:</b> Αρχιτεκτονική Nambla-N	14
<b>Εικόνα 3.1:</b> Κατανομή δεδομένων 2019 και 2020	15
<b>Εικόνα 3.2:</b> Κατανομή συνολικών δεδομένων σε κατηγορίες-κλάσεις	16
<b>Εικόνα 3.3:</b> Ενδεικτικά αποτελέσματα επαύξησης δεδομένων	18
<b>Εικόνα 4.1:</b> Αρχιτεκτονική backbone μοντέλου του ensemble	20
<b>Εικόνα 4.2:</b> Focal loss για διαφορετικές τιμές της παραμέτρου $\gamma$	22
<b>Εικόνα 5.1:</b> Σύγκριση αποτελεσμάτων κατά AUC	25
<b>Εικόνα 5.2:</b> Σύγκριση αποτελεσμάτων κατά AUC 2020	26

# Κατάλογος Πινάκων

<b>Πίνακας 2.1:</b> Score κάθε μοντέλου του ensemble του [1]	10
<b>Πίνακας 2.2:</b> Score κάθε Transfer learning Backbone του [4]	11
<b>Πίνακας 2.3:</b> 5-fold CV Score για ResNet50 από το [4]	11
<b>Πίνακας 2.4:</b> Score του Inception ResNet v2 + SA για HAM10000 dataset	13
<b>Πίνακας 2.5:</b> Score του Inception ResNet v2 + SA για ISIC Melanoma Challenge dataset	13
<b>Πίνακας 2.6:</b> Αποτελέσματα Nambla-N στο ISIC dataset 2018	14
<b>Πίνακας 3.1:</b> Συχνότητα σημείων προέλευσης εικόνων από σημεία του σώματος	17
<b>Πίνακας 3.2:</b> Μετασχηματισμοί επαύξησης δεδομένων	18
<b>Πίνακας 4.1:</b> Class Mapping [1]	21

## Περίληψη

Αντικείμενο της συγκεκριμένης εργασίας είναι η ανάπτυξη εφαρμογής Android για την ταξινόμηση μελανωμάτων του δέρματος. Για την υλοποίηση του classification task χρησιμοποιήσαμε τα δεδομένα που παρέχονται από τον International Skin Imaging Collaboration για τον διαγωνισμό Melanoma Classification Challenge. Τα δεδομένα είναι αρχεία JPEG. Αναλύσαμε τα δεδομένα και εξασφαλίσουμε την μη ύπαρξη διπλότυπων στα σύνολα εκπαίδευσης και ελέγχου ώστε να διασφαλίσουμε το unbiasedness του μοντέλου. Αναπτύξαμε ένα μοντέλο ensemble από βαθιά CNN με τη μέθοδο του Transfer Learning. Λόγω υπολογιστικού κόστους υλοποιήθηκε ένα μοντέλο για κάθε backbone μοντέλο του ensemble κάνοντας χρήση σε κάποια μοντέλα και των μεταδεδομένων που παρέχονται στο dataset. Επεκτείναμε τη μεθοδολογία της βιβλιογραφίας εφαρμόζοντας τεχνικές εξισορρόπησης δεδομένων μέσω subsampling και τροποίησης της loss function του νευρωνικού. Έγινε tuning των ποσοστών εκπροσώπησης κάθε κλάσης στη μέθοδο subsampling καθώς και της παραμέτρου ποινής του focal loss. Πραγματοποιήθηκε εξαγωγή επεξήγησης των αποτελεσμάτων με τη χρήση του LIME framework. Το μοντέλο που υλοποιήσαμε χρησιμοποιήθηκε για την ανάπτυξη Android εφαρμογής με τη χρήση του Android studio, το οποίο ενημερώνει τον χρήστη για την πιθανότητα ύπαρξης μελανώματος στην φωτογραφία που ανέβασε στην εφαρμογή καθώς και την LIME επεξήγηση του αποτελέσματος είτε για προσωπική χρήση, είτε για ενημέρωση του επιβλέποντα ιατρού.

# 1. Εισαγωγή

## 1.1. Περιγραφή αντικειμένου

Ο καρκίνος είναι μια από τις κύριες αιτίες θανάτου σε όλο τον κόσμο. Από τους πολλούς τύπους καρκίνου, το μελάνωμα είναι ένας τύπος καρκίνου που επηρεάζει το δέρμα. Ευτυχώς, εάν το μελάνωμα ανιχνευθεί στα αρχικά του στάδια, τότε η σωστή θεραπεία μπορεί να εξασφαλίσει πλήρη ανάρρωση. Σαν αποτέλεσμα, το ποσοστό επιβίωσης για το μελάνωμα είναι πολύ υψηλό. Η έγκαιρη ανίχνευση αυτού του καρκίνου του δέρματος είναι επομένως εξαιρετικά σημαντική.

Ο ακριβής εξοπλισμός είναι απαραίτητος για να εξασφαλιστεί ακριβής και έγκαιρη ανίχνευση κακοηθειών στο δέρμα. Τα τελευταία χρόνια παρουσιάζεται μια τάση για αποκέντρωση της διαδικασίας διάγνωσης καρκίνου του δέρματος μέσω εφαρμογών που δίνουν τη δυνατότητα στο μέσο άνθρωπο να έχει μία πρώτη εικόνα της για την φύση των μελανωμάτων στο δέρμα του.

Παράλληλα, παρατηρείται σημαντική εξέλιξη στο πεδίο της αυτοματοποιημένης διάγνωσης ή της επικουρούμενης από υπολογιστή διάγνωσης (Computer Aided Diagnosis, CAD). Η πρόοδος αυτή ταυτίζεται χρονικά με την πρόοδο της μηχανικής μάθησης (Machine Learning) και την εμφάνιση της βαθιάς μηχανικής μάθησης (Deep Learning) η οποία έχει συντελέσει στην επίτευξη εξαιρετικών αποτελεσμάτων σε πληθώρα εφαρμογών. Σήμερα, η μηχανική μάθηση χρησιμοποιείται για την αυτοματοποίηση όλο και περισσότερων διαδικασιών στο πεδίο της ιατρικής επιστήμης.

Το παραπάνω γεγονός έχει εξαιρετική σημασία, καθώς η ιατρική περίθαλψη είναι ένα προνόμιο που απολαμβάνει ένα περιορισμένο ποσοστό του παγκόσμιου πληθυσμού, ενώ κατά τη γνώμη μας θα έπρεπε να αποτελεί θεμελιώδες δικαίωμα. Επίσης ακόμα και σε περιοχές του πλανήτη όπου η ιατρική περίθαλψη είναι εύκολα προσβάσιμη, δεν είναι εύκολο για κάποιον να επισκέπτεται έναν γιατρό για οποιοδήποτε θέμα αντιμετωπίζει καθώς είναι και χρονοβόρο και πολλές φορές κοστοβόρο. Η πρόοδος της τεχνητής νοημοσύνης στον τομέα της βιοϊατρικής δίνει δυνατότητες αποκέντρωσης της ιατρικής διάγνωσης και την καθιστά προσβάσιμη σε όλο και μεγαλύτερο μέρος του πληθυσμού.

Βασικός σκοπός της παρούσας εργασίας είναι η υλοποίηση ενός μοντέλου διάγνωσης καρκίνου του δέρματος από εικόνες διαφόρων σημείων του σώματος με την χρήση βαθιών συνελικτικών νευρωνικών δικτύων. Στη συνέχεια, σκοπεύουμε να χρησιμοποιήσουμε το μοντέλο που αναπτύξαμε για την ανάπτυξη μίας εφαρμογής Android που θα δίνει τη δυνατότητα στο χρήστη της συγκεκριμένης εφαρμογής να έχει πρόσβαση στο μοντέλο αυτό μέσω μίας εύχρηστης διεπαφής χρήστη. Εκτός αυτού, σε ένα γενικότερο πλαίσιο, σκοπεύουμε να διερευνήσουμε τις επικρατούσες τάσεις στο συγκεκριμένο πεδίο, να τις αξιολογήσουμε και να προσπαθήσουμε να επεκτείνουμε κάποιες ήδη υπάρχουσες μεθοδολογίες συνδυάζοντας τις προσεγγίσεις που μελετήσαμε.

## 1.2. Κλινική συνεισφορά

Κάνοντας χρήση της εφαρμογής και στην συνέχεια του συστήματος που αναφέραμε παραπάνω, ένας άνθρωπος θα μπορεί να είναι σε θέση να λάβει μία διάγνωση μέσω του κινητού του και μία πρώτη ιδέα για το εάν η υποψία του αν μία ελιά θα μπορούσε στην πραγματικότητα να είναι καρκίνος του δέρματος. Συνεπώς, μία τέτοια εφαρμογή θα βοηθούσε ώστε κάποιος να μπορεί να ανιχνεύσει εάν έχει καρκίνο και να οδηγηθεί στον γιατρό εγκαίρως, ώστε να λάβει την θεραπεία του πολύ πιο σύντομα και εν τέλει να θεραπευτεί. Επίσης, η χρήση της εφαρμογής θα μπορούσε να βοηθήσει και ανθρώπους, οι οποίοι είναι είτε αναβλητικοί με την επίσκεψη στον γιατρό, είτε αρνητικοί. Χρησιμοποιώντας την εφαρμογή, θα μπορούσε να είναι ένα σημαντικό κίνητρο ώστε να επισκεφτεί τον γιατρό του.

## 1.3. Δομή αναφοράς

Η συγκεκριμένη εργασία αποτελείται από 5 δομικές ενότητες. Η πρώτη υποενότητα αφορά την ήδη υπάρχουσα βιβλιογραφία που μελετήθηκε και αποτέλεσε το background για την υλοποίηση των μοντέλων που θα αναφερθούν στη συνέχεια. Στη συγκεκριμένη ενότητα γίνεται εκτενής αναφορά στις επικρατούσες τάσεις στο πεδίο της ταξινόμησης μελανωμάτων και στα πλεονεκτήματα που προσφέρει η κάθε προσέγγιση.

Η δεύτερη ενότητα αποσκοπεί στην ολοκληρωμένη παρουσίαση των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων που θα παρουσιαστούν στη συνέχεια. Γίνεται

αναφορά στο είδος των δεδομένων, την κατανομή τους και τα προβλήματα που προκύπτουν από αυτή.

Η τρίτη ενότητα αφορά τη μεθοδολογία που εφαρμόστηκε για την εκπόνηση της συγκεκριμένης εργασίας. Παρουσιάζονται τα μοντέλα, οι τεχνικές εκπαίδευσης και λεπτομέρειες υλοποίησης, όπως οι τεχνολογίες που χρησιμοποιήθηκαν.

Στην τέταρτη ενότητα παρουσιάζονται τα αποτελέσματα των διαφόρων πειραμάτων που εκτελέστηκαν ενώ γίνεται και σύντομος σχολιασμός των σημαντικότερων αποτελεσμάτων.

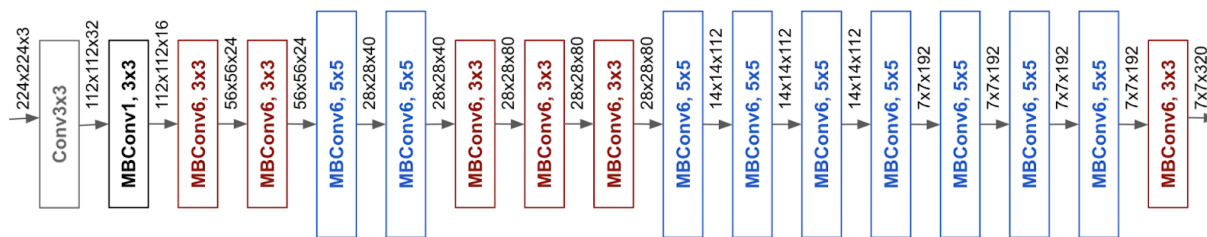
Στην πέμπτη ενότητα γίνεται μία ανακεφαλαίωση των προηγούμενων ενοτήτων, παρουσιάζονται τα βασικά συμπεράσματα που προέκυψαν από την εκπόνηση της εργασίας και γίνεται αναφορά στις μελλοντικές επεκτάσεις και ιδέες που θα μπορούσαν να υλοποιηθούν βασιζόμενες στα αποτελέσματα της συγκεκριμένης εργασίας.



## 2. Βιβλιογραφική επισκόπηση

Λόγω των επιτυχιών που έχουν προκύψει από εφαρμογές βαθιάς μηχανικής μάθησης στο συγκεκριμένο πεδίο όλο και περισσότερες έρευνες και δημοσιεύσεις έρχονται στο προσκήνιο και κατά συνέπεια η βιβλιογραφία είναι αρκετά εκτενής. Παρόλα αυτά, λόγω της φύσης του προβλήματος και λόγω των εξαιρετικών επιδόσεων των βαθιών συνελικτικών δικτύων σε εφαρμογές που εμπεριέχουν δεδομένα εικόνων, οι προσεγγίσεις που εμφανίζονται στη βιβλιογραφία δε διαφέρουν πολύ μεταξύ τους, αντιθέτως παρατηρούνται ομοιότητες στις τεχνικές και τεχνολογίες που χρησιμοποιούνται. Στο σημείο αυτό γίνεται μία προσπάθεια επισκόπησης της ήδη υπάρχουσας βιβλιογραφίας και παρατίθεται ένας σύντομος σχολιασμός των βασικών στοιχείων κάθε αρχιτεκτονικής και μεθοδολογίας.

Με μία γρήγορη ματιά στη σχετική βιβλιογραφία γίνεται εύκολα αντιληπτή η κυριαρχία των βαθιών συνελικτικών νευρωνικών δικτύων (Deep Convolutional Neural Networks) ως αρχιτεκτονική και της μεταφοράς μάθησης (Transfer Learning) ως μεθόδου εκπαίδευσης. Επίσης, μία ακόμη τεχνική που βελτιώνει τις επιδόσεις των διαφόρων μοντέλων που αναφέρονται στη σχετική βιβλιογραφία είναι η μέθοδος ensemble, δηλαδή ο συνδυασμός των προβλέψεων επιμέρους μοντέλων για την εξαγωγή μίας από κοινού πρόβλεψης. Ο Qishen Ha et al. [1] κατάφεραν να νικήσουν στο διαγωνισμό του International Skin Imaging Collaboration “SIIM - ISIC Melanoma Classification Challenge” εφαρμόζοντας τις παραπάνω τεχνικές. Συγκεκριμένα δημιούργησαν ένα ensemble από 18 μοντέλα μεταφοράς μάθησης από πολύ βαθιά συνελικτικά νευρωνικά δίκτυα (16 EfficientNets [10], 2 ResNets[11]). Η τελική πρόβλεψη προκύπτει ως σταθμισμένος μέσος όρος των προβλέψεων των επιμέρους μοντέλων, όπου ο παράγοντας βάρους αντιστοιχεί στην συνολική ακρίβεια του κάθε μοντέλου που συγκροτεί το ensemble.

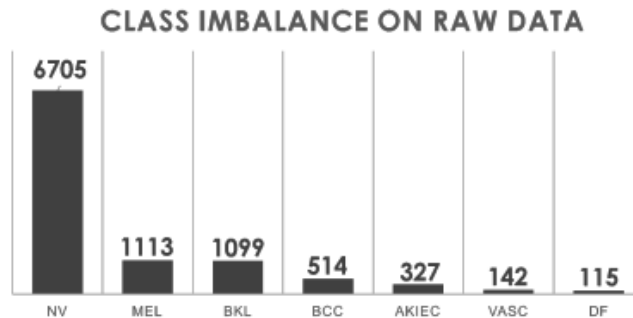


Εικόνα 2.1 : Αρχιτεκτονική EfficientNet

Model	Backbone	Target	Input	Resize	Metadata	Init	lr	Epochs	cv_all	cv_2020	private_LB	public_LB
1	B3	9c	768	512	yes	3e-5	18	0.9762	0.9300	0.9305	0.9182	
2	B4	9c	768	640		3e-5	15	0.9767	0.9400	0.9299	0.9342	
3	B4	9c	768	768		3e-5	15	0.9771	0.9408	0.9264	0.9251	
4	B4	9c	768	640	yes	3e-5	15	0.9765	0.9408	0.9302	0.9221	
5	B4	9c	1024	896		2e-5	15	0.9744	0.9390	0.9320	0.9281	
6	B4	9c	512	448		3e-5	15	0.9748	0.9307	0.9213	0.9002	
7	B5	9c	512	384	yes	3e-5	15	0.9752	0.9329	0.9167	0.9350	
8	B5	9c	768	640		1.5e-5	15	0.9771	0.9428	0.9291	0.9216	
9	B5	4c	768	640		1.5e-5	15	0.9765	0.9384	0.9362	0.9260	
10	B5	9c	512	448	yes	3e-5	15	0.9751	0.9397	0.9363	0.9387	
11	B6	9c	768	640		3e-5	15	0.9756	0.9444	0.9408	0.9283	
12	B6	9c	768	576		3e-5	15	0.9761	0.9443	0.9266	0.9245	
13	B6	9c	512	448		3e-5	15	0.9742	0.9383	0.9261	0.9154	
14	B7	9c	512	384		3e-5	15	0.9748	0.9394	0.9193	0.9170	
15	B7	9c	768	576		1e-5	15	0.9764	0.9432	0.9260	0.9271	
16	B7	9c	768	640		1e-5	15	0.9754	0.9440	0.9304	0.9115	
17	SE_X101	9c	768	640		3e-5	15	0.9739	0.9428	0.9295	0.9337	
18	Nest101	9c	768	640		2e-5	15	0.9728	0.9396	0.9320	0.9267	
Ensemble									0.9845	0.9600	0.9490	0.9442

Πίνακας 2.1 : Score κάθε μοντέλου του ensemble του [1]

Στη δημοσίευση των Duyen N.T. Le et al. παρουσιάζεται μία ακόμα προσέγγιση η οποία βασίζεται στην τεχνική της μεταφοράς μάθησης από βαθιά συνελικτικά νευρωνικά δίκτυα. Συγκεκριμένα, οι Duyen N.T. Le et al. χρησιμοποίησαν ένα βαθύ residual CNN (ResNet50), χωρίς να δημιουργήσουν κάποιο ensemble από classifiers όπως στην προηγούμενη βιβλιογραφική αναφορά. Στη συγκεκριμένη προσέγγιση δόθηκε έμφαση στην αντιμετώπιση του προβλήματος της ανισορροπίας δεδομένων. Στη συγκεκριμένη έρευνα χρησιμοποιήθηκε το dataset HAM10000, το οποίο δεν είναι το ίδιο με το σύνολο δεδομένων που χρησιμοποιήθηκε στην παρούσα εργασία, όμως παρουσιάζει εξίσου μεγάλη ανισορροπία στο πλήθος δειγμάτων καθ' κλάσης.



Εικόνα 2.2: Ανισορροπία δεδομένων στο HAM10000 dataset

Οι Duyen N.T. Le et al. προτίνουν σα λύση σε αυτό το φαινόμενο την τροποποίηση της συνάρτησης κόστους του νευρωνικού δικτύου, έτσι ώστε να δίνει μεγαλύτερο βάρος (ποινή) στα σφάλματα που αφορούν δείματα που ανήκουν στις κλάσεις που είναι δύσκολα διαχωρίσιμες λόγω μη επαρκούς υποστήριξης στο σύνολο εκπαίδευσης. Η συγκεκριμένη συνάρτηση κόστους ονομάζεται focal loss και αποτελεί μία παραλλαγή της συνάρτησης κόστους cross entropy

$$p_t = \begin{cases} p, y = 1 \\ 1 - p, otherwise \end{cases}$$

Cross - Entropy Loss

$$CE(p, y) = CE(p_t) = -\log(p_t)$$

Focal Loss

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

Exp	Methods	Accuracy	Precision	Recall	F1-score
6	ResNet50	<b>90</b>	<b>81</b>	<b>80</b>	<b>80</b>
7	VGG16	85	73	61	63
8	MobileNet	88	76	72	73
9	EfficientNet B1	89	76	76	75

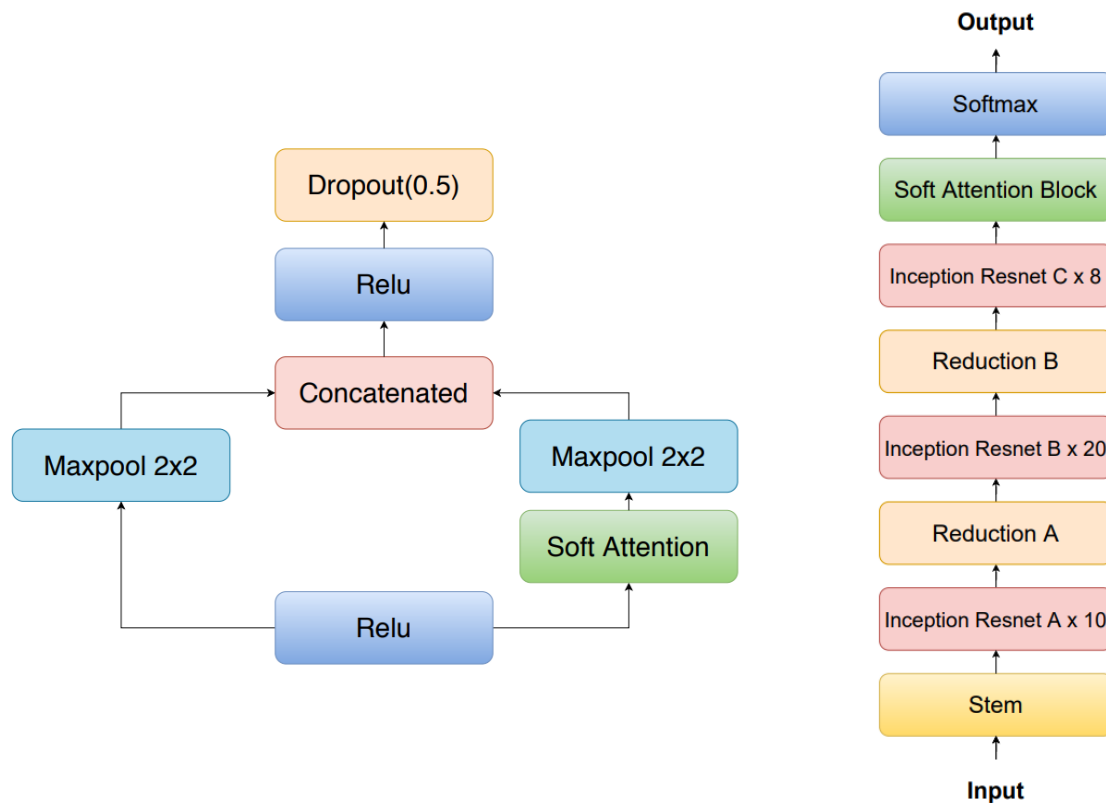
Πίνακας 2.2: Score κάθε Transfer learning Backbone του [4]

Fold	Acc	Pre	Recall	F1-score	Top 2	Top 3	AUC
1	90	85	80	82	96.85	98.73	98.62
2	90	82	77	79	96.18	98.73	98.39
3	91	83	81	82	96.99	99.00	98.32
4	91	83	82	83	96.85	98.46	98.48
5	91	85	78	81	<b>97.46</b>	<b>99.33</b>	98.62
Ensemble	<b>93</b>	<b>88</b>	84	86	<b>97.46</b>	<b>99.33</b>	<b>98.63</b>
By TTA	<b>93</b>	<b>88</b>	<b>85</b>	<b>87</b>	96.85	99.20	98.41

Πίνακας 2.3: 5-fold CV Score για ResNet50 από το [4]

Οι Soumyya Kanti Datta et al. προτείνουν τη χρήση ενός μηχανισμού προσοχής (Soft Attention) και την εφαρμογή του σε ένα μοντέλο μεταφοράς μάθησης από ένα βαθύ συνελικτικό

νευρωνικό δίκτυο. Η ιδέα του Soft Attention πηγάζει από το γεγονός ότι στις εικόνες που χρησιμοποιούνται συνήθως μόνο ένα μικρό μέρος παρέχει πληροφορία για την ταξινόμηση που θέλουμε να πραγματοποιήσουμε. Έτσι, ο μηχανισμός Soft Attention χρησιμοποιείται για να δοθεί έμφαση από το μοντέλο στις περιοχές αυτές και κατ' επέκταση να οδηγηθεί σε καλύτερα αποτελέσματα. Ο μηχανισμός προσοχής εφαρμόζεται στο τελικό επίπεδο του μοντέλου, αφότου δηλαδή έχει εξαχθεί η τελική διανυσματική αναπαράσταση και αποσκοπεί στον περιορισμό της επιρροής στοιχείων του διανύσματος που δεν αφορούν το πρόβλημα της ταξινόμησης μελανώματος και στην επισήμανση των στοιχείων της διανυσματικής αναπαράστασης που περιέχουν πληροφορία σχετικά με την περιοχή ενδιαφέροντος.



Εικόνα 2.3: Soft Attention block και End to End αρχιτεκτονική Inception ResNet v2 + SA

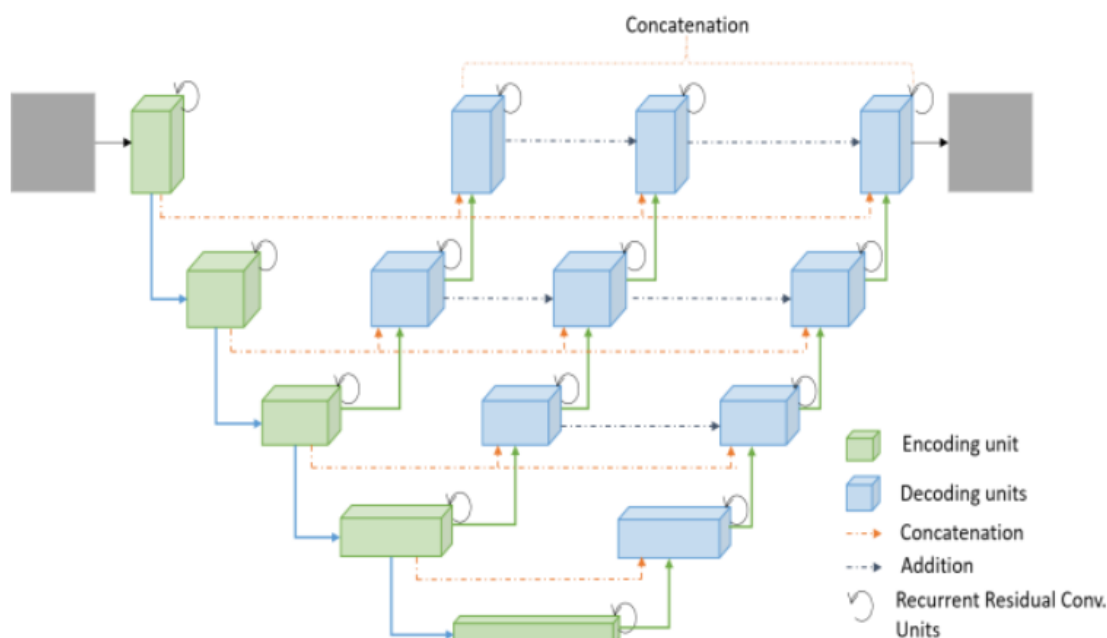
Model	Avg AUC	Precision	Accuracy
Loss balancing and ensemble[5]	0.941	-	0.926
Single Model Deep Learning[29]	0.974	-	0.864
Data classification augmentation[19]	0.975	-	0.853
Two path CNN model[14]	-	-	0.886
Various Deep CNN (Baseline) [16]	0.979	0.890	-
<b>IRv2+SA(Proposed Approach)</b>	<b>0.984</b>	<b>0.937</b>	<b>0.934</b>

Πίνακας 2.4: Score του Inception ResNet v2 + SA για HAM10000 dataset

Networks	AUC	Accuracy	Sensitivity	Specificity
ResNet50 [6]	0.948	0.842	0.867	0.837
RAN50 [27]	0.942	0.862	0.878	0.859
SEnet50 [7]	0.952	0.863	0.856	0.865
ARL-CNN50[31]	0.958	0.868	0.878	<b>0.867</b>
IRv2 <sub>12x12</sub> +SA	0.935	0.898	<b>0.945</b>	0.711
<b>IRv2<sub>5x5</sub>+SA</b>	<b>0.959</b>	<b>0.904</b>	0.916	0.833

Πίνακας 2.5: Score του Inception ResNet v2 + SA για ISIC Melanoma Challenge dataset

Στην επόμενη δημοσίευση γίνεται μία προσπάθεια προσέγγισης του προβλήματος ταξινόμησης μελανωμάτων με τη μέθοδο του semantic segmentation, δηλαδή της κατάτμησης της εικόνας σε περιοχές όπου εντοπίζονται σημεία ενδιαφέροντος. Η προτεινόμενη αρχιτεκτονική από τους Zahangir Alom et al. [2] είναι το δίκτυο NABLA-N, μια παραλλαγή του διαδομένου UNet. Η διαφοροποίηση του συγκεκριμένου μοντέλου έγκειται στο γεγονός ότι αντί να γίνεται αποκωδικοποίηση μόνο της τελικής αναπαράστασης, γίνεται αποκωδικοποίηση και άλλων αναπαραστάσεων από προηγούμενα επίπεδα του δικτύου και παράγεται μία τελική αναπαράσταση από την συνένωση των επιμέρους αναπαραστάσεων. Αυτό έχει σαν αποτέλεσμα να αξιοποιούνται και άλλα λιγότερο high level χαρακτηριστικά που εξάγονται από το δίκτυο και που μπορούν να οδηγήσουν σε βελτίωση της ποιότητας της τελικής αναπαράστασης.



Εικόνα 2.4: Αρχιτεκτονική Nambla-N

Method	Precision	Recall	Accuracy	F1-score	IoU	DI
R2U-Net	0.9629	0.9556	0.9556	0.9553	0.8719	0.8841
$\nabla^2$ -Net <sub>A</sub>	0.9631	0.9545	0.9545	0.9551	0.8688	0.8777
$\nabla^2$ -Net <sub>B</sub>	0.9642	0.9570	0.9570	0.9572	0.8751	0.8877
$\nabla^2$ -Net <sub>AB</sub>	0.9659	0.9582	0.9581	0.9582	0.8803	0.8927
$\nabla^2$ -Net <sub>AB</sub> + TL	<b>0.9707</b>	<b>0.9636</b>	<b>0.9636</b>	<b>0.9644</b>	<b>0.8883</b>	<b>0.8960</b>
$\nabla^2$ -Net <sub>AB</sub> + TL + Data Aug.	0.9668	0.9603	0.9603	0.9603	0.8821	0.8929

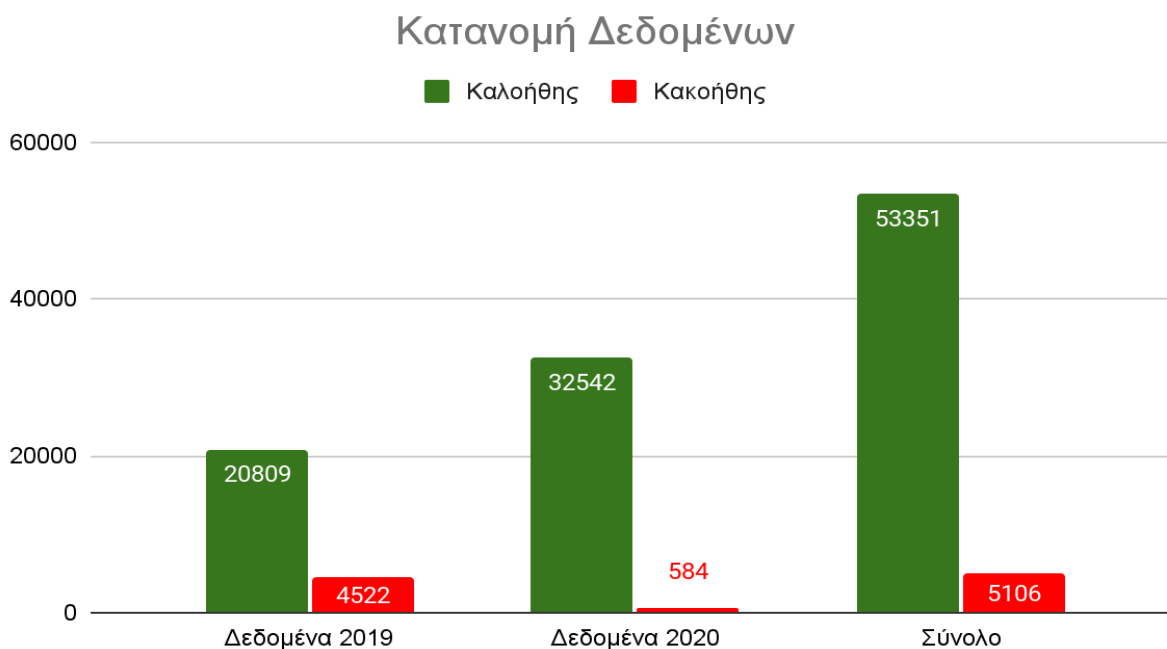
Πίνακας 2.6: Αποτελέσματα Nambla-N στο ISIC dataset 2018

Η πλειονότητα από τις παραπάνω μεθοδολογίες έχει χρησιμοποιηθεί για την ανάπτυξη μοντέλων με σκοπό τη συμμετοχή στο ετήσιο διαγωνισμό του International Skin Imaging Collaboration “SIIM - ISIC Melanoma Classification Challenge”. Το State-of-the Art μοντέλο, κρίνοντας με βάση τα leaderboards του διαγωνισμού, είναι το ensemble από πολλά μοντέλα μεταφοράς μάθησης, ενώ το μοντέλο που χρησιμοποιείται κατά κόρον σε αυτή την περίπτωση είναι το EfficientNet με το ResNet να ακολουθεί. Για το λόγο αυτό, η προσέγγισή μας πάνω στο πρόβλημα ακολουθεί τις προτάσεις των Qishen Ha et al. [1] προσπαθώντας να εφαρμόσουμε μεθόδους εξισορρόπησης δεδομένων από το [4] καθώς και άλλες μεθόδους που εμφανίζονται στη σχετική βιβλιογραφία.

### 3. Δεδομένα

#### 3.1. Παρουσίαση και Ανάλυση Δεδομένων

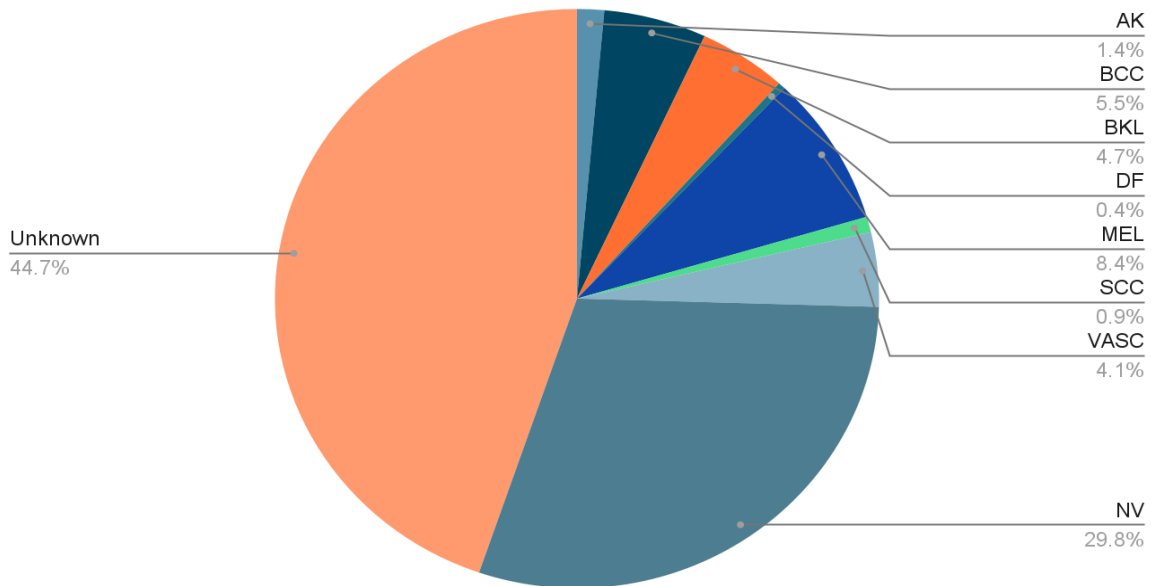
Τα δεδομένα που χρησιμοποιήθηκαν για τα πειράματα που πραγματοποιήθηκαν και θα παρουσιαστούν στην επόμενη ενότητα, προέρχονται από το [Society for Imaging Informatics in Medicine \(SIIM\)](#) σε συνεργασία με το [International Skin Imaging Collaboration \(ISIC\)](#) και μπορούν να βρεθούν στον ακόλουθο σύνδεσμο<sup>1</sup>. Πιο συγκεκριμένα, χρησιμοποιήθηκαν τα διαθέσιμα δεδομένα των ετών 2019 και 2020. Τα δεδομένα αποτελούνται από συνολικά 58.457 εικόνες, εκ των οποίων οι 25.331 ανήκουν στο dataset του 2019, ενώ οι υπόλοιπες 33.126 ανήκουν στο dataset του 2020. Ακόμη, οι εικόνες συνοδεύονται και από μεταδεδομένα, τα οποία περιέχουν πληροφορίες σχετικά με το όνομα της εικόνας που αντιπροσωπεύουν, το id του ασθενή, το φύλο, την ηλικία, το σημείο του σώματος που προέρχεται η φωτογραφία, την διάγνωση (9 κατηγορίες), εάν είναι καλοήθης ή κακοήθης καρκίνος, καθώς και πληροφορίες για την εικόνα όπως ύψος ή πλάτος. Στην συνέχεια, παρουσιάζεται η κατανομή των δεδομένων ανά έτος και ανά κατηγορίες.



Εικόνα 3.1: Κατανομή δεδομένων 2019 και 2020

<sup>1</sup> <https://challenge.isic-archive.com/data>

## Κατανομή Κατηγοριών



Εικόνα 3.2: Κατανομή συνολικών δεδομένων σε κατηγορίες-κλάσεις

Στην εικόνα 3.1, όπου παρουσιάζεται η κατανομή των δεδομένων σε καλοήγη και κακοήγη καρκίνο, παρατηρείται ότι υπάρχει μεγάλη ανισορροπία στα δεδομένα, καθώς η μεγάλη πλειονότητα αυτών ανήκει σε μία κλάση. Αυτή η ανισορροπία είναι ακόμη πιο εμφανής στα δεδομένα του 2020, όπου το ποσοστό των εικόνων που αντιστοιχούν σε κακοήγη καρκίνο ανέρχεται σε μόλις 1,7% επί του συνόλου των δεδομένων εκείνης της χρονιάς.

Ακόμη, στην εικόνα 3.2, όπου παρουσιάζεται η κατανομή των δεδομένων σε διαφορετικές κατηγορίες, παρατηρείται ότι οι 2 επικρατέστερες κατηγορίες είναι τα δεδομένα χωρίς διάγνωση που όμως είναι γνωστό ότι είναι καλοήγη καρκίνος και τα δεδομένα με ετικέτα *Nevus*. Η τρίτη μεγαλύτερη κατηγορία είναι η ετικέτα *Melanoma*, που είναι και η μοναδική που αντιπροσωπεύει τον κακοήγη καρκίνο.

Τέλος, σύμφωνα με το aggregation που έγινε στα μεταδεδομένα, το πλήθος των ανδρών ήταν 30.366 και των γυναικών 27.642, το εύρος ηλικίας των ασθενών ήταν από 5 έως τα 90 έτη και η μέση τιμή τους ήταν τα 51 έτη, ενώ οι συχνότητες των σημείων του σώματος από όπου προέρχονται οι εικόνες παρουσιάζονται στον πίνακα 3.1.



Σημείο Σώματος	Συχνότητα Εμφάνισης
anterior torso	6915
head/neck	6442
lateral torso	54
lower extremity	13407
oral/genital	183
palms/soles	773
posterior torso	2787
torso	16845
upper extremity	7893

Πίνακας 3.1: Συχνότητα σημείων προέλευσης εικόνων από σημεία του σώματος

### 3.2. Επαύξηση δεδομένων

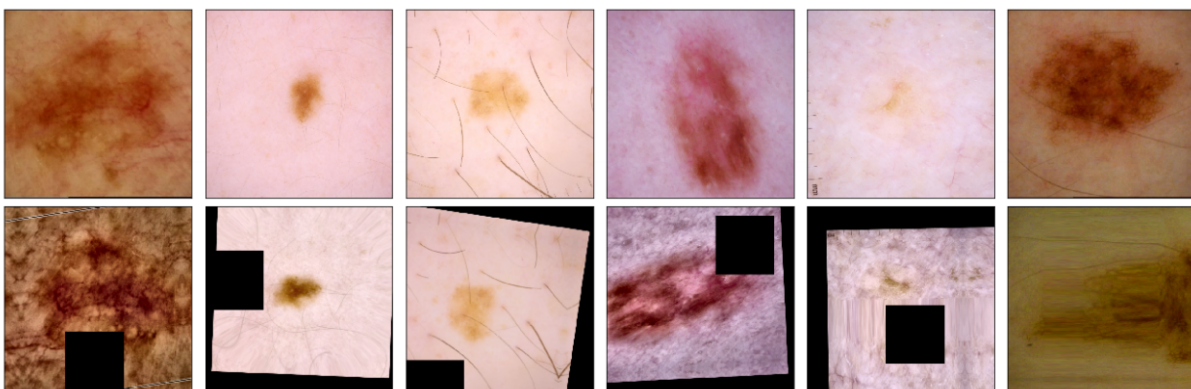
Η χρήση βαθιών νευρωνικών δικτύων με πολύ μεγάλο αριθμό εκπαιδευσιμων παραμέτρων καθιστά μείζον θέμα την θωράκιση του μοντέλου απέναντι στο φαινόμενο της υπερεκπαίδευσης. Το συγκεκριμένο πρόβλημα προκύπτει σε περιπτώσεις όπου το δίκτυο διαθέτει πολύ μεγάλη περιγραφική ικανότητα ενώ δεν υπάρχουν αρκετά δεδομένα για να την αξιοποιήσουν. Έτσι παρατηρείται το φαινόμενο πολλές από τις εκπαιδευσιμες παραμέτρους του δικτύου να χρησιμοποιούνται για την μοντελοποίηση του θορύβου των δεδομένων και όχι της πραγματική κατανομής. Το μοντέλο έτσι χάνει την ικανότητα γενίκευσής του και παρουσιάζει πολύ χειρότερα αποτελέσματα σε unseen data, παρά την εξαιρετική του επίδοση στο σύνολο εκπαίδευσης.

Το παραπάνω φαινόμενο οδηγεί στην ανάγκη αναζήτησης περισσότερων δεδομένων. Αυτό δεν είναι πάντα εύκολο γιατί τα δεδομένα και ιδίως τα επισημειωμένα δεδομένα είναι δυσεύρετα. Για το λόγο αυτό συχνά οδηγούμαστε στη λύση της επαύξησης των δεδομένων εκπαίδευσης, δηλαδή στη δημιουργία νέων δεδομένων από τα ήδη υπάρχοντα, που ακολουθούν την ίδια κατανομή και παρουσιάζουν τις ίδιες ιδιότητες. Στα dataset που αποτελούνται από εικόνες η διαδικασία αυτή είναι

αρκετά πιο εύκολη σε σχέση με άλλες εφαρμογές, και αυτό έγκειται στο γεγονός ότι μπορούμε να πραγματοποιήσουμε διάφορους μετασχηματισμούς σε μία εικόνα χωρίς αυτή να χάνει το νόημα της. Στον πίνακα 3.2 παρουσιάζονται οι μετασχηματισμοί προτείνονται από τη σχετική βιβλιογραφία [6]. Οι μετασχηματισμοί αυτοί χρησιμοποιούνται ευρέως για πληθώρα εφαρμογών και επιλέχθηκαν για την παρούσα εργασία.

Augmentation	Probability / Details
Horizontal Flipping	50% / -
Vertical Flipping	50% / -
Color Jitter	<p>Each pixel's modification is chosen uniformly from</p> <p><math>[ \max (0, 1 - \text{brightness}), 1 + \text{brightness} ]</math>  <math>[ \max (0, 1 - \text{contrast}), 1 + \text{contrast} ]</math>  <math>[ \max (0, 1 - \text{saturation}), 1 + \text{saturation} ]</math>  <math>[ -\text{hue}, \text{hue} ]</math></p> <p>Where brightness, contrast, saturation and hue are the original values of each pixel</p>
Random Erasing	50% / scaled uniformly from: $[0.02, 0.10]$

Πίνακας 3.2: Μετασχηματισμοί επαύξησης δεδομένων



Εικόνα 3.3: Ενδεικτικά αποτελέσματα επαύξησης δεδομένων

## 4. Μεθοδολογία

### 4.1. Δομή Android Εφαρμογής

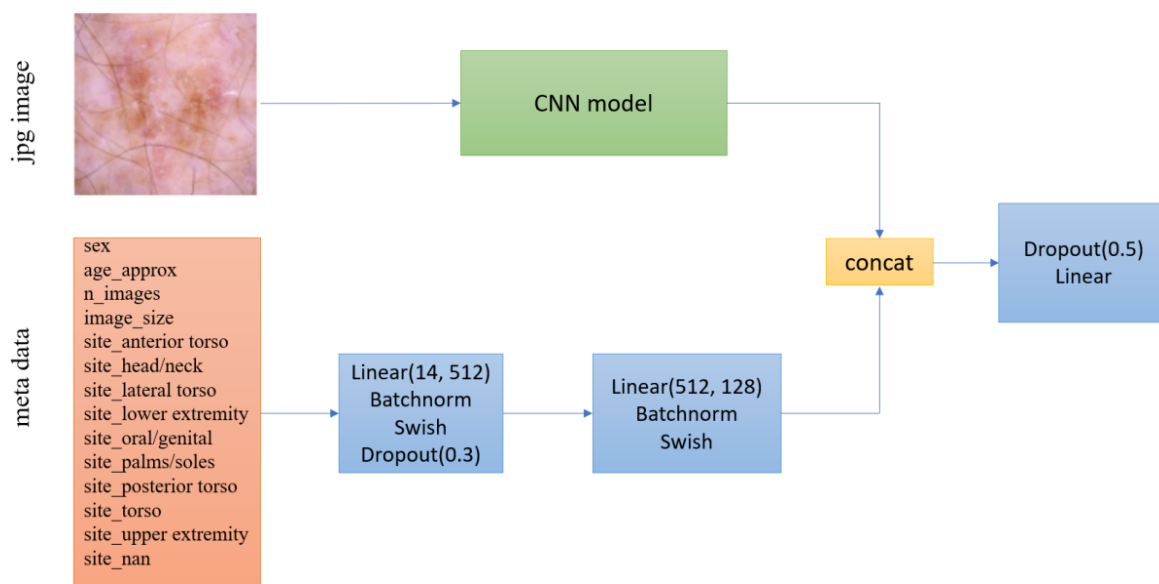
Σκοπός της εργασίας αυτής είναι η ανάπτυξη μίας Android εφαρμογής η οποία θα δίνει τη δυνατότητα στο χρήστη να τραβάει μία φωτογραφία με την κάμερα του κινητού του τηλεφώνου και στη συνέχεια να την ανεβάζει στην εφαρμογή, η οποία με τη σειρά της θα του επιστρέφει την πιθανότητα να υπάρχει κάποιο μελάνωμα στην φωτογραφία αυτή. Το inference από ένα βαθύ νευρωνικό δίκτυο είναι μία ακριβή υπολογιστικά διαδικασία. Για το λόγο αυτό, είναι αρκετά σύνηθες να χρησιμοποιούνται μοντέλα που απαιτούν λιγότερους υπολογιστικούς πόρους. Η πιο διαδεδομένη μέθοδος είναι η χρήση depthwise separable convolution. Χαρακτηριστικό παράδειγμα τέτοιου δικτύου είναι το ευρέως διαδεδομένο MobileNet, το οποίο δημιουργήθηκε για να χρησιμοποιηθεί σε εφαρμογές βαθιάς μάθησης σε φορητές συσκευές (smartphones ως επί το πλείστον) που μοιραία έχουν περιορισμένη υπολογιστική ισχύ.

Στην παρούσα εργασία ακολουθήθηκε μία διαφορετική στρατηγική. Σκοπός μας ήταν να μην κάνουμε συμβιβασμούς στην επίδοση των μοντέλων που θα χρησιμοποιήσουμε. Για το λόγο αυτό δημιουργήσαμε μία online εφαρμογή, η οποία θα επικοινωνεί με έναν απομακρυσμένο server, ο οποίος έχει επιφορτιστεί με το καθήκον της υλοποίησης του inference από το μοντέλο, καθώς έχει την απαραίτητη υπολογιστική ισχύ. Αυτό θέτει έναν περιορισμό στο χρήστη, καθώς θα πρέπει να έχει σύνδεση στο διαδίκτυο για να μπορεί να τρέξει την εφαρμογή, αλλά την καθιστά πολύ πιο ελαφριά και μας δίνεται η δυνατότητα να αναπτύξουμε πολύ βαθιά και περίπλοκα νευρωνικά χωρίς να αντιμετωπίσουμε bottleneck από την πλευρά του hardware του smartphone. Επίσης, η σύνδεση με έναν κεντρικό server μας δίνει επιπλέον δυνατότητες οι οποίες θα αναλυθούν στην ενότητα που αφορά τις μελλοντικές επεκτάσεις.

Η εφαρμογή που παρουσιάζουμε στη συγκεκριμένη εργασία είναι μία απλή εφαρμογή για smartphones με λειτουργικό Android. Η λειτουργία της κρατήθηκε όσο πιο απλή γίνεται. Ο χρήστης επιλέγει μία φωτογραφία. Μπορεί να επιλέξει μεταξύ της λήψης μία νέας φωτογραφίας ή της αποστολής μίας ήδη υπάρχουσας φωτογραφίας από το κινητό του. Η εφαρμογή στέλνει στον server την εικόνα και ξεκινάει το inference. Από τη διαδικασία αυτή θα προκύψει ένας float αριθμός ο οποίος αντιστοιχεί στην πιθανότητα ταξινόμησης ως μελανώματος της εικόνας. Ακόμη, δίνεται η δυνατότητα στον χρήστη να δει την εικόνα που προέκυψε από το lime explainability, ώστε να διακρίνει ποιες περιοχές έπαιξαν μεγαλύτερη βαρύτητα στην έκβαση του αποτελέσματος του prediction.

## 4.2. Ανάπτυξη μοντέλου

Όπως προαναφέρθηκε και στην ενότητα 2, η προσέγγισή μας βασίζεται στο State-of-the-Art μοντέλο των Qishen Ha et al. που συμμετείχε στον διαγωνισμό του International Skin Imaging Collaboration “SIIM - ISIC Melanoma Classification Challenge 2020” και κατέκτησε την πρώτη θέση. Το μοντέλο επί της ουσίας αποτελείται από ένα ensemble από 18 μοντέλα μεταφοράς μάθησης από πολύ βαθιά συνελικτικά νευρωνικά δίκτυα (16 EfficientNets, 2 ResNets). Η τελική πρόβλεψη προκύπτει ως σταθμισμένος μέσος όρος των προβλέψεων των επιμέρους μοντέλων, όπου ο παράγοντας βάρους αντιστοιχεί στην συνολική ακρίβεια του κάθε μοντέλου που συγκροτεί το ensemble. Η επιτυχία των Qishen Ha et al. έγκειται σε μεγάλο βαθμό στη χρήση του συνελικτικού νευρωνικού δικτύου EfficientNet της GoogleAI το οποίο αποτελεί το State-of-the-Art μοντέλο στο πεδίο της ταξινόμησης εικόνων.



Εικόνα 4.1: Αρχιτεκτονική backbone μοντέλου του ensemble

Στην ενότητα 3 πραγματοποιήθηκε μία εκτενής περιγραφή των δεδομένων που χρησιμοποιήθηκαν για την ανάπτυξη των μοντέλων. Η τελική ταξινόμηση που θέλουμε να πραγματοποιήσουμε και που είναι και το ζητούμενο του διαγωνισμού είναι η ταξινόμηση των δειγμάτων σε καλοήγη και κακοήγη. Όπως προκύπτει από την εικόνα 3.1, τα δεδομένα εμφανίζουν

ανισορροπία σε ακραίο βαθμό. Οι Qishen Ha et al. προτείνουν την ταξινόμηση με βάση τη διάγνωση. Στο σύνολο δεδομένων που έχουμε στη διάθεση μας η διάγνωση αποτελείται από 9 κλάσεις, όπως φαίνεται στην εικόνα 3.2. Όπως προκύπτει όμως από την κατανομή των δειγμάτων στην εικόνα 3.2 τα δεδομένα είναι εμφανίζουν έντονη ανισορροπία και με το συγκεκριμένο διαχωρισμό. Οι Qishen Ha et al. δεν προβαίνουν σε περαιτέρω υλοποίηση τεχνικών εξισορρόπησης δεδομένων και αφήνουν τα πανίσχυρα EfficientNet να φέρουν σε πέρας το multi-class classification task.

Σε πρώτο χρόνο, αυτό που κάναμε ήταν να αναπαράγουμε τα αποτελέσματα των Qishen Ha et al. για να διαπιστώσουμε αφενός την ορθότητα των ισχυρισμών τους και αφετέρου να συγκρίνουμε την μέθοδο εκπαίδευσης που χρησιμοποίησαν, δηλαδή την εκπαίδευση με βάση τη διάγνωση και όχι με βάση την binary target variable (benign - malignant). Επειδή οι κλάσεις των 2 dataset (2019 και 2020) είναι διαφορετικές, πραγματοποιείται το mapping του πίνακα 4.1.

2019 Diagnosis	2020 Diagnosis	Target
NV	nevus	NV
MEL	melanoma	MEL
BCC		BCC*
BKL	seborrheic keratosis lichenoid keratosis solar lentigo lentigo NOS	BKL
AK		AK*
SCC		SCC*
VASC		VASC*
DF		DF*
	cafe-au-lait macule atypical melanocytic proliferation	Unknown

Πίνακας 4.1 : Class Mapping [1]

Όπως προαναφέραμε, στην [1] δεν έχει γίνει κάποια εξισορρόπηση δεδομένων εκπαίδευσης. Στην παρούσα εργασία επικεντρωνόμαστε σε δύο μεθόδους εξισορρόπησης δεδομένων. Η πρώτη είναι η προσέγγιση balanced subsampling. Στην προσέγγιση αυτή η μεταβλητή στόχος είναι μία binary μεταβλητή που υποδηλώνει την πιθανότητα το skin lesion που εμφανίζεται στην εικόνα να είναι benign ή malignant. Για την υλοποίηση της παραπάνω τεχνικής αναθέσαμε βάρη στα δείγματα που αντιστοιχούν στην πιθανότητα εμφάνισης του δείγματος στα δεδομένα της κάθε εποχής. Αντιμετωπίσαμε τα βάρη αυτά σαν υπερπαραμέτρο του μοντέλου ώστε να πετύχουμε διαφορετικά ποσοστά και να βρούμε το ποσοστό που πετυχαίνει τα καλύτερα αποτελέσματα. Για ίσα βάρη για όλα τα δείγματα η παραπάνω τεχνική εκφυλίζεται σε απλό binary classification με την αρχική target

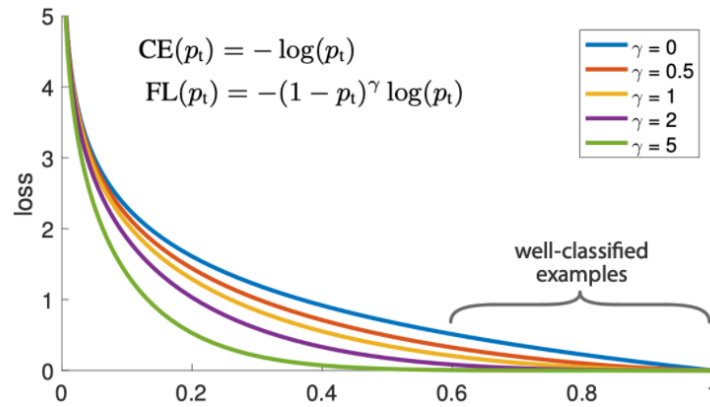
variable. Περισσότερες λεπτομέρειες για της συγκεκριμένη τεχνική και τα αποτελέσματά που προέκυψαν από την εφαρμογή της υπάρχουν στην ενότητα 5.2..

Στη συνέχεια προβήκαμε στην εκπαίδευση του μοντέλου με τη χρήση του focal loss σαν συνάρτηση κόστους στη θέση της categorical cross-entropy που εφαρμόζεται στην περίπτωση του multi-class classification από τους [1]. Παρακάτω παραθέτουμε πως υπολογίζεται το focal loss σε σχέση με τη συνάρτηση κόστους cross-entropy καθώς και την λειτουργία της υπερπαραμέτρου  $\gamma$  στο υπολογισμό του κόστους.

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (1)$$

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (2)$$

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (3)$$



Εικόνα 4.2: Focal loss για διαφορετικές τιμές της παραμέτρου  $\gamma$

Από τους παραπάνω μαθηματικούς τύπους γίνεται πιο ξεκάθαρη η λειτουργία του focal loss. Συγκεκριμένα η διαφοροποίηση του focal loss (3) από το cross-entropy loss (2) έγγειται στον πολλαπλασιασμό του cross-entropy loss με τον όρο  $(1 - p_t)^\gamma$ . Η  $p_t$  δηλώνει την πιθανότητα ένα

δείγμα να ανήκει σε μία κλάση. Αρχικά η  $p_i$  της κλάσης με το μεγάλο support στο dataset είναι μεγάλη καθώς το μοντέλο μαθαίνει να προβλέπει αυτή σαν σωστή για να μειώσει τη συνάρτηση μέσω του όρου  $\log(p_i)$ . Όμως ταυτόχρονα αυξάνεται ο όρος  $(1 - p_i)$ . Θέτοντας την παράμετρο  $\gamma > 0$  μειώνει το relative loss για τα ευκόλως ταξινομήσιμα δείγματα της πλειοψηφούσας κλάσης δίνοντας μεγαλύτερο βάρος στα δυσκόλως διαχωρίσιμα δείγματα τα οποία κάνει misclassify. Η συγκεκριμένη μεθοδολογία εφαρμόζεται ευρέως σε προβλήματα του πεδίου της όραση υπολογιστών και ιδιαίτερα σε εφαρμογές semantic segmentation για να δοθεί βάρος στα αντικείμενα ενδιαφέροντος που αποτελούνται από λίγα pixel σε σχέση με το background που αποτελεί το μεγαλύτερο μέρος της εικόνας.

Για τη συγκεκριμένη μεθοδολογία δεν έγινε επιπλέον κάποιου είδους subsampling. Έτσι σε κάθε εποχή τροφοδοτούμε στο δίκτυο ολόκληρο το σύνολο των δεδομένων στο δίκτυο. Εν τέλει, και για τον παραπάνω λόγο, εκπαιδεύουμε το δίκτυο για μικρότερο αριθμό εποχών σε σχέση με το subsampling.

Για όλες τις μεθόδους που αναφέρουμε παραπάνω, πραγματοποιούμε train - test split με τον ήδη υπάρχοντα διαχωρισμό των δεδομένων. Στη συνέχεια πραγματοποιούμε 5-fold cross validation για την ρύθμιση των υπερπαραμέτρων. Για κάθε μοντέλο υπολογίζουμε δύο μετρικές, την cv\_all και cv\_2020 όπου αντιστοιχούν στην τιμή του κριτηρίου AUC για το σύνολο όλων των δεδομένων και των δεδομένων του 2020 dataset αντίστοιχα. Εκπαιδεύουμε τα μοντέλα για προκαθορισμένο αριθμό εποχών. Ελέγχουμε μετά από κάθε εποχή το validation loss και δημιουργούμε νέο checkpoint για το μοντέλο αν το validation loss είναι μικρότερο από την μικρότερη τιμή μέχρι τη συγκεκριμένη εποχή.

## 5. Αποτελέσματα - Σχολιασμός

### 5.1. Μετρικές αξιολόγησης

Όπως παρουσιάστηκε στην ενότητα 3.1, υπάρχει έντονη ανισορροπία στα δεδομένα που χρησιμοποιήθηκαν και έτσι, μετρικές όπως το accuracy θα μπορούσαν να μας οδηγήσουν σε λάθος συμπεράσματα. Συνεπώς, η ύπαρξη ανισορροπίας μεταξύ των κλάσεων του συνόλου δεδομένων καθιστά αναγκαία την επιλογή αξιόπιστων μετρικών αξιολόγησης. Μια τέτοια μετρική είναι το Area Under the Curve (AUC).

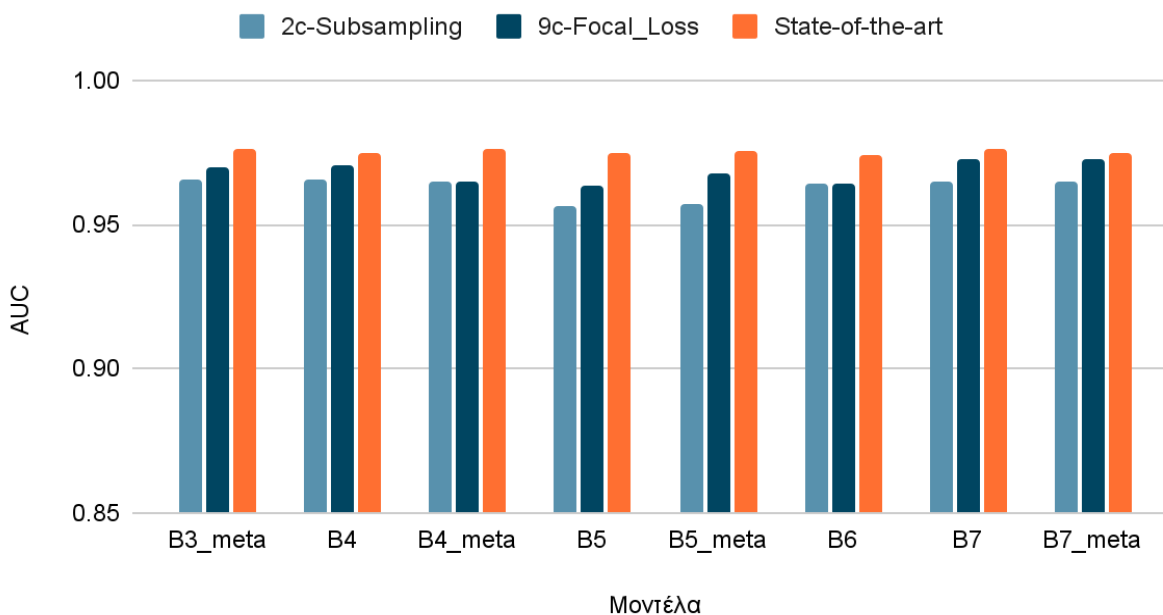
Η καμπύλη Precision-Recall (PR) είναι παρόμοια με την καμπύλη ROC (Receiver Operation Characteristics), αλλά με έναν από τους άξονες να αλλάζει από False Positive Rate σε precision. Συγκεκριμένα, η καμπύλη Precision-Recall μπορεί να χρησιμοποιηθεί ως εναλλακτική μετρική για την αξιολόγηση του ταξινομητή όταν τα δεδομένα είναι μη ισορροπημένα. Στο ROC, η καμπύλη αποτελείται από τον FPR (άξονας x) και τον True Positive Rate / recall (άξονας y). Η περιοχή κάτω από την καμπύλη ROC (AUC) είναι μια ευρέως χρησιμοποιούμενη μετρική για την αξιολόγηση της συνολικής απόδοσης του μοντέλου. Η τιμή του AUC κυμαίνεται από 0 έως 1, όσο μεγαλύτερο είναι το καλύτερο. Η καμπύλη ROC θα κινείται προς την πάνω δεξιά γωνία όταν βελτιώνεται η απόδοση του μοντέλου. Από την άλλη πλευρά, η καμπύλη PR αποτελείται από το recall/True Positive Rate (άξονας x) και την precision (άξονας y). Η περιοχή κάτω από την καμπύλη PR ROC (PR AUC) παρέχει μια διαφορετική προοπτική για την αξιολόγηση του αποτελέσματος του δυαδικού ταξινομητή. Όσο αυξάνει η τιμή του PR AUC τόσο καλύτερη είναι η απόδοση του μοντέλου - η καμπύλη PR θα κινηθεί προς την επάνω αριστερή γωνία. Ωστόσο, δεν είναι εφικτό για την PR AUC να λάβει οποιαδήποτε τιμή μέσα στο εύρος [0, 1]. Ανάλογα με τα δεδομένα, η βασική γραμμή της καμπύλης PR είναι η οριζόντια γραμμή, όπου y ισούται με την τιμή του  $P/(P+N)$ , δηλαδή την μικρότερη τιμή του precision. Όταν το κατώφλι φτάνει στο 0 ή αλλιώς το πιο δεξιό σημείο στο γράφημα, όλα τα δείγματα ταξινομούνται ως θετικά δείγματα. Δηλαδή, ο αριθμός των πραγματικών θετικών είναι ίσος με τον αριθμό των θετικών δειγμάτων. Έτσι, η τιμή της γραμμής αναφοράς μειώνεται όταν τα δεδομένα γίνονται ακόμη περισσότερο μη ισορροπημένα.



## 5.2. Αποτελέσματα

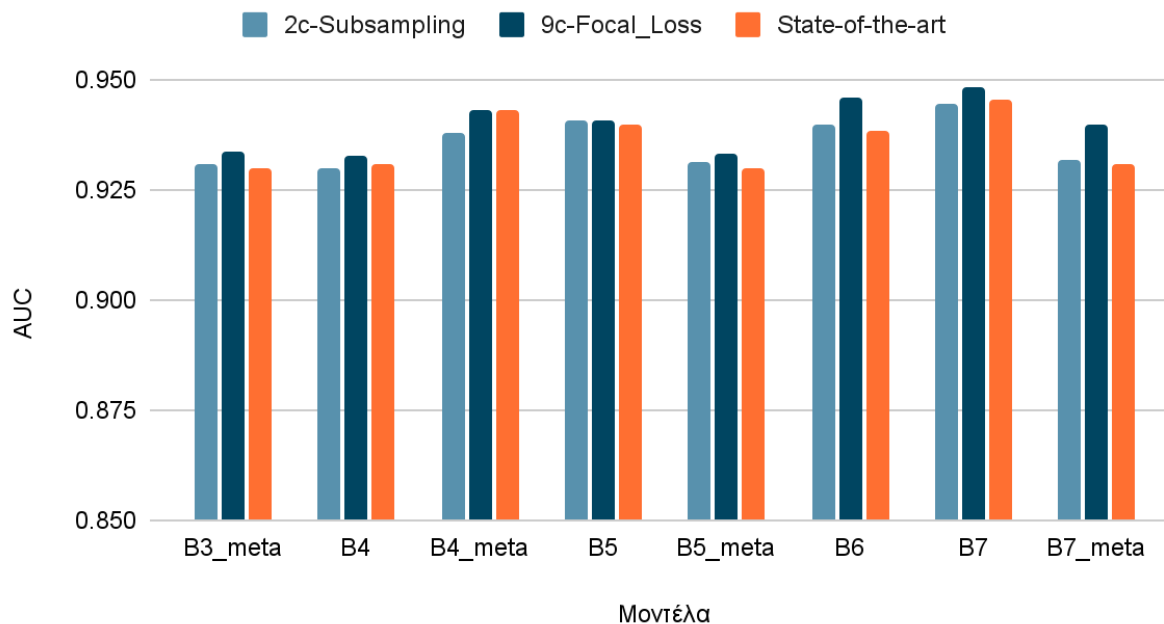
Σε αυτήν την ενότητα, παρουσιάζονται αναλυτικά τα αποτελέσματα που προέκυψαν από τα μοντέλα που εκπαιδεύτηκαν και γίνεται σύγκριση με τα αντίστοιχα αποτελέσματα των Qishen Ha et al.. Η σύγκριση γίνεται μεταξύ των τροποποιημένων μοντέλων για δυαδική ταξινόμηση, όπως παρουσιάστηκε στην ενότητα 4.2, και των μοντέλων που αναπτύχθηκαν από τους Qishen Ha et al. για ταξινόμηση σε 9 κατηγορίες-κλάσεις. Σε όλα τα μοντέλα που χρησιμοποιήθηκαν για δυαδική ταξινόμηση χρησιμοποιήθηκε το subsampling ως μία μέθοδος εξισορρόπησης των δεδομένων και την βελτίωση των αποτελεσμάτων. Ακόμη, λόγω περιορισμού μνήμης της κάρτας γραφικών που χρησιμοποιήθηκε, το batch\_size ορίστηκε σε 8, ενώ χρησιμοποιήθηκαν οι εικόνες μεγέθους 512 με resize σε 384. Όλα τα πειράματα έγιναν με χρήση μίας κάρτας γραφικών Nvidia RTX 3070. Για τις συγκρίσεις που παρουσιάζονται παρακάτω, σε όσα μοντέλα των Qishen Ha et al. δεν ήταν διαθέσιμη η επιλογή των resized εικόνων σε 384, έγινε χρήση του πιο κοντινού μεγέθους resized εικόνων σε αυτό.

### Σύγκριση αποτελεσμάτων κατά AUC



Εικόνα 5.1: Σύγκριση αποτελεσμάτων κατά AUC

## Σύγκριση αποτελεσμάτων κατά AUC\_2020



Εικόνα 5.2: Σύγκριση αποτελεσμάτων κατά AUC 2020

Από την εικόνα 5.1, παρατηρείται ότι τα μοντέλα που αναπτύχθηκαν από τους Qishen Ha et al. και ταξινομούν σε 9 κατηγορίες, επιτυγχάνουν μεγαλύτερη επίδοση κατά την μετρική AUC. Αποδίδουμε αυτήν την διαφορά στα αποτελέσματα, στον περιορισμό που υπήρξε στο μέγεθος των εικόνων λόγω περιορισμένης μνήμης στην κάρτα γραφικών, καθώς με μεγαλύτερο μέγεθος, θα υπήρχε και περισσότερη πληροφορία διαθέσιμη ως προς τα νευρωνικά δίκτυα. Ακόμη, εξαιτίας και πάλι του περιορισμένου χώρου στην κάρτα γραφικών, χρησιμοποιήθηκε ένα πολύ μικρότερο batch size μεγέθους 8, σε σύγκριση με το batch size που χρησιμοποιείται από το State-of-the-Art σύστημα μεγέθους 64. Αυτή η διαφορά μπορεί να είναι ικανή ώστε να επηρεάσει τα αποτελέσματα αρνητικά, αφού το batch size επηρεάζει άμεσα τα αποτελέσματα του loss function και συνεπώς, την ανανέωση των βαρών του νευρωνικού δικτύου.

Τέλος, στην εικόνα 6.2 παρουσιάζονται τα συγκριτικά αποτελέσματα κατά την μετρική AUC μόνο στο dataset του 2020, μεταξύ των μοντέλων δυαδικής ταξινόμησης που κατασκευάστηκαν με τα μοντέλα ταξινόμησης σε 9 κατηγορίες των Qishen Ha et al.. Παρατηρείται ότι όλα τα τροποποιημένα μοντέλα που κατασκευάσαμε επιτυγχάνουν μεγαλύτερη επίδοση σε σύγκριση με το State-of-the-Art. Αποδίδουμε αυτήν την διαφορά στο subsampling approach, αφού ειδικά στο dataset του 2020 τα δεδομένα ήταν εξαιρετικά ανισόρροπα.

Συνοψίζοντας, τα αποτελέσματα των τροποποιημένων μοντέλων θα μπορούσαν να βελτιωθούν περαιτέρω με την αύξηση του batch size ή του μεγέθους των εικόνων, αφού μας ήταν διαθέσιμη μόνο μία κάρτα γραφικών συγκριτικά με τις 8 κάρτες γραφικών που είχαν στην διάθεση τους οι Qishen Ha et al..

### 5.3. Επεξήγηση αποτελεσμάτων - LIME

Σε αυτή την ενότητα, δίνουμε ιδιαίτερη βαρύτητα στην ικανότητα επεξήγησης των αποτελεσμάτων του μοντέλου μας, η οποία επιτυγχάνεται με τη χρήση της τεχνικής LIME. Οι τεχνικές επεξηγηματικότητας δεν είναι ανθεκτικές σε λάθη, αλλά εξαρτώνται άμεσα από την άριστη κατανόηση της εκάστοτε μεθόδου, έτσι ώστε να μην οδηγούμαστε σε λάθος εκτιμήσεις.

Τα αρχικά του LIME προέρχονται από τις λέξεις Local, Interpretable, Model-Agnostic και Explanations. Αναλύοντας ξεχωριστά τον κάθε όρο, θα χαρακτηρίζαμε το Local ως την τοπική προσέγγιση ή την γειτονιά των προβλέψεων προς εξήγηση. Το Interpretable αφορά την φύση των εξηγήσεων, οι οποίες θα πρέπει να γίνονται εύκολα αντιληπτές και κατανοητές από έναν μη σχετικό με το αντικείμενο άνθρωπο. Όσον αφορά τον όρο Model-Agnostic, θα λέγαμε πρακτικά ότι το LIME εφαρμόζεται καλά σε κάθε SVM μοντέλο, νευρωνικά δίκτυα κλπ., όπως και στην περίπτωση μας που έχουμε image classification με τη χρήση μοντέλων βαθιάς μάθησης. Τέλος με τον όρο Explanations, υποδηλώνεται η παροχή επεξηγήσεων των προβλέψεων του μοντέλου και ουσιαστικά η τοπική γραμμική εξήγηση της συμπεριφοράς του μοντέλου.

Συνεπώς, το LIME ανήκει στην κατηγορία Model-Agnostic, το οποίο σημαίνει ότι πρακτικά λειτουργεί καλά σε κάθε Machine Learning μοντέλο. Η τεχνική προσπαθεί να καταλάβει το αρχικό μας μοντέλο, μεταβάλλοντας τα δείγματα εισόδου και πως ανάλογα επηρεάζονται οι προβλέψεις. Θα χαρακτηρίζαμε το LIME ως ένα black box machine learning model, το οποίο αναζητά τη σχέση μεταξύ εισόδου και εξόδου.

Οι model-specific προσεγγίσεις έχουν στόχο να ερμηνεύσουν αυτό το black box, αναλύοντας τα εσωτερικά στοιχεία του μοντέλου και τους τρόπους που αυτά αλληλεπιδρούν. Για παράδειγμα, σε μοντέλα deep learning είναι εφικτό κανείς να εξερευνήσει activation units και να συνδέσει internal activations πίσω στο input.

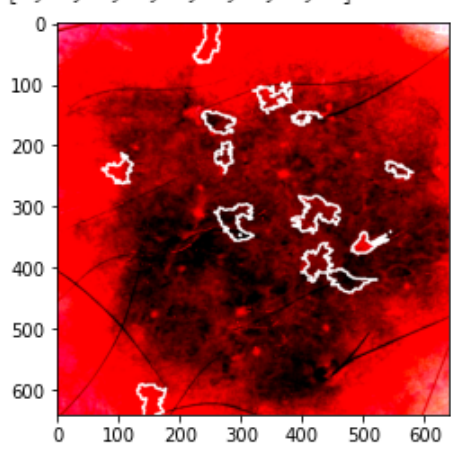
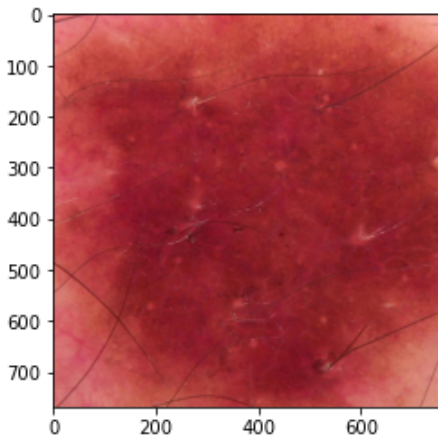
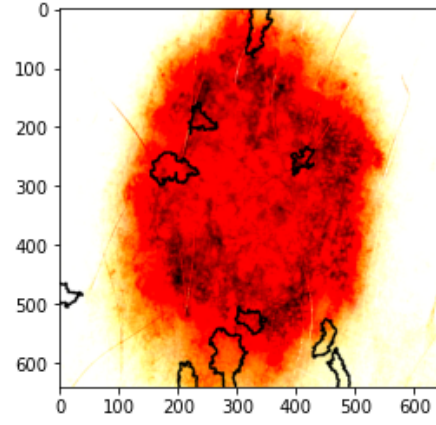
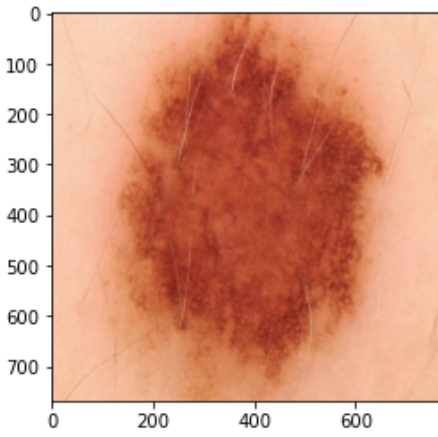
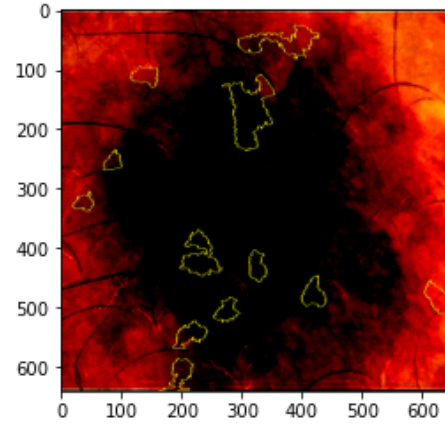
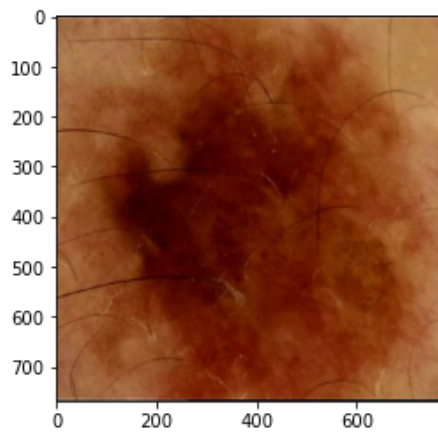
Σε αντίθεση με άλλες τεχνικές ερμηνείας των αποτελεσμάτων ενός μοντέλου που εξηγούν την απόφαση σε επίπεδο ολοκληρωμένου dataset, το LIME επιτρέπει στον χρήστη να κατανοήσει και ποιά χαρακτηριστικά θεωρούνται περισσότερο σημαντικά. Επιτρέπει την επαλήθευση των

υποθέσεων και εάν το μοντέλο κάνει υπερπροσαρμογή στο θόρυβο, αλλά καθιστά δύσκολη τη διάγνωση συγκεκριμένων προβλέψεων του μοντέλου.

Μια εξήγηση δημιουργείται προσεγγίζοντας τοπικά το underlying μοντέλο με ένα ερμηνεύσιμο μοντέλο. Τέτοια ερμηνεύσιμα μοντέλα μπορεί να είναι γραμμικά μοντέλα με regularization, δέντρα αποφάσεων κλπ. Τα ερμηνεύσιμα μοντέλα εκπαιδεύονται σε μικρές μεταβολές του αρχικού δείγματος εισόδου και πρέπει να παρέχουν μόνο μια καλή τοπική προσέγγιση. Το σύνολο δεδομένων μπορεί να δημιουργηθεί είτε μέσω της προσθήκης θορύβου σε συνεχή χαρακτηριστικά, είτε αφαιρώντας λέξεις ή όπως στην περίπτωση μας, κρύβοντας κομμάτια μιας εικόνας. Προσεγγίζοντας το black box μόνο τοπικά, η εργασία απλοποιείται σημαντικά, κερδίζοντας χρόνο και υπολογιστικούς πόρους.

Αν και η γενική ιδέα του LIME ακούγεται εύκολη στην εφαρμογή της, υπάρχουν και μερικά μειονεκτήματα. Στο LIME χρησιμοποιούνται μόνο γραμμικά μοντέλα για την προσέγγιση τοπικής συμπεριφοράς. Σε κάποιο βαθμό, αυτή η υπόθεση είναι σωστή όταν εξετάζουμε μια πολύ μικρή περιοχή γύρω από τα δείγματα εισόδου. Ωστόσο, με την επέκταση αυτής της περιοχής, είναι πιθανό ένα γραμμικό μοντέλο να μην είναι αρκετά ισχυρό για να εξηγήσει τη συμπεριφορά του αρχικού μας μοντέλου. Η μη γραμμικότητα στις τοπικές περιοχές συμβαίνει για εκείνα τα σύνολα δεδομένων που απαιτούν περίπλοκα και μη ερμηνεύσιμα μοντέλα. Συνεπώς, αυτό αποτελεί ένα σημαντικό μειονέκτημα του LIME.

Στις παρακάτω εικόνες αναφέρουμε κάποια ενδεικτικά αποτελέσματα από την εφαρμογή του LIME για 3 εικόνες του συνόλου ελέγχου. Παρατηρούμε ότι οι περιοχές ενδιαφέροντος περιορίζονται ως επί το πλείστον εντός της γενικότερη περιοχής του skin lesion.



## 6. Συμπεράσματα - Μελλοντικές επεκτάσεις

### 6.1. Συμπεράσματα

Όπως προαναφέραμε, στην παρούσα εργασία προσπαθήσαμε να εκμεταλλευτούμε τις ιδέες και προσεγγίσεις που έχουν διατυπωθεί πάνω στο πρόβλημα της ταξινόμησης μελανωμάτων και με τη σειρά μας να συνδυάσουμε θετικά στοιχεία κάθε προσέγγισης και να ερευνήσουμε κατά πόσο οι συνδυασμοί αυτοί μπορούν να οδηγήσουν στην ανάπτυξη ακόμα καλύτερων μοντέλων. Η εξαντλητική εξερεύνηση των συνδυασμών αυτών είναι εξαιρετικά χρονοβόρα και απαιτεί πολλούς υπολογιστικούς πόρους. Στο σημείο αυτό θεωρούμε σκόπιμο να αναφέρουμε ότι τα μοντέλα που αναπτύχθηκαν στην παρούσα εργασία αποτελούν πολύ βαθιά συνελκτικά νευρωνικά δίκτυα τα οποία απαιτούν μεγάλη υπολογιστική ισχύ για να εκπαιδευτούν. Χαρακτηριστικό παράδειγμα είναι το γεγονός ότι στην σχετική βιβλιογραφία γίνεται χρήση συστοιχίας καρτών γραφικών (GPUs) μέχρι και 8 καρτών για την εκπαίδευση κάποιων από τα μοντέλα. Ως εκ τούτου, η ανεπάρκεια υπολογιστικών πόρων δεν μας επέτρεψε να εξαντλήσουμε τις πιθανότητες βελτιστοποίησης των μοντέλων είτε γιατί μας περιόριζε χρονικά είτε ήταν απαγορευτική από άποψη μνήμης.

Έγινε προσπάθεια να εφαρμοστούν καινοτόμες προσεγγίσεις οι οποίες θα συνάδουν με τις ιδιαιτερότητες του εγχειρήματος μας. Συγκεκριμένα, η επιλογή μικρότερων από άποψη ανάλυσης εικόνων βοήθησε ως προς τις απαιτήσεις σε μνήμα αλλά κρίνεται σκόπιμη καθώς οι κάμερες των κινητών τηλεφώνων δεν είναι σε θέση να τραβούν φωτογραφίες πολύ υψηλής ανάλυσης, όπως αρκετές από το σύνολο δεδομένων εκπαίδευσης. Αυτό, κρίνοντας και από τα αποτελέσματα επέφερε ένα μικρό κόστος στην επίδοση του μοντέλου αλλά είναι ένα αναγκαίο tradeoff. Επίσης, αντίθετα με την επικρατούσα τάση να χρησιμοποιούνται ελαφριά μοντέλα όπως το MobileNet σε εφαρμογές κινητών τηλεφώνων.

Η μέθοδος ensemble από μοντέλα μηχανικής μάθησης είναι η επικρατούσα τάση και τα State-of-the-Art μοντέλα αποτελούν αποτελέσματα της μεθόδου αυτής. Εμείς από τη μεριά μας προσπαθήσαμε να εφαρμόσουμε τεχνικές εξισορρόπησης δεδομένων σε ένα τέτοιο μοντέλο αντί να περιμένουμε από αυτό να ανταποκριθεί χάρη στην ικανότητα του να πετυχαίνει εξαιρετικά αποτελέσματα. Δοκιμάσαμε να κάνουμε subsampling των δεδομένων με διαφορετικά ποσοστά εκπροσώπησης των κλάσεων καθώς και τη χρήση διαφορετικής loss function, του focal loss που αυξάνει τη σημασία της ταξινόμησης των δύσκολων δειγμάτων της κλάσης με μικρό support. Παρατηρήσαμε ότι ανάλογα με το δίκτυο και τις υπερπαραμέτρους που επιλέχθηκαν (όπως πχ το

μέγεθος εικόνας εισόδου και η χρήση ή όχι μεταδεδομένων) οδηγούμαστε σε ανάλογα και μερικές φορές καλύτερα αποτελέσματα, ειδικότερα σε σύνολα ελέγχου που εμφανίζουν μεγαλύτερη ανισορροπία.

## 6.2. Μελλοντικές επεκτάσεις

Η παρούσα εργασία είναι μία πρώτη απόπειρα δημιουργίας μία εφαρμογής ταξινόμησης μελανωμάτων για Android smartphones. Όπως κάθε πρώτη απόπειρα έτσι και η εφαρμογή που παρουσιάζουμε στην εργασία αυτή επιδέχεται περαιτέρω τροποποιήσεις αλλά και επεκτάσεις που θα την καταστήσουν ακόμα καλύτερη και πληρέστερη. Στην ενότητα αυτή θα αναφέρουμε ιδέες για μελλοντικές επεκτάσεις της παρούσας εργασίας. Οι επεκτάσεις αυτές αφορούν και το σκέλος της ανάπτυξης της εφαρμογής και του μοντέλου που χρησιμοποιεί η εφαρμογή για την εξαγωγή προβλέψεων.

Σε ότι αφορά την εφαρμογή, κάτι που θα είχε μεγάλη σημασία αλλά δεν προλάβουμε να υλοποιήσουμε και θεωρούμε ότι πρέπει να προστεθεί στην ήδη υπάρχουσα εφαρμογή είναι η δυνατότητα upload ιατρικής διάγνωσης για μία υπάρχουσα φωτογραφία. Πιο συγκεκριμένα, ο χρήστης θα έχει τη δυνατότητα για μία φωτογραφία που τράβηξε να μπορεί να ανεβάσει τη διάγνωση ενός δερματολόγου στον οποίο απευθύνθηκε. Αυτό είναι εξαιρετικά χρήσιμο, καθώς αποτελεί μία μορφή crowdsourcing και θα δώσει τη δυνατότητα να συλλεχθούν πάρα πολλές εικόνες μαζί με τις επισημειώσεις τους. Η ιατρική γνωμάτευση είναι απαραίτητη για να διασφαλίζεται η αδιαβλητότητα της διαδικασίας και η ορθότητα των επισημειώσεων. Αυτός είναι και ένας από τους λόγους που θέλαμε η εφαρμογή να είναι διαδικτυακή.

Σε ότι έχει να κάνει με το μοντέλο που αναπτύχθηκε, εκεί οι βελτιώσεις που μπορούν να γίνουν δεν μας είναι γνωστές όμως εκτός από τα πειράματα που διεξάγαμε στα πλαίσια της συγκεκριμένης εργασίας υπάρχουν κάποιες ακόμα προσεγγίσεις οι οποίες χρήζουν περαιτέρω διερεύνησης. Συγκεκριμένα, η παρούσα εργασία επικεντρώθηκε στην ανάπτυξη ενός μοντέλου ensemble και στη διερεύνηση τεχνικών αντιμετώπισης του φαινομένου της ανισορροπίας δεδομένων. Στη σχετική βιβλιογραφία που παρουσιάσαμε στην ενότητα 3 γίνεται αναφορά σε κάποια πιο sophisticated μοντέλα, όπως το Nambla-N και το Inception ResNet v2 SA. Τα μοντέλα αυτά πετυχαίνουν αρκετά καλά αποτελέσματα, αλλά όχι καλύτερα από το EfficientNet. Πιστεύουμε ότι θα έχει ιδιαίτερο ενδιαφέρον η χρήση του μηχανισμού Soft Attention σε συνδυασμό με ένα μοντέλο μεταφοράς μάθησης από το EfficientNet ή τη χρήση ενός semantic segmentation μοντέλου όπως το Nambla-N, καθώς στη σχετική βιβλιογραφία για την παραπάνω εφαρμογή έχουν χρησιμοποιηθεί

μόνο μοντέλα ResNet και παραλλαγές αυτών. Επίσης θα ήταν ενδιαφέρον να χρησιμοποιήσουμε τέτοια μοντέλα για τη δημιουργία ενός ensemble ώστε να το συγκρίνουμε με τα αποτελέσματα που δίνει το EfficientNet ensemble.

### 6.3. Τεχνικές λεπτομέρειες

Τα μοντέλα που παρουσιάστηκαν στην παρούσα εργασία αναπτύχθηκαν με τη χρήση του framework PyTorch της Python. Η εφαρμογή Android δημιουργήθηκε με χρήση της εφαρμογής Android Studio και η γλώσσα προγραμματισμού που χρησιμοποιήθηκε είναι η Java. Η σύνδεση της εφαρμογής με τον server που hostάει το μοντέλο για το inference έγινε μέσω του web framework Flask της Python. Για το LIME χρησιμοποιήσαμε τη βιβλιοθήκη lime της python που προσφέρει έτοιμη υλοποίηση του LIME.

### 6.4. Συνεισφορά Μελών Ομάδας

Η συνεισφορά των μελών της ομάδας στην εκπόνηση της παρούσας εξαμηνιαίας εργασίας στο μάθημα 'Ανάλυση Βιο-Δεδομένων' αναλύεται ως εξής: οι Συρράφος Δημήτρης και Χατσιακίαν Μελκόν ασχολήθηκαν με την τροποποίηση του κώδικα των νευρωνικών δικτύων, καθώς και τον σχεδιασμό και την εκτέλεση των πειραμάτων σχετικά με το balanced subsampling και το focal loss. Ο Μπριλλάκης Βασίλειος ασχολήθηκε με την σχεδίαση, κατασκευή και εκτέλεση του Android application που παρουσιάστηκε παραπάνω. Ο Κετσέας Σταύρος έλαβε συμμετοχή στην ανάπτυξη του Android application, καθώς και με την επεξήγηση των αποτελεσμάτων του νευρωνικού δικτύου μέσω του lime explainability και την παρουσίαση των μετρικών αξιολόγησης. Τέλος, όλα τα μέλη της ομάδας ασχολήθηκαν με την αναζήτηση και την επισκόπηση της βιβλιογραφίας που παρατέθηκε.

**Disclaimer:** Τα παραπάνω δεν αποτελούν ορθογώνιο διαχωρισμό των εργασιών που εκτελέστηκε από κάθε άτομο, καθώς όλοι μας είχαμε συνεισφορά και παρείχαμε βοήθεια και σε άλλα κομμάτια της εργασίας, εκτός αυτών που αναφέρονται παραπάνω.



## Βιβλιογραφία

- [1] Identifying Melanoma Images using EfficientNet Ensemble: Winning Solution to the SIIM-ISIC Melanoma Classification Challenge, 2020, Qishen Ha (LINE), Corp Bo Liu (NVIDIA), Fuxu Liu (ReadSense Ltd), ArXiv preprint arXiv:2010.05351v1 [cs.CV] 11 Oct 2020
- [2] Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks Md Zahangir Alom, Theus Aspiras, Tarek M. Taha, and Vijayan K. Asari, arXiv:1904.11126, 2018
- [3] Soft-Attention Improves Skin Cancer Classification Performance, Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari, Mingchen Gao, arXiv:2105.03358v3 [eess.IV] 4 Jun 2020
- [4] Transfer learning with class-weighted and focal loss function for automatic skin cancer classification Duyen N.T. Le , Hieu X. Le , Lua T. Ngo , and Hoan T. Ngo, arXiv:2009.05977v1 [cs.AI] 13 Sep 2020
- [5] Accessible Melanoma Detection using Smartphones and Mobile Image Analysis Thanh-Toan Do, Tuan Hoang, Victor Pomponiu, Yiren Zhou, Zhao Chen, Ngai-Man Cheung, Dawn Koh, Aaron Tan and Suat-Hoon Tan, arXiv:1711.09553v2 [cs.CV] 26 Feb 2018
- [6] E. Khvedchenya V. I. Iglovikov A. Buslaev, A. Parinov and A. A. Kalinin. Albumentations: fast and flexible image augmentations. ArXiv e-prints, 2018
- [7] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017
- [8] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018
- [9] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983, 2016
- [10] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946, 2019
- [11] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955, 2020