

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Σχολή Θετικών Επιστημών
Τμήμα Φυσικής

Πρόγραμμα Μεταπτυχιακών Σπουδών Ηλεκτρονικής Φυσικής
(Ραδιοηλεκτρολογίας)
Κατεύθυνση Τηλεπικοινωνιών

Ταξινόμηση Οπωροκηπευτικών με χρήση Βαθιών Συνελικτικών Δικτύων

Διπλωματική Εργασία
της
Θεοδώρας Γκαρτζονίκα

Επιβλέπων καθηγητής: Σωτήριος Γούδος

Θεσσαλονίκη 2021

Ευχαριστίες

Πρώτα από όλου θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Σωτήριο Γούδο για την εμπιστοσύνη που μου έδειξε και την συμβολή του στη διεκπεραίωση της διπλωματικής μου εργασίας.

Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω την αδερφή μου Ιωάννα για την καθοριστική της βοήθεια, την υπομονή και συνεχή ενθάρρυνση που μου παρείχε.

Τέλος ευχαριστώ την οικογένεια και τους φίλους μου για την συμπαράστασή και στήριξή τους.

Περίληψη

Στην παρούσα διπλωματική εργασία γίνεται χρήση βαθιών συνελικτικών δικτύων για την ταξινόμηση οπωροκηπευτικών προϊόντων. Το σύνολο δεδομένων που χρησιμοποιείται περιείχε 131 είδη φρούτων και λαχανικών και αποτελούνταν από RGB εικόνες.

Κατά την διαδικασία της εκπαίδευσης των δικτύων τα δεδομένα αυξάνονται με αλλαγή της θέσης των εικονοστοιχείων και της τιμής τους. Τα παραπάνω επιτυγχάνονται με αναστροφή των εικόνων και προσαρμογή του κορεσμού και της απόχρωσης των εικονοστοιχείων.

Για την ταξινόμηση των οπωροκηπευτικών προτείνεται ένα συνελικτικό νευρωνικό δίκτυο. Η εξαγωγή των χαρακτηριστικών γίνεται με τέσσερα συνελικτικά επίπεδα, το καθένα ακολουθούμενο από επίπεδο υποδειγματολήψιας. Ο ταξινομητής του προτεινόμενου δικτύου αποτελείται από δύο πλήρως συνδεδεμένα επίπεδα, το καθένα ακολουθούμενο από επίπεδα απόσυρσης. Επίσης εφαρμόζεται μέθοδος μεταφορά εκπαίδευσης κάνοντας χρήση του προ εκπαιδευμένου μοντέλου ResNet-50. Για την ταξινόμηση με ResNet-50, χρησιμοποιούνται τα βάρη του προ εκπαιδευμένου δικτύου σε ένα παραλλαγμένο ResNet-50 δίκτυο. Στο δίκτυο παραλλαγής ResNet-50, η εξαγωγή των χαρακτηριστικών γίνεται χρησιμοποιώντας την αρχιτεκτονική του ResNet-50, ενώ χρησιμοποιείται ο ταξινομητής του προτεινόμενου δικτύου.

Πραγματοποιήθηκαν δοκιμές μεταβάλλοντας τον χρωματικό χώρο των εικόνων εισόδου. Επίσης μελετάται η επίδραση της εξισορρόπησης του συνόλου δεδομένων.

Στο προτεινόμενο δίκτυο με είσοδο εικόνων στον RGB χρωματικό χώρο, επιτεύχθηκε ακρίβεια 97,76%. Τέλος, τα καλύτερα αποτελέσματα δόθηκαν με χρήση της παραλλαγής του ResNet-50 και είσοδο RGB εικόνες, πετυχαίνοντας ακρίβεια 99.55%.

Abstract

In this master thesis, Convolutional Neural Networks are used for the classification of fruits and vegetables. The dataset that is used has 131 species of fruits and vegetables and is consisted of RGB images.

During the training process of the networks, data are augmented by changing the position and the value of pixels. This is achieved by flipping the image and adjusting the value of saturation and hue of pixels.

A Convolutional Neural Network is proposed for the classification of fruits and vegetables. Feature extraction is done by four convolutional layers, each on followed by a pooling layer. The classifier of this network is consisted by two fully connected layers, each one followed by a dropout layer. Also transfer learning method is applied by using the pre-trained network ResNet-50. The classification is done by using the weights of the pretrained network to a modified ResNet-50. The modified ResNet-50 network uses as feature extraction the architecture of ResNet-50 and as classifier the one that was proposed.

Tests were performed by changing the colour space if the input images. Also, the effect of balanced dataset is studied.

The proposed network achieved accuracy of 97,76% by using images in RGB colour space. Finally, best results were given by modified ResNet-50 with input RGB images, by achieving accuracy of 99,55%.

Περιεχόμενα

1. Εισαγωγή	8
1.1 Περιγραφή του προβλήματος	8
1.2 Στόχος εργασίας	8
2. Βιβλιογραφική Επισκόπηση.....	9
2.1 Υπάρχουσες προσεγγίσεις ταξινόμησης οπωροκηπευτικών	9
2.1.1 Ταξινόμηση με μεθόδους Μηχανικής Μάθησης	9
2.1.2 Ταξινόμηση με Συνελκτικά Νευρωνικά Δίκτυα	10
2.2 Παραδείγματα χρήσης του Fruits-360	11
3. Θεωρητικό υπόβαθρο και εργαλεία	13
3.1 Τεχνητή Νοημοσύνη.....	13
3.2 Μηχανική Μάθηση.....	14
3.3 Τεχνητά Νευρωνικά Δίκτυα.....	14
3.3.1 Δομικά στοιχεία Τεχνητών Νευρωνικών Δικτύων.....	14
3.3.2 Εκπαίδευση μοντέλου σε προβλήματα ταξινόμησης.....	18
3.4 Συνελκτικά Νευρωνικά Δίκτυα.....	21
3.5 Υπολειμματικά Νευρωνικά Δίκτυα	27
3.5.1 ResNet-50	28
3.6 Μεταφορά Μάθησης	29
3.7 Προγραμματιστικές πλατφόρμες και εργαλεία	29
4. Σύνολα δεδομένων και προ-επεξεργασία εικόνων	31
4.1 Σύνολο δεδομένων Fruits-360.....	31
4.2 Επεξεργασία εικόνων	34
4.2.1 Χρωματικοί Χώροι	34
4.2.2 Τεχνικές Αύξησης Δεδομένων	35
4.3 Ασύμμετρη κατανομή κλάσεων	37
5. Πειραματική Διαδικασία	38
5.1 Αρχιτεκτονική Μοντέλων	38
5.1.1 Προτεινόμενο CNN.....	38
5.1.2 Προτεινόμενη παραλλαγή ResNet-50	40
5.2 Μετρικές.....	40
5.2.2 Sparse Categorical Cross Entropy Loss.....	40
5.2.3 Accuracy.....	41

5.3	Πειράματα	42
5.3.2	Υλοποίηση σε χρωματικούς χώρους	42
5.3.3	Υλοποίηση με ισορροπημένο σύνολο δεδομένων	44
5.3.4	Υλοποίηση με παραλλαγή ResNet-50	45
6	Αποτελέσματα.....	46
6.1	Αξιολόγηση – Αποτελέσματα	46
6.2	Οπτικοποίηση με Score-CAM	48
6.2.1	Θεωρητική Εισαγωγή	48
6.2.2	Χάρτες Χαρακτηριστικών	50
6.2.3	Αποτελέσματα οπτικοποίησης	51
6.3	Μελλοντικές επεκτάσεις	53
	Βιβλιογραφία	54
	Παράρτημα Α	57
	Παράρτημα Β	61
	Παράρτημα Γ	67

1. Εισαγωγή

1.1 Περιγραφή του προβλήματος

Η γεωργία είναι μια από τις πιο σημαντικές βιομηχανίες του κόσμου, αφού τα προϊόντα που παράγει είναι προϊόντα άμεσης χρήσης από τον άνθρωπο. Αποτελείται από ένα σύνολο δραστηριοτήτων όπως η καλλιέργεια του εδάφους, η παραγωγή και η συλλογή αγροτικών προϊόντων και η πρωτογενής επεξεργασία τους. Οι εικόνες αποτελούν σημαντική πηγή πληροφοριών για την επιστήμη της γεωπονίας. Τα τελευταία χρόνια, κυρίως λόγω της προόδου στις τεχνικές βαθιάς μάθησης και συγκεκριμένα των συνελκτικών νευρωνικών δικτύων, η χρήση μηχανικής όρασης έχει ενσωματωθεί στη γεωργική βιομηχανία. Η ενσωμάτωση αυτή έχει ως στόχο την αυτοματοποίηση διαδικασιών, τη μείωση του κόστους παραγωγής και την αύξηση της παραγωγικότητας.

Στον αγροτικό τομέα έχουν εφαρμοστεί διάφοροι μέθοδοι βαθιάς μάθησης α) για την εκτίμηση της απόδοσής καλλιεργειών [1], β) για την ταξινόμηση εκτάσεων σπαρτών με χρήση δορυφορικών [2] και UAV [3] εικόνων, γ) για την ανίχνευση αγριόχορτων σε καλλιεργήσιμα χωράφια [4]. Στη συλλογή και επεξεργασία των αγροτικών προϊόντων, έχει γίνει χρήση μεθόδων βαθιάς μάθησης για την καταμέτρηση φρούτων σε οπωρώνες [5], την ταξινόμηση τους ανάλογα με το επίπεδο ωρίμανσης [6] και τη συγκομιδή με χρήση ρομπότ [7].

Η παρούσα διπλωματική πραγματεύεται την υλοποίηση ενός μοντέλου για την αυτόματη ταξινόμηση οπωροκηπευτικών, φρούτων και λαχανικών. Η ταξινόμηση φρούτων και λαχανικών είναι ένα περίπλοκο πρόβλημα λόγω της τεράστιας ποικιλίας σε είδη και του ακανόνιστου σχήματος, χρώματος, μεγέθους και υφής τους. Μία τέτοια εφαρμογή θα μπορούσε να χρησιμοποιηθεί σε συνδυασμό με άλλα εργαλεία, σε προβλήματα που σχετίζονται με τη συλλογή, την επεξεργασία και τη διάθεση φρούτων και λαχανικών. Τέτοια προβλήματα είναι η εκτίμηση της σοδειάς οπωρώνων και η αυτόματη συγκομιδή φρούτων με χρήση ρομπότ. Επίσης τα οπωροκηπευτικά προϊόντα αποτελούν μεγάλο μέρος της βιομηχανίας τροφίμων, έτσι θα μπορούσε να χρησιμοποιηθεί στον τομέα λιανικής, σε super-market και μανάβικα για την αυτόματη αναγνώριση και κοστολόγηση των φρούτων στα ταμεία [8].

1.2 Στόχος εργασίας

Στόχος της εργασίας είναι η υλοποίηση ενός αλγορίθμου για την αυτόματη ταξινόμηση 131 ειδών φρούτων και λαχανικών με χρήση Βαθιών Συνελκτικών Νευρωνικών Δικτύων. Παράλληλα εφαρμόστηκε μιας μέθοδος Επεξηγούμενης Τεχνητής Νοημοσύνης.

2. Βιβλιογραφική Επισκόπηση

2.1 Υπάρχουσες προσεγγίσεις ταξινόμησης οπωροκηπευτικών

2.1.1 Ταξινόμηση με μεθόδους Μηχανικής Μάθησης

Ένας μεγάλος αριθμός από έρευνες έχουν γίνει για την ταξινόμηση φρούτων και λαχανικών. Χρησιμοποιώντας τεχνικές επεξεργασίας, οι εικόνες φρούτων και λαχανικών μπορούν να περιγραφούν με ένα σύνολο χαρακτηριστικών όπως το χρώμα, το μέγεθος, το σχήμα και την υφή. Στη συνέχεια στα εξαγόμενα χαρακτηριστικά εφαρμόζεται ένας αλγόριθμος Μηχανικής Μάθησης με σκοπό την ταξινόμηση τους στην αντίστοιχη κλάση [9]. Δημοφιλής αλγόριθμοι Μηχανικής Μάθησης είναι τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Network), οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines), Γραμμική Παλινδρόμηση (Linear Regression), Λογιστική Παλινδρόμηση (Logistic Regression), Δίκτυα Bayes (Naive Bayes), Γραμμική Διακριτική Ανάλυση (Linear Discriminant Analysis), Δέντρα Απόφασης (Decision Trees), K - κοντινότεροι Γείτονες (K-nearest Neighbor).

Η πρώτη προσπάθεια κατασκευής συστήματος αναγνώρισης και ταξινόμησης φρούτων και λαχανικών σε super market πραγματοποιήθηκε το 1997 από το τμήμα έρευνας της IBM [10]. Το σύστημα ονομάστηκε Veggie-Vision και είχε την ικανότητα να ταξινομεί τα φρούτα και τα λαχανικά με βάση το χρώμα, την υφή και την πυκνότητά τους. Η ταξινόμηση επιτυγχάνονταν με τον αλγόριθμο Μηχανικής Μάθησης K-Κοντινότεροι Γείτονες (K-Nearest Neighbor). Όταν συνδυάζονταν τα χαρακτηριστικά του χρώματος και την υφής, η ακρίβεια έφτανε το 95%, για τον υπολογισμό της οποίας λαμβάνονταν υπόψη οι πέντε πρώτες πιθανές προβλέψεις.

Στην συνέχεια οι Arivazhagan et al. [11], κατηγοριοποίησαν 15 τύπους φρούτων. Με χρήση αλγορίθμων εξήγαγαν τα χαρακτηριστικά του χρώματος και της υφής, τα οποία συνδύασαν και με τον στατιστικό αλγόριθμο ελάχιστης απόστασης πέτυχαν ακρίβεια 86% .

Οι Zhang and Wu [12] πρότειναν μια άλλη προσέγγιση ταξινόμησης φρούτων. Μελέτησαν διάφορες μεθόδους για εξαγωγή των χαρακτηριστικών του χρώματος, της υφής και του σχήματος. Έπειτα εφάρμοσαν τρεις διαφορετικούς ταξινομητές Διανυσμάτων Υποστήριξης (Support Vector Machines). Το μοντέλο τους σημείωσε ακρίβεια 82,6% στην ταξινόμηση 18 τύπων φρούτων.

Οι Suresha et al. [13] κατάφεραν να ταξινομήσουν με ακρίβεια 95% 8 τύπους λαχανικών, εξαγοντας τα χαρακτηριστικά χρώμα και υφή και χρησιμοποιώντας τον αλγόριθμο Απόφασης Δέντρου (Decision Tree).

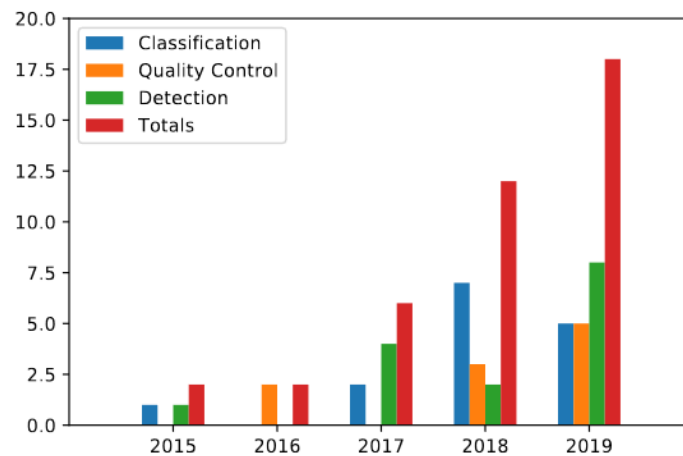
Οι Chowdhury et al. [14] χρησιμοποίησαν αλγόριθμους εξαγωγής των χαρακτηριστικών χρώματος και υφής σε συνδυασμό με ένα Νευρωνικό Δίκτυο. Η ταξινόμηση έγινε σε 10 είδη λαχανικών με 96,55% ακρίβεια.

Οι Danti et al. [15], ταξινόμησαν 10 είδη λαχανικών με ακρίβεια 96,40%. Τα χαρακτηριστικά που τροφοδότησαν στον ταξινομητή αποτελούνταν από τον μέσο όρο και το εύρος των τιμών της απόχρωσης και του κορεσμού σε HSV εικόνες. Ο ταξινομητής που χρησιμοποιήθηκε ήταν ένα Οπίσθιας Διάδοσης Νευρωνικό Δίκτυο (Back Propagation Neural Network).

2.1.2 Ταξινόμηση με Συνελκτικά Νευρωνικά Δίκτυα

Η ταξινόμηση οπωροκηπευτικών έχει πολλές προκλήσεις που σχετίζονται κυρίως με το ακανόνιστο σχήμα και μέγεθος, καθώς και το μεταβλητό χρώμα τους. Πρακτικά όλες οι φυσικές πτυχές (σχήμα, χρώμα, υφή) των φρούτων και λαχανικών αποτελούν φυσικά χαρακτηριστικά και πρέπει να συμπεριληφθούν με τα υπόλοιπα χαρακτηριστικά του φρούτου στην διαδικασία της ταξινόμησης [9]. Η χρήση των Συνελκτικών Δικτύων βοήθησε στην συμπερίληψη αυτών των χαρακτηριστικών, αφού το δίκτυο είναι σε θέση να αναγνωρίζει χαρακτηριστικά μέσα από τα φίλτρα που μαθαίνει κατά την εκπαίδευση.

Στο παρακάτω Διάγραμμα φαίνεται πως εξελίχθηκε η χρήση των Συνελκτικών Νευρωνικών Δικτύων για την ταξινόμηση, τον ποιοτικό έλεγχο και την ανίχνευση οπωροκηπευτικών, το διάστημα 2015-2019 (Σχήμα 1.1). Τα δεδομένα προέκυψαν ψάχνοντας τις εργασίες που δημοσιεύτηκαν το διάστημα αυτό σε επιστημονικές βάσεις δεδομένων και χρησιμοποιώντας ως λέξεις κλειδιά “Fruits”, “Convolution Neural Network”, “Deep learning”.



Σχήμα 1. 1: Διάγραμμα πλήθους εργασιών συναρτήσει του έτους έκδοσής τους.

Κοιτώντας τις εργασίες που εκπονήθηκαν με σκοπό την ταξινόμηση φρούτων φαίνεται ότι οι συγγραφείς είτε πρότειναν μικρά συνελκτικά δίκτυα των 5-9 επιπέδων, είτε επέλεξαν την χρήση προ εκπαιδευμένων μοντέλων.

Παραδείγματος χάρη οι Wang and Chen [16] δημιούργησαν ένα μοντέλο αποτελούμενο από 8 συνελκτικά επίπεδα, που χρησιμοποιεί ReLU και έχει ένα επίπεδο απόρριψης πριν από κάθε πλήρως συνδεδεμένο επίπεδο. Το σύνολο δεδομένων αποτελείται 3600 εικόνες που ανήκουν σε 18 τύπους φρούτων, με εικόνες που συλλέξαν από το Google και το Baidu. Η προεπεξεργασία των εικόνων περιελάμβανε μεταφορά του φρούτου στο κέντρο, κόψιμο των άκρων, αλλαγή μεγέθους και αφαίρεση φόντου. Η ακρίβεια του μοντέλου ήταν 95,67%.

Ένα παράδειγμα μεταφοράς μάθησης είναι η εργασία των Patino-Saucedo et al. [17]. Οι συγγραφείς χρησιμοποίησαν το μοντέλο βασισμένο στον αλγόριθμο AlexNet [18] με σκοπό να ταξινομήσουν τροπικά φρούτα. Χρησιμοποίησαν το σύνολο δεδομένων “The Supermarket Produce dataset” [19], που αποτελείται από 2633 εικόνες με 15 κλάσεις. Η ακρίβεια του μοντέλου ήταν 99,56% και 100% με την χρήση χαρακτηριστικών χρώματος και υφής, αντίστοιχα.

2.2 Παραδείγματα χρήσης του Fruits-360

Το σύνολο δεδομένων που χρησιμοποιήθηκε στην εργασία είναι το “Fruits-360: A dataset of images containing fruits and vegetables”. Από τη δημιουργία του το 2018 και μετά έχει χρησιμοποιηθεί σε διάφορες εργασίες που ερευνούν την ταξινόμηση φρούτων και λαχανικών. Παρακάτω κάποια παραδείγματα της βιβλιογραφίας, στα οποία έχει γίνει χρήση του Fruits-360.

Οι Muresan και Oltean [20], είναι η δημιουργοί του Fruits-360 dataset, το οποίο αποτελείται από RGB εικόνες φρούτων και λαχανικών μεγέθους 100x100 [21]. Οι συγγραφείς πρότειναν ένα δίκτυο αποτελούμενο από τρία συνελικτικά επίπεδα με συνάρτηση ενεργοποίησης την ReLU, καθένα από τα οποία ακολουθούνταν από επίπεδο μέγιστης υποδειγματοληψίας. Στην συνέχεια υπήρχε ένα επίπεδο εξισορρόπησης και τρία πλήρως συνδεδεμένα επίπεδα, τα δύο πρώτα χρησιμοποιούσαν ReLU και ακολουθούνταν από επίπεδα απόρριψης, ενώ το τελευταίο αποτελούνταν από ένα ταξινομητή με συνάρτηση ενεργοποίησης Softmax. Επίσης έγιναν μελέτες και σε διαφορετικούς χρωματικούς χώρους. Με τον κατάλληλη αρχιτεκτονική κατάφεραν να πετύχουν ακρίβεια 98.66% στην ταξινόμηση 131 φρούτων. Η παρούσα διπλωματική βασίστηκε στην παραπάνω εργασία.

Οι Shakib et al. [22] κατασκεύασαν ένα δίκτυο αποτελούμενο από δύο συνελικτικά επίπεδα με συνάρτηση ενεργοποίησης την ReLU, καθένα από τα οποία ακολουθούνταν από ένα επίπεδο μέγιστης υποδειγματοληψίας. Στην συνέχεια τα δεδομένα πέρασαν από ένα επίπεδο απόρριψης και δύο πλήρως συνδεδεμένα επίπεδα, με το δεύτερο να αποτελεί έναν Softmax ταξινομητή. Μελέτησαν την ακρίβεια του δικτύου διαμορφώνοντας τις παραμέτρους των συνελικτικών επιπέδων, όπως τον αριθμό των φίλτρων, τη χρήση ή όχι επιπέδου απόρριψης και πειραματίστηκαν με τον αριθμό των εποχών. Με τις κατάλληλες συνθέσεις κατάφεραν να ταξινομήσουν με ακρίβεια 99,79% για εκπαίδευση με 15 εποχές και με τη χρήση επιπέδου απόρριψης. Χρησιμοποίησαν 25 κατηγορίες φρούτων από το σύνολο δεδομένων Fruits-360.

Οι Rajasekare et al. [23] εφάρμοσαν μεθόδους εξαγωγής των χαρακτηριστικών του χρώματος, σχήματος και υφής. Μελέτησαν την ταξινόμηση των εικόνων χρησιμοποιώντας πρώτα κάθε χαρακτηριστικό μόνο του, στην συνέχεια συνδυασμούς ανά δύο (χρώμα-σχήμα, σχήμα-υφή, χρώμα-υφή) και τέλος συνδέοντας όλα τα χαρακτηριστικά μαζί. Η ταξινόμηση έγινε με τους αλγορίθμους Μηχανικής Μάθησης K-Κοντινότεροι Γείτονες, Ανάλυση Γραμμικής Διάκρισης (Linear Discriminant Analysis), Δίκτυα Bayes, Κωδικοποίηση Διόρθωσης Λαθών Εξόδου (Error Output Control Classifier – ECOC) και Δέντρο Απόφασης. Καλύτερα αποτελέσματα ακρίβειας έδωσε ο αλγόριθμος K-NN. Όσον αφορά τα χαρακτηριστικά εισόδου, στη χρήση ενός μόνο χαρακτηριστικού μεγαλύτερη ακρίβεια επιτεύχθηκε με το χρώμα (ακρίβεια 99,65%) και στη χρήση συνδυασμού χαρακτηριστικών με τον συνδυασμό χρώματος-σχήματος (ακρίβεια 99,93%). Τέλος, στην περίπτωση συνδυασμού όλων των χαρακτηριστικών επιτεύχθηκε ακρίβεια 99,93%. Από το Fruits-360 dataset επιλέχθηκαν 24 κατηγορίες φρούτων μαζί με 100 εικόνες τους, έτσι το σύνολο δεδομένων αποτελούνταν από 2400 εικόνες φρούτων. Η αξιολόγηση των αποτελεσμάτων έγινε με την μέθοδο 10-fold cross-validation, δηλαδή για 10 φορές το σύνολο δεδομένων χωρίζονταν τυχαία με αναλογία 60-40, 60% των δεδομένων ανήκε στο σύνολο εκπαίδευσης και 40% στο σύνολο ελέγχου και υπολογίζονταν η ακρίβεια. Η τελική τιμή της ακρίβειας ήταν ο μέσος όρος των τιμών ακρίβειας κάθε προσπάθειας.

Τέλος, οι Narsir et al. [24], μελέτησαν την ταξινόμηση 95 κατηγοριών φρούτων. Χρησιμοποιώντας το VGG19 και με τη μέθοδο PHOG εξήγαγαν τους χάρτες χαρακτηριστικών. Έπειτα με την μέθοδο mRMR επιλέχθηκαν κάποια από αυτά τα χαρακτηριστικά, τα οποία τροφοδοτήθηκαν σε επτά διαφορετικούς ταξινομητές. Καλύτερα αποτελέσματα παρουσίασε η ταξινόμηση με αλγόριθμο Κυβικής Διανυσματικής Υποστήριξης (Cubic SVM) και ακρίβεια 99,6%. Η αξιολόγηση των αποτελεσμάτων έγινε με την μέθοδο 10-fold cross-validation [25].

3. Θεωρητικό υπόβαθρο και εργαλεία

3.1 Τεχνητή Νοημοσύνη

Από την αρχαιότητα ο άνθρωπος προσπάθησε να κατασκευάσει μηχανές, οι οποίες αυτοματοποιούσαν καθημερινές δραστηριότητες. Με την πάροδο του χρόνου και την εξέλιξη της τεχνολογίας οι επιστήμονες ξεκίνησαν ενεργά να ερευνούν την εύρεση αλγορίθμων, οι οποίοι θα μιμούνται λειτουργίες του ανθρώπινου εγκεφάλου όπως η μάθηση, η εξαγωγή συμπερασμάτων, η αντίληψη, ο συλλογισμός, κ.α. Έτσι η επιστήμη της Τεχνητής Νοημοσύνης - TN (Artificial Intelligence – AI) γεννήθηκε.

Ολόκληρο το εύρος της τεχνητής νοημοσύνης μπορεί να διαιρεθεί σε δύο κατηγορίες, την γενική / δυνατή TN (Artificial Strong Intelligence, AGI, or Strong AI) και την αδύναμη/στενή TN (Artificial Narrow Intelligence, ANI, or Weak AI). Στην πρώτη κατηγορία, ανήκουν οι μηχανές που είναι ικανές να αναπτύξουν νόηση και συμπεριφορά ίδια με την ανθρώπινη. Μια μηχανή που διαθέτει γενική TN, έχει την ικανότητα να αποδίδει νοητικές καταστάσεις στον εαυτό της και τους άλλους και ταυτόχρονα να χρησιμοποιεί αυτές τις νοητικές καταστάσεις με στόχο να ερμηνεύσει την συμπεριφορά τους (Θεωρία του Νου - Premack and Woodruff 1987 [26]). Η έλλειψη βαθιάς γνώσης στον τρόπο λειτουργίας του ανθρώπινου μυαλού δυσκολεύει το έργο των επιστημόνων, οι οποίοι δεν έχουν κατορθώσει ακόμη να κατασκευάσουν μια μηχανή που να κατέχει γενική TN. Στη δεύτερη κατηγορία, όπως υποδηλώνει το όνομά της, ανήκουν οι μηχανές που εκτελούν ένα συγκεκριμένο έργο, τα συστήματα αυτά δεν διαθέτουν ευφυΐα, η λειτουργία τους είναι περιορισμένη και δεν μπορεί να συγκριθεί με αυτή του ανθρώπινου εγκεφάλου. Στην καθημερινότητά του ο σύγχρονος άνθρωπος χρησιμοποιεί πολλές εφαρμογές αδύναμης TN.

Η Αδύναμη TN ανάλογα με το πεδίο μελέτης της χωρίζεται στα παρακάτω υποσύνολα:

- Μηχανική Όραση (Computer Vision), το πεδίο που οι μηχανές εκπαιδεύονται στην αναγνώριση και ερμηνεία του οπτικού κόσμου χρησιμοποιώντας ψηφιακές εικόνες.
- Μηχανική Ομιλία (Computer Speech), το πεδίο που οι μηχανές εκπαιδεύονται να κατανοήσουν τον προφορικό λόγο και να τον μεταφέρουν σε γραπτό και το αντίστροφο.
- Μηχανική Μάθηση (Machine Learning), το πεδίο που δίνεται η ικανότητα σε μηχανές να μαθαίνουν και να βελτιώνονται, χωρίς να έχουν προγραμματιστεί ρητά για την εκάστοτε λειτουργία.
- Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP), το πεδίο που οι μηχανές εκπαιδεύονται να “καταλαβαίνουν” τον ανθρώπινο λόγο, να εξάγουν συμπεράσματα και να παράγουν λόγο, δηλαδή να είναι σε θέση να πάρουν μέρος σε μια συζήτηση, να ρωτούν και να απαντούν.
- Ρομποτική (Robotics), το πεδίο που ασχολείται με τον σχεδιασμό και την κατασκευή μηχανών που αλληλοεπιδρούν με τον υλικό κόσμο με σκοπό τη συλλογή, μετακίνηση και τροποποίηση των αντικειμένων, απαλλάσσοντας το ανθρώπινο δυναμικό από τις επαναλαμβανόμενες εργασίες.
- Έμπειρα συστήματα (Expert Systems), το πεδίο που ασχολείται με την κατασκευή συστημάτων, τα οποία μιμούνται τη δυνατότητα λήψης αποφάσεων ενός ανθρώπου και χρησιμοποιούνται για την επίλυση σύνθετων προβλημάτων.

3.2 Μηχανική Μάθηση

Σύμφωνα με τον ορισμό που δόθηκε από τον T. Mitchell (1997) [27] η Μηχανική Μάθηση είναι η διαδικασία που:

“Ένα πρόγραμμα υπολογιστή μαθαίνει από μια εμπειρία E ως προς κάποια κλάση διεργασιών T και μέτρο απόδοσης P , εάν η απόδοσή του στις διεργασίες T , όπως μετριέται από το P , βελτιώνεται μέσω της εμπειρίας E ”.

Στον κλάδο της Μηχανικής Μάθησης διακρίνονται τρεις μορφές μάθησης:

- **Επιβλεπόμενη Μάθηση (Supervised Learning).** Στους αλγόριθμους Επιβλεπόμενης Μάθησης δίνεται ένα σύνολο δεδομένων σε συνδυασμό με την επιθυμητή έξοδο. Σκοπός είναι ο αλγόριθμος να εξαγεί συσχετίσεις μεταξύ εισόδου και εξόδου. Η Επιβλεπόμενη Μάθηση χρησιμοποιείται κυρίως σε προβλήματα:
 1. Ταξινόμησης (Classification)
 2. Πρόγνωσης (Prediction)
 3. Διερμηνείας (Interpretation)
- **Μη-Επιβλεπόμενη Μάθηση (Unsupervised Learning).** Στους αλγόριθμους Μη-Επιβλεπόμενης Μάθησης εισάγονται μόνο τα δεδομένα, χωρίς την επιθυμητή έξοδο, έτσι το μοντέλο βρίσκει συσχετίσεις - μοτίβα μεταξύ των εισόδων. Χρησιμοποιείται κυρίως σε προβλήματα:
 1. Ανάλυσης Συσχετισμών (Association Analysis)
 2. Ομαδοποίησης (Clustering)
- **Ενισχυτική Μάθηση (Reinforcement Learning),** ο αλγόριθμος εκπαιδεύεται μέσα από ακολουθίες ζευγών κατάστασης – δράσης. Για κάθε ζεύγος προκύπτει ένα αποτέλεσμα-ανταμοιβή. Στόχος του αλγορίθμου είναι για μια κατάσταση να προβλέψει με ακρίβεια τη δράση με την μεγαλύτερη ανταμοιβή. Χρησιμοποιείται κυρίως σε προβλήματα:
 1. Σχεδιασμού (Planning)

3.3 Τεχνητά Νευρωνικά Δίκτυα

Ένας από τους σημαντικότερους αλγόριθμους Επιβλεπόμενης Μηχανικής Μάθησης είναι τα Τεχνητά Νευρωνικά Δίκτυα.

3.3.1 Δομικά στοιχεία Τεχνητών Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Network-ANN), μιμούνται τη δομή των βιολογικών νευρώνων από τα οποία και έχουν εμπνευστεί. Πιο συγκεκριμένα, οι νευρώνες είναι κύτταρα στους εγκεφάλους ζώντων οργανισμών, που συνδέονται μεταξύ τους δημιουργώντας δίκτυα και επικοινωνούν με ηλεκτρικούς παλμούς. Οι συνδέσεις που σχηματίζουν ονομάζονται συνάψεις και δεν είναι ούτε σταθερές ούτε στατικές, αλλά δημιουργούνται, αλλάζουν και καταστρέφονται συνεχώς.

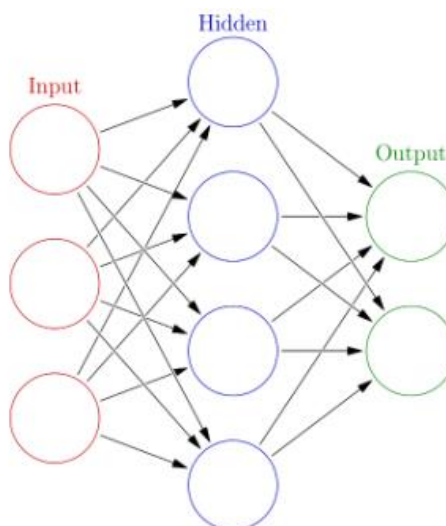
Οι νευρώνες είναι μέρος του νευρικού δικτύου, το οποίο δέχεται ερεθίσματα από το εξωτερικό περιβάλλον. Η πληροφορία των ερεθισμάτων επεξεργάζεται στους νευρώνες, οι οποίοι αποφασίζουν και παράγουν την έξοδο για την αντίστοιχη αντίδραση. Ο Haykin [28] πρόσφερε τον παρακάτω ορισμό για τα νευρωνικά δίκτυα:

“Ένα νευρωνικό δίκτυο είναι ένας μαζικός, παράλληλα κατανεμημένος επεξεργαστής κατασκευασμένος από απλές επεξεργαστικές μονάδες που έχουν τη φυσική ροπή να αποθηκεύουν εμπειρική γνώση και να τη διαθέτουν προς χρήση. Θυμίζουν τον εγκέφαλο σε δύο σημεία:

1. Το δίκτυο αποκτά τη γνώση από το περιβάλλον μέσω μιας διαδικασίας μάθησης.
2. Οι διασυνδέσεις μεταξύ των νευρώνων, γνωστές και ως συνοπτικά βάρη, χρησιμοποιούνται για την αποθήκευση της γνώσης”.

Η πιο απλή δομή ενός ANN παρουσιάζεται στο Σχήμα 3.1, όπου διακρίνονται τρία επίπεδα:

- το επίπεδο εισόδου (Input Layer)
- το κρυμμένο επίπεδο (Hidden Layer)
- το επίπεδο εξόδου (Output Layer)



Σχήμα 3. 1: Βασική αρχιτεκτονική Τεχνητού Νευρωνικού Δικτύου.

Οι νευρώνες είναι οι συνδεδεμένοι κόμβοι μεταξύ των επιπέδων, ενώ οι συνδέσεις μεταξύ τους ονομάζονται ακμές. Οι νευρώνες ενός επιπέδου είναι συνδεδεμένοι με όλους ή κάποιους από τους νευρώνες του προηγούμενου ή του επόμενου επιπέδου.

Κάθε νευρώνας δέχεται ένα σύνολο αριθμητικών εισόδων είτε από άλλους νευρώνες είτε από το περιβάλλον. Οι νευρώνες εισόδου και εξόδου εκπληρώνουν διαφορετικούς ρόλους από αυτούς των νευρώνων που βρίσκονται στα κρυμμένα επίπεδα. Οι πρώτοι αναλαμβάνουν να τροφοδοτήσουν το δίκτυο με την είσοδο και οι τελευταίοι αναλαμβάνουν να διοχετεύσουν στο περιβάλλον τις εξόδους.

Οι νευρώνες που βρίσκονται στα κρυμμένα επίπεδα δέχονται εισόδους και εκτελούν υπολογισμούς, οι οποίοι παράγουν τις εξόδους. Κάθε νευρώνας εκτελεί μια συνάρτηση μεταφοράς f_i της μορφής:

$$y_i = f_i \left(\sum_{j=1}^n w_{ij} x_j + b_i \right) \quad (3.1)$$

, όπου y_i η είσοδος του νευρώνα i , x_j η j^{th} είσοδος στον i νευρώνα, w_{ij} το βάρος που συνδέει τους νευρώνες i και j , b_i το κατώφλι (bias) του νευρώνα μέσω του οποίου μπορεί να ρυθμιστεί περεταίρω η ενεργοποίησή του.

Συνάρτηση Ενεργοποίησης

Η f_i ονομάζεται συνάρτηση ενεργοποίησης και βοηθά το ANN να εντοπίσει και να μάθει σύνθετα μοτίβα στα δεδομένα. Η συνάρτηση αυτή μπορεί να συγκριθεί με τη διαδικασία διέγερσης ενός βιολογικού νευρώνα. Είναι συνήθως μη γραμμική, ώστε να μαθαίνει σύνθετα και πολύπλοκα χαρακτηριστικά και διαφορίσιμη, ώστε το δίκτυο να μπορεί να ανανεώνει τα βάρη και τα κατώφλια των νευρώνων κατά την οπισθοδιάδοση της απώλειας. Τέλος είναι θεμιτό να πληροί το κριτήριο της μονοτονίας, ώστε να εγγυάται η εύρεση του ολικού ελαχίστου. Συνήθως χρησιμοποιούνται οι ακόλουθες συναρτήσεις:

- Γραμμικός Ανορθωτής (Linear Rectifier, ReLU). Η συνάρτηση ανόρθωσης μηδενίζει τις αρνητικές εισόδους, ενώ στις μη-αρνητικές παρουσιάζει μία γραμμική σχέση μεταξύ εισόδου-εξόδου. Έχει σύνολο τιμών το $[0, +\infty)$ και συνάρτηση:

$$\varphi(x) = \max(0, x) \quad (3.2)$$

Το χαρακτηριστικό να μηδενίζει κάποιους νευρώνες είναι αρκετά χρήσιμο αφού εισάγει μη γραμμικότητα στην έξοδο. Ωστόσο, το χαρακτηριστικό αυτό έχει ως αποτέλεσμα κατά την οπισθοδιάδοση τα βάρη και τα κατώφλια των ανενεργών νευρώνων να μην ενημερώνονται και να μην συνεισφέρουν στη διαδικασία της εκπαίδευσης. Το πρόβλημα αυτό καλείται το πρόβλημα του νεκρού νευρώνα.

- Λογιστική Συγμοειδής (Logistic Sigmoid). Είναι μια συνεχής συνάρτηση, γνησίως αύξουσα και παραγωγίσιμη σε όλο το πεδίο ορισμού της. Η συνάρτηση παρουσιάζεται στο Σχήμα 3.2 και περιγράφεται από την παρακάτω σχέση:

$$\sigma(x) = \frac{1}{1 + e^{-ax}} \quad (3.3)$$

Η παράμετρος a ονομάζεται παράμετρος κλίσης και καθορίζει την κλίση της καμπύλης. Το σύνολο τιμών της είναι $(0, 1)$ και χρησιμοποιείται κυρίως σε προβλήματα δυαδικής ταξινόμησης. Το γεγονός ότι η έξοδος κυμαίνεται μεταξύ 0 και 1 σημαίνει ότι εισοδοί με μεγάλες διακυμάνσεις στην τιμή τους προκαλούν πολύ μικρές αλλαγές στην έξοδο. Η σιγμοειδής, αν και παλαιότερα ήταν πολύ δημοφιλής, έχει αντικατασταθεί από τη ReLU κυρίως στα Βαθιά Νευρωνικά Δίκτυα εξαιτίας του προβλήματος της εξαφάνισης κλίσης (vanishing gradient) που προκαλεί.

- Υπερβολική Εφαπτομένη (Hyperbolic Tanget). Είναι μια σιγμοειδής συνάρτηση, που σημαίνει ότι είναι συνεχής, γνησίως αύξουσα και παραγωγίσιμη στο πεδίο ορισμού της. Το σύνολο τιμών εξαρτάται από την τιμή του a και ισούται με $(-a, a)$. Η συνάρτηση που την περιγράφει είναι η παρακάτω:

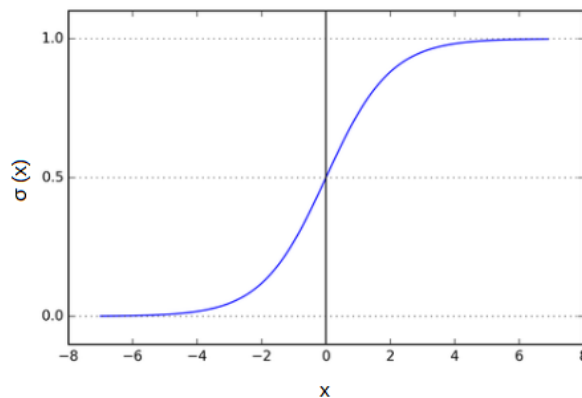
$$\varphi(x) = a * \tanh(\beta x) \quad (3.4)$$

ε περιπτώσεις που η τιμή του a είναι μικρή και το σύνολο τιμών της συνάρτησης αντιστοίχως μικρό εμφανίζεται το πρόβλημα της εξαφάνισης της κλίσης.

- Κανονικοποιημένη εκθετική (Softmax). Αποτελεί μια γενικευμένη μορφή της σιγμοειδούς συνάρτησης, είναι συνεχής, γνησίως αύξουσα και παραγωγίσιμη στο πεδίο ορισμού της. Η συνάρτηση που την περιγράφει είναι η παρακάτω:

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i^N e^{x_i}} \quad (3.5)$$

Ουσιαστικά μετατρέπει ένα διάνυσμα από αριθμούς σε διάνυσμα πιθανοτήτων. Το σύνολο τιμών της είναι $[0,1]$ με το άθροισμα των πιθανοτήτων να είναι ίσο με 1. Για αυτόν τον λόγο χρησιμοποιείται σε προβλήματα ταξινόμησης πολλών κλάσεων και συγκεκριμένα στο τελευταίο επίπεδο του δικτύου. Ο αριθμός των νευρώνων της εξόδου είναι ίσος με τον αριθμό των κλάσεων.



Σχήμα 3. 2 Σιγμοειδής συνάρτηση.

3.3.2 Εκπαίδευση μοντέλου σε προβλήματα ταξινόμησης

Εκπαίδευση καλείται η διαδικασία κατά την οποία οι ελεύθεροι παράμετροι ενός νευρωνικού δικτύου αναπροσαρμόζονται μέσα από μια διαδικασία διέγερσης από το περιβάλλον του δικτύου [29].

Κατά τη διαδικασία εκπαίδευσης ενός μοντέλου τα δεδομένα περνάνε από όλα τα επίπεδα και υπολογίζεται η πρόβλεψη του δικτύου. Στη συνέχεια, συγκρίνεται η πραγματική κλάση στην οποία ανήκουν τα δεδομένα με την πρόβλεψη του δικτύου και υπολογίζεται το διάνυσμα απώλειας. Τέλος, το διάνυσμα απώλειας ανατροφοδοτείται (backpropagation) διαδοχικά στα προηγούμενα επίπεδα, από το επίπεδο εξόδου μέχρι το επίπεδο εισόδου, και οι παράμετροι αναπροσαρμόζονται προς την κατεύθυνση μείωσης του σφάλματος. Αυτή η διαδικασία επαναλαμβάνεται μέχρι να ολοκληρωθεί η εκπαίδευση. Αναλυτικά:

1. Πρώτα τα βάρη και τα κατώφλια του δικτύου αρχικοποιούνται. Οι τιμές δεν πρέπει να είναι μηδενικές, ώστε να μην απενεργοποιηθούν οι νευρώνες και δεν πρέπει να έχουν την ίδια τιμή μεταξύ τους.
2. Αφού η είσοδος έχει περάσει από όλα τα επίπεδα (forward pass), υπολογίζεται η συνάρτηση απώλειας.
3. Στη συνέχεια με χρήση μιας συνάρτησης βελτιστοποίησης (optimizer) υπολογίζεται η νέα τιμή των παραμέτρων.
4. Τέλος οι τιμές των βαρών και κατωφλίων ενημερώνονται και το δίκτυο αναμένει την επόμενη είσοδο.

A. Συνάρτηση απώλειας

Στο δεύτερο βήμα της διαδικασίας εκπαίδευσης αναφέρθηκε η συνάρτηση απώλειας (loss function). Η συνάρτηση απώλειας συναντάται και ως συνάρτηση σφάλματος ή κόστους. Αξιολογεί πόσο λανθασμένες είναι οι προβλέψεις του αλγορίθμου και η έξοδός της είναι συνεχής αριθμός ή πιθανότητα. Η τιμή της ρυθμίζει το πόσο και προς τα που πρέπει να τροποποιηθούν οι παράμετροι του μοντέλου για να επιτευχθούν καλύτερες προβλέψεις.

Υπάρχουν τρεις κατηγορίες συναρτήσεων απώλειας:

- Απώλεια παλινδρόμησης (Regression loss). Χρησιμοποιείται όταν η πρόβλεψη παίρνει συνεχείς τιμές, π.χ. οι πωλήσεις μιας εταιρείας. Η απώλεια υπολογίζεται με μια από τις παρακάτω συναρτήσεις:
 - Απώλεια μέσου τετραγωνικού σφάλματος (Mean square error)
 - Απώλεια μέσου τετραγωνικού λογαριθμικού σφάλματος (Mean Squared Logarithmic Error loss)
 - Απώλεια μέσου απόλυτου σφάλματος (Mean Absolute Error Loss)
- Απώλεια Δυαδικής Ταξινόμησης (Binary Classification Loss). Χρησιμοποιείται σε περιπτώσεις που η έξοδος έχει δύο κλάσεις, π.χ. σε προβλήματα Ναι/Όχι. Η απώλεια υπολογίζεται με μια από τις παρακάτω συναρτήσεις:
 - Απώλεια δυαδικής Διασταυρωμένης Εντροπίας (Binary Cross-Entropy Loss).
 - Απώλεια Hinge (Hinge Loss).

- Απώλεια Ταξινόμησης Πολλών Κλάσεων (Multi-Class Classification Loss). Χρησιμοποιείται σε περιπτώσεις που το πρόβλημα έχει παραπάνω από δύο κλάσεις. Η συνάρτηση ενεργοποίησης του τελευταίου επιπέδου είναι μία Softmax. Η απώλεια υπολογίζεται με μια από τις παρακάτω συναρτήσεις:
 - ο Απώλεια διασταυρωμένης εντροπίας (Categorical Cross-Entropy Loss). Η έξοδος είναι διάνυσμα $1 \times N$, όπου N ο αριθμός των κλάσεων (one-hot encoding)
 - ο Απώλεια αραιής διασταυρωμένης εντροπίας (Sparse Categorical Cross-Entropy Loss). Η έξοδος είναι ένας ακέραιος αριθμός με εύρος $[0, N-1]$, όπου N ο αριθμός των κλάσεων.

B. Συνάρτηση Βελτιστοποίησης

Στο τρίτο βήμα της διαδικασίας εκπαίδευσης αναφέρθηκε η χρήση μιας συνάρτησης βελτιστοποίησης, που σκοπός της είναι η μείωση της απώλειας μέσω τροποποιήσεων στις παραμέτρους του δικτύου. Για να επιτευχθεί αυτό, υπολογίζονται οι κλίσεις της συνάρτησης απώλειας ως προς τις ελεύθερους παραμέτρους και οι παράμετροι αυτοί τροποποιούνται προς την κατεύθυνση αρνητικότερων κλίσεων (φθίνουσα κλίση - gradient descent).

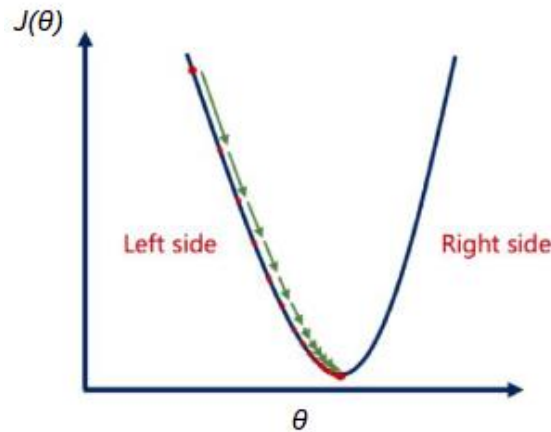
Μία από τις πιο διαδεδομένες συναρτήσεις βελτιστοποίησης και η οποία χρησιμοποιήθηκε για την παρούσα διπλωματική είναι η Stochastic Gradient Descent (SGD) και η οποία περιγράφεται από τον παρακάτω τύπο:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (3.6)$$

, όπου J η συνάρτηση απώλειας, α ο ρυθμός εκμάθησης (learning rate) και θ η ρυθμιζόμενη παράμετρος.

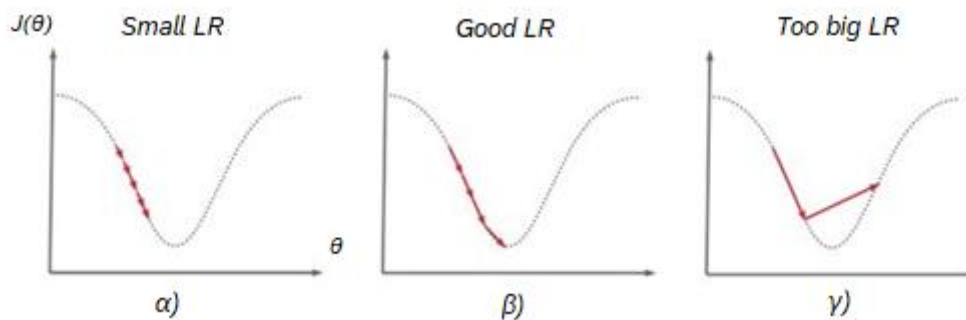
Στον αλγόριθμο SGD οι παράμετροι ανανεώνονται σε κάθε δείγμα ή σε κάθε παρτίδα (batch) του συνόλου εκπαίδευσης. Στην περίπτωση ανανέωσης των παραμέτρων σε κάθε δείγμα, ως απώλεια δίνεται η απώλεια του δείγματος αυτού. Στην περίπτωση ανανέωσης των παραμέτρων σε κάθε batch, η απώλεια δίνεται από τον μέσο όρο των απωλειών των δειγμάτων του batch. Οι αλλαγές που γίνονται δεν οδηγούν πάντοτε σε χαμηλότερες τιμές απώλειας, αλλά μακροπρόθεσμα η απώλεια μειώνεται. Ο αλγόριθμος είναι κατάλληλος για μεγάλα σύνολα δεδομένων. Η συχνές ανανεώσεις, ανά δείγμα ή ανά batch, προκαλούν μεγάλες διακυμάνσεις στην τιμή της απώλειας, παρόλα αυτά σε περίπτωση που η συνάρτηση έχει εγκλωβιστεί σε ένα τοπικό ελάχιστο, ο θόρυβος που εισάγουν αυτές οι ταλαντώσεις, μπορεί να είναι ικανός να κατευθύνει την συνάρτηση απώλειας στο ολικό ελάχιστο ή σε καλύτερο τοπικό ελάχιστο.

Σκοπός της συνάρτησης βελτιστοποίησης είναι η προσαρμογή των παραμέτρων προς την κατεύθυνση της πιο αρνητικής παραγώγου. Όπως φαίνεται στο Σχήμα 3.3, αν η παράγωγος είναι αρνητική τότε η απώλεια βρίσκεται στην αριστερή πλευρά της καμπύλης (left side) και το θ θα αυξηθεί με κατεύθυνση προς το ελάχιστο. Αντίθετα αν η παράγωγος είναι θετική τότε η απώλεια βρίσκεται στη δεξιά πλευρά (right side) της καμπύλης και το θ θα μειωθεί με κατεύθυνση προς το ελάχιστο. Στόχος της εκπαίδευσης είναι η συνάρτηση απώλειας να φτάσει στο ολικό της ελάχιστο.



Σχήμα 3. 3: Συνάρτηση απώλειας συναρτήσει της παραμέτρου θ .

Στη συνάρτηση βελτιστοποίησης, εμφανίζεται ο ρυθμός εκμάθησης. Ο ρυθμός εκμάθησης αποτελεί το βήμα με το οποίο η συνάρτηση βελτιστοποίησης θα κατευθυνθεί προς το ελάχιστο. Η επιλογή μικρού ρυθμού εκμάθησης έχει ως αποτέλεσμα αργή σύγκλιση στο ελάχιστο (Σχήμα 3.4(α)). Η επιλογή μεγάλου ρυθμού εκμάθησης σημαίνει ταχύτερη εκπαίδευση, αλλά είναι πιθανό η συνάρτησης απώλειας να ταλαντώνεται γύρω από το ελάχιστο της, χωρίς να καταφέρνει να συγκλίνει (Σχήμα 3.4(γ)) ή ακόμη και να αποκλίνει από αυτό.



Σχήμα 3. 4: Συνάρτηση απώλειας συναρτήσει της παραμέτρου θ με α) Μεγάλο ρυθμό εκμάθησης, β) Καλό ρυθμό εκμάθησης, γ) Μεγάλο ρυθμό εκμάθησης.

Για την περεταίρω βελτίωση του SGD έχουν προταθεί αλγόριθμοι που είτε ρυθμίζουν την Gradient descent (πχ momentum, Nesterov) , είτε τον ρυθμό μάθησης (πχ. AdaGrad, RMSprop, Adadelta), είτε και τα δύο (Adam, AMSGrad).

Στην εργασία χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης Adadelata, ο οποίος προσαρμόζει τον ρυθμό εκμάθησης στις παραμέτρους. Οι παράμετροι που σχετίζονται με συχνά εμφανιζόμενα χαρακτηριστικά ανανεώνονται λιγότερο σε σχέση με τις παραμέτρους των οποίων τα χαρακτηριστικά δεν εμφανίζονται συχνά. Ουσιαστικά οι παράμετροι με μεγάλες κλίσεις θα έχουν μικρό ρυθμό μάθησης ενώ αυτές με μικρές κλίσεις θα έχουν μεγάλο ρυθμό εκμάθησης.

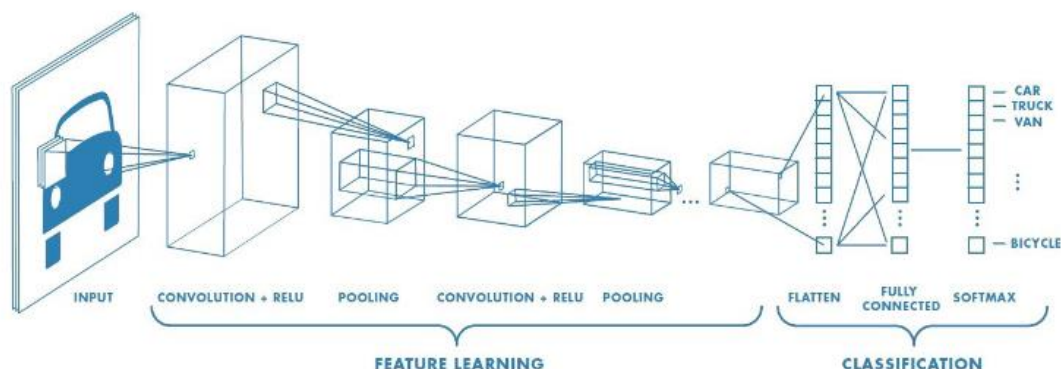
Με σκοπό την περεταίρω προσαρμογή του ρυθμού εκμάθησης εφαρμόστηκε η συνάρτηση ReduceLROnPlateau. Η συνάρτηση αυτή μειώνει τον ρυθμό εκμάθησης όταν οι τιμές μιας ορισμένης μετρικής δεν βελτιωθούν για ένα ορισμένο διάστημα (patience). Στην παρούσα εργασία η μετρική που ελέγχεται είναι η απώλεια του συνόλου επιβεβαίωσης (validation loss). Συγκεκριμένα ο ρυθμός εκμάθησης μειώνεται κατά 0,5 αν η τιμή της απώλειας δεν βελτιωθεί σε πάροδο τριών εποχών.

3.4 Συνελικτικά Νευρωνικά Δίκτυα

Η ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων (Convolution Neural Networks – CNNs) βασίστηκε στην επιστήμη της νευρολογίας και συγκεκριμένα σε μια έκθεση των Hubel και Wiesel (1962) για τον τρόπο λειτουργίας των οπτικών υποδοχέων μιας γάτας [30].

Τα CNNs δημιουργήθηκαν με σκοπό την εξαγωγή χαρακτηριστικών από εικόνες και τελικά την ταξινόμησή τους. Αποτελούνται από ένα ή περισσότερα επίπεδα συνέλιξης, ακολουθούμενα συνήθως από επίπεδα υποδειγματοληψίας. Τα τελευταία επίπεδα του δικτύου συνήθως περιέχουν πλήρη συνδεδεμένα επίπεδα για την ταξινόμηση των εισόδων στις κλάσεις. Η περιγραφή των παραπάνω επιπέδων θα δοθεί σε επόμενες παραγράφους.

Η δομή των CNNs δεν μοιάζει με αυτή των Πλήρως Συνδεδεμένων Δικτύων, στα οποία οι νευρώνες ενός επιπέδου είναι συνδεδεμένοι με όλους τους νευρώνες του προηγούμενου και του επόμενου επιπέδου. Συγκεκριμένα, ένα CNN περιέχει σταθερά φίλτρα διαφόρων διαστάσεων ανεξάρτητα από τις διαστάσεις των εικόνων εισόδου. Για παράδειγμα, ένα πλήρως συνδεδεμένο δίκτυο για μία μικρή εικόνα 100x100 απαιτεί 10.000 παραμέτρους για κάθε νευρώνα στο δεύτερο επίπεδο. Αντίθετα με τη χρήση CNNs ο αριθμός των παραμέτρων των φίλτρων είναι ανεξάρτητος του μεγέθους της εικόνας εισόδου και επιλέγεται κατά τον ορισμό της αρχιτεκτονικής του μοντέλου.



Σχήμα 3. 5: Τυπική δομή ενός Συνελικτικού Νευρωνικού Δικτύου

Όπως φαίνεται στο Σχήμα 3.5 ένα τυπικό CNN περιέχει δύο στάδια: το στάδιο της εκμάθησης χαρακτηριστικών (feature learning) και το στάδιο της ταξινόμησης (classification).

Οι εικόνες εισέρχονται στο δίκτυο και μέσω των φίλτρων μαθαίνονται τα χαρακτηριστικά των εικόνων (feature learning). Συγκεκριμένα, το CNN στα πρώτα επίπεδα εκμάθησης μαθαίνει βασικά σχήματα και ακμές. Όσο προχωράει σε βάθος τα φίλτρα προσαρμόζονται και εξάγουν πιο σύνθετα χαρακτηριστικά, αναγνωρίζοντας πιο σύνθετες δομές. Τέλος τα χαρακτηριστικά που εξάχθηκαν εισέρχονται στον ταξινομητή (classification) και βάση αυτών των χαρακτηριστικών η είσοδος αντιστοιχίζεται σε μια ή παραπάνω κλάσεις. Παρακάτω αναλύονται τα δομικά στοιχεία των επιπέδων.

Συνελικτικό επίπεδο

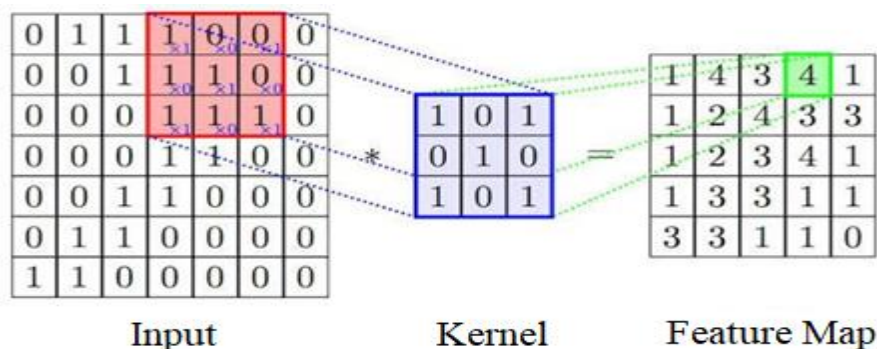
Το συνελικτικό επίπεδο περιέχει φίλτρα (kernel) τα οποία αρχικοποιούνται κατά τον ορισμό της αρχιτεκτονικής του μοντέλου και μαθαίνονται κατά τη διάρκεια της εκπαίδευσης. Κατά την εκπαίδευση τα φίλτρα εφαρμόζονται στην είσοδο του συνελικτικού επιπέδου και από τη συνέλιξη της εισόδου και του φίλτρου παράγονται οι χάρτες χαρακτηριστικών (feature maps), οι οποίοι περιέχουν τις θέσεις ενεργοποίησης των φίλτρων.

Πιο αναλυτικά η διαδικασία της συνέλιξης μπορεί να περιγραφεί ως εξής: Για ένα συγκεκριμένο εικονοστοιχείο της εισόδου του συνελικτικού επιπέδου υπολογίζεται το γινόμενο (στοιχείο προς στοιχείο) του φίλτρου με ένα παράθυρο της εισόδου ίδιων διαστάσεων με αυτές του φίλτρου και με κέντρο παραθύρου για την είσοδο το συγκεκριμένο εικονοστοιχείο (Σχήμα 3.6). Για να ολοκληρωθεί η διαδικασία της συνέλιξης πάνω στην είσοδο το φίλτρο θα περάσει από όλα τα εικονοστοιχεία της εικόνας.

Οι διαστάσεις των φίλτρων (ύψος και πλάτος), πρέπει να είναι μικρότερες από αυτές της εικόνας. Γενικά η έξοδος του συνελικτικού επιπέδου είναι της μορφής:

$$\alpha_l(i, j) = \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C W_l(h, w, c) F(i - h, j - w, c) + b_l(c) \quad (3.7)$$

, όπου $\alpha_l(i, j)$ είναι ένα εικονοστοιχείο του χάρτη ενεργοποίησης, W_l είναι το φίλτρο (kernel) με μέγεθος $H \times W \times C$ και b_l το κατώφλι.



Σχήμα 3. 6: Συνέλιξη εισόδου με φίλτρο (kernel).

Οι διαστάσεις της εξόδου του συνελικτικού επιπέδου εξαρτώνται από τρεις παραμέτρους: το βάθος, το βήμα και το γέμισμα περιθωρίου. Αναλυτικά:

- Το βάθος (depth) δηλώνει τον αριθμό και το μέγεθος των φίλτρων (kernel) που χρησιμοποιούνται σε κάθε επίπεδο. Αυξάνοντας το πλήθος των φίλτρων αυξάνεται και ο αριθμός των νευρώνων που επιπέδου αυτού. Το μέγεθος (ύψος και πλάτος) του φίλτρου ορίζει τον αριθμό των βαρών (weights) ανά φίλτρο. Τα μικρού μεγέθους φίλτρα συγκεντρώνουν μεγάλο όγκο τοπικής πληροφορίας, έχοντας την ικανότητα να αναγνωρίζουν λεπτομέρειες κάποιου μέρους της εικόνας. Παράλληλα τα μεγάλου μεγέθους φίλτρα αποτυπώνουν μια πιο γενική άποψη της εικόνας. Ο χαρακτηρισμός μικρό και μεγάλο φίλτρο γίνεται έχοντας ως σημείο αναφοράς το μέγεθος της εισόδου. Τέλος, το μέγεθος του φίλτρου είναι τετραγωνικό με περιττό αριθμό μήκους και πλάτους και συνήθεις τιμές 3x3, 5x5, 7x7.
- Ο βηματισμός (strides) ορίζει τον τρόπο με τον οποίο το φίλτρο ολισθαίνει κατά μήκος και πλάτος της εισόδου. Η τιμή του βήματος ορίζει τα εικονοστοιχεία που θα προσπεράσει το φίλτρο κατά τη συνέλιξή του με την είσοδο. Στην περίπτωση που το βήμα είναι (1, 1), το φίλτρο ολισθαίνει ομαλά, χωρίς να προσπεράσει κάποιο εικονοστοιχείο. Σε κάθε άλλη περίπτωση ο βηματισμός προκαλεί μείωση των εικονοστοιχείων της εισόδου.
- Το γέμισμα περιθωρίου (padding). Γενικά η συνέλιξη προκαλεί μείωση των διαστάσεων της εισόδου, λόγω των εικονοστοιχείων των άκρων, καθώς δεν υπάρχουν οι αντίστοιχοι γείτονες για να πολλαπλασιαστεί το φίλτρο. Στο γέμισμα περιθωρίου υπάρχουν δύο επιλογές: έγκυρο (valid) και όμοιο (same). Η επιλογή valid, έχει ως αποτέλεσμα το ύψος και το πλάτος της εισόδου να μειωθεί ανάλογα με το ύψος και το πλάτος του φίλτρου συνέλιξης. Στην επιλογή same, εικονικά μηδενικά εικονοστοιχεία προστίθενται γύρω από την εικόνα, έτσι δεν μειώνονται οι διαστάσεις της εισόδου.

Έστω μία εικόνα $M \times M \times C$, όπου $M \times M$ το ύψος και πλάτος της εικόνας και C ο αριθμός των καναλιών της, η συνέλιξη με N_f φίλτρα $J \times K$ διάστασης, βηματισμό $S \times S$ θα δώσει ως αποτέλεσμα χάρτες χαρακτηριστικών διαστάσεων

$$\left(\frac{M - J + 2P}{S} + 1 \right) \times \left(\frac{M - K + 2P}{S} + 1 \right) \times N_f \quad (3.8)$$

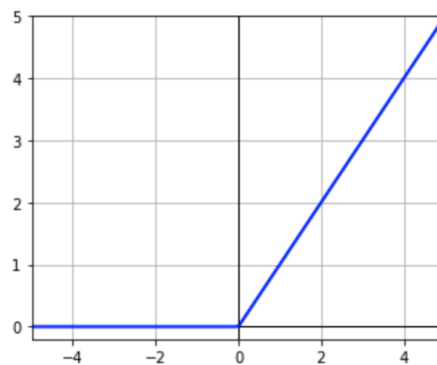
, όπου για padding = valid $\rightarrow P = 0$ και για padding = same $\rightarrow P = 1$.

Για το παραπάνω φίλτρο ο αριθμό των παραμέτρων εκμάθησης (trainable parameters), δηλαδή των παραμέτρων που θα εκπαιδευτούν κατά την διάρκεια της εκμάθησης, δίνεται από τον τύπο:

$$parameters = (J * K * C + 1) * N_f \quad (3.9)$$

Συνάρτηση Ενεργοποίησης

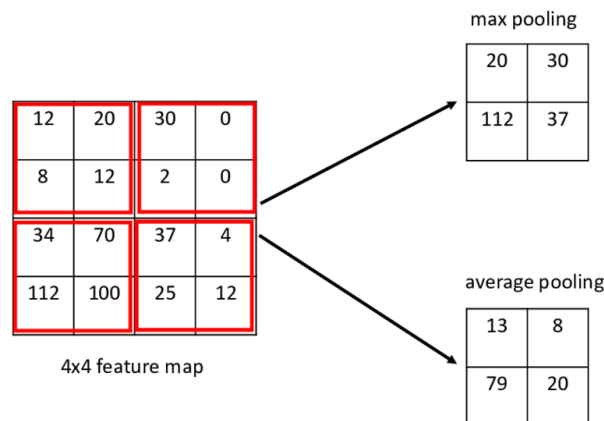
Η συνάρτηση ενεργοποίησης (activation function) εφαρμόζεται στο συνελκτικό επίπεδο, εισάγοντας μη γραμμικότητα. Σκοπός της είναι η ενεργοποίηση ή απενεργοποίηση νευρώνων. Στις εφαρμογές ταξινόμησης εικόνων συνήθως επιλέγεται η συνάρτηση ανόρθωσης – ReLU (Σχήμα 3.7).



Σχήμα 3. 7: Συνάρτηση Ανόρθωσης (ReLU).

Επίπεδο Υποδειγματοληψίας

Το επίπεδο υποδειγματοληψίας (Pooling Layer) συνήθως εφαρμόζεται μετά από ένα συνελκτικό επίπεδο. Κύρια λειτουργία του είναι η μείωση των χωρικών διαστάσεων των feature maps. Ουσιαστικά μέρος της πληροφορίας χάνεται, καθώς συνδεδεμένες ομάδες εικονοστοιχείων μετατρέπονται σε μοναδικές τιμές που εκφράζουν τις ομάδες αυτές. Συγκεκριμένα, χωρίζοντας την είσοδο σε L παράθυρα διαστάσεων $N \times N$, υπολογίζεται μία μοναδική τιμή για κάθε παράθυρο σύμφωνα με τη μέθοδο pooling που επιλέγεται.



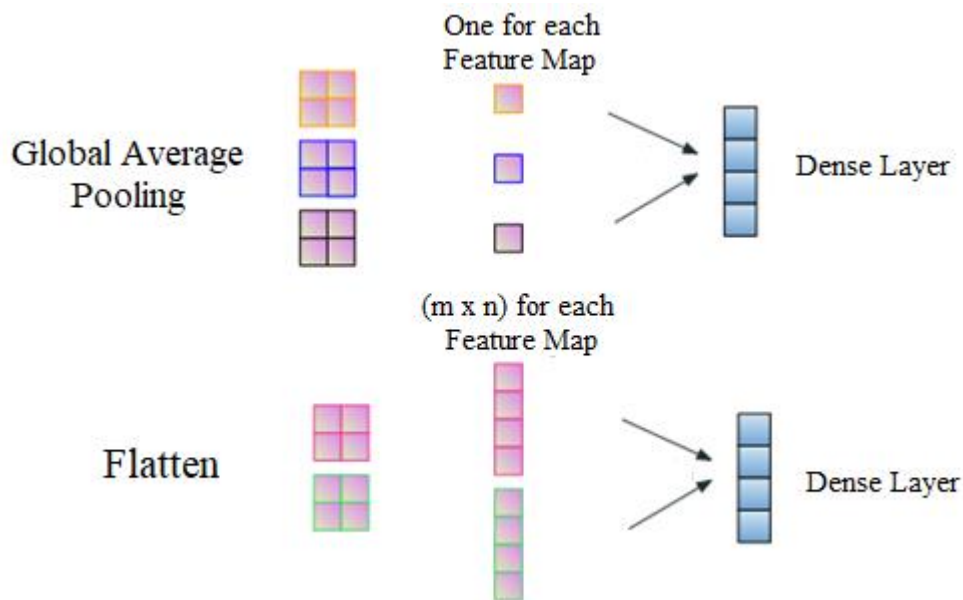
Σχήμα 3. 8: Υποδειγματοληψία μέγιστης (επάνω) και μέσης (κάτω) τιμής.

Τρεις είναι οι κύριες μέθοδοι υποδειγματοληψίας: η υποδειγματοληψία μεγίστου (Max Pooling), η υποδειγματοληψία μέσης τιμής (Average Pooling) και η καθολική υποδειγματοληψία μέσου όρου (Global Average Pooling). Αναλυτικά:

- **Max Pooling.** Η συνάρτηση αυτή επιλέγει τη μέγιστη τιμή για κάθε ένα από τα L παράθυρα της εισόδου (Σχήμα 3.8). Χρησιμοποιείται κυρίως στα πρώτα επίπεδα του συνελκτικού επιπέδου, καθώς απορρίπτει τα αδύναμα εικονοστοιχεία ενώ διατηρεί τα πιο ισχυρά, τα οποία συνήθως αναπαριστούν γωνίες και ακμές.

- Average Pooling. Η συνάρτηση αυτή υπολογίζει τον μέσο όρο των τιμών όλων των εικονοστοιχείων για κάθε ένα από τα L παράθυρα της εισόδου (Σχήμα 3.8). Έτσι, στην έξοδο συνεισφέρουν όλα τα εικονοστοιχεία και έχει ως αποτέλεσμα μια ομαλοποιημένη μορφή της εισόδου.
- Global Average Pooling. Αποτελεί ειδική περίπτωση υποδειγματοληψίας καθώς ως παράθυρο χρησιμοποιείται όλη η είσοδος (Σχήμα 3.9). Μια είσοδος με διαστάσεις $M \times M \times C$, μετατρέπεται σε ένα διάνυσμα $1 \times C$. Συνήθως χρησιμοποιείται μετά το τελευταίο συνελκτικό επίπεδο πριν τη χρήση ενός πλήρως συνδεδεμένου επιπέδου.

Η χρήση pooling επιπέδου μειώνει την πιθανότητα υπερεκπαίδευσης του μοντέλου, καθώς ένα κομμάτι της εισόδου χάνεται, προσφέροντας γενίκευση στο μοντέλο. Υπερεκπαίδευση συμβαίνει όταν κατά την διάρκεια της εκπαίδευσης το μοντέλο “αποστηθίζει” το σύνολο εκπαίδευσης με αποτέλεσμα να παρουσιάζει μειωμένη ικανότητα γενίκευσης. Συνήθως προκαλείται όταν χρησιμοποιούνται βαθιά μοντέλα για πιο απλά προβλήματα ή όταν το σύνολο εκπαίδευσης είναι πολύ μικρό για ένα δεδομένο πρόβλημα.



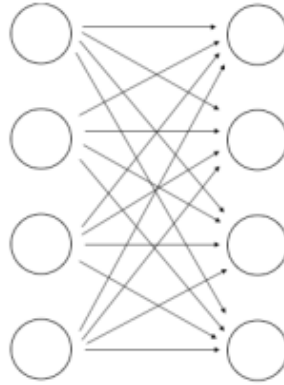
Σχήμα 3. 9: Global Average Pooling επίπεδο (πάνω) και Flatten επίπεδο (κάτω).

Επίπεδο Εξισορρόπησης

Μετά το τελευταίο συνελκτικό επίπεδο τα δεδομένα εισέρχονται στον ταξινομητή. Για να είναι δυνατή η είσοδος των δεδομένων στον ταξινομητή εφαρμόζεται είτε επίπεδο global average pooling είτε επίπεδο εξισορρόπησης (Flatten Layer). Το flatten επίπεδο μετατρέπει την είσοδό του, διάστασης $M \times M \times C$, σε διάνυσμα διάστασης $1 \times M * M * C$ (Σχήμα 3.9). Συνήθως χρησιμοποιείται όταν το μέγεθος της εισόδου του είναι σχετικά μικρό, καθώς ο αριθμός των παραμέτρων ισούται με το γινόμενο τους. Αντίθετα η συνάρτηση global average pooling οδηγεί σε διάνυσμα διάστασης C .

Πλήρως Συνδεδεμένο Επίπεδο

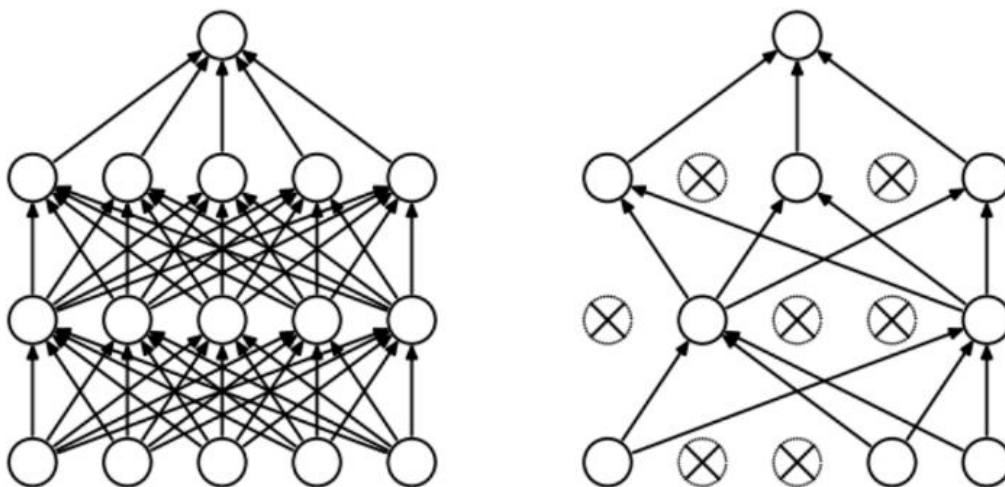
Το πλήρως συνδεδεμένο επίπεδο (Fully Connected Layer) αποτελείται από νευρώνες, οι οποίοι συνδέονται με όλους τους νευρώνες του προηγούμενου και του επόμενου επιπέδου. Λόγω της πλήρους σύνδεσης των νευρώνων απαιτείται μεγάλος αριθμός παραμέτρων. Το τελευταίο επίπεδο περιέχει τόσους νευρώνες όσες και οι κλάσεις του εκάστοτε προβλήματος ταξινόμησης.



Σχήμα 3. 10: Πλήρως συνδεδεμένο επίπεδο.

Επίπεδο Απόσυρσης (Dropout)

Στο επίπεδο της απόσυρσης αγνοείται τυχαία ένα ποσοστό νευρώνων. Οι νευρώνες που αγνοούνται παράγουν μηδενική έξοδο και οι παράμετροι τους δεν ενημερώνονται για την τρέχουσα εποχή για το τρέχον δείγμα. Η απενεργοποίηση τυχαίων νευρώνων προκαλεί μείωση της πληροφορίας εισόδου, με αποτέλεσμα τη γενίκευση του μοντέλου, ενώ παράλληλα η εκπαίδευση γίνεται λιγότερη ευαίσθητη στην υπερκεπαίδευση.



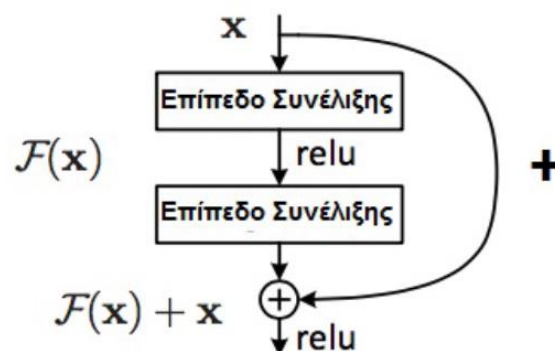
Σχήμα 3. 11: Πλήρως συνδεδεμένο δίκτυο χωρίς απόσυρση (αριστερά) και με απόσυρση (δεξιά).

3.5 Υπολειμματικά Νευρωνικά Δίκτυα

Στα τέλη του 2015, οι Kaiming He et al [31] παρουσίασαν μια νέα αρχιτεκτονική βαθιάς μάθησης. Σκοπός τους ήταν η κατασκευή ενός βαθιού νευρωνικού δικτύου, η απόδοση του οποίου δεν θα υποβαθμιζόταν με την αύξηση των κρυφών επιπέδων και δεν θα αντιμετώπιζε το πρόβλημα των εξαφανιζόμενων κλίσεων. Το δίκτυο αυτό ονομάστηκε Υπολειμματικό Νευρωνικό Δίκτυο (Residual Neural Networks - ResNet).

Στα CNNs δίκτυα η μετάδοση της πληροφορίας είναι ακολουθιακή (sequential), δηλαδή η έξοδος ενός επιπέδου συνδέεται με την είσοδο του ακριβώς επόμενου επιπέδου.

Αντίθετα στα ResNet δίκτυα η πληροφορία έχει τη δυνατότητα να παρακάμψει διασυνδέσεις (skip connection), δηλαδή η είσοδος ενός επιπέδου συνδέεται με την έξοδο του ακριβώς προηγούμενου επιπέδου και την είσοδο n-φορές προηγούμενου επιπέδου. Η παράκαμψη διασυνδέσεων, υλοποιείται με την χρήση υπολειμματικών block (residual block) (Σχήμα 3.12). Η έξοδος ενός block τροφοδοτείται ως είσοδος στο επόμενο residual block, κ.ο.κ. Πολλαπλά residual blocks συνδεδεμένα γραμμικά αποτελούν ένα ResNet δίκτυο.



Σχήμα 3. 12: Αρχιτεκτονική residual block.

Αναλυτικά, κάθε block αποτελείται από διαδοχικά συνελκτικά επίπεδα, συνήθως δύο ή τρία. Η είσοδος x του block αποθηκεύεται σε έναν τοπικό buffer και αφού περάσει από τα διαδοχικά συνελκτικά επίπεδα προστίθεται στην έξοδο τους.

Ουσιαστικά σε κάθε residual block το δίκτυο μαθαίνει αν η είσοδος πρέπει να διατηρηθεί (skip connection) ή να περάσει μέσα από τα συνελκτικά επίπεδα του block. Με αυτόν τον τρόπο επιτρέπεται η χρήση πολλών συνελκτικών επιπέδων, χωρίς τη μείωση της απόδοσης του δικτύου. Επίσης με τη παράκαμψη διασυνδέσεων αντιμετωπίζεται και το πρόβλημα των εξαφανιζόμενων κλίσεων, αφού η παράκαμψη εφαρμόζεται και στη διαδικασία της οπισθοδιάδοσης.

3.5.1 ResNet-50

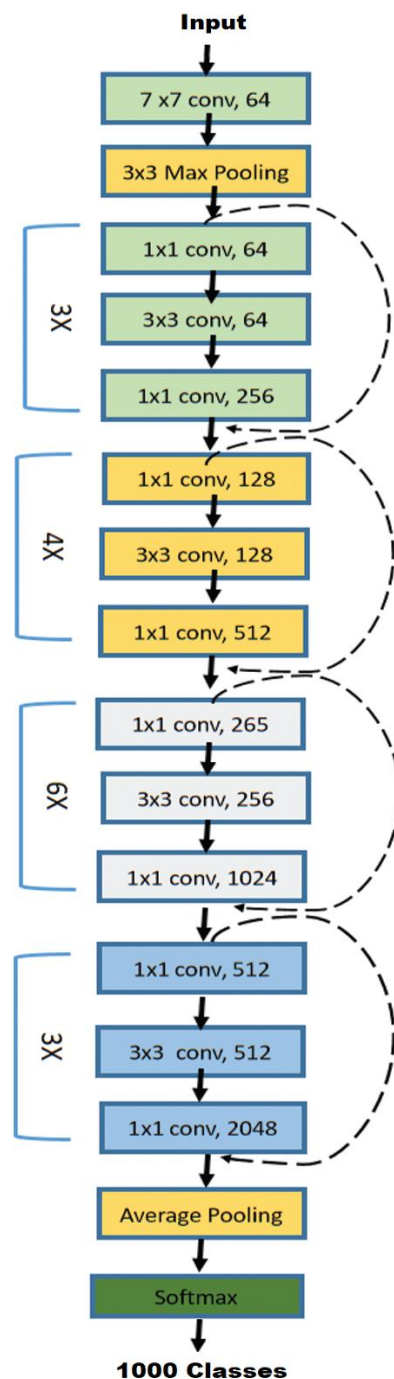
Στο paper των Kaiming He et al [31], παρουσιάστηκαν πέντε ResNet δίκτυα με διαφορετικό αριθμό επιπέδων και διαφορετική αρχιτεκτονική των blocks, τα ResNet-18, ResNet-34 ResNet-50 και ResNet-152. Τα δίκτυα αυτά εκπαιδεύτηκαν στο σύνολο δεδομένων του ImageNet, που περιέχει περίπου 1,2 εκατομμύριο RGB εικόνες διαφόρων διαστάσεων.

Η αρχιτεκτονική ενός ResNet δικτύου δομείται ως εξής:

1. Feature extractor: Η είσοδος περνάει από ένα συνελικτικό επίπεδο με kernel $64 \times 7 \times 7$ και στην συνέχεια από ένα επίπεδο μέγιστης υποδειγματοληψίας με kernel 2×2 και βήμα 2. Ακολουθούν τα residual blocks, που αποτελούνται από συνελικτικά επίπεδα με kernel 1×1 ή 3×3 και η αριθμό φίλτρων. Ο αριθμός των blocks και των συνελικτικών επιπέδων ανά block εξαρτάται από το μοντέλο που χρησιμοποιείται.
2. Classifier: Στη συνέχεια εφαρμόζεται ένα επίπεδο μέσης καθολικής υποδειγματοληψίας, στην έξοδο του οποίου εφαρμόζεται συνάρτηση ενεργοποίησης Softmax που καταλήγει σε έξοδο 1×1000 .

Το ResNet-50 αποτελείται από 50 συνελικτικά επίπεδα και η αρχιτεκτονική του παρουσιάζεται στο Σχήμα 3.13.

Κάθε residual block διαθέτει τρία διαδοχικά συνελικτικά επίπεδα.



Σχήμα 3. 13: Αρχιτεκτονική ResNet-50.

3.6 Μεταφορά Μάθησης

Μεταφορά Μάθησης (Transfer Learning) [32], είναι η διαδικασία κατά την οποία ένα μοντέλο που έχει υλοποιηθεί για ένα πρόβλημα επαναχρησιμοποιείται για ένα διαφορετικό, αλλά παρόμοιο πρόβλημα.

Στη μεταφορά μάθησης μπορεί να μεταφερθεί η αρχιτεκτονική του μοντέλου, όπως επίσης και τα βάρη και τα κατώφλια του. Κατά τη μεταφορά μάθησης το προ εκπαιδευμένο δίκτυο μπορεί να χρησιμοποιηθεί:

- Ως ταξινομητής. Το προ εκπαιδευμένο μοντέλο χρησιμοποιείται αυτούσιο, χωρίς καμία αλλαγή στα διάφορα επίπεδα και χωρίς περεταίρω εκπαίδευση. Αυτή η χρήση προϋποθέτει ομοιότητες μεταξύ των συνόλων δεδομένων.
- Για ρύθμιση βαρών. Το προ εκπαιδευμένο μοντέλο εκπαιδεύεται ολόκληρο ή μερικώς. Πιο συγκεκριμένα υπάρχει η δυνατότητα παγώματος των βαρών κάποιων επιπέδων, με την ταυτόχρονη εκπαίδευση των βαρών των υπολοίπων επιπέδων. Συνήθως επιλέγεται να παγώσουν τα πρώτα επίπεδα ενός CNN, στα οποία εντοπίζονται τα γενικότερα χαρακτηριστικά μιας εικόνας, π.χ. ακμές και γωνίες, καθώς αποτελούν κοινά χαρακτηριστικά σε όλες τις εικόνες. Τα τελευταία πιο εξειδικευμένα επίπεδα εκπαιδεύονται περεταίρω.
- Για εξαγωγή χαρακτηριστικών. Από το προ εκπαιδευμένο μοντέλο αφαιρείται ο ταξινομητής και προσαρμόζεται στο τρέχον πρόβλημα με προσθήκη των κατάλληλων επιπέδων. Τα βάρη του σταδίου εκμάθησης χαρακτηριστικών κρατούνται παγωμένα.

3.7 Προγραμματιστικές πλατφόρμες και εργαλεία

Στην παρούσα διπλωματική έγινε χρήση της βιβλιοθήκης Keras πάνω σε TensorFlow. Συνδυαστικά χρησιμοποιήθηκαν διάφορες βιβλιοθήκες της Python. Ο κώδικας των πειραμάτων είναι γραμμένος σε Python και εκτελέστηκε στο περιβάλλον της διαδικτυακής πλατφόρμας Google Colaboratory Pro σε μορφή Jupyter Notebook. Ακολουθεί μια σύντομη περιγραφή για κάθε εργαλείο που χρησιμοποιήθηκε.

Το **Jupyter Notebook** είναι μία διαδικτυακή εφαρμογή ανοιχτού κώδικα και ελεύθερου λογισμικού για διαδραστικό προγραμματισμό σε πολλές γλώσσες προγραμματισμού. Επιτρέπει στους χρήστες να δημιουργήσουν και να μοιράσουν αρχεία που περιέχουν εκτελέσιμο κώδικα, εξισώσεις, γραφήματα, απεικονίσεις και επεξηγηματικό κείμενο.

Το **Google Colaboratory (Colab)** αποτελεί μια έκδοση του Jupiter Notebook. Είναι προϊόν της Google Research και προσφέρει δωρεάν στους χρήστες την δυνατότητα να γράψουν και να τρέξουν Python κώδικα. Είναι κατάλληλο για εφαρμογές Μηχανικής Μάθησης, Ανάλυσης Δεδομένων, για προσωπική χρήση και για εκπαιδευτικούς σκοπούς. Επίσης έχει μεγάλη ποικιλία προ-εγκατεστημένων βιβλιοθηκών. Η έκδοση που χρησιμοποιήθηκε είναι η Pro στην οποία προσφέρεται:

- Επεξεργαστής Intel® Xeon® CPU 2.30 GHz
- Μνήμη RAM 32 GB or 52 GB
- Κάρτα Γραφικών Tesla P100-PCIE-16GB

Το **TensorFlow** είναι μια δωρεάν, ανοιχτού κώδικα βιβλιοθήκη για την ανάπτυξη και την εκπαίδευση μοντέλων Μηχανικής Μάθησης. Διαθέτει ένα ολοκληρωμένο και ευέλικτο οικοσύστημα από εργαλεία, βιβλιοθήκες και πόρους που επιτρέπουν στους ερευνητές να υλοποιήσουν εφαρμογές Μηχανικής Μάθησης. Είναι φιλικό στον χρήστη και αποτελεί μια από τις πιο δημοφιλείς επιλογές για την ανάπτυξη νευρωνικών δικτύων από ερευνητές και βιομηχανίες. Το TensorFlow μπορεί να τρέξει σε πολλαπλές CPUs και GPUs ενώ υποστηρίζει 64-bit Linux, Mac OS, Windows, και πλατφόρμες κινητών όπως Android and iOS. Η έκδοση του TensorFlow που χρησιμοποιήθηκε είναι η TensorFlow 2.6.0.

Το **Keras** είναι μια υψηλού επιπέδου βιβλιοθήκη βαθιών νευρωνικών δικτύων γραμμένη σε Python. Αποτελεί μια διεπαφή προγραμματισμού εφαρμογών (API) υψηλού επιπέδου που τρέχει σε TensorFlow, Theano ή CNTK. Αναπτύχθηκε με σκοπό να δίνει στον χρήστη τη δυνατότητα να πειραματίζεται γρήγορα, με βάση τη φιλοσοφία: “Το να μπορείς να πηγαίνεις από την ιδέα στο αποτέλεσμα όσο πιο γρήγορα, γίνεται κλειδί για να κάνεις καλή έρευνα”. Περιέχει έτοιμες υλοποιήσεις από πολλά δομικά στοιχεία των νευρωνικών δικτύων. Η έκδοση του Keras που χρησιμοποιήθηκε είναι η Keras 2.6.0.

Τέλος χρησιμοποιήθηκαν οι βιβλιοθήκες:

- NumPy: Μαθηματική βιβλιοθήκη για την υλοποίηση σύνθετων μαθηματικών υπολογισμών μεταξύ πινάκων.
- Pandas: Βιβλιοθήκη για ανάλυση και επεξεργασία δεδομένων.
- Matplotlib: Βιβλιοθήκη γραφικών απεικονίσεων.
- Scikit-learn: Βιβλιοθήκη αλγορίθμων μηχανικής μάθησης και στατιστικής μοντελοποίησης, π.χ. προβλήματα παλινδρόμησης, ταξινόμησης.
- Seaborn: Βιβλιοθήκη γραφικών απεικονίσεων.

4. Σύνολα δεδομένων και προ-επεξεργασία εικόνων

4.1 Σύνολο δεδομένων Fruits-360

Το σύνολο δεδομένων (dataset) που χρησιμοποιήθηκε είναι το “Fruits-360 DataSet” και δημιουργήθηκε από τους Horea Muresan και Mihai Otlean το 2018 . Έκτοτε είναι διαθέσιμο στο Github και στο Kaggle [21]. Περιέχει εικόνες από 131 είδη φρούτων και λαχανικών. Οι εικόνες προέκυψαν από την παρακάτω διαδικασία.

Τα διάφορα οπωροκηπευτικά τοποθετήθηκαν σε άξονα κινητήρα που περιστρεφόταν με χαμηλή ταχύτητα, τρεις περιστροφές ανά λεπτό (3 rpm). Με χρήση κάμερας Logitech C920 κινηματογραφήθηκαν βίντεο διάρκειας 20 δευτερολέπτων [33], από τα οποία έγινε εξαγωγή των εικόνων (frames). Ως φόντο τοποθετήθηκε ένα λευκό φύλλο χαρτιού.

Παρόλα αυτά δεν υπήρχε ομοιομορφία στο φόντο των εικόνων, κυρίως λόγω των διαφορών στις συνθήκες φωτισμού. Για αυτόν τον λόγο οι δημιουργοί υλοποίησαν έναν αλγόριθμο τύπου γεμίσματος (flood fill), ο οποίος εξάγει το αντικείμενο της φωτογραφία από το φόντο.



Σχήμα 4. 1: Αριστερά: Αρχική εικόνα, όπου φαίνεται το φόντο και ο άξονας.
Δεξιά: Η εικόνα μετά την εφαρμογή αλγόριθμος flood-fill και αλλαγή διαστάσεων σε 100x100.

Ο flood fill αλγόριθμος καθορίζει μια περιοχή που αποτελείται από συνδεδεμένα εικονοστοιχεία, με στόχο τον διαχωρισμό της από την υπόλοιπη εικόνα βάση ενός χαρακτηριστικού. Δίνοντας ως δεδομένο ένα αρχικό σημείο στην εικόνα, ο αλγόριθμος εξετάζει τα γειτονικά εικονοστοιχεία προς όλες τις κατευθύνσεις ελέγχοντας αν υπάρχει σχέση μεταξύ τους με βάση το επιθυμητό χαρακτηριστικό. Η διαδικασία αυτή επαναλαμβάνεται έως όπου τα εναπομείναντα εικονοστοιχεία δεν παρουσιάζουν σύνδεση με τα προηγούμενα.

Για τον συγκεκριμένο αλγόριθμο, ως σημεία αναφοράς επιλέχθηκαν οι άκρες της εικόνας και ως χαρακτηριστικό διαχωρισμού το χρώμα. Προχωρώντας προς το εσωτερικό σημειώθηκαν όλα τα σημεία των οποίων η χρωματική απόσταση ήταν μικρότερη από ένα κατώφλι και ανήκαν στη γειτονιά των σημείων αναφοράς. Τα σημειωμένα εικονοστοιχεία θεωρήθηκαν μέρος του φόντου και χρωματίστηκαν λευκά, ενώ τα υπόλοιπα θεωρήθηκαν μέρος του αντικείμενου. Οι δημιουργοί σημειώνουν ότι για κάθε αντικείμενο η τιμή του κατωφλίου ορίστηκε με την μέθοδο δοκιμής-λάθους (trial-error) και είναι διαφορετική για κάθε αντικείμενο.

Η συγκεκριμένη επεξεργασία εισήγαγε θόρυβο στις εικόνες, λόγω της δυσκολίας να διακριθεί το αντικείμενο από το φόντο. Στο Σχήμα 4.2 φαίνονται δύο τέτοιες περιπτώσεις, που έχει κοπεί το κάτω άκρο του αντικειμένου και διακρίνεται ο άξονας περιστροφής στο κάτω άκρο του αντικειμένου.



Σχήμα 4. 2: Εικόνες του συνόλου δεδομένων, όπου έχει κοπεί το κάτω άκρο του αντικείμενου (αριστερά) και διακρίνεται ο άξονας περιστροφής στο κάτω άκρο του (δεξιά).

Τέλος, το μέγεθος των εικόνων μετατράπηκε 100x100. Όπως φαίνεται στο Σχήμα 4.3 δεν κρατήθηκαν οι σχετικές αναλογίες των αντικειμένων, αλλά μέσω μεγέθυνσης ή σμίκρυνσης προσαρμόστηκαν σε εικόνα διαστάσεων 100x100. Το μέγεθος αυτό κρίθηκε κατάλληλο ώστε να γίνει εφικτός ο διαχωρισμός των διάφορων κλάσεων.



Σχήμα 4. 3: Εικόνες συνόλου δεδομένων της κλάσης σταφύλι λευκό (αριστερά) και καρπούζι (δεξιά) μεγέθους 100x100.

Το Fruits-360 dataset αποτελείται από εικόνες οπωροκηπευτικών που ανήκουν σε 131 κλάσεις. Οι εικόνες είναι τοποθετημένες ανά κλάση σε φακέλους. Κάθε εικόνα περιέχει ένα μόνο αντικείμενο. Στις περιπτώσεις που οι δημιουργοί του dataset δεν είχαν την επιστημονική ονομασία του φρούτου ή του λαχανικού ως όνομα κλάσης δόθηκε η κοινή ονομασία (πχ. μήλο, δαμάσκηνο) και ένα ψηφίο. Παραδείγματος χάρη, στην περίπτωση του δαμάσκηνου υπάρχουν τρία είδη με άγνωστη επιστημονική ονομασία.



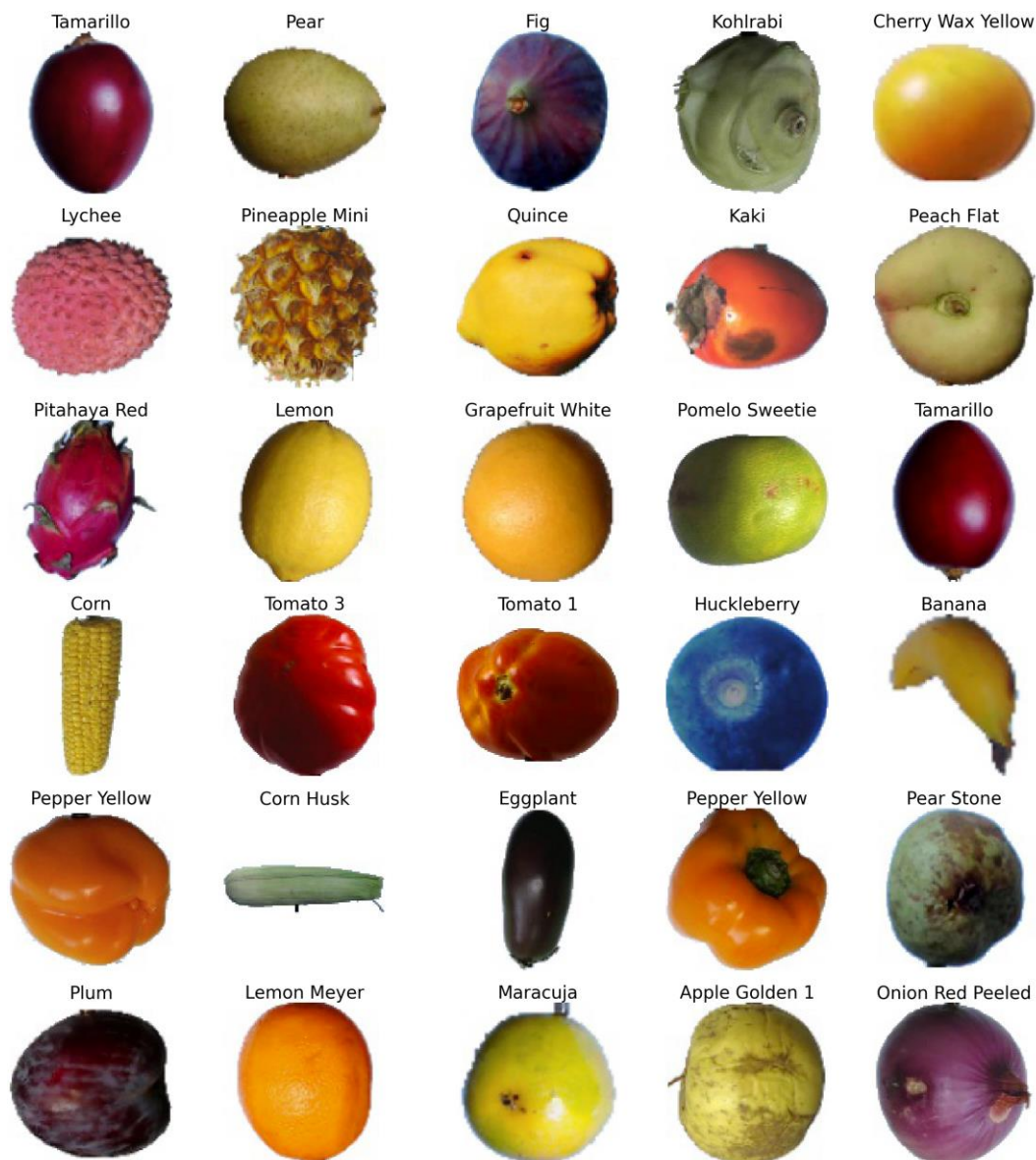
Σχήμα 4. 4: Εικόνες δαμάσκηνων διαφορετικού είδους, που ανήκουν στις κλάσεις Plum, Plum 2 και Plum 3.

Το σύνολο δεδομένων περιέχει **90.380 RGB εικόνες**, οι οποίες οργανώθηκαν σε δύο φακέλους, τον φάκελο Training και τον φάκελο Test. Στο Παράρτημα Α παρουσιάζεται ο αριθμός των δεδομένων ανά κλάση για κάθε ένα από τους δύο φακέλους.

Τέλος οι εικόνες χωρίστηκαν σε σύνολο εκπαίδευσης (training dataset), σύνολο επιβεβαίωσης (validation dataset) και σύνολο ελέγχου (test dataset). Στον παρακάτω πίνακα δίνεται ο αριθμός των εικόνων ανά σύνολο.

	Σύνολο εικόνων
Training Dataset	60955
Validation Dataset	6737
Test Dataset	22688
Total	90380

Πίνακας 1: Διαχωρισμός του συνόλου δεδομένων ανά υποσύνολο.



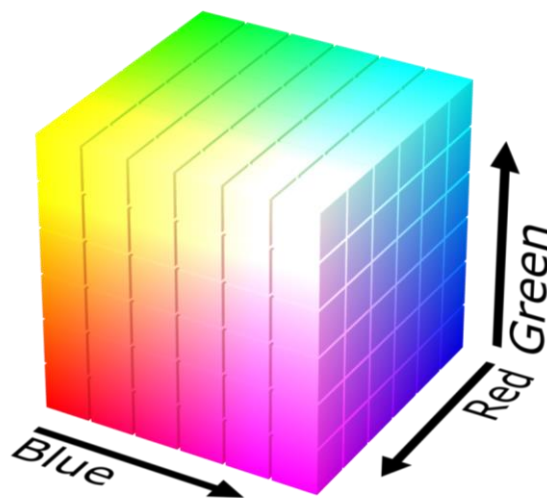
Σχήμα 4. 5: Δείγμα του συνόλου δεδομένων.

4.2 Επεξεργασία εικόνων

4.2.1 Χρωματικοί Χώροι

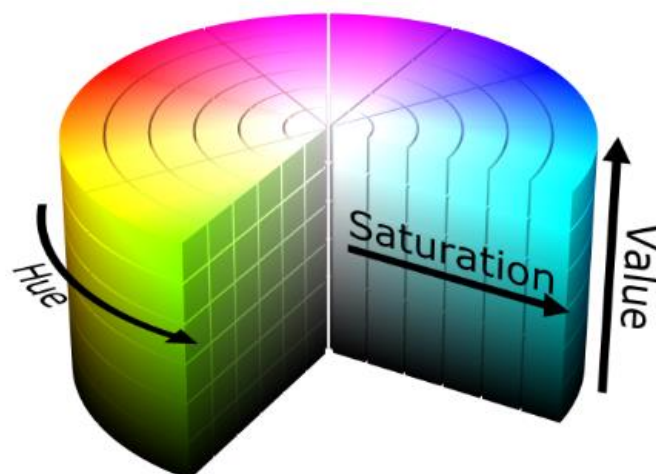
Οι εικόνες του Fruits-360 dataset είναι RGB εικόνες. Στην παρούσα εργασία έγιναν πειράματα στους χρωματικούς χώρους RGB, HSV και Grayscale και στους συνδυασμούς τους.

Ο χρωματικός χώρος RGB, ένας από τους πιο δημοφιλείς χρωματικούς χώρους, αποτελείται από 3 κανάλια χρώματος (Σχήμα 4.6). Κάθε ένα αντιπροσωπεύει ένα από τα βασικά χρώματα, R για το κόκκινο (Red), G για το πράσινο (Green) και B για το μπλε (Blue). Αναπαρίσταται με 24bit, δηλαδή σε κάθε κανάλι δίνονται 8bit με εύρος τιμών $[0, 255]$ με το 0 να αντιπροσωπεύει την απουσία φωτός και το 255 την μέγιστη ένταση φωτός. Η δημιουργία του χρωματικού χώρου RGB βασίστηκε στο γεγονός ότι όλα τα χρώματα μπορούν να αναπαρασταθούν χρησιμοποιώντας διαφορετικές αποχρώσεις του κόκκινου, πράσινου και μπλε.



Σχήμα 4. 6: Χρωματικός χώρος RGB.

Ο χρωματικός χώρος HSV αποτελεί εναλλακτική αναπαράσταση του χρωματικού χώρου RGB και σχεδιάστηκε για να αναπαραστήσει καλύτερα τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται τα χρώματα. Αποτελείται από 3 κανάλια, κάθε ένα αντιπροσωπεύει ένα χαρακτηριστικό του χρώματος, H για την απόχρωση (Hue), S για τον κορεσμό (Saturation) και V για την τιμή της φωτεινότητας (Value).



Σχήμα 4. 7: Χρωματικός χώρος HSV.

Η αναπαράσταση του HSV χρωματικού χώρου έχει κυλινδρική γεωμετρία (Σχήμα 4.7). Τα χρώματα περιγράφονται μέσω του καναλιού της απόχρωσης (Hue). Το κανάλι της απόχρωσης αποτελεί τη γωνιακή συντεταγμένη. Ξεκινώντας από το κόκκινο στις 0° περνάει στο κίτρινο, το πράσινο, το γαλάζιο, το μπλε και το μωβ και καταλήγει μετά από πλήρη περιστροφή ξανά στο κόκκινο. Το κανάλι κορεσμού αποτελεί την ακτινική διάσταση και περιγράφει την ποσότητα γκρι σε ένα συγκεκριμένο χρώμα με τιμές στο [0,1]. Όσο πιο κοντά στο 0, τόσο πιο μεγάλη ποσότητα γκρι περιέχεται στο χρώμα. Το κανάλι φωτεινότητας αποτελεί τη διάσταση του ύψους και περιγράφει την ένταση του χρώματος σε κλίμακα [0-100]. Στο 0 αντιστοιχίζεται το μαύρο και στο 100 έχουμε τη μεγαλύτερη ένταση του χρώματος.

Η μετατροπή της εικόνας από RGB σε HSV έγινε χρησιμοποιώντας τη συνάρτηση του TensorFlow, `tf.image.rgb_to_hsv` η έξοδος της οποίας δίνει τιμές στο [0,1] για κάθε κανάλι.

Ο χρωματικός χώρος Grayscale αποτελεί μια ασπρόμαυρη απεικόνιση και η εικόνα αποτελείται από διάφορες διαβαθμίσεις του γκρι (Σχήμα 4.8). Δεν μεταφέρει καμία χρωματική πληροφορία αλλά απεικονίζει την ένταση της φωτεινότητας σε κάθε σημείο. Κάθε εικονοστοιχείο αναπαρίσταται με 8bit παρέχοντας 256 αποχρώσεις του γκρι. Τα εικονοστοιχεία με τιμή 0 δηλώνουν απουσία φωτός (μαύρο), ενώ τα εικονοστοιχεία με τιμή 1 δηλώνουν μεγάλη ένταση φωτεινότητας (λευκό). Η μετατροπή της εικόνας από RGB σε Grayscale έγινε χρησιμοποιώντας τη συνάρτηση του TensorFlow `tf.image.rgb_to_grayscale`.



Σχήμα 4. 8: Χρωματικός χώρος Grayscale 8-bit.

4.2.2 Τεχνικές Αύξησης Δεδομένων

Για αποφυγή της υπερεκπαίδευσης και αύξηση της ακρίβειας του μοντέλου εφαρμόστηκαν τεχνικές αύξησης δεδομένων. Σε κάθε εποχή οι εικόνες μετασχηματίζονταν, έτσι το μοντέλο εκπαιδεύονταν με τυχαίο τρόπο είτε στις αρχικές εικόνες είτε στις επεξεργασμένες. Οι τεχνικές αυτές δεν αύξησαν τον αριθμό των δεδομένων του dataset, καθώς ο αριθμός των εικόνων παρέμεινε σταθερός. Οι αλλαγές αυτές εφαρμόστηκαν σε όλα το σύνολο δεδομένων μέσω της συνάρτησης επεξεργασία (augmentation Function).

Στο Keras, εκτός των έτοιμων συναρτήσεων για μετασχηματισμό των δεδομένων, δίνεται η δυνατότητα ορισμού μιας συνάρτησης επεξεργασίας που θα καλείται για κάθε εικόνα κάθε εποχή και μπορεί να περιέχει οποιοδήποτε κώδικα. Η συνάρτηση επεξεργασίας είναι ένα από τα πιο σημαντικά χαρακτηριστικά που προσφέρει το Keras για αύξηση δεδομένων, καθώς δίνει τη δυνατότητα πειραματισμού με διάφορους μετασχηματισμούς (αλλαγή χρωματικού χώρου, εξισορρόπηση χρώματος) χωρίς να απαιτείται ο μετασχηματισμός των εικόνων και η αποθήκευσή τους πριν την έναρξη της εκπαίδευσης.

Στην παρούσα διπλωματική χρησιμοποιήθηκε η βιβλιοθήκη `Preprocessing.Image` και συγκεκριμένα η συνάρτηση `ImageDataGenerator()`. Η `ImageDataGenerator()` πραγματοποιεί μετασχηματισμούς σε πραγματικό χρόνο. Αρχικά, στα δεδομένα με μορφή 100x100x3 RGB, εφαρμόστηκε τυχαία κάθετη και οριζόντια αναστροφή. Επίσης, εφαρμόστηκε η συνάρτηση τυχαίας προσαρμογής της Απόχρωσης και του Κορεσμού. Παρακάτω εξηγούνται αναλυτικά οι δύο τεχνικές.

Οριζόντια και Κάθετη αναστροφή

Επιλέχθηκε η χρήση αύξησης δεδομένων με οριζόντια και κάθετη αναστροφή ως προς το κέντρο της εικόνας. Κατά την δημιουργία του dataset πάρθηκαν φωτογραφίες από διάφορες οπτικές γωνίες, παρόλα αυτά ένα αντικείμενο μπορεί να έχει οποιοδήποτε προσανατολισμό στην εικόνα. Η οριζόντια αναστροφή αντιστρέφει το ενεργό πεδίο οριζόντια, δηλαδή από αριστερά προς τα δεξιά. Η κάθετη αναστροφή αντιστρέφει το ενεργό επίπεδο κάθετα, δηλαδή από κάτω προς τα πάνω.

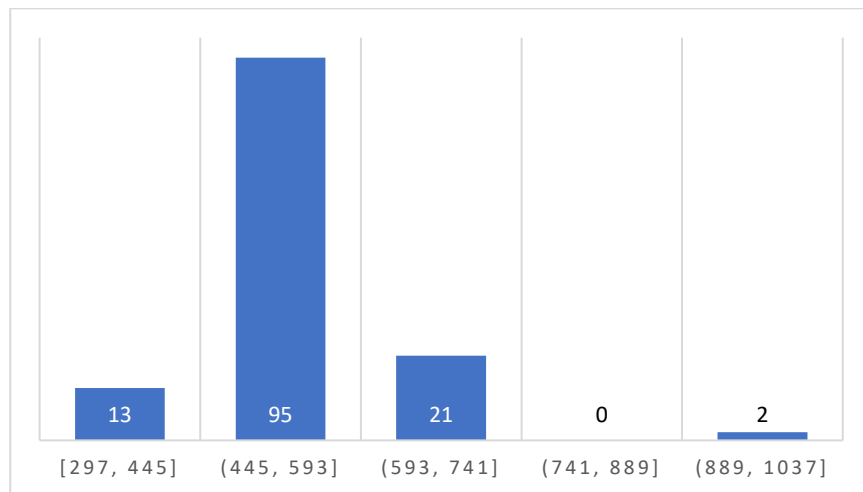
Αύξηση Απόχρωσης και Κορεσμού

Στην προσαρμογή της απόχρωσης η συνάρτηση που χρησιμοποιήθηκε λαμβάνει ως είσοδο μια RGB εικόνα, τη μετατρέπει σε float μορφή και στη συνέχεια στον χρωματικό χώρο HSV. Στο κανάλι της απόχρωσης προστίθεται σε όλα τα εικονοστοιχεία μιας εικόνας ένα τυχαίος αριθμός στο διάστημα $[-0,02, +0,02]$. Αφού εφαρμοστούν οι παραπάνω αλλαγές, επιστρέφεται η εικόνα σε RGB.

Για την προσαρμογή του κορεσμού η συνάρτηση που χρησιμοποιήθηκε λαμβάνει ως είσοδο μια RGB εικόνα, τη μετατρέπει σε float και στη συνέχεια στον χρωματικό χώρο HSV. Στο κανάλι του κορεσμού προστίθεται σε όλα τα εικονοστοιχεία μιας εικόνας ένα τυχαίος αριθμός στο διάστημα $[0.9, 1.2]$. Αφού εφαρμοστούν οι παραπάνω αλλαγές, επιστρέφεται η εικόνα σε RGB.

4.3 Ασύμμετρη κατανομή κλάσεων

Εξετάζοντας το σύνολο δεδομένων εκπαίδευσης του Fruits-360 υπολογίστηκε ότι κάθε κλάση διαθέτει κατά μέσο όρο 520 δείγματα. Στο παρακάτω ιστόγραμμα (Σχήμα 4.9) απεικονίζεται η κατανομή των κλάσεων με βάση τον αριθμό των δειγμάτων τους. Ο αριθμός δειγμάτων των περισσότερων κλάσεων ανήκει στο διάστημα (444, 593]. Παρόλα αυτά φαίνεται ότι υπάρχουν με μεγάλη διαφορά δειγμάτων. Για παράδειγμα η κλάση Μπλε Σταφύλι έχει 984 δείγματα, ενώ η κλάση ρίζα τζίντζερ έχει μόλις 297.



Σχήμα 4. 9: Ιστόγραμμα κατανομής των κλάσεων με βάση τον αριθμό δειγμάτων ανά κλάση.

Στο πρόβλημα της ταξινόμησης η ασύμμετρη κατανομή των κλάσεων κάνει τον ταξινομητή μεροληπτικό απέναντι στις κλάσεις πλειοψηφίας, λαμβάνοντας λιγότερο υπόψη τις μειοψηφικές κλάσεις. Για να αποφευχθεί αυτό το πρόβλημα, εφαρμόζονται μέθοδοι ισορρόπησης.

Μια τέτοια μέθοδος είναι η εξισορρόπηση δεδομένων με βάρη. Η μέθοδος αυτή υπολογίζει τα βάρη κάθε κλάσης και κατά τη διάρκεια της εκπαίδευσης δίνει βαρύτητα ανάλογη του αριθμού των δειγμάτων της συγκεκριμένης κλάσης. Τα βάρη κάθε κλάσης υπολογίζονται με τον παρακάτω τύπο:

$$w_j = \frac{n_{samples}}{n + classes * n_{samples_j}} \quad (4.1)$$

, όπου w_j το βάρος της κλάσης j , $n_{samples}$ ο συνολικός αριθμός δειγμάτων του συνόλου εκπαίδευσης, $n_{classes}$ ο αριθμός των κλάσεων του προβλήματος και $n_{samples_j}$ ο συνολικός αριθμός των δειγμάτων της κλάσης j .

Η συνάρτηση με την οποία πραγματοποιήθηκε η ισορρόπηση των δεδομένων, είναι η `class_weight.compute_class_weight(class_weight,*,classes,y)` και ανήκει στη βιβλιοθήκη `sklearn`.

Σε αυτό το σημείο να σημειωθεί ότι το σύνολο δεδομένων δεν είναι μη-ισορροπημένο. Ένα σύνολο δεδομένων καλείται μη-ισορροπημένο (*imbalanced dataset*) όταν ο αριθμός των δειγμάτων σε μια κλάση είναι σημαντικά μεγαλύτερος από αυτόν μιας άλλης κλάσης [34]. Παρόλα αυτά με σκοπό της μελέτης επιρροής της άνισης κατανομής στην ακρίβεια του ταξινομητή πραγματοποιήθηκε το πείραμα που θα περιγραφεί στην συνέχεια, Κεφ. 5.3.

5. Πειραματική Διαδικασία

5.1 Αρχιτεκτονική Μοντέλων

5.1.1. Προτεινόμενο CNN

Το προτεινόμενο συνελκτικό δίκτυο αποτελείται από τέσσερα συνελκτικά επίπεδα, κάθε ένα από αυτά ακολουθείται από ένα επίπεδο υποδειγματοληψίας, ένα επίπεδο εξισορρόπησης και δύο πλήρως συνδεδεμένα επίπεδα σε συνδυασμό με επίπεδα απόσυρσης. Στην έξοδο εφαρμόζεται συνάρτηση ενεργοποίησης Softmax. Η παραπάνω αρχιτεκτονική αποδίδεται σχηματικά με το Σχήμα 5.1. Αναλυτικά το δίκτυο αποτελείται από:

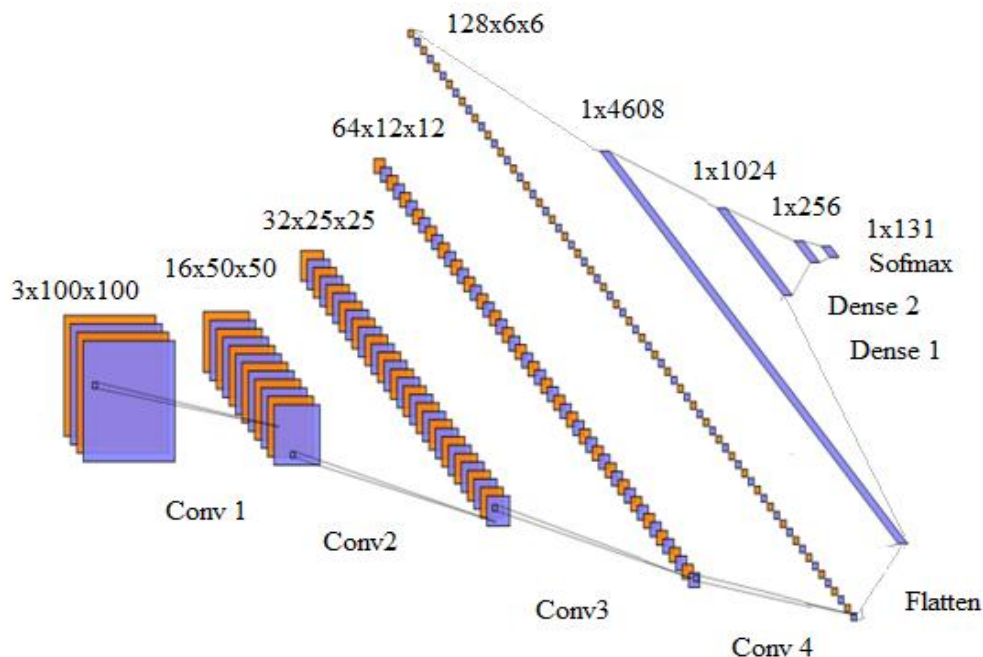
Feature Extractor

1st: Στο πρώτο συνελκτικό επίπεδο εφαρμόζεται kernel διαστάσεων $16 \times 5 \times 5$, με γέμισμα, χωρίς βήμα και συνάρτηση ενεργοποίησης ReLU. Στη συνέχεια εφαρμόζεται υποδειγματοληψία μεγίστου, με kernel 2×2 , βήμα 2 (ώστε το φίλτρο να μην επικαλύπτει ίδιες περιοχές) και χωρίς γέμισμα. Η έξοδος που προκύπτει είναι $16 \times 50 \times 50$.

2nd: Στο δεύτερο συνελκτικό επίπεδο εφαρμόζεται kernel $32 \times 5 \times 5$, με γέμισμα και συνάρτηση ενεργοποίησης ReLU. Στο επίπεδο της υποδειγματοληψίας μεγίστου εφαρμόζεται kernel 2×2 , με βήμα 2 και χωρίς γέμισμα. Η έξοδος που προκύπτει είναι $32 \times 25 \times 25$.

3rd: Στο τρίτο συνελκτικό επίπεδο εφαρμόζεται kernel $64 \times 5 \times 5$, με γέμισμα και συνάρτηση ενεργοποίησης ReLU. Στο επίπεδο της υποδειγματοληψίας μεγίστου εφαρμόζεται kernel 2×2 , με βήμα 2 και χωρίς γέμισμα. Η έξοδος που προκύπτει είναι $64 \times 12 \times 12$.

4th: Στο τέταρτο και τελευταίο συνελκτικό επίπεδο εφαρμόζεται kernel $128 \times 5 \times 5$, με γέμισμα και συνάρτηση ενεργοποίησης ReLU. Στο επίπεδο της υποδειγματοληψίας μεγίστου, εφαρμόζεται kernel 2×2 , με βήμα 2 και χωρίς γέμισμα. Η έξοδος που προκύπτει είναι $128 \times 6 \times 6$.



Σχήμα 5. 1: Αρχιτεκτονική προτεινόμενου CNN.

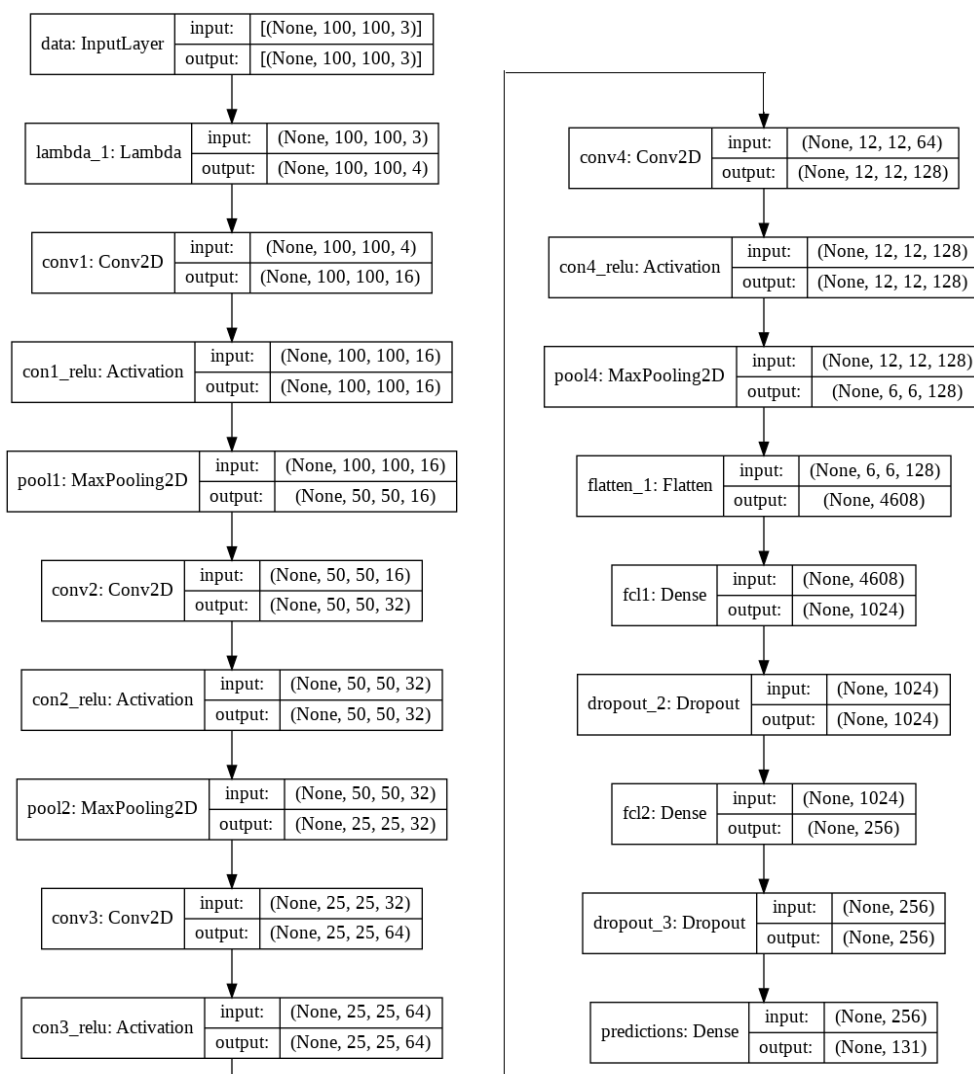
Classifier

Πρώτα εφαρμόζεται το επίπεδο εξισορρόπησης που μετατρέπει τα δεδομένα σε έναν μονοδιάστατο πίνακα 1x4608.

Στη συνέχεια τα δεδομένα εισάγονται στο πρώτο πλήρως συνδεδεμένο επίπεδο, εφαρμόζεται συνάρτηση ενεργοποίησης ReLU και η έξοδος που προκύπτει είναι 1x1024. Ακολουθεί το επίπεδο απόσυρσης, με το 20% των νευρώνων απενεργοποιείτε τυχαία.

Στο δεύτερο πλήρως συνδεδεμένο επίπεδο εφαρμόζεται συνάρτηση ενεργοποίησης ReLU και η έξοδος που προκύπτει είναι 1x1024. Ακολουθεί το επίπεδο απόσυρσης, με το 20% των νευρώνων απενεργοποιείτε τυχαία.

Στο τρίτο και τελευταίο πλήρως συνδεδεμένο επίπεδο με 131 νευρώνες, όσες και οι κλάσεις του προβλήματος, εφαρμόζεται συνάρτηση ενεργοποίησης Softmax.



Σχήμα 5. 2: Απεικόνιση σε blocks των επιπέδων του προτεινόμενου CNN.

5.1.2 Προτεινόμενη παραλλαγή ResNet-50

Στο δίκτυο αυτό έγινε χρήση του προ εκπαιδευμένου μοντέλου ResNet-50 και συγκεκριμένα χρησιμοποιήθηκε η αρχιτεκτονική του feature extractor του μοντέλου. Το προ εκπαιδευμένο μοντέλο εκπαιδεύτηκε σε εικόνες 3x224x224. Για το παρόν πρόβλημα οι διαστάσεις των εικόνων εισόδου άλλαξαν σε 3x100x100, ώστε να ταιριάζουν με τις διαστάσεις των εικόνων του dataset Fruits-360.

Η έξοδος του τελευταίου συνελκτικού επιπέδου του feature extractor του ResNet-50 έχει διαστάσεις 2028x4x4. Στη συνέχεια προστέθηκε ένα επίπεδο μέσης καθολικής υποδειγματοληψίας με έξοδο 2028 .

Έπειτα τα δεδομένα εισάγονται στο πρώτο πλήρως συνδεδεμένο επίπεδο, εφαρμόζεται συνάρτηση ενεργοποίησης ReLU και η έξοδος που προκύπτει είναι 1x1024. Ακολουθεί το επίπεδο απόσυρσης, με το 20% των νευρώνων απενεργοποιείτε τυχαία.

Στο δεύτερο πλήρως συνδεδεμένο επίπεδο εφαρμόζεται συνάρτηση ενεργοποίησης ReLU και η έξοδος που προκύπτει είναι 1x1024. Ακολουθεί το επίπεδο απόσυρσης, με το 20% των νευρώνων απενεργοποιείτε τυχαία.

Στο τρίτο και τελευταίο πλήρως συνδεδεμένο επίπεδο με 131 νευρώνες, όσες και οι κλάσεις του προβλήματος, εφαρμόζεται συνάρτηση ενεργοποίησης Softmax.

Ο classifier του περιεγράφηκε παραπάνω είναι ίδιο με αυτόν του προτεινόμενου CNN (Κεφ.5.1.1).

5.2 Μετρικές

Κατά την εκπαίδευση έγινε χρήση δύο μετρικών, την Sparse Categorical Cross Entropy Loss και της Ακρίβειας (Accuracy). Οι δύο αυτές μετρικές χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου στο σύνολο εκπαίδευσης και στο σύνολο επιβεβαίωσης, κατά την εκπαίδευση και στο σύνολο ελέγχου, στο εκπαιδευμένο δίκτυο.

5.2.2 Sparse Categorical Cross Entropy Loss

Η Sparse Categorical Cross Entropy (SCCE) χρησιμοποιείται σε προβλήματα ταξινόμησης πολλών κλάσεων (Multi-Class Classification). Τέτοιο είναι και το πρόβλημα που μελετάται στην παρούσα εργασία. Η συνάρτηση απώλειας περιγράφεται από την παρακάτω σχέση:

$$J(w) = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (5.1)$$

, όπου w τα βάρη του μοντέλου, y_i η πραγματική κλάση του δείγματος i, \hat{y}_i η πιθανότητα που πρόβλεψε το μοντέλο να ανήκει το αντικείμενο στην συγκεκριμένη κλάση, N ο αριθμός των εικόνων κάθε παρτίδας (batch).

Οι κλάσεις του προβλήματος αντιστοιχήθηκαν σε ακέραιους αριθμούς [0, 130]. Ο υπολογισμός της απώλειας γίνεται σε batches 50 εικόνων.

5.2.3 Accuracy

Κατά την διάρκεια της εκπαίδευσης ως εποπτικό μέσο για την επίδοση του μοντέλου ορίστηκε η ακρίβεια (accuracy). Η μετρική της ακρίβειας υπολογίζει το ποσοστό των δειγμάτων που ο ταξινομητής προέβλεψε ορθά την κλάση τους, προς τον αριθμό των δεδομένων που χρησιμοποιήθηκαν ($N_{samples}$).

$$accuracy_{model} = \frac{\sum_{class} TP_{class}}{N_{samples}} \quad (5.2)$$

Επίσης υπολογίστηκαν οι μετρικές precision, recall και F1-score για κάθε κλάση. Για τον υπολογισμό των μετρικών για μία κλάση A ορίστηκαν τα παρακάτω σύνολα:

- True Positive - TP: Ο αριθμός των δειγμάτων που ο ταξινομητής προέβλεψε ορθά ότι ανήκουν στην κλάση A.
- False Negative - FN: Ο αριθμός των δειγμάτων που ο ταξινομητής προέβλεψε λανθασμένα ότι ανήκουν σε άλλη κλάση.
- False Positive – FP: Ο αριθμός των δειγμάτων που ο ταξινομητής προέβλεψε λανθασμένα ότι ανήκουν στην κλάση A.

Η ακρίβεια (precision) μιας κλάσης δηλώνει τον λόγο των αληθώς θετικών δειγμάτων της κλάσης (TP) προς όλα τα δείγματα που ο ταξινομητής προέβλεψε, ορθά (TP) ή λανθασμένα (FP), ότι ανήκουν σε αυτήν.

$$precision_{class} = \frac{TP_{class}}{TP_{class} + FP_{class}} \quad (5.3)$$

Η ανάκλαση (recall) μιας κλάσης δηλώνει τον λόγο των αληθώς θετικών δειγμάτων της κλάσης (TP) προς τον αριθμό των δειγμάτων που ανήκουν σε αυτήν, ανεξάρτητα της πρόβλεψης του ταξινομητή (TP+FN).

$$recall_{class} = \frac{TP_{class}}{TP_{class} + FN_{class}} \quad (5.4)$$

F1-Score: Η μετρική αυτή συνδυάζει την ακρίβεια και την ανάκλαση μια κλάσης, καθώς αποτελεί τον αρμονικό τους μέσο.

$$F1 - score_{class} = \frac{2}{\frac{1}{precision_{class}} + \frac{1}{recall_{class}}} \quad (5.5)$$

Επίσης υπολογίστηκε ο μέσος όρος (macro average) των precision, recall και F1-score για το κάθε μοντέλο.

5.3 Πειράματα

Στην παρούσα διπλωματική πραγματοποιήθηκαν έξι πειράματα. Στα πρώτα πέντε πειράματα χρησιμοποιήθηκε το προτεινόμενο CNN (Κεφ. 5.1.1) και εκπαιδεύτηκε για 25 εποχές. Στο τελευταίο πείραμα χρησιμοποιήθηκε μια παραλλαγή του ResNet-50 (Κεφ. 5.1.2) και εκπαιδεύτηκε για 10 εποχές. Εφαρμόστηκαν οι τεχνικές αύξησης δεδομένων με τις τιμές των παραμέτρων που αναφέρθηκαν στο Κεφ. 4.2.2. Επίσης όλα τα μοντέλα αξιολογήθηκαν με βάση τις ίδιες μετρικές. Τα δεδομένα χωρίστηκαν σε batches των 50 δειγμάτων, άρα ο αριθμός των iterations ανά εποχή ήταν 1354.

Κατά τη διάρκεια της εκπαίδευσης έγινε χρήση του ModelCheckpoint του Keras, το οποίο ελέγχει κάθε εποχή κάποια επιλεγμένη μετρική και δίνει τη δυνατότητα αποθήκευσης του μοντέλου βάση των τιμών αυτής της μετρικής. Για την παρούσα διπλωματική επιλέχθηκε έλεγχος της μετρικής της ακρίβειας επιβεβαίωσης (Validation accuracy) και αποθήκευση του μοντέλου με την μεγαλύτερη ακρίβεια.

Η εκπαίδευση του νευρωνικού έγινε σε μονάδα επεξεργασίας γραφικών Tesla P100-PCIE-16GB. Επιπλέον, έγινε χρήση Python version 3.7.12, TensorFlow version 2.6.0 και Keras version 2.6.0. Παρακάτω παρουσιάζονται αναλυτικά τα πειράματα. Ο χρόνος διεξαγωγής του κάθε πειράματος δεν είναι χαρακτηριστικός, καθώς η πλατφόρμα Google Colab κάνει καταμερισμό των πόρων ανάλογα του αριθμού των χρηστών μια δεδομένη στιγμή.

5.3.2 Υλοποίηση σε χρωματικούς χώρους

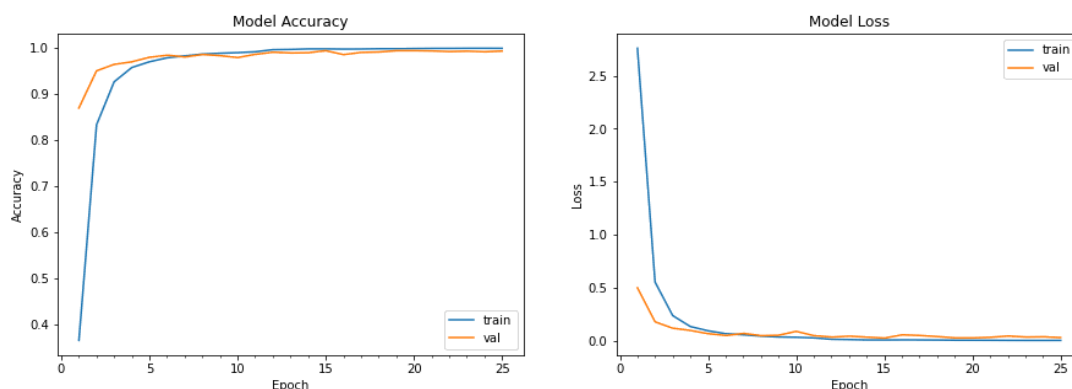
Αρχικά χρησιμοποιώντας το προτεινόμενο CNN εφαρμόστηκαν τέσσερα πειράματα, στα οποία μεταβάλλεται ο χρωματικός χώρος του συνόλου δεδομένων.

Πείραμα 1: HSV + Grayscale

Στο πείραμα αυτό τα δεδομένα από RGB μετατράπηκαν σε HSV και Grayscale. Τα κανάλια των χρωματικών χώρων ενώθηκαν, τροφοδοτώντας το νευρωνικό με είσοδο μεγέθους 4x100x100, στα πρώτα τρία κανάλια υπήρχαν οι τιμές της απόχρωσης, του κορεσμού και της τιμής της φωτεινότητας και στο τέταρτο οι τόνοι του γκρι.

Το μοντέλο είχε 5.286.323 εκπαιδεύσιμους παραμέτρους. Η εκπαίδευση διήρκεσε 9 ώρες και 32 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 22 λεπτά.

Παρακάτω παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 7 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσής σημειώθηκε στην 15^η εποχή.



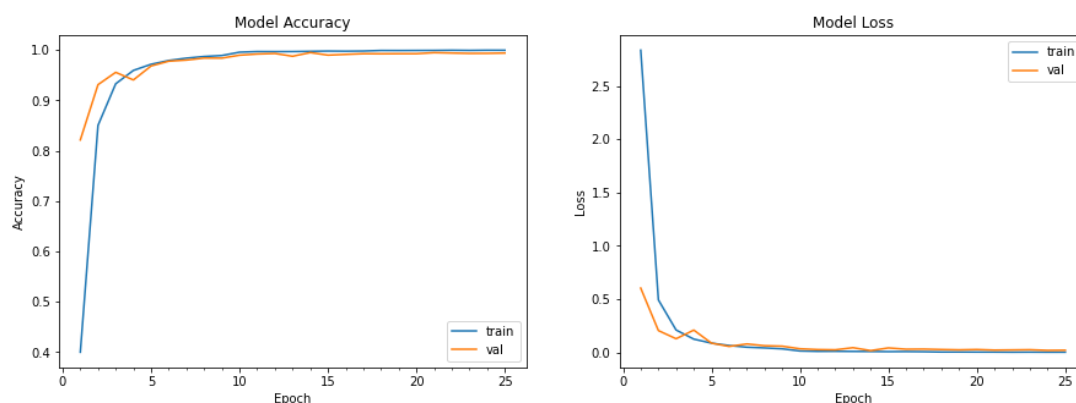
Σχήμα 5. 3: Μετρικές accuracy και loss συναρτήσει των εποχών για το Πείραμα 1.

Πείραμα 2: RGB + Grayscale

Στο πείραμα αυτό τα δεδομένα από RGB μετατράπηκαν σε Grayscale. Τα κανάλια των χρωματικών χώρων RGB και Grayscale ενώθηκαν, τροφοδοτώντας το νευρωνικό με είσοδο μεγέθους 4x100x100, στα πρώτα τρία κανάλια υπήρχαν οι τιμές της έντασης του φωτός του κόκκινου – πράσινου – μπλε και στο τέταρτο οι τόνοι του γκρι.

Το μοντέλο είχε 5.286.323 εκπαιδεύσιμους παραμέτρους. Η εκπαίδευση διήρκησε 5 ώρες και 3 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 12 λεπτά.

Στο Σχήμα 5.4 παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 10 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσης σημειώθηκε στην 14^η εποχή.



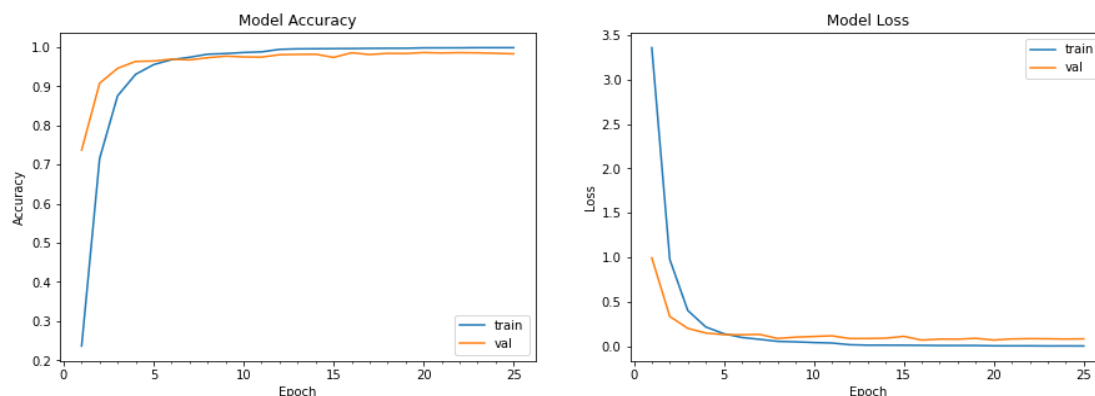
Σχήμα 5. 4: Μετρικές accuracy και loss για το σύνολο επιβεβαίωσης συναρτήσει των εποχών για το Πείραμα 2.

Πείραμα 3: HSV

Στο πείραμα αυτό τα δεδομένα από RGB μετατράπηκαν σε HSV, τροφοδοτώντας το νευρωνικό με εικόνες μεγέθους 3x100x100.

Το μοντέλο έχει 5.285.923 εκπαιδεύσιμους παραμέτρους. Η εκπαίδευση διήρκησε 9 ώρες και 43 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 23 λεπτά.

Στο Σχήμα 5.5 παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 10 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσης σημειώθηκε στην 20^η εποχή.



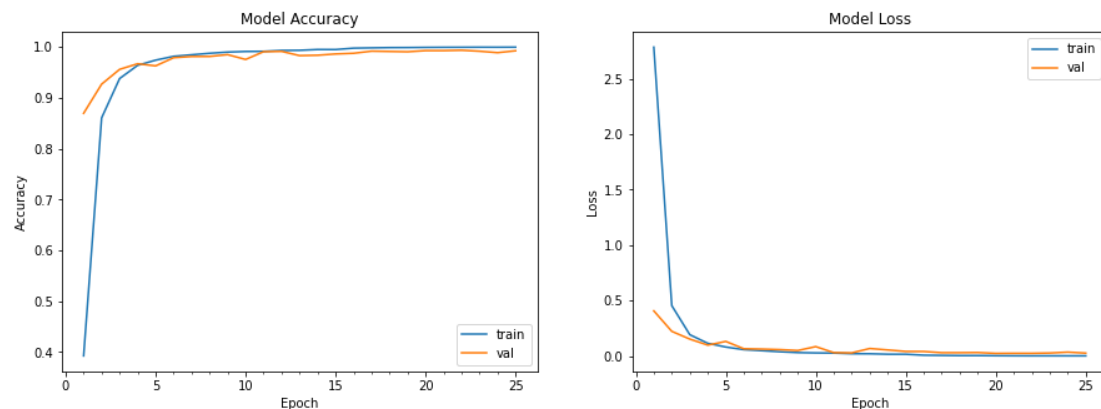
Σχήμα 5. 5: Μετρικές accuracy και loss για το σύνολο επιβεβαίωσης συναρτήσει των εποχών για το Πείραμα 3.

Πείραμα 4: RGB

Στο πείραμα αυτό τα δεδομένα τροφοδοτήθηκαν σε RGB.

Το μοντέλο έχει 5.285.923 εκπαιδεύσιμους παραμέτρους. Η εκπαίδευση διήρκησε 9 ώρες και 49 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 23 λεπτά.

Στο Σχήμα 5.6 παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 7 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσης σημειώθηκε στην 20^η εποχή.



Σχήμα 5. 6: Μετρικές accuracy και loss για το σύνολο επιβεβαίωσης συναρτήσει των εποχών για το Πείραμα 4.

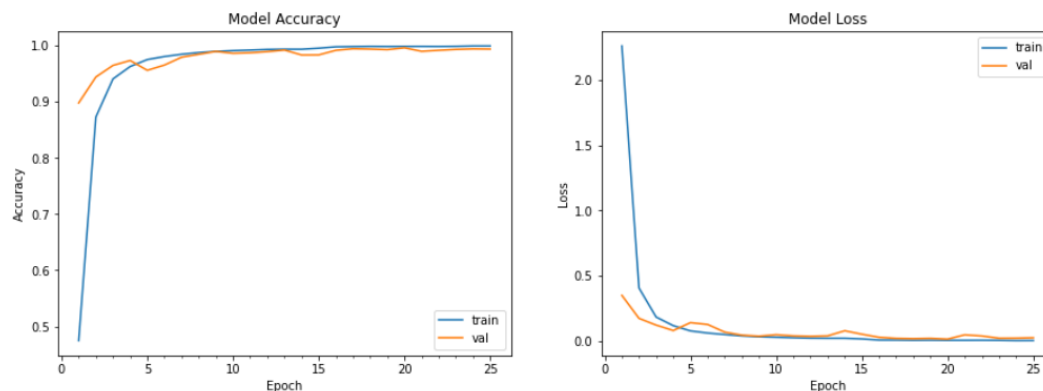
5.3.3 Υλοποίηση με ισορροπημένο σύνολο δεδομένων

Πείραμα 5: Balanced RGB

Στο πείραμα αυτό τα δεδομένα τροφοδοτήθηκαν στο προτεινόμενο CNN σε RGB χρωματικό χώρο. Η διαδικασία που ακολουθήθηκε περιγράφεται στο Κεφ 4.3.

Το μοντέλο έχει 5.285.923 εκπαιδεύσιμους παραμέτρους. Η εκπαίδευση διήρκησε 5 ώρες και 10 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 12 λεπτά.

Στο Σχήμα 5.7 παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 10 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσης σημειώθηκε στην 20^η εποχή.



Σχήμα 5. 7: Μετρικές accuracy και loss για το σύνολο επιβεβαίωσης συναρτήσει των εποχών για το Πείραμα 5.

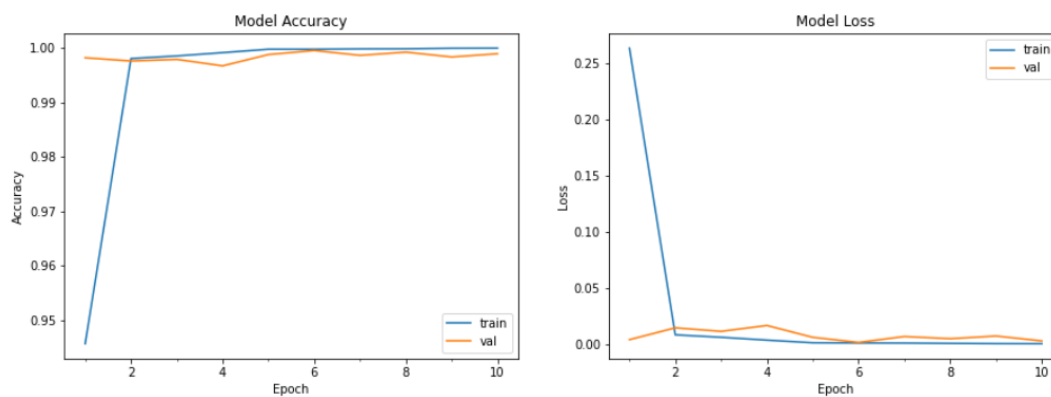
5.3.4 Υλοποίηση με παραλλαγή ResNet-50

Πείραμα 6: Παραλλαγή του ResNet-50

Η παραλλαγή του ResNet-50 (Κεφ. 5.1.2) εκπαιδεύτηκε εξολοκλήρου, δηλαδή κανένα από τα βάρη του μοντέλου δεν κρατήθηκε παγωμένο. Συγκεκριμένα, η εκμάθηση χαρακτηριστικών αρχικοποιήθηκε με τα βάρη του ResNet-50.

Το μοντέλο έχει 25.981.955 παραμέτρους. Η εκπαίδευση διήρκησε 6 ώρες και 42 λεπτά, με κάθε εποχή να διαρκεί κατά μέσο όρο 40 λεπτά.

Στο Σχήμα 5.8 παρουσιάζονται τα διαγράμματα της ακρίβειας και της απώλειας για το σύνολο εκπαίδευσης και επιβεβαίωσης συναρτήσει των εποχών. Όπως φαίνεται το δίκτυο συγκλίνει μετά από 6 περίπου εποχές. Συγκεκριμένα μέγιστη ακρίβεια για το σύνολο επιβεβαίωσης σημειώθηκε στην 6^η εποχή.



Σχήμα 5. 8: Μετρικές accuracy και loss για το σύνολο επιβεβαίωσης συναρτήσει των εποχών για το Πείραμα 6.

6 Αποτελέσματα

6.1 Αξιολόγηση – Αποτελέσματα

Η μετρική με βάση την οποία αξιολογήθηκαν τα μοντέλα στα πειράματα είναι το accuracy, τα average precision, average recall, average F1-Score.

Παρακάτω παρουσιάζονται οι τιμές τους για κάθε πείραμα που πραγματοποιήθηκε. Η στρογγυλοποίηση των τιμών έγινε στο δεύτερο δεκαδικό ψηφίο.

	Πείραμα	Accuracy	Precision	Recall	F1-Score
1	HSV and Grayscale	94,88%	0.95	0.95	0.95
2	RGB and Grayscale	96,78%	0.97	0.97	0.97
3	HSV	94,20%	0.94	0.94	0.94
4	RGB	97,76%	0.98	0.98	0.98
5	RGB Balanced	97,39%	0.98	0.97	0.97
6	ResNet50	99,55%	1.00	1.00	1.00

Πίνακας 2: Μετρικές των μοντέλων των Πειραμάτων 1 - 6.

Στο Παράρτημα Β παρουσιάζονται οι τιμές των μετρικών precision, recall και F1-score για κάθε κλάση ανά πείραμα.

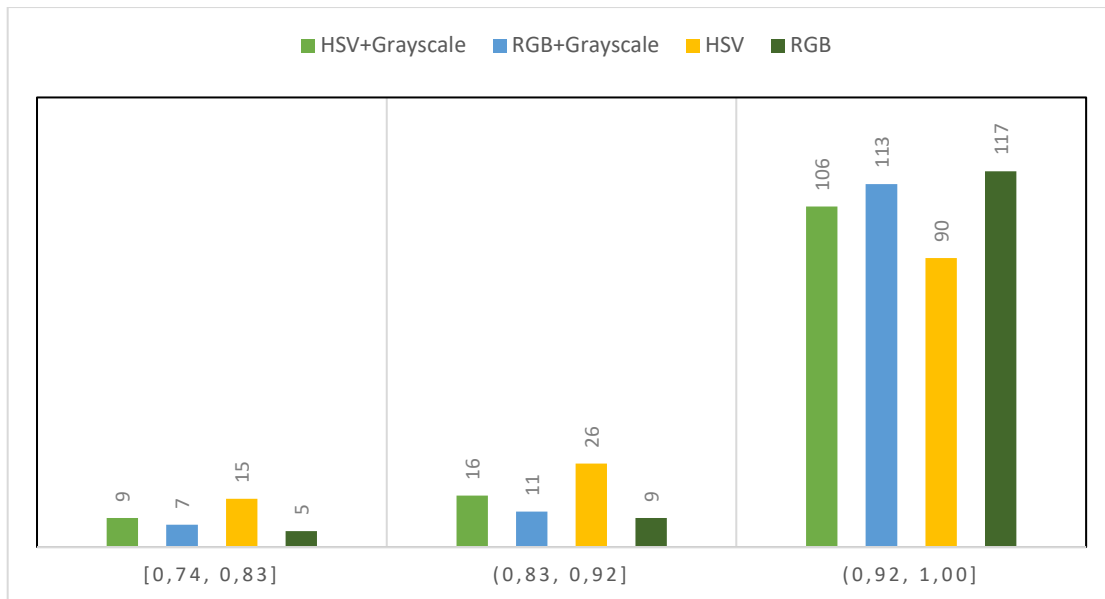
Στο Παράρτημα Γ παρουσιάζονται οι πίνακες σύγχυσης (confusion matrix) για τα πειράματα 4 και 6.

Πειράματα 1 – 4

Συγκρίνοντας τα αποτελέσματα των πειραμάτων 1 - 4 που αφορούν υλοποιήσεις σε διαφορετικούς χρωματικούς χώρους φαίνεται ότι μεγαλύτερη ακρίβεια δίνει το μοντέλο με RGB εικόνες.

Η προσθήκη του Grayscale καναλιού στις RGB εικόνες δεν πρόσφερε μεγαλύτερη ακρίβεια στο μοντέλο, αντίθετα μείωσε ελάχιστα την τιμή της. Στην περίπτωση της εκπαίδευσης με HSV εικόνες η προσθήκη του Grayscale καναλιού αύξησε την ακρίβεια του μοντέλου.

Με σκοπό την πληρέστερη παρουσίαση των αποτελεσμάτων δίνεται το ιστόγραμμα (Σχήμα 5.9) στο οποίο φαίνεται η κατανομή των κλάσεων με βάση τη μετρική precision για τα τέσσερα μοντέλα. Η χαμηλότερη τιμή ακρίβειας ήταν 0,74 και η υψηλότερη 1, συνεπώς η ακρίβεια χωρίστηκε σε τρία ίσα διαστήματα τιμών τα οποία είναι [0.74, 0.83], [0.83, 0.92], [0.92, 1.0].

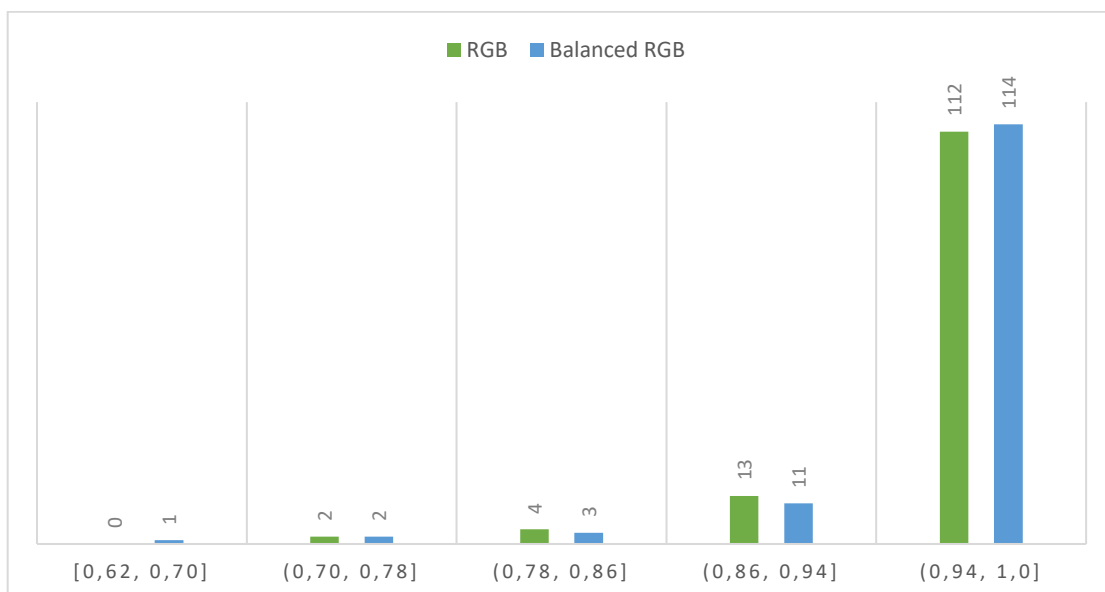


Σχήμα 5. 9: Ιστόγραμμα κατανομής των κλάσεων συναρτήσει της μετρικής precision για τα μοντέλα των πειραμάτων 1 - 4.

Πείραμα 5

Από τον Πίνακα 2 φαίνεται ότι η εξισορρόπηση των δειγμάτων στις κλάσεις δεν αύξησε την ακρίβεια του δικτύου. Συγκεκριμένα η μετρική precision δεν μεταβλήθηκε σε 79 κλάσεις, ενώ σε 30 μειώθηκε και σε 22 αυξήθηκε.

Παρακάτω δίνεται το ιστόγραμμα της κατανομής των κλάσεων με βάση το precision για τα μοντέλα RGB και Balanced RGB (Σχήμα 5.10).

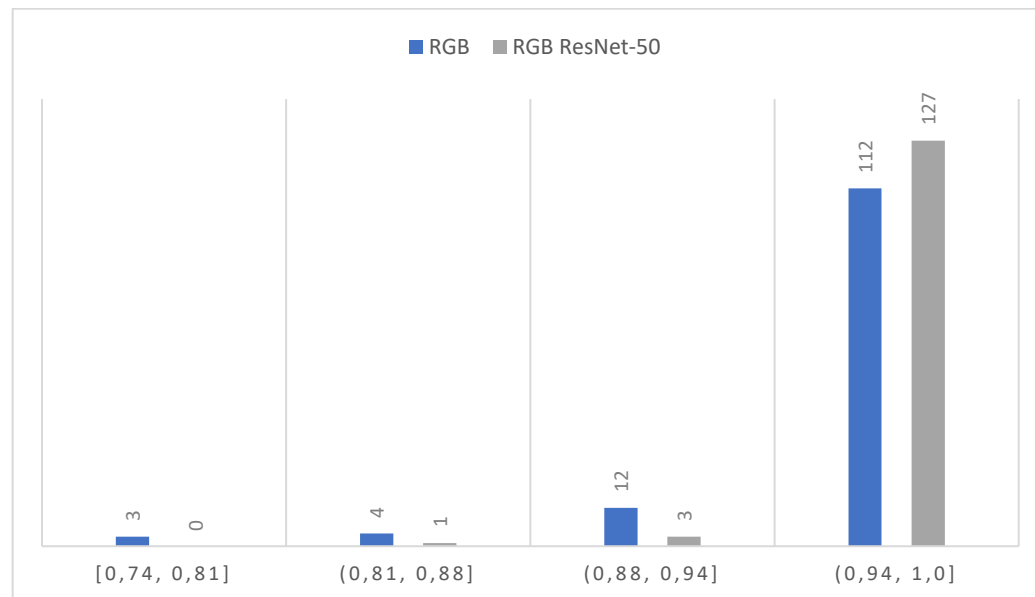


Σχήμα 5. 10: Ιστόγραμμα κατανομής των κλάσεων συναρτήσει της μετρικής precision για τα μοντέλα των πειραμάτων 3 και 5.

Πείραμα 6

Το πείραμα μεταφοράς μάθησης με ResNet-50 έδωσε τα καλύτερα αποτελέσματα ακρίβειας, φτάνοντας το 99,56%, αύξηση κατά 2% σε σχέση με την ακρίβεια του προτεινόμενου CNN με RGB εικόνες. Συγκεκριμένα η μετρική precision μεταβλήθηκε σε 85 κλάσεις, ενώ σε 9 μειώθηκε και σε 37 αυξήθηκε.

Παρακάτω δίνεται το ιστόγραμμα της κατανομής των κλάσεων με βάση το precision για τα μοντέλα RGB και RGB με παραλλαγή ResNet-50 (Σχήμα 5.11).



Σχήμα 5. 11: Ιστόγραμμα κατανομής των κλάσεων συναρτήσει της μετρικής precision για τα μοντέλα των πειραμάτων 3 και 6..

6.2 Οπτικοποίηση με Score-CAM

6.2.1 Θεωρητική Εισαγωγή

Η επεξήγηση των Βαθιών Συνελικτικών Δικτύων βοηθά στη διαφάνεια της λειτουργίας τους, καθώς δίνει ερμηνεύσιμες από τους ανθρώπους εξηγήσεις σχετικά με τον τρόπο που το μοντέλο παίρνει αποφάσεις.

Στα CNNs η επεξήγηση δίνεται με μορφή επεξηγηματικού χάρτη (saliency map) οπτικοποιώντας τις σημαντικές περιοχές. Συγκεκριμένα, δίνοντας μια εικόνα στον αλγόριθμο επεξήγησης παράγεται ένας χάρτης με ενεργοποιημένες τις περιοχές που εστίασε το νευρωνικό για να προβλέψει μια συγκεκριμένη κλάση.

Οι πιο διαδεδομένες κατηγορίες επεξήγησης στα CNNs είναι:

- Μέθοδοι οπισθοδιάδοσης (Backpropagation methods). Οι μέθοδοι αυτοί οπισθοδιαδίδουν την κλίση της κλάσης που ανήκει η εικόνα έως την είσοδο, ώστε να επισημάνουν τις περιοχές που επηρέασαν την απόφαση. Οι μέθοδοι αυτοί δίνουν αποτελέσματα χαμηλής ποιότητας και με θόρυβο.
- Μέθοδοι Διαταραχής (Perturbation methods). Διαταράσσουν την εικόνα εισόδου πολλαπλασιάζοντας την με μια μάσκα, έπειτα την περνούν από το μοντέλο και εξετάζουν την αλλαγή στην έξοδο του μοντέλου, σε σχέση την έξοδο της αρχικής εικόνας. Υπολογίζουν επεξηγηματικούς χάρτες ως ένα σταθμισμένο μέσο όρο των μασκών. Τα βάρη του σταθμισμένου μέσου όρου υπολογίζονται από την επιρροή των πολλαπλασιασμένων με την εικόνα μασκών στην έξοδο.
- Μέθοδοι κλίσης (Gradient methods). Υπολογίζουν επεξηγηματικούς χάρτες ως ένα σταθμισμένο μέσο όρο των feature maps. Τα βάρη του σταθμισμένου μέσου όρου υπολογίζονται από την κλίση της εξόδου ως προς ένα ενδιάμεσο επίπεδο.

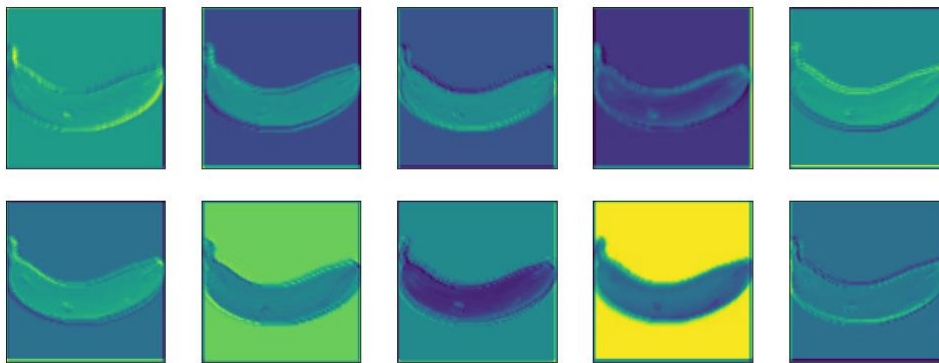
Στην παρούσα εργασία χρησιμοποιήθηκαν μέθοδοι διαταραχής και συγκεκριμένα ο αλγόριθμος Score-CAM [35]. Ο αλγόριθμος Score-CAM χρησιμοποιεί ως μάσκες διαταραχής τα feature maps της εικόνας για ένα ορισμένο επίπεδο, συνήθως το τελευταίο συνελικτικό.

Η διαδικασία με βάση την οποία εξάγεται ο saliency map είναι η εξής:

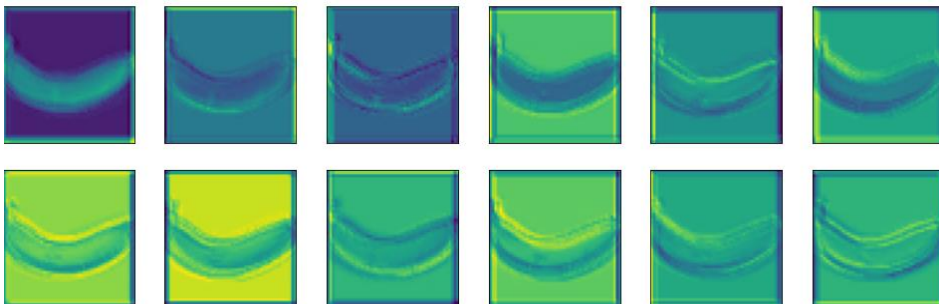
- Εισάγεται η εικόνα στο μοντέλο και εξάγονται τα k feature maps της για ένα ορισμένο επίπεδο l .
- Για κάθε k feature map:
 1. Αλλάζει η διάσταση του ώστε να ταυτίζεται με αυτή της αρχικής εικόνας.
 2. Κανονικοποιείται στο $[0,1]$.
 3. Πολλαπλασιάζεται στοιχείο με στοιχείο με την αρχική εικόνα.
- Η κάθε masked εικόνα που προκύπτει εισάγεται στο μοντέλο και υπολογίζεται η διαφορά μεταξύ της εξόδου της και της αρχικής εικόνας για την κλάση ενδιαφέροντος.
- Στις k διαφορές που προκύπτουν για την κλάση ενδιαφέροντος, εφαρμόζεται η συνάρτηση Softmax και οι τιμές του διανύσματος που προκύπτει αποτελούν τα βάρη των feature maps.
- Υπολογίζεται ο σταθμισμένος μέσος όρος των feature maps με χρήση των βαρών που προέκυψαν στο προηγούμενο βήμα.
- Στο αποτέλεσμα εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU.

6.2.2 Χάρτες Χαρακτηριστικών

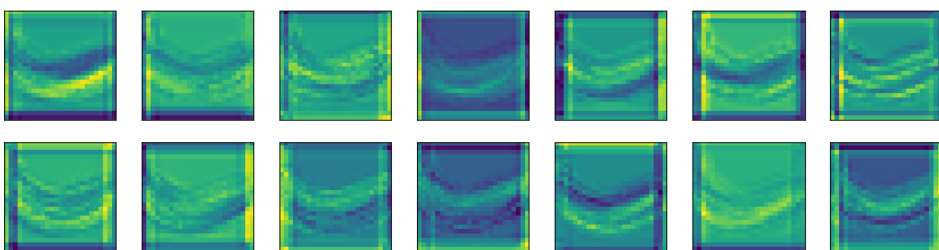
Παρακάτω παρουσιάζονται κάποιοι χάρτες χαρακτηριστικών για τα διάφορα συνελκτικά επίπεδα του προτεινόμενου μοντέλου με RGB είσοδο. Ως είσοδος εφαρμόστηκε μια εικόνα της κλάσης “Banana” από το σύνολο ελέγχου (Σχήμα 16.α). Η εικόνα εισόδου είναι η ίδια για όλους τους χάρτες χαρακτηριστικών. Συγκρίνοντας τους χάρτες χαρακτηριστικών φαίνεται ότι στα πρώτα συνελκτικά επίπεδα σκιαγραφείται το αντικείμενο της εικόνας, μιας το νευρωνικό εντοπίζει γωνίες και ακμές και οι διαστάσεις των χαρτών είναι συγκρίσιμες με τις διαστάσεις της αρχικής εικόνας. Ωστόσο στα τελευταία συνελκτικά επίπεδα οι διαστάσεις των χαρτών χαρακτηριστικών μειώνονται σημαντικά και πλέον περιέχουν περισσότερη πληροφορία σχετικά με την ενεργοποίηση των κλάσεων.



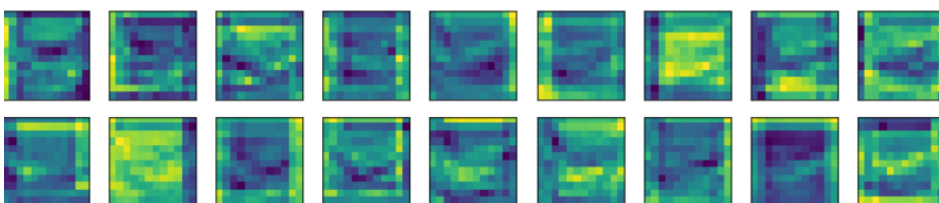
Σχήμα 5. 12: Feature Maps διαστάσεων 100x100 από το πρώτο συνελκτικό επίπεδο.



Σχήμα 5. 13: Feature Maps διαστάσεων 50x50 από το δεύτερο συνελκτικό επίπεδο.



Σχήμα 5. 14: Feature Maps διαστάσεων 25x25 από το τρίτο συνελκτικό επίπεδο.



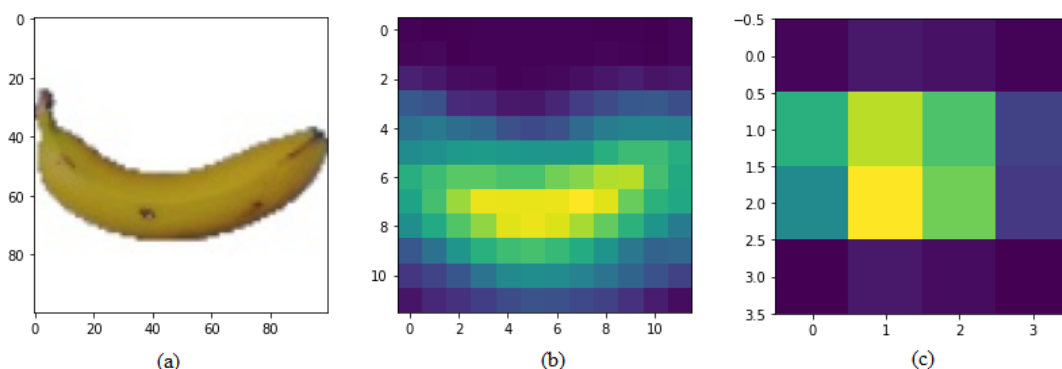
Σχήμα 5. 15: Feature Maps διαστάσεων 12x12 από το τέταρτο συνελκτικό επίπεδο.

6.2.3 Αποτελέσματα οπτικοποίησης

Για την οπτικοποίηση του προτεινόμενου CNN με RGB εικόνες, χρησιμοποιήθηκαν οι χάρτες χαρακτηριστικών του τελευταίου συνελκτικού επιπέδου, διαστάσεων $128 \times 12 \times 12$. Η οπτικοποίηση με Score-cam παράγει μια εικόνα διαστάσεων 12×12 , η οποία περιέχει τις περιοχές ενεργοποίησης της κλάσης που μελετάται.

Ομοίως για την οπτικοποίηση του μοντέλου παραλλαγής ResNet-50 με RGB εικόνες, χρησιμοποιήθηκαν οι χάρτες χαρακτηριστικών του τελευταίου residual block, με διαστάσεις $2048 \times 4 \times 4$. Η εφαρμογή Score-cam παράγει μια εικόνα διαστάσεων 4×4 .

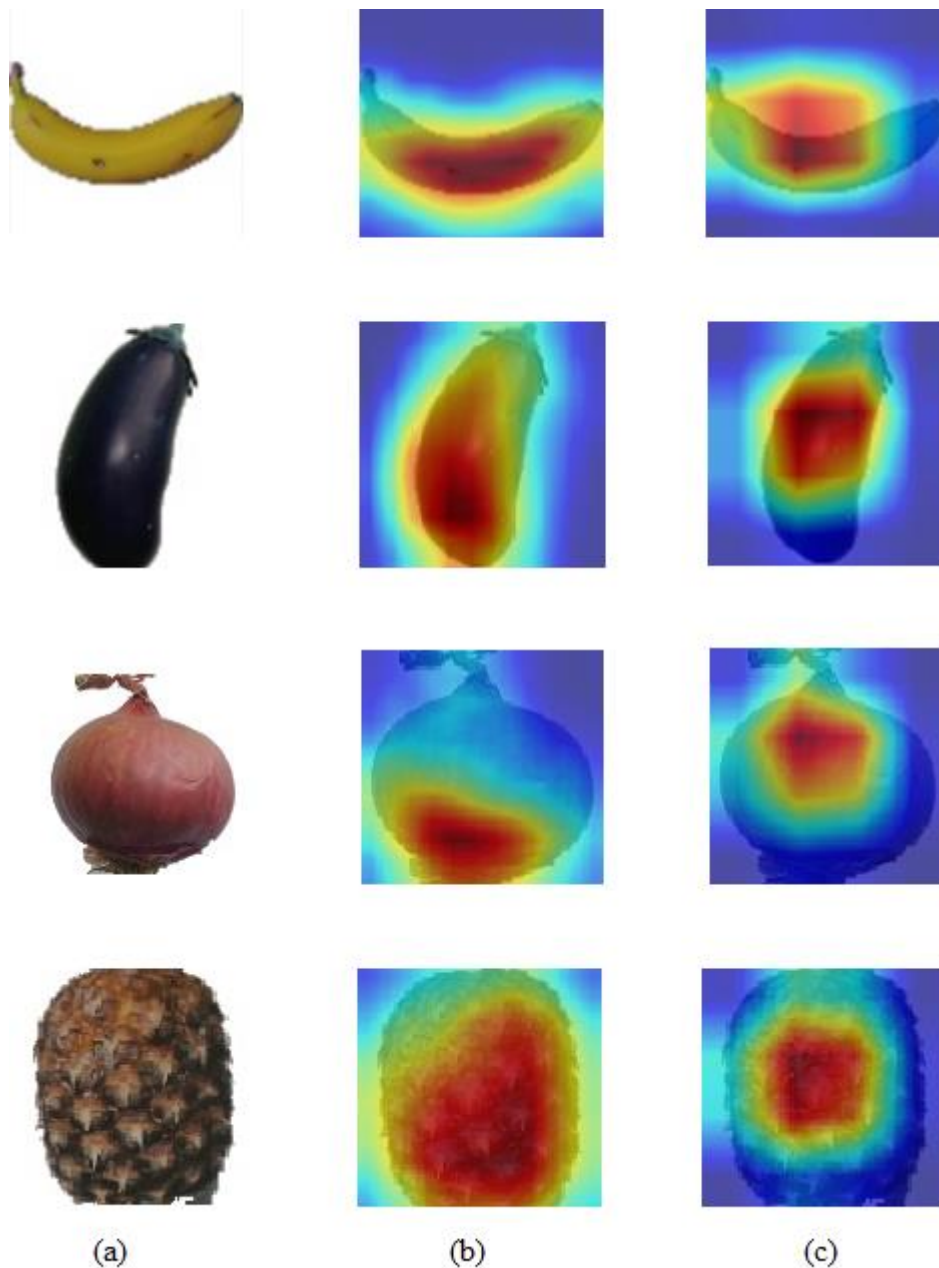
Παρακάτω παρουσιάζονται τα saliency maps που παράγονται μετά την εφαρμογή του αλγόριθμου Score-cam, για είσοδο μια τυχαία εικόνα του συνόλου ελέγχου (Σχήμα 5.16). Όπως αναφέρθηκε και παραπάνω το saliency map του προτεινόμενου CNN έχει διαστάσεις 12×12 , ενώ το saliency map του μοντέλου παραλλαγής ResNet-50, 4×4 . Αυτό έχει ως αποτέλεσμα στο δεύτερο να μην διακρίνονται καθαρά οι περιοχές ενεργοποίησης της εικόνας. Μια λύση σε αυτό το πρόβλημα θα ήταν η χρήση χαρτών χαρακτηριστικών προηγούμενων residual block. Παρόλα αυτά για να προκύψουν saliency maps μεγέθους 13×13 χρειάζεται να χρησιμοποιηθούν οι χάρτες χαρακτηριστικών της εξόδου του δεύτερου residual block, δηλαδή 27 συνελκτικά επίπεδα πίσω. Αυτό έχει ως αποτέλεσμα την απώλεια μεγάλου όγκου πληροφορίας σχετικά με τις περιοχές ενεργοποίησης της κλάσης και για αυτό τον λόγο δεν χρησιμοποιείται.



Σχήμα 5. 16: (α) Εικόνα του συνόλου δεδομένων ελέγχου της κλάσης “Banana”.
(β) Saliency map της εικόνας (α), μετά από εφαρμογή Score-cam στο μοντέλο του προτεινόμενου CNN με RGB.
(γ) Saliency map της εικόνας (α), μετά από εφαρμογή Score-cam στο μοντέλο παραλλαγής ResNet-50 με RGB.

Στη συνέχεια η γραφική αναπαράσταση των περιοχών ενδιαφέροντος έγινε με χρήση heatmap. Τα heatmaps αποτελούν χάρτες, οι τιμές των οποίων αναπαρίστανται με χρώματα. Οι τιμές του χάρτη κατανέμονται σε μια διαβαθμισμένη χρωματική διάταξη, με τις περιοχές υψηλής δραστηριότητας να απεικονίζονται με κόκκινο και οι περιοχές χαμηλότερης με μπλε.

Με σκοπό την υπέρθεση των αρχικών εικόνων και των heatmaps, αυξήθηκε το μέγεθος των δεύτερων από 12x12 και 4x4 σε 100x100. Μέσω της υπέρθεσης των δύο εικόνων γίνεται πιο εύκολος ο εντοπισμός των σημείων ενεργοποίησης. Παρόλα αυτά λόγω της μεγέθυνσης του heatmap, οι περιοχές ενδιαφέροντος εντοπίζονται σε μεγαλύτερη επιφάνεια και κάποιες φορές στις περιοχές ενεργοποίησης περιλαμβάνεται και το φόντο της εικόνας (Σχήμα 5.17, 1^η και 2^η σειρά). Το πρόβλημα αυτό είναι πιο εμφανή στην οπτικοποίηση του μοντέλου ResNet-50, που λόγω του πολύ μικρού μεγέθους του saliency map περιοχές έντονης ενεργοποίησης βρίσκονται έξω από το αντικείμενο.



Σχήμα 5. 17: (α) Εικόνες δειγμάτων του συνόλου ελέγχου.
Οπτικοποίηση των περιοχών ενεργοποίησης με Score-Cam στο προτεινόμενο μοντέλο (b) και στο μοντέλο παραλλαγής ResNet-50.

6.3 Μελλοντικές επεκτάσεις

Καθώς το σύνολο δεδομένων που χρησιμοποιήθηκε κατά την εκπαίδευση αποτελείται από εικόνες διαστάσεων 100x100, μια μελλοντική επέκταση θα ήταν η εφαρμογή των αλγορίθμων που εξετάστηκαν στην παρούσα διπλωματική σε ένα νέο σύνολο δεδομένων που θα περιέχει μεγαλύτερες εικόνες.

Επιπλέον, μια επέκταση της εργασίας θα ήταν η χρήση ενός πιο ελαφρού συνελικτικού δικτύου, δηλαδή θα έχει μικρότερο αριθμό παραμέτρων αλλά θα δίνει τα ίδια αποτελέσματα ακρίβειας. Τέτοια δίκτυα είναι το MobileNet και SquixeNet. Σκοπός αυτής της εφαρμογής είναι η ταξινόμησή σε πραγματικό χρόνο μέσω εφαρμογών σε smart-phone.

Βιβλιογραφία

- [1] Kuwata, K., & Shibasaki, R. (2015, July). Estimating crop yields with deep learning and remotely sensed data. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 858-861). IEEE.
- [2] Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782.
- [3] Rebetez, J., Satizábal, H. F., Mota, M., Noll, D., Büchi, L., Wendling, M., ... & Burgos, S. (2016, April). Augmenting a convolutional neural network with local histograms-A case study in crop classification from high-resolution UAV imagery. In *ESANN*.
- [4] Dyrmann, M., Mortensen, A. K., Midtiby, H. S., & Jørgensen, R. N. (2016, June). Pixel-wise classification of weeds and crops in images by using a fully convolutional neural network. In *Proceedings of the International Conference on Agricultural Engineering, Aarhus, Denmark* (pp. 26-29).
- [5] Bargouti, S., & Underwood, J. (2017, May). Deep fruit detection in orchards. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3626-3633). IEEE.
- [6] Herman, H., Susanto, A., Cenggoro, T. W., Suharjito, S., & Pardamean, B. (2020). Oil Palm Fruit Image Ripeness Classification with Computer Vision using Deep Learning and Visual Attention. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 12(2), 21-27.
- [7] Xiong, J., Lin, R., Liu, Z., He, Z., Tang, L., Yang, Z., & Zou, X. (2018). The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment. *biosystems engineering*, 166, 44-57.
- [8] Rojas-Aranda, J. L., Nunez-Varela, J. I., Cuevas-Tello, J. C., & Rangel-Ramirez, G. (2020, June). Fruit Classification for Retail Stores Using Deep Learning. In *Mexican Conference on Pattern Recognition* (pp. 3-13). Springer, Cham.
- [9] Hameed, K., Chai, D., & Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing*, 80, 24-44.
- [10] Bolle, R. M., Connell, J. H., Haas, N., Mohan, R., & Taubin, G. (1996, December). Veggievision: A produce recognition system. In *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96* (pp. 244-251). IEEE.
- [11] Arivazhagan, S., Shebiah, R. N., Nidhyandhan, S. S., & Ganesan, L. (2010). Fruit recognition using color and texture features. *Journal of Emerging Trends in Computing and Information Sciences*, 1(2), 90-94.
- [12] Zhang, Y., & Wu, L. (2012). Classification of fruits using computer vision and a multiclass support vector machine. *sensors*, 12(9), 12489-12505.
- [13] Suresha, M., Kumar, K. S. S., & Kumar, G. S. (2012). Texture features and decision trees based vegetables classification. *International Journal of Computer Applications*, 975, 8878.

- [14] Chowdhury, M. T., Alam, M. S., Hasan, M. A., & Khan, M. I. (2013). Vegetables detection from the glossary shop for the blind. *IOSR Journal of Electrical and Electronics Engineering*, 8(3), 43-53.
- [15] Danti, A., Madgi, M., & Anami, B. S. (2012). Mean and range color features based identification of common Indian leafy vegetables. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 5(3), 151-160.
- [16] Wang, S. H., & Chen, Y. (2020). Fruit category classification via an eight-layer convolutional neural network with parametric rectified linear unit and dropout technique. *Multimedia Tools and Applications*, 79(21), 15117-15133.
- [17] Patino-Saucedo, A., Rostro-Gonzalez, H., & Conradt, J. (2018, December). Tropical fruits classification using an alexnet-type convolutional neural network and image augmentation. In *International Conference on Neural Information Processing* (pp. 371-379). Springer, Cham.
- [18] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [19] Rocha, A., Hauagge, D. C., Wainer, J., & Goldenstein, S. (2008, October). Automatic produce classification from images using color, texture and appearance cues. In *2008 XXI Brazilian Symposium on Computer Graphics and Image Processing* (pp. 3-10). IEEE.
- [20] Mureşan, H., & Oltean, M. (2017). Fruit recognition from images using deep learning. *arXiv preprint arXiv:1712.00580*.
- [21] Fruits-360: A dataset of images containing fruits and vegetables.
<https://github.com/Horea94/Fruit-Images-Dataset>
<https://www.kaggle.com/moltean/fruits>
- [22] Sakib, S., Ashrafi, Z., Siddique, M., & Bakr, A. (2019). Implementation of fruits recognition classifier using convolutional neural network algorithm for observation of accuracies for various hidden layers. *arXiv preprint arXiv:1904.00783*.
- [23] Rajasekar, L., & Sharmila, D. (2019). Performance analysis of soft computing techniques for the automatic classification of fruits dataset. *Soft Computing*, 23(8), 2773-2788.
- [24] Nasir, I. M., Bibi, A., Shah, J. H., Khan, M. A., Sharif, M., Iqbal, K., ... & Kadry, S. (2021). Deep learning-based classification of fruit diseases: An application for precision agriculture. *CMC-Computers Materials & Continua*, 66(2), 1949-1962.
- [25] <https://github.com/Rahul-Sridhar/VGG-16-net-on-Fruits-360-dataset>
- [26] Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, 1(4), 515-526.
- [27] Mitchell, T. M. (1997). Does machine learning really work?. *AI magazine*, 18(3), 11-11.
- [28] Haykin, S., & Network, N. (2004). A comprehensive foundation. *Neural networks*, 2(2004), 41.
- [29] Mendel, J. M. (1970). Adaptive learning and pattern recognition systems. *Theory and applications*.

- [30] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1), 106-154.
- [31] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [32] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1), 1-40.
- [33] https://www.youtube.com/watch?v=_HFKJ144JuU
- [34] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10, pp. 978-3). Berlin: Springer.
- [35] Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 24-25).

Παράρτημα Α

Στο παρακάτω πίνακα βρίσκονται ο αριθμός των δειγμάτων του συνόλου εκπαίδευσης (περιέχει και το σύνολο επιβεβαίωσης) και του συνόλου ελέγχου ανά κλάση.

Κλάση	Αριθμός εικόνων του συνόλου εκπαίδευσης	Αριθμός εικόνων του συνόλου ελέγχου
Apple Braeburn	492	164
Apple Crimson Snow	444	148
Apple Golden 1	480	160
Apple Golden 2	492	164
Apple Golden 3	481	161
Apple Granny Smith	492	164
Apple Pink Lady	456	152
Apple Red 1	492	164
Apple Red 2	492	164
Apple Red 3	429	144
Apple Red Delicious	490	166
Apple Red Yellow 1	492	164
Apple Red Yellow 2	672	219
Apricot	492	164
Avocado	427	143
Avocado ripe	491	166
Banana	490	166
Banana Lady Finger	450	152
Banana Red	490	166
Beetroot	450	150
Blueberry	462	154
Cactus fruit	490	166
Cantaloupe 1	492	164
Cantaloupe 2	492	164
Carambola	490	166
Cauliflower	702	234
Cherry 1	492	164
Cherry 2	738	246
Cherry Rainier	738	246
Cherry Wax Black	492	164
Cherry Wax Red	492	164
Cherry Wax Yellow	492	164
Chestnut	450	153
Clementine	490	166
Cocos	490	166
Corn	450	150
Corn Husk	462	154
Cucumber Ripe	392	130
Cucumber Ripe 2	468	156

Κλάση	Αριθμός εικόνων του συνόλου εκπαίδευσης	Αριθμός εικόνων του συνόλου ελέγχου
Dates	490	166
Eggplant	468	156
Fig	702	234
Ginger Root	297	99
Granadilla	490	166
Grape Blue	984	328
Grape Pink	492	164
Grape White	490	166
Grape White 2	490	166
Grape White 3	492	164
Grape White 4	471	158
Grapefruit Pink	490	166
Grapefruit White	492	164
Guava	490	166
Hazelnut	464	157
Huckleberry	490	166
Kaki	490	166
Kiwi	466	156
Kohlrabi	471	157
Kumquats	490	166
Lemon	492	164
Lemon Meyer	490	166
Limes	490	166
Lychee	490	166
Mandarine	490	166
Mango	490	166
Mango Red	426	142
Mangostan	300	102
Maracuja	490	166
Melon Piel de Sapo	738	246
Mulberry	492	164
Nectarine	492	164
Nectarine Flat	480	160
Nut Forest	654	218
Nut Pecan	534	178
Onion Red	450	150
Onion Red Peeled	445	155
Onion White	438	146
Orange	479	160
Papaya	492	164
Passion Fruit	490	166
Peach	492	164
Peach 2	738	246
Peach Flat	492	164

Κλάση	Αριθμός εικόνων του συνόλου εκπαίδευσης	Αριθμός εικόνων του συνόλου ελέγχου
Pear	492	164
Pear 2	696	232
Pear Abate	490	166
Pear Forelle	702	234
Pear Kaiser	300	102
Pear Monster	490	166
Pear Red	666	222
Pear Stone	711	237
Pear Williams	490	166
Pepino	490	166
Pepper Green	444	148
Pepper Orange	702	234
Pepper Red	666	222
Pepper Yellow	666	222
Physalis	492	164
Physalis with Husk	492	164
Pineapple	490	166
Pineapple Mini	493	163
Pitahaya Red	490	166
Plum	447	151
Plum 2	420	142
Plum 3	900	304
Pomegranate	492	164
Pomelo Sweetie	450	153
Potato Red	450	150
Potato Red Washed	453	151
Potato Sweet	450	150
Potato White	450	150
Quince	490	166
Rambutan	492	164
Raspberry	490	166
Redcurrant	492	164
Salak	490	162
Strawberry	492	164
Strawberry Wedge	738	246
Tamarillo	490	166
Tangelo	490	166
Tomato 1	738	246
Tomato 2	672	225
Tomato 3	738	246
Tomato 4	479	160
Tomato Cherry Red	492	164
Tomato Heart	684	228
Tomato Maroon	367	127

Κλάση	Αριθμός εικόνων του συνόλου εκπαίδευσης	Αριθμός εικόνων του συνόλου ελέγχου
Tomato not Ripened	474	158
Tomato Yellow	459	153
Walnut	735	249
Watermelon	475	157
Total	67692	22688
Average per class	520	174

Πίνακας 3: Αριθμός εικόνων των συνόλων εκπαίδευσης-επιβεβαίωσης και του συνόλου ελέγχου.

Παράρτημα Β

Στους παρακάτω πίνακες βρίσκονται οι τιμές των μετρικών precision, recall και F1-Score για κάθε κλάση, όπως προέκυψαν από τα Πειράματα 1 – 6.

	HSV and Grayscale			RGB and Grayscale			HSV		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Apple Braeburn	0,95	0,76	0,85	1,00	0,68	0,81	0,81	0,79	0,80
Apple Crimson Snow	0,83	1,00	0,91	0,98	0,95	0,96	1,00	0,74	0,85
Apple Golden 1	1,00	0,78	0,88	0,98	1,00	0,99	0,99	1,00	1,00
Apple Golden 2	1,00	1,00	1,00	1,00	1,00	1,00	0,90	1,00	0,95
Apple Golden 3	0,95	1,00	0,98	0,90	1,00	0,95	0,94	0,91	0,92
Apple Granny Smith	1,00	1,00	1,00	1,00	1,00	1,00	0,99	0,88	0,94
Apple Pink Lady	0,84	1,00	0,92	0,97	1,00	0,99	0,75	1,00	0,86
Apple Red 1	0,76	1,00	0,86	0,95	1,00	0,98	0,86	1,00	0,92
Apple Red 2	0,99	0,88	0,94	1,00	1,00	1,00	0,82	0,93	0,87
Apple Red 3	1,00	0,97	0,99	0,99	1,00	0,99	1,00	1,00	1,00
Apple Red Delicious	0,97	1,00	0,98	0,98	1,00	0,99	0,97	1,00	0,99
Apple Red Yellow 1	1,00	0,99	1,00	0,89	1,00	0,94	0,99	0,81	0,89
Apple Red Yellow 2	1,00	0,98	0,99	1,00	1,00	1,00	0,99	0,87	0,93
Apricot	0,96	1,00	0,98	0,98	1,00	0,99	0,95	1,00	0,98
Avocado	0,99	1,00	1,00	1,00	1,00	1,00	0,98	0,91	0,95
Avocado ripe	1,00	1,00	1,00	0,97	1,00	0,98	1,00	1,00	1,00
Banana	0,81	0,75	0,78	1,00	0,67	0,80	0,78	0,74	0,76
Banana Lady Finger	0,78	0,91	0,84	0,91	0,89	0,90	0,80	0,83	0,81
Banana Red	0,97	0,67	0,79	0,92	0,93	0,92	0,98	0,61	0,76
Beetroot	0,96	0,59	0,73	0,94	0,99	0,96	0,82	0,57	0,67
Blueberry	0,92	1,00	0,96	1,00	1,00	1,00	1,00	1,00	1,00
Cactus fruit	0,84	1,00	0,91	0,82	1,00	0,90	0,94	1,00	0,97
Cantaloupe 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cantaloupe 2	1,00	1,00	1,00	0,95	1,00	0,98	1,00	1,00	1,00
Carambula	1,00	0,90	0,95	1,00	0,95	0,97	0,95	1,00	0,98
Cauliflower	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry 2	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,99
Cherry Rainier	1,00	0,80	0,89	0,94	0,79	0,86	1,00	1,00	1,00
Cherry Wax Black	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry Wax Red	1,00	1,00	1,00	1,00	1,00	1,00	0,95	1,00	0,97
Cherry Wax Yellow	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Chestnut	0,98	0,84	0,90	0,98	0,97	0,98	0,99	0,95	0,97
Clementine	1,00	0,77	0,87	1,00	0,81	0,90	1,00	0,89	0,94
Cocos	0,96	0,94	0,95	0,99	0,95	0,97	0,87	0,90	0,88
Corn	0,94	0,48	0,63	0,81	0,44	0,57	0,99	0,53	0,69
Corn Husk	0,92	0,65	0,76	0,99	0,71	0,83	0,71	0,62	0,66
Cucumber Ripe	0,86	1,00	0,93	1,00	1,00	1,00	1,00	1,00	1,00
Cucumber Ripe 2	1,00	0,88	0,94	1,00	0,97	0,99	0,95	0,85	0,89
Dates	0,94	1,00	0,97	0,99	1,00	1,00	0,85	1,00	0,92
Eggplant	1,00	0,76	0,87	0,93	0,73	0,82	0,92	0,71	0,80
Fig	1,00	1,00	1,00	1,00	1,00	1,00	0,91	0,99	0,95
Ginger Root	0,77	0,86	0,81	1,00	1,00	1,00	0,98	0,99	0,98
Granadilla	0,87	1,00	0,93	1,00	1,00	1,00	0,97	1,00	0,98
Grape Blue	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grape Pink	1,00	1,00	1,00	1,00	1,00	1,00	0,92	1,00	0,96

	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Grape White	0,98	1,00	0,99	0,98	1,00	0,99	1,00	1,00	1,00
Grape White 2	0,78	1,00	0,87	0,97	1,00	0,99	1,00	1,00	1,00
Grape White 3	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grape White 4	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grapefruit Pink	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grapefruit White	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Guava	0,88	1,00	0,94	1,00	1,00	1,00	0,91	1,00	0,95
Hazelnut	0,99	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00
Huckleberry	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Kaki	0,74	1,00	0,85	0,84	1,00	0,91	0,97	0,88	0,92
Kiwi	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00
Kohlrabi	1,00	1,00	1,00	1,00	1,00	1,00	0,99	0,94	0,96
Kumquats	1,00	1,00	1,00	1,00	1,00	1,00	0,89	0,95	0,92
Lemon	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,99
Lemon Meyer	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Limes	0,98	1,00	0,99	1,00	1,00	1,00	0,92	1,00	0,96
Lychee	0,83	1,00	0,90	1,00	1,00	1,00	0,94	1,00	0,97
Mandarine	1,00	1,00	1,00	1,00	0,92	0,96	1,00	0,73	0,84
Mango	1,00	1,00	1,00	1,00	1,00	1,00	0,95	1,00	0,97
Mango Red	0,98	0,86	0,91	1,00	0,96	0,98	1,00	0,94	0,97
Mangostan	1,00	1,00	1,00	1,00	1,00	1,00	0,84	1,00	0,91
Maracuja	0,83	0,84	0,83	0,99	0,92	0,95	0,81	0,91	0,86
Melon Piel de Sapo	0,94	1,00	0,97	0,99	1,00	0,99	1,00	1,00	1,00
Mulberry	0,94	1,00	0,97	1,00	1,00	1,00	0,96	1,00	0,98
Nectarine	0,77	0,67	0,72	0,70	0,80	0,75	0,76	0,49	0,59
Nectarine Flat	0,91	0,94	0,92	0,84	0,95	0,89	0,88	1,00	0,94
Nut Forest	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,98
Nut Pecan	0,95	1,00	0,98	0,99	1,00	0,99	0,97	0,87	0,91
Onion Red	0,94	1,00	0,97	0,99	1,00	1,00	0,77	1,00	0,87
Onion Red Peeled	0,99	0,94	0,96	1,00	1,00	1,00	0,91	0,90	0,91
Onion White	0,72	1,00	0,84	0,65	1,00	0,79	0,86	0,99	0,92
Orange	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Papaya	1,00	1,00	1,00	0,95	1,00	0,98	1,00	1,00	1,00
Passion Fruit	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Peach	0,72	0,84	0,78	0,90	0,80	0,85	0,74	0,71	0,72
Peach 2	0,95	1,00	0,97	0,97	1,00	0,98	1,00	1,00	1,00
Peach Flat	0,97	0,79	0,87	0,91	0,95	0,93	0,91	0,97	0,94
Pear	1,00	0,95	0,97	0,99	0,89	0,94	1,00	0,84	0,91
Pear 2	0,89	0,59	0,71	0,95	0,48	0,64	0,94	0,60	0,73
Pear Abate	0,98	0,86	0,91	0,81	0,91	0,86	1,00	0,86	0,92
Pear Forelle	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00
Pear Kaiser	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pear Monster	0,99	0,96	0,98	0,88	1,00	0,94	0,96	0,93	0,94
Pear Red	0,99	1,00	0,99	1,00	1,00	1,00	0,87	1,00	0,93
Pear Stone	0,82	1,00	0,90	0,80	1,00	0,89	0,96	1,00	0,98
Pear Williams	0,85	1,00	0,92	0,98	1,00	0,99	0,91	1,00	0,95
Pepino	0,60	0,82	0,69	0,73	0,98	0,84	0,96	0,79	0,87
Pepper Green	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pepper Orange	1,00	0,82	0,90	1,00	1,00	1,00	0,83	0,86	0,84
Pepper Red	0,97	1,00	0,98	1,00	1,00	1,00	0,86	1,00	0,92
Pepper Yellow	0,98	0,99	0,99	1,00	1,00	1,00	0,97	0,88	0,92
Physalis	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Physalis with Husk	0,90	0,99	0,94	0,88	1,00	0,93	0,92	0,99	0,95

	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Pineapple	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pineapple Mini	1,00	1,00	1,00	1,00	1,00	1,00	0,96	1,00	0,98
Pitahaya Red	0,98	1,00	0,99	1,00	1,00	1,00	0,99	1,00	1,00
Plum	1,00	1,00	1,00	1,00	1,00	1,00	0,90	1,00	0,95
Plum 2	1,00	0,95	0,97	1,00	1,00	1,00	0,89	1,00	0,94
Plum 3	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	0,99
Pomegranate	0,92	0,87	0,90	0,95	1,00	0,98	0,81	0,99	0,89
Pomelo Sweetie	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,91	0,95
Potato Red	0,94	0,93	0,93	0,87	0,97	0,92	0,97	0,93	0,95
Potato Red Washed	1,00	0,92	0,96	0,99	0,99	0,99	1,00	0,89	0,94
Potato Sweet	0,96	0,89	0,92	1,00	0,83	0,91	0,82	0,74	0,78
Potato White	0,97	0,95	0,96	1,00	1,00	1,00	0,91	0,85	0,88
Quince	1,00	0,80	0,89	1,00	1,00	1,00	0,87	1,00	0,93
Rambutan	1,00	1,00	1,00	0,98	1,00	0,99	0,88	1,00	0,93
Raspberry	1,00	1,00	1,00	1,00	1,00	1,00	0,98	1,00	0,99
Redcurrant	1,00	1,00	1,00	1,00	1,00	1,00	0,94	1,00	0,97
Salak	0,99	1,00	0,99	0,99	1,00	0,99	0,96	0,96	0,96
Strawberry	1,00	0,88	0,94	1,00	1,00	1,00	0,98	1,00	0,99
Strawberry Wedge	1,00	0,99	0,99	1,00	0,99	0,99	0,95	0,97	0,96
Tamarillo	0,96	1,00	0,98	1,00	1,00	1,00	0,93	1,00	0,96
Tangelo	1,00	1,00	1,00	0,93	1,00	0,96	0,79	1,00	0,88
Tomato 1	1,00	0,95	0,97	1,00	1,00	1,00	0,98	1,00	0,99
Tomato 2	0,92	1,00	0,96	1,00	1,00	1,00	0,96	1,00	0,98
Tomato 3	0,95	1,00	0,98	1,00	1,00	1,00	1,00	1,00	1,00
Tomato 4	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato Cherry Red	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato Heart	1,00	0,97	0,98	1,00	1,00	1,00	0,97	1,00	0,98
Tomato Maroon	0,92	1,00	0,96	1,00	1,00	1,00	0,95	1,00	0,97
Tomato not Ripened	0,95	0,97	0,96	0,99	1,00	0,99	0,97	0,87	0,92
Tomato Yellow	0,99	1,00	1,00	1,00	1,00	1,00	0,86	1,00	0,92
Walnut	1,00	1,00	1,00	0,96	1,00	0,98	1,00	1,00	1,00
Watermelon	0,77	1,00	0,87	0,96	1,00	0,98	0,89	0,99	0,94

Πίνακας 4: Μετρικές precision, recall, F1-Score για τα Πειράματα 1-3.

	RGB			Balanced RGB			ResNet		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Apple Braeburn	1,00	0,71	0,83	0,93	0,92	0,93	1,00	1,00	1,00
Apple Crimson Snow	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Apple Golden 1	1,00	1,00	1,00	0,97	1,00	0,98	1,00	1,00	1,00
Apple Golden 2	0,88	1,00	0,93	1,00	1,00	1,00	1,00	1,00	1,00
Apple Golden 3	0,95	1,00	0,98	1,00	1,00	1,00	1,00	1,00	1,00
Apple Granny Smith	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Apple Pink Lady	1,00	1,00	1,00	0,96	1,00	0,98	1,00	1,00	1,00
Apple Red 1	0,89	1,00	0,94	0,95	1,00	0,97	1,00	1,00	1,00
Apple Red 2	0,93	1,00	0,96	1,00	0,93	0,97	1,00	1,00	1,00
Apple Red 3	0,99	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00
Apple Red Delicious	1,00	1,00	1,00	0,97	1,00	0,99	1,00	1,00	1,00
Apple Red Yellow 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Apple Red Yellow 2	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Apricot	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Avocado	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Avocado ripe	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Banana	0,99	0,81	0,89	1,00	0,74	0,85	1,00	1,00	1,00
Banana Lady Finger	0,96	0,87	0,91	0,87	0,81	0,84	1,00	1,00	1,00
Banana Red	0,92	0,94	0,93	1,00	0,92	0,96	1,00	1,00	1,00
Beetroot	0,93	0,99	0,96	0,93	0,74	0,83	1,00	1,00	1,00
Blueberry	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cactus fruit	0,89	1,00	0,94	0,89	1,00	0,94	1,00	1,00	1,00
Cantaloupe 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cantaloupe 2	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00
Carambula	1,00	1,00	1,00	1,00	0,98	0,99	1,00	1,00	1,00
Cauliflower	0,99	1,00	0,99	0,97	1,00	0,99	1,00	1,00	1,00
Cherry 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry 2	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry Rainier	1,00	0,93	0,96	1,00	1,00	1,00	1,00	1,00	1,00
Cherry Wax Black	0,99	1,00	0,99	1,00	1,00	1,00	1,00	1,00	1,00
Cherry Wax Red	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Cherry Wax Yellow	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Chestnut	0,99	0,99	0,99	0,98	0,97	0,98	1,00	0,99	1,00
Clementine	1,00	0,92	0,96	1,00	0,87	0,93	1,00	1,00	1,00
Cocos	1,00	0,93	0,96	1,00	0,95	0,98	1,00	1,00	1,00
Corn	0,74	0,50	0,60	0,62	0,47	0,53	1,00	0,88	0,94
Corn Husk	1,00	0,73	0,84	1,00	0,71	0,83	1,00	0,86	0,92
Cucumber Ripe	0,94	1,00	0,97	0,96	1,00	0,98	1,00	1,00	1,00
Cucumber Ripe 2	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Dates	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Eggplant	1,00	0,79	0,88	0,96	0,78	0,86	0,99	0,90	0,94
Fig	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Ginger Root	1,00	0,87	0,93	1,00	0,93	0,96	0,89	1,00	0,94
Granadilla	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grape Blue	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grape Pink	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grape White	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00
Grape White 2	0,93	1,00	0,96	0,91	1,00	0,95	1,00	1,00	1,00
Grape White 3	1,00	1,00	1,00	1,00	1,00	1,00	0,98	1,00	0,99
Grape White 4	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grapefruit Pink	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Grapefruit White	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Guava	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Hazelnut	0,96	1,00	0,98	1,00	1,00	1,00	1,00	1,00	1,00
Huckleberry	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Kaki	0,92	1,00	0,96	0,88	1,00	0,94	1,00	1,00	1,00
Kiwi	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Kohlrabi	1,00	1,00	1,00	1,00	1,00	1,00	0,94	1,00	0,97
Kumquats	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Lemon	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Lemon Meyer	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Limes	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Lychee	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Mandarine	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Mango	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Mango Red	1,00	1,00	1,00	1,00	0,89	0,94	1,00	1,00	1,00
Mangostan	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Maracuja	0,99	0,87	0,93	0,99	0,99	0,99	1,00	1,00	1,00
Melon Piel de Sapo	1,00	1,00	1,00	0,99	1,00	0,99	1,00	1,00	1,00

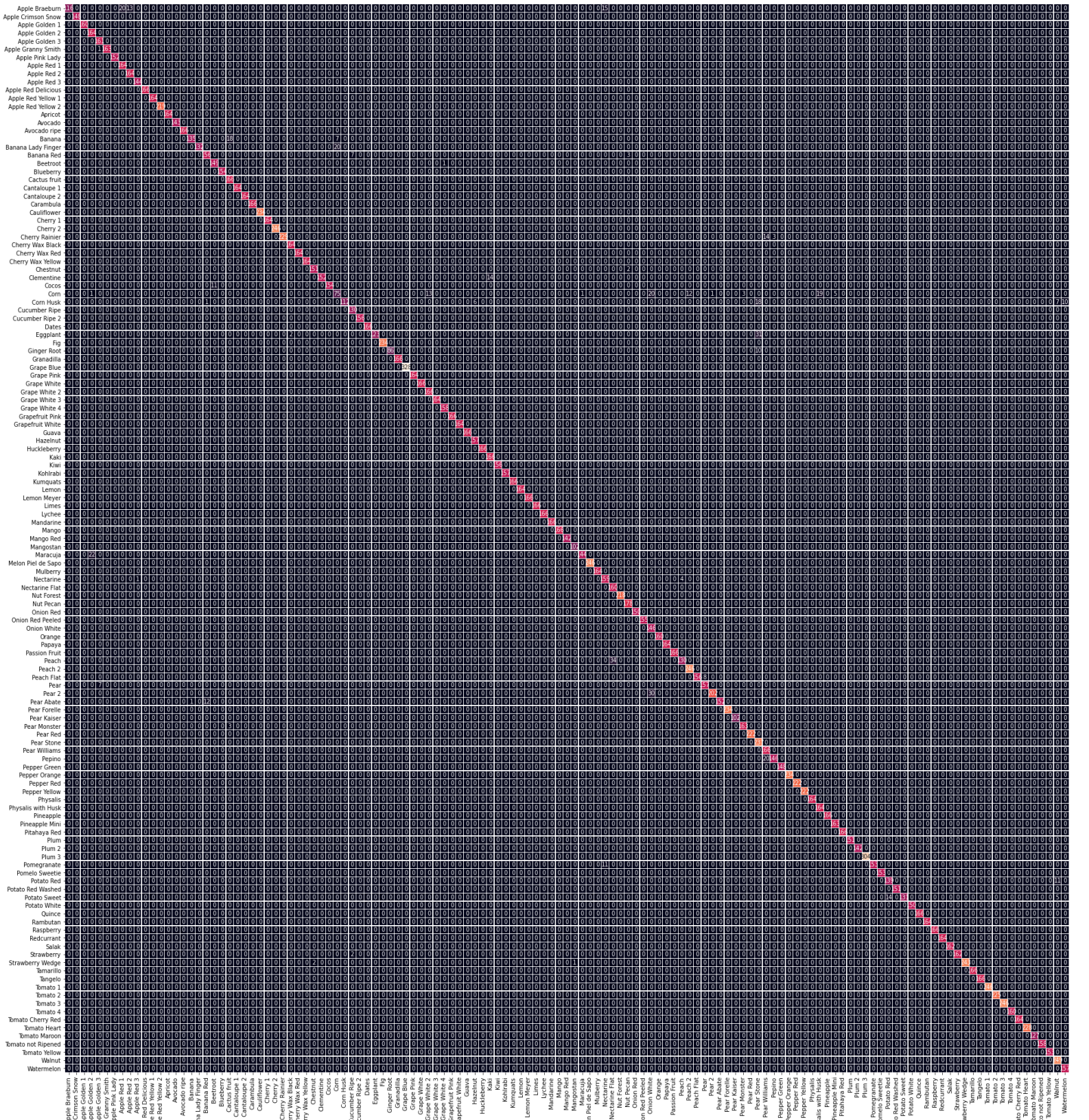
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Mulberry	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Nectarine	0,86	0,97	0,91	0,86	0,79	0,82	0,83	1,00	0,91
Nectarine Flat	0,82	1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00
Nut Forest	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Nut Pecan	0,97	1,00	0,99	0,99	1,00	1,00	1,00	1,00	1,00
Onion Red	1,00	1,00	1,00	0,99	1,00	0,99	1,00	1,00	1,00
Onion Red Peeled	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Onion White	0,74	1,00	0,85	0,73	1,00	0,84	0,99	1,00	1,00
Orange	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Papaya	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Passion Fruit	1,00	1,00	1,00	0,99	1,00	0,99	1,00	1,00	1,00
Peach	0,97	0,79	0,87	1,00	1,00	1,00	1,00	0,79	0,88
Peach 2	0,95	1,00	0,98	0,96	1,00	0,98	1,00	1,00	1,00
Peach Flat	1,00	0,95	0,97	1,00	0,65	0,79	1,00	1,00	1,00
Pear	0,95	0,96	0,95	0,95	1,00	0,97	1,00	1,00	1,00
Pear 2	1,00	0,87	0,93	0,80	0,68	0,73	1,00	1,00	1,00
Pear Abate	0,99	0,92	0,95	1,00	0,87	0,93	1,00	1,00	1,00
Pear Forelle	1,00	1,00	1,00	0,87	1,00	0,93	1,00	1,00	1,00
Pear Kaiser	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pear Monster	0,98	0,98	0,98	0,97	1,00	0,99	1,00	1,00	1,00
Pear Red	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pear Stone	0,80	1,00	0,89	0,74	1,00	0,85	0,94	1,00	0,97
Pear Williams	0,83	1,00	0,91	0,92	1,00	0,96	0,97	1,00	0,99
Pepino	0,98	0,88	0,93	0,99	0,94	0,96	1,00	1,00	1,00
Pepper Green	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pepper Orange	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,95	0,97
Pepper Red	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pepper Yellow	1,00	1,00	1,00	1,00	1,00	1,00	0,95	1,00	0,97
Physalis	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Physalis with Husk	0,90	1,00	0,95	0,88	1,00	0,94	1,00	1,00	1,00
Pineapple	1,00	1,00	1,00	0,81	1,00	0,90	1,00	1,00	1,00
Pineapple Mini	0,97	1,00	0,98	1,00	1,00	1,00	1,00	1,00	1,00
Pitahaya Red	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00
Plum	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Plum 2	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Plum 3	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Pomegranate	1,00	0,93	0,97	1,00	0,95	0,97	1,00	1,00	1,00
Pomelo Sweetie	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Potato Red	0,90	0,93	0,91	0,99	0,94	0,97	1,00	1,00	1,00
Potato Red Washed	1,00	1,00	1,00	1,00	0,98	0,99	1,00	1,00	1,00
Potato Sweet	1,00	0,87	0,93	0,99	0,99	0,99	1,00	1,00	1,00
Potato White	1,00	1,00	1,00	0,93	1,00	0,96	1,00	1,00	1,00
Quince	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Rambutan	1,00	1,00	1,00	0,99	1,00	1,00	0,99	1,00	1,00
Raspberry	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Redcurrant	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Salak	1,00	1,00	1,00	0,99	1,00	0,99	1,00	1,00	1,00
Strawberry	1,00	0,99	0,99	0,87	1,00	0,93	1,00	1,00	1,00
Strawberry Wedge	1,00	0,99	0,99	1,00	0,99	0,99	1,00	1,00	1,00
Tamarillo	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tangelo	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato 1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato 2	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Tomato 3	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato 4	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato Cherry Red	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato Heart	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Tomato Maroon	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00
Tomato not Ripened	1,00	1,00	1,00	0,96	1,00	0,98	0,97	1,00	0,98
Tomato Yellow	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Walnut	0,92	1,00	0,96	0,97	1,00	0,98	1,00	1,00	1,00
Watermelon	0,94	1,00	0,97	0,99	1,00	1,00	1,00	1,00	1,00

Πίνακας 5: Μετρικές precision, recall, F1-score για το πειράματα 4-6.

Παράστημα Γ

Παρακάτω βρίσκονται οι πίνακες σύγκρισης για τα Πειράματα 1 – 6.



Εικόνα 1: Confusion Matrix του πειράματος 4.

Εικόνα 2: Confusion matrix πειράματος 6