



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη Συστήματος Αυτοματοποιημένης Αναγνώρισης Εικόνων Τροφής
Προσαρμοσμένο στις Ανάγκες των Ατόμων με Σακχαρώδη Διαβήτη**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κλεομένης Κόγιας

Επιβλέπουσα: Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Νοέμβριος 2017



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ
ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Ανάπτυξη Συστήματος Αυτοματοποιημένης Αναγνώρισης Εικόνων Τροφής Προσαρμοσμένο στις Ανάγκες των Ατόμων με Σακχαρώδη Διαβήτη

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κλεομένης Κόγιας

Επιβλέπουσα: Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

(Υπογραφή)

.....
Κωνσταντίνα Νικήτα
Καθηγήτρια Ε.Μ.Π.

(Υπογραφή)

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2017

(Υπογραφή)

.....

Κόγιας Κλεομένης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Κόγιας Κλεομένης, 2017 – All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξολοκλήρου ή μέρους αυτής, για εμπορικό ή κερδοσκοπικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Ερωτήματα που αφορούν τη χρήση της εργασίας για εμπορικό- κερδοσκοπικό σκοπό πρέπει να απευθύνονται αποκλειστικά στους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτή την εργασία εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου συμπεριλαμβανόμενων Σχολών, Τομέων και Μονάδων αυτού.

Περίληψη

Ο υπολογισμός της περιεχόμενης ποσότητας υδατανθράκων στα γεύματα των ατόμων με σακχαρώδη διαβήτη, αποτελεί πληροφορία μείζονος σημασίας καθώς καθορίζει την μεταγευματική δοσολογία της ινσουλίνης. Η παρούσα διπλωματική εργασία παρουσιάζει ένα μοντέλο ταξινόμησης φωτογραφικών στιγμιότυπων τροφής προσαρμοσμένο στις διατροφικές ανάγκες των ατόμων με σακχαρώδη διαβήτη.

Η μέθοδος που χρησιμοποιείται για την ταξινόμηση των εικόνων στηρίζεται στα Συνελκτικά Νευρωνικά Δίκτυα. Συγκεκριμένα δημιουργείται ταξινομητής δύο επιπέδων, ο οποίος διακρίνει αρχικά τις εικόνες σε 8 ευρύτερες κατηγορίες: Γαλακτοκομικά και Προϊόντα, Φρούτα, Λαχανικά, Όσπρια και Αμυλούχα Λαχανικά, Κρέας και Υποκατάστατα, Ψωμί-Δημητριακά, Breaded Food, Λιπαρά Φαγητά και στην συνέχεια γίνεται ταξινόμηση της εικόνας ως ένα συγκεκριμένο φαγητό εντός της κατηγορίας. Η ταξινόμηση πραγματοποιείται σε σύνολο δεδομένων ειδικά διαμορφωμένο για άτομα με σακχαρώδη διαβήτη, καθώς επίσης δημιουργείται μετρική αξιολόγησης των αποτελεσμάτων ειδικά για το σκοπό αυτό.

Η μέθοδος που προτείνεται βελτιώνει τα αποτελέσματα στο σύνολο δεδομένων Sarantea-2016, σημειώνοντας ακρίβεια ταξινόμησης 97,08%. Στο νέο σύνολο δεδομένων NTUA-Food 2017 η ακρίβεια ανά κατηγορία φθάνει το 84,18% ενώ ανά τροφή το 72,34%. Όπως φαίνεται από την μετρική αξιολόγησης των αποτελεσμάτων, η πρόβλεψη του μοντέλου που παρουσιάζεται, εμφανίζει μέση απόκλιση υπολογισμού των υδατανθράκων μικρότερη των 2γρ. ανά ένα ισοδύναμο τροφής.

Συνεπώς, στο πλαίσιο της εργασίας παρουσιάζεται μοντέλο ειδικά διαμορφωμένο στις ανάγκες των ατόμων με σακχαρώδη διαβήτη. Τα Συνελκτικά Νευρωνικά Δίκτυα, το σύνολο δεδομένων NTUA-Food 2017, η ταξινόμηση δύο επιπέδων και η μετρική αξιολόγησης αποτελεσμάτων που προτείνονται αποτελούν τη βάση για την ανάπτυξη συστήματος αναγνώρισης τροφής. Η ενσωμάτωση του μοντέλου αυτού σε ένα ολοκληρωμένο αυτοματοποιημένο σύστημα πρόβλεψης υδατανθράκων θα επιτρέψει την καλύτερη αξιολόγηση του συστήματος.

Λέξεις Κλειδιά

Σακχαρώδης Διαβήτης, Συνελκτικά Νευρωνικά Δίκτυα, Προεκπαίδευση Συνελκτικού Νευρωνικού Δικτύου, Ειδική Προσαρμογή (Fine-Tuning), Σύστημα Αναγνώρισης Τροφίμων, Ταξινόμηση Εικόνας, Εξαγωγή Χαρακτηριστικών, Μείωση Διάστασης Χαρακτηριστικών, Νευρωνικά Δίκτυα, Μηχανή Διανυσμάτων Υποστήριξης, Εκτίμηση Υδατανθράκων, Σύνολο Δεδομένων Φαγητών, Sarantea-2016, NTUA-Food 2017, Ισοδύναμα Υδατάνθρακα, Ισοδύναμα Τροφών

Abstract

The accurate estimation of food carbohydrate content for people diagnosed with diabetes mellitus is of great importance, as it determines post-meal insulin dosage. The present thesis introduces a classification model for food images that is tailored to the nutritional needs of people with diabetes mellitus.

The two-level image classification algorithm leverages Convolutional Neural Networks, first to classify an image in one of 8 broad food categories: Dairy Products, Fruits, Vegetables, Legumes and Starchy Vegetables, Meat and Meat Substitutes, Bread and Grains, Breaded Foods and High-Fat Foods and then to classify it as a specific food within that category. The dataset used to train the model, as well as the metric used to evaluate it, are tailored to people with diabetes mellitus.

The proposed algorithm improves classification accuracy on the Sarantea-2016 dataset to 97.08%. It further achieves 84.18% food category classification accuracy and 72.34% food classification accuracy on the new NTUA-Food 2017 dataset. The algorithm's mean error at carbohydrate content estimation is an impressive less than 2gr. per food serving.

The method proposed is specifically designed for the needs of people with diabetes mellitus, and forms the basis for the development of a food recognition system. Incorporating the proposed model into a complete automatic carbohydrate prediction system will allow its better evaluation.

Keywords

Diabetes Mellitus, Convolutional Neural Networks, Pre-training Convolutional Neural Network, Fine-Tuning, Food Recognition System, Image Classification, Feature Extraction, Dimensionality Reduction, Neural Networks, Support Vector Machines, Carbohydrate Estimation, Food Dataset, Sarantea-2016, NTUA-Food 2017, Carbohydrate Equivalent, Food Equivalent

Ευχαριστίες

Γράφοντας αυτό το σημείωμα μπαίνουν οι τελευταίες λέξεις στο έγγραφο που αποκαλείται προπτυχιακή διπλωματική εργασία. Η εργασία αυτή αποτελεί ένα αποτέλεσμα για το οποίο είμαι πολύ περήφανός καθώς κατέθεσα πολύ προσπάθεια για να δημιουργηθεί και αισθάνομαι ότι συνέβαλα με το δικό μου τρόπο όχι μόνο σε ένα επιστημονικό τομέα αλλά σε ένα τομέα που καθορίζει και θα καθορίσει τη δική μου υπόλοιπη ζωή, τον σακχαρώδη διαβήτη.

Θα ήθελα να ευχαριστήσω θερμά την καθηγήτρια Κωνσταντίνα Νικήτα η οποία με εμπιστεύτηκε με αυτή την διπλωματική εργασία και μου έδωσε την ευκαιρία να συνδυάσω τη Πληροφορική με την Βιοϊατρική. Επίσης θα ήθελα να ευχαριστήσω τους υποψήφιους διδάκτορες Γιάννη Ανδρεάδη και την Καλλιόπη Δαλακλείδη, οι οποίοι στάθηκαν δίπλα μου καθ' όλη τη διάρκεια πραγμάτωσης αυτής της εργασίας, με πολλή υπομονή, πολλές συναντήσεις και με πολλή διάθεση για να παραχθεί το βέλτιστο αποτέλεσμα.

Για την συγγραφή της διπλωματικής εργασίας θέλω επίσης να ευχαριστήσω δύο από τους πιο στενούς μου φίλους, τον Σταύρο Τσαγκάρη και τον Νικόλα Πλατιά. Ο πρώτος με τις ιατρικές του γνώσεις μου επέτρεψε να έχω μια βαθύτερη κατανόηση της νόσου του σακχαρώδη διαβήτη. Ο Νικόλας από την άλλη μεριά με τις γνώσεις του στην Πληροφορική και στην Μηχανική Μάθηση, με συμβούλευσε σε πιο τεχνικά ζητήματα. Επίσης δεν γίνεται να παραλείψω την αδερφή μου, Ηλέκτρα, η οποία και η ίδια με τις ιατρικές τις γνώσεις συνέβαλε στην εργασία.

Με το τέλος αυτής της διπλωματικής εργασίας δεν σημαίνεται μόνο η ολοκλήρωση της σχολής των Ηλεκτρολόγων Μηχανικών και Μηχανολόγων Μηχανικών αλλά και ένα μεγάλο κεφάλαιο της ζωής μου. Θα ήθελα θερμά να ευχαριστήσω πρώτα από όλους την οικογένεια μου, τον πατέρα μου, την μητέρα μου και την αδερφή μου, γιατί καθ' όλη την πορεία της ζωής μου έχουν δώσει πάρα πολλά για μένα. Επίσης θέλω να ευχαριστήσω όλους τους ανθρώπους με τους οποίους πέρασα πολλά μαζί τους αυτά τα χρόνια και με έχουν διαμορφώσει σε αυτό που είμαι σήμερα. Ιδίως θέλω να ευχαριστήσω τους στενούς μου φίλους, την παρέα μου, με τους οποίους ήρθαμε πολύ κοντά αυτά τα χρόνια και δημιουργήσαμε κάτι το οποίο ελπίζω να κρατήσει για πάντα.

Περιεχόμενα

1	Εισαγωγή.....	17
1.1	Κίνητρο	17
1.2	Ανασκόπηση	17
1.3	Συνεισφορά της Διπλωματικής Εργασίας.....	18
1.4	Δομή.....	18
2	Σακχαρώδης Διαβήτης.....	20
2.1	Ορισμός.....	20
2.2	Τύποι Σακχαρώδη Διαβήτη.....	20
2.2.1	Σακχαρώδης Διαβήτης Τύπου 1 (ΣΔΤ1).....	20
2.2.2	Σακχαρώδης Διαβήτης τύπου 2 (ΣΔΤ2)	20
2.2.3	Σακχαρώδης Διαβήτης Κύησης	21
2.3	Διάγνωση και Θεραπεία.....	21
2.3.1	Διάγνωση	21
2.3.2	Θεραπεία.....	21
2.4	Επιπλοκές του Σακχαρώδη Διαβήτη.....	22
2.4.1	Οξείες Επιπλοκές.....	22
2.4.2	Χρόνιες Επιπλοκές.....	23
2.5	Στατιστικά Στοιχεία	24
2.6	Γλυκόζη.....	26
2.6.1	Ομοιόσταση Γλυκόζης.....	26
3	Βιβλιογραφική Ανασκόπηση	30
3.1	Καταμέτρηση μακροθρεπτικών συστατικών	30
3.1.1	Συμβατικός Τρόπος Καταγραφής Διατροφικού Περιεχομένου	30
3.1.2	Ηλεκτρονική Καταγραφή Διατροφικού Περιεχομένου	31
3.2	Προετοιμασία Περιβάλλοντος	33
3.3	Προ-επεξεργασία Εικόνας	34

3.4	Κατάτμηση της Εικόνας.....	35
3.5	Ταξινόμηση Εικόνας.....	37
3.5.1	Εξαγωγή Χαρακτηριστικών.....	37
3.5.2	Μείωση Διάστασης Χαρακτηριστικών.....	40
3.5.3	Ταξινόμηση.....	42
3.5.4	Συνελικτικά Νευρωνικά Δίκτυα.....	43
3.6	Υπολογισμός Όγκου.....	45
3.7	Συζήτηση.....	45
4	Μεθοδολογία.....	48
4.1	Νευρωνικά Δίκτυα.....	48
4.1.1	Το Perceptron.....	48
4.1.2	Δομή Νευρωνικού Δικτύου.....	50
4.1.3	Εκπαίδευση Νευρωνικού Δικτύου.....	51
4.2	Συνελικτικά Νευρωνικά Δίκτυα.....	62
4.2.1	Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου.....	62
4.2.2	Λειτουργίες Συνελικτικών Νευρωνικών Δικτύων.....	75
4.2.3	Γνωστά Συνελικτικά Νευρωνικά Δίκτυα.....	77
4.3	Μηχανές Διανυσμάτων Υποστήριξης.....	81
4.3.1	Ταξινόμηση Μηχανών Διανυσμάτων Υποστήριξης πολλών κλάσεων.....	86
4.4	Μέθοδοι Επικύρωσης.....	87
4.4.1	Μέθοδος Παρακράτησης (Holdout Method).....	87
4.4.2	Διασταυρωμένη Επικύρωση σε k-μέρη (k-fold Cross Validation).....	87
4.5	Μέτρα Αξιολόγησης.....	88
4.5.1	Ακρίβεια.....	88
4.5.2	Μήτρα Σύγχυσης (Confusion Matrix).....	88
4.5.3	Μετρική βασισμένη στην Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων.....	90
5	Σύνολο Δεδομένων.....	94

5.1	Γνωστά Σύνολα Δεδομένων.....	94
5.2	Ανάγκη δημιουργίας νέου συνόλου δεδομένων.....	96
5.3	Σύνολο Δεδομένων Sarantea-2016	97
5.4	Υπολογισμός δοσολογίας ινσουλίνης.....	98
5.4.1	Ισοδύναμα Υδατάνθρακα.....	98
5.4.2	Ισοδύναμα Λίπους και Πρωτεΐνης.....	99
5.5	Ισοδύναμα Τροφών	100
5.6	Νέο Σύνολο Δεδομένων – NTUA Food 2017.....	101
6	Ανάπτυξη Συστήματος Αυτόματης Αναγνώρισης Τροφής.....	107
6.1	Ταξινόμηση στο Σύνολο Δεδομένων Sarantea-2016.....	107
6.1.1	Εξαγωγή χαρακτηριστικών από Συνελκτικό Νευρωνικό Δίκτυο	107
6.1.2	Fine-Tuning.....	111
6.1.3	Αποτελέσματα.....	116
6.2	Ταξινόμηση στο Νέο Σύνολο Δεδομένων NTUA-Food 2017	117
6.2.1	Ταξινόμηση Επιπέδου 1	118
6.2.2	Ανάλυση Σφαλμάτων.....	123
6.2.3	Ταξινόμηση Επιπέδου 2.....	125
6.2.4	Αποτελέσματα.....	128
7	Συμβολή της Εργασίας και Προτεινόμενη Μελλοντική Έρευνα.....	132
7.1	Συμβολή της Εργασίας.....	132
7.2	Προτάσεις για περαιτέρω Έρευνα.....	133
8	Παραρτήματα.....	136
8.1	Παράρτημα Α: Πίνακας Τεχνικών Εξαγωγής Χαρακτηριστικών.....	136
8.2	Παράρτημα Β: Πίνακας Σύγκρισης Σημαντικότερων Δημοσιεύσεων.....	138
9	Βιβλιογραφία	141

Ευρετήριο Εικόνων

Εικόνα 2.1. Πλήθος ανθρώπων με διαβήτη στον κόσμο ανά περιοχή το έτος 2015 και το έτος 2040 (International Diabetes Federation, 2015)	25
Εικόνα 2.2. Αναπαράσταση Ομοιοστατικού Μηχανισμού Γλυκόζης (2014).....	28
Εικόνα 2.3. Αναπαράσταση των ομοιοστατικών μηχανισμών (Orhardt, 2003).....	29
Εικόνα 3.1. . Τα βασικά βήματα επεξεργασίας εικόνας φαγητού για τον υπολογισμό της περιεκτικότητας του σε υδατάνθρακες.	32
Εικόνα 3.2 Αναπαράσταση των βημάτων εξαγωγής χαρακτηριστικών από την εφαρμογή GoCarb (Anthimopoulos M. , et al., 2015).....	33
Εικόνα 3.3. Χρήση του αντίχειρα ως μέτρο αναφοράς για τον υπολογισμό των κανονικών διαστάσεων των τροφών (Pouladzadeh, Yassine, & Shirmohammadi, 2015).....	33
Εικόνα 3.4. Βαθμονόμηση κάμερας με σκακιέρα, με σκοπό τον υπολογισμό των διαστάσεων των νομισμάτων της εικόνας (MathWorks, n.d.).....	34
Εικόνα 3.5 Το φαινόμενο του specular highlight.....	35
Εικόνα 3.6. Αλγόριθμος κατάτμησης των φαγητών που υπάρχουν σε μια εικόνα (α) αρχική εικόνα (β) η εικόνα μετά το mean-shift-filtering (γ) εικόνα μετά το Region Growing (δ) εικόνα μετά το Region Merging (Anthimopoulos M. , Dehais, Diem, & Mougiakakou, 2013)	37
Εικόνα 3.7. Σχηματική αναπαράσταση του αλγορίθμου HOG (α) διαχωρισμός της εικόνας σε μπλοκ μέσω πλέγματος για την εξαγωγή τοπικών χαρακτηριστικών (β) Διανυσματική αναπαράσταση της HOG τιμής κάθε εικονοστοιχείου (γ) πίνακας πλάτους και διεύθυνσης για κάθε εικονοστοιχείο (δ) οπτικοποίηση της μεθόδου HOG (Satya, 2016).....	39
Εικόνα 3.8. Επιλογή σημείων εξαγωγής χαρακτηριστικών με την μέθοδο SIFT (α) Αφού έχουν εντοπισθεί τα scale-space extrema (β) απόρριψη των σημείων χαμηλής αντίθεσης (γ) απομάκρυνση σημείων κοντά στις ακμές (Wikipedia c. , 2017).....	39
Εικόνα 3.9. (α) Δημιουργία Οπτικού Λεξιλογίου από την ομαδοποίηση των διανυσμάτων χαρακτηριστικών του συνόλου εκπαίδευσης (β) Περιγραφή κάθε εικόνας από το οπτικό λεξιλόγιο που έχει παραχθεί (MathWorks, n.d.).....	41
Εικόνα 3.10. Αρχιτεκτονική εξαγωγής χαρακτηριστικών μέσω της ακολουθίας Εξαγωγή Χαρακτηριστικών- Μείωση Διάστασης Χαρακτηριστικών- Ταξινόμηση (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014)	42
Εικόνα 3.11. Σχήμα εξαγωγής χαρακτηριστικών μέσω Συνελκτικών Νευρωνικών Δικτύων (Martinel, Piciarelli, Micheloni, & Foresti, 2015)	44
Εικόνα 4.1. Το Perceptron του Rosenblatt (Raschka, 2015).....	48
Εικόνα 4.2. Γραμμικός Διαχωρισμός του χώρου από το Perceptron (Haykin, 2010).....	49

Εικόνα 4.3. Γραμμικά διαχωρίσιμες λογικές συναρτήσεις (Stanford, n.d.).....	49
Εικόνα 4.4. Λογικός τελεστής XOR (Stanford, n.d.).....	50
Εικόνα 4.5. (α) νευρωνικό δίκτυο για την επίλυση του προβλήματος του XOR (Neel, 2015) (β) Το υπερπίεδο απόφασης που δημιουργεί το νευρωνικό δίκτυο.....	50
Εικόνα 4.6. Αναπαράσταση νευρωνικού δικτύου 3 επιπέδων, αποτελούμενο από δύο κρυφά επίπεδα και ένα επίπεδο εξόδου (Nielsen, 2015)	51
Εικόνα 4.7. Νευρώνας j (Haykin, 2010).....	54
Εικόνα 4.8. Σύνδεση δύο νευρώνων διαφορετικού επιπέδου, με το νευρώνα j να ανήκει σε κρυφό επίπεδο ενώ τον νευρώνα k να ανήκει στο επίπεδο εξόδου (Haykin, 2010).....	56
Εικόνα 4.9. Σχηματική αναπαράσταση των μεταβλητών από τις οποίες επηρεάζεται το σφάλμα σε νευρώνα του κρυφού επιπέδου (Haykin, 2010).....	58
Εικόνα 4.10. Βαθμωτή κατάβαση στο χώρο των βαρών (Raschka, Machine Learning FAQ, n.d.).....	59
Εικόνα 4.11. Επιπτώσεις των διαφορετικών ρυθμών μάθησης στην συνάρτηση σφάλματος (Karpathy, Neural Networks Part 3: Learning and Evaluation, 2017).....	59
Εικόνα 4.12. Βαθμωτή κατάβαση με την προσθήκη του όρου της ορμής.....	60
Εικόνα 4.13. Αντιπροσωπευτική αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου (Britz, 2015).....	62
Εικόνα 4.14. Βασικά στρώματα Συνελκτικού Νευρωνικού Δικτύου	63
Εικόνα 4.15. (α) Πίνακας αναπαράστασης εικόνας στην κλίμακα του γκρι (β) η αντίστοιχη εικόνα.....	64
Εικόνα 4.16. (α) Αναπαράσταση φίλτρου σε πίνακα (β) οπτικοποίηση του πίνακα (Desphande, 2016)...	65
Εικόνα 4.17. Εσωτερικό γινόμενο φίλτρου που εφαρμόζεται στην εικόνα με χωρίο της εικόνας, το οποίο επιστρέφει μεγάλη απόκριση (Desphande, 2016).....	65
Εικόνα 4.18. Εσωτερικό γινόμενο χωρίου της εικόνας με φίλτρο, το οποίο επιστρέφει μηδενική απόκριση (Desphande, 2016)	66
Εικόνα 4.19. Στήλη βάθους συνελκτικού νευρωνικού δικτύου (Karpathy, Convolutional Neural Networks, 2017)	66
Εικόνα 4.20. Αναπαράσταση συνέλιξης σε πίνακα $5 \times 5 \times 3$, με μια σειρά παραγεμίματος με 0, όταν εφαρμόζονται δύο φίλτρα W_0 και W_1 , διαστάσεων $3 \times 3 \times 3$ σε αυτό. Με πράσινο απεικονίζεται ο χάρτης ενεργοποίησης $3 \times 3 \times 2$ που παράγεται (Karpathy, Convolutional Neural Networks, 2017).....	70
Εικόνα 4.21. Γραφική παράσταση της συνάρτησης ReLU, με τύπο $Output = Max(zero, Input)$	71
Εικόνα 4.22. (α) Λογιστική συνάρτηση ενεργοποίησης (β) Συνάρτηση Υπερβολικής εφαπτομένης (Karpathy, 2017).....	72
Εικόνα 4.23. Βελτιωμένος ρυθμός μάθησης που επιτυγχάνεται μέσω της χρήσης της συνάρτησης ενεργοποίησης ReLU (συνεχόμενη γραμμή), συγκριτικά με την συνάρτηση της υπερβολικής εφαπτομένης (διακεκομμένη γραμμή). Πεδίο εφαρμογής των παραπάνω συναρτήσεων είναι συνελκτικό νευρωνικό	

δίκτυο 4 επιπέδων που εκπαιδεύεται στο σύνολο δεδομένων CIFAR-10 (Krizhevsky, Sutskever, & Hinton, 2012).	73
Εικόνα 4.24. Εφαρμογή συνάρτησης ReLU σε εικόνα (Fergus, 2015)	73
Εικόνα 4.25. (α) η μείωση των διαστάσεων που επιτυγχάνεται μέσω του Στρώματος Δειγματοληψίας (β) Αναπαράσταση του max-pooling χρησιμοποιώντας φίλτρο 2x2 και βήμα 2 (Karpathy, Convolutional Neural Networks, 2017).	74
Εικόνα 4.26. Διαφορετικοί χάρτες ενεργοποίησης και διαφορετικά αποτελέσματα ταξινόμησης που επιτυγχάνονται στην περίπτωση του pre-trained και του fine-tuned μοντέλου. (Ao & Ling, 2015)	77
Εικόνα 4.27. Αναπαράσταση της διαδικασίας του Local Response Normalization (Normalization in Neural Networks, 2016).	78
Εικόνα 4.28. Αρχιτεκτονική του Συνελκτικού Νευρωνικού Δικτύου AlexNet (Krizhevsky, Sutskever, & Hinton, 2012).	79
Εικόνα 4.29. Αρχιτεκτονική του VGGnet (Cord, 2016)	81
Εικόνα 4.30. Απεικόνιση γραμμικού διαχωρισμού από SVM, δείχνοντας το περιθώριο διαχωρισμού και την συνάρτηση του υπερεπιπέδου που διαχωρίζει τις δύο κλάσεις (Caragea, Cook, & Honavar, 2005)	82
Εικόνα 4.31. Γεωμετρική ερμηνεία της απόστασης σημείου από το βέλτιστο υπερεπίπεδο (Haykin, 2010)	83
Εικόνα 4.32. Οπτικοποίηση των σημαντικότερων μεγεθών του SVM (Wikipedia, n.d.)	84
Εικόνα 4.33. (Αριστερά) Ένα σύνολο δεδομένων μη γραμμικά διαχωρίσιμο στον δισδιάστατο χώρο. (Δεξιά) Το ίδιο σύνολο δεδομένων στον τρισδιάστατο χώρο μέσω της μετατροπής $[x_1, x_2] = [x_1 \ x_2 \ x_1^2 + x_2^2]$	85
Εικόνα 4.34.(Αριστερά) Υπερεπίπεδο που διαχωρίζει τις δύο κλάσεις στον τρισδιάστατο χώρο (Δεξιά) Μετατροπή του υπερεπιπέδου στον δισδιάστατο χώρο (Kim, 2013)	85
Εικόνα 4.35.Αναπαράσταση Μήτρας Σύγχυσης	89
Εικόνα 4.36. Διαγραμματική αναπαράσταση αντιστοίχισης συγκεκριμένων μακροθρεπτικών συστατικών σε ι μονάδα ινσουλίνης, ανάγοντας τα τελικώς σε 15γρ. υδατανθράκων.	91
Εικόνα 4.37. Μήτρα Αξιολόγησης	92
Εικόνα 4.38. Διαδικασία εύρεσης της τιμής κάθε κελιού της μήτρας αξιολόγησης	92
Εικόνα 5.1. Γνωστά Σύνολα Δεδομένων και οι βασικές πληροφορίες που σχετίζονται με αυτά. Στην στήλη Εικόνες ανά Κατηγορία αναγράφονται κατά μέσο όρο οι εικόνες ανά κατηγορία για κάθε φαγητό.	94
Εικόνα 5.2. Φωτογραφία από κάθε κατηγορία του συνόλου δεδομένων (α) Hamburger (Breaded Food), (β) Γάλα (Γαλακτοκομικά και Προϊόντα) (γ) Μπριζόλα (Κρέας και Υποκατάστατα), (δ) Μπρόκολο (Λαχανικά) (ε) Brownie (Λιπαρά Φαγητά) (ζ) Φασολάκια (Όσπρια και Αμυλούχα Λαχανικά) (η) Μήλο (Φρούτα) (θ) Ψωμί Ολικής Αλέσεως (Ψωμί- Δημητριακά)	104

Εικόνα 6.1. Δομή του AlexNet	108
Εικόνα 6.2. Δομή VGG-19 Net.....	110
Εικόνα 6.3. Δομή fine-tuned Συνελκτικού Νευρωνικού Δικτύου	112
Εικόνα 6.4. Διαγραμματική Αναπαράσταση της ακρίβειας που επιτυγχάνεται στις πρώτες 10 εποχές στις περιπτώσεις διαχωρισμού του συνόλου δεδομένων σε 40%-60%, 60%-40% και 80%-20%.....	115
Εικόνα 6.5. Ταξινόμηση δύο επιπέδων	117
Εικόνα 6.6. (α) Πίνακας ακρίβειας των 13 εποχών εκπαίδευσης (β) Το αντίστοιχο γράφημα αναπαράστασης της ακρίβειας συναρτήσει της εποχής.....	119
Εικόνα 6.7. Διαγραμματική αναπαράσταση του ποσοστού των σωστά αναγνωρισμένων εικόνων ανά εποχή και ανά κατηγορία.....	120
Εικόνα 6.8. Μήτρα Αξιολόγησης	121
Εικόνα 6.9. Διαγραμματική αναπαράσταση του αποτελέσματος της μετρικής ανά εποχή	121
Εικόνα 6.10. Μήτρα Σύγχυσης κατηγοριών που δεν αποτελούνται από ίσο πλήθος εικόνων	126

Ευρετήριο Πινάκων

Πίνακας 4.1. Πίνακας αναγωγής των ισοδυνάμων κάθε ομάδας φαγητών σε μονάδες ινσουλίνης και αντίστοιχα σε γραμμάρια υδατάνθρακα.....	91
Πίνακας 5.1. Ισοδύναμο πρωτεΐνης και λίπους	100
Πίνακας 5.2. Αντιστοίχιση των ισοδυνάμων πρωτεΐνης και λίπους με ένα ισοδύναμο υδατάνθρακα	100
Πίνακας 5.3. Ισοδύναμα των βασικών κατηγοριών τροφών. Οι συντομογραφίες (χ.π.λ), (μ.π.λ),(υ.π.λ) αντιστοιχούν σε χαμηλή, μέτρια και υψηλή περιεκτικότητα σε λίπος.	101
Πίνακας 5.4. Κατηγοριοποιημένα τα φαγητά που αποτελούν το νέο σύνολο δεδομένων στις 8 βασικές ομάδες: Γάλα και Προϊόντα, Λαχανικά, Φρούτα, Όσπρια και Αμυλούχα Λαχανικά, Κρέας και Υποκατάστατα, Breaded Food, Ψωμί-Δημητριακά, Λιπαρά Φαγητά	106
Πίνακας 6.1. Μήτρα Σύγκρισης χρησιμοποιώντας εξαγωγή χαρακτηριστικών από το AlexNet	109
Πίνακας 6.2. Μήτρα Σύγκρισης χρησιμοποιώντας εξαγωγή χαρακτηριστικών από το VGG-19 Net.....	111
Πίνακας 6.3. Μήτρα Σύγκρισης της 8 ^{ης} εποχής στο σχήμα 40%-60% του συνόλου δεδομένων.....	113
Πίνακας 6.4. Μήτρα Σύγκρισης της 6 ^{ης} εποχής στο σχήμα 40%-60% του συνόλου δεδομένων.....	113
Πίνακας 6.5. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 40-60% του συνόλου δεδομένων	113
Πίνακας 6.6. Μήτρα Σύγκρισης της 7 ^{ης} εποχής στο σχήμα 60%-40% του συνόλου δεδομένων.....	114
Πίνακας 6.7. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 60%-40% του συνόλου δεδομένων	114
Πίνακας 6.8. Μήτρα Σύγκρισης της 7 ^{ης} εποχής στο σχήμα 80%-20% του συνόλου δεδομένων.....	114
Πίνακας 6.9. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 80%-20% του συνόλου δεδομένων	114
Πίνακας 6.10. Παρουσίαση αποτελεσμάτων για τις διαφορετικές τεχνικές ταξινόμησης	117
Πίνακας 6.11. Παρουσιάζεται το ποσοστό ακρίβειας που επιτυγχάνεται ανά κατηγορία φαγητών ανά εποχή. Με κόκκινο σημειώνεται η βέλτιστη ακρίβεια που επιτυγχάνεται ανά κατηγορία, με πράσινο χρωματίζεται η βέλτιστη ακρίβεια ανά ομάδα μεταξύ των εποχών 8 και 10, ενώ με μπλε σημειώνεται η καλύτερη ακρίβεια ανάμεσα στις εποχές 10 και 12.	120
Πίνακας 6.12. Στις πρώτες δύο στήλες γίνεται παρουσίαση των αποτελεσμάτων της μετρικής ανά εποχή. Στην τρίτη στήλη παρουσιάζεται το πλήθος σφαλμάτων από τις 1296 συνολικά εικόνες. Στην τέταρτη στήλη φαίνεται το μέσο σφάλμα ανά εποχή, με κόκκινο χρωματίζονται τα κελιά με το υψηλότερο μέσο σφάλμα, ενώ με μπλε αυτά με το μικρότερο.	121
Πίνακας 6.13. Παρουσίαση των αποτελεσμάτων της μετρικής ανά εποχή. Συμπληρωματικά παρατίθενται η μέση απόκλιση των υδατανθράκων ανά φαγητό και το πλήθος φαγητών που κατά μέσο όρο χρειάζονται για απόκλιση μέχρι 10γρ και μέχρι 20γρ. υδατανθράκων αντιστοίχως.	122

Πίνακας 6.14. Μήτρα Σύγκυσης της εποχής 10.....	122
Πίνακας 6.15. Μήτρα Σύγκυσης της εποχής 10 που αναπαριστά τα αποτελέσματα της ταξινόμησης σε μορφή πλήθους εικόνων και όχι με ποσοστιαία αναπαράσταση	123
Πίνακας 6.16.Επαυξημένη Μήτρα Σύγκυσης της Εποχής 10. Στο Άθροισμα κατά στήλη και κατά γραμμή δεν εμπεριέχονται τα στοιχεία της κυρίας διαγωνίου	125
Πίνακας 6.17. Μήτρα Σύγκυσης των φαγητών της κατηγορίας του Breaded Food, το σύνολο ελέγχου αποτελείται από 11 κατηγορίες για κάθε τρόφιμο.	127
Πίνακας 6.18. Ταξινόμηση με διαφορετικές συναρτήσεις πυρήνα SVM στην ομάδα των Breaded Food	128
Πίνακας 6.19. Ακρίβεια ταξινόμησης δεύτερου επιπέδου.....	128
Πίνακας 6.20. Μήτρα Σύγκυσης για τα φαγητά της κατηγορίας Breaded Food.....	129
Πίνακας 6.21. Μήτρα Σύγκυσης για τα τρόφιμα της κατηγορίας Φρούτα	130
Πίνακας 6.22. Μήτρα Σύγκυσης για τα φαγητά της κατηγορίας Λιπαρά Φαγητά	130
Πίνακας 6.23.Μήτρα Σύγκυσης για τα φαγητά της κατηγορίας Όσπρια και Αμυλούχα Λαχανικά	130
Πίνακας 6.24. Μήτρα Σύγκυσης για τα φαγητά της κατηγορίας Κρέας και Υποκατάστατα	130
Πίνακας 6.25. Μήτρα Σύγκυσης για τα τρόφιμα της κατηγορίας Γαλακτοκομικά Προϊόντα	130
Πίνακας 6.26. Μήτρα Σύγκυσης για τα τρόφιμα της κατηγορίας Λαχανικά.....	131
Πίνακας 6.27. Μήτρα Σύγκυσης για τα τρόφιμα της κατηγορίας Ψωμί-Δημητριακά.....	131
Πίνακας 8.1. Αναλυτική παρουσίαση όλων των τεχνικών εξαγωγής χαρακτηριστικών που συναντήθηκαν στην διπλωματική εργασία. Με πράσινο χρώμα σημειώνονται οι πιο γνωστές μέθοδοι εξαγωγής χαρακτηριστικών	137
Πίνακας 8.2. Παρουσίαση των σημαντικότερων δημοσιεύσεων και των σχημάτων ταξινόμησης που ακολουθούν. Στην στήλη της ακρίβειας αναγράφεται η βέλτιστη ακρίβεια που επιτυγχάνεται σε κάθε δημοσίευση, ενώ με bold εμφανίζεται η τεχνική η οποία σημειώνει την ακρίβεια αυτή.	140

1 Εισαγωγή

1.1 Κίνητρο

Ο σακχαρώδης διαβήτης αποτελεί μια διαδεδομένη ασθένεια, επιφέροντας μακροχρόνια έως και θανάσιμες συνέπειες στους νοσούντες. Δεν έχει βρεθεί ακόμα θεραπεία για το σακχαρώδη διαβήτη, παρά μόνο τρόποι διαχείρισης του. Οι τρόποι αυτοί επικεντρώνονται στην διατήρηση του σακχάρου κοντά στα φυσιολογικά επίπεδα, με σκοπό να αποφευχθούν οι οξείες επιπτώσεις του και να καθυστερήσει η εμφάνιση των μακροχρόνιων. Ο σημαντικότερος τρόπος για να γίνει αυτό από τα άτομα με σακχαρώδη διαβήτη, είναι ο υγιεινός τρόπος ζωής, ο οποίος στηρίζεται στην υγιεινή διατροφή και άσκηση. Η άσκηση επηρεάζει τον ρυθμό με τον οποίο μεταβολίζεται η ενέργεια στον ανθρώπινο οργανισμό, ενώ η διατροφή καθορίζει το είδος του καυσίμου με το οποίο θα τροφοδοτηθεί. Ο συνδυασμός των δύο παραπάνω και ιδίως της διατροφής, καθώς τα άτομα με σακχαρώδη διαβήτη αδυνατούν να μεταβολίσουν τους υδατάνθρακες οι οποίοι βρίσκονται στο γεύμα τους, προσδιορίζουν τις παραμέτρους από τις οποίες εξαρτάται η ποιότητα ζωής των ατόμων με σακχαρώδη διαβήτη.

Σκοπός της παρούσας διπλωματικής εργασίας είναι να συμβάλει στην ανάπτυξη ενός μοντέλου εκτίμησης της περιεχόμενης ποσότητας υδατανθράκων στα λαμβανόμενα γεύματα από φωτογραφικά στιγμιότυπα. Το μοντέλο αυτό κατά κανόνα αποτελείται από τρία στάδια, κατάτμηση εικόνας, ταξινόμηση και υπολογισμός του όγκου. Δεδομένου, ότι η προαναφερθείσα αλληλουχία διεργασιών είναι ιδιαίτερα απαιτητική, στην συγκεκριμένη διπλωματική εργασία αναπτύσσεται ο αλγόριθμος ταξινόμησης ο οποίος θα χρησιμοποιηθεί σε ένα τέτοιο μοντέλο. Ο αλγόριθμος λαμβάνει ως είσοδο εικόνες φαγητών και τις ταξινομεί σε κάποια από τις υπάρχουσες κλάσεις που έχει προεπιλεγθεί να αναγνωρίζει ο ταξινομητής. Ο αλγόριθμος ταξινόμησης που επιλέγεται στη παρούσα διπλωματική εργασία είναι τα Συνελικτικά Νευρωνικά Δίκτυα.

1.2 Ανασκόπηση

Την παρούσα στιγμή, στην βιβλιογραφία έχουν ήδη αναπτυχθεί συστήματα ταξινόμησης εικόνων φαγητών βασισμένα στα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ). Πολλά από τα οποία εκπαιδεύονται σε Σύνολα Δεδομένων με υπέρογκο πλήθος εικόνων και χρησιμοποιούν Συνελικτικά Νευρωνικά Δίκτυα με πολύ σύνθετες αρχιτεκτονικές δομές. Η μεγάλη χωρητικότητα των καρτών γραφικών (GPU) που χρησιμοποιείται από τα συστήματα εκπαίδευσης επιτρέπουν την εκπαίδευση των παραπάνω ΣΝΔ σε αποδεκτά χρονικά πλαίσια, που κυμαίνονται κοντά στην μια εβδομάδα.

Επίσης έχει δημιουργηθεί ένα σημαντικό πλήθος συνόλων δεδομένων που αποτελείται από φαγητά. Το κάθε ένα με ένα διαφορετικό πλήθος εικόνων, με διαφορετικά φαγητά και διαφορετική ποιότητα φωτογραφιών. Το μόνο σύνολο δεδομένων βασισμένο στις διατροφικές ανάγκες των ατόμων με σακχαρώδη διαβήτη είναι το Diabetes (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014) το οποίο δεν είναι προσβάσιμο.

1.3 Συνεισφορά της Διπλωματικής Εργασίας

Σύμφωνα με όλα τα παραπάνω, η παρούσα διπλωματική εργασία προωθεί την ανάπτυξη συστήματος αυτόματης αναγνώρισης τροφής προς δύο άξονες. Αρχικά εκπαιδεύεται Συνελικτικό Νευρωνικό Δίκτυο χρησιμοποιώντας μια state-of-art αρχιτεκτονική δομή, βασισμένη όμως πλέον στις επιπτώσεις που προκαλεί κάθε φαγητό στο μεταγευματικό σάκχαρο, δηλαδή προσαρμοσμένο στις ανάγκες των ατόμων με σακχαρώδη διαβήτη. Δευτερευόντως, προτείνεται το νέο σύνολο δεδομένων NTUA-Food 2017 το οποίο την παρούσα στιγμή αποτελεί το μόνο διαθέσιμο προσβάσιμο σύνολο δεδομένων οργανωμένο σύμφωνα με τις επιπτώσεις κάθε τρόφιμου στο σάκχαρο. Τα παραπάνω συνδυάζονται μεταξύ τους, καθώς υλοποιείται ταξινόμηση δύο επιπέδων. Η ταξινόμηση του πρώτου επιπέδου γίνεται σύμφωνα με τις ομάδες φαγητών που προκαλούν παρόμοιες μεταγευματικές μεταβολές στο σάκχαρο, ενώ στο δεύτερο επίπεδο αναγνωρίζεται κάθε φαγητό μεμονωμένα.

1.4 Δομή

Η δομή της παρούσας διπλωματικής εργασίας διακρίνεται σε 6 νοηματικές ενότητες. Στο Κεφάλαιο 2 αναπτύσσεται ο Σακχαρώδης Διαβήτης αναλύοντας τους βασικούς τύπους εμφάνισης του, τις συνέπειες του και την διατάραξη των σημαντικότερων ομοιοστατικών μηχανισμών που επιτελεί. Στην συνέχεια ακολουθεί το κεφάλαιο της Βιβλιογραφικής Ανασκόπησης, στο οποίο αναλύονται όλα τα στάδια για τον υπολογισμό της περιεχόμενης ποσότητας υδατανθράκων σε φωτογραφικά στιγμιότυπα γεύματος, δίνοντας ιδιαίτερη έμφαση στο στάδιο της ταξινόμησης. Ακολούθως αναλύεται η μεθοδολογία στην οποία θα στηριχτεί η παρούσα διπλωματική εργασία, ξεκινώντας από τα Νευρωνικά Δίκτυα, προχωρώντας στην ανάλυση των Συνελικτικών Νευρωνικών Δικτύων και κλείνοντας με την ανάλυση της Μηχανής Διανυσμάτων Υποστήριξης. Στο Κεφάλαιο 5 γίνεται ενδελεχής παρουσίαση των σημαντικότερων συνόλων δεδομένων φαγητού, παρουσιάζοντας και αναλύοντας την συλλογιστική δημιουργίας του νέου συνόλου δεδομένων NTUA-Food 2017, το οποίο εισάγεται από την παρούσα διπλωματική εργασία. Επιπλέον στο Κεφάλαιο 6 υπογραμμίζεται η βελτίωση της ακρίβειας που επιτυγχάνεται με την χρήση Συνελικτικών Νευρωνικών Δικτύων για την ανάπτυξη συστήματος αυτόματης αναγνώρισης τροφίμων. Συγκεκριμένα αρχικά παρατίθεται η αύξηση στην ακρίβεια ταξινόμησης που επιτυγχάνεται στο σύνολο δεδομένων

Sarantea-2016 και στην συνέχεια περιγράφεται η ταξινόμηση δύο επιπέδων η οποία εφαρμόζεται στο προτεινόμενο σύνολο δεδομένων NTUA-Food 2017. Σε αμφότερα τα δύο αυτά υποκεφάλαια ο πυρήνας εξαγωγής χαρακτηριστικών και ταξινόμησης είναι τα Συνελικτικά Νευρωνικά Δίκτυα. Τέλος ακολουθεί ένα συγκεντρωτικό κεφάλαιο, στο οποίο αναδεικνύεται η συμβολή της παρούσας διπλωματικής εργασίας και προτείνονται πεδία για μελλοντική έρευνα.

2 Σακχαρώδης Διαβήτης

2.1 Ορισμός

Σύμφωνα με την ΕΛΛΗΝΙΚΗ ΕΝΔΟΚΡΙΝΟΛΟΓΙΚΗ ΕΤΑΙΡΕΙΑ (2017):

Ο Σακχαρώδης Διαβήτης (ΣΔ) είναι μεταβολική ασθένεια η οποία χαρακτηρίζεται από αύξηση της συγκέντρωσης του σακχάρου στο αίμα (υπεργλυκαιμία) και από διαταραχή του μεταβολισμού της γλυκόζης, των λιπιδίων και των πρωτεϊνών, είτε ως αποτέλεσμα ελαττωμένης έκκρισης ινσουλίνης είτε λόγω ελάττωσης της ευαισθησίας των κυττάρων του σώματος στην ινσουλίνη.

2.2 Τύποι Σακχαρώδη Διαβήτη

2.2.1 Σακχαρώδης Διαβήτης Τύπου 1 (ΣΔΤ1)

Το πάγκρεας αποτελεί το κύριο όργανο του ανθρωπίνου σώματος που εμπλέκεται στην παθοφυσιολογία του ΣΔ, εμπεριέχει μεταξύ άλλων τα Α-κύτταρα και τα Β-κύτταρα. Τα πρώτα είναι υπεύθυνα για την παραγωγή γλυκαγόνης ενώ τα δεύτερα για την έκκριση προϊνσουλίνης. Η προϊνσουλίνη αποτελεί την πρόιμη και ανενεργή μορφή της ινσουλίνης, η οποία μετατρέπεται σε ινσουλίνη στην κυκλοφορία του αίματος. Στην περίπτωση του ΣΔΤ1 το ανοσοποιητικό σύστημα του οργανισμού επιτίθεται στα Β-κύτταρα του παγκρέατος και τα καταστρέφει, με αποτέλεσμα να υπάρχει ελάχιστη έως και καθόλου παραγωγή ινσουλίνης. Συνεπώς η γλυκόζη αδυνατεί να δεσμευτεί από τα κύτταρα και να μεταβολιστεί. Η έλλειψη ινσουλίνης οδηγεί στην αδυναμία διατήρησης του σακχάρου στο αίμα σε φυσιολογικά επίπεδα, δημιουργώντας πολλές επιπλοκές στον ανθρώπινο οργανισμό. Ο ασθενής είναι απόλυτα εξαρτημένος από την εξωγενή χορήγηση ινσουλίνης, καθώς ο ίδιος δεν μπορεί να την παράξει σε επαρκείς ποσότητες. Ο ΣΔΤ1 εμφανίζεται κυρίως σε παιδιά και νεότερες ηλικιακές ομάδες.

2.2.2 Σακχαρώδης Διαβήτης τύπου 2 (ΣΔΤ2)

Ο σακχαρώδης διαβήτης τύπου 2 είναι γνωστός ως μη ινσουλινο-εξαρτώμενος σακχαρώδης διαβήτης ή διαβήτης των ενηλίκων. Πρόκειται για πάθηση που προκύπτει όταν το σώμα δε διαχειρίζεται σωστά την παραγόμενη ινσουλίνη ή δεν παράγει αρκετή ινσουλίνη. Στην πρώτη περίπτωση, τα κύτταρα δεν έχουν την δυνατότητα να χρησιμοποιήσουν αποτελεσματικά τα φυσιολογικά επίπεδα ινσουλίνης στο αίμα, φαινόμενο το οποίο αποκαλείται ινσουλινο-αντίσταση. Το φαινόμενο αυτό εμφανίζεται πρωτίστως σε κύτταρα των μυών, του ήπατος και σε λιπώδεις ιστούς. Στην δεύτερη περίπτωση τα β' κύτταρα του

παγκρέατος δεν παράγουν επαρκής ποσότητα ινσουλίνης. Συνήθως στα άτομα με ΣΔΤ2, συνδυάζονται οι δύο παραπάνω παράγοντες, με άλλες φορές το κυρίαρχο αίτιο της περίσσειας γλυκόζης να είναι η ινσουλινο-αντίσταση, ενώ άλλες φορές η αδυναμία των κυττάρων β του παγκρέατος να εκχύσουν την κατάλληλη ποσότητα ινσουλίνης. Αποτελεί τη συχνότερη μορφή ΣΔ και εμφανίζεται κυρίως σε άτομα με παχυσαρκία (YPSOMED, 2017).

2.2.3 Σακχαρώδης Διαβήτης Κύησης

Ο σακχαρώδης διαβήτης κύησης αποτελεί μορφή ΣΔ που αναπτύσσεται κατά τη διάρκεια της εγκυμοσύνης στις γυναίκες. Τα χαρακτηριστικά του σακχαρώδη διαβήτη κύησης προσομοιάζουν τα χαρακτηριστικά του ΣΔΤ2, δηλαδή χαρακτηρίζεται τόσο από ινσουλινοαντίσταση των κυττάρων όσο και μειωμένη έκκριση ινσουλίνης από τα πάγκρεας. Οι γυναίκες που εμφανίζουν ΣΔ κύησης έχουν έως και 60% πιθανότητα να αναπτύξουν ΣΔ τύπου 2 στο μέλλον (Noctor & Dunne, 2015).

2.3 **Διάγνωση και Θεραπεία**

2.3.1 Διάγνωση

Η διάγνωση του ΣΔ στηρίζεται στο ιστορικό του ασθενούς και επαληθεύεται με τον μεταβολικό του έλεγχο. Συμπτώματα όπως η πολουρία, η πολυδιψία και η ανεξήγητη απώλεια σωματικού βάρους είναι ενδεικτικά της ύπαρξης υπεργλυκαιμίας και άρα ΣΔ. Ο μεταβολικός έλεγχος, ο οποίος αποτελεί το βασικό διαγνωστικό στοιχείο στηρίζεται στην εργαστηριακή μέτρηση τόσο της γλυκόζης πλάσματος όσο και της γλυκοζυλιωμένης αιμοσφαιρίνης (HbA1c). Η εξέταση της γλυκόζης του πλάσματος υποδεικνύει δυσλειτουργία στον μηχανισμό ομοιόστασης της, όταν η γλυκόζης νηστείας 8 ωρών είναι μεγαλύτερη ή ίση των 126mg/dl ή σε μια τυχαία στιγμή ξεπερνά τα 200mg/dl. Από την άλλη μεριά, η γλυκοζυλιωμένη αιμοσφαιρίνη δείχνει το ποσοστό της αιμοσφαιρίνης το οποίο προκύπτει από την μη αναστρέψιμη ενζυματική γλυκοζυλίωση της αιμοσφαιρίνης του αίματος. Καθώς ο χρόνος ζωής του ερυθροκυτάρου είναι 128 μέρες, η τιμή αυτή χρησιμοποιείται για να καθορίσει τα επίπεδα της μέσης τιμής της γλυκόζης στο αίμα τους 3 τελευταίους μήνες από την στιγμή εξέτασης. Η διακύμανση των φυσιολογικών ορίων της είναι 4,8-6,0%. Οι παραπάνω ενδείξεις λαμβάνονται με απλή αιμοληψία (Βρότσος & Καρούσης, 2016).

2.3.2 Θεραπεία

Μέχρι στιγμής δεν έχει βρεθεί επίσημα θεραπεία για τον ΣΔ, αλλά χρησιμοποιούνται διάφοροι τρόποι αντιμετώπισης και διαχείρισης της νόσου, τα οποία μεταβάλλουν σημαντικά την κλινική πορεία του ασθενή. Τα θεραπευτικά μέσα που χρησιμοποιούνται εξατομικεύονται ανάλογα με τον τύπο του ΣΔ αλλά

και τον ίδιο τον ασθενή. Στον ΣΔΤ1 (ινσουλινοεξαρτώμενος) , απαιτείται η λήψη ινσουλίνης διαφορετικού είδους και σε διαφορετικά σχήματα, με τους επικρατέστερους τρόπους χορήγησης της ινσουλίνης να είναι οι πένες και οι αντλίες ινσουλίνης .

Η διαχείριση του ΣΔΤ2, στηρίζεται σε λιγότερο επεμβατικά μέσα, ξεκινώντας από τον έλεγχο της διαίτας και της φυσικής άσκησης των ασθενών. Συγκεκριμένα, στην περίπτωση των παχύσαρκων ασθενών η απώλεια βάρους θα οδηγήσει στην αύξηση του αριθμού των υποδοχέων ινσουλίνης και θα αυξηθεί η ευαισθησία των ιστών στην ινσουλίνη και η ανεκτικότητα στην γλυκόζη. Επίσης η φαρμακοβιομηχανία έχει πλέον δημιουργήσει πληθώρα φαρμάκων (Διγουανίδια-Μετφορμίνη, Θιαζολιδινεδιόνες) τα οποία είτε μειώνουν την γλυκόζη του πλάσματος, είτε ευαισθητοποιούν τους περιφερειακούς ιστούς ως προς την δράση της ινσουλίνης είτε διεγείρουν την έκκριση ινσουλίνης από τα β παγκρεατικά κύτταρα. Παρ' όλα αυτά μακροπρόθεσμα οι ασθενείς αυτοί ακολουθούν το σχήμα θεραπείας των ινσουλινοεξαρτώμενων ασθενών το οποίο απαιτεί εξωγενείς ενέσεις ινσουλίνης για την αποτελεσματική ρύθμιση του σακχάρου του αίματος (Devlin, 2007).

2.4 Επιπλοκές του Σακχαρώδη Διαβήτη

Ο ΣΔΤ1 αποτελεί αυτοάνοσο νόσημα, δηλαδή ανήκει στις νόσους που προκαλούνται από υπερβολική και λανθασμένη απόκριση του ανοσοποιητικού συστήματος έναντι στον ίδιο τον οργανισμό (National Library of Medicine, 2011). Αντισώματα λοιπόν του οργανισμού στρέφονται έναντι συστατικών των φυσιολογικών ιστών και όχι έναντι μικροβίων ή κυττάρων που παρουσιάζουν πρόβλημα, όπως θα ήταν το φυσιολογικό. Ο ΣΔ πρόκειται για μια ασθένεια που πλήττει σημαντική μερίδα του παγκόσμιου πληθυσμού και που στην περίπτωση που δεν επιδεικνύεται η κατάλληλη προσοχή από τον νοσούντα, μπορεί να οδηγήσει σε σοβαρές επιπλοκές ακόμη και στο θάνατο.

Οι επιπλοκές που προκαλούνται από τον ΣΔ, χωρίζονται σε δύο βασικές κατηγορίες: στις οξείες και στις χρόνιες. Οι οξείες οφείλονται στις τιμές του σακχάρου του παρόντος και του άμεσου παρελθόντος, ενώ οι χρόνιες εμφανίζονται εξαιτίας των αποκλίσεων από το φυσιολογικό που σημειώνει το σάκχαρο για μεγάλο χρονικό διάστημα.

2.4.1 Οξείες Επιπλοκές

Διαβητική κετοξέωση:

Πρόκειται για επιπλοκή που εμφανίζεται όταν υπάρχει έλλειψη ινσουλίνης και σημαντικά αυξημένη τιμή γλυκόζης στο αίμα. Ο οργανισμός όταν υπάρχει πολύ χαμηλή έκκριση ινσουλίνης δεν μπορεί να μεταφέρει τη γλυκόζη εντός των κυττάρων κι επομένως εμφανίζει ενεργειακό έλλειμμα. Προκειμένου να αντιμετωπιστεί αυτό το έλλειμμα, το ήπαρ μετατρέπει λιπαρά οξέα σε κετόνες, μέσω της διαδικασίας της κέτωσης. Οι κετόνες αποτελούν μια προσωρινή μορφή ενέργειας, αλλά εάν παραμείνουν ως η κύρια πηγή ενέργειας μπορεί να προκαλέσουν σοβαρές επιπλοκές. Η σημαντικότερη από αυτές είναι η διαβητική κετοξέωση και χαρακτηρίζεται από μείωση του pH του αίματος. Συμπτώματα και σημεία της κετοξέωσης αποτελούν αρχικά η αφυδάτωση και τα κοιλιακά άλγη και σε απώτερο στάδιο η απώλεια συνείδησης καθώς και το κόμα.

Υπογλυκαιμία:

Ο όρος υπογλυκαιμία περιγράφει τη χαμηλή τιμή γλυκόζης στο αίμα. Το φαινόμενο αυτό εμφανίζεται σε περιπτώσεις λανθασμένης δοσολογίας ινσουλίνης, έντονη άσκηση ή μειωμένη λήψη υδατανθράκων από την τροφή. Τα πρώτα σημάδια της υπογλυκαιμίας είναι εκνευρισμός, σύγχυση και αδυναμία που στην πορεία μετατρέπονται σε εφίδρωση και πανικό. Εάν δεν καταναλωθεί άμεσα φαγητό από τον διαβητικό ασθενή μπορεί να προκληθεί απώλεια συνείδησης μέχρι και κόμα.

2.4.2 Χρόνιες Επιπλοκές

Η περίσσεια γλυκόζης στο αίμα καταστρέφει τα αιμοφόρα αγγεία και προκαλεί διαβητική αγγειακή νόσο, η οποία διακρίνεται σε μικροαγγειοπάθεια και μακροαγγειοπάθεια. Οι σημαντικότερες επιπτώσεις της μικροαγγειοπάθειας είναι η διαβητική αμφιβληστροειδοπάθεια, η διαβητική νεφροπάθεια καθώς και η διαβητική νευροπάθεια:

Διαβητική Αμφιβληστροειδοπάθεια:

Η διαβητική αμφιβληστροειδοπάθεια χαρακτηρίζεται από βλάβη των αιμοφόρων αγγείων στον αμφιβληστροειδή χιτώνα του οφθαλμού. Ιδιαίτερα από την πάθηση αυτή επηρεάζεται η ωχρή κηλίδα του οφθαλμού, ένας σχηματισμός του οπισθίου τμήματος του οφθαλμού που σχετίζεται με την όραση. Η πάθηση της ωχρής κηλίδας ονομάζεται εκφύλιση της ωχρής κηλίδας. Η διαβητική αμφιβληστροειδοπάθεια οδηγεί σε τελικό στάδιο στην τύφλωση και επηρεάζει μέχρι και το 80% των νοσούντων από ΣΔ για περισσότερο από 20 χρόνια (Kertes & Johnson, 2007).

Διαβητική Νεφροπάθεια:

Πρόκειται για πάθηση που επηρεάζει τη λειτουργία του νεφρού μέχρι τελικώς να οδηγήσει σε νεφρική ανεπάρκεια. Η πάθηση αυτή οφείλεται στις διαρκώς υψηλές τιμές του σακχάρου και στις βλάβες

που δημιουργεί η περίσσεια γλυκόζης στο νεφρικό παρέγχυμα. Συγκεκριμένα καταστρέφονται τα σπειράματα, τα οποία αποτελούν τις βασικές δομικές και λειτουργικές ομάδες του νεφρού. Η διαβητική νεφροπάθεια υπολογίζεται ότι αφορά έναν στους τέσσερις διαβητικούς και αποτελεί την πιο συχνή αιτία τελικού σταδίου χρόνιας νεφρικής ανεπάρκειας (Χριστοδούλου, 2017).

Διαβητική Νευροπάθεια:

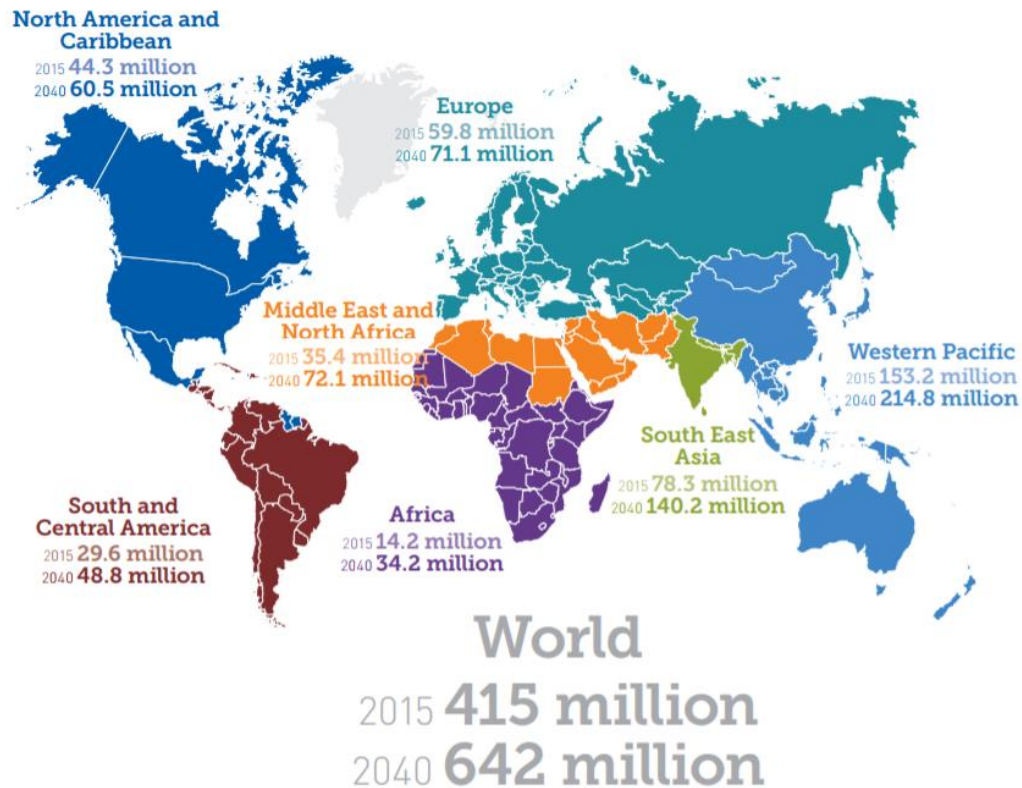
Η διαβητική νευροπάθεια αφορά την καταστροφή των περιφερικών νεύρων και γίνεται αντιληπτή μέσω της απώλειας αισθητικότητας του ασθενή. Η περίσσεια γλυκόζης οδηγεί σε ωσμωτική καταστροφή κυττάρων που είναι υπεύθυνα για την παραγωγή μιας προστατευτικής για τα νεύρα ουσίας, της μυελίνης. Σχετικές ασθένειες που μπορεί να προκληθούν είναι μονονευροπάθεια, πολυνευροπάθεια καθώς και μορφές νευροπάθειας του αυτόνομου νευρικού συστήματος.

Άλλες σημαντικές ασθένειες που συνδέονται με την διαβητική μικροαγγειοπάθεια είναι η περιοδοντίτιδα και η διαβητική καρδιομυοπάθεια.

Τέλος η μακροαγγειοπάθεια του ΣΔ και η επακόλουθη αθηροσκλήρωση που προκαλεί είναι ο λόγος της αυξημένης επίπτωσης καρδιακών εμφραγμάτων, εγκεφαλικών επεισοδίων και γάγγραινας των κάτω άκρων (ΓΝΑ Ελληνικός Ερυθρός Σταυρός, 2005). Από τα παραπάνω ξεχωρίζει το έμφραγμα μυοκαρδίου, το οποίο παρουσιάζεται 3-5 φορές συχνότερα σε διαβητικούς ασθενείς και αποτελεί τη κύρια αιτία θανάτου σε αυτούς (Χριστοδούλου, 2017).

2.5 Στατιστικά Στοιχεία

Για την καλύτερη κατανόηση της εξάπλωσης και της σημασίας της νόσου του ΣΔ παρουσιάζονται μερικά στατιστικά στοιχεία που σχετίζονται μαζί του. Αρχικά παρατίθεται χάρτης με τον υπολογισμό των ενηλίκων άνω των 20 χρονών, που πλήττονται από την νόσο του ΣΔ. Αναλυτικά αναγράφονται τα ποσοστά των διαβητικών, σε 7 γεωγραφικές ζώνες, στην Βόρεια Αμερική και Καραϊβική, στην Νότια και Κεντρική Αμερική, στην Ευρώπη, στην μέση Ανατολή και Βόρεια Αφρική, στην Αφρική, στην Νοτιο-Δυτική Ασία καθώς και στην ζώνη του Δυτικού Ειρηνικού, το έτος 2015. Πέρα από τα στοιχεία αυτά, γίνεται και μια προσπάθεια πρόβλεψης της εξάπλωσης του ΣΔ μέχρι το έτος 2040. Χαρακτηριστικά φαίνεται ότι 415 εκατομμύρια άνθρωποι παγκοσμίως πλήττονται από την νόσο αυτή, ενώ μέχρι το έτος 2040 θεωρείται ότι ο αριθμός αυτός θα φτάσει τα 642 εκατομμύρια. Το παραπάνω αποτελεί αύξηση της τάξεως του 55%, σύμφωνα με τα στοιχεία που δημοσίευσε η International Diabetes Federation (IDF), στην ετήσια έκθεση της (2015).



Εικόνα 2.1. Πλήθος ανθρώπων με διαβήτη στον κόσμο ανά περιοχή το έτος 2015 και το έτος 2040 (International Diabetes Federation, 2015)

Επίσης αξίζει να αναφερθεί ότι παγκοσμίως ο ΣΔ αποτελεί μια από τις σημαντικότερες αιτίες θανάτων. Συγκεκριμένα σύμφωνα με την ίδια έκθεση που εκπονήθηκε από το IDF, βρέθηκε ότι 5 εκατομμύρια θάνατοι ετησίως έχουν ως βασική αιτία τον ΣΔ. Αντίστοιχα άλλες νόσοι που μαστίζουν την ανθρωπότητα όπως ο ιός του HIV/AIDS, η φυματίωση, και η ελονοσία ευθύνονται για 1,5 εκατομμύρια θανάτους κάθε μια από τις δύο πρώτες και για 0,6 εκατομμύρια η τελευταία, σύμφωνα με το World Health Organization (2013).

Υπολογίζεται ότι το έτος 2015, ένας στους 11 ανθρώπους μαστιζόταν από την νόσο ΣΔ, ενώ μέχρι το 2040 το ποσοστό θα αυξηθεί σε έναν στους 10 ανθρώπους. Από το συνολικό πλήθος των διαβητικών φαίνεται ότι 52% είναι άντρες ενώ το υπόλοιπο 48% γυναίκες. Αξιοσημείωτο και ίσως το πιο ανησυχητικό στατιστικό, αποτελεί το γεγονός ότι περίπου ένας στους δύο ανθρώπους που νοσεί με ΣΔ, δεν έχει διαγνωστεί ακόμα. Ο ΣΔ εξαπλώνεται και στις παιδικές ηλικίες, κυρίως μέσω του παιδικού σακχαρώδη διαβήτη (ΣΔΤ1), όπου το 2015 ήταν η πρώτη χρονιά όπου το συνολικό πλήθος παιδιών με ΣΔ ξεπέρασε το μισό εκατομμύριο. Ο παιδικός σακχαρώδης διαβήτης παρουσιάζει αύξηση της τάξεως του 3% ετησίως (International Diabetes Federation, 2015).

Όλα τα παρακάτω τονίζουν τις διαστάσεις που έχει λάβει η νόσος του ΣΔ και επιδεικνύουν την επιτακτική ανάγκη της πρόληψης και της καλύτερης αντιμετώπισης του. Αν και σύμφωνα με τα στατιστικά του 2015 φαίνεται ότι το 12% των παγκοσμίων εξόδων στον τομέα της υγείας επενδύονται στην αντιμετώπιση του ΣΔ, φαίνεται η ανάγκη για περισσότερη αφοσίωση στο ζήτημα αυτό (International Diabetes Federation, 2015).

2.6 Γλυκόζη

Η βασική πηγή γλυκόζης στον οργανισμό προέρχεται από τα μακρομόρια των υδατανθράκων, καθώς είναι ταχεία η διάσπαση τους σε μονοσακχαρίτες, για την άμεση κάλυψη των ενεργειακών αναγκών των κυττάρων. Οι υδατάνθρακες αποτελούνται από αλυσίδες σακχάρων που μπορεί να έχουν πολύ απλή μορφή ξεκινώντας από μονοσακχαρίτες και δισακχαρίτες, μέχρι σύνθετες οργανικές ουσίες αποτελούμενες από πολλά σάκχαρα όπως οι πολυσακχαρίτες. Οι πολυσακχαρίτες δεν είναι άμεσα επεξεργάσιμοι από τον ανθρώπινο οργανισμό για αυτό μέσω της διαδικασίας της πέψης τα σύνθετα αυτά μόρια μετατρέπονται σε απλές δομές σακχάρου. Συγκεκριμένα η διάσπαση των υδατανθράκων ξεκινάει μέσω των ενζύμων που βρίσκονται στο σίελο. Στην συνέχεια μέσω της διαδικασίας της πέψης προχωρούν στον οισοφάγο, στο στομάχι και καταλήγουν στο λεπτό έντερο. Στο σημείο αυτό, μέσω ενζύμων διασπώνται στις απλούστερες μορφές τους, τους μονοσακχαρίτες, εκ των οποίων οι σημαντικότεροι είναι η γλυκόζη, η φρουκτόζη και η γαλακτόζη. Η γλυκόζη απορροφάται από τα τοιχώματα του λεπτού εντέρου και με αυτό τον τρόπο εισέρχεται στην κυκλοφορία του αίματος. Από εκεί και πέρα ξεκινάει η διαδικασία του μεταβολισμού της γλυκόζης. Ο μεταβολισμός της φρουκτόζης και της γαλακτόζης γίνονται αντίστοιχα μέσω των διαδικασιών της φρουκτόλυσης και του μεταβολισμού της γαλακτόζης που έχουν ως παράγωγα και αυτά γλυκόζη.

2.6.1 Ομοιόσταση Γλυκόζης

Η γλυκόζη αποτελεί την κυριότερη πηγή ενέργειας του οργανισμού και είναι ουσία απαραίτητη για τη λειτουργία του εγκεφάλου. Μάλιστα, πρόκειται για το υποχρεωτικό «καύσιμο» του εγκεφάλου σε φυσιολογικές συνθήκες. Συνυπολογίζοντας και το γεγονός ότι ο εγκέφαλος δεν μπορεί να την παράξει ή να την αποθηκεύσει ο ίδιος, συνάγεται εύκολα το συμπέρασμα ότι η γλυκόζη θα πρέπει να ρυθμίζεται προσεκτικά και με ακρίβεια ώστε να παρέχεται στα κύτταρα του εγκεφάλου σε επαρκείς ποσότητες.

Η γλυκόζη εντός των κυττάρων χρησιμοποιείται για την παραγωγή ενέργειας, μέσω της διαδικασίας της γλυκόλυσης. Τα τελικά μόρια που παράγονται από αυτή τη διαδικασία θα εισέλθουν με τη σειρά τους

σε έναν έτερο καταρράκτη χημικών αντιδράσεων που θα έχει ως τελικό αποτέλεσμα την παραγωγή μορίων ATP. Αυτά είναι μόρια υψηλής ενέργειας που συμμετέχουν σε πληθώρα περαιτέρω διεργασιών.

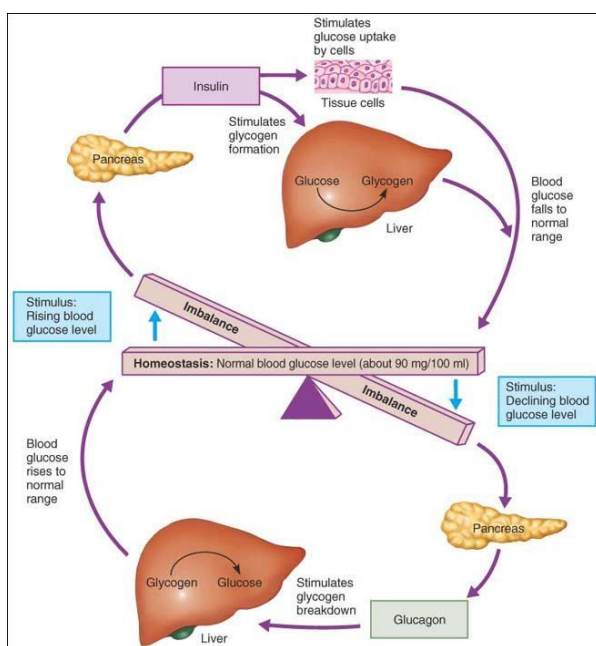
Ο ανθρώπινος οργανισμός διαθέτει ομοιοστατικούς μηχανισμούς, οι οποίοι του επιτρέπουν να διατηρεί την τιμή της γλυκόζης εντός φυσιολογικών επιπέδων (70-110mg/dl σε νηστεία, με παροδικές αυξήσεις έπειτα από γεύματα), προκειμένου να αποφεύγονται επικίνδυνες ακραίες αυξομειώσεις (υπογλυκαιμία – υπεργλυκαιμία).

Η ομοίωση της γλυκόζης επηρεάζεται από ποικίλους παράγοντες. Εστιάζοντας όμως στους βασικότερους, δύο είναι τα κύρια όργανα του ανθρώπινου σώματος που την επηρεάζουν (το πάγκρεας και το ήπαρ) και ισάριθμες είναι και οι κύριες ρυθμιστικές ορμόνες (η ινσουλίνη και η γλυκαγόνη).

Ας ξεκινήσουμε από την περίπτωση της αντιρρόπησης της αύξησης της γλυκόζης στο αίμα. Μετά από τα γεύματα, η ποσότητα της γλυκόζης στην αιματική κυκλοφορία αυξάνεται. Το γεγονός αυτό ανιχνεύεται από τα β-κύτταρα του παγκρέατος. Συγκεκριμένα, η γλυκόζη προσδέεται σε έναν υποδοχέα γλυκόζης που εδράζεται στην επιφάνεια των β-κυττάρων και έτσι σηματοδοτείται η παραγωγή ινσουλίνης από τα κύτταρα αυτά. Η ινσουλίνη είναι μία ορμόνη που προωθεί την είσοδο της γλυκόζης στα κύτταρα των περιφερικών ιστών (πχ μύες) και παράλληλα αναχαιτίζει τις διαδικασίες της γλυκονεογένεσης και της γλυκογονόλυσης. Το συνολικό αποτέλεσμα είναι η μείωση των επιπέδων γλυκόζης στο αίμα και η αύξηση των επιπέδων αυτής ενδοκυττάρια, όπου είτε χρησιμοποιείται είτε αποθηκεύεται.

Εξίσου άμεση κι αποτελεσματική είναι η αντίδραση του οργανισμού στην περίπτωση της μείωσης της γλυκόζης στο αίμα, όπως σε περιπτώσεις νηστείας ή αρκετές ώρες μετά από κάποιο γεύμα. Στις καταστάσεις αυτές, η πρώτη κίνηση του οργανισμού είναι να μειώσει την παραγωγή ινσουλίνης. Ακολουθώς, η ελάττωση της γλυκόζης ανιχνεύεται από τα α-κύτταρα του παγκρέατος (γειτονικά ως προς τα β-κύτταρα), τα οποία αντιδρούν παράγοντας μία άλλη ορμόνη, τη γλυκαγόνη. Η γλυκαγόνη ουσιαστικά επιτελεί δράσεις αντίθετες της ινσουλίνης: συμβάλλει στη γλυκογονόλυση, τη γλυκονεογένεση και τη μείωση της χρήσης της γλυκόζης από τους περιφερικούς ιστούς. Πιο συγκεκριμένα, ενισχύει την αποδόμηση του αποθηκευμένου γλυκογόνου προς μόρια γλυκόζης (γλυκογονόλυση), μια διαδικασία που το μεγαλύτερο μέρος της διενεργείται στο ήπαρ. Επίσης, προωθεί τη δημιουργία γλυκόζης από άλλες πηγές, όπως τα αμινοξέα των πρωτεϊνών, η γλυκερόλη των λιπιδίων, καθώς και άλλες πηγές όπως το γαλακτικό οξύ (νεογλυκογένεση). Η νεογλυκογένεση λαμβάνει χώρα λόγω στο ήπαρ και στους νεφρούς. Χρονικά, στα αρχικά στάδια νηστείας η γλυκογονόλυση επικρατεί, ενώ όσο η νηστεία παρατείνεται τόσο αυξάνεται ο βαθμός συμμετοχής της γλυκονεογένεσης στην παραγωγή γλυκόζης. Η κοινή συνισταμένη αυτών των διαδικασιών είναι η αύξηση της γλυκόζης στο αίμα και συνεπώς η επαναφορά της στα επιθυμητά επίπεδα.

Από τα παραπάνω συνάγεται ότι η ινσουλίνη και η γλυκαγόνη αποτελούν αντίστοιχα μία αναβολική και μία καταβολική ορμόνη, οι οποίες επιτελούν αντιδιαμετρικά αντίθετες λειτουργίες. Η ισορροπία μεταξύ αυτών των δύο εξασφαλίζει τη σωστή ρύθμιση της γλυκόζης στο αίμα. Πέρα από τις δύο αυτές ορμόνες, στην ομοίωση της γλυκόζης συμμετέχουν - σε μικρότερο όμως βαθμό και με λιγότερο καθοριστικό ρόλο - και άλλες ουσίες, όπως οι κατεχολαμίνες, τα γλυκορτικοειδή, η αυξητική ορμόνη και οι θυρεοειδικές ορμόνες. Η σημασία αυτών των ουσιών γίνεται περισσότερο εμφανής σε διάφορες παθήσεις που τις αφορούν, σε περιπτώσεις σχετικής φαρμακευτικής αγωγής, αλλά και σε ακραίες καταστάσεις υπέρ- ή υπογλυκαιμίας.



Εικόνα 2.2. Αναπαράσταση Ομοιοστατικού Μηχανισμού Γλυκόζης (2014)

Βασικές λειτουργίες της ομοίωσης της γλυκόζης αποτελούν η γλυκογένεση, η γλυκογονόλυση, η γλυκονογένεση και η γλυκόλυση. Αν και έχουν αναφερθεί παραπάνω οι όροι αυτοί, θα αναλυθεί στην συνέχεια κάθε ένας από αυτούς πιο προσεχτικά.

2.6.1.1 Γλυκογένεση

Η γλυκογένεση αποτελεί τον σχηματισμό γλυκογόνου από την γλυκόζη. Το γλυκογόνο σχηματίζεται ανάλογα με τη ζήτηση γλυκόζης και ATP από τα κύτταρα. Αν υπάρχει περίσσεια γλυκόζης μετά την κάλυψη των ενεργειακών αναγκών, τότε αποθηκεύεται σε μορφή γλυκογόνου στο ήπαρ και στα κύτταρα των μυών

2.6.1.2 Γλυκογονόλυση

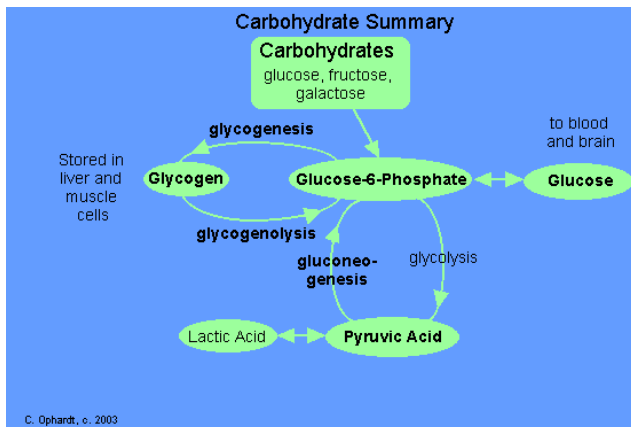
Η γλυκογονόλυση, αποτελεί την αντίστροφη διαδικασία της γλυκογένεσης, προκειμένου να παραχθεί γλυκόζη από το γλυκογόνο που είναι αποθηκευμένο στο στους μύες και στο συκώτι. Οι δύο ορμόνες που ρυθμίζουν αυτή την διαδικασία είναι η γλυκαγόνη από το πάγκρεας και η επινεφρίνη από τα επινεφρίδια. Η γλυκαγόνη εκκρίνεται όταν παρατηρούνται χαμηλά επίπεδα γλυκόζης στο αίμα ενώ η επινεφρίνη ως απόκριση στο άγχος. Αμφότερες οι δύο ορμόνες ενεργοποιούν ένζυμα προκειμένου να εκκινήσει η διαδικασία της γλυκογονόλυσης και παράλληλα να ανασταλεί η διαδικασία της γλυκογένεσης.

2.6.1.3 Γλυκονεογένεση

Με τον όρο γλυκονεογένεση, (glyconeogenesis), χαρακτηρίζεται η βιοχημική διεργασία παραγωγής γλυκόζης, από πηγές που δεν προέρχονται από υδατάνθρακες. Το βασικό σημείο εκκίνησης της διεργασίας αυτής είναι το πυροσταφυλικό οξύ. Το γαλακτικό οξύ, κάποια αμινοξέα από πρωτεΐνες και η γλυκερόλη από το λίπος μπορούν να μετατραπούν σε γλυκόζη. Η γλυκονεογένεση αποτελεί διαδικασία με αντίστροφη ροή από την Γλυκόλυση, παρ' όλα αυτά οι δύο διαδικασίες δεν αποτελούνται από τα ίδια βήματα.

2.6.1.4 Γλυκόλυση

Η γλυκόλυση είναι το μεταβολικό μονοπάτι μετατροπής της γλυκόζης σε πυροσταφυλικό οξύ (pyruvate), που αποτελεί σημαντικό μόριο της κυτταρικής αναπνοής. Κάθε μόριο γλυκόζης αποδίδει δύο μόρια πυροσταφυλικού οξέος. Η γλυκόλυση λαμβάνει χώρα στο κυτταρόπλασμα των κυττάρων κι αποτελείται από μια αλληλουχία 10 αντιδράσεων ενζυματικής κατάλυσης. Τα ενδιάμεσα παράγωγα της γλυκόλυσης είναι ιδιαίτερα χρήσιμα καθώς μπορούν να παραχθούν και από άλλα μεταβολικά μονοπάτια όπως από αυτά της φρουκτόζης και της γαλακτόζης.



Εικόνα 2.3. Αναπαράσταση των ομοιοστατικών μηχανισμών (Ophardt, 2003)

3 Βιβλιογραφική Ανασκόπηση

3.1 Καταμέτρηση μακροθρεπτικών συστατικών

Η διατροφή αποτελεί έναν από τους σημαντικότερους παράγοντες που επηρεάζει την υγεία των ανθρώπων. Τους επηρεάζει είτε άμεσα, καθορίζοντας για παράδειγμα την αντίστοιχη δοσολογία ινσουλίνης σε περίπτωση ατόμων με διαβήτη, είτε εμμέσως έχοντας επιπτώσεις στο βάρος τους ή προκαλώντας σημαντικά προβλήματα υγείας. Σύμφωνα με όλα τα παραπάνω, αναδύεται η ανάγκη καταγραφής και καταμέτρησης των θερμίδων και μακροθρεπτικών συστατικών που καταναλώνονται καθημερινά από τον άνθρωπο.

3.1.1 Συμβατικός Τρόπος Καταγραφής Διατροφικού Περιεχομένου

Ο συμβατικός τρόπος καταγραφής του διατροφικού περιεχομένου των γευμάτων αποτελεί μια σύνθετη διαδικασία. Αρχικά απαιτείται ζύγιση κάθε διαφορετικού συστατικού που εμπεριέχεται στο γεύμα, ενώ στη συνέχεια ενδείκνυται η γνώση της σύστασης της εκάστοτε τροφής σε θρεπτικά συστατικά. Αντίστοιχα σε περίπτωση συσκευασμένων φαγητών, τα μακροθρεπτικά συστατικά αναγράφονται στην συσκευασία και η ζύγιση της τροφής που καταναλώνεται, είναι το στάδιο που ακολουθεί. Σύμφωνα με όλα τα παραπάνω γίνεται εμφανές ότι η προαναφερθείσα διαδικασία είναι σχετικά σύνθετη και χρονοβόρα. Αυτό συμβαίνει καθώς η πρόσβαση σε ηλεκτρονική ζυγαριά φαγητών δεν είναι πάντα διαθέσιμη και οι παραπάνω υπολογισμοί σε σύνθετα φαγητά αποκλίνουν σημαντικά. Επίσης σε κάθε περίπτωση που δεν είναι γνωστός ο τρόπος παρασκευής του γεύματος, όπως η κατανάλωση τροφής σε εστιατόρια και fast food, καθιστά την συγκεκριμένη διαδικασία έως και απαγορευτική. Συνοψίζοντας, η σημαντικότερη δυσκολία της παραπάνω μεθόδου έγκειται στο χρόνο που απαιτείται, στα διαθέσιμα μέσα που χρειάζονται και στο γνωστικό υπόβαθρο, το οποίο είναι απαραίτητο.

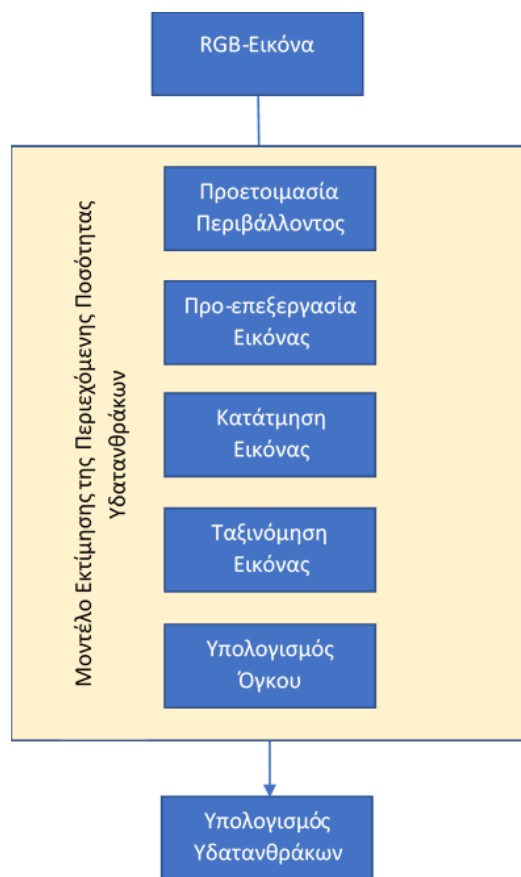
Συγκεκριμένα στην περίπτωση του ΣΔ, ιδίως στη μορφή του ινσουλινοεξαρτώμενου, κάθε γεύμα πρέπει να ακολουθείται από μια δοσολογία ινσουλίνης. Ο υπολογισμός των υδατανθράκων που εμπεριέχονται στο γεύμα και τελικώς θα καθορίσουν την συνολική ποσότητα ινσουλίνης που θα χρησιμοποιήσει το άτομο με ΣΔ, ιδανικά θα μπορούσε να επιτευχθεί με την μεθοδολογία που περιγράφηκε παραπάνω. Καθώς όμως η διαδικασία επιτελείται καθημερινώς από τους νοσούντες με ΣΔ, χρησιμοποιούνται συνήθως διαφορετικοί πιο εμπειρικοί τρόποι, για την καταμέτρηση των υδατανθράκων. Σύμφωνα με έρευνα που έγινε στο πανεπιστημιακό νοσοκομείο της Βέρνης σε 19 εθελοντές με ΣΔΤ1, παρατηρήθηκε ότι το μέσο απόλυτο σφάλμα στην πρόβλεψη των υδατανθράκων είναι 27,89 γραμμάρια (Rhyner, et al., 2016, p. 101). Σφάλμα ιδιαίτερα σοβαρό, δεδομένου ότι σφάλμα πρόβλεψης του σακχάρου

με απόκλιση μεγαλύτερη των ± 20 γραμμαρίων, επιφέρει σοβαρές επιπτώσεις στο μεταγευματικό σάκχαρο (Smart, King, McElduff, & Collins, 2012) . Αντιθέτως αποκλίσεις στον υπολογισμό των υδατανθράκων του γεύματος μικρότερες των ± 10 γραμμαρίων, δεν προκαλούν αποσυντονισμό του μεταγευματικού σακχάρου (Smart, et al., 2009).

3.1.2 Ηλεκτρονική Καταγραφή Διατροφικού Περιεχομένου

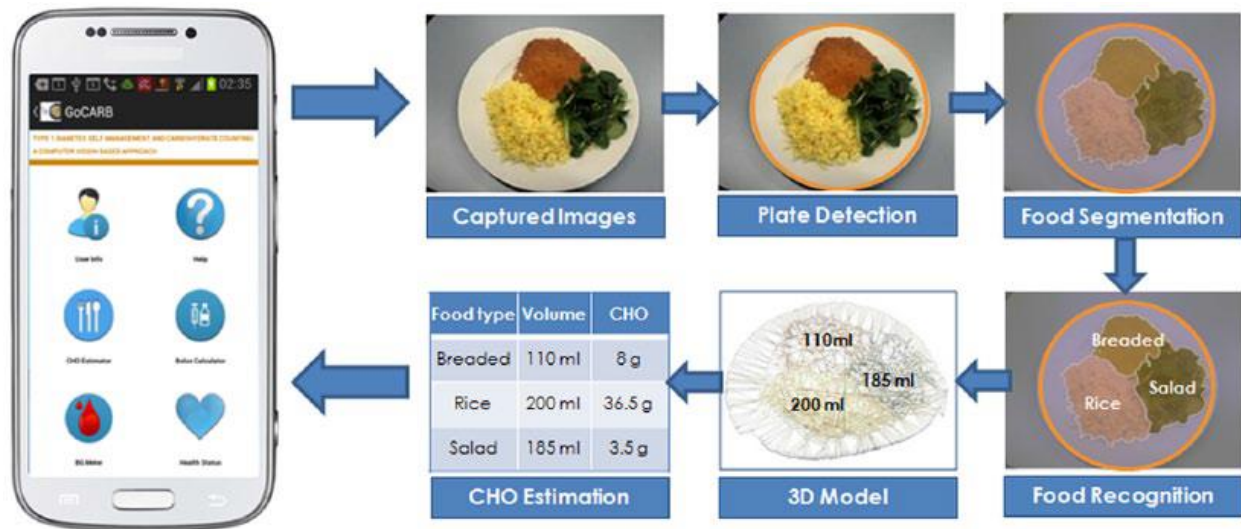
Για την επίλυση πολλών εκ των παραπάνω δυσκολιών και δεδομένου της τεχνολογικής προόδου που έχει επιτευχθεί τα τελευταία χρόνια, γίνεται σταδιακά η μετάβαση στη νέα γενιά διατροφικής καταγραφής και παρακολούθησης (Fioravanti, et al., 2014). Η διαδικασία αυτή πλέον υλοποιείται ηλεκτρονικά μέσω τριών βασικών βημάτων. Αρχικά ως είσοδος, δίνεται ένα φωτογραφικό στιγμιότυπο του γεύματος που πρόκειται να καταναλώσει ο χρήστης. Στην συνέχεια, μέσω της μεθόδου κατάτμησης (segmentation method), διακρίνονται τα επιμέρους τμήματα της φωτογραφίας και σκιαγραφείται το περίγραμμά τους. Ως δεύτερο βήμα της συγκεκριμένης διαδικασίας, τα σκιαγραφημένα τμήματα που αναγνωρίστηκαν προηγουμένως δίνονται ως είσοδο σε έναν ταξινομητή, προκειμένου να ταυτιστούν με κάποια κατηγορία φαγητού. Από την στιγμή που κάθε τροφή της φωτογραφίας διακριθεί από τις υπόλοιπες της εικόνας και ταυτιστεί με κάποιο γνωστό φαγητό, μεταβαίνουμε στο στάδιο υπολογισμού του όγκου. Για κάθε καταχωρημένο φαγητό της εφαρμογής, καταγράφονται τα μακροθρεπτικά συστατικά που περιέχει ανά μονάδα όγκου. Έτσι προκειμένου να υπολογισθούν τα μακροθρεπτικά συστατικά κάθε τμήματος της εικόνας και εν τέλει τα μακροθρεπτικά συστατικά ολόκληρης της εικόνας, απαιτείται ο υπολογισμός του όγκου του εκάστοτε τμήματος.

Η παραπάνω βασική δομή που χρησιμοποιείται για την ηλεκτρονική καταγραφή του διατροφικού περιεχομένου, πλαισιώνεται από δύο προκαταρκτικά βήματα, την προετοιμασία του περιβάλλοντος και την προεπεξεργασία της εικόνας. Στο πρώτο λαμβάνονται κάποια βοηθητικά μέτρα κατά την λήψη της εικόνας, για την αξιοποίηση τους από τα μελλοντικά βήματα. Το συνηθέστερο είναι ο εμπλουτισμός της εικόνας με κάποιο αντικείμενο αναφοράς, έτσι ώστε να πραγματοποιηθεί με μεγαλύτερη ακρίβεια ο υπολογισμός του όγκου. Όσον αφορά το στάδιο της προεπεξεργασίας της εικόνας, είναι επιθυμητή η μείωση του θορύβου που εμπεριέχεται σε αυτή ή η εφαρμογή κάποιου βοηθητικού φίλτρου για την επίτευξη καλύτερων αποτελεσμάτων ταξινόμησης. Σχηματικά τα βήματα που ακολουθούνται σε ένα μοντέλο εκτίμησης της περιεχόμενης ποσότητας υδατανθράκων από φωτογραφίες αναπαρίσταται στην Εικόνα 3.1



Εικόνα 3.1. . Τα βασικά βήματα επεξεργασίας εικόνας φαγητού για τον υπολογισμό της περιεκτικότητας του σε υδατάνθρακες.

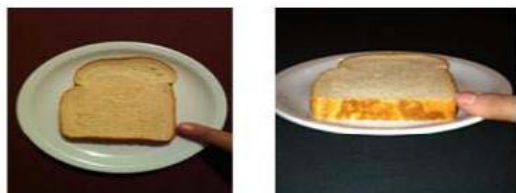
Συνοψίζοντας, τα πέντε στάδια που απαιτούνται για την καταγραφή των μακροθρεπτικών συστατικών που εμπεριέχονται σε ένα φωτογραφικό στιγμιότυπο ενός γεύματος είναι η προετοιμασία περιβάλλοντος, η προεπεξεργασία εικόνας, η κατάτμηση, η ταξινόμηση και ο υπολογισμός του όγκου. Με τον τρόπο αυτό δίνοντας ως είσοδο μια εικόνα στο σύστημα, γίνεται αυτοματοποιημένα ο υπολογισμός των μακροθρεπτικών συστατικών που εμπεριέχονται σε αυτή. Όπως φαίνεται, η παραπάνω διαδικασία σημειώνει σημαντική πρόοδο ως προς την χρονική πολυπλοκότητα και το πλήθος των μέσων που απαιτούνται. Από εκεί και πέρα, η ακρίβεια που επιτυγχάνεται μέσω αυτής της μεθόδου εξαρτάται από πολλές παραμέτρους και αποτελεί το κομμάτι που χρήζει βελτιστοποίησης και μελέτης. Ενδεικτικά στην Εικόνα 3.2, φαίνονται τα βήματα που ακολουθούνται από την εφαρμογή GoCarb, η οποία υπολογίζει την περιεκτικότητα του γεύματος σε υδατάνθρακες (Anthimopoulos M. , et al., 2015). Στο παράδειγμα αυτό, ως μέτρο προεπεξεργασίας της εικόνας θεωρείται η ανάπτυξη αλγορίθμου εντοπισμού των πιάτων (plate detection) .



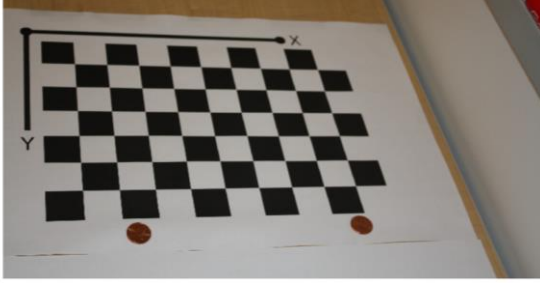
Εικόνα 3.2 Αναπαράσταση των βημάτων εξαγωγής χαρακτηριστικών από την εφαρμογή GoCarb (Anthimopoulos M. , et al., 2015)

3.2 Προετοιμασία Περιβάλλοντος

Σε πολλές εφαρμογές είναι απαραίτητη η προετοιμασία του περιβάλλοντος πριν την λήψη της φωτογραφίας. Οι Rouladzadeh, Shirmohammadi και Al-Maghrabi απαιτούσαν τον αντίχειρα του χρήστη να συμπεριλαμβάνεται στην φωτογραφία ως μέγεθος αναφοράς (2014). Σε άλλες περιπτώσεις προτάθηκε η χρήση άλλων βασικών αντικειμένων στην θέση του αντίχειρα (Zhu, et al., 2008). Επιπλέον άλλες προσεγγίσεις αναπτύχθηκαν, που βασίζονται σε κάρτες αναφοράς προκειμένου να γίνει η βαθμονόμηση της κάμερας (Martin, Kaya, & Gunturk, 2009). Παρόμοια βαθμονόμηση μπορεί να επιτευχθεί και με την χρήση ενός υφάσματος σκακιέρας, με γνωστές διαστάσεις, στο πίσω μέρος της φωτογραφίας. Τέλος έχουν χρησιμοποιηθεί πιάτα και σκεύη με γνωστές διαστάσεις προς αυτή την κατεύθυνση. Ο σκοπός της βαθμονόμησης της κάμερας μέσω αντικειμένων γνωστών διαστάσεων, αποσκοπεί στην υποβοήθηση του υπολογισμού του όγκου των τροφών που εμπεριέχονται στο φωτογραφικό στιγμιότυπο.



Εικόνα 3.3. Χρήση του αντίχειρα ως μέτρο αναφοράς για τον υπολογισμό των κανονικών διαστάσεων των τροφών (Rouladzadeh, Yassine, & Shirmohammadi, 2015)

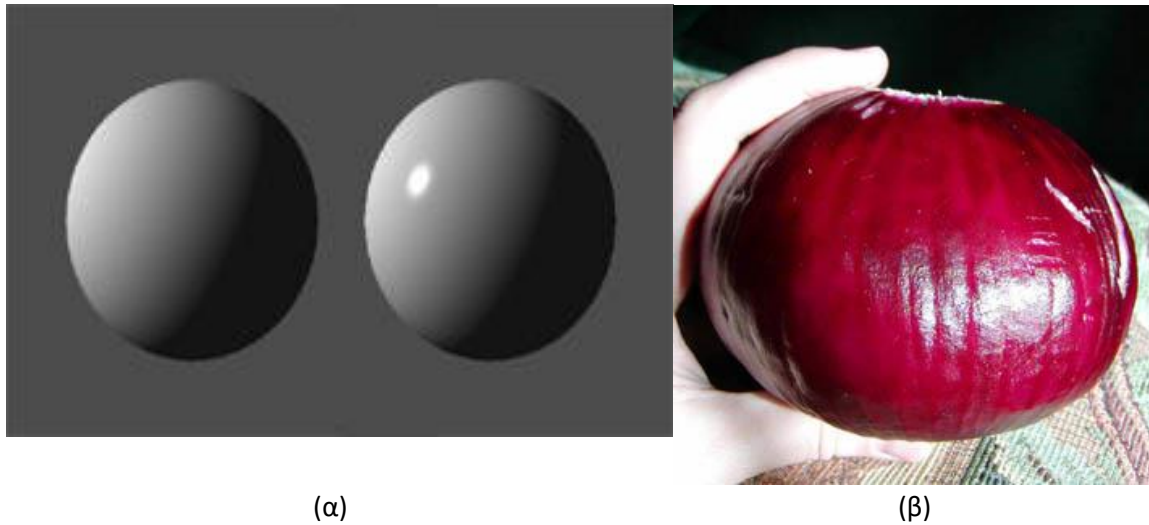


Εικόνα 3.4. Βαθμονόμηση κάμερας με σκακιέρα, με σκοπό τον υπολογισμό των διαστάσεων των νομισμάτων της εικόνας (MathWorks, n.d.)

3.3 Προ-επεξεργασία Εικόνας

Η εικόνα αποτελεί την είσοδο του συστήματος, οπότε η ποιότητα της, ο θόρυβος που εμπεριέχεται σε αυτήν, καθώς και η διαστρέβλωση σημαντικών της στοιχείων είναι παράμετροι οι οποίες θα πρέπει να λαμβάνονται υπόψιν. Για το λόγο αυτό σε πολλές εφαρμογές, επιτελείται αρχικά ένα βήμα προ-επεξεργασίας της εικόνας. Οι Patel, Jain και Joshi βασίστηκαν στην μορφολογική επεξεργασία της εικόνας για την μείωση του θορύβου (2012). Άλλοι ερευνητές χρησιμοποίησαν τεχνικές μείωσης της θόλωσης της εικόνας ως μέσο προ-επεξεργασίας (Neelamegam, Abirami, Vishnu Priya, & Rubalya Valentina, 2013; Xu, He, Khanna, Boushey, & Delp, 2013). Ενώ σε άλλες περιπτώσεις, ερευνητές δημιούργησαν ένα σύστημα αλληλεπίδρασης μεταξύ της εφαρμογής και του χρήστη, ζητώντας από τον χρήστη να επαναλάβει την διαδικασία λήψης της φωτογραφίας, σε περίπτωση που η φωτογραφία δεν πληροί κάποια ποιοτικά χαρακτηριστικά (Ahmad, Khanna, Kerr, Boushey, & Delp, 2014).

Άλλες προσεγγίσεις που έχουν αναπτυχθεί σχετίζονται με την εξάλειψη των σκιών και της ανάκλασης του φωτός (Sun & Du, 2004). Συγκεκριμένα ένα από τα σημαντικότερα προβλήματα που δημιουργούνται στις φωτογραφίες φαγητών είναι το κατοπτρικό αποκορύφωμα (specular highlight), που πρόκειται για την δημιουργία ενός φωτεινού σημείου πάνω σε γυαλιστερά αντικείμενα όταν προσκρούουν ακτίνες φωτός σε αυτά. Δεδομένης της ανακλαστικής φύσης των πιάτων, σε συνδυασμό με την υφή κάποιων φαγητών και το φωτισμό του περιβάλλοντος, το φαινόμενο του κατοπτρικού αποκορυφώματος είναι αρκετά συχνό. Προς την κατεύθυνση αυτή, για την αντιμετώπιση αυτής της τεχνικής δυσκολίας, έχουν αναπτυχθεί φίλτρα σύμφωνα με τους Khanna, Boushey και Delp (2012).



Εικόνα 3.5 Το φαινόμενο του specular highlight

3.4 Κατάτμηση της Εικόνας

Η κατάτμηση της εικόνας αποτελεί το πρώτο ουσιαστικό βήμα επεξεργασίας της εικόνας προκειμένου να εξαχθούν οι υδατάνθρακες που εμπεριέχονται σε αυτή. Το 2012 οι Matsuda, Hoashi και Yanai επικεντρώνονται στην συσχέτιση του εντοπισμού των υποψήφιων περιοχών (candidate regions) της εικόνας με την ακρίβεια ταξινόμησης (Matsuda, Hoashi, & Yanai, 2012). Δηλαδή δίνεται έμφαση στην επίδραση που έχει η μέθοδος εντοπισμού των διαφόρων φαγητών στην εικόνα, στο τελικό αποτέλεσμα του ταξινομητή. Για τον εντοπισμό των διάφορων φαγητών που υπάρχουν σε μια εικόνα, καθώς και ο διαχωρισμός τους από το υπόβαθρο (background) χρησιμοποιήθηκαν οι τεχνικές: Whole Image (Whole), Deformable Part Model(DPM), Circle Detector (Circle), Region Segmentation(JSEG) και Ground Truth Bonding Boxes of the test images (GTBB). Στην περίπτωση της Whole Image, ολόκληρη η εικόνα δίνεται ως είσοδος στον ταξινομητή, ενώ σε στην περίπτωση του GTBB, θεωρείται η ιδανική περίπτωση που το κάθε φαγητό είναι τέλεια διακεκριμένο από το περιβάλλον του. Οι υπόλοιπες τεχνικές σημειώνουν ακρίβεια στο ενδιάμεσο των δύο προηγούμενων.

Έχουν προταθεί πολλοί άλλοι διαφορετικοί τρόποι για την επίτευξη αυτής της σύνθετης διεργασίας. Το 2013 οι He, Khana, Boushey και Delp επιχείρησαν να διεξάγουν μια έρευνα σύγκρισης των επικρατέστερων μεθόδων κατάτμησης εικόνων με φαγητά (He, Khanna, Boushey, & Delp, 2013). Εκείνη την χρονική στιγμή οι τρεις βασικές μέθοδοι σε ελληνική μετάφραση είναι το μοντέλο ενεργού περιγράμματος (active contour), οι κανονικοποιημένες περικοπές (normalized cuts) και το μοντέλο τοπικής διακύμανση (local variation). Η μέθοδος με τα καλύτερα αποτελέσματα ήταν το μοντέλο τοπικής

διακύμανσης, όμως σε πολλές περιπτώσεις ανάλογα με τις αρχικές συνθήκες, το μοντέλο ενεργού περιγράμματος έφθανε σε παρόμοια επίπεδα ακρίβειας. Το πρόβλημα κατάτμησης της εικόνας αποτελεί ένα πολυπαραμετρικό ζήτημα, συνεπώς το αποτέλεσμα αυτό δεν μπορεί να γενικευθεί. Αξίζει να αναφερθεί ότι στην έρευνα αυτή δεν λήφθηκε υπόψιν η χρονική πολυπλοκότητα των παραπάνω τεχνικών.

Ένα ακόμα ζήτημα που προκύπτει στο θέμα κατάτμησης εικόνων φαγητού, είναι αυτό της επικάλυψης, που δεν επιτρέπει την ακριβή οριοθέτηση των φαγητών στο πιάτο. Το συγκεκριμένο θέμα το διερεύνησαν οι He, Xu, Khanna, Boushey, and Delp (2013).

Το πρόβλημα της κατάτμησης επιχειρήθηκε να επιλυθεί με πολλές ακόμα μεθόδους. Το 2014 οι Pouladzadeh, Shirmohammadi, and Yassine χρησιμοποίησαν την μέθοδο αποκοπής γραφημάτων (graph cut) (2014), ενώ το 2012 αντιμετωπίστηκε το ζήτημα της κατάτμησης ως πρόβλημα βελτιστοποίησης, έχοντας ως βάση το μοντέλο των ενεργών περιγραμμάτων. Σε δημοσίευση που έγινε το 2004 για τον διαχωρισμό εικόνων με πολλά είδη φαγητών επιχειρήθηκε η μεθοδολογία επέκτασης και συνένωσης (stick growing and merging) (Sun & Du). Τέλος οι Anthimopoulos, Dehais, Diem και Mouggiakakou για την κατάτμηση των τροφών στην εικόνα χρησιμοποίησαν ένα συνδυασμό των αλγορίθμων Mean Shift-filtering, επέκταση περιοχών (Region Growing) και συγχώνευση περιοχών (Region Merging), όπως φαίνεται στην Εικόνα 3.6.

Οι παραπάνω μέθοδοι είναι κάποιες από τις σημαντικότερες που εμφανίστηκαν μέχρι το 2015. Στα επόμενα χρόνια το κομμάτι της κατάτμησης άρχισε να υλοποιείται και με αλγορίθμους μηχανικής μάθησης. Συγκεκριμένα σε δημοσίευση σχετικά με το σύστημα Im2Calories (Myers, et al., 2015), η κατάτμηση επιτυγχάνεται βασιζόμενη στο DeepLab, έναν αλγόριθμο κατάτμησης στηριζόμενο σε συνελκτικά νευρωνικά δίκτυα (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2017).



(α)



(β)



(γ)



(δ)

Εικόνα 3.6. Αλγόριθμος κατάτμησης των φαγητών που υπάρχουν σε μια εικόνα (α) αρχική εικόνα (β) η εικόνα μετά το *mean-shift-filtering* (γ) εικόνα μετά το *Region Growing* (δ) εικόνα μετά το *Region Merging* (Anthimopoulos M. , Dehais, Diem, & Mouggiakakou, 2013)

3.5 Ταξινόμηση Εικόνας

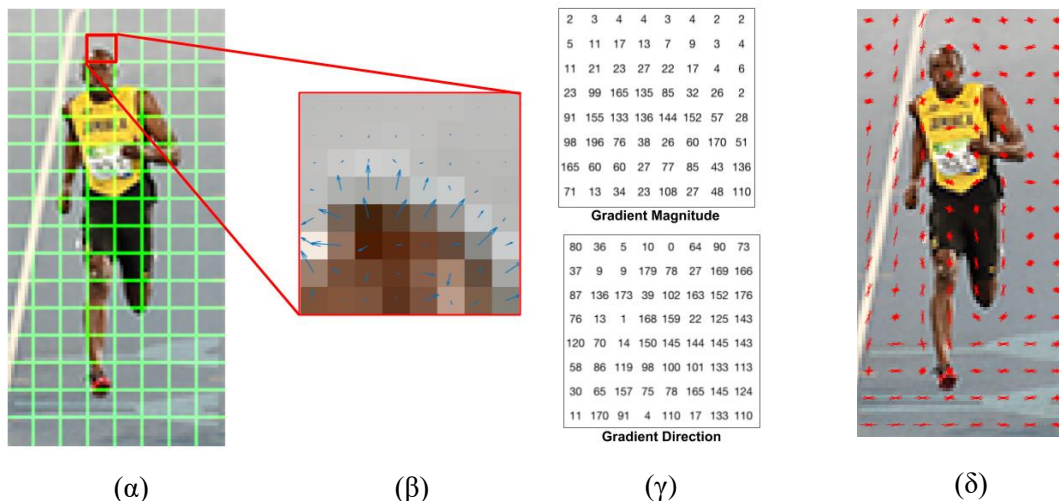
Η ταξινόμηση της εικόνας αποτελεί ένα από τα τρία βασικά στάδια για την δημιουργία εφαρμογής υπολογισμού των υδατανθράκων σε φωτογραφικά στιγμιότυπα, καθώς επίσης αποτελεί το στάδιο το οποίο θα μελετηθεί στην παρούσα διπλωματική εργασία. Η διαδικασία της ταξινόμησης διαφέρει από περίπτωση σε περίπτωση, αλλά γενικότερα το σχήμα το οποίο χρησιμοποιείται αποτελείται από τρία στάδια: Εξαγωγή Χαρακτηριστικών, Μείωση Διάστασης Χαρακτηριστικών και Ταξινόμηση. Οι δύο βασικοί μέθοδοι στην ταξινόμηση εικόνων είναι ο συμβατικός τρόπος που ακολουθεί την αλληλουχία των τριών σταδίων και τα Συνελκτικά Νευρωνικά Δίκτυα.

3.5.1 Εξαγωγή Χαρακτηριστικών

Ο συμβατικός τρόπος εξαγωγής χαρακτηριστικών ή εξαγωγή «χειροποίητων» χαρακτηριστικών (*hand-crafted features*) αποτελεί την εφαρμογή προκαθορισμένων φίλτρων ή ανάλυση της εικόνας με μια προκαθορισμένη αλληλουχία βημάτων για την περιγραφή της. Διαφέρει από τις νεότερες τεχνικές

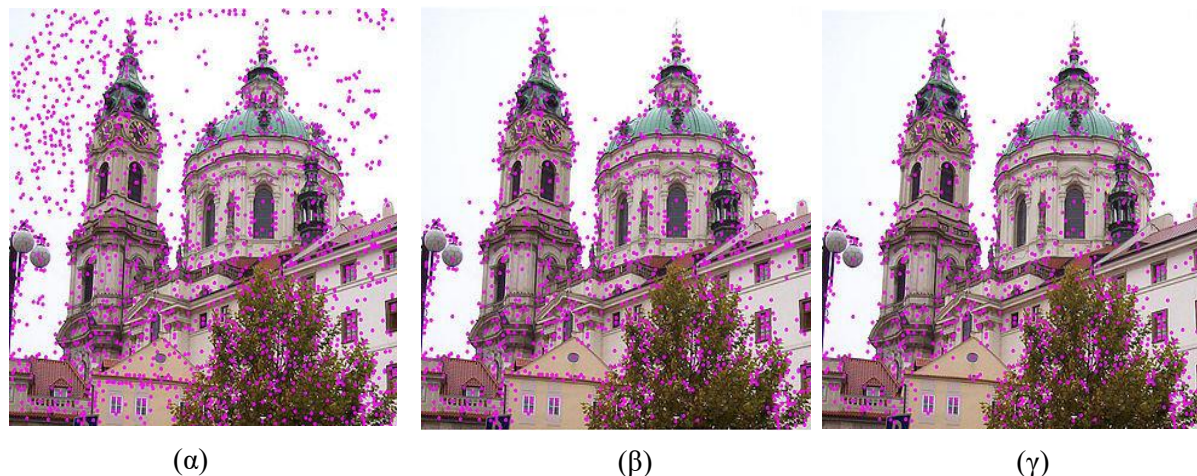
εξαγωγής χαρακτηριστικών, που χρησιμοποιούνται για παράδειγμα στα Συνελκτικά Νευρωνικά Δίκτυα, οι οποίες προσαρμόζουν τα φίλτρα τους στις ανάγκες του συνόλου δεδομένων.

Ο συμβατικός τρόπος εξαγωγής χαρακτηριστικών χρησιμοποιεί τοπικούς ή καθολικούς περιγραφείς για να περιγράψει την εικόνα (Bosch, Zhu, Khanna, Boushey, & Delp, 2011). Οι τοπικοί περιγραφείς στηρίζονται στην ιδέα προσδιορισμού της εικόνας μέσω των επιμέρους χωρίων που την αποτελούν. Δηλαδή γίνεται εξαγωγή χαρακτηριστικών από περιοχές μικρότερων διαστάσεων συγκριτικά με την εικόνα και η συνένωση τους αποτελεί μια αναπαράσταση της εικόνας. Από την άλλη μεριά, στην περίπτωση των καθολικών περιγραφέων, η φωτογραφία δε διασπάται σε επιμέρους χωρία, αλλά η επεξεργασία της γίνεται ως μια αδιάσπαστη ολότητα. Τα βασικά χαρακτηριστικά που συνήθως λαμβάνονται υπόψιν στην εικόνα είναι το χρώμα, η υφή (Asvestas, Matsopoulos, & Nikita, 1999) καθώς και οι μεταβολές οι οποίες γίνονται σε αυτά (Mougiakakou, Valavanis, Nikita, & Kelekis, 2003). Για την επεξεργασία αυτών των χαρακτηριστικών και άλλων έχουν αναπτυχθεί πολλές διαφορετικές τεχνικές. Οι σημαντικότερες μέθοδοι που επικρατούν στον τομέα της αναγνώρισης φαγητών σύμφωνα με το Παράρτημα Α είναι Ιστόγραμμα Χρώματος (Ciocca, Napoletano, & Schettini, 2016; Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014; Hoashi, Joutou, & Yanai, 2010), Scale Invariant Feature Transform (SIFT) (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014; Farinella, Allegra, & Stanco, 2015; Farinella, Moltisanti, & Battiato, 2014; Zhang, Yu, Siddiquie, Divakaran, & Sawhney, 2015), Speeded up Robust Features (SURF) (Kawano & Yanai, 2013; Miyazaki, De Silva, & Aizawa, 2011), Local Binary Pattern (LBP) (Ciocca, Napoletano, & Schettini, 2016; Anthimopoulos M. , Dehais, Diem, & Mougiakakou, 2013), Gabor Filters (Ciocca, Napoletano, & Schettini, 2016; Hoashi, Joutou, & Yanai, 2010; Chen, et al., 2012; Martinel, Piciarelli, Micheloni, & Foresti, 2015) και Histogram of Oriented Gradients (HOG) (Ciocca, Napoletano, & Schettini, 2016; Kawano & Yanai, 2013; Zhang, Yu, Siddiquie, Divakaran, & Sawhney, 2015). Στην Εικόνα 3.7 παρουσιάζεται ενδεικτικά η εξαγωγή χαρακτηριστικών με την μέθοδο HOG.



Εικόνα 3.7. Σχηματική αναπαράσταση του αλγορίθμου HOG (α) διαχωρισμός της εικόνας σε μπλοκ μέσω πλέγματος για την εξαγωγή τοπικών χαρακτηριστικών (β) Διανυσματική αναπαράσταση της HOG τιμής κάθε εικονοστοιχείου (γ) πίνακας πλάτους και διεύθυνσης για κάθε εικονοστοιχείο (δ) οπτικοποίηση της μεθόδου HOG (Satya, 2016).

Στην περίπτωση του αλγορίθμου SIFT, όπως και σε άλλους αλγόριθμους, δεν γίνεται απεικόνιση κάθε εικονοστοιχείου της εικόνας, αλλά επιλέγονται τα σημαντικότερα. Στην Εικόνα 3.8 απεικονίζεται μια σειρά βημάτων η οποία υλοποιείται κατά τον αλγόριθμο SIFT, για την επιλογή των πιο αντιπροσωπευτικών σημείων.



Εικόνα 3.8. Επιλογή σημείων εξαγωγής χαρακτηριστικών με την μέθοδο SIFT (α) Αφού έχουν εντοπισθεί τα scale-space extrema (β) απόρριψη των σημείων χαμηλής αντίθεσης (γ) απομάκρυνση σημείων κοντά στις ακμές (Wikipedia c. , 2017)

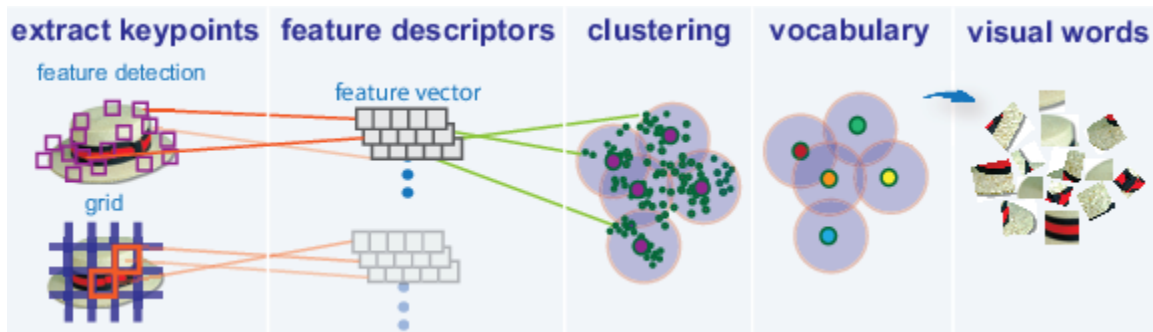
Οι περισσότερες δημοσιεύσεις συγκρίνουν κάποια από τα παραπάνω χαρακτηριστικά μεταξύ τους και μελετούν τις παραμέτρους εκείνες, οι οποίες επιστρέφουν τα καλύτερα αποτελέσματα ταξινόμησης, για κάθε χαρακτηριστικό (Joutou & Yanai, 2010; Hoashi, Joutou, & Yanai, 2010). Ίσως η πιο ενδεικτική παράμετρος που μπορεί να καθοριστεί στην περίπτωση των τοπικών χαρακτηριστικών είναι το μέγεθος των μπλοκ των εικονοστοιχείων από τα οποία θα εξαχθούν. Επιπρόσθετα ιδιαίτερο ενδιαφέρον εμφανίζουν

η μέθοδος των Textons που προτάθηκε για την περιγραφή εικόνων με φαγητά (Farinella, Allegra, & Stanco, 2015) και τα φίλτρα υφής που χρησιμοποιούνται από τους Martinel, Piciarelli, Micheloni και Foresti για την περιγραφή της εικόνας (2015). Τα τελευταία δεν καταφέρνουν να σημειώσουν αύξηση στην ακρίβεια ταξινόμησης.

Τέλος, αναπτύσσονται μέθοδοι εξαγωγής χαρακτηριστικών οι οποίες ξεφεύγουν από την τοπική περιγραφή κάθε εικονοστοιχείου και επιδιώκουν την χωρική συσχέτιση μεταξύ τους. Με τον τρόπο αυτό επιχειρείται να αναπαρασταθεί η χωρική διάταξη της εικόνας. Η βασική τεχνική προς αυτή την κατεύθυνση είναι τα Τοπικά Χαρακτηριστικά ανά Ζεύγη (Pairwise Local Features), η οποία αναπτύσσει μετρικές ανάμεσα σε ζεύγη ταξινομημένων εικονοστοιχείων. Οι μετρικές αυτές σχετίζονται με την απόσταση, την διεύθυνση και το είδος των εικονοστοιχείων που παρεμβάλλονται ανάμεσα στα ζεύγη εικονοστοιχείων (Yang, Chen, Pomerleau, & Sukthankar, 2010).

3.5.2 Μείωση Διάστασης Χαρακτηριστικών

Ανάλογα με τις διαστάσεις της εικόνας αλλά και της μεθοδολογίας εξαγωγής χαρακτηριστικών που χρησιμοποιείται κάθε φορά, συχνά δημιουργείται ένα πίνακας για την περιγραφή της εικόνας. Κάθε γραμμή του πίνακα αντιπροσωπεύει κάποιο σημείο της εικόνας, ενώ το πλήθος των στηλών αποτελεί ένα διάλυμα περιγραφής του. Εξαιτίας των πολλών χαρακτηριστικών που μπορούν να εξαχθούν από τις εικόνες, δημιουργείται η ανάγκη εξαγωγής των σημαντικότερων, με σκοπό τη δημιουργία ενός οπτικού λεξιλογίου. Προκειμένου να δημιουργηθεί το «λεξιλόγιο», δηλαδή ένα αντιπροσωπευτικό πλήθος χαρακτηριστικών που εξάγει η κάθε μέθοδος ταξινόμησης, χρησιμοποιείται η τεχνική της μείωσης χαρακτηριστικών. Σύμφωνα με την οποία, τα χαρακτηριστικά που εξάγονται από τις εικόνες του συνόλου εκπαίδευσης ομαδοποιούνται σε κ-κλάσεις και αντιστοιχίζονται στο κέντρο της κλάσης στην οποία ανήκουν, το οποίο αποτελεί τον αντιπρόσωπό της. Ο αντιπρόσωπος κάθε κλάσης αποτελεί «λέξη» του οπτικού λεξιλογίου και η διαδικασία αυτή γίνεται μέσω ενός αλγορίθμου ομαδοποίησης (k-mean clustering). Με τον τρόπο αυτό παράγεται το οπτικό λεξιλόγιο, όπως φαίνεται στην Εικόνα 3.9(α). Στην συνέχεια τα χαρακτηριστικά που εξάγονται από τις εικόνες αντιστοιχίζονται με έναν αλγόριθμο ταξινόμησης, όπως των κ-κοντινότερων γειτόνων (k-nearest neighbors), σε μια από τις ήδη κ-υπάρχουσες λέξεις του λεξιλογίου. Με αποτέλεσμα κάθε εικόνα να περιγράφεται από μια σειρά των κ λέξεων του λεξιλογίου. Η συχνότητα εμφάνισης της κάθε λέξης, η θέση της στην εικόνα καθώς και άλλοι συσχετισμοί αποδίδουν μια καλύτερη αναπαράσταση των εξαγόμενων χαρακτηριστικών, παράγοντας τελικά ένα διάλυμα περιγραφής της εικόνας (Εικόνα 3.9(β)).



(α)



(β)

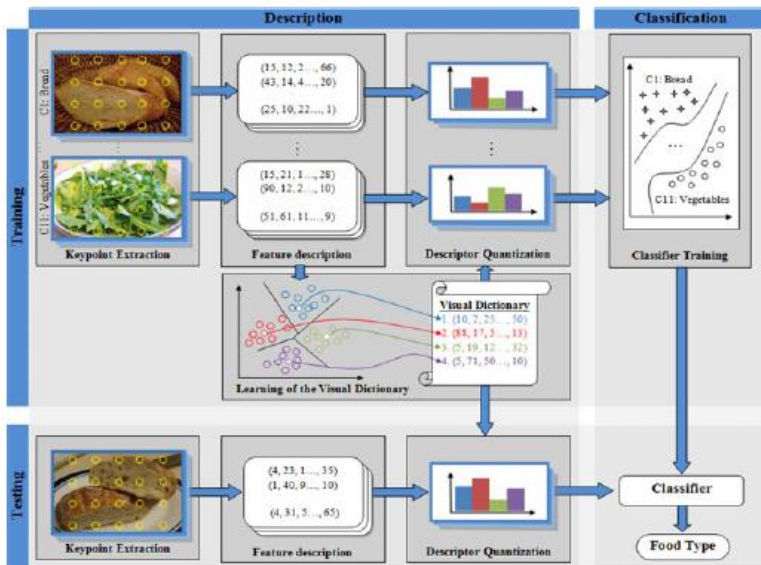
Εικόνα 3.9. (α) Δημιουργία Οπτικού Λεξιλογίου από την ομαδοποίηση των διανυσμάτων χαρακτηριστικών του συνόλου εκπαίδευσης (β) Περιγραφή κάθε εικόνας από το οπτικό λεξιλόγιο που έχει παραχθεί (MathWorks, n.d.).

Η βασικότερη τεχνική εξαγωγής χαρακτηριστικών είναι η Bag of Features (BoF), η οποία συνδυάζεται συνήθως με την μέθοδο εξαγωγής χαρακτηριστικών SIFT. Ενδελεχής μελέτη των παραμέτρων για την βελτιστοποίηση της μεθόδου γίνεται από τους Anthimopoulos, Gianola, Scarnato, Diem και Mougiakakou (2014). Στην συνέχεια οι Kawano και Yanai, προτείνουν την υλοποίηση της μείωσης των χαρακτηριστικών με την τεχνική των Fisher Vectors, σημειώνοντας καλύτερα αποτελέσματα (2013). Τέλος σε διάφορες δημοσιεύσεις τροποποιούνται οι παράμετροι της μεθόδου Fisher Vector, δημιουργώντας μια διαδικασία μείωσης χαρακτηριστικών με βελτιωμένα αποτελέσματα, την Improved Fisher Vector (IFV) (Yanai & Kawano, 2014).

Τέλος, μια επιπλέον μέθοδος που χρησιμοποιείται για την μείωση των διαστάσεων των εξαγόμενων χαρακτηριστικών από κάθε εικόνα είναι αυτή της Ανάλυσης Κύριων Συνιστωσών (Principal Component Analysis) (Bajwa, Naweed, Asif, & Hyder, 2009). Μέσω της μεθόδου αυτής επιτυγχάνεται ορθογώνιος μετασχηματισμός πιθανών συσχετιζόμενων μεταβλητών σε ένα πλήθος γραμμικών ανεξάρτητων μεταβλητών μεταξύ τους. Με τον τρόπο αυτό υλοποιείται μείωση του διανύσματος περιγραφής κάθε εικόνας, διαγράφοντας τα χαρακτηριστικά που δεν παρουσιάζουν ικανοποιητική διακύμανση και συνδυάζοντας χαρακτηριστικά με παρόμοια συμπεριφορά.

3.5.3 Ταξινόμηση

Το τελευταίο στάδιο στην διαδικασία της ταξινόμησης των εικόνων, είναι η ίδια η ταξινόμηση. Τα χαρακτηριστικά που προκύπτουν από κάθε μέθοδο εξαγωγής χαρακτηριστικών, είτε αυτά περάσουν από το στάδιο της ομαδοποίησης είτε όχι, αναλόγως την μέθοδο, τροφοδοτούν έναν ταξινομητή. Η συνηθέστερη μέθοδος ταξινόμησης στην αναγνώριση φαγητών είναι οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines) (Παράρτημα Β), οι οποίες πολλές φορές χρησιμοποιούνται με συνδυασμό πολλαπλών πυρήνων (Multiple Kernel Learning) (Joutou & Yanai, 2010; Hoashi, Joutou, & Yanai, 2010). Άλλες μέθοδοι ταξινόμησης που χρησιμοποιούνται σε εφαρμογές αναγνώρισης τροφίμων είναι τα Τυχαία Δάση (Martinel, Picciarelli, Micheloni, & Foresti, 2015; Bossard, Guillaumin, & Van Gool, 2014), Τεχνητά Νευρωνικά Δίκτυα (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014; Mougiakakou, Valavanis, Nikita, & Kelekis, 2003) και ο αλγόριθμος των κ-κοντινότερων γειτόνων (Merler, Wu, Uceda-Sosa, Nguyen, & Smith, 2016; Zhu, Bosch, Khanna, Boushey, & Delp, 2015; Ciocca, Napolitano, & Schettini, 2016; Papageorgiou, et al., 2001). Η μέθοδος εξαγωγής χαρακτηριστικών, η τεχνική αναπαράστασή τους και εν τέλει η επιλογή του ταξινομητή αποτελούν τις τρεις καθοριστικές επιλογές στην ακρίβεια που θα σημειώσει το σύστημα. Ένα ολοκληρωμένο σύστημα το οποίο επιτελεί σειριακά και τις τρεις διαδικασίες απεικονίζεται στην Εικόνα 3.10.



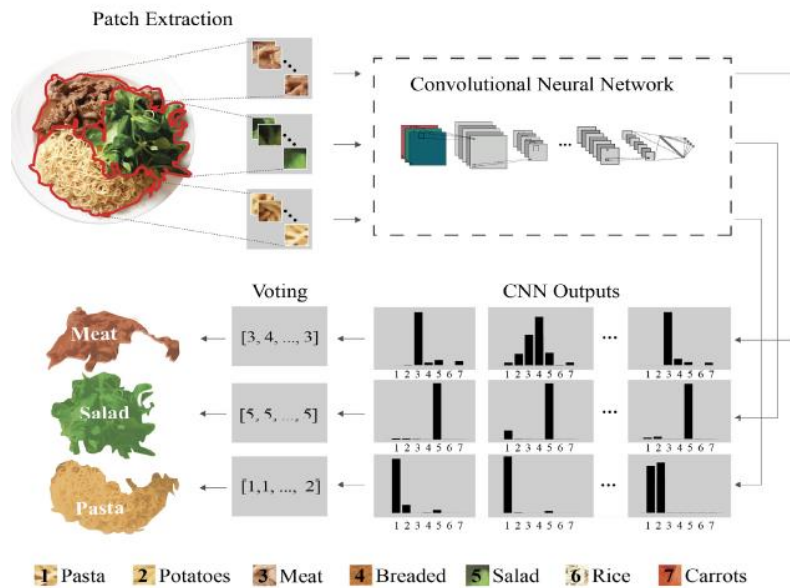
Εικόνα 3.10. Αρχιτεκτονική εξαγωγής χαρακτηριστικών μέσω της ακολουθίας Εξαγωγή Χαρακτηριστικών- Μείωση Διάστασης Χαρακτηριστικών- Ταξινόμηση (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014)

3.5.4 Συνελκτικά Νευρωνικά Δίκτυα

Τα Συνελκτικά Νευρωνικά Δίκτυα (ΣΝΔ) βασίζουν την λειτουργία τους στους τρεις βασικούς πυλώνες, οι οποίοι περιγράφηκαν προηγουμένως. Η βασική διαφοροποίηση τους είναι ότι όλες οι παραπάνω λειτουργίες μπορούν να υλοποιηθούν από το ίδιο το ΣΝΔ. Υπάρχουν όμως και περιπτώσεις στις οποίες το ΣΝΔ μπορεί να συνδυαστεί με ταξινομητές όπως τα SVM. Το βασικό πλεονέκτημα των ΣΝΔ είναι ότι η εξαγωγή χαρακτηριστικών δεν γίνεται από τυποποιημένες μεθόδους όπως προηγουμένως. Αντιθέτως η μέθοδος εξαγωγής χαρακτηριστικών, ή αλλιώς φίλτρα όπως αποκαλούνται στα ΣΝΔ, προσομοιώνονται από τα βάρη των νευρώνων του δικτύου και προσαρμόζονται στις ανάγκες κάθε συνόλου δεδομένων, μέσω της διαδικασίας της επιβλεπόμενης μάθησης.

Συγκεκριμένα, η εξαγωγή των χαρακτηριστικών γίνεται από τα στάδια συνέλιξης του νευρωνικού δικτύου, όπου μέσω της αναβάθμισης των βαρών των νευρώνων, προσδιορίζονται τα βέλτιστα φίλτρα για την περιγραφή των εικόνων. Η μείωση των χαρακτηριστικών αυτών επιτυγχάνεται μέσω δύο μεθόδων. Αρχικά χρησιμοποιούνται τα στρώματα δειγματοληψίας (pooling layers), τα οποία συγκρατούν μόνο την πιο χρήσιμη πληροφορία, μειώνοντας έτσι τις διαστάσεις, άρα και τις παραμέτρους, του δικτύου. Συμπληρωματικά, μετά τα στρώματα συνέλιξης του δικτύου, μπορεί να ακολουθήσουν ένα ή περισσότερα επίπεδα πλήρης σύνδεσης. Στην περίπτωση που υπάρχουν περισσότερα του ενός επίπεδα πλήρης σύνδεσης, γίνεται με έναν ακόμα τρόπο η μείωση των χαρακτηριστικών. Στην περίπτωση αυτή, δεν γίνεται επιλεκτική συγκράτηση της πιο σημαντικής γνώσης, όπως στο στάδιο δειγματοληψίας. Αντιθέτως ακολουθείται μια πιο αντιπροσωπευτική αναπαράσταση των ήδη εξαγόμενων χαρακτηριστικών. Αυτό επιτυγχάνεται μέσω δύο διαδοχικών επιπέδων νευρώνων πλήρης σύνδεσης, με το πρώτο να αποτελείται από περισσότερους νευρώνες συγκριτικά με το δεύτερο. Αν το πρώτο επίπεδο αποτελείται από X νευρώνες και το δεύτερο από Y , η δομή αυτή μεταφράζεται ως μετάβαση από από ένα διάνυσμα X χαρακτηριστικών της εικόνας, σε διάνυσμα Y χαρακτηριστικών. Τα παραπάνω θα αναλυθούν στο Κεφάλαιο 4.

Τέλος η μέχρι στιγμής αναπαράσταση των χαρακτηριστικών που χρησιμοποιείται από τα ΣΝΔ, μπορεί να επεξεργαστεί με δύο τρόπους. Μπορεί να δοθεί είτε ως διάνυσμα εισόδου σε έναν συμβατικό ταξινομητή ή διαμορφώνοντας κατάλληλα τα τελευταία στάδια του ΣΝΔ, προκειμένου να επιστρέφουν την πιθανότητα ταξινόμησης της εικόνας σε κάθε μια από τις πιθανές κλάσεις. Μια αναπαράσταση της ταξινόμησης εικόνων από ΣΝΔ, το οποίο στην προκειμένη περίπτωση ακολουθείται και από ένα σχήμα ψηφοφορίας, εμφανίζεται στην Εικόνα 3.11.



Εικόνα 3.11. Σχήμα εξαγωγής χαρακτηριστικών μέσω Συνελκτικών Νευρωνικών Δικτύων (Martinel, Piciarelli, Micheloni, & Foresti, 2015)

Στην βιβλιογραφία δοκιμάζονται διαφορετικές αρχιτεκτονικές Συνελκτικών Νευρωνικών Δικτύων για να βρεθεί εκείνη που θα παράγει τα καλύτερα αποτελέσματα. Οι Kagaya, Aizawa και Ogawa δημιουργούν Συνελκτικά Νευρωνικά Δίκτυα με διαφορετικό πλήθος στρωμάτων και διαφορετική επιλογή υπερπαραμέτρων κάθε φορά (2014). Αντίστοιχη διαδικασία ακολουθούν και οι Christodoulidis, Anthimopoulos και Mouggiakakou οι οποίοι πειραματίζονται στις υπερπαραμέτρους ΣΝΔ 6 επιπέδων (2015). Σημαντική διαφοροποίηση του συγκεκριμένου σχήματος με άλλα ΣΝΔ, αποτελεί το γεγονός ότι το δίκτυο τροφοδοτείται από τμήματα της εικόνας και όχι ολόκληρη την εικόνα. Αυτό έχει ως αποτέλεσμα να ακολουθεί ένα σχήμα ψηφοφορίας μετά το ΣΝΔ, προκειμένου να γίνει η ταξινόμηση των φαγητών της εικόνας. Η διαδικασία εξαγωγής χαρακτηριστικών από αυτή την εφαρμογή φαίνεται στην Εικόνα 3.11.

Στην συνέχεια χρησιμοποιούνται τα διακεκριμένα Συνελκτικά Νευρωνικά Δίκτυα AlexNet και GoogleNet, για την ταξινόμηση εικόνων φαγητού. Οι Kawano και Yanai, βασιζόμενοι στην δομή του ΣΝΔ AlexNet, μελέτησαν την σημασία των ιδιοτήτων του συνόλου δεδομένων με το οποίο θα προ-εκπαιδευθεί το δίκτυο, ενώ παράλληλα ανέδειξαν την ακρίβεια ταξινόμησης των ΣΝΔ συγκριτικά με άλλες μεθόδους ταξινόμησης (Kawano & Yanai, 2015). Επιπροσθέτως πέρα από το δίκτυο AlexNet, στην ταξινόμηση εικόνων φαγητού δοκιμάζεται το αναγνωρισμένο ΣΝΔ GoogleNet (Ao & Ling, 2015). Το GoogleNet στην συγκεκριμένη δημοσίευση συγκρίνεται με το AlexNet και με άλλους ταξινομητές.

3.6 Υπολογισμός Όγκου

Το τελευταίο στάδιο στην διαδικασία του υπολογισμού των μακροθρεπτικών συστατικών μιας εικόνας φαγητού, είναι ο υπολογισμός του όγκου των τροφίμων που απεικονίζονται. Στα πρώτα στάδια των εφαρμογών αναγνώρισης τροφής κάθε τροφή είτε συνδεόταν με προκαθορισμένη περιεκτικότητα σε μακροθρεπτικά συστατικά ή ο υπολογισμός αυτός γινόταν με την χρήση απλών γεωμετρικών σχημάτων. Τα τελευταία χρόνια επιχειρείται η τρισδιάστατη απεικόνιση των φαγητών για τον πιο ακριβή προσδιορισμό του όγκου, άρα και των μακροθρεπτικών συστατικών.

Ένας απλός τρόπος για τον υπολογισμό του όγκου της εικόνας είναι μέσω του on-line υπολογισμού του, από ανθρώπους, για παράδειγμα εκμεταλλεύοντας χρήστες από το Amazon Mechanical Turk (Noronha, Hysen, Zhang, & Gajos, 2011). Η μέθοδος αυτή είναι ικανοποιητική, παρ' όλα αυτά εμπεριέχει το ανθρώπινο σφάλμα. Μια άλλη ιδιαίτερα αποδοτική τεχνική είναι αυτή της κάμερας βάθους (depth camera), η οποία επιτρέπει την αναπαράσταση του βάθους άρα και του όγκου των φαγητών, αν και είναι δύσκολα προσβάσιμη (Chen, et al., 2012). Σύμφωνα με τα παρόντα τεχνολογικά μέσα η πιο βιώσιμη λύση είναι η λήψη πολλαπλών φωτογραφιών του φαγητού. Οι φωτογραφίες αυτές στην συνέχεια συνδυάζονται, μέσω της εύρεσης των κοινών σημείων μεταξύ τους από τις διάφορες λήψεις, δημιουργώντας στο τέλος μια τρισδιάστατη αναπαράσταση (Puri, Zhu, Yu, Divakaran, & Sawhney, 2009). Ακόμα μέσω των μεθόδων της προετοιμασίας του περιβάλλοντος, δηλαδή με την χρήση του αντίχειρα, αντικειμένων γνωστών διαστάσεων ή και σκακιέρας βαθμονόμησης υποβοηθείται ο υπολογισμός του όγκου των αντικειμένων που απεικονίζονται. Τέλος όταν ο χρήστης τρώει σε εστιατόριο, μέσω της εύρεσης της γεωγραφικής του τοποθεσίας αλλά και της ταύτισης του φαγητού που καταναλώνει από το κατάλογο (menu) του εστιατορίου, θα μπορεί να αντιστοιχίζεται αυτόματα το γεύμα του σε μακροθρεπτικά συστατικά (Bejibom, Joshi, Morris, Saronas, & Khullar, 2015). Αυτό προϋποθέτει την γνωστοποίηση του καταλόγου (menu) κάθε εστιατορίου αλλά και των μακροθρεπτικών συστατικών κάθε πιάτου.

3.7 Συζήτηση

Παραπάνω αναφέρθηκαν τα βασικά βήματα που χρησιμοποιούνται στις εφαρμογές αυτόματης αναγνώρισης μακροθρεπτικών συστατικών από φωτογραφικά στιγμιότυπα φαγητών. Στην συγκεκριμένη εργασία θα γίνει εστίαση στο στάδιο ταξινόμησης της εικόνας και ιδιαίτερα στο κομμάτι υπολογισμού της περιεχόμενης ποσότητας υδατανθράκων. Όπως υπογραμμίστηκε προηγουμένως υπάρχουν δύο εναλλακτικές στην διαδικασία ταξινόμησης των εικόνων. Η πρώτη αποτελείται από την εξαγωγή «χειροποίητων» χαρακτηριστικών, ακολουθούμενη από έναν ταξινομητή. Στην δεύτερη περίπτωση

χρησιμοποιούνται πιο προσαρμοστικές τεχνικές εξαγωγής χαρακτηριστικών, με βασικό τους αντιπρόσωπο τα Συνελικτικά Νευρωνικά Δίκτυα.

Στο Παράρτημα Β, απεικονίζονται κάποιες από τις σημαντικότερες δημοσιεύσεις στο κομμάτι της αυτοματοποιημένης αναγνώρισης τροφίμων από φωτογραφικά στιγμιότυπα. Για κάθε δημοσίευση παρατίθεται το σύνολο δεδομένων στο οποίο βασίζεται, οι μέθοδοι εξαγωγής χαρακτηριστικών, μείωσης χαρακτηριστικών και ταξινόμησης που χρησιμοποιούνται καθώς και το βέλτιστο ποσοστό ακρίβειας το οποίο επιτυγχάνεται. Οι δημοσιεύσεις ταξινομούνται σε φθίνουσα χρονολογική σειρά και παρουσιάζουν τις επικρατέστερες τάσεις που έχουν αναπτυχθεί τα τελευταία επτά χρόνια. Αξίζει να σημειωθεί ότι οι δημοσιεύσεις αυτές αποτελούν ένα μικρό τμήμα της βιβλιογραφίας.

Παρουσιάζοντας πολύ συνοπτικά την πληροφορία που εμπεριέχεται στο Παράρτημα Β, φαίνεται ότι στις πρώτες δημοσιεύσεις χρησιμοποιούνται συμβατικές μέθοδοι εξαγωγής χαρακτηριστικών (handcrafted methods), που δεν εμπεριέχουν καμία προσαρμοστικότητα ως προς το σύνολο δεδομένων στο οποίο απευθύνονται. Οι μέθοδοι αυτές συγκρίνονται μεταξύ τους και συνδυάζονται δημιουργώντας μεγαλύτερα διανύσματα περιγραφής, τα οποία ακολούθως τροφοδοτούν κάποιον ταξινομητή. Στην συνέχεια μελετήθηκαν οι μέθοδοι μείωσης των χαρακτηριστικών περιγραφής της εικόνας, ξεκινώντας από την πιο συνήθη, την Bag of Features, και καταλήγοντας στην μείωση των χαρακτηριστικών μέσω της τεχνικής των Improved Fisher Vectors. Οι μεταβάσεις αυτές επέφεραν βελτιώσεις στην ακρίβεια ταξινόμησης, αλλά με περιορισμένη προοπτική. Τελικώς μετά την δοκιμή διαφόρων τεχνικών εξαγωγής χαρακτηριστικών και αλγορίθμων ταξινόμησης εμφανίστηκε η μεθοδολογία των Συνελικτικών Νευρωνικών Δικτύων, η οποία εμφανίζεται σε όλο και μεγαλύτερο πλήθος δημοσιεύσεων τα τελευταία χρόνια.

Αν και ακόμα δεν έχουν παρουσιαστεί τα σύνολα δεδομένων, τα οποία θα αναλυθούν στο Κεφάλαιο 5, φαίνεται ότι κάποια από τα σύνολα δεδομένων που αναφέρονται στο Παράρτημα Β είναι τα UEC-FOOD 100, UEC-FOOD 256, Food-101 και UNIMIB-2016. Η βέλτιστη ακρίβεια στα δύο πρώτα σύνολα δεδομένων είναι αντίστοιχα 78,77% και 67,57%, η οποία προέκυψε από fine-tuned Συνελικτικό Νευρωνικό Δίκτυο (Kawano & Yanai, 2015). Αντίστοιχα με την χρήση του Συνελικτικού Νευρωνικού Δικτύου GoogleNet, επιτυγχάνεται ακρίβεια 78,11% στο σύνολο δεδομένων Food-101. Τέλος αυτή τη φορά, χρησιμοποιώντας τα Συνελικτικό Νευρωνικό Δίκτυο για την εξαγωγή χαρακτηριστικών επιτυγχάνεται ακρίβεια 82,00% στο σύνολο δεδομένων UNIMIB-2016. Σύμφωνα με τα προηγούμενα φαίνεται ότι η πιο υψηλή ακρίβεια, σε κάποια από τα γνωστά σύνολα δεδομένων, μονοπωλείται από την εφαρμογή των Συνελικτικών Νευρωνικών Δικτύων. Άρα, το σημαντικότερο συμπέρασμα που προκύπτει από την βιβλιογραφική ανασκόπηση είναι, ότι αν και έχουν αναπτυχθεί κατά καιρούς πολλές διαφορετικές μέθοδοι

ταξινόμησης, αυτή τη στιγμή ο επικρατέστερος μηχανισμός για την εξαγωγή χαρακτηριστικών και ταξινόμηση εικόνων είναι τα Συνελικτικά Νευρωνικά Δίκτυα.

4 Μεθοδολογία

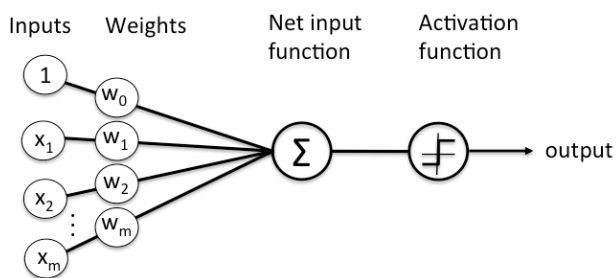
4.1 Νευρωνικά Δίκτυα

4.1.1 Το Perceptron

Οι νευρώνες αποτελούν τις στοιχειώδεις μονάδες των νευρωνικών δικτύων. Με τον όρο νευρώνα συνήθως γίνεται αναφορά στο perceptron του Rosenblatt, το οποίο βασίζεται σε έναν μη γραμμικό νευρώνα. Το νευρωνικό μοντέλο αυτό αποτελείται από έναν γραμμικό συνδυαστή, ακολουθούμενο από έναν απότομο περιοριστή (συνάρτηση ενεργοποίησης). Ο κόμβος άθροισης του νευρωνικού μοντέλου υπολογίζει ένα γραμμικό συνδυασμό των εισόδων που εφαρμόζονται στις συνάψεις του και ενσωματώνει επίσης μια εξωτερικά εφαρμοζόμενη «προδιάθεση» ή «πόλωση». Το προκύπτον άθροισμα εφαρμόζεται σε έναν απότομο περιοριστή (Haykin,2010).

Όπως απεικονίζεται στην Εικόνα 4.1 οι εισόδοι που δίνονται στο perceptron συμβολίζονται με x_1, x_2, \dots, x_m και πολλαπλασιάζονται με τα βάρη w_0, w_1, \dots, w_m . Το βάρος w_0 πολλαπλασιάζεται με την μονάδα και ουσιαστικά αντιπροσωπεύει την πόλωση.

Σαν μονάδα το perceptron, μπορεί να λειτουργήσει ως γραμμικός ταξινομητής δύο κλάσεων. Ως απότομος περιοριστής στο συγκεκριμένο παράδειγμα χρησιμοποιείται η συνάρτηση προσήμου, για την καλύτερη αναπαράσταση του διαχωρισμού σε δύο κλάσεις. Συγκεκριμένα αντιστοιχίζεται η έξοδος +1 στα στοιχεία της μια κλάσης C_1 ενώ η έξοδος -1 στην δεύτερη κλάση C_2 .



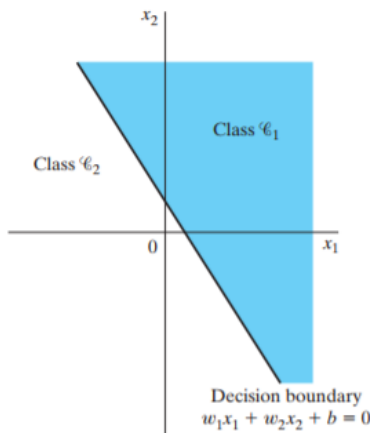
Εικόνα 4.1. Το Perceptron του Rosenblatt (Raschka, 2015)

Για να γίνει καλύτερα κατανοητή η διαδικασία ταξινόμησης που επιτελεί το perceptron, μπορεί τα στοιχεία εισόδου του να απεικονιστούν σε ένα m -διάστατο χώρο, δηλαδή όσο και το μήκος τους. Το perceptron ορίζει ένα υπερεπίπεδο στην περιοχή αυτή μέσω της συνάρτησης (4.1). Το βάρος w_0 , το οποίο

αναπαριστά την πόλωση, μπορεί να συμβολιστεί ως b (bias), για να διακρίνεται πιο εύκολα από τα υπόλοιπα βάρη.

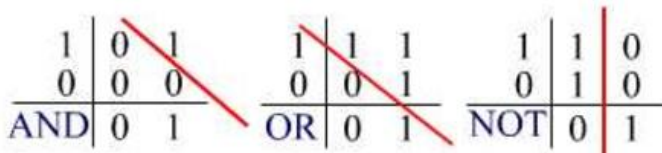
$$\sum_{i=1}^m w_i x_i + b = 0 \quad (4.1)$$

Στην απλούστερη περίπτωση, η είσοδος του perceptron αποτελείται από δύο μεταβλητές. Έτσι δημιουργείται διδιάστατος χώρος για την απεικόνιση των σημείων εισόδου και το perceptron λειτουργεί σαν ευθεία γραμμή προκειμένου να διακρίνει τις δύο κλάσεις του επιπέδου. Όπως φαίνεται στην Εικόνα 4.2, το perceptron τελικώς δημιουργεί ένα γραμμικό διαχωρισμό του χώρου, με τον όρο b να προσδιορίζει την απόσταση του υπερεπιπέδου από την αρχή των αξόνων. Η προσαρμογή των βαρών επιτελείται μέσω μια επαναληπτικής διαδικασίας.



Εικόνα 4.2. Γραμμικός Διαχωρισμός του χώρου από το Perceptron (Haykin, 2010)

Όπως αναφέρθηκε προηγουμένως, το perceptron, μπορεί να ταξινομεί μεταβλητές εισόδου όταν αυτές είναι γραμμικά διαχωρίσιμες. Οπότε λογικοί συντελεστές όπως το AND, OR και το NOT, μπορούν να επιλυθούν με την χρήση ενός perceptron. Μια απλή αναπαράσταση του διαχωρισμού αυτού, φαίνεται στην Εικόνα 4.3



Εικόνα 4.3. Γραμμικά διαχωρίσιμες λογικές συναρτήσεις (Stanford, n.d.)

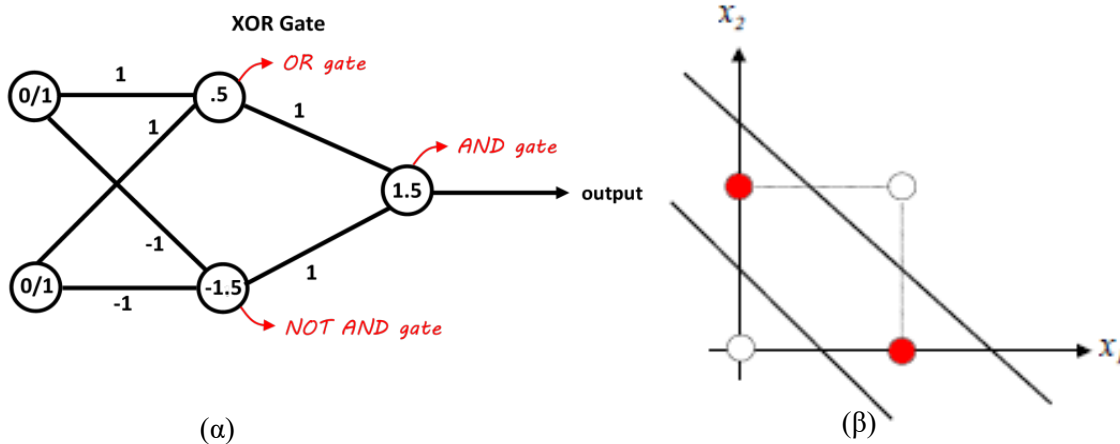
Τα περισσότερα όμως προβλήματα του πραγματικού κόσμου δεν είναι γραμμικά διαχωρίσιμα. Ένα απλό παράδειγμα είναι ο λογικός τελεστής XOR, που απεικονίζεται στην Εικόνα 4.4. Για το λόγο αυτό τα

perceptrons οργανώθηκαν σε επίπεδα, δημιουργώντας δίκτυα, για την επίλυση πιο σύνθετων μη γραμμικών δικτύων, που ονομάστηκαν νευρωνικά δίκτυα.

$$\begin{array}{c|cc} 1 & 1 & 0 \\ 0 & 0 & 1 \\ \hline \text{XOR} & 0 & 1 \end{array}$$

Εικόνα 4.4. Λογικός τελεστής XOR (Stanford, n.d.)

Παρακάτω, απεικονίζεται ένα νευρωνικό δίκτυο, δύο επιπέδων, που μπορεί να επιλύσει το πρόβλημα του XOR. Κάθε νευρώνας του μεσαίου επιπέδου, αποτελεί ένα perceptron, συνεπώς ορίζει μια ευθεία στο επίπεδο. Η λύση για το πρόβλημα του XOR, αποτελεί την ενδιάμεση λωρίδα που ορίζεται μεταξύ των δύο ευθειών που εμφανίζονται στην Εικόνα 4.5β. Συνεπώς για να ορισθεί το χωρίο αυτό, προστίθεται στο τελευταίο επίπεδο ένα perceptron που επιτελεί την λειτουργία του λογικού AND, επιστρέφοντας δηλαδή 1, στην περίπτωση που μόνο και τα δύο perceptrons του μεσαίου επιπέδου επιστρέφουν 1.

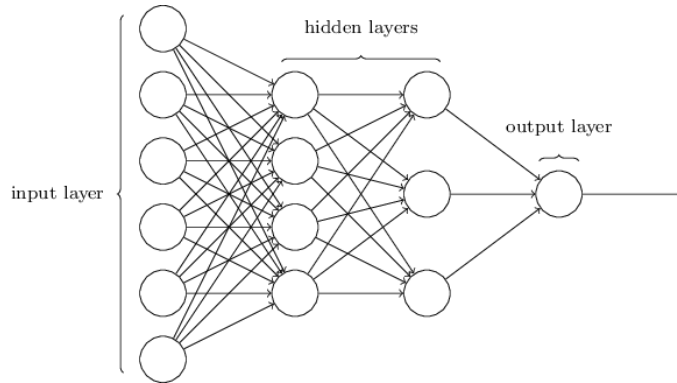


Εικόνα 4.5. (α) νευρωνικό δίκτυο για την επίλυση του προβλήματος του XOR (Neel, 2015) (β) Το υπερεπίπεδο απόφασης που δημιουργεί το νευρωνικό δίκτυο

4.1.2 Δομή Νευρωνικού Δικτύου

Όπως περιγράφηκε παραπάνω, τα perceptrons οργανώνονται σε επίπεδα. Οι νευρώνες κάθε επιπέδου συνδέονται με ένα σύνολο των νευρώνων του προηγούμενου επιπέδου και παράγουν μια έξοδο, που θα αποτελέσει την είσοδο για τους νευρώνες του επόμενου επιπέδου. Δεν υπάρχει καμία σύναψη, δηλαδή καμία σύνδεση, ανάμεσα στους νευρώνες του ίδιου επιπέδου. Οι μεταβλητές εισόδου εισάγονται στο δίκτυο μέσω του στρώματος εισόδου (input layer). Στην συνέχεια τα επόμενα επίπεδα του δικτύου, εκτός του τελικού, ονομάζονται κρυφά επίπεδα (hidden layers), τα οποία επεξεργάζονται τις μεταβλητές

εισόδου, μέσω συνδέσεων που καθεμία έχει το δικό της βάρος. Το τελευταίο επίπεδο του δικτύου, ονομάζεται επίπεδο εξόδου, όπου αναπαρίσταται η έξοδος, το αποτέλεσμα δηλαδή που επιστρέφει το δίκτυο. Η δομή ενός νευρωνικού δικτύου απεικονίζεται στην Εικόνα 4.6.



Εικόνα 4.6. Αναπαράσταση νευρωνικού δικτύου 3 επιπέδων, αποτελούμενο από δύο κρυφά επίπεδα και ένα επίπεδο εξόδου (Nielsen, 2015)

4.1.3 Εκπαίδευση Νευρωνικού Δικτύου

Σκοπός του νευρωνικού δικτύου είναι ο προσδιορισμός των βαρών κάθε νευρώνα, προκειμένου για κάθε διάνυσμα εισόδου να επιστρέφεται το αναμενόμενο αποτέλεσμα από το επίπεδο εξόδου. Η εκπαίδευση του νευρωνικού δικτύου επιτυγχάνεται με την μέθοδο της επιβλεπόμενης μάθησης (supervised learning).

4.1.3.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση αποτελεί τη διαδικασία εκτίμησης μιας συνάρτησης με αξιοποίηση ενός συνόλου εκπαίδευσης μαζί με την επιθυμητή απόκριση. Στην περίπτωση της επιβλεπόμενης μάθησης το σύνολο εκπαίδευσης δίνεται στην μορφή:

$$J = \{ x(n), d(n) \}_{n=1}^N$$

δηλαδή αποτελείται από ένα πλήθος N ζευγαριών, όπου $x(n)$ είναι το διάνυσμα εισόδου και $d(n)$ να η αναμενόμενη απόκριση του συστήματος. Ένας αλγόριθμος επιβλεπόμενης μάθησης αναλύει το σύνολο εκπαίδευσης και παράγει ένα σύστημα που, δεδομένου των διανυσμάτων εισόδου, μπορεί να εξάγει την έξοδο που αντιστοιχεί σε αυτό. Το παραγόμενο σύστημα ιδανικά θα πρέπει να μπορεί, μέσω γενίκευσης, να υπολογίζει την απόκριση διανυσμάτων εισόδου τα οποία δεν εμπεριέχονται στο σύνολο εκπαίδευσης.

Το σήμα σφάλματος που παράγεται από ένα νευρωνικό δίκτυο, αναπαρίσταται από τον τύπο (4.2). Με τον όρο $y_j(n)$, συμβολίζεται η απόκριση που παράγεται στον j -νευρώνα εξόδου από την είσοδο $x(n)$.

Αντίστοιχα η μεταβλητή $d_j(n)$ αναπαριστά την αναμενόμενη έξοδο, από τον συγκεκριμένο νευρώνα. Συνεπώς ως σφάλμα εξόδου ενός νευρώνα εξόδου, ορίζεται η διαφορά ανάμεσα στην αναμενόμενη και την προβλεπόμενη τιμή εξόδου.

$$e_j(n) = d_j(n) - y_j(n) \quad (4.2)$$

Η συνάρτηση κόστους που επιλέγεται εκφράζεται ως άθροισμα των τετραγωνικών σφαλμάτων, ενώ συμπεριλαμβάνεται και ο αριθμός $\frac{1}{2}$ για την απλοποίηση των συντελεστών μετά τη διαδικασία παραγωγισής της. Η συνάρτηση κόστους παίρνει την μορφή της σχέσης (4.3) για ένα νευρώνα του επιπέδου εξόδου του νευρωνικού δικτύου και την μορφή (4.4) για το σύνολο των νευρώνων αυτού του επιπέδου.

$$E_j(n) = \frac{1}{2} e_j^2(n) \quad (4.3)$$

$$E(n) = \sum_{j \in C} E_j(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (4.4)$$

Ενώ η μέση ενέργεια εξόδου, δεδομένου ότι το δείγμα εκπαίδευσης αποτελείται από N παραδείγματα, ορίζεται από την σχέση (4.5):

$$E_{av}(n) = \frac{1}{N} \sum_{n=1}^N E(n) = \frac{1}{2} \sum_{n=1}^N \sum_{j \in C} e_j^2(n) \quad (4.5)$$

4.1.3.1.1 Μαζική Μάθηση – Batch Gradient Descent

Στην μαζική επιβλεπόμενη μάθηση, οι προσαρμογές στα συναπτικά βάρη του νευρωνικού δικτύου, γίνονται αφού αναλυθούν και οι N παρατηρήσεις του συνόλου δεδομένων, οι οποίες συνιστούν μια εποχή. Με αποτέλεσμα στην περίπτωση αυτή, η συνάρτηση κόστους να καθορίζεται από την μέση ενέργεια σφάλματος E_{av} και η προσαρμογή των συνοπτικών βαρών κάθε perceptron να γίνεται στο τέλος της εποχής. Η μέθοδος αυτή παράγει αντιπροσωπευτική εκτίμηση του διανύσματος κλίσης, επειδή λαμβάνει υπόψιν όλο το σύνολο εκπαίδευσης. Επιπλέον, αφού οι αλλαγές στις τιμές των συναπτικών βαρών δεν εκτελούνται, παρά μόνο όταν υπολογιστούν τα σφάλματα όλων των παρατηρήσεων εισόδου, για το λόγο αυτό επιτρέπεται η παράλληλη επεξεργασία. Κατά την διάρκεια της μαζικής μάθησης χρειάζεται να μεταβεί ολόκληρο το σύνολο δεδομένων στην μνήμη RAM, δημιουργώντας σημαντικά προβλήματα στην περίπτωση που αυτό ξεπερνάει σε μέγεθος την μνήμη. Τέλος η μειωμένη συχνότητα αναβάθμισης των βαρών οδηγεί από τη μια πλευρά σε πιο σταθερή κλίση σφάλματος, από την άλλη πλευρά όμως αυτό μπορεί να έχει ως αποτέλεσμα πρώιμη σύγκλιση (Ruder, 2016). Η ανανέωση των βαρών περιγράφεται από τον ακόλουθο τύπο (4.6), με το η να αποτελεί το ρυθμό μάθησης.

$$w = w - \eta \nabla_w E(w) \quad (4.6)$$

4.1.3.1.2 Στοχαστική Βαθμωτή Κατάβαση (ΣΒΚ) - Stochastic Gradient Descent (SGD)

Στην περίπτωση της στοχαστικής βαθμωτής κατάβασης οι προσαρμογές στα συναπτικά βάρη εκτελούνται για κάθε διάνυσμα εισόδου της εκπαίδευσης. Στην περίπτωση της μαζικής μάθησης γίνονται περιττοί υπολογισμοί, καθώς επαν-υπολογίζεται η κλίση παρόμοιων παραδειγμάτων χωρίς πρώτα να έχουν αναβαθμιστεί τα βάρη. Το πρόβλημα αυτό διορθώνεται με το ΣΒΚ, οδηγώντας σε πιο γρήγορη εκπαίδευση. Από την άλλη πλευρά οι συνεχείς ανανεώσεις των παραμέτρων του δικτύου από κάθε διάνυσμα εισόδου, οδηγούν σε μεγάλες διακυμάνσεις στην κλίση της συνάρτησης σφάλματος. Αυτό έχει ως αποτέλεσμα από την μια μεριά την δυνατότητα μεταβίβασης σε καλύτερα τοπικά ελάχιστα, ενώ από την άλλη μεριά δυσχεραίνει την διαδικασία σύγκλισης σε αυτά. Ο τύπος (4.7) περιγράφει την ανανέωση των βαρών, με κάθε παρατήρηση εισόδου να συμβολίζεται ως $x^{(i)}$ και ως $y^{(i)}$ να συμβολίζεται η κλάση στην οποία ανήκει (Ruder, 2016).

$$w = w - \eta \nabla_w E(w; x^{(i)}; y^{(i)}) \quad (4.7)$$

4.1.3.1.3 Βαθμωτή Κατάβαση Μικρής Παρτίδας - Mini Batch Gradient

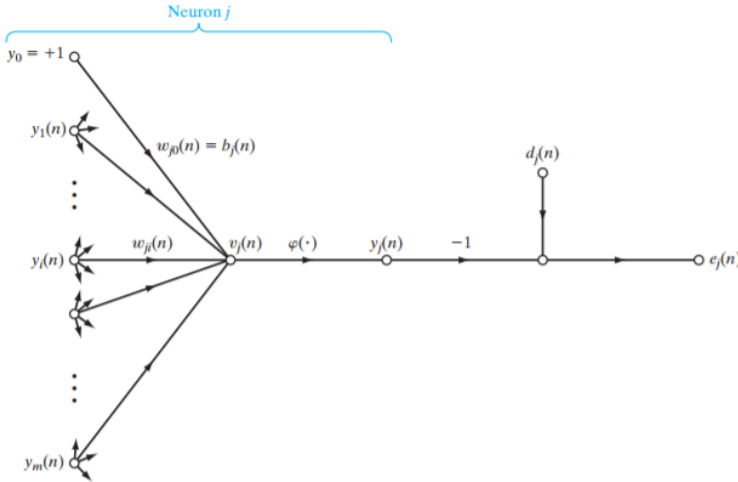
Η μέθοδος της βαθμωτής κατάβασης μικρής παρτίδας διορθώνει τις παραμέτρους του δικτύου με μια μερίδα από το σύνολο των παρατηρήσεων εισόδου. Με τον τρόπο αυτό συνδυάζει τα πλεονεκτήματα και των δύο προαναφερθέντων τεχνικών. Συγκεκριμένα από την μια μεριά δεν παρουσιάζει τις μεγάλες διακυμάνσεις που υπήρχαν στην στοχαστική βαθμωτή κατάβαση, γιατί κάθε μερίδα παρατηρήσεων αποτελεί αντιπροσωπευτική απεικόνιση του συνόλου δεδομένων, οδηγώντας έτσι σε πιο σταθερή σύγκλιση, ενώ καθίσταται δυνατή η μετάβαση σε καλύτερα τοπικά ελάχιστα της συνάρτησης σφάλματος. Η σχέση ανανέωσης των βαρών φαίνεται στην εξίσωση (4.8), με κάθε παρτίδα να αποτελείται από $k + 1$ παρατηρήσεις.

$$w = w - \eta \nabla_w E(w; x^{(i:i+k)}; y^{(i:i+k)}) \quad (4.8)$$

Οι διαδικασίες της στοχαστικής βαθμωτής κατάβασης και της βαθμωτής κατάβασης μικρής παρτίδας ορίζουν την on-line μάθηση στα νευρωνικά δίκτυα, δηλαδή την ανανέωση των βαρών του δικτύου χωρίς να προσπελαστεί ολόκληρο το σύνολο δεδομένων, αλλά ένα μέρος του (Ruder, 2016).

4.1.3.2 Backpropagation

Ένας ακόμα λόγος για την εδραίωση της on-line μάθησης, αποτελεί ο αλγόριθμος του backpropagation, σύμφωνα με τον οποίο γίνεται η ανανέωση των συναπτικών βαρών.



Εικόνα 4.7. Νευρώνας j (Haykin, 2010)

Στην Εικόνα 4.7 απεικονίζεται ένας τυχαίος νευρώνας j του δικτύου. Όπως έχει αναφερθεί, η είσοδος που φτάνει στην συνάρτηση ενεργοποίησης του perceptron, ορίζεται μέσω της σχέσης (4.9) και ονομάζεται τοπικό πεδίο $u_j(n)$.

$$u_j(n) = \sum_{i=0}^m w_{ji}(n)y_i(n) \quad (4.9)$$

Η πόλωση απεικονίζεται, ως βάρος w_{j0} με μοναδιαία είσοδο. Ο συμβολισμός w_{ji} αναφέρεται στο βάρος του νευρώνα j , που λαμβάνει ως είσοδο την έξοδο του νευρώνα i του προηγούμενου επιπέδου. Η έξοδος του νευρώνα j κατά την επανάληψη n (on-line μάθηση), προκύπτει ως το αποτέλεσμα του τοπικού πεδίου $u_j(n)$ αφού περάσει από την συνάρτηση ενεργοποίησης και συμβολίζεται ως $y_j(n)$.

$$y_j(n) = \varphi_j(u_j(n)) \quad (4.10)$$

Η διόρθωση των βαρών, σύμφωνα με τον αλγόριθμο του backpropagation, είναι ανάλογη της μερικής παραγώγου της συνάρτησης κόστους ως προς τα βάρη και σύμφωνα με τον κανόνα της αλυσίδας διαφορικού μπορεί να γραφεί στην μορφή της σχέσης (4.11).

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (4.11)$$

Η μερική παράγωγος των βαρών $\frac{\partial E(n)}{\partial w_{ji}(n)}$ αντιπροσωπεύει έναν συντελεστή ευαισθησίας, ο οποίος καθορίζει την κατεύθυνση της αναζήτησης στο χώρο βαρών για το συναπτικό βάρος w_{ji} (Haykin, 2010).

Μέσω των σχέσεων (4.2) και (4.11) προκύπτει ότι ο πρώτος παράγοντας της αλυσίδας παραγοντικού ισούται με $e_j(n)$.

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n) \quad (4.12)$$

Επίσης, με παραγωγή της σχέσης (4.2) ως προς y_j , δίνεται

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (4.13)$$

Τέλος διαφορίζοντας τις σχέσεις (4.5) και (4.4) αντιστοίχως προκύπτουν οι εξισώσεις (4.14) και (4.15), όπως φαίνεται παρακάτω.

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi_j'(u_j(n)) \quad (4.14)$$

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (4.15)$$

Αναλύοντας κάθε έναν από τους 4 παράγοντες της αλυσίδας διαφορικού, τελικά προκύπτει ότι η εξίσωση (4.11), παίρνει την μορφή της εξίσωσης (4.16).

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -e_j(n) \varphi_j'(u_j(n)) y_i(n) \quad (4.16)$$

Από τον κανόνα δέλτα η γενική μορφή της διόρθωσης των βαρών γίνεται

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \quad (4.17)$$

με το η να συμβολίζει την παράμετρο του ρυθμού μάθησης, ενώ το αρνητικό πρόσημο να υποδηλώνει την κατάβαση στο χώρο των βαρών. Αντικαθιστώντας την σχέση (4.16) στην σχέση (4.17), προκύπτει η σχέση (4.18), η οποία περιγράφει την προσαρμογή του i -οστού βάρους του j νευρώνα, συναρτήσει της εισόδου $y_i(n)$.

$$\Delta w_{ji}(n) = -\eta \delta_j(n) y_i(n) \quad (4.18)$$

Η τοπική κλίση $\delta_j(n)$, έχει την μορφή της σχέσης (4.19) και όπως φαίνεται εξαρτάται από το σφάλμα $e_j(n)$ του αντίστοιχου νευρώνα και την παράγωγο της σχετιζόμενης συνάρτησης ενεργοποίησης.

$$\delta_j(n) = \frac{\partial E(n)}{\partial u_j(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = e_j(n) \varphi'_j(u_j(n)) \quad (4.19)$$

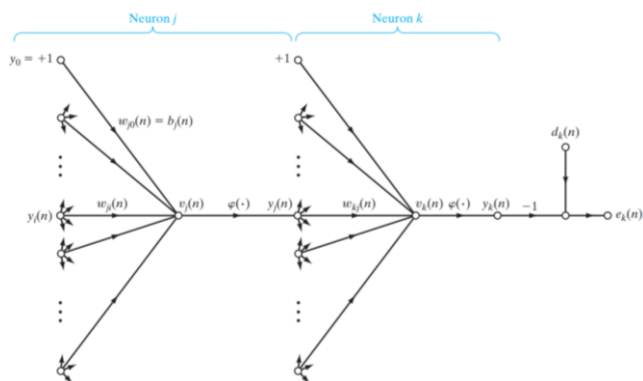
Στην περίπτωση που ο νευρώνας βρίσκεται στο επίπεδο εξόδου, είναι εύκολος ο υπολογισμός της μεταβλητής του σφάλματος $e_j(n)$. Αντιθέτως ο υπολογισμός της τοπικής κλίσης αποτελεί πιο σύνθετη διαδικασία στη περίπτωση των νευρώνων του κρυφού επιπέδου, καθώς δεν έχει προσδιοριστεί η μεταβλητή του σφάλματος

Νευρώνας εξόδου

Όταν ο κόμβος j , είναι κόμβος εξόδου τότε η μεταβλητή $e_j(n)$ υπολογίζεται κατευθείαν από την σχέση (4.2). Αφού υπολογιστεί το σήμα σφάλματος, στην συνέχεια ο υπολογισμός της τοπικής κλίσης $\delta_j(n)$ γίνεται από την σχέση (4.19).

Νευρώνας Κρυφού Επιπέδου

Για τους νευρώνες κρυφού επιπέδου δεν υπάρχει καθορισμένη επιθυμητή απόκριση, συνεπώς το σφάλμα εξόδου δεν μπορεί να υπολογιστεί με τον τρόπο που υπολογίζεται σε νευρώνες εξόδου. Μέσω μιας διαδικασίας οπισθοδρόμησης θα υπολογισθεί το σφάλμα αυτό, καθώς επίσης παρατίθεται η Εικόνα 4.8 για την καλύτερη κατανόηση της. Στην Εικόνα 4.8 παρατίθενται δύο νευρώνες διαδοχικών επιπέδων, με τον νευρώνα j να ανήκει σε κρυφό επίπεδο ενώ ο νευρώνας k να αποτελεί νευρώνα του επιπέδου εξόδου.



Εικόνα 4.8. Σύνδεση δύο νευρώνων διαφορετικού επιπέδου, με το νευρώνα j να ανήκει σε κρυφό επίπεδο ενώ τον νευρώνα k να ανήκει στο επίπεδο εξόδου (Haykin, 2010)

Η σχέση (4.19) μετασχηματίζεται για ακόμη μια φορά με τον κανόνα της αλυσίδας και η τοπική κλίση υπολογίζεται ως:

$$\delta_j(n) = - \frac{\partial E(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = - \frac{\partial E(n)}{\partial y_j(n)} \varphi'_j(u_j(n)) \quad (4.20)$$

Για τον υπολογισμό της παραμέτρου $\frac{\partial E(n)}{\partial e_j(n)}$ μπορεί να οριστεί η συνάρτηση κόστους του επιπέδου εξόδου με k , όπως φαίνεται και στην Εικόνα 4.8, προκειμένου να μην δημιουργείται σύγχυση με την συνάρτηση κόστους των υπόλοιπων επιπέδων.

$$E(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (4.21)$$

Παραγωγίζοντας την σχέση (4.21), ως προς το $y_j(n)$ προκύπτει

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \quad (4.22)$$

Χρησιμοποιώντας και πάλι τον κανόνα της αλυσίδας η σχέση (4.22) μετασχηματίζεται σε

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial u_k(n)} \frac{\partial u_k(n)}{\partial y_j(n)} \quad (4.23)$$

με τον όρο $\frac{\partial e_k(n)}{\partial u_k(n)}$ να απευθύνεται στο επίπεδο εξόδου του δικτύου. Για το επίπεδο εξόδου είναι γνωστό ότι

$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \varphi'_k(v_k(n))$, οπότε παραγωγίζοντας

$$\frac{\partial e_k(n)}{\partial u_k(n)} = -\varphi'_k(v_k(n)) \quad (4.24)$$

Επιπρόσθετα, από την Εικόνα 4.8

$$v_k(n) = \sum_{j=0}^m w_{kj}(n) y_j(n) \quad (4.25)$$

όπου το m , αντιστοιχεί στο πλήθος των βαρών του νευρώνα εξόδου. Διαφορίζοντας την συνάρτηση (4.25) προκύπτει

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \quad (4.26)$$

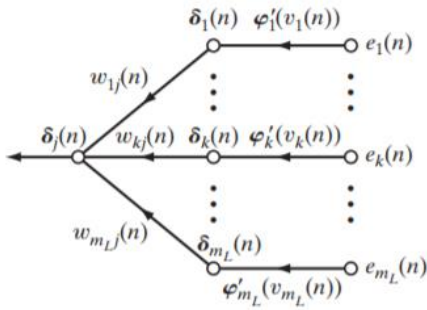
Συνδυάζοντας τις σχέσεις (4.24) και (4.26), η εξίσωση (4.23) παίρνει την μορφή

$$\frac{\partial E(n)}{\partial y_j(n)} = - \sum_k e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) = - \sum_k \delta_k(n) w_{kj}(n) \quad (4.27)$$

Τέλος, η εξίσωση (4.27) αντικαθίσταται στην εξίσωση (4.20), δίνοντας την οπισθοδιάδοση της τοπικής κλίσης $\delta_j(n)$ για κρυφό νευρώνα j

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (4.28)$$

Όπως φαίνεται ο υπολογισμός της κλίσης ενός νευρώνα του κρυφού επιπέδου, εξαρτάται από τρεις όρους. Ο πρώτος είναι η συνάρτηση ενεργοποίησης του νευρώνα και δεν επηρεάζεται από τους γειτονικούς νευρώνες. Οι δύο εναπομείναντες όροι είναι η κλίση που έχει υπολογισθεί για κάθε νευρώνα του πιο δεξιού επιπέδου, που συνδέεται με τον νευρώνα j και το βάρος της αντίστοιχης σύνδεσης. Με τον τρόπο αυτό, επιμερίζεται το τελικό σφάλμα αναλογικά στους νευρώνες των κρυφών. Η διαδικασία αυτή απεικονίζεται στην Εικόνα 4.9.



Εικόνα 4.9. Σχηματική αναπαράσταση των μεταβλητών από τις οποίες επηρεάζεται το σφάλμα σε νευρώνα του κρυφού επιπέδου (Haykin, 2010)

Συνοψίζοντας, η διόρθωση των βαρών κάθε νευρώνα δίνεται από την σχέση:

$$\Delta w_{ij} = \eta \delta_j(n) y_i(n) \quad (4.29)$$

με η να αποτελεί τον ρυθμό μάθησης, $\delta_j(n)$ την τοπική κλίση και $y_i(n)$ το σήμα εισόδου του νευρώνα. Η τιμή της τοπικής κλίσης λαμβάνει διαφορετική μορφή, σύμφωνα με το αν πρόκειται για νευρώνα εξόδου ή νευρώνα κρυφού επιπέδου. Ο υπολογισμός της τοπικής κλίσης δίνεται από την σχέση:

$$\delta_j(n) = \begin{cases} e_j(n) \varphi'_j(u_j(n)), & \text{νευρώνας στο επίπεδο εξόδου} \\ \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n), & \text{νευρώνας στο κρυφό επίπεδο} \end{cases} \quad (4.30)$$

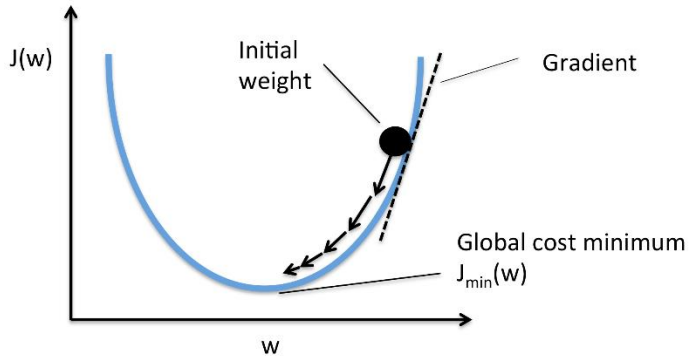
Η σχέση (4.29) που περιγράφει την διόρθωση των βαρών, ονομάζεται Κανόνας Δέλτα.

4.1.3.3 Ρυθμός Μάθησης – Γενικευμένος Κανόνας Δέλτα

Ο κανόνας Δέλτα όπως έχει αναφερθεί στην σχέση (4.12), στην αρχική του μορφή, περιγράφεται από την εξίσωση

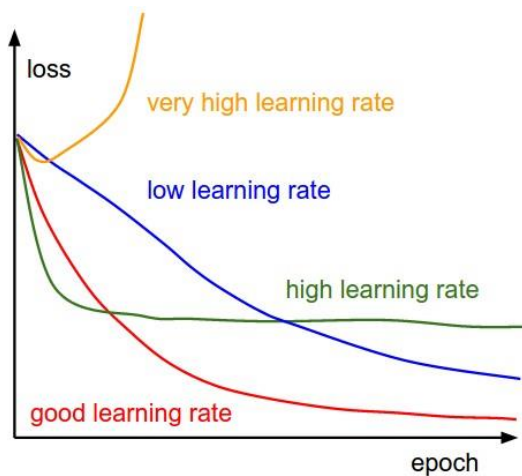
$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)}$$

Άρα ο κανόνας Δέλτα αποσκοπεί στην ελαχιστοποίηση της συνάρτησης $E(n)$, μεταβάλλοντας τις τιμές των βαρών από τις οποίες εξαρτάται, μέσω της τεχνικής της βαθμωτής κατάβασης (gradient descent). Η τεχνική αυτή στηρίζεται στην ιδέα εύρεσης του τοπικού ελαχίστου μια συνάρτησης, κάνοντας «βήματα» ανάλογα του αρνητικού της κλίσης στο εκάστοτε σημείο (Εικόνα 4.10).



Εικόνα 4.10. Βαθμωτή κατάβαση στο χώρο των βαρών (Raschka, Machine Learning FAQ, n.d.)

Το μέγεθος των βημάτων αυτών ονομάζονται ρυθμός μάθησης και απεικονίζεται μέσω της παραμέτρου η . Η χρήση μεγάλης τιμής για τον ρυθμό μάθησης μεταφράζεται σε γρήγορη σύγκλιση του αλγορίθμου, έχοντας την δυνατότητα αποφυγής τοπικών ελαχίστων, με κίνδυνο όμως το δίκτυο να γίνει ασταθές. Αυτό συμβαίνει καθώς μπορεί η τιμή της συνάρτησης σφάλματος να ταλαντώνεται πάνω από το τοπικό ελάχιστο, χωρίς να μπορεί να το προσεγγίσει. Από την άλλη μεριά, επιλέγοντας μικρή τιμή για τον ρυθμό μάθησης, επιτυγχάνεται πιο καλή προσέγγιση του τοπικού ελαχίστου, υστερώντας σημαντικά στον χρόνο σύγκλισης. Οπότε η επιλογή του ρυθμού μάθησης αποτελεί μια ανταλλαγή (trade-off) ανάμεσα στην ακρίβεια και την ταχύτητα σύγκλισης, όπως φαίνεται από την Εικόνα 4.11. Για τον λόγο αυτό, γίνεται αντιληπτό ότι η επιλογή σταθερού ρυθμού μάθησης δεν αποτελεί καλή επιλογή.



Εικόνα 4.11. Επιπτώσεις των διαφορετικών ρυθμών μάθησης στην συνάρτηση σφάλματος (Karpathy, Neural Networks Part 3: Learning and Evaluation, 2017)

Για να μπορέσει να αντιμετωπιστεί η αδυναμία του κανόνα Δέλτα να συνδυάσει ταχύτητα σύγκλισης και ακρίβεια προσέγγισης του τοπικού ελαχίστου, προστίθεται ένας όρος ορμής. Η νέα μορφή του κανόνα Δέλτα δίνεται από την σχέση (4.31) και αποκαλείται Γενικευμένος Κανόνας Δέλτα, με την παράμετρο a να αποτελεί την σταθερά ορμής.

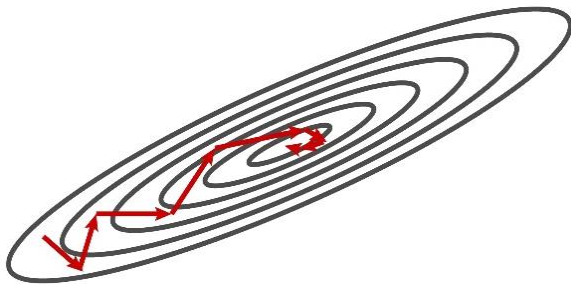
$$\Delta w_{ji}(n) = a\Delta w_{ji}(n-1) - \eta\delta_j(n)y_i(n) \quad (4.31)$$

Η εξίσωση (4.31) μπορεί να αντιμετωπιστεί ως μια εξίσωση διαφορών πρώτης τάξης και να γραφτεί στην μορφή της εξίσωσης (4.32), δεδομένου ότι ο όρος $\delta_j(n)y_i(n)$ ισούται με $-\frac{\partial E(n)}{\partial w_{ji}(n)}$.

$$\Delta w_{ji}(n) = -\eta \sum_{t=0}^n a^{n-t} \frac{\partial E(t)}{\partial w_{ji}(t)} \quad (4.32)$$

Η εξίσωση (4.32) αναπαριστά μια χρονοσειρά, στην μορφή αυτή είναι πιο εμφανής η λειτουργία της παραμέτρου της ορμής. Για να μπορεί να συγκλίνει η σειρά, θα πρέπει η τιμή του a να είναι μεταξύ του 0 και 1, ενώ στην περίπτωση που το a μηδενιστεί, ο γενικευμένος αλγόριθμος Δέλτα αποκτά την αρχική του μορφή, χωρίς τον όρο ορμής.

Αυτό που γίνεται εμφανές από την εξίσωση (4.32) είναι ότι η μεταβολή των βαρών εξαρτάται από τις προηγούμενες τιμές της μερικής παραγώγου $\frac{\partial E(t)}{\partial w_{ji}(t)}$, δηλαδή της κλίσης της συνάρτησης σφάλματος. Συνεπώς, όταν για πολλές διαδοχικές επαναλήψεις η μερική παράγωγος $\frac{\partial E(t)}{\partial w_{ji}(t)}$ έχει σταθερό πρόσημο, η σχέση (4.32) αυξάνεται σε μέτρο, άρα επιταχύνεται η κατάβαση. Αντιθέτως, αν η μερική παράγωγος της συνάρτησης σφάλματος παρουσιάζει διαφορετικό αλγεβρικό πρόσημο σε διαδοχικές επαναλήψεις, τότε το μέτρο του αθροίσματος είναι μικρό.



Εικόνα 4.12. Βαθμωτή κατάβαση με την προσθήκη του όρου της ορμής

Η παραπάνω διαπίστωση απεικονίζεται διαγραμματικά στην Εικόνα 4.12. Αρχικά, όταν η μερική παράγωγος σφάλματος διατηρεί σταθερό πρόσημο, σημαίνει ότι για διαδοχικές επαναλήψεις γίνεται κατάβαση στα τοιχώματα της κοιλάδας της συνάρτησης, οπότε μεγαλώνει ο ρυθμός κατάβασης, μέχρι να

προσεγγισθεί το τοπικό ελάχιστο. Στην περίπτωση αυτή η ορμή, επιταχύνει σημαντικά την κατάβαση, επιτυγχάνοντας βελτιστοποιημένους χρόνους σύγκλισης. Από την άλλη μεριά, όταν η μερική παράγωγος σφάλματος παρουσιάζει διαφορετικό πρόσημο για διαδοχικές επαναλήψεις, σημαίνει ότι υπάρχει ταλάντωση γύρω από κάποιο τοπικό ελάχιστο. Με τον τρόπο αυτό το άθροισμα μειώνεται σημαντικά σαν μέτρο, λειτουργώντας έτσι σαν σταθεροποιητικός παράγοντας προσέγγισης του τοπικού ελαχίστου. Με τον τρόπο αυτό συνδυάζονται τα θετικά του μεγάλου ρυθμού μάθησης για την ταχύτητα σύγκλισης και του μικρού ρυθμού μάθησης για την ακρίβεια σύγκλισης στο τοπικό ελάχιστο.

Στη συγκεκριμένη παράγραφο παρουσιάστηκε εκτενώς η θεμελιώδης μονάδα ενός νευρωνικού δικτύου, η δομή του καθώς και ο τρόπος εκπαίδευσής του. Παρ' όλα αυτά τα παραδοσιακά νευρωνικά δίκτυα δεν ανταποκρίνονται καλά στην ταξινόμηση εικόνων (Karpathy, 2017). Αυτό συμβαίνει, εξαιτίας του πολύ μεγάλου πλήθους παραμέτρων που προσδίδει στο σύστημα η πλήρης συνδεσιμότητα των επιπέδων, όταν εισέρχεται ως είσοδος η τρισδιάστατη αναπαράσταση των εικόνων. Συγκεκριμένα για μια εικόνα μέτριας ανάλυσης 200x200x3, θα χρειαστούν 120.000 βάρη για την σύνδεση της με ένα νευρώνα του κρυφού επιπέδου. Δεδομένου του μεγάλου πλήθους νευρώνων κάθε επιπέδου και του βάθους του νευρωνικού δικτύου, που συνήθως ξεπερνάει το ένα κρυμμένο επίπεδο, το σύστημα των νευρωνικών δικτύων καθίσταται απαγορευτικό για την ταξινόμηση εικόνων. Τα συνελκτικά νευρωνικά δίκτυα, στηριζόμενα στην δομή των perceptrons, συνιστούν μια διαφορετική αρχιτεκτονική για την διαχείριση σύνθετων αναπαραστάσεων όπως οι εικόνες.

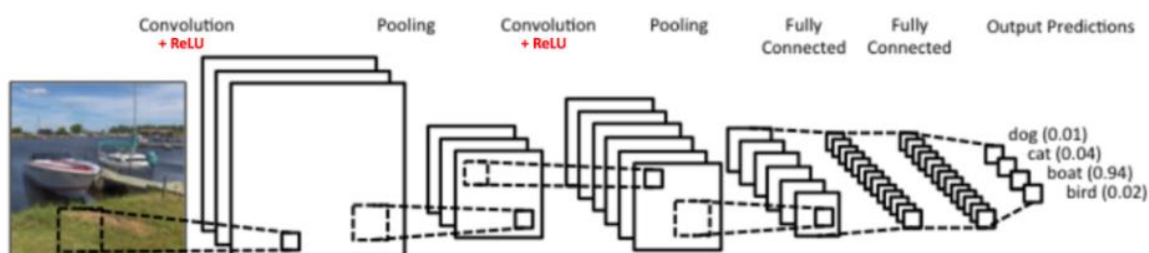
4.2 Συνελκτικά Νευρωνικά Δίκτυα

Το 2012 ήταν η χρονιά κατά την οποία τα Συνελκτικά Νευρωνικά Δίκτυα, άρχισαν να εδραιώνονται στον τομέα της ταξινόμησης των εικόνων. Όταν χρησιμοποιήθηκαν από τον Alex Krizhevsky για την βελτίωση του σφάλματος ταξινόμησης από 26% στο 15% στο διαγωνισμό του ImageNet Large Scale Visual Recognition Competition (ILSVRC), ο οποίος πρόκειται για έναν από τους μεγαλύτερους διαγωνισμούς ταξινόμησης εικόνων.

4.2.1 Αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου

Στην Εικόνα 4.13 παρουσιάζεται η δομή ενός αντιπροσωπευτικού Συνελκτικού Νευρωνικού Δικτύου. Οι βασικές λειτουργίες που επιτελούνται σε αυτό είναι

1. Συνέλιξη (Convolution)
2. Μη γραμμικότητα (Non – Linearity)
3. Δειγματοληψία (Pooling - Subsampling)
4. Ταξινόμηση μέσω του πλήρους συνδεδεμένου επιπέδου (fully connected layer)



Εικόνα 4.13. Αντιπροσωπευτική αρχιτεκτονική Συνελκτικού Νευρωνικού Δικτύου (Britz, 2015)

Όπως φαίνεται στην Εικόνα 4.13 το Συνελκτικό Νευρωνικό Δίκτυο τροφοδοτείται με εικόνες, καθώς μπορεί να διαχειριστεί αποδοτικά τα μεγάλα διανύσματα εισόδου που τις αντιπροσωπεύουν. Στην συνέχεια ακολουθεί μια πληθώρα λειτουργιών οργανωμένες σε επίπεδα. Τα πρώτα στρώματα του ΣΝΔ αφορούν συνελκτικά στρώματα, τα οποία επιλεκτικά συνδυάζονται με κάποια από τα βοηθητικά στρώματα της μη-γραμμικότητας και δειγματοληψίας. Ο συνδυασμός των παραπάνω στρωμάτων αξιοποιείται για εξαγωγή των χαρακτηριστικών της εικόνας εισόδου. Ο συνδυασμός των παραπάνω χαρακτηριστικών για την ταξινόμηση της εικόνας υλοποιείται με την χρήση στρωμάτων πλήρους σύνδεσης. Το τελευταίο στρώμα πλήρους σύνδεσης αποτελείται με τόσους νευρώνες όσες και οι κλάσεις εξόδου και επιστρέφει την πιθανότητα ταξινόμησης της εικόνας σε κάθε μια από τις κλάσεις. Τα βασικά στρώματα του Συνελκτικού Νευρωνικού Δικτύου, φαίνονται στην Εικόνα 4.14. Αξίζει να αναφερθεί ότι πριν το

στρώμα της πλήρης συνδεσιμότητας μπορούν να ακολουθήσουν πολλά συνελικτικά στρώματα, καθώς επίσης ότι συνήθως χρησιμοποιείται παραπάνω από ένα στρώμα πλήρης σύνδεσης στο τέλος του δικτύου.



Εικόνα 4.14. Βασικά στρώματα Συνελικτικού Νευρωνικού Δικτύου

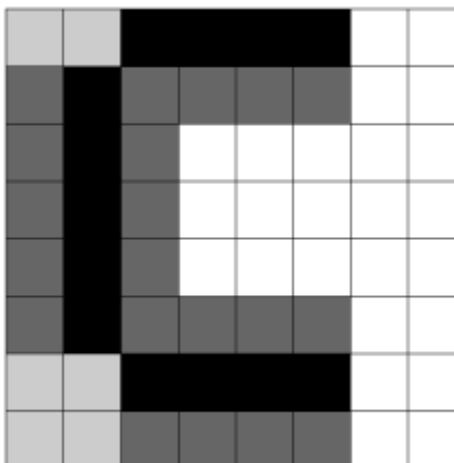
4.2.1.1 Είσοδος του Συνελικτικού Νευρωνικού Δικτύου (ΣΝΔ)

Τα ΣΝΔ δέχονται στην είσοδο τους εικόνες. Στην περίπτωση που η εικόνα είναι στην κλίμακα του γκρι, αναπαρίστανται από έναν διδιάστατο πίνακα, με σύνολο τιμών από 0 έως 255, όπως στην Εικόνα 4.15. Το 0 αντιπροσωπεύει το μαύρο ενώ το 255 αντιπροσωπεύει το άσπρο και όλες οι τιμές στο ενδιάμεσο, αποτελούν αποχρώσεις του γκρι.

Στην περίπτωση που η εικόνα είναι RGB-έγχρωμη χρειάζονται τρία κανάλια για την απεικόνιση της, το κανάλι του κόκκινου, του πράσινου και του μπλε. Στην περίπτωση αυτή η εικόνα αναπαρίστανται ως τρισδιάστατος πίνακας. Ουσιαστικά πρόκειται για 3 διδιάστατους πίνακες, έναν για κάθε κανάλι, με τιμές από 0 έως και 255 σε κάθε θέση τους και ο συνδυασμός τους αποδίδει τελικά την περιγραφή του εικονοστοιχείου.

100	100	0	0	0	0	255	255
50	0	50	50	50	50	255	255
50	0	50	255	255	255	255	255
50	0	50	255	255	255	255	255
50	0	50	255	255	255	255	255
50	0	50	50	50	50	255	255
100	100	0	0	0	0	255	255
100	100	50	50	50	50	255	255

(α)



(β)

Εικόνα 4.15. (α) Πίνακας αναπαράστασης εικόνας στην κλίμακα του γκρι (β) η αντίστοιχη εικόνα

4.2.1.2 Συνελικτικό Στρώμα

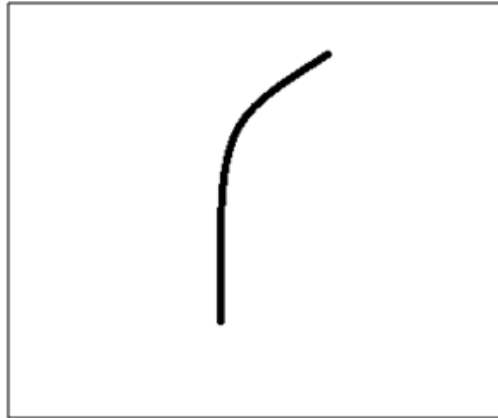
Όπως προκύπτει και από την ονομασία το συνελικτικό στρώμα αποτελεί τον πυρήνα του ΣΝΔ. Οι παράμετροι του επιπέδου αυτού αποτελούνται από φίλτρα που εφαρμόζονται στην εικόνα. Κάθε φίλτρο έχει μικρές διαστάσεις μήκους και πλάτους, συγκριτικά με τις διαστάσεις της εικόνας εισόδου. Για παράδειγμα, τυπικές διαστάσεις του φίλτρου είναι $5 \times 5 \times 3$, με το 3 να είναι το βάθος της εικόνας. Το φίλτρο μετατοπίζεται κατά μήκος και κατά πλάτος της εικόνας, υπολογίζοντας το γινόμενο κάθε θέσης της εικόνας με το φίλτρο. Η μετατόπιση γίνεται πάντα κατά ένα σταθερό αριθμό εικονοστοιχείων και μπορεί να περιγραφεί σαν ένα κυλιόμενο παράθυρό, που όταν φτάνει στο τέλος της κάθε γραμμής, κατεβαίνει πάλι τόσα εικονοστοιχεία κάτω. Το φαινόμενο αυτό ονομάζεται συνέλιξη του φίλτρου με μια εικόνα και παράγει ένα δισδιάστατο χάρτη ενεργοποίησης (activation map). Με τον τρόπο αυτό γίνεται η εξαγωγή των χαρακτηριστικών.

4.2.1.2.1 Παράδειγμα

Με τον όρο εξαγωγή χαρακτηριστικών, αναφέρονται χαρακτηριστικά όπως ακμές, απλά χρώματα και καμπύλες στο πρώτο επίπεδο του συνελικτικού δικτύου, ενώ σε επόμενα στάδια μπορεί να γίνει αναγνώριση πολύ πιο σύνθετων χαρακτηριστικών. Συγκεκριμένα ένα απλό φίλτρο $7 \times 7 \times 3$ αναγνώρισης καμπύλης με κλίση προς τα δεξιά απεικονίζεται στην Εικόνα 4.16(β) ενώ δίπλα φαίνεται ο πίνακας που το αναπαριστά.

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

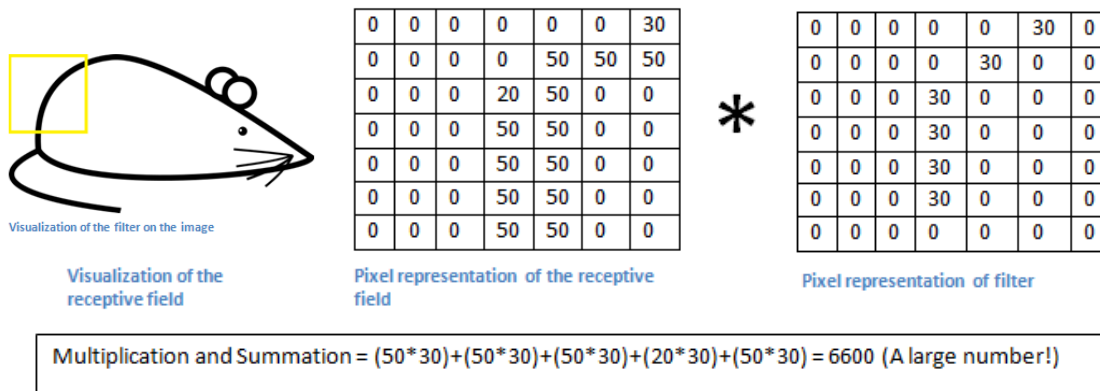
(α)



(β)

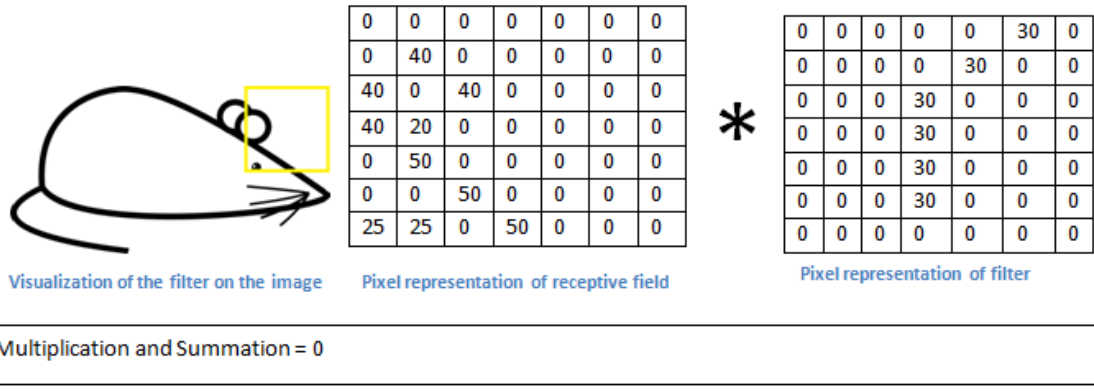
Εικόνα 4.16. (α) Αναπαράσταση φίλτρου σε πίνακα (β) οπτικοποίηση του πίνακα (Desphande, 2016)

Όπως φαίνεται στο παρακάτω απλό παράδειγμα, το φίλτρο αυτό συνελίσσεται με την εικόνα ενός ποντικιού. Στην Εικόνα 4.17 όταν υπολογίζεται το γινόμενο (dot product) ανάμεσα στο πάνω δεξιά 7x7 τμήμα της εικόνας με το φίλτρο, φαίνεται ότι το αποτέλεσμα του εσωτερικού γινομένου έχει μεγάλη απόκριση, γεγονός που σηματοδοτεί ότι εντοπίστηκε κάποιο χωρίο της εικόνας με παρόμοια μορφή με αυτού του φίλτρου.



Εικόνα 4.17. Εσωτερικό γινόμενο φίλτρου που εφαρμόζεται στην εικόνα με χωρίο της εικόνας, το οποίο επιστρέφει μεγάλη απόκριση (Desphande, 2016)

Αντιθέτως στην Εικόνα 4.18 όταν το φίλτρο έχει μετατοπιστεί προς τα δεξιά, το γινόμενο που υπολογίζεται είναι πολύ μικρότερο, το οποίο σημαίνει ότι συναντώνται χωρία που δεν απεικονίζουν καμπύλες μορφολογικά παρόμοιες με αυτές του φίλτρου.



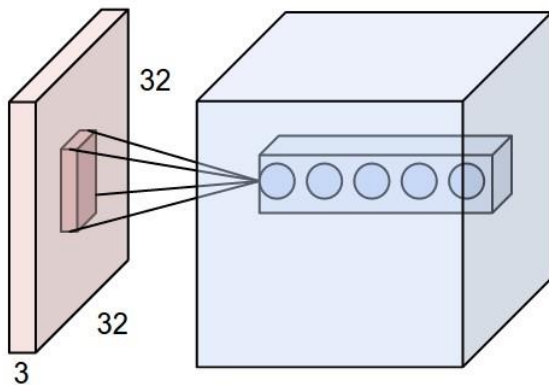
Εικόνα 4.18. Εσωτερικό γινόμενο χωρίου της εικόνας με φίλτρο, το οποίο επιστρέφει μηδενική απόκριση (Desphande, 2016)

4.2.1.2.2 Υπερπαράμετροι

Η χωρική διάταξη που παράγεται μετά την πράξη της συνέλιξης καθορίζεται από 3 υπερπαράμετρους, το βάθος (depth), το βήμα (stride) και το παραγέμισμα (padding).

Βάθος

Το βάθος που παράγεται μετά από κάθε επίπεδο συνέλιξης σχετίζεται με την διάσταση του βάθους της τρισδιάστατης απεικόνισης και είναι ανάλογο του πλήθους των φίλτρων που χρησιμοποιείται στο επίπεδο συνέλιξης. Με την εφαρμογή ενός φίλτρου στην εικόνα παράγεται χάρτης ενεργοποίησης δύο διαστάσεων. Συνεπώς το βάθος του εξαγόμενου όγκου που παράγεται στο στρώμα της συνέλιξης είναι ίσο με το πλήθος των φίλτρων που ενεργοποιούνται. Οι νευρώνες του επιπέδου συνέλιξης που δέχονται ως είσοδο το ίδιο χωρίο υποδοχής (receptive field) ονομάζονται στήλη βάθους (depth column) και σχηματικά απεικονίζονται στην Εικόνα 4.19.



Εικόνα 4.19. Στήλη βάθους συνελκτικού νευρωνικού δικτύου (Karpathy, Convolutional Neural Networks, 2017)

Βήμα (Stride)

Η δεύτερη υπερπαράμετρος (hyperparameter) που πρέπει να οριστεί αφορά τη μετατόπιση του κυλιόμενου παραθύρου, ή απλώς το βήμα (stride). Πρόκειται δηλαδή για το πόσα εικονοστοιχεία μετακινείται κάθε φορά το κυλιόμενο παράθυρο προκειμένου να υπολογιστεί το γινόμενο του φίλτρου με το αντίστοιχο χωρίο υποδοχής. Όσο μεγαλύτερο είναι το βήμα αυτό τόσο μικρότερο είναι το μήκος και το πλάτος του χάρτη ενεργοποίησης που παράγεται.

Παραγέμισμα με μηδέν (Zero-Padding)

Τέλος, υπάρχει η υπερπαράμετρος του μηδενικού παραγεμίματος (zero padding). Μέσω της τεχνικής αυτής προστίθενται μηδενικά γύρω από την εικόνα, πριν την συνέλιξη, με σκοπό να καθοριστούν οι διαστάσεις του εξαγόμενου χάρτη ενεργοποίησης.

4.2.1.2.3 Υπολογισμός διαστάσεων χάρτη ενεργοποίησης

Δεδομένου ότι γνωρίζουμε τις διαστάσεις του όγκου εισόδου (W), τις διαστάσεις του χωρίου υποδοχής (F), το βήμα (S) και το μέγεθος του παραγεμίματος (P), τότε υπολογίζεται κάθε διάσταση του στρώματος συνέλιξης, εκτός του βάρους, από τον τύπο (4.33). Όπως αναφέρθηκε και προηγουμένως, η διάσταση του βάρους εξαρτάται μόνο από το πλήθος των φίλτρων που θα χρησιμοποιηθούν.

$$\frac{W - F + 2P}{S} + 1 \quad (4.33)$$

Για παράδειγμα, στην αρχιτεκτονική που χρησιμοποίησε ο Krizhensky, στο διαγωνισμό του ILSVRC 2012, οι εικόνες εισόδου είχαν διαστάσεις $227 \times 227 \times 3$, άρα το $W=227$. Επιπρόσθετα το πρώτο συνελκτικό δίκτυο έχει χωρίο υποδοχής μεγέθους $11 \times 11 \times 3$ ($F=11$), το βήμα είναι 4 ($S = 4$) και δεν έχει χρησιμοποιηθεί παραγέμισμα ($P=0$). Χρησιμοποιώντας τον τύπο (4.33), προκύπτει ότι $\frac{227-11+0}{4} + 1 = 55$. Δεδομένου ότι ενεργοποιούνται 96 φίλτρα, καθώς και οι διαστάσεις εισόδου της εικόνας κατά μήκος και κατά πλάτος είναι ίσες, προκύπτει ότι στο στρώμα συνέλιξης παράγεται όγκος εξόδου $55 \times 55 \times 96$. Κάθε νευρώνας του $55 \times 55 \times 96$ όγκου, συνδέεται με μια περιοχή $11 \times 11 \times 3$ του όγκου εισόδου, ίση δηλαδή με το χωρίο υποδοχής. Καθώς επίσης κάθε 96 νευρώνες της στήλης βάρους, συνδέονται με την ίδια $11 \times 11 \times 3$ περιοχή εισόδου, με διαφορετικά βάρη που καθορίζονται από το αντίστοιχο φίλτρο κάθε φορά.

4.2.1.2.4 Πλεονεκτήματα Συνελκτικού Στρώματος

Ένας από τους κινδύνους των νευρωνικών δικτύων είναι το πολύ μεγάλο πλήθος των παραμέτρων που διαθέτουν, το οποίο μπορεί να οδηγήσει στο φαινόμενο της υπερπροσαρμογής (overfitting). Παρακάτω παρατίθενται οι τεχνικές της τοπικής συνδεσιμότητας και μοίρασμα παραμέτρων που χρησιμοποιούνται

στα ΣΝΔ, με σκοπό την μείωση των παραμέτρων, καθώς επίσης γίνεται ο ορισμός του φαινομένου υπερπροσαρμογής (overfitting). Αξίζει να αναφερθεί, ότι παρ' όλου που οι παρακάτω τεχνικές μειώνουν σημαντικά το πλήθος των παραμέτρων, το μεγάλο βάθος των ΣΝΔ, δεν επιτρέπει πλήρως την εξάλειψη του φαινομένου.

Φαινόμενο Υπερπροσαρμογής (Overfitting)

Το φαινόμενο της υπερπροσαρμογής σε οποιοδήποτε μοντέλο, όπως και στα νευρωνικά δίκτυα, συμβαίνει όταν μαθαίνεται λεπτομέρεια, θόρυβος καθώς και τυχαία χαρακτηριστικά από το σύνολο εκπαίδευσης, με αποτέλεσμα να επηρεάζει αρνητικά την απόδοση του δικτύου σε νέα δεδομένα. Συνήθως το φαινόμενο αυτό εμφανίζεται σε πολυπαραμετρικά συστήματα, που εξειδικεύονται στην αναγνώριση των χαρακτηριστικών του συνόλου εκπαίδευσης χάνοντας έτσι τη δυνατότητα γενίκευσης στο σύνολο ελέγχου. Η συχνότερη ένδειξη υπερπροσαρμογής εμφανίζεται όταν αρχίζει να αυξάνεται το σφάλμα στο σύνολο ελέγχου μετά από ένα πλήθος επαναλήψεων. Ο επικρατέστερος τρόπος αντιμετώπισης του φαινομένου στον τομέα των νευρωνικών δικτύων, είναι μέσω της χρήσης συνόλου επικύρωσης. Το σύνολο επικύρωσης (validation set) αποτελεί ένα σύνολο δεδομένων που δεν χρησιμοποιείται ούτε για εκπαίδευση, ούτε για έλεγχο, αλλά ως μέσο επικύρωσης ότι μειώνεται το σφάλμα εκπαίδευσης χωρίς να χάνεται η ιδιότητα της γενίκευσης του δικτύου. Συνήθως θεωρείται ότι η προηγούμενη εποχή από την οποία η ακρίβεια ταξινόμησης στο μοντέλο επικύρωσης αρχίζει να μειώνεται, αποτελεί την εποχή μέγιστης γενίκευσης του δικτύου.

Τοπική συνδεσιμότητα (Local Connectivity)

Όταν δίνεται πολυδιάστατο διάνυσμα εισόδου στο δίκτυο, όπως οι εικόνες, η μέθοδος της πλήρους συνδεσιμότητας (fully connectivity) δεν είναι αποδοτική, καθώς δημιουργεί ένα τεράστιο πλήθος παραμέτρων. Για το λόγο αυτό δεν είναι αποτελεσματικό κάθε νευρώνας να συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου. Έτσι στα ΣΝΔ, κάθε νευρώνας συνδέεται με ένα περιορισμένο πλήθος νευρώνων του προηγούμενου επιπέδου που ονομάζεται χωρίο υποδοχής. Με τον τρόπο αυτό επιτυγχάνεται τοπική συνδεσιμότητα (local connectivity) κατά μήκος και πλάτος της εικόνας. Το βάθος του χωρίου υποδοχής είναι ίσο με το βάθος του στρώματος στο οποίο ανήκει.

Μοίρασμα Παραμέτρων

Στο δίκτυο του Krizhensky, όπως αναφέρθηκε παραπάνω, υπάρχουν $55 \times 55 \times 96 = 290.400$ νευρώνες στο πρώτο συνελκτικό στρώμα. Κάθε νευρώνας έχει $11 \times 11 \times 3 = 363$ βάρη συν την πόλωση. Συνεπώς παράγεται ένα σύνολο $290.400 \times 364 = 105.705.600$ παραμέτρων, μόνο για το πρώτο συνελκτικό επίπεδο. Είναι προφανές ότι ένα τόσο μεγάλο πλήθος παραμέτρων δεν είναι διαχειρίσιμο.

Για να μειωθεί σημαντικά το πλήθος των παραμέτρων χρησιμοποιείται η μέθοδος του μοιράσματος παραμέτρων (parameter sharing). Σύμφωνα με την μέθοδο αυτή, όταν είναι χρήσιμο να εξαχθεί ένα χαρακτηριστικό στην θέση (x_1, y_1) θα είναι χρήσιμο να εξαχθεί το ίδιο χαρακτηριστικό και σε οποιαδήποτε άλλη θέση (x_2, y_2) της εικόνας. Έτσι οι νευρώνες που βρίσκονται στο ίδιο βάθος έχουν ίδια βάρη, με την διαφορά ότι τα βάρη εφαρμόζονται σε διαφορετικά χωρία υποδοχής. Με τον τρόπο αυτό, υπάρχει ένα διάγραμμα βαρών $11 \times 11 \times 3$ συν την πόλωση, δηλαδή ένα κοινό φίλτρο για κάθε «φέτα» διαφορετικού βάθους (depth slice). Καθώς συνολικά το βάθος είναι 96, δηλαδή το επίπεδο αυτό αποτελείται από 96 διαφορετικές φέτες, προκύπτει ότι το συνολικό πλήθος των παραμέτρων μετατρέπεται σε $(11 \times 11 \times 3 + 1) \times 96 = 34.944$ παραμέτρους. Παρατηρείται συνεπώς μια εντυπωσιακή πτώση στο πλήθος των παραμέτρων που χρειάζονται.

Έτσι κατά την διαδικασία του backpropagation, κάθε επίπεδο 55×55 νευρώνων του ίδιου βάθους θα υπολογίζει την κλίση των βαρών του. Με την διαφορά όμως ότι οι κλίσεις αυτές θα αθροίζονται για κάθε «φέτα» του ίδιου βάθους, ανανεώνοντας τελικά το κοινό διάγραμμα βαρών που έχουν (Karpathy, Convolutional Neural Networks, 2017).

4.2.1.2.5 Σύνοψη

Συνοψίζοντας, το συνελκτικό στρώμα δέχεται σαν είσοδο ένα τρισδιάστατο πίνακα διαστάσεων $W_1 \times H_1 \times D_1$. Ορίζονται σε αυτό οι υπερπαραμέτροι του πλήθους των φίλτρων K , διαστάσεις εισόδου F_w και F_H , το βήμα S και τέλος το μέγεθος του μηδενικού γεμίματος. Ως έξοδος της διαδικασίας της συνέλιξης, δίνεται τρισδιάστατος πίνακας $W_2 \times H_2 \times D_2$. Με κάθε μια από τις τιμές να προσδιορίζεται ως:

$$W_2 = \frac{W_1 - F_w + 2P}{S} + 1 \quad (4.34)$$

$$H_2 = \frac{H_1 - F_H + 2P}{S} + 1 \quad (4.35)$$

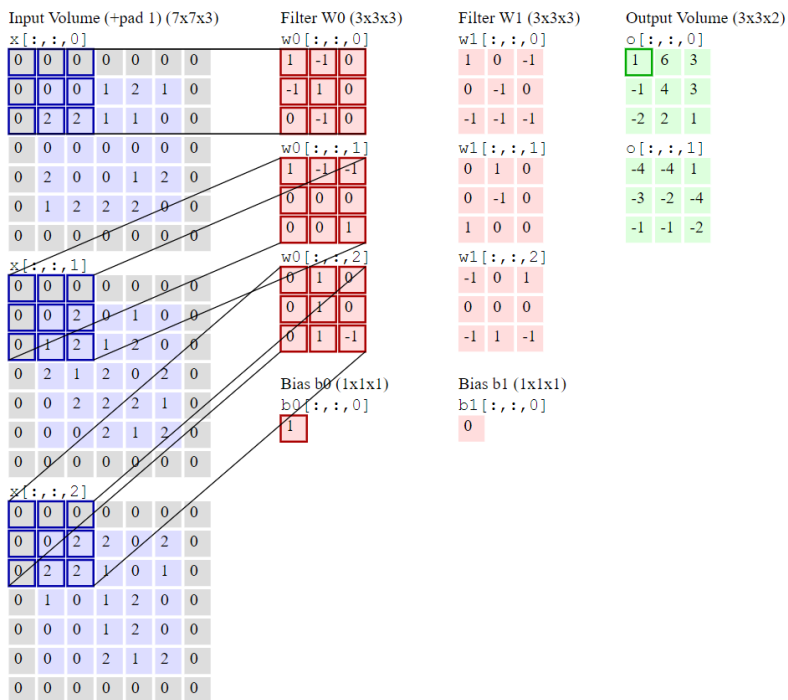
$$D_2 = K \quad (4.36)$$

Συνήθως οι διαστάσεις του ύψους (μήκους) και του πλάτους είναι ίσες, οπότε $F_w = F_H = F$.

Μέσω της τεχνικής μοιράσματος παραμέτρων, κάθε φίλτρο έχει διαστάσεις $F \times F \times D_1$, δηλαδή τόσο είναι και το διάγραμμα βαρών κάθε νευρώνα συν την πόλωση b . Έτσι προκύπτουν $(F \times F \times D_1) \times K$ βάρη και αντίστοιχα K πολώσεις.

Αφού για κάθε χάρτη ενεργοποίησης του ίδιου βάθους, κάθε νευρώνας χρησιμοποιεί τα ίδια βάρη, ουσιαστικά επιτελείται συνέλιξη. Η διαδικασία της συνέλιξης αποτελείται από μια σειρά γινομένων πινάκων. Στην Εικόνα 4.20 αναπαρίσταται αναλυτικά η διαδικασία, με πίνακα εισόδου $5 \times 5 \times 3$ και

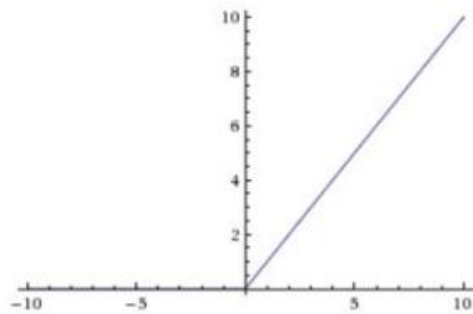
παραμέτρους $K=2$, $F=3$, $S=2$, $P=1$. Με πράσινο φαίνονται τα δύο επίπεδα του τρισδιάστατου πίνακα εξόδου.



Εικόνα 4.20. Αναπαράσταση συνέλιξης σε πίνακα $5 \times 5 \times 3$, με μια σειρά παραγεμίματος με 0, όταν εφαρμόζονται δύο φίλτρα $W0$ και $W1$, διαστάσεων $3 \times 3 \times 3$ σε αυτό. Με πράσινο απεικονίζεται ο χάρτης ενεργοποίησης $3 \times 3 \times 2$ που παράγεται (Karpathy, Convolutional Neural Networks, 2017)

4.2.1.3 Μη Γραμμικότητα - ReLU (Rectified Linear Unit)

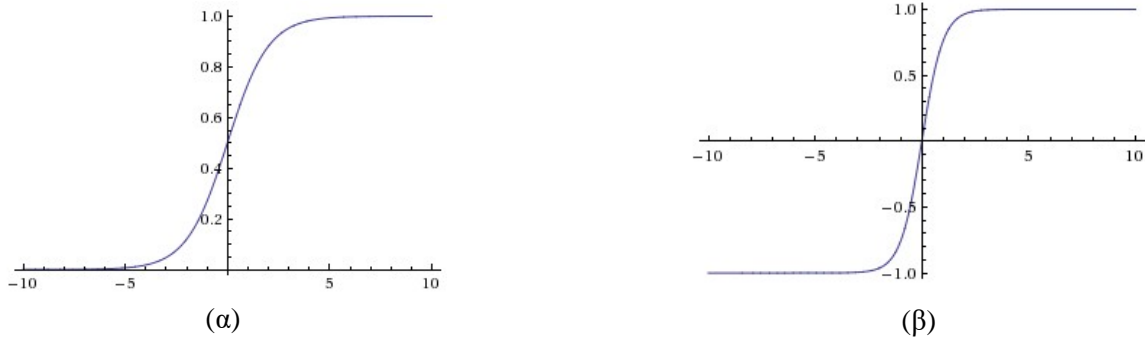
Μετά το στάδιο της συνέλιξης συνήθως ακολουθεί η συνάρτηση ReLU, η οποία εισάγει μη γραμμικότητα στο δίκτυο. Η συνάρτηση αυτή ουσιαστικά κρατάει μόνο τις μη-αρνητικές τιμές του πίνακα που προέκυψαν από την συνέλιξη, δηλαδή συγκρατεί το μέγιστο ανάμεσα σε κάθε στοιχείο του πίνακα και του 0. Η γραφική της παράσταση φαίνεται στην Εικόνα 4.21.



Εικόνα 4.21. Γραφική παράσταση της συνάρτησης ReLU, με τύπο $Output = Max(zero, Input)$

Ο λόγος για τον οποίο επιθυμείται η μη γραμμικότητα είναι επειδή τα πιο πολλά αντικείμενα του πραγματικού κόσμου δεν είναι γραμμικά. Ο πολλαπλασιασμός των βαρών που επιτελείται από στην είσοδο των νευρώνων είναι μια γραμμική μετατροπή, ακόμα και με την προσθήκη της πόλωσης. Έτσι πολλές τέτοιες μετατροπές που επιτελούνται από τα διάφορα στάδια των νευρωνικών δικτύων, θα μπορούσαν να αναπαρασταθούν από μια γραμμική σχέση ανάμεσα στην είσοδο και την έξοδο του δικτύου, στην περίπτωση που δεν χρησιμοποιούταν μη γραμμικές συναρτήσεις. Για την υλοποίηση μη γραμμικών συσχετίσεων ανάμεσα στην είσοδο και έξοδο, δηλαδή την εκμάθηση πιο σύνθετων μοτίβων από το δίκτυο, χρησιμοποιούνται οι μη γραμμικές συναρτήσεις ενεργοποίησης.

Οι συνήθεις συναρτήσεις για την προσθήκη μη γραμμικότητας είναι η λογιστική συνάρτηση $f(x) = \frac{1}{1+e^{-x}}$, η υπερβολική εφαπτομένη $f(x) = \tanh(x)$ και η ReLU, με τις δύο πρώτες να απεικονίζονται στην Εικόνα 4.22. Λόγω της πολύ μεγαλύτερης ταχύτητας σύγκλισης που προσφέρει η συνάρτηση ReLU, είναι αυτή που έχει εδραιωθεί στα Συνελικτικά Νευρωνικά Δίκτυα. Αυτό συμβαίνει επειδή, μέσω της μεθόδου της βαθμωτής κατάβασης (gradient descent) που χρησιμοποιείται για την ανανέωση των βαρών, παρατηρήθηκε ότι η συνάρτηση ReLU, λόγω της μη κορεσμένη μη-γραμμικότητας που παρουσιάζει, έχει πολύ μεγαλύτερη ταχύτητα εκπαίδευσης από τις συναρτήσεις που παρουσιάζουν κορεσμό (Krizhevsky, Sutskever, & Hinton, 2012).



Εικόνα 4.22. (α) Λογιστική συνάρτηση ενεργοποίησης (β) Συνάρτηση Υπερβολικής εφαπτομένης (Karpathy, 2017)

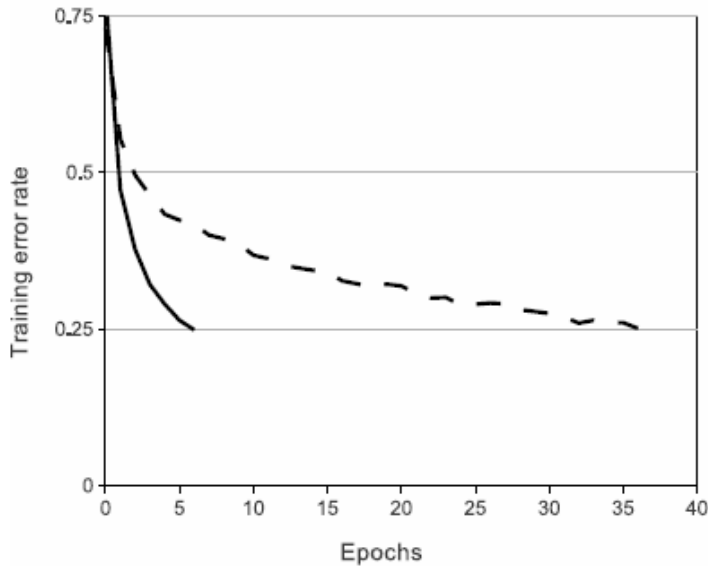
Συγκεκριμένα οι μη-κορεσμένες συναρτήσεις είναι αυτές που όταν το πεδίο τιμών απειρίζεται, τότε απειρίζεται σε τουλάχιστον μια περίπτωση και το σύνολο τιμών.

$$\left(\lim_{x \rightarrow -\infty} f(x) = -\infty \right) \vee \left(\lim_{x \rightarrow \infty} f(x) = +\infty \right) \quad (4.37)$$

Σε περίπτωση που δεν ικανοποιείται η παραπάνω συνθήκη, η συνάρτηση ονομάζεται κορεσμένη, δηλαδή το σύνολο τιμών συγκλίνει σε κάποια τιμή καθώς το πεδίο ορισμού απειρίζεται.

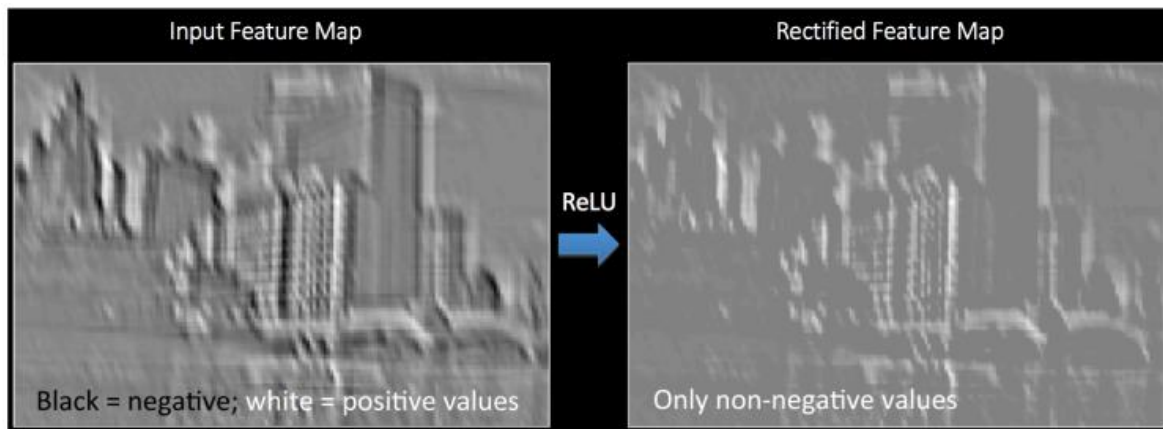
Στα νευρωνικά δίκτυα, μέσω της βαθμωτής κατάβασης, η ανανέωση των βαρών είναι ανάλογη της κλίσης της συνάρτησης ενεργοποίησης. Παρατηρείται λοιπόν ότι για μεγάλες τιμές του πεδίου ορισμού, όταν δηλαδή η τιμή επιστροφής της συνάρτησης είναι κοντά στο 0 ή στο 1, για την λογιστική συνάρτησης και κοντά στο -1 και στο 1, για την υπερβολική εφαπτομένη, τότε η κλίση της συνάρτησης τείνει στο 0. Συνεπώς αφού η κλίση της συνάρτησης είναι πολύ μικρή τα βάρη ουσιαστικά θα σταματήσουν να ανανεώνονται. Το φαινόμενο αυτό ονομάζεται κορεσμός και δημιουργεί σημαντικά προβλήματα στην περίπτωση που εμφανίζεται πολύ νωρίς στην διαδικασία εκπαίδευσης του δικτύου.

Η σημαντικά βελτιωμένη ταχύτητα εκπαίδευσης των ΣΝΔ, χρησιμοποιώντας της συνάρτηση ReLU, φαίνεται στο πλήθος των εποχών που χρειάστηκε το CIFAR-10 dataset ώσπου να φτάσει στο 25% το σφάλμα εκπαίδευσης (Krizhevsky, Sutskever, & Hinton, 2012). Συγκεκριμένα έγινε σύγκριση ανάμεσα σε συνελκτικό νευρωνικό δίκτυο 4 επιπέδων, όπου τη μια φορά χρησιμοποιεί ως συνάρτηση ενεργοποίησης τη ReLU ενώ τη δεύτερη φορά, αυτή της υπερβολικής εφαπτομένης. Όπως φαίνεται στην Εικόνα 4.23 το σφάλμα εκπαίδευσης μειώνεται με σημαντικά μεγαλύτερο ρυθμό στην περίπτωση της συνάρτησης ReLU, η οποία απεικονίζεται με συνεχόμενη γραμμή. Αυτό αποτελεί το λόγο καθιέρωσής της ως τη βασική συνάρτηση ενεργοποίησης στην περίπτωση των συνελκτικών νευρωνικών δικτύων.



Εικόνα 4.23. Βελτιωμένος ρυθμός μάθησης που επιτυγχάνεται μέσω της χρήσης της συνάρτησης ενεργοποίησης ReLU (συνεχόμενη γραμμή), συγκριτικά με την συνάρτηση της υπερβολικής εφαπτομένης (διακεκομμένη γραμμή). Πεδίο εφαρμογής των παραπάνω συναρτήσεων είναι συνεκτικό νευρωνικό δίκτυο 4 επιπέδων που εκπαιδεύεται στο σύνολο δεδομένων CIFAR-10 (Krizhevsky, Sutskever, & Hinton, 2012).

Παρακάτω απεικονίζεται η μεταβολή που σημειώνεται σε μια εικόνα εφαρμόζοντας τον τελεστή του ReLU, δεδομένου ότι με μαύρο απεικονίζονται οι αρνητικές τιμές των εικονοστοιχείων, ενώ με άσπρο οι θετικές.



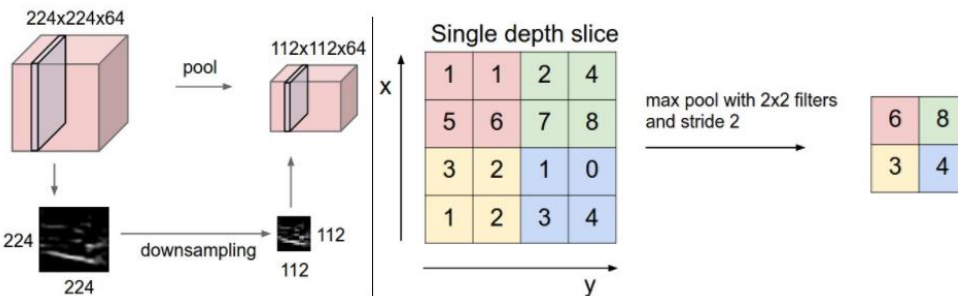
Εικόνα 4.24. Εφαρμογή συνάρτησης ReLU σε εικόνα (Fergus, 2015)

4.2.1.4 Στρώμα Δειγματοληψίας

Μεταξύ των στρωμάτων συνέλιξης, συνήθως προστίθεται ένα επίπεδο δειγματοληψίας (pooling layer). Η λειτουργία του είναι σταδιακά να μειώσει το χωρικό μέγεθος της εικόνας (ύψος και πλάτος), ενώ παράλληλα διατηρείται η πιο σημαντική πληροφορία (Εικόνα 4.25). Με την τρόπο αυτό η σημαντική

πληροφορία δεν χάνεται, μειώνεται η πολυπλοκότητα καθώς επίσης επιτυγχάνεται καλύτερη γενίκευση, αποφεύγοντας το φαινόμενο της υπερπροσαρμογής. Το στρώμα δειγματοληψίας λειτουργεί ξεχωριστά για κάθε επίπεδο του τρισδιάστατου πίνακα, χρησιμοποιώντας διαφορετικές συναρτήσεις όπως το Max, Sum και Average.

Ο τελεστής που χρησιμοποιείται πιο συχνά είναι ο MAX όπως φαίνεται στην Εικόνα 4.25β, κρατώντας δηλαδή τα χαρακτηριστικά με την μεγαλύτερη απόκριση από κάθε χάρτη ενεργοποίησης. Συνήθως χρησιμοποιείται σε φίλτρα διαστάσεων 2x2, με βήμα 2, μειώνοντας έτσι στο μισό τις διαστάσεις του πλάτους και του μήκους. Με τον τρόπο αυτό, αφαιρείται το 75% της πληροφορίας κάθε χάρτη ενεργοποίησης, ενώ η διάσταση του βάθους παραμένει ανέπαφη. Άλλες συναρτήσεις, οι οποίες μπορούν να χρησιμοποιηθούν είναι οι SUM και AVERAGE, οι οποίες αντιστοίχως υπολογίζουν το άθροισμα και την μέση τιμή στο φίλτρο διαστάσεων που θα ορισθεί. Στις δύο αυτές περιπτώσεις επιτυγχάνεται και πάλι μείωση των διαστάσεων, διατηρώντας όμως διαφορετική πληροφορία.



Εικόνα 4.25. (α) η μείωση των διαστάσεων που επιτυγχάνεται μέσω του Στρώματος Δειγματοληψίας (β) Αναπαράσταση του max-pooling χρησιμοποιώντας φίλτρο 2x2 και βήμα 2 (Karpathy, Convolutional Neural Networks, 2017).

4.2.1.5 Στρώμα Πλήρης Συνδεσιμότητας (Fully Connected Layer)

Το fully connected layer, ή σε ελληνική μετάφραση το στρώμα πλήρης συνδεσιμότητας δέχεται σαν είσοδο ένα τρισδιάστατο πίνακα, που προέκυψε μετά από τα επίπεδα συνέλιξης, μη γραμμικότητας και δειγματοληψίας. Επίσης επιστρέφει ένα N-διάστατο διάνυσμα, όπου N είναι το πλήθος των κλάσεων στις οποίες θα γίνει η ταξινόμηση. Ο όρος «πλήρης συνδεσιμότητα» υποδηλώνει ότι κάθε νευρώνας αυτού του επιπέδου συνδέεται με όλους τους νευρώνες του προηγούμενου.

Αναλυτικά, το στρώμα πλήρης συνδεσιμότητας επεξεργάζεται την έξοδο του προηγούμενου επιπέδου, το οποίο σύμφωνα με τα όσα αναφέρθηκαν παραπάνω περιέχει τους χάρτες ενεργοποίησης των υψηλού επιπέδου χαρακτηριστικών. Έτσι αφού οι νευρώνες του επιπέδου πλήρης συνδεσιμότητας συνδέονται με όλους τους νευρώνες του προηγούμενου επιπέδου, ουσιαστικά προσδιορίζουν την εξάρτηση της κάθε κλάσης του επιπέδου εξόδου, από τα χαρακτηριστικά υψηλού επιπέδου που έχουν εξαχθεί από το

τελευταίο στρώμα συνέλιξης. Για παράδειγμα, αν μια από τις κλάσεις ήταν ένα πτηνό τότε θα υπήρχαν μεγάλη βάρη στις συνδέσεις των χαρτών ενεργοποίησης που περιγράφουν το ράμφος και τα φτερά, ενώ για την κλάση σκύλος θα δινόταν μεγαλύτερη έμφαση στα τέσσερα πόδια και τα κρεμαστά αυτιά.

Το στρώμα πλήρης σύνδεσης δίνει ως έξοδο ένα διάνυσμα διαστάσεων N , όσες και οι κλάσεις ταξινόμησης. Στην συνέχεια στο N -διάστατο διάνυσμα εφαρμόζεται η συνάρτηση κανονικοποίησης softmax. Η συνάρτηση softmax, αποτελεί γενίκευση της σιγμοειδής συνάρτησης η οποία αποτελεί την συχνότερη επιλογή συνάρτησης ενεργοποίησης στα νευρωνικά δίκτυα, με αποτέλεσμα να παρέχει τα πλεονεκτήματα της εύκολης διαφορισιμότητας και της φραγμένης εξόδου στο διάστημα $[0,1]$. Η συνάρτηση softmax εμπεριέχει ακόμα την ιδιότητα άθροισης των αποτελεσμάτων της στην μονάδα, για τον λόγο αυτό αποτελεί ενδεικτική επιλογή για την αναπαράσταση πιθανοτήτων. Τελικώς το αποτέλεσμα που εξάγεται από την συνάρτηση softmax, άρα και ολόκληρου του στρώματος πλήρης συνδεσιμότητας, περιγράφει την πιθανότητα ένταξης της εικόνας σε κάθε μια από τις κλάσεις.

4.2.2 Λειτουργίες Συνελικτικών Νευρωνικών Δικτύων

Τα ΣΝΔ παρουσιάζουν διττό ρόλο, καθώς μπορούν να χρησιμοποιηθούν τόσο για την εξαγωγή χαρακτηριστικών όσο και για την ταξινόμηση εικόνων. Παρακάτω, αναλύονται αυτές οι δύο λειτουργίες τους, αφού πρώτα αναλυθεί η τεχνική της προ-εκπαίδευσης, που αποσκοπεί στην καλύτερη αρχικοποίηση του δικτύου.

4.2.2.1 *Προ-εκπαίδευση (pre-Training):*

Τα ΣΝΔ για να παρουσιάσουν ικανοποιητικά αποτελέσματα, χρειάζονται ένα πολύ μεγάλο σύνολο δεδομένων για εκπαίδευση, ενώ τα δεδομένα αυτά θα πρέπει να είναι προσεχτικά επιλεγμένα, προκειμένου να αποφευχθεί το φαινόμενο του overfitting. Για το λόγο αυτό είναι πολύ σπάνιο ένα ΣΝΔ να εκπαιδευθεί απευθείας στο σύνολο εκπαίδευσης χωρίς πρώτα να έχει προηγηθεί η διαδικασία της προ-εκπαίδευσης.

Η διαδικασία της προ-εκπαίδευσης, αποτελεί την εκπαίδευση ενός ΣΝΔ από ένα πολυπληθές σύνολο δεδομένων. Τα σύνολα δεδομένων που επιλέγονται για τον σκοπό αυτό, από την μια μεριά θα πρέπει να αποτελούνται από μεγάλο αριθμό εικόνων για την σωστή εκπαίδευση του ΣΝΔ και από την άλλη να αποτελούνται από πολλές κατηγορίες, προκειμένου να επιτυγχάνεται καλύτερη γενίκευση εξαγωγής χαρακτηριστικών. Η διαδικασία της προ-εκπαίδευσης, λόγω του μεγάλου πλήθους εικόνων στο οποίο στηρίζεται, αποτελεί μια ιδιαίτερα χρονικά δαπανηρή διαδικασία. Η πιο γνωστή και διαδεδομένη βάση δεδομένων για την προ-εκπαίδευση των δικτύων είναι το ImageNet (Russakovsky, et al., 2015). Με πλούσια σύνολα δεδομένων, όπως το ImageNet, γίνεται εκπαίδευση ΣΝΔ, όχι για κάποιο συγκεκριμένο σκοπό, αλλά προκειμένου να επιτευχθεί μια πολύ καλή αρχικοποίηση των βαρών (Ao & Ling, 2015). Με

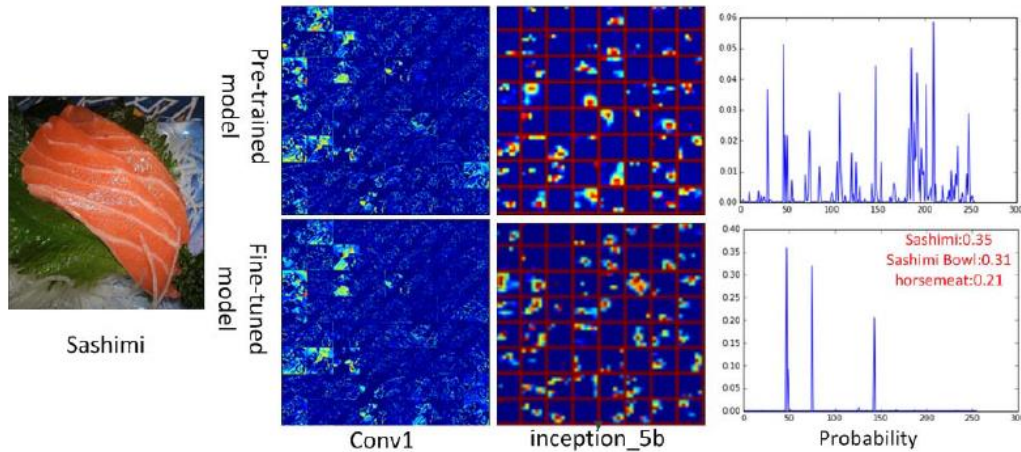
τον τρόπο αυτό ορίζονται «γενικευμένα» φίλτρα εξαγωγής χαρακτηριστικών και παράγεται ένα Προ-Εκπαιδευμένο ΣΝΔ.

4.2.2.2 *Εξαγωγή Χαρακτηριστικών (Feature-Extraction):*

Τα ΣΝΔ, αποτελούνται ουσιαστικά από δύο μέρη, από ένα μεγάλο σύνολο στρωμάτων συνέλιξης και καταλήγουν σε ένα περιορισμένο πλήθος στρωμάτων πλήρης σύνδεσης. Όπως έχει αναλυθεί και προηγουμένως, τα πρώτα λειτουργούν για την εξαγωγή χαρακτηριστικών, ενώ τα δεύτερα συνδυάζουν τα χαρακτηριστικά αυτά, προκειμένου να προσδιορίσουν με τον καλύτερο τρόπο τις κλάσεις του επιπέδου εξόδου. Σύμφωνα με την μέθοδο αυτή, χρησιμοποιούνται μόνο τα συνελκτικά επίπεδα ενός προ-εκπαιδευμένου δικτύου, δηλαδή τα επίπεδα μέχρι το πρώτα στρώμα πλήρης σύνδεσης, προκειμένου να εξαχθεί ένα διάνυσμα περιγραφής της εικόνας. Το διάνυσμα που προκύπτει στην συνέχεια μπορεί να τροφοδοτήσει κάποιο ταξινομητή, έτσι ώστε να υλοποιηθεί η διαδικασία της ταξινόμησης. Συνεπώς η τεχνική αυτή αποτελεί ένα γρήγορο τρόπο για να χρησιμοποιηθεί η αποτελεσματικότητα των ΣΝΔ, χωρίς να χρειαστεί να υλοποιηθεί η χρονοβόρα διαδικασία της εκπαίδευσης. Η εξαγωγή χαρακτηριστικών μέσω προ-εκπαιδευμένων ΣΝΔ έχει παρατηρηθεί ότι παρουσιάζει καλύτερα αποτελέσματα από τις συμβατικές τεχνικές εξαγωγής χαρακτηριστικών, όπως οι Histogram Oriented Gradient (HOG), Local Binary Pattern (LBP) και ιστογράμματα χρώματος (Color Histogram) (Christodoulidis, Anthimopoulos, & Mougiakakou, 2015; Kawano & Yanai, 2015).

4.2.2.3 *Fine-Tuning*

Η μέθοδος του fine-tuning επιτελείται σε ένα υπάρχον προ-εκπαιδευμένο ΣΝΔ, με σκοπό την επανεκπαίδευση ενός μέρους ή συνολικά ολόκληρου του δικτύου. Ουσιαστικά με το προ-εκπαιδευμένο δίκτυο γίνεται μια καλή αρχικοποίηση των βαρών του ΣΝΔ. Από εκεί και πέρα, μέσω του fine-tuning γίνεται προσαρμογή των βαρών του δικτύου στο πρόβλημα ταξινόμησης το οποίο αντιμετωπίζεται. Υπάρχουν φορές, που τα πρώτα συνελκτικά επίπεδα παραμένουν ίδια, για τον εντοπισμό των χαρακτηριστικών χαμηλού επιπέδου, ενώ τροποποιούνται τα βάρη των πιο ψηλών επιπέδων. Ενώ άλλες φορές επανεκπαιδεύεται ολόκληρο το δίκτυο. Ανεξαρτήτως την περίπτωση, κατά την διαδικασία του fine-tuning, μεταβάλλεται το επίπεδο πλήρης συνδεσιμότητας, έτσι ώστε στην έξοδο του να υπάρχουν τόσοι νευρώνες όσες και οι κλάσεις του συνόλου δεδομένων. Οπότε η επανεκπαίδευση των βαρών του fully-connected layer είναι απαραίτητη. Η μέθοδος fine-tuning αποτελεί την επικρατούσα τεχνική για την χρήση των ΣΝΔ, καθώς γίνεται αυστηρή προσαρμογή των βαρών στο σκοπό της ταξινόμησης επιτυγχάνοντας με αυτό τον τρόπο τα καλύτερα αποτελέσματα.



Εικόνα 4.26. Διαφορετικοί χάρτες ενεργοποίησης και διαφορετικά αποτελέσματα ταξινόμησης που επιτυγχάνονται στην περίπτωση του pre-trained και του fine-tuned μοντέλου. (Ao & Ling, 2015)

4.2.3 Γνωστά Συνελκτικά Νευρωνικά Δίκτυα

Έχει γίνει σημαντική μελέτη για την εύρεση αρχιτεκτονικών ΣΝΔ, τα οποία αποδίδουν τα καλύτερα αποτελέσματα. Δύο από τα πιο γνωστά είναι αυτά του AlexNet και VGGnet, τα οποία διακρίθηκαν στην ταξινόμηση του συνόλου δεδομένων ImageNet (Schlag, 2016; Deshpande, 2016).

Το ImageNet είναι σύνολο δεδομένων αποτελούμενο από 15 εκατομμύρια υψηλής ανάλυσης εικόνες, που ανήκουν σε 22.000 περίπου κατηγορίες. Η συλλογή των εικόνων έχει γίνει από το διαδίκτυο και η κατηγοριοποίηση τους έχει γίνει από ανθρώπους μέσω του Amazon's Mechanical Turk crowd-sourcing tool. Από το 2010, βασιζόμενο στις εικόνες του ImageNet, ξεκίνησε ο ετήσιος διαγωνισμός ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). Στον διαγωνισμό αυτό χρησιμοποιείται ένα υποσύνολο των εικόνων του ImageNet, που αποτελείται από 1000 εικόνες από 1000 διαφορετικές κατηγορίες περίπου. Συνολικά υπάρχουν περίπου 1,2 εκατομμύρια εικόνων στο σύνολο εκπαίδευσης, 50.000 εικόνες ως σύνολο δεδομένων επικύρωσης (validation dataset), ενώ το σύνολο ελέγχου (test set), αποτελείται από 150,000 εικόνες.

4.2.3.1 AlexNet

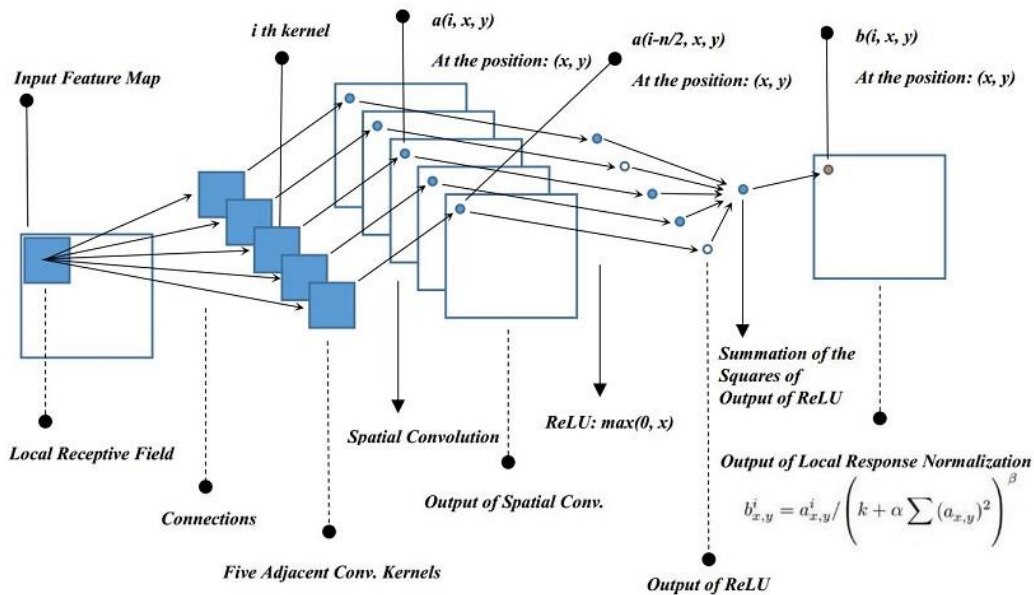
Το AlexNet αποτελεί ένα ΣΝΔ αποτελούμενο από 8 στρώματα, εκ των οποίων τα 5 είναι συνελκτικά και τα εναπομείναντα 3 είναι στρώματα πλήρης σύνδεσης (Krizhevsky, Sutskever, & Hinton, 2012). Η συνάρτηση ενεργοποίησης που χρησιμοποιείται από τους νευρώνες του δικτύου είναι η ReLU (Rectified Linear Units). Αξίζει να σημειωθεί, ότι η συνάρτηση ReLU, ως μη γραμμική συνάρτηση των νευρώνων προτάθηκε πρώτη φορά κατά την δημιουργία του ΣΝΔ AlexNet.

Κανονικοποίηση

Όταν χρησιμοποιείται η συνάρτηση ενεργοποίησης ReLU, δεν είναι απαραίτητη η κανονικοποίηση προκειμένου να αποφευχθεί ο κορεσμός. Παρ' όλα αυτά φαίνεται ότι η κανονικοποίηση του Local Response Normalization, βοηθάει στην καλύτερη γενίκευση του δικτύου. Συμβολίζοντας με $a_{x,y}^i$ το αποτέλεσμα που παράγεται εφαρμόζοντας φίλτρο i στην θέση (x,y) του χάρτη ενεργοποίησης(activation map) και στην συνέχεια εφαρμόζοντας τη συνάρτηση ReLU, το αποτέλεσμα της κανονικοποίησης είναι το $b_{x,y}^i$ και αναπαρίσταται μέσω της εξίσωσης (4.38).

$$b_{x,y}^i = a_{x,y}^i / \left(k + a \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} (a_{x,y}^j)^2 \right)^\beta \quad (4.38)$$

Με την μεταβλητή n απεικονίζονται τα γειτονικά φίλτρα που ορίζονται στην θέση (x,y) , από το συνολικό πλήθος N φίλτρων. Ο χρήστης ορίζει τα φίλτρα τα οποία θα είναι γειτονικά μεταξύ τους. Οι μεταβλητές k , n , a , b αποτελούν υπερπαραμέτρους και στην περίπτωση του AlexNet λαμβάνουν τις τιμές 2, 5, 10^{-4} και 0,75 αντιστοίχως. Διαγραμματικά η διαδικασία που ακολουθείται φαίνεται στην Εικόνα 4.27.



Εικόνα 4.27. Αναπαράσταση της διαδικασίας του Local Response Normalization (Normalization in Neural Networks, 2016)

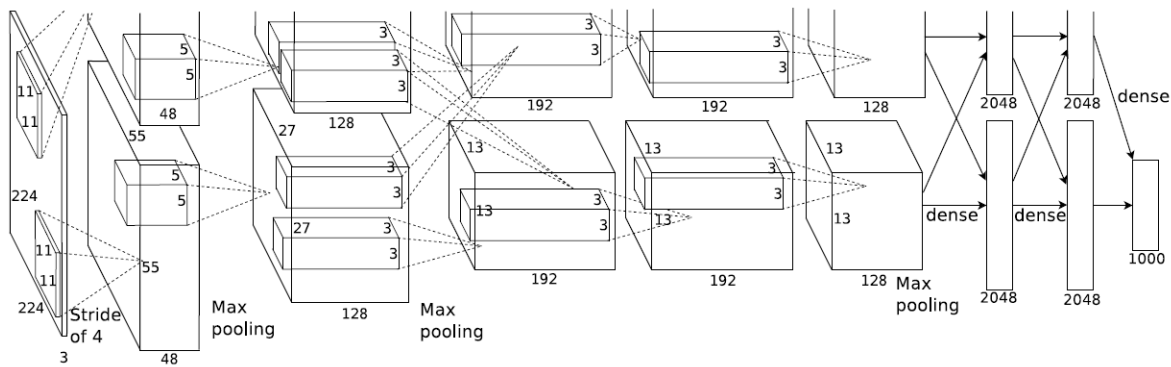
Μέσω της κανονικοποίησης επιχειρείται να υλοποιηθεί το φαινόμενο της παράπλευρης καταστολής (lateral inhibition) ανάμεσα στα perceptrons, φαινόμενο που προσομοιώνει την λειτουργία των βιολογικών νευρώνων. Η παράπλευρη καταστολή αποτελεί μια μορφή ανταγωνισμού μεταξύ των νευρώνων. Συγκεκριμένα αναφέρεται στην ικανότητα των διεγερμένων νευρώνων να παρεμποδίζουν την δραστηριότητα των γειτονικών τους. Συνεπώς μόνο οι νευρώνες που είναι περισσότερο διεγερμένοι και λιγότερο παρεμποδισμένοι αποκρίνονται, δημιουργώντας έτσι καλύτερη αντίθεση άρα και ανάλυση (Kreutzer, DeLuca, & Caplan, 2010).

Στρώμα Δειγματοληψίας (Pooling Layer)

Το ΣΝΔ AlexNet, χρησιμοποιεί επικαλυπτόμενη δειγματοληψία. Καθώς η περιοχή στην οποία γίνεται το pooling έχει μικρότερη διάσταση από το βήμα. Συγκεκριμένα η γειτονιά στο οποίο γίνεται το pooling, έχει διαστάσεις 3x3, ενώ το βήμα είναι 2. Παρατηρήθηκε ότι το ποσοστό σφάλματος μειώνεται κατά 0,3% χρησιμοποιώντας επικαλυπτόμενη δειγματοληψία, συγκριτικά με την επικαλυπτόμενη.

Αρχιτεκτονική

Η αρχιτεκτονική του AlexNet απεικονίζεται στην Εικόνα 4.28. Όπως φαίνεται, αποτελείται από 8 στρώματα, με τα πρώτα 5 να είναι συνελκτικά και τα 3 τελευταία να είναι fully connected. Η έξοδος του τελευταίου fully-connected layer επεξεργάζεται από την συνάρτηση softmax, επιστρέφοντας την κατανομή στις 1000 κλάσεις εξόδου του ILSVRC.



Εικόνα 4.28. Αρχιτεκτονική του Συνελκτικού Νευρωνικού Δικτύου AlexNet (Krizhevsky, Sutskever, & Hinton, 2012).

Συγκεκριμένα, μετά το πρώτο και το δεύτερο συνελκτικό στρώμα ακολουθεί Local Response κανονικοποίηση καθώς και επίπεδο δειγματοληψίας, το οποίο χρησιμοποιεί την συνάρτηση Max (Max-pooling layer). Το στρώμα του max pooling εφαρμόζεται και μετά από το πέμπτο επίπεδο συνέλιξης. Επίσης το στρώμα ReLU εφαρμόζεται μετά από κάθε νευρώνα τόσο των συνελκτικών επιπέδων όσο και των επιπέδων πλήρης συνδεσιμότητας.

Ως είσοδος του δικτύου δίνονται εικόνες διαστάσεων 224x224x3. Στο πρώτο συνελκτικό στρώμα εφαρμόζονται 96 φίλτρα σε αυτές, διαστάσεων 11x11x3 με βήμα 4. Το δεύτερο συνελκτικό στρώμα δέχεται σαν είσοδο το αποτέλεσμα που παράγεται μετά την κανονικοποίηση και χρησιμοποιεί 256 φίλτρα διαστάσεων 5x5x48. Το τρίτο και τέταρτο συνελκτικό στρώμα συνδέονται χωρίς να παρεμβάλλεται κάτι μεταξύ τους. Στο τρίτο συνελκτικό στρώμα εφαρμόζονται 384 φίλτρα διαστάσεων 3x3x256 ενώ στο

τέταρτο 384 φίλτρα διαστάσεων 3x3x192. Τέλος το πέμπτο συνελκτικό στρώμα διαθέτει 256 φίλτρα διαστάσεων 3x3x192, ενώ τα fully connected layers αποτελούνται από 4096 νευρώνες.

Overfitting στο AlexNet

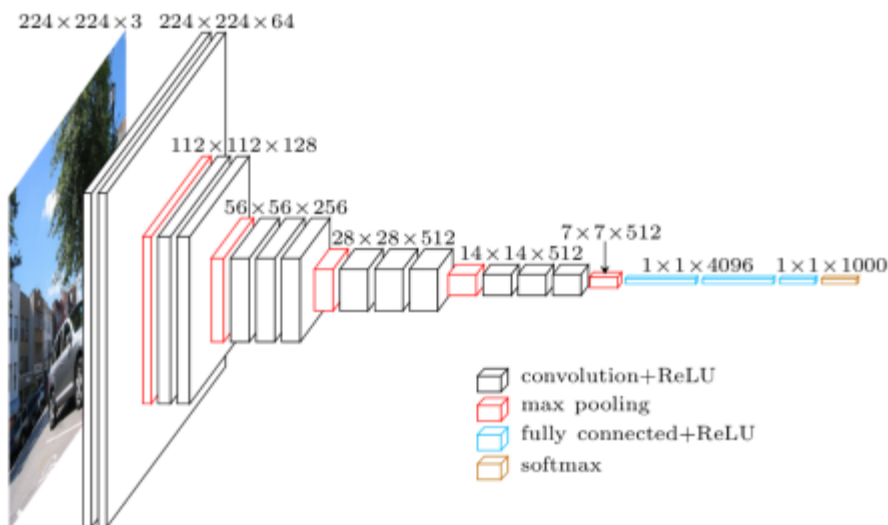
Η αρχιτεκτονική του AlexNet αποτελείται από 60 εκατομμύρια παραμέτρους, με αποτέλεσμα η διαδικασία του overfitting να είναι αναπόφευκτη. Για την αντιμετώπιση του φαινομένου αυτού προτάθηκαν δύο τεχνικές:

Αρχικά χρησιμοποιήθηκε η επαύξηση του συνόλου φωτογραφιών. Αυτό επιτυγχάνεται δημιουργώντας είτε νέες φωτογραφίες λαμβάνοντας τυχαίες περιοχές 224x224 εικονοστοιχείων από τις εικόνες διαστάσεων 256x256 ή αλλάζοντας τον προσανατολισμό των φωτογραφιών.

Η δεύτερη μέθοδος είναι αυτή του dropout, που για πρώτη φορά υλοποιήθηκε στο AlexNet. Η μέθοδος αυτή προτείνει την μηδενική συμβολή κάθε νευρώνα στο δίκτυο, με πιθανότητα 0,5. Οι νευρώνες αυτοί δεν συμμετέχουν ούτε στο εμπρόσθιο πέρασμα ούτε κατά την διάρκεια του backpropagation για την αναβάθμιση των βαρών τους. Με τον τρόπο αυτό για κάθε εικόνα εισόδου, το δίκτυο σχηματίζει μια διαφορετική αρχιτεκτονική. Έτσι δεν δημιουργούνται «συνεργασίες» μεταξύ των νευρώνων, καθώς ένας νευρώνας δεν μπορεί να βασιστεί στην παρουσία κάποιου άλλου. Κατά την διάρκεια της ταξινόμησης, συμμετέχουν όλοι οι νευρώνες πολλαπλασιάζοντας την έξοδο τους με 0,5 (Krizhevsky, Sutskever, Hinton, 2012).

4.2.3.2 VGG Net

Το VGG Net αποτελεί ΣΝΔ, που έλαβε μέρος στο διαγωνισμό ILSVRC 2014, λαμβάνοντας της πρώτη και δεύτερη θέση στον εντοπισμό και ταξινόμηση εικόνων αντιστοίχως (Simonyan & Zisserman, 2015). Αποτελείται από δύο εκδόσεις το VGG-16, με 16 επίπεδα, και το VGG-19 με 19 επίπεδα. Στο πλήθος των επιπέδων προσμετρούνται μόνο τα συνελκτικά και τα fully-connected layers. Η σημαντικότερη συμβολή της δομής του VGG net είναι ότι ανέδειξε πως το βάθος του δικτύου αποτελεί βασικό στοιχείο της καλής απόδοσης. Χρησιμοποιεί συνέλιξη φίλτρων 3x3 και δειγματοληψία σε 2x2 κελιά από την αρχή μέχρι το τέλος του αλγορίθμου. Σημαντικό μειονέκτημα του VGG-Net είναι το τεράστιο πλήθος παραμέτρων, που φτάνει τα 140 εκατομμύρια στο VGG-16 και τα 160 εκατομμύρια στο VGG-19, με τις περισσότερες από αυτές να βρίσκονται στα επίπεδα πλήρους σύνδεσης.



Εικόνα 4.29. Αρχιτεκτονική του VGGnet (Cord, 2016)

4.3 Μηχανές Διανυσμάτων Υποστήριξης

Παραπάνω αναφέρθηκε η λειτουργία των ΣΝΔ, τα οποία λειτουργούν τόσο ως εξαγωγείς χαρακτηριστικών όσο και ως ταξινομητές. Το επόμενο είδος ταξινομητή το οποίο χρησιμοποιείται για την ανάπτυξη του συστήματος αυτόματης αναγνώρισης τροφής της παρούσας εργασίας είναι οι Μηχανές Διανυσμάτων Υποστήριξης ή αλλιώς SVMs.

Οι μηχανές διανυσμάτων υποστήριξης είναι δυαδικές μηχανές με δυνατότητα μάθησης, επιδεικνύοντας κάποιες εξαιρετικές ιδιότητες συγκριτικά με άλλες τεχνικές ταξινόμησης. Η βασική ιδέα των SVM σύμφωνα με τον Haykin είναι: Δοθέντος ενός δείγματος εκπαίδευσης, οι μηχανές διανυσμάτων υποστήριξης κατασκευάζουν ένα υπερεπίπεδο ως επιφάνεια απόφασης με τρόπο ώστε το περιθώριο διαχωρισμού μεταξύ θετικών και αρνητικών παραδειγμάτων να μεγιστοποιείται (2010).

Έστω ότι υπάρχει ένα δείγμα εκπαίδευσης (training set) της μορφής $\{(x_i, d_i)\}$, που αποτελείται από N στοιχεία, με x_i να αντιστοιχεί στο διάνυσμα εισόδου της i -οστής παρατήρησης και το d_i να αποτελεί την αντίστοιχη επιθυμητή απόκριση. Θεωρείται ότι οι αποκρίσεις μπορεί να διακριθούν σε δύο κλάσεις με τιμές $d_i = 1$ ή $d_i = -1$, καθώς επίσης και ότι οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες. Η εξίσωση κάθε υπερεπιπέδου διαχωρισμού, δύο γραμμικά διαχωρίσιμων κλάσεων, δίνεται από τον τύπο:

$$w^T x + b = 0 \quad (4.39)$$

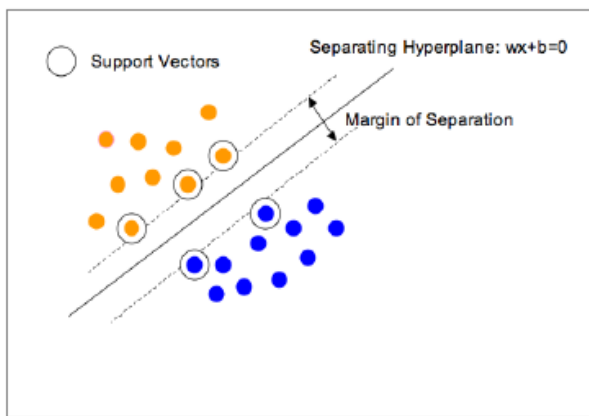
με w να αποτελεί κανονικό διάνυσμα του υπερεπιπέδου, x το διάνυσμα εισόδου και η παράμετρος $\frac{b}{\|w\|}$ να προσδιορίζει την απόσταση του υπερεπιπέδου από την αρχή των αξόνων, με διεύθυνση εκείνη του w .

Δεδομένου ότι οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες από την προαναφερθείσα εξίσωση διαχωρισμού, τότε τα δύο επιμέρους υπερεπίπεδα που προκύπτουν μπορούν να περιγραφούν από τις παρακάτω ανισώσεις και εμπεριέχουν μια από τις δύο κλάσεις το καθένα.

$$w^T x + b \geq 0 \quad \text{για } d_i = 1 \quad (4.40)$$

$$w^T x + b \leq 0 \quad \text{για } d_i = -1 \quad (4.41)$$

Για δεδομένες τιμές των παραμέτρων w και b η απόσταση που ορίζεται από το υπερεπίπεδο της εξίσωσης (4.39) και του πλησιέστερου σημείου δεδομένων ονομάζεται περιθώριο διαχωρισμού (margin of separation), και απεικονίζεται στην Εικόνα 4.30.



Εικόνα 4.30. Απεικόνιση γραμμικού διαχωρισμού από SVM, δείχνοντας το περιθώριο διαχωρισμού και την συνάρτηση του υπερεπίπεδου που διαχωρίζει τις δύο κλάσεις (Caragea, Cook, & Honavar, 2005)

Σκοπός του SVM είναι η εύρεση του υπερεπιπέδου εκείνου που μεγιστοποιεί το περιθώριο διαχωρισμού, το επίπεδο αυτό ονομάζεται βέλτιστο υπερεπίπεδο. Αν οι τιμές του διανύσματος βαρών και της πόλωσης που χρησιμοποιούνται στην εξίσωση του υπερεπιπέδου είναι w_0 και b_0 , τότε το βέλτιστο υπερεπίπεδο περιγράφεται από την εξίσωση

$$w_0^T x + b_0 = 0 \quad (4.42)$$

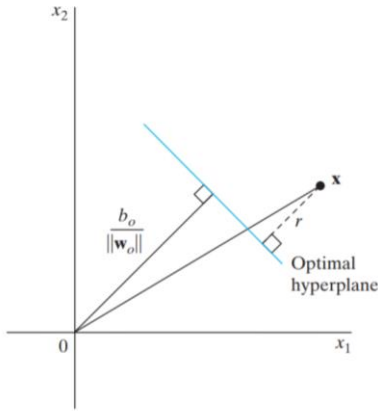
Η απόσταση των διανυσμάτων εισόδου από το βέλτιστο υπερεπίπεδο σύμφωνα με τους Duda και Hart (1973), δίνεται από την συνάρτηση:

$$g(x) = w_0^T x + b_0 \quad (4.43)$$

Η συνάρτηση αυτή ονομάζεται συνάρτηση διάκρισης.

Αναλύοντας γεωμετρικά τον χώρο του SVM (Εικόνα 4.31), προκύπτει ότι η απόσταση ενός διανύσματος εισόδου από το υπερεπίπεδο διαχωρισμού δίνεται από την σχέση (4.44) και περιγράφεται μέσω της μεταβλητής r :

$$r = \frac{g(x)}{\|w_0\|} \quad (4.44)$$



Εικόνα 4.31. Γεωμετρική ερμηνεία της απόστασης σημείου από το βέλτιστο υπερεπίπεδο (Haykin, 2010)

Τέλος ισχύει ότι τα w_0 και b_0 πρέπει να ακολουθούν τους περιορισμούς που περιγράφονται από τις συναρτήσεις:

$$w_0^T x + b_0 \geq 1 \quad \text{για } d_i = 1 \quad (4.45)$$

$$w_0^T x + b_0 \leq -1 \quad \text{για } d_i = -1 \quad (4.46)$$

Τα σημεία εκείνα για τα οποία ικανοποιείται το σύμβολο ισότητας των σχέσεων (4.45) και (4.46) αποτελούν τα διανύσματα υποστήριξης (support vector). Πρόκειται δηλαδή για τα σημεία εκείνα, που βρίσκονται κοντινότερα στα σύνορα διαχωρισμού των δύο κλάσεων και η θέση τους είναι αυτή που προσδιορίζει το σχεδιασμό του βέλτιστου υπερεπιπέδου. Συνεπώς μέσω του SVM επιτυγχάνεται ταξινόμηση όλων των παρατηρήσεων εισόδου, παρ' όλα αυτά μόνο ένα μικρό υποσύνολο των σημείων αυτών προσδιορίζουν το βέλτιστο υπερεπίπεδο, τα οποία στην διανυσματική τους απεικόνιση ονομάζονται διανύσματα υποστήριξης.

Μέσω της σχέσης (4.44) έχει οριστεί η αλγεβρική απόσταση ενός διανύσματος εισόδου από το υπερεπίπεδο διαχωρισμού, η οποία λαμβάνει ειδική μορφή στην περίπτωση των διανυσμάτων υποστήριξης. Συγκεκριμένα μέσω των σχέσεων (4.45), (4.46) η συνάρτηση διάκρισης λαμβάνει την τιμή:

$$g(x^{(s)}) = w_0^T x^{(s)} + b_0 = \pm 1, \quad \text{για } d_s = \pm 1$$

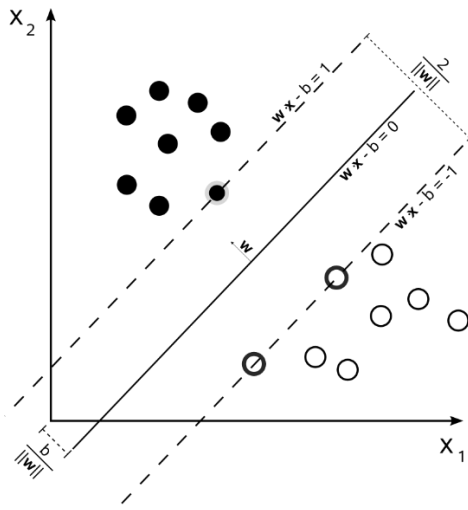
Έτσι αντικαθιστώντας στην (4.44) προκύπτει η μορφή:

$$r^{(s)} = \frac{g(x^{(s)})}{\|w_0\|} = \begin{cases} \frac{1}{\|w_0\|}, & \text{εάν } d_s = 1 \\ -\frac{1}{\|w_0\|}, & \text{εάν } d_s = -1 \end{cases} \quad (4.47)$$

όπου το πρόσημο προσδιορίζει αν το διάνυσμα υποστήριξης βρίσκεται στην αρνητική ή θετική πλευρά του βέλτιστου υπερεπιπέδου. Με αποτέλεσμα η βέλτιστη τιμή του επιπέδου διαχωρισμού να προσδιορίζεται ως το διπλάσιο της απόστασης ενός διανύσματος υποστήριξης από το βέλτιστο υπερεπίπεδο. Αυτό συμβαίνει καθώς τόσο τα διανύσματα υποστήριξης που βρίσκονται στα αρνητικά όσο και αυτά που βρίσκονται στα θετικά του υπερεπιπέδου, ισαπέχουν από αυτό. Η απόσταση αυτή στην περίπτωση των διανυσμάτων υποστήριξης λαμβάνει την μορφή και φαίνεται στην Εικόνα 4.32.

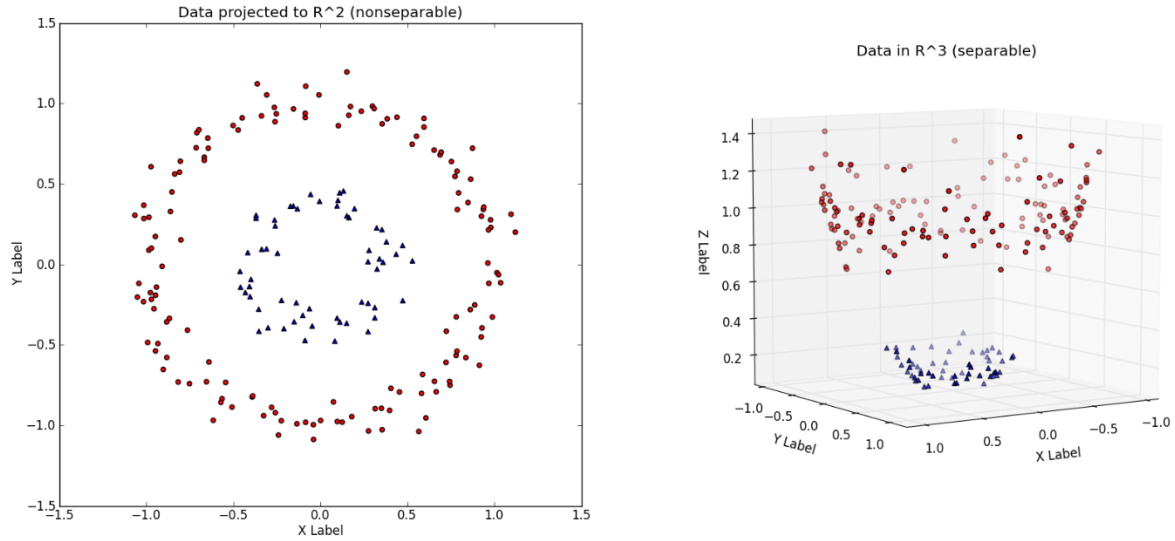
$$\rho = 2r^{(s)} = \frac{2}{\|w_0\|} \quad (4.48)$$

Η εξίσωση (4.48) δηλώνει ότι η μεγιστοποίηση του περιθωρίου διαχωρισμού μεταξύ δυαδικών κλάσεων είναι ισοδύναμη με την ελαχιστοποίηση της Ευκλείδειας νόρμας του διανύσματος βαρών w (Haykin, 2010).



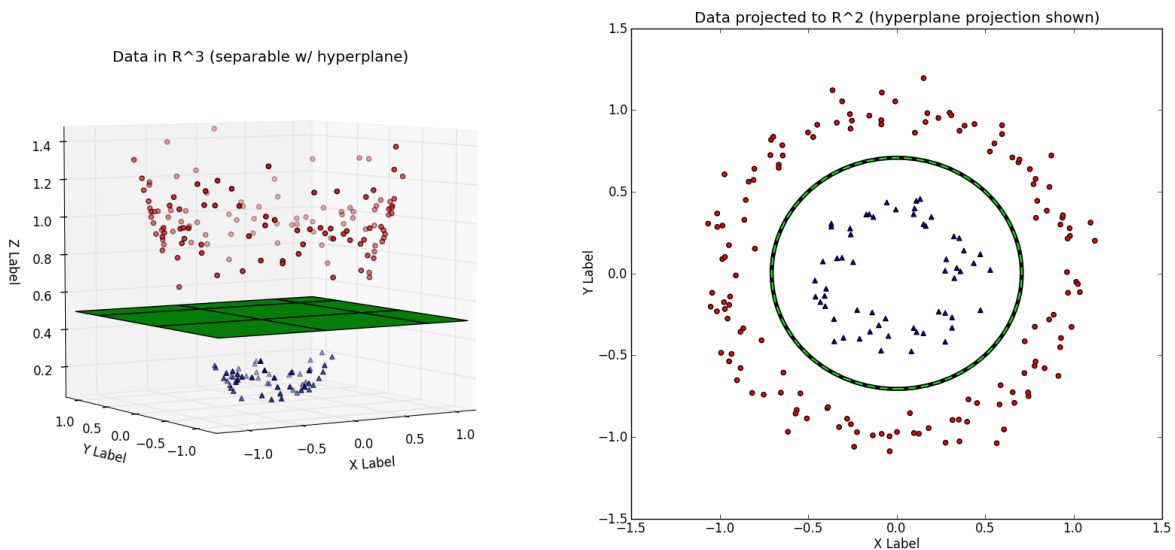
Εικόνα 4.32. Οπτικοποίηση των σημαντικότερων μεγεθών του SVM (Wikipedia, n.d.)

Τις περισσότερες φορές όμως τα προβλήματα που τίθενται προς ταξινόμηση από τα SVM, δεν είναι γραμμικά διαχωρίσιμα. Για το λόγο αυτό η τεχνική που χρησιμοποιείται, είναι η επίλυση του προβλήματος σε ένα χώρο μεγαλύτερων διαστάσεων, στον οποίο υπάρχει υπερεπίπεδο, το οποίο διαχωρίζει γραμμικά τις παρατηρήσεις των δύο κλάσεων.



Εικόνα 4.33. (Αριστερά) Ένα σύνολο δεδομένων μη γραμμικά διαχωρίσιμο στον δισδιάστατο χώρο. (Δεξιά) Το ίδιο σύνολο δεδομένων στον τρισδιάστατο χώρο μέσω της μετατροπής $[x_1, x_2] = [x_1 \ x_2 \ x_1^2 + x_2^2]$

Όπως φαίνεται στην Εικόνα 4.33 είναι δυνατόν μέσω ενός μετασχηματισμού ένα δισδιάστατο μη γραμμικό πρόβλημα να μεταφερθεί σε ένα χώρο μεγαλύτερων διαστάσεων στον οποίο μπορεί να υπάρχει γραμμική διαχωρισιμότητα (Εικόνα 4.34). Η συνάρτηση μετάβασης συνήθως περιγράφεται από το σύμβολο ϕ και στο συγκεκριμένο παράδειγμα έχει τον τύπο $\phi([x_1, x_2]) = [x_1 \ x_2 \ x_1^2 + x_2^2]$.



Εικόνα 4.34.(Αριστερά) Υπερεπίπεδο που διαχωρίζει τις δύο κλάσεις στον τρισδιάστατο χώρο (Δεξιά) Μετατροπή του υπερεπιπέδου στον δισδιάστατο χώρο (Kim, 2013)

Η ιδιότητα αυτή μπορεί να γενικευθεί για την μετάβαση από ένα N -διάστατο χώρο σε ένα M -διάστατο ($M > N$), προκειμένου να βρεθεί υπερεπίπεδο διαχωρισμού των σημείων των δύο κλάσεων. Παρ' όλα αυτά η μετατροπή σε χώρο μεγαλύτερων διαστάσεων αποτελεί υπολογιστικά σύνθετη διεργασία. Για

το λόγο αυτό δεν είναι δυνατή η μετατροπή των σημείων εισόδου αρχικά στον χώρο M διαστάσεων και στην συνέχεια μετά την εύρεση του βέλτιστου υπερεπιπέδου, η αποτύπωση του στον N -διάστατο χώρο (Kim, Everything You Wanted to Know about the Kernel Trick, 2013).

Έχει παρατηρηθεί λοιπόν ότι δεν είναι απαραίτητη η μετατροπή κάθε σημείου εισόδου, σε ένα χώρο περισσότερων διαστάσεων προκειμένου να χρησιμοποιηθεί ο ταξινομητής SVM. Αντιθέτως έχει αποδειχθεί ότι το πρόβλημα βελτιστοποίησης μπορεί να επιλυθεί χρησιμοποιώντας μόνο το εσωτερικό γινόμενο ανάμεσα στις παρατηρήσεις εισόδου στον νέο χώρο M διαστάσεων. Αν η συνάρτηση φ στέλνει τις παρατηρήσεις εισόδου σε ένα νέο M -διάστατο χώρο, τότε η επίλυση της βελτιστοποίησης απαιτεί μόνο τον υπολογισμό του $\varphi(x)^T \varphi(z)$ για παρατηρήσεις εισόδου x, z . Ο υπολογισμός εσωτερικών γινομένων στον νέο χώρο επιτυγχάνεται μέσω μιας συνάρτησης πυρήνα K που ορίζεται στον αρχικό χώρο και ικανοποιεί τη σχέση $K(x, z) = \varphi(x)^T \varphi(z)$ για οποιαδήποτε σημεία x, z του αρχικού χώρου. Έτσι επιτυγχάνεται επίλυση στο νέο χώρο μεγαλύτερων διαστάσεων χωρίς τον υπολογισμό της αναπαράστασης των δεδομένων στο χώρο αυτό.

Κάποιες από τις σημαντικότερες συναρτήσεις πυρήνα είναι οι:

Γραμμικός πυρήνας

$$k(x, y) = x^T y + c \quad (4.49)$$

Γκαουσιανός πυρήνας ή Ακτινικής βάσης Συνάρτηση Πυρήνα

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (4.50)$$

Πολυωνυμικός πυρήνας

$$k(x, y) = (ax^T y + c)^d \quad (4.51)$$

Με το d να αποτελεί το βαθμό του πολυωνύμου, το a να επηρεάζει την κλίση ενώ τέλος το c να είναι σταθερός όρος (Souza, 2010).

4.3.1 Ταξινόμηση Μηχανών Διανυσμάτων Υποστήριξης πολλών κλάσεων

Οι Μηχανές Διανυσμάτων Υποστήριξης μπορούν να επεκταθούν στην ταξινόμηση συνόλων δεδομένων που ανήκουν σε παραπάνω από δύο κλάσεις. Οι δύο επικρατούσες μέθοδοι στην ταξινόμηση παρατηρήσεων N κλάσεων, με SVM είναι:

One-vs-All:

Στην περίπτωση του one-vs-all θα αξιοποιηθούν τόσο ταξινομητές όσες και οι διαφορετικές κλάσεις. Κάθε ταξινομητής θεωρεί θετικές τις ετικέτες μιας μόνο κλάσης και όλες τις υπόλοιπες αρνητικές. Έτσι χρησιμοποιούνται διαφορετικά SVM, ένα για κάθε κλάση, προκειμένου να υλοποιηθεί η ταξινόμηση. Το σημαντικότερο μειονέκτημα της μεθόδου αυτής είναι ότι δημιουργούνται μη ισορροπημένα σύνολα δεδομένων, για κάθε δυαδικό ταξινομητή.

One-vs-One:

Στην περίπτωση της τεχνικής one-vs-one χρησιμοποιείται ένα SVM ανά δύο κλάσεις. Συνεπώς συνολικά χρειάζονται $\frac{N \cdot (N-1)}{2}$ ταξινομητές, δεδομένου ότι υπάρχουν N κλάσεις στο σύνολο δεδομένων. Η χρήση του μεγάλου πλήθους ταξινομητών καθιστά την μέθοδο αυτή αρκετά πιο χρονοβόρα.

4.4 Μέθοδοι Επικύρωσης

4.4.1 Μέθοδος Παρακράτησης (Holdout Method)

Στην περίπτωση της μεθόδου παρακράτησης, το σύνολο δεδομένων χωρίζεται σε δύο υποσύνολα, το σύνολο εκπαίδευσης και το σύνολο ελέγχου. Το μέγεθος κάθε ένα από τα δύο σύνολα είναι αυθαίρετα ορισμένα, αν και συνήθως προτιμάται το σύνολο εκπαίδευσης να είναι μεγαλύτερο από το σύνολο ελέγχου. Το σύστημα διαμορφώνει την συνάρτηση συμπεριφοράς του, σύμφωνα με τα παραδείγματα που υπάρχουν στο σύνολο εκπαίδευσης και στην συνέχεια προβλέπει τις τιμές για το σύνολο ελέγχου. Σημαντικό μειονέκτημα της τεχνικής αυτής είναι ότι δεν μπορεί να χρησιμοποιηθεί όταν υπάρχει μικρό πλήθος δεδομένων. Από την άλλη μεριά ενδείκνυται για πολύ μεγάλα σύνολα δεδομένων, λόγω της περιορισμένης πολυπλοκότητας της μεθόδου.

4.4.2 Διασταυρωμένη Επικύρωση σε k-μέρη (k-fold Cross Validation)

Η μέθοδος της διασταυρωμένης επικύρωσης αποτελεί βελτίωση της μεθόδου παρακράτησης. Το σύνολο δεδομένων διακρίνεται σε κ-υποσύνολα με ίσο πλήθος παρατηρήσεων. Κάθε φορά ένα από τα κ υποσύνολα χρησιμοποιείται ως σύνολο ελέγχου και τα υπόλοιπα κ-1 υποσύνολα χρησιμοποιούνται ως σύνολο εκπαίδευσης. Η διαδικασία αυτή επαναλαμβάνεται κ-φορές και υπολογίζεται το μέσο σφάλμα από τις κ δοκιμές. Η διαδικασία αυτή έχει το πλεονέκτημα ότι δεν έχει σημασία ο τρόπος με τον οποίο χωρίζεται το σύνολο, καθώς όλες οι παρατηρήσεις θα χρησιμοποιηθούν τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο ελέγχου. Όσο μεγαλύτερη είναι η τιμή κ, τόσο μικρότερη διακύμανση υπάρχει στο τελικό

αποτέλεσμα. Το μειονέκτημα αυτής της διαδικασίας είναι ότι πρέπει να εκτελεσθεί κ-φορές, κάνοντας την δύσχρηστη σε περιπτώσεις με πολύ μεγάλο σύνολο δεδομένων (Schneider, 1997).

4.5 Μέτρα Αξιολόγησης

4.5.1 Ακρίβεια

Το στοιχειώδες μέτρο της απόδοσης ενός συστήματος είναι αυτό της ακρίβειας. Προκύπτει ως το πηλίκο των σωστά ταξινομημένων παρατηρήσεων, προς το συνολικό πλήθος των παρατηρήσεων. Αποτελεί την πρώτη ένδειξη για την αποτελεσματικότητα ενός αλγορίθμου, χωρίς να δίνει περισσότερη πληροφορία σχετικά με τα είδη των σφαλμάτων που έχουν γίνει. Η χρήση της μετρικής της ακρίβειας ευνοείται στην παρούσα διπλωματική εργασία, επειδή επιλέγεται κάθε κλάση να αποτελείται από ίσο πλήθος εικόνων.

4.5.2 Μήτρα Σύγχυσης (Confusion Matrix)

Ο βασικότερος σκοπός της παρούσας διπλωματικής εργασίας είναι η υλοποίηση ενός ταξινομητή πολλαπλών κλάσεων (multiclass classifier), ο οποίος θα επιτυγχάνει την βέλτιστη ακρίβεια. Για τον λόγο αυτό δεν αρκεί να παρουσιάζεται μόνο ένα ποσοστό των σωστά ταξινομημένων παρατηρήσεων, αλλά αναλυτικά θα πρέπει να σκιαγραφείται η απόδοση του ταξινομητή σε κάθε κατηγορία, αλλά και το είδος των λαθών που συμβαίνουν. Το είδος των λαθών φέρει μεγάλη σημασία στο πλαίσιο υπολογισμού της περιεχόμενης ποσότητας υδατανθράκων. Αυτό συμβαίνει επειδή κάθε κατηγορία φαγητών εμπεριέχει διαφορετική περιεκτικότητα υδατανθράκων, με κάποιες από αυτές να εμφανίζουν μικρές αποκλίσεις μεταξύ τους και άλλες μεγαλύτερες, με αποτέλεσμα κάθε λάθος να επιβαρύνει με διαφορετικό τρόπο το τελικό αποτέλεσμα.

Η μήτρα σύγχυσης αποτελεί ένα μέτρο αξιολόγησης που στην περίπτωση της δυαδικής ταξινόμησης χρησιμοποιείται για την καλύτερη περιγραφή των αποτελεσμάτων ταξινόμησης. Αυτό συμβαίνει καθώς στην περίπτωση της δυαδικής ταξινόμησης, όταν οι δύο κλάσεις διαφέρουν σημαντικά στο πλήθος των παρατηρήσεων από τις οποίες αποτελούνται, το αποτέλεσμα της ακρίβειας μπορεί να διαστρεβλώσει την λειτουργία του ταξινομητή. Στην περίπτωση της ταξινόμησης πολλών κλάσεων δίνεται η δυνατότητα παρακολούθησης της ακρίβειας κάθε κλάσης αλλά και τον εντοπισμό της σύγχυσης ανάμεσα σε δύο κλάσεις.

Διαγραμματικά η Μήτρα Σύγχυσης απεικονίζεται στην Εικόνα 4.35. Όπως φαίνεται η μήτρα σύγχυσης αναπαριστά είτε μια δυαδική ταξινόμηση, ή στην περίπτωση ταξινόμησης πολλών κλάσεων την ταξινόμηση ανάμεσα στην κλάση X και όχι στην κλάση \bar{X} . Ουσιαστικά μέσω της αναπαράστασης της

Εικόνα 4.35 ορίζονται κάποιοι θεμελιώδεις όροι που επιτρέπουν την εξαγωγή περιγραφικών μέτρων αξιολόγησης της ταξινόμησης. Οι όροι αυτοί είναι:

Αληθώς θετικό: Η περίπτωση που η προβλέφθηκε θετική μια παρατήρηση και όντως είναι θετική

Αληθώς αρνητικό: Η περίπτωση στην οποία προβλέφθηκε ως αρνητική μια κλάση και όντως είναι αρνητική

Ψευδώς θετικό: Η περίπτωση κατά την οποία προβλέφθηκε θετική μια παρατήρηση ενώ είναι αρνητική

Ψευδώς αρνητικό: Η περίπτωση κατά την οποία προβλέφθηκε αρνητική μια παρατήρηση ενώ είναι θετική

		Κλάση που προβλέφθηκε	
		Θετικό	Αρνητικό
Πραγματική κλάση	Θετικό	Αληθώς θετικό	Ψευδώς αρνητικό
	Αρνητικό	Ψευδώς θετικό	Αληθώς αρνητικό

Εικόνα 4.35. Αναπαράσταση Μήτρας Σύγχυσης

Σύμφωνα με τους όρους αυτούς και την πληροφορία που περιέχεται στην μήτρα σύγχυσης μπορεί να δημιουργηθούν διάφορες μετρικές αξιολόγησης της ταξινόμησης. Οι δύο βασικότερες είναι η ακρίβεια και η ανάκληση (recall). Η ακρίβεια αποτελεί την πρώτη μετρική που έχει οριστεί στο υποκεφάλαιο αυτό και ορίζεται ως το πλήθος των σωστά ταξινομημένων παρατηρήσεων (στην κλάση A), από όλες τις παρατηρήσεις που έχουν προβλεφθεί σε αυτή την κλάση. Δηλαδή:

$$\frac{\text{Αληθώς_Θετικό}}{\text{Αληθώς_Θετικό} + \text{Ψευδώς_Αρνητικό}} \quad (4.52)$$

Η ανάκληση ορίζεται ως το πλήθος των παρατηρήσεων οι οποίες έχουν προβλεφθεί σωστά για μία κλάση, προς το συνολικό πλήθος των παρατηρήσεων που έχουν προβλεφθεί στην κλάση αυτή και περιγράφεται από το λόγο (4.53).

$$\frac{\text{Αληθώς_Θετικό}}{\text{Αληθώς_Θετικό} + \text{Ψευδώς_Θετικό}} \quad (4.53)$$

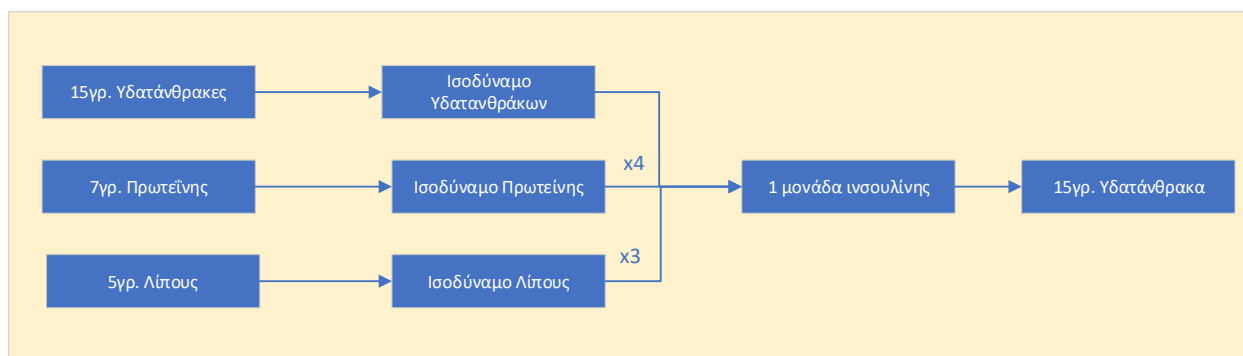
Παρατηρείται συνεπώς ότι η πρώτη μετρική δίνεται ως τα Αληθώς Θετικά προς την πρώτη γραμμή της Εικόνα 4.35, ενώ η δεύτερη μετρική προκύπτει ως το πηλίκο των αληθώς θετικών προς την πρώτη στήλη της ίδιας εικόνας. Στην περίπτωση της ταξινόμησης σε περισσότερες από δύο κλάσεις, η μήτρα σύγχυσης αποτελείται από τόσες γραμμές και στήλες όσο και το πλήθος των κλάσεων. Με τον τρόπο ο

οποίος παρουσιάστηκε παραπάνω, γενικεύεται η μεθοδολογία υπολογισμού της ακρίβειας και της ανάκλησης σε περισσότερες από δύο κλάσεις.

4.5.3 Μετρική βασισμένη στην Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων

Οι παραπάνω μετρικές αποτελούν γνωστές τεχνικές αξιολόγησης ταξινομητών. Στην παρούσα διπλωματική εργασία δημιουργείται μήτρα αξιολόγησης, η οποία επεξεργάζεται τα αποτελέσματα της Μήτρα Σύγχυσης, και έχει αναπτυχθεί ειδικά για την αξιολόγηση των αποτελεσμάτων ταξινόμησης στον κλάδο αναγνώρισης της περιεχόμενης ποσότητας υδατανθράκων από φωτογραφικά στιγμιότυπα. Η μήτρα αυτή θα αποκαλείται Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων. Το αποτέλεσμα που επιστρέφει η μετρική, η οποία βασίζεται στην Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων, αποτελεί ένδειξη της μέσης απόκλισης υδατανθράκων ανά φαγητό.

Στο Κεφάλαιο 5, αναφέρεται ότι το μεταγευματικό σάκχαρο δεν εξαρτάται μόνο από τους υδατάνθρακες (ισοδύναμα υδατανθράκων) αλλά και από την περιεκτικότητα του γεύματος σε πρωτεΐνες (ισοδύναμα πρωτεϊνών) και λίπος (ισοδύναμα λίπους). Συγκεκριμένα ορίζεται ως ισοδύναμο υδατανθράκων τα 15γρ. υδατανθράκων, ισοδύναμο πρωτεΐνης τα 7γρ. πρωτεΐνης και ισοδύναμο λίπους τα 5γρ. λίπους. Κάθε άτομο με ΣΔΤ1, έχει διαφορετική ευαισθησία στην ινσουλίνη καθώς επίσης η κατανάλωση των μακροθρεπτικών συστατικών των υδατανθράκων, του λίπους και της πρωτεΐνης προκαλεί διαφορετικές μεταβολές στο σάκχαρο του. Παρ' όλα αυτά, μια απλή γενίκευση είναι ότι ένα ισοδύναμο υδατανθράκων ή 4 ισοδύναμα πρωτεΐνης ή 3 ισοδύναμα λίπους ισοδυναμούν μεταξύ τους και αντιστοιχίζονται σε μια μονάδα ινσουλίνης. Στην προκειμένη περίπτωση δεδομένου ότι αναζητάται μια δυαδική σχέση ανάμεσα ενός μεγέθους και των επιπτώσεων του στο σάκχαρο, όπως φαίνεται στην Εικόνα 4.36, μπορούν να αναχθούν όλα τα μακροθρεπτικά συστατικά σε μια ποσότητα υδατανθράκων. Αυτό ουσιαστικά σημαίνει ότι 15γρ.υδατάνθρακα ή 28γρ. πρωτεΐνης ή 15 γραμμάρια λίπους δημιουργούν παρόμοιες μεταγευματικές μεταβολές. Σύμβαση η οποία εν μέρει ισχύει αν και είναι αρκετά απλουστευμένη. Με τον τρόπο αυτό, παρ' όλο που τόσο οι πρωτεΐνες όσο και το λίπος επηρεάζουν το μεταγευματικό σάκχαρο, την επίδραση τους την ανάγουμε σε μια ποσότητα υδατανθράκων. Με σκοπό η μεταβολή του σακχάρου των ατόμων με ΣΔ μετά από κάθε γεύμα, να εξαρτάται μόνο από τον γενικευμένο όρο των υδατανθράκων.



Εικόνα 4.36. Διαγραμματική αναπαράσταση αντιστοίχισης συγκεκριμένων μακροθρεπτικών συστατικών σε 1 μονάδα ινσουλίνης, ανάγοντας τα τελικώς σε 15γρ. υδατανθράκων.

Αντίστοιχα με τα ισοδύναμα του υδατάνθρακα, της πρωτεΐνης και του λίπους, ορίζονται και ισοδύναμα για κάθε ομάδα φαγητών. Πρόκειται για ομάδες φαγητών με παρόμοια μακροθρεπτικά συστατικά, δίνοντας έτσι την δυνατότητα στον χρήστη να εμπλουτίσει τη διατροφή του με εναλλακτικές επιλογές, οι οποίες έχουν παρόμοιες επιδράσεις στο σάκχαρο. Τα ισοδύναμα του Γάλακτος, των Λαχανικών, των Φρούτων, των Ψωμί-Δημητριακών, των Οσπρίων και Αμυλούχων Λαχανικών και του Κρέατος έχουν οριστεί από την Ελληνική Διαβητολογική Εταιρεία (Ελληνική Διαβητολογική Εταιρεία, 2013). Από την άλλη μεριά τα Ισοδύναμα των Breaded Food και Λιπαρών Φαγητών ορίζονται στα πλαίσια της παρούσας διπλωματικής εργασίας. Δεν υπάρχει σταθερή μεταβολή του μεταγευματικού σακχάρου από όλα τα τρόφιμα αυτών των κατηγοριών, ούτε έχει βρεθεί μια προκαθορισμένη ποσότητα γραμμαρίων στην οποία αντιστοιχίζεται το ισοδύναμο κάθε ομάδας φαγητών. Για τον λόγο αυτό εμπειρικά και με ανάλυση του διατροφικού περιεχομένου κάποιων φαγητών αυτών των δύο ομάδων, ορίζεται μια ενδεικτική τιμή ινσουλίνης και γραμμαρίων υδατάνθρακα, όπως φαίνεται στον Πίνακα 4.1. Οι μονάδες ινσουλίνης για κάθε ισοδύναμο δίνεται από τον τύπο (4.54).

$$\text{Μονάδες Ινσουλίνης} = \frac{\text{Υδατάνθρακες}}{15} + \frac{\text{Πρωτεΐνη}}{28} + \frac{\text{Λίπος}}{15} = \frac{\text{Γραμμάρια Υδατάνθρακα}}{15} \quad (4.54)$$

Ισοδύναμο	Υδατάνθρακες (γρ)	Πρωτεΐνη (γρ)	Λίπος (γρ)	Μονάδες ινσουλίνης	Γραμμάρια Υδατάνθρακα
Ισοδύναμο Γάλακτος	12	8		1.1	16.29
Ισοδύναμο Λαχανικών	5	2		0.4	6.07
Ισοδύναμο Φρούτων	15			1.0	15.00
Ισοδύναμο Ψωμί- Δημητριακών	15		0	1.0	15.00
Ισοδύναμο Οσπρίων και Αμυλούχων Λαχανικών	10	7		0.9	13.75
Ισοδύναμο Κρέατος		7	5	0.6	8.75
Ισοδύναμο Breaded Food				2.0	30.00
Ισοδύναμο Λιπαρών Φαγητών				2.5	37.50

Πίνακας 4.1. Πίνακας αναγωγής των ισοδυνάμων κάθε ομάδας φαγητών σε μονάδες ινσουλίνης και αντίστοιχα σε γραμμάρια υδατάνθρακα.

Σύμφωνα με όλα τα παραπάνω ορίζεται η Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων, η οποία απεικονίζεται στην Εικόνα 4.37. Αρχικά στον οριζόντιο και στον κατακόρυφο άξονα αναγράφονται οι ομάδες φαγητών και δίπλα τους τα γραμμάρια υδατανθράκων τα οποία αντιστοιχούνται σε αυτές. Η μήτρα

αξιολόγησης διαβάζεται με παρόμοιο τρόπο όπως οι μήτρες σύγκυσης, με την διαφορά όμως ότι περιέχει συντελεστές. Η τιμή για κάθε κελί του πίνακα προσδιορίζεται ως η απόλυτη διαφορά ανάμεσα στην αναγραφόμενη τιμή γραμμαρίων υδατάνθρακα, της γραμμής και τη στήλης, οι οποίες προσδιορίζουν το εκάστοτε κελί.

Μήτρα Αξιολόγησης	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτομικα	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
	30.00	13.75	16.29	8.75	6.07	37.50	15.00	15.00
Breaded Food	30.00	0.00	16.25	13.71	21.25	23.93	7.50	15.00
Όσπρια και Αμυλούχα Λαχανικά	13.75	16.25	0.00	2.54	5.00	7.68	23.75	1.25
Γαλακτομικα	16.29	13.71	2.54	0.00	7.54	10.22	21.21	1.29
Κρέας και Υποκατάστατα	8.75	21.25	5.00	7.54	0.00	2.68	28.75	6.25
Λαχανικά	6.07	23.93	7.68	10.22	2.68	0.00	31.43	8.93
Λιπαρά Φαγητά	37.50	7.50	23.75	21.21	28.75	31.43	0.00	22.50
Φρούτα	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00
Ψωμί Δημητριακά	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00

Εικόνα 4.37. Μήτρα Αξιολόγησης

Για παράδειγμα όπως φαίνεται στην Εικόνα 4.38 για τον υπολογισμό του στοιχείου που έχει κυκλωθεί, η τιμή του προέκυψε ως η απόλυτη διαφορά των τιμών 8,75 και 37,5. Με όρους μήτρας σύγκυσης, το στοιχείο αυτό θα έπρεπε να ταξινομηθεί στην κατηγορία του Κρέατος και Υποκατάστατων, η οποία αποτελείται από 8,75γρ. υδατάνθρακα ανά ισοδύναμο. Παρ' όλα αυτά, ταξινομείται στην κατηγορία των Λιπαρών Φαγητών, η οποία αποτελείται από 37,5γρ. ανά ισοδύναμο. Η διαφορά αυτή μεταξύ των ισοδυνάμων των φαγητών, αναπαριστά την απόκλιση του σακχάρου που τελικά θα εμφανιστεί στην προβλεπόμενη περιεχόμενη ποσότητα υδατανθράκων από τον ταξινομητή για μια εικόνα. Η σύμβαση που γίνεται στην προκειμένη περίπτωση είναι ότι κάθε φωτογραφία, κάθε ομάδα, περιέχει τα μακροθρεπτικά συστατικά του ισοδύναμου στο οποίο ανήκει. Σύμβαση αποδεκτή, καθώς στην παρούσα εργασία δεν έχει αναπτυχθεί μοντέλο υπολογισμού του όγκου.

abs (8.75-37.50)

Μήτρα Αξιολόγησης	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτομικα	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
	30.00	13.75	16.29	8.75	6.07	37.50	15.00	15.00
Breaded Food	30.00	0.00	16.25	13.71	21.25	23.93	7.50	15.00
Όσπρια και Αμυλούχα Λαχανικά	13.75	16.25	0.00	2.54	5.00	7.68	23.75	1.25
Γαλακτομικα	16.29	13.71	2.54	0.00	7.54	10.22	21.21	1.29
Κρέας και Υποκατάστατα	8.75	21.25	5.00	7.54	0.00	2.68	28.75	6.25
Λαχανικά	6.07	23.93	7.68	10.22	2.68	0.00	31.43	8.93
Λιπαρά Φαγητά	37.50	7.50	23.75	21.21	28.75	31.43	0.00	22.50
Φρούτα	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00
Ψωμί Δημητριακά	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00

Εικόνα 4.38. Διαδικασία εύρεσης της τιμής κάθε κελιού της μήτρας αξιολόγησης.

Η μετρική που χρησιμοποιείται, αποτελεί συνάρτηση ανάμεσα σε δύο πίνακες ίδιων διαστάσεων. Αναλυτικά, υπολογίζεται το γινόμενο των κελιών των δύο πινάκων που βρίσκονται στην ίδια θέση, έστω (i,j), και το αποτέλεσμα αποθηκεύεται σε πίνακα ίδιων διαστάσεων στην θέση (i,j). Υλοποιείται ουσιαστικά γινόμενο πινάκων ανά στοιχείο. Τέλος υπολογίζεται το άθροισμα όλων των στοιχείων του παραγόμενου πίνακα. Συνεπώς υλοποιείται ένας τελεστής του αθροίσματος του γινομένου των κελιών της ίδιας θέσης δύο πινάκων, ο οποίος αποτελεί μια ποσοτική αναπαράσταση της απόκλισης της μήτρας σύγκυσης από τα αναμενόμενα αποτελέσματα και περιγράφεται από την σχέση (4.55).

$$P_3 = e^T(P_1 \circ P_2)e \quad (4.55)$$

Με τον τελεστή \circ δηλώνεται το Hadamard γινόμενο και το σύμβολο e αναπαριστά το μοναδιαίο πίνακα στήλη.

Συγκεκριμένα, όταν η μετρική εφαρμόζεται ανάμεσα στην Μήτρα Αξιολόγησης και σε Μήτρα Σύγχυσης, η οποία θα πρέπει να αναφέρεται στο πλήθος των εικόνων και να μην ακολουθεί ποσοστιαία αναπαράσταση, τότε η μετρική υπολογίζει το συνολικό σφάλμα ταξινόμησης. Διαιρώντας τον αριθμό αυτό, με το πλήθος των εικόνων που ταξινομήθηκαν, προκύπτει η μέση απόκλιση υδατανθράκων ανά φαγητό. Σύμφωνα με όσα αναλύθηκαν προηγουμένως, επειδή δεν γίνεται υπολογισμός του όγκου των εικόνων, αλλά και επειδή τα τρόφιμα σε κάθε ομάδα δεν έχουν την ίδια ακριβώς περιεκτικότητα σε υδατάνθρακες, το αποτέλεσμα της μετρικής είναι μια ένδειξη και όχι ο απόλυτος αριθμός σφάλματος. Αν και ένδειξη, προσθέτει για κάθε σφάλμα ταξινόμησης ποινή ανάλογη των επιπτώσεων που θα υπάρξουν στο μεταγευματικό σάκχαρο του ασθενή, για το λόγο αυτό αποτελεί ισχυρό μέτρο αξιολόγησης.

5 Σύνολο Δεδομένων

Ένα από τους σημαντικότερους παράγοντες στην διαδικασία της ταξινόμησης των φαγητών, είναι αυτό της επιλογής του συνόλου δεδομένων στο οποίο θα εφαρμοστεί ο αλγόριθμος ταξινόμησης. Για τον σκοπό αυτό έχουν προταθεί πολλά σύνολα δεδομένων (datasets).

5.1 Γνωστά Σύνολα Δεδομένων

Σύνολο Δεδομένων	Κατηγορίες Φαγητών	Εικόνες ανά Κατηγορία	Συνολικό Αριθμός Φαγητών	Φαγητά στην Εικόνα	Διαθεσιμότητα
Food 101	101	1000	101,000	Ένα	Ναι
PFID	101	45	4,545	Ένα	Ναι
UNICT-FD889	889	4	3,583	Ένα	Ναι
UNICT-FD1200	1200	4	4,745	Ένα	Ναι
UEC Food 100	100	90	9060	Ένα	Ναι
UEC Food 256	256	122	31,397	Ένα	Ναι
FoDD	23		3,000	Ένα και Πολλαπλά	Ναι
Diabetes	11		4,868	Ένα	Άγνωστο
UNIMIB 2015	15		2,000	Πολλαπλά	Άγνωστο
UNIMIB 2016	73		1027 (3616)	Πολλαπλά	Ναι
Food-11K	11		16643		Ναι

Εικόνα 5.1. Γνωστά Σύνολα Δεδομένων και οι βασικές πληροφορίες που σχετίζονται με αυτά. Στην στήλη Εικόνες ανά Κατηγορία αναγράφονται κατά μέσο όρο οι εικόνες ανά κατηγορία για κάθε φαγητό.

Food 101: Το Food 101 αποτελεί σύνολο δεδομένων που αποτελείται από 101 κατηγορίες διαφορετικών φαγητών, με 1000 φωτογραφίες ανά κατηγορία (Bossard, Guillaumin, & Van Gool, 2014). Η πληρότητα του συνόλου δεδομένων αυτού, σε συνδυασμό με το μεγάλο πλήθος εικόνων που περιέχει το έχουν αναδείξει σε ένα από τα επικρατέστερα για εφαρμογές αναγνώρισης τροφής.

Pittsburgh Fast Food Image Dataset (PFID): Αποτελείται από 101 διαφορετικά φαγητά που προέρχονται από 11 δημοφιλείς αλυσίδες καταστημάτων fast food (Chen, et al., 2009). Οι φωτογραφίες αυτές έχουν ληφθεί τόσο σε περιβάλλον εργαστηρίου με άσπρο υπόβαθρο, όσο και σε πραγματικές συνθήκες. Για περαιτέρω ανάλυση εμπεριέχονται ακόμα stereo φωτογραφίες καθώς και βίντεο. Για κάθε φαγητό υπάρχουν διαφορετικά στιγμιότυπα λήψης του.

UEC – FOOD 100/256: Αρχικά δημοσιεύτηκε το UEC-FOOD 100, με 100 είδη φαγητών. Στην συνέχεια το σύνολο δεδομένων αυτό επεκτάθηκε, δημιουργώντας το UEC-FOOD 256, αποτελούμενο από 256 κατηγορίες φαγητών (Kawano & Yanai, 2014). Για κάθε φωτογραφία των δύο συνόλων δεδομένων υπάρχουν συντεταγμένες οριοθέτησης (bounding box) του φαγητού μέσα στην εικόνα. Οι περισσότερες κατηγορίες φαγητών αποτελούν δημοφιλή πιάτα της Ιαπωνέζικης κουζίνας.

UNIMIB 2015: Το UNIMIB 2015 αποτελεί σύνολο δεδομένων αποτελούμενο από 2000 εικόνες από 15 κατηγορίες διαφορετικών φαγητών. Σε κάθε φωτογραφία εμπεριέχονται πολλαπλά φαγητά. Η

ιδιαιτερότητα του συνόλου δεδομένων αυτού είναι ότι περιέχει ζεύγη εικόνων, ένα με την αρχική ποσότητα των φαγητών και ένα μετά το τέλος του γεύματος. Με τον τρόπο αυτό μπορεί να υπολογισθεί το φαγητό που καταναλώθηκε από τον χρήστη και όχι αναγκαστικά η συνολική ποσότητα φαγητού που εμπεριέχεται στην εικόνα.

UNIMIB 2016: Το UNIMIB 2016 αποτελεί ένα από τα τελευταία δημοσιευμένα σύνολα δεδομένων. Αποτελείται από 73 κατηγορίες φαγητών σε 1027 φωτογραφίες (Ciocca, Napoletano, & Schettini, 2017). Αποτελεί ένα από τα ελάχιστα σύνολα δεδομένων, που σε κάθε εικόνα απεικονίζεται ένα σύνολο διαφορετικών πιάτων, άρα και φαγητών. Με αποτέλεσμα αν και συνολικά αποτελείται από 1027 εικόνες, το συνολικό πλήθος διαφορετικών πιάτων είναι 3616. Το σύνολο δεδομένων συνοδεύεται από χαρτογράφηση των εικόνων σε γλώσσα προγραμματισμού MATLAB που περιγράφει την θέση και τα χαρακτηριστικά κάθε πιάτου στην εικόνα.

UNICT-FD889/1200: Πρώτα έγινε η δημοσίευση του UNICT-FD 889 που συντίθεται από 889 διαφορετικά πιάτα φαγητού, ενώ στην πορεία επεκτάθηκε σε 1200 πιάτα με την μορφή του UNICT-FD 1200 (M., Allegra, Moltisanti, Stanco, & Battiato, 2016). Για κάθε είδος φαγητού εμπεριέχονται κατά μέσο όρο 4 φωτογραφίες, για τις οποίες όμως δεν αναγράφεται το όνομα του πιάτου που απεικονίζουν. Συνήθως γίνεται ομαδοποίηση των φαγητών αυτών και επιτελείται ταξινόμηση σε μεγαλύτερες κατηγορίες. Η βασική κατηγοριοποίηση που έχει επικρατήσει είναι σε: Ορεκτικό (Appetizer), Κυρίως Πιάτο (Main Course), Δεύτερο Πιάτο (Second Course), Single Course, Συνοδευτικό (Side Dish), Επιδόρπιο (Dessert), Πρωινό (Breakfast) και Φρούτα (Fruit). Τέλος, δεν υπάρχουν σαφείς διευκρινίσεις για την αντιστοίχιση των εικόνων του συνόλου δεδομένων στις παραπάνω ευρύτερες κλάσεις.

Food-11K (συνδυάζεται με Food 5K): Αποτελεί ένα σχετικά καινούριο σύνολο δεδομένων φωτογραφιών αποτελούμενο από 16643 φωτογραφίες διακεκριμένες σε 11 κλάσεις (Ecole Polytechnique Fédérale de Lausanne (EPFL), n.d.). Οι κλάσεις αυτές είναι Γλυκό, Αυγά, Τηγανητό Φαγητό, Κρέας, Noodles-Ζυμαρικά, Θαλασσινό φαγητό, Σούπες και Λαχανικά-Φρούτα. Συνδυάζεται με το Food-5K, που αποτελεί σύνολο δεδομένων 5000 εικόνων, μισές εκ των οποίων απεικονίζουν φαγητά ενώ οι υπόλοιπες δεν σχετίζονται με φαγητά. Το σύνολο δεδομένων αυτό χρησιμοποιείται για την εκπαίδευση δυαδικού ταξινομητή για να διακρίνει τις φωτογραφίες που εμπεριέχουν φαγητό από τις υπόλοιπες, με σκοπό οι μη σχετικές φωτογραφίες με το φαγητό να φιλτράρονται και να μην μεταβαίνουν στο δεύτερο επίπεδο ταξινόμησης.

Diabetes: Αποτελεί σύνολο δεδομένων ειδικά διαμορφωμένο στις ανάγκες των ατόμων με ΣΔ. Το σύνολο δεδομένων αυτό, στην δημοσίευση που προτάθηκε (Anthimopoulos, Gianola, Scarnato, Diem, & Mougiakakou, 2014), δεν έχει επίσημη ονομασία, αλλά θα χρησιμοποιείται η ονομασία του Diabetes για να διακρίνεται από τα υπόλοιπα. Αποτελείται από 4868 φωτογραφίες οι οποίες οργανώνονται σε 11

κλάσεις: Ψωμί, Ψωμο-ειδή φαγητά (Breaded Food), Τυρί, Αυγά και προϊόντα, Όσπρια, Κρέας, Ζυμαρικά, Πίτσα, Πατάτες, Ρύζι και Λαχανικά.

Τα παραπάνω αποτελούν τα σημαντικότερα σύνολα δεδομένων που χρησιμοποιούνται στην διεθνή βιβλιογραφία. Έχουν προταθεί και άλλα σύνολα δεδομένων, τα οποία ωστόσο, είτε επειδή αφορούν παραλλαγές των παραπάνω ή επειδή δεν είναι δημοσίως διαθέσιμα, δεν συζητώνται εκτενέστερα στο πλαίσιο της τρέχουσας εργασίας.

5.2 Ανάγκη δημιουργίας νέου συνόλου δεδομένων

Σκοπός των παραπάνω συνόλων δεδομένων είναι να αποτελέσουν την βάση για την εφαρμογή του αλγορίθμου ταξινόμησης. Παρ' όλα αυτά δεν έχει γίνει εφικτό να εδραιωθεί καθολικά κάποιο από τα παραπάνω, προκειμένου να γίνεται άμεση σύγκριση των αποτελεσμάτων κάθε αλγορίθμου ταξινόμησης. Ως εκ τούτου, δημοσιεύονται διαφορετικές τεχνικές ταξινόμησης σε διαφορετικά σύνολα δεδομένων, δημιουργώντας έτσι σύγχυση σχετικά με τις επικρατούσες τεχνικές στον τομέα της αναγνώρισης φαγητών. Δυστυχώς, δεν προβλέπεται η σύγκλιση των συνόλων δεδομένων σε ένα, καθώς συνήθως διαφορετικά σύνολα δεδομένων, εξυπηρετούν διαφορετικούς σκοπούς.

Τα παραπάνω σύνολα δεδομένων αν και περιέχουν σημαντική πληροφορία, δεν χρησιμοποιήθηκαν σε αυτή την διπλωματική εργασία. Το Food 101, μετά την εδραίωση των ΣΝΔ στην ταξινόμηση εικόνων, ίσως αποτελεί το επικρατέστερο σύνολο δεδομένων. Παρ' όλα αυτά το μεγάλο πλήθος των εικόνων και των κατηγοριών που περιέχει, καθιστούσε απαραίτητη την πρόσβαση σε υπολογιστές με εξειδικευμένες κάρτες γραφικών, προκειμένου να γίνει η επεξεργασία των δεδομένων σε θεμιτούς υπολογιστικά χρόνους.

Στην συνέχεια το PFID αποτελεί σύνολο δεδομένων βασισμένο σε μη-υγιεινές διατροφικές συνήθειες, τις οποίες θα πρέπει να μην ακολουθούν οι διαβητικοί. Οπότε η δημιουργία ταξινομητή βασισμένο σε αυτό το σύνολο δεδομένων θα απέκλινε σημαντικά από το βασικό στόχο της διπλωματικής εργασίας, που αφορά άτομα με ΣΔ.

Τα UEC-FOOD 100 και 256, αν και αποτελούν σύνολα δεδομένων με ικανοποιητικό αριθμό εικόνων και κατηγοριών, έχουν στην πλειονότητα τους εικόνες από ξένες κουζίνες. Από την άλλη μεριά, τα σύνολα δεδομένων των UNICT-FD, αποτελούνται από ελλιπή αριθμό εικόνων για κάθε φαγητό. Παράλληλα δεν υπάρχει σαφής διαχωρισμός του συνόλου δεδομένων σε κατηγορίες, προκειμένου να υλοποιηθεί ταξινόμηση υψηλότερου επιπέδου.

Όσον αφορά το σύνολο δεδομένων του Food-11K δεν έχει χρησιμοποιηθεί ιδιαίτερα στην βιβλιογραφία. Παράλληλα οι 11 κλάσεις στις οποίες διαχωρίζεται δεν είναι απόλυτα σύμφωνες με τα

ισοδύναμα υδατανθράκων των διαβητικών. Ακόμα, η διαδικασία ταξινόμησης σε φαγητά και όχι ομάδες φαγητών απαιτεί επιπρόσθετη επεξεργασία των εικόνων, καθώς δεν αναγράφεται σε αυτές το φαγητό που απεικονίζουν.

Το UNIMIB 2016 απεικονίζει πολλά φαγητά σε κάθε εικόνα, μη επιτρέποντας την ταύτιση κάθε εικόνας με μια κατηγορία, αλλά καθιστώντας τον αλγόριθμο της κατάτμησης (segmentation) απαραίτητο. Τα φαγητά του UNIMIB-2016 προέρχονται από καντίνια σε πραγματικές συνθήκες, με αποτέλεσμα να μην συνίστανται για διαβητικούς. Επιπλέον η χρήση της Ιταλικής ορολογίας, προσθέτει ακόμα μια τεχνική δυσκολία.

Τέλος το σύνολο δεδομένων του Diabetes, παρά το γεγονός ότι ικανοποιεί τις ανάγκες της διπλωματικής εργασίας, δεν είναι ελεύθερα διαθέσιμο. Συνεπώς βγαίνει το συμπέρασμα ότι θα ήταν πολύ χρήσιμο να δημιουργηθεί ένα σύνολο δεδομένων σύμφωνα με τις ανάγκες της διπλωματικής εργασίας.

5.3 Σύνολο Δεδομένων Sarantea-2016

Αξίζει να γίνει αναφορά στο αδημοσίευτο σύνολο δεδομένων Sarantea-2016. Το Sarantea-2016 σύνολο δεδομένων δημιουργήθηκε στα πλαίσια διπλωματικής εργασίας με τίτλο «Ανάπτυξη Μοντέλου Εκτίμησης της Περιεχόμενης Ποσότητας Υδατανθράκων στα Λαμβανόμενα Γεύματα από Φωτογραφικά Στιγμιότυπα» (Σαραντέα-Μαγγαλούση, 2016). Η διπλωματική εργασία αυτή αποσκοπεί στην ταξινόμηση εικόνων σε 6 μεγάλες κλάσεις Ψωμί, Κρέας, Ζυμαρικά, Πατάτες, Ρύζι και Λαχανικά. Κάθε κατηγορία αποτελείται από 200 φωτογραφίες, με ένα φαγητό σε κάθε μια από αυτές.

Σημαντικό πλεονέκτημα του συνόλου δεδομένων αυτού, είναι ότι οι φωτογραφίες αυτές χειροκίνητα έχουν διαχωριστεί από το υπόβαθρο τους, με αποτέλεσμα σε κάθε εικόνα να απεικονίζεται ένα φαγητό με λευκό υπόβαθρο, χωρίς επιπρόσθετο θόρυβο. Έτσι γίνεται μια προσομοίωση του αλγορίθμου κατάτμησης, που κανονικά θα έπρεπε να έχει αναπτυχθεί σε μια ολοκληρωμένη εφαρμογή, ώστε να διαχωρίζει τα διαφορετικά φαγητά στην εικόνα, αλλά και να αφαιρεί τον περιττό θόρυβο που υπάρχει σε αυτή. Με τον τρόπο αυτό ελέγχεται ο αλγόριθμος εξαγωγής χαρακτηριστικών και ταξινόμησης, χωρίς τα τελικά αποτελέσματα να επηρεάζονται από άλλα στοιχεία της φωτογραφίας του γεύματος.

Επιπρόσθετα, σημαντικό πλεονέκτημα του συνόλου δεδομένων αυτού αποτελεί ο τρόπος διάκρισης των φαγητών σε κατηγορίες, που αν και δεν προσομοιώνει το πλήρες φάσμα των διατροφικών συνηθειών, αποτελεί μια αντιπροσωπευτική διάκριση των τροφών σε Κρέας, Ζυμαρικά, Λαχανικά και Ψωμί. Οι κατηγορίες αυτές αποτελούν σημαντικές διατροφικές ομάδες για όλους τους ανθρώπους και λόγω της διαφορετικής περιεκτικότητας τους σε υδατάνθρακες, αποτελούν μια ενδεικτική πρώτη ταξινόμηση των

φαγητών στις ανάγκες των ατόμων με ΣΔ. Τέλος το περιορισμένο πλήθος εικόνων ανά κατηγορία, καθώς και το γεγονός ότι κάθε κατηγορία αποτελείται από ίσο πλήθος εικόνων, καθιστά το σύνολο δεδομένων εύκολα επεξεργάσιμο.

Ως εκ τούτου, το συγκεκριμένο σύνολο δεδομένων ικανοποιεί τις απαιτήσεις που έχουν τεθεί για τις ανάγκες των ατόμων με ΣΔ και μπορεί να προσπελαστεί σχετικά γρήγορα, δίνοντας έτσι ενδεικτικές κατευθυντήριες για την ανάπτυξη ταξινομητή τροφών από φωτογραφικά στιγμιότυπα για άτομα με ΣΔ.

5.4 Υπολογισμός δοσολογίας ινσουλίνης

5.4.1 Ισοδύναμα Υδατάνθρακα

Ένα άτομο με ΣΔ ακολουθεί τρία βήματα για τον υπολογισμό της δόσης ινσουλίνης που θα λάβει. Αυτά είναι: ο υπολογισμός της διόρθωσης του σακχάρου, ο υπολογισμός της δόσης του φαγητού και ο υπολογισμός της τελικής δόσης ινσουλίνης (Ελληνική Διαβητολογική Εταιρεία, 2013). Προτού αναλυθούν τα παραπάνω βήματα, συμβατικά έχει οριστεί ως ισοδύναμο υδατανθράκων η ποσότητα που περιέχει 15 γραμμάρια υδατανθράκων, όπως ένα φλιτζάνι γάλα ή ένα μικρό μήλο ή μια φέτα ψωμί.

Κάθε άτομο με ΣΔ χαρακτηρίζεται από έναν μεταβαλλόμενο δείκτη ευαισθησίας στην ινσουλίνη, δηλαδή την επίδραση που έχει μια μονάδα ινσουλίνης στο σάκχαρο του. Καθώς ακόμα γνωρίζει για κάθε ισοδύναμο υδατανθράκων που καταναλώνει την αντιστοίχισή του σε ποσότητα ινσουλίνης, προκειμένου να μείνει το σάκχαρο του αμετάβλητο. Διατυπώνεται παρακάτω ένα ενδεικτικό παράδειγμα υπολογισμού της κατάλληλης ποσότητας ινσουλίνης για άτομο με σάκχαρο 150mg/dl, ευαισθησία 40 και αντιστοίχιση 1,5 μονάδας ινσουλίνης ανά ισοδύναμο υδατανθράκων. Το γεύμα που καταναλώνει το άτομο αυτό αποτελείται από 78 γραμμάρια υδατανθράκων. Θεωρείται ότι η φυσιολογική τιμή του σακχάρου είναι 100mg/dl στα υγιή άτομα.

5.4.1.1 Υπολογισμός διόρθωσης του σακχάρου

Η τιμή του σακχάρου θα πρέπει να μειωθεί από 150mg/dl στα 100mg/dl. Δεδομένου ότι για το άτομο που μελετάται μια μονάδα ινσουλίνης μειώνει το σάκχαρο κατά 40 mg/dl, αυτό σημαίνει ότι θα χρειαστεί $\frac{50}{40} = 1,25$ μονάδες ινσουλίνης. Η δόση αυτή δεν σχετίζεται με το γεύμα που θα καταναλωθεί, αλλά πρόκειται για διορθωτική δοσολογία.

5.4.1.2 Υπολογισμός της δόσης του φαγητού

Κάθε φαγητό αντιστοιχίζεται με κάποιο πλήθος ισοδυνάμων υδατάνθρακα. Στην συνέχεια του κεφαλαίου αυτού θα γίνει η παρουσίαση και ομαδοποίηση των φαγητών σύμφωνα με την περιεκτικότητα τους σε υδατάνθρακες. Ενδεικτικά, στο συγκεκριμένο παράδειγμα θεωρούμε ότι ο χρήστης έχει υπολογίσει ότι το γεύμα του αποτελείται από 78 γραμμάρια υδατανθράκων. Τα 78 γραμμάρια υδατανθράκων αντιστοιχίζονται σε $\frac{78}{15} = 5,2$ ισοδύναμα υδατανθράκων. Αν για κάθε ισοδύναμο υδατανθράκων, ο χρήστης χρειάζεται 1,5 μονάδα ινσουλίνης, τότε συνολικά για όλο το γεύμα θέλει $5,2 \times 1,5 = 7,8$ μονάδες ινσουλίνης.

5.4.1.3 Υπολογισμός της τελικής δόσης της ινσουλίνης

Όπως φάνηκε από τα παραπάνω, η τελική δοσολογία της ινσουλίνης εξαρτάται από την προγευματική ένδειξη του σακχάρου, σε συνδυασμό με την περιεκτικότητα του γεύματος σε υδατάνθρακες. Στο συγκεκριμένο παράδειγμα χρειάζονται 1,25 μονάδες ως διορθωτική δοσολογία στο σάκχαρο του ασθενή και επιπρόσθετα 7,8 μονάδες εξαιτίας του γεύματος. Συνολικά ο ασθενής χρειάζεται $1,25 + 7,8 = 9,05$ μονάδες. Από εκεί και πέρα μια σημαντική παράμετρος είναι η μεταγευματική δραστηριότητα του ασθενή. Αν μετά το γεύμα ο χρήστης ακολουθήσει στατική ζωή, τότε θα στρογγυλέψει την δοσολογία στις 9 μονάδες, ενώ στην περίπτωση που θα ακολουθήσει σωματική άσκηση μπορεί να χρειαστούν 7 ή 8 μονάδες. Γενικότερα η στρογγυλοποίηση γίνεται προς τα κάτω, για να αποφεύγονται φαινόμενα υπογλυκαιμίας.

Μέσω του παραπάνω παραδείγματος, αναλύθηκε ο τρόπος υπολογισμού της συνολικής ποσότητας ινσουλίνης. Σημαντική παράμετρος αυτής της διαδικασίας αποτελεί η περιεκτικότητα σε υδατάνθρακες κάθε φαγητού, που είναι και το κριτήριο οργάνωσης του προτεινόμενου συνόλου δεδομένων.

5.4.2 Ισοδύναμα Λίπους και Πρωτεΐνης

Είναι κοινώς αποδεκτή μέθοδος ο υπολογισμός της δοσολογίας της ινσουλίνης σύμφωνα με τα ισοδύναμα των υδατανθράκων που βρίσκονται στο φαγητό. Νεότερες μελέτες όμως δείχνουν ότι άτομα που καταναλώνουν σημαντικές ποσότητες λίπους στο γεύμα τους, εμφανίζουν υψηλό σάκχαρο μετά από πολλές ώρες (Ελληνική Διαβητολογική Εταιρεία, 2013). Καθώς επίσης ότι μετά από την κατανάλωση κάποιας ποσότητας πρωτεΐνης, υπάρχουν επιδράσεις στο μεταγευματικό σάκχαρο. Έχει καθιερωθεί η εκδοχή ότι για κάθε 120 γραμμάρια κρέατος να προστίθεται μια μονάδα ινσουλίνης, καθώς επίσης για κάθε 15 γραμμάρια λίπους.

Σύμφωνα με τα παραπάνω δημιουργείται η ανάγκη προσδιορισμού του ισοδύναμου πρωτεΐνης και λίπους. Ένα ισοδύναμο πρωτεΐνης είναι η ποσότητα τροφής που εμπεριέχει 7 γραμμάρια πρωτεΐνης και

βρίσκεται σε 30 γραμμάρια κρέατος ή ψαριού. Αντίστοιχα, ένα ισοδύναμο λίπους προσδιορίζεται ως η ποσότητα της τροφής που περιέχει 5 γραμμάρια λίπους και βρίσκεται για παράδειγμα σε μια κουταλιά του γλυκού λαδιού.

1 Ισοδύναμο	Αντιστοιχεί σε:	Βρίσκεται σε:
Πρωτεΐνης	7 gr πρωτεΐνης	30gr κρέας/ψάρι/κοτόπουλο
Λίπους	5gr λίπους	1 κουταλιά του γλυκού λάδι

Πίνακας 5.1. Ισοδύναμο πρωτεΐνης και λίπους

Τα ισοδύναμα των πρωτεϊνών και του λίπους δεν προστίθενται σε αυτά των υδατανθράκων για τον συνολικό υπολογισμό της ποσότητας της ινσουλίνης. Τα ισοδύναμα των υδατανθράκων υπολογίζονται ξεχωριστά και επεξεργάζονται με τον τρόπο που αναλύθηκε παραπάνω. Ενώ για κάθε 4 ισοδύναμα πρωτεΐνης στην περίπτωση του κρέατος (120 γραμμάρια), 5 ισοδύναμα πρωτεΐνης στην περίπτωση του ψαριού (150 γραμμάρια) και για κάθε 3 ισοδύναμα λίπους (1 κουταλιά της σούπας λάδι – 15 ml) προστίθεται ένα ισοδύναμο υδατανθράκων.

	1 Ισοδύναμο Υδατάνθρακα
Ισοδύναμα Πρωτεΐνης	4 (Κρέας)
	5 (Ψάρι)
Ισοδύναμα Λίπους	3

Πίνακας 5.2. Αντιστοίχιση των ισοδυνάμων πρωτεΐνης και λίπους με ένα ισοδύναμο υδατάνθρακα

5.5 Ισοδύναμα Τροφών

Η Ελληνική Διαβητολογική Εταιρεία, σύμφωνα με την περιεκτικότητα των τροφών σε υδατάνθρακες, λίπος και πρωτεΐνες, διακρίνει τα φαγητά σε 6 μεγάλες κατηγορίες. Οι κατηγορίες αυτές είναι Γάλα και Προϊόντα, Λαχανικά, Φρούτα και Χυμοί, Ψωμί-Δημητριακά-Όσπρια-Αμυλούχα Λαχανικά, Κρέας και Υποκατάστατα, Λίπος. Η κάθε κατηγορία χαρακτηρίζεται από το αντίστοιχο ισοδύναμο, που περιγράφει την περιεκτικότητα των φαγητών κάθε ομάδας σε υδατάνθρακες, πρωτεΐνη και λίπος ανά στοιχειώδεις ποσότητες.

Ισοδύναμα	Υδατάνθρακες (γρ.)	Πρωτεΐνη (γρ.)	Λίπος (γρ.)	Θερμίδες	Φυτικές ίνες (γρ.)	Προϊόντα
Ισοδύναμο Γάλακτος	12	8				Γάλα και προϊόντα
Ισοδύναμο Λαχανικών	5	2		25	1 με 4	Λαχανικά
Ισοδύναμο Φρούτων και Χυμών	15			60		Φρούτα και χυμοί
Ισοδύναμο Αμυλούχων	15	3	0 με 1	80		Ψωμι-Δημητριακά - Όσπρια - Αμυλούχα λαχανικά
Ισοδύναμο Κρέατος (χ.π.λ)		7	3	55		Κρέας και υποκατάστατα
(μ.π.λ)		7	5	75		
(υ.π.λ)		7	8	100		

Πίνακας 5.3. Ισοδύναμα των βασικών κατηγοριών τροφών. Οι συντομογραφίες (χ.π.λ), (μ.π.λ), (υ.π.λ) αντιστοιχούν σε χαμηλή, μέτρια και υψηλή περιεκτικότητα σε λίπος.

5.6 Νέο Σύνολο Δεδομένων – NTUA Food 2017

Στο πλαίσιο της τρέχουσας διπλωματικής εργασίας προτείνεται ένα νέο σύνολο δεδομένων, που αποσκοπεί στην προσέγγιση των διατροφικών συνηθειών του διαβητικού. Πρόκειται για ένα σύνολο δεδομένων με πληθώρα υγιεινών φαγητών, που κατά κανόνα αποτελούν τον κορμό της διατροφής των ατόμων με ΣΔ. Ωστόσο, συμπεριλαμβάνονται κάποια από τα πιο γνωστά φαγητά πλούσια σε λίπος και ζάχαρη ώστε να αποτελέσουν μια κατηγορία επιβλαβών τροφών που συνίσταται να αποφεύγουν τα άτομα με ΣΔ. Το σύνολο δεδομένων αυτό δημιουργήθηκε σύμφωνα με τον «Οδηγό Διατροφής για την ρύθμιση του Διαβήτη», που εκδόθηκε το 2013 από την Ελληνική Διαβητολογική Εταιρεία (ΕΔΕ) και αποτελείται από 8 ομάδες τροφών (Ελληνική Διαβητολογική Εταιρεία, 2013).

Η πρώτη ομάδα αποτελείται από γαλακτοκομικά προϊόντα, με τα σημαντικότερα από αυτά να είναι το γάλα και το γιαούρτι. Ένα φλιτζάνι γάλακτος χαμηλών λιπαρών, καθώς και 180γρ. γιαούρτι με 2% λίπος, περιέχουν 12γρ. υδατάνθρακες και 8γρ. πρωτεΐνη. Συνεπώς η κατηγορία Γάλα και Προϊόντα προσδιορίζεται από σχεδόν ένα ισοδύναμο υδατανθράκων και μια σημαντική ποσότητα πρωτεΐνης. Τα τυριά αν και γαλακτοκομικά προϊόντα, δεν έχουν την ίδια περιεκτικότητα σε μακροθρεπτικά συστατικά με το γάλα και το γιαούρτι. Παρ' όλα αυτά επειδή είναι πλούσια σε λίπος, προκαλούν αντίστοιχες μεταβολές στο σάκχαρο του ατόμου με ΣΔ. Τα κανονικά τυριά ανά 30γρ., κατά μέσο όρο, περιέχουν 7γρ. πρωτεΐνης και 8γρ. λίπους. Η κατηγορία των τυριών που χρησιμοποιήθηκε αποτελείται από γραβιέρα, ένταμ, κασέρι, τσένταρ, φέτα, παρμεζάνα και μοτσαρέλα.

Η επόμενη βασική ομάδα είναι τα λαχανικά. Ένα ισοδύναμο λαχανικών αντιστοιχεί σε μισό φλιτζάνι του τσαγιού βρασμένα λαχανικά ή ένα φλιτζάνι ωμά λαχανικά (100γρ.) και περιέχει 5γρ. υδατάνθρακες, 2γρ. πρωτεΐνες και 1-4γρ. φυτικές ίνες. Οι φυτικές ίνες αν και δεν επιδρούν στην μεταβολή του σακχάρου, επιτρέπουν την ομαλή απορρόφηση των μακροθρεπτικών συστατικών από τον οργανισμό, με αποτέλεσμα να μην προκαλείται απότομη αύξηση του σακχάρου. Άρα τα λαχανικά αποτελούν μια κατηγορία που δεν έχει έντονη επίδραση στο σάκχαρο του διαβητικού. Τα αντιπροσωπευτικότερα λαχανικά που επιλέχθηκαν για αυτή την κατηγορία είναι τα αγγούρι, βλίτα, καρότα, κολοκυθάκια,

κουνουπίδι, λάχανο, μελιτζάνες, μπρόκολο, παντζάρια και ντομάτες. Στην κατηγορία αυτή επίσης εμπεριέχονται φωτογραφίες από σαλάτες.

Η επόμενη κατηγορία είναι τα φρούτα, που περιλαμβάνουν: ανανά, αχλάδι, βερίκοκα, καρπούζι, κεράσια, μήλο, μπανάνα, πεπόνι, πορτοκάλι, ροδάκινο, σταφύλια, φράουλες καθώς και πιατέλες με φρούτα. Ένα ισοδύναμο φρούτων αποτελείται από 15γρ. υδατάνθρακες. Η ποσότητα στην οποία εντοπίζεται το ισοδύναμο του φρούτου, διαφέρει από περίπτωση σε περίπτωση, καθώς 15γρ. υδατάνθρακα μπορούν να εντοπιστούν τόσο σε ένα μικρό μήλο ή αχλάδι, όσο και σε μια φέτα καρπουζιού και σε 4 ολόκληρα βερίκοκα. Ανεξαρτήτως την περίπτωση, παρατηρείται ότι τα φρούτα είναι πλούσια σε υδατάνθρακες, με αποτέλεσμα συνήθως να αντιστοιχούν σε ένα ισοδύναμο υδατάνθρακα, που συνήθως μεταφράζεται σε μια μονάδα ινσουλίνης.

Στην συνέχεια ορίζεται μια ιδιαίτερα ευρεία κατηγορία, αυτή των Ψωμί-Δημητριακά-Όσπρια-Αμυλούχα λαχανικά. Ένα ισοδύναμο αμυλούχων, όπως θα αποκαλείται η παραπάνω κατηγορία, περιέχει 15γρ. υδατάνθρακες, 3γρ. πρωτεΐνες και ελάχιστο λίπος. Καθώς η κατηγορία αυτή είναι πολυπληθής επιλέγεται να διασπαστεί σε δύο επιμέρους. Η πρώτη κατηγορία είναι τα Όσπρια και Αμυλούχα λαχανικά, ενώ η δεύτερη αποτελείται από Ψωμί και Δημητριακά. Αρχικά η διαφοροποίηση προέκυψε από το γεγονός ότι σε αντίθεση με τις υπόλοιπες υποομάδες των αμυλούχων, τα όσπρια είναι πλούσια σε πρωτεΐνη, περιέχοντας 7γρ. ανά μισή κούπα. Το συμπληρωματικό γεγονός που ολοκληρώνει τον διαχωρισμό αυτό, είναι η χαμηλότερη περιεκτικότητα σε υδατάνθρακες των οσπρίων και αμυλούχων λαχανικών, συγκριτικά με το Ψωμί και το Ρύζι. Αυτό συμβαίνει καθώς, το ένα ισοδύναμο αμυλούχων στην πρώτη περίπτωση βρίσκεται σε ποσότητα που αντιστοιχεί στο $\frac{1}{3}$ του φλιτζανιού, ενώ στην δεύτερη περίπτωση σε ποσότητα ίση με το $\frac{1}{2}$ του φλιτζανιού. Παράλληλα διαμορφώνονται δυο κλάσεις με παρόμοιο πλήθος τροφίμων. Τα τρόφιμα κάθε κατηγορίας αναγράφονται στο Πίνακα 5.4.

Η τελευταία κατηγορία του νέου συνόλου δεδομένων με τον διαχωρισμό φαγητών που προτείνεται από την Ελληνική Διαβητολογική Εταιρεία, είναι το Κρέας και Υποκατάστατα. Το ισοδύναμο πρωτεϊνών, που αντιπροσωπεύει αυτή την κατηγορία, αποτελείται βασικά από 7γρ. πρωτεΐνης. Παρ' όλα αυτά εμπεριέχει φαγητά με σημαντικές διαφοροποιήσεις ως προς το λίπος που περιέχουν. Αρχικά, ορίζονται τα τρόφιμα της κατηγορίας αυτής, με χαμηλή περιεκτικότητα σε λιπαρά(χ.π.λ.), όπως κάποια από τα μέρη του χοιρινού και μοσχαρίσιου κρέατος, το κοτόπουλο και τα ψάρια. Τα κρέατα αυτά περιέχουν 3γρ. λίπους ανά 30γρ. κρέατος. Στην κατηγορία κρεάτων μέσης περιεκτικότητας σε λίπος(μ.π.λ), ανήκουν τα αυγά, το αρνί, ο κιμάς καθώς και οι χοιρινές και μοσχαρίσιες μπριζόλες. Η υποκατηγορία αυτή εμπεριέχει 5 γρ. λίπους ανά 30γρ. κρέατος. Τέλος στην κατηγορία των λιπαρών κρεάτων εμπεριέχονται τα φαγητά με 8γρ. λίπους ανά 30γρ. Χαρακτηριστικός αντιπρόσωπος της υποομάδας αυτής είναι το λουκάνικο. Παρατηρείται ότι το

κρέας δεν επιδρά ιδιαίτερα στο επίπεδο του σακχάρου, παρά μόνο σε πολύ μεγάλες ποσότητες ή όταν είναι πλούσιο σε λίπος.

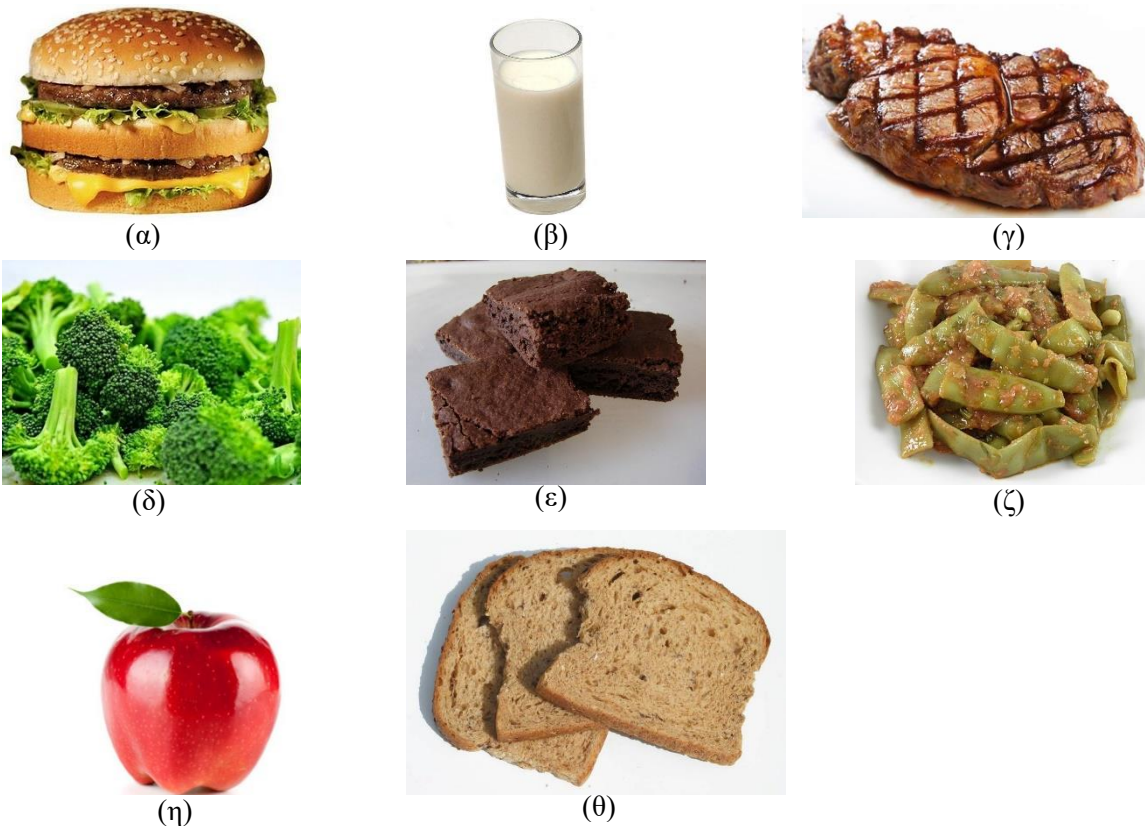
Οι παραπάνω κατηγορίες, αν και διαφέρουν στα μακροθρεπτικά συστατικά που περιέχουν, περιλαμβάνουν θρεπτικά τρόφιμα που ενδείκνυνται να καταναλωθούν από άτομα με ΣΔ. Οπότε επιτελούν το σκοπό του να προσομοιάζουν τις προτεινόμενες διατροφικές συνήθειες των διαβητικών. Παρ' όλα αυτά ο σκοπός του συνόλου δεδομένων είναι διττός, καθώς αποσκοπεί επιπλέον στην διάκριση ανάμεσα σε προτεινόμενες υγιείς τροφές και σε ανθυγιεινά τρόφιμα. Για το λόγο αυτό, εισάγονται οι ομάδες των λιπαρών φαγητών και των Breaded Food.

Η κατηγορία των Breaded food, αποτελείται από φαγητά τα οποία έχουν ως βάση το ψωμί ή άλλα αμυλούχα παράγωγα, όπως η ζύμη. Ουσιαστικά στην ομάδα αυτή επιδιώχθηκε να ενταχθούν τα πιο αντιπροσωπευτικά fast foods, που έχουν ως βάση το ψωμί. Η δημιουργία της κατηγορίας αυτής προέκυψε εξαιτίας δύο λόγων. Αρχικά καθώς η κατανάλωση τέτοιων φαγητών αποτελεί αναπόσπαστο κομμάτι των εβδομαδιαίων διατροφικών συνηθειών των περισσότερων ανθρώπων. Επίσης αμυλούχα παράγωγα όπως το ψωμί του χοτ ντογκ, η κρέπα και το ψωμί του χάμπουργκερ που υπάγονται στην κατηγορία του Ψωμιού είναι σπάνιο να βρεθούν μόνα τους προς κατανάλωση. Για το λόγο αυτό επιλέχθηκε να συμπεριληφθούν στο σύνολο δεδομένων, αλλά ως συστατικά ενός φαγητού.

Συγκεκριμένα τα φαγητά που ανήκουν στην κατηγορία αυτή είναι η αραβική πίτα, το χάμπουργκερ, το χοτ-ντογκ, το τοστ, η κρέπα, το σουβλάκι τυλιχτό, η πίτσα και τα σάντουιτς. Τα φαγητά αυτά αν και δεν περιέχουν ακριβώς τα ίδια μακροθρεπτικά συστατικά, έχουν μια κοινή δομή. Από τη μια μεριά είναι πλούσια σε άμυλο, καθώς είναι πλούσια σε ψωμί όπως στην περίπτωση του χάμπουργκερ, του χοτ-ντόγκ και του σάντουιτς ή σε μορφή πίτας στα σουβλάκια ή ακόμα σε μορφή ζύμης για τη πίτσα. Παράλληλα είναι και ιδιαίτερα λιπαρά φαγητά, λόγω του τρόπου παρασκευής τους, του λαδιού και του τυριού που χρησιμοποιείται. Τα παραπάνω χαρακτηριστικά αποτελούν την κοινή βάση των φαγητών αυτών.

Τέλος ορίζεται η κατηγορία των λιπαρών φαγητών, η οποία ουσιαστικά αντιπροσωπεύει τα επιδόρπια. Τα επιδόρπια αποτελούν φαγητά πλούσια σε λίπος και υδατάνθρακες, έχοντας σοβαρές επιπτώσεις στις μεταγευματικές τιμές του σακχάρου του ανθρώπου. Στη κατηγορία αυτή εμπεριέχονται τα κρουασάν, brownies, μηλόπιτα, μπακλαβάς, κέικ σοκολάτας, donuts, παγωτό, πανακότα και το τιραμισού. Τα γλυκά αυτά θεωρήθηκε ότι αποτελούν κάποια από τα πιο γνωστά γλυκά, αποτελώντας έτσι ένα αντιπροσωπευτικό δείγμα των επιδορπίων. Στην ομάδα αυτή θεωρήθηκε σκόπιμο να προστεθούν και οι πατάτες τηγανητές. Σε καμιά περίπτωση δεν αποτελούν κάποιο είδος επιδορπίου, αλλά αποτελεί ένα φαγητό που δεν θα μπορούσε να λείπει από κανένα σύνολο δεδομένων, είναι πλούσιο σε λίπος και δεν ικανοποιεί την δομή των breaded food. Με το σκεπτικό αυτό, θεωρήθηκε σκόπιμο να προστεθεί σε μια κατηγορία φαγητών ακατάλληλων για τα άτομα με το ΣΔ και ως καταλληλότερη επιλέχθηκε η κατηγορία

αυτή. Ενδεικτικά μια φωτογραφία από κάθε κατηγορία του συνόλου δεδομένων απεικονίζεται στην Εικόνα 5.2.



Εικόνα 5.2. Φωτογραφία από κάθε κατηγορία του συνόλου δεδομένων (α) Hamburger (Breaded Food), (β) Γάλα (Γαλακτοκομικά και Προϊόντα) (γ) Μπριζόλα (Κρέας και Υποκατάστατα), (δ) Μπρόκολο (Λαχανικά) (ε) Brownie (Λιπαρά Φαγητά) (ζ) Φασολάκια (Οσπρια και Αμυλούχα Λαχανικά) (η) Μήλο (Φρούτα) (θ) Ψωμί Ολικής Αλέσεως (Ψωμί- Δημητριακά)

Ο τρόπος συλλογής των εικόνων προήλθε από διαφορετικές πηγές, με την βασική να είναι τα αποτελέσματα αναζήτησης εικόνων της μηχανής Google. Πέρα όμως από αυτό, χρησιμοποιήθηκαν κάποιες φωτογραφίες από τα ήδη δημοσιευμένα σύνολα δεδομένων Food-101 και UEC-FOOD256 (Bossard, Guillaumin, & Van Gool, 2014; Kawano & Yanai, 2014). Τέλος χρησιμοποιήθηκε ο ιστότοπος αναζήτησης φαγητών foodspotting, λόγω της υψηλής ευκρίνειας εικόνων που παρείχε (Foodspotting, 2017).

Αναλυτικά η δομή του συνόλου δεδομένων που προτείνεται παρουσιάζεται στον Πίνακα 5.4.

Γάλα και Προϊόντα
Γάλα
Γιαούρτι
Τυριά
Γραβιέρα
Ένταμ
Κασέρι
Τσένταρ
Φέτα
Παρμεζάνα
Μοτσαρέλα

Λαχανικά
Αγγούρι
Βλίτα
Καρότα
Κολοκυθάκια
Κουνουπίδι
Λάχανο
Μελιτζάνες
Μπρόκολο
Ντομάτες
Παντζάρια
Σαλάτες

Φρούτα
Ανανάς
Αχλάδι
Βερίκοκα
Καρπούζι
Κεράσια
Μήλο
Μπανάνα
Πεπόνι
Πορτοκάλι
Ροδάκινο
Σταφύλια
Φράουλες
Πιατέλες με φρούτα

Όσπρια και Αμυλούχα Λαχανικά
Όσπρια
Φάβα
Ρεβίθια
Φασολάκια
Φακές
Φασολάδα
Αμυλούχα λαχανικά
Αρακάς
Γλυκοπατάτα
Καλαμπόκι Βραστό
Καλαμπόκι Ολόκληρο
Κάστανα
Πατάτα ψητή με την φλούδα
Πατάτα βραστή
Πατάτας Πουρές

Κρέας και Υποκατάστατα
Αρνί
Αυγό Βραστό
Ζαμπόν
Κοτόπουλα Υπόλοιπα
Κοτόπουλο Φιλέτο
Κρέας Κοκκινιστό
Κρέας Χοιρινό
Λουκάνικο
Μοσχαρίσιο Κρέας
Μπιφτέκι
Μπριζόλα
Ομελέτα
Φιλέτο
Ψάρι

Breaded Food
Sandwich
Αραβική Πίτα
Κρέπα
Πίτσα
Σουβλάκι Τυλιχτό
Τοστ
Χάμπουργκερ
Χοτ ντόγκ

Ψωμί - Δημητριακά
Ψωμί
Κουλούρι Θεσσαλονίκης
Κριτσίνια
Παξιμάδια - Φρυγανιές
Ψωμί Λευκό
Ψωμί Ολικής Αλέσεως
Δημητριακά - Ρύζι
Corn Flakes
Βρώμη
Κριθαράκι
Μακαρόνια
Ρύζι

Λιπαρά Φαγητά
Brownie
Cupcakes
Donuts
Βάφλα
Κέικ Σοκολάτας
Κρουασαν
Μηλόπιτα
Μπακλαβάς
Παγωτό
Πανακότα
Πατάτες τηγανητές
Τιραμισού

Πίνακας 5.4. Κατηγοριοποιημένα τα φαγητά που αποτελούν το νέο σύνολο δεδομένων στις 8 βασικές ομάδες: Γάλα και Προϊόντα, Λαχανικά, Φρούτα, Οσπρια και Αμυλόχυα Λαχανικά, Κρέας και Υποκατάστατα, Breaded Food, Ψωμί-Δημητριακά, Λιπαρά Φαγητά

6 Ανάπτυξη Συστήματος Αυτόματης Αναγνώρισης Τροφής

Μέχρι στιγμής έχουν παρουσιαστεί τα σημαντικότερα σύνολα δεδομένων στον τομέα της ταξινόμησης φαγητών, έχει γίνει ανάλυση των μεθόδων εξαγωγής χαρακτηριστικών από φωτογραφίες, καθώς και των μεθόδων ταξινόμησης. Επιπλέον έχει γίνει μια εκτενής βιβλιογραφική ανασκόπηση, στην οποία παρουσιάζονται τα αποτελέσματα των τεχνικών αυτών σε διαφορετικά σύνολα δεδομένων και με διαφορετικές παραμέτρους. Λαμβάνοντας υπόψιν όλα τα παραπάνω αναπτύσσεται ταξινομητής αυτόματης αναγνώρισης τροφής.

6.1 Ταξινόμηση στο Σύνολο Δεδομένων Sarantea-2016

Όπως έχει αναφερθεί στο Κεφάλαιο 5, κανένα από τα διαθέσιμα σύνολα δεδομένων δεν ικανοποιούσε πλήρως τις προδιαγραφές της παρούσας διπλωματικής εργασίας, είτε επειδή περιείχαν υπερβολικά μεγάλο πλήθος εικόνων ή οι εικόνες ήταν νοηματικά εκτός του πλαισίου της εργασίας ή εμφάνιζαν κάποια μη ικανοποιητικά ποιοτικά χαρακτηριστικά. Για τους παραπάνω λόγους, ως σύνολο δεδομένων αναφοράς χρησιμοποιείται το αδημοσίευτο Sarantea-2016.

Στην διπλωματική εργασία στην οποία προτάθηκε το σύνολο δεδομένων Sarantea-2016, πραγματοποιήθηκε σύγκριση των μεθόδων εξαγωγής χαρακτηριστικών SURF, Color Extractor και LBP συναρτήσει οπτικού λεξιλογίου διαφορετικού μήκους κάθε φορά. Ως ταξινομητές χρησιμοποιήθηκαν μια βελτιστοποιημένη δομή της μεθόδου Directed Acyclic Graph (DAG) SVM και το SVM πολλών κλάσεων (Σαραντέα-Μαγγαλούση, 2016). Τα βέλτιστα αποτελέσματα επιτεύχθηκαν όταν έγινε η σύνθεση των χαρακτηριστικών του SURF, Color Extractor και LBP, χρησιμοποιώντας λεξικό 500 και 300 λέξεων αντιστοίχως για τους δύο πρώτους αλγορίθμους και εξαγωγή 59 χαρακτηριστικών από την μέθοδο LBP. Ως ταξινομητής χρησιμοποιήθηκε SVM πολλών κλάσεων (multiclass SVM) και η βέλτιστη ακρίβεια ταξινόμησης που επιτεύχθηκε είναι 86,92%. Η ακρίβεια αυτή θα αποτελέσει το σημείο αναφοράς για την ανάπτυξη αλγορίθμου ταξινόμησης κάνοντας χρήση ΣΝΔ, στο σύνολο δεδομένων αυτό.

6.1.1 Εξαγωγή χαρακτηριστικών από Συνελκτικό Νευρωνικό Δίκτυο

6.1.1.1 AlexNet

Η πρώτη μέθοδος ταξινόμησης που επιλέγεται είναι μέσω της εξαγωγής χαρακτηριστικών από ΣΝΔ, ακολουθούμενη από SVM ταξινομητή. Αρχικά φορτώνεται στο περιβάλλον MATLAB, το δίκτυο AlexNet, το οποίο έχει προ-εκπαιδευθεί με ένα υποσύνολο εικόνων της βάσης δεδομένων ImageNet. Η δομή του φαίνεται στην Εικόνα 6.1.

25x1 Layer array with layers:

1	'data'	Image Input	227x227x3 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	softmax
25	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

Εικόνα 6.1. Δομή του AlexNet

Το επόμενο βήμα είναι να φορτωθούν στο περιβάλλον MATLAB οι εικόνες του Sarantea-2016. Το γεγονός ότι όλες οι κατηγορίες του συνόλου δεδομένων είναι ισοπληθείς, αποτελούμενες από 200 εικόνες, επιτρέπει την εκμετάλλευση ολόκληρου του συνόλου δεδομένων για την εκπαίδευση και αξιολόγηση του ταξινομητή. Αναλυτικά διακρίνεται το σύνολο δεδομένων Sarantea-2016 σε σύνολο εκπαίδευσης και σύνολο ελέγχου με τυχαίο τρόπο, αποτελούμενο κάθε ένα από αυτά με το 60% και 40% του συνόλου δεδομένων αντιστοίχως. Δημιουργούνται έτσι δύο υποσύνολα του αρχικού συνόλου, με το πρώτο να εμπεριέχει 720 εικόνες, ενώ το δεύτερο 480.

Στην συνέχεια επιλέγεται για κάθε εικόνα του συνόλου εκπαίδευσης να επιστραφεί ο χάρτης ενεργοποίησης του «fc7» στρώματος, δηλαδή του προτελευταίου fully-connected layer του νευρωνικού δικτύου. Το επίπεδο αυτό αποτελείται από 4096 νευρώνες, με αποτέλεσμα να επιστρέφεται ένα διάνυσμα 4096 τιμών για την περιγραφή της κάθε εικόνας του συνόλου εκπαίδευσης. Με τον τρόπο αυτό, χρησιμοποιείται η προκαθορισμένη δομή του προ-εκπαιδευμένου δικτύου AlexNet, προκειμένου οι εικόνες εισόδου να μεταφραστούν σε ένα διάνυσμα 4096 χαρακτηριστικών. Τελικά, το σύνολο εκπαίδευσης περιγράφεται από ένα πίνακα διαστάσεων 4096x720.

Ακολούθως εκπαιδεύεται Μηχανή Διανυσμάτων Υποστήριξης με γραμμικό πυρήνα «ΓΤ» με τον πίνακα που εξήχθη στο προηγούμενο βήμα, χρησιμοποιώντας την τεχνική one-vs-all. Η ίδια διαδικασία που χρησιμοποιήθηκε για την εξαγωγή χαρακτηριστικών των εικόνων του συνόλου εκπαίδευσης,

χρησιμοποιείται για την περιγραφή των εικόνων του συνόλου ελέγχου, παράγοντας αυτή την φορά έναν πίνακα διαστάσεων 4096x480. Ο διδιάστατος πίνακας αυτός τροφοδοτεί τον ταξινομητή «ΓΤ», προκειμένου να ελεγχθεί η ακρίβεια της μεθόδου. Τα αποτελέσματα απεικονίζονται στον Πίνακα 6.1, σημειώνοντας μέση ακρίβεια για τις 6 κατηγορίες 93,96%

	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	92.50%	3.75%	1.25%	1.25%	1.25%	0.00%
Κρέας	8.75%	86.25%	0.00%	5.00%	0.00%	0.00%
Ζυμαρικά	1.25%	0.00%	96.25%	1.25%	0.00%	1.25%
Πατάτες	3.75%	1.25%	1.25%	93.75%	0.00%	0.00%
Ρύζι	1.25%	0.00%	0.00%	0.00%	98.75%	0.00%
Λαχανικά	0.00%	1.25%	2.50%	0.00%	0.00%	96.25%

Πίνακας 6.1. Μήτρα Σύγκρισης χρησιμοποιώντας εξαγωγή χαρακτηριστικών από το AlexNet

6.1.1.2 VGG-19net

Στην συνέχεια, η μέθοδος εξαγωγής χαρακτηριστικών υλοποιείται από το πιο βαθύ, μεταγενέστερο ΣΝΔ VGG-19net, η δομή του οποίου φαίνεται στην Εικόνα 6.2. Η μόνη σημαντική διαφοροποίηση που παρατηρείται στην επεξεργασία των εικόνων, πριν τροφοδοτήσουν τα δίκτυα Alexnet και VGG-19net, είναι ότι το πρώτο δέχεται σαν είσοδο εικόνες διαστάσεων 227x227x3, ενώ το δεύτερο 224x224x3. Ο διαχωρισμός των εικόνων σε σύνολο εκπαίδευσης και σύνολο ελέγχου ακολουθεί και αυτή την φορά τον διαχωρισμό 60%-40%. Η εξαγωγή των χαρακτηριστικών για κάθε εικόνα εισόδου, γίνεται και πάλι από το «fc7» fully-connected layer, επιστρέφοντας ένα διάνυσμα 4096 θέσεων για την περιγραφή της κάθε εικόνας.

1	'input'	Image Input	224x224x3 images with 'zerocenter' normalization
2	'conv1_1'	Convolution	64 3x3x3 convolutions with stride [1 1] and padding [1 1]
3	'relu_1'	ReLU	ReLU
4	'conv1_2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1]
5	'relu_2'	ReLU	ReLU
6	'pool1'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
7	'conv2_1'	Convolution	128 3x3x64 convolutions with stride [1 1] and padding [1 1]
8	'relu_2_1'	ReLU	ReLU
9	'conv2_2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1]
10	'relu_2_2'	ReLU	ReLU
11	'pool2'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
12	'conv3_1'	Convolution	256 3x3x128 convolutions with stride [1 1] and padding [1 1]
13	'relu_3_1'	ReLU	ReLU
14	'conv3_2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1]
15	'relu_3_2'	ReLU	ReLU
16	'conv3_3'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1]
17	'relu_3_3'	ReLU	ReLU
18	'conv3_4'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1]
19	'relu_3_4'	ReLU	ReLU
20	'pool3'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
21	'conv4_1'	Convolution	512 3x3x256 convolutions with stride [1 1] and padding [1 1]
22	'relu_4_1'	ReLU	ReLU
23	'conv4_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
24	'relu_4_2'	ReLU	ReLU
25	'conv4_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
26	'relu_4_3'	ReLU	ReLU
27	'conv4_4'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
28	'relu_4_4'	ReLU	ReLU
29	'pool4'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
30	'conv5_1'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
31	'relu_5_1'	ReLU	ReLU
32	'conv5_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
33	'relu_5_2'	ReLU	ReLU
34	'conv5_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
35	'relu_5_3'	ReLU	ReLU
36	'conv5_4'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1]
37	'relu_5_4'	ReLU	ReLU
38	'pool5'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0]
39	'fc6'	Fully Connected	4096 fully connected layer
40	'relu6'	ReLU	ReLU
41	'drop6'	Dropout	50% dropout
42	'fc7'	Fully Connected	4096 fully connected layer
43	'relu7'	ReLU	ReLU
44	'drop7'	Dropout	50% dropout
45	'fc8'	Fully Connected	1000 fully connected layer
46	'prob'	Softmax	softmax
47	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

Εικόνα 6.2. Δομή VGG-19 Net

Στην περίπτωση του VGG-19net, επιτυγχάνεται ακρίβεια 94,79%, ελαφρώς βελτιωμένη από το 93,96% του AlexNet. Το αποτέλεσμα είναι αναμενόμενο, καθώς το VGG-19net έχει πολύ μεγαλύτερο βάθος. Η μικρή αυτή βελτίωση στην ακρίβεια, δεν επιτρέπει την εδραίωση του VGG-19net ως βασικού ΣΝΔ, καθώς το πολύ μεγάλο βάθος του, συνεπάγεται χρονοβόρες διαδικασίες εξαγωγής χαρακτηριστικών και εκπαίδευσης του δικτύου. Η διαδικασία εκπαίδευσης του ΣΝΔ AlexNet για την εξαγωγή χαρακτηριστικών, διήρκεσε κοντά στα 30 λεπτά, ενώ η ίδια διαδικασία με την χρήση του VGGnet έφθασε

κοντά στις 6 ώρες. Αναλυτικά, η μήτρα σύγχυσης της συγκεκριμένης ταξινόμησης φαίνεται στον Πίνακα 6.2.

	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	97.50%	1.25%	0.00%	0.00%	1.25%	0.00%
Κρέας	10.00%	86.25%	0.00%	2.50%	1.25%	0.00%
Ζυμαρικά	0.00%	0.00%	98.75%	0.00%	1.25%	0.00%
Πατάτες	1.25%	1.25%	1.25%	95.00%	1.25%	0.00%
Ρύζι	1.25%	0.00%	1.25%	0.00%	95.00%	2.50%
Λαχανικά	0.00%	1.25%	2.50%	0.00%	0.00%	96.25%

Πίνακας 6.2. Μήτρα Σύγχυσης χρησιμοποιώντας εξαγωγή χαρακτηριστικών από το VGG-19 Net

6.1.2 Fine-Tuning

Η δεύτερη τεχνική ταξινόμησης βασισμένη στα ΣΝΔ, που χρησιμοποιήθηκε για την βελτίωση της ακρίβειας στο σύνολο δεδομένων Sarantea-2016, είναι αυτή του fine-tuning. Όπως στην αρχική περίπτωση, χρησιμοποιείται το προ-εκπαιδευμένο ΣΝΔ AlexNet και το σύνολο δεδομένων διακρίνεται σε δύο κατηγορίες, στο σύνολο εκπαίδευσης και στο σύνολο ελέγχου.

Στην μέθοδο του fine-tuning τόσο η εξαγωγή χαρακτηριστικών όσο και η ταξινόμηση γίνεται από το ΣΝΔ. Για το λόγο αυτό τα 3 τελευταία στρώματα του AlexNet τροποποιούνται, προκειμένως το τελευταίο fully-connected layer να οδηγεί σε τόσες εξόδους όσες και οι κλάσεις του συνόλου δεδομένων. Έτσι όπως φαίνεται στην Εικόνα 6.3, το τελευταίο fully-connected layer του AlexNet αντικαθίσταται με άλλο, αποτελούμενο από 6 νευρώνες και ακολουθούμενο από την συνάρτηση softmax και το στρώμα εξόδου.

25x1 **Layer** array with layers:

1	'data'	Image Input	227x227x3 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	''	Fully Connected	6 fully connected layer
24	''	Softmax	softmax
25	''	Classification Output	crossentropyex

Εικόνα 6.3. Δομή fine-tuned Συνελκτικού Νευρωνικού Δικτύου

Στην συνέχεια ακολουθεί η εκπαίδευση του fine-tuned δικτύου, ορίζοντας έναν μεγάλο συντελεστή ρυθμού εκπαίδευσης για τα βάρη των νευρώνων του τελευταίου fully-connected layer, έτσι ώστε να δημιουργηθούν σημαντικές μεταβολές στις τιμές τους, κατά την διάρκεια της εκπαίδευσης. Από την άλλη μεριά, χρησιμοποιείται ένας μικρός αρχικός ρυθμός μάθησης για το υπόλοιπο δίκτυο, προκειμένου να μην μεταβληθούν σημαντικά τα χαρακτηριστικά που εξάγονται από το AlexNet.

Πριν την εκπαίδευση του ΣΝΔ, μεταβάλλονται οι διαστάσεις των εικόνων του συνόλου δεδομένων σε 227x227x3, καθώς το δίκτυο AlexNet δέχεται εικόνες τέτοιων διαστάσεων στην είσοδο του. Αφού γίνουν όλα τα παραπάνω, εκπαιδεύεται το fine-tuned δίκτυο με το σύνολο εικόνων εκπαίδευσης, χρησιμοποιώντας διαφορετικές παραμέτρους ως προς το πλήθος των εποχών και τον διαχωρισμό σε σύνολο εκπαίδευσης και ελέγχου, ερευνώντας τις τιμές εκείνες οι οποίες θα επιτύχουν την βέλτιστη ακρίβεια.

Αναλυτικά, οι παράμετροι που θα μελετηθούν στο ΣΝΔ είναι το πλήθος των εποχών και ο τρόπος διαχωρισμού του συνόλου δεδομένων σε σύνολο εκπαίδευσης και ελέγχου. Καθώς η πλατφόρμα MATLAB R2017a δεν διαθέτει την δυνατότητα εκπαίδευσης ΣΝΔ με την χρήση συνόλου επικύρωσης, για το λόγο αυτό η ακρίβεια του δικτύου θα μελετηθεί ως προς το πλήθος εποχών από 1 έως 10. Ο αριθμός 10 είναι ενδεικτικός και επιλέχθηκε γιατί θεωρείται ότι οι πρώτες 10 εποχές δείχνουν ένα αντιπροσωπευτικό δείγμα της βέλτιστης ταξινόμησης που μπορεί να πραγματοποιηθεί καθώς και αποτελεί ένα άνω όριο, επειδή η εκπαίδευση ΣΝΔ είναι χρονικά απαιτητική διαδικασία. Με τον τρόπο αυτό το σύνολο ελέγχου, λειτουργεί και ως σύνολο επικύρωσης, προσδιορίζοντας την εποχή στην οποία επιτυγχάνεται η μέγιστη ακρίβεια.

Όσον αφορά τον διαχωρισμό του συνόλου δεδομένων, είναι σε 40%-60%, 60%-40% και 80%-20%, με το πρώτο ποσοστό να αντιστοιχεί στο σύνολο εκπαίδευσης και το δεύτερο στο σύνολο ελέγχου. Ο διαχωρισμός των φωτογραφιών σε κάθε ένα από τα παραπάνω σύνολα έχει γίνει με τυχαίο τρόπο.

6.1.2.1 Διαχωρισμός του Συνόλου Δεδομένων σε 40%-60%

Αρχικά υλοποιείται ο διαχωρισμός του συνόλου δεδομένων σε 40%-60%. Ο Πίνακας 6.5 παρουσιάζει αναλυτικά την ακρίβεια που επιτυγχάνεται σε κάθε εποχή. Όπως αναδεικνύει ο πίνακας αυτός υπάρχει σταδιακή αύξηση της ακρίβειας μέχρι την 6^η εποχή ενώ στην συνέχεια η ακρίβεια κάνει μικρές διακυμάνσεις γύρω από την τιμή του 91%. Οι εποχές στις οποίες πραγματοποιείται η μέγιστη ακρίβεια είναι η 6^η και η 8^η, για τις οποίες αναλυτικά παρουσιάζονται οι Μήτρες Σύγκρισης στους Πίνακας 6.4 και Πίνακας 6.3 αντίστοιχως. Η μέγιστη ακρίβεια που σημειώνεται είναι 91,67%.

Αξίζει να γίνει αναφορά στα αποτελέσματα που περιγράφονται από τις Μήτρες Σύγκρισης των εποχών 6 και 8. Αν και στις δύο περιπτώσεις επιτυγχάνεται η ίδια τελική ακρίβεια παρατηρείται ότι στις δύο περιπτώσεις ο ταξινομητής παρουσιάζει πολύ διαφορετική συμπεριφορά. Στην Εποχή 6 αναγνωρίζονται με ακρίβεια σχεδόν 86% τα Ζυμαρικά ενώ παρόμοια ακρίβεια εμφανίζεται και για τις Πατάτες. Αντιθέτως στην περίπτωση της εποχής 8, τα Ζυμαρικά εμφανίζουν χαμηλότερη ακρίβεια 78,33% ενώ σημειώνεται μεγάλη αύξηση στην ταξινόμηση των Πατατών που ξεπερνάει το 90%. Στις υπόλοιπες κλάσεις οι διαφορές ανάμεσα στους δύο ταξινομητές είναι αμελητέες. Άρα παρατηρείται ότι οι ταξινομητές των εποχών 6 και 8 αν και σημειώνουν την ίδια ακρίβεια παρουσιάζουν διαφορετική συμπεριφορά στην ταξινόμηση δύο κλάσεων.

Εποχή	Ακρίβεια
1	74.86%
2	88.75%
3	89.72%
4	90.69%
5	90.83%
6	91.67%
7	90.97%
8	91.67%
9	91.25%
10	91.25%

Πίνακας 6.5. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 40-60% του συνόλου δεδομένων

Εποχή 6	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	99.17%	0.83%	0.00%	0.00%	0.00%	0.00%
Κρέας	13.33%	83.33%	0.00%	1.67%	1.67%	0.00%
Ζυμαρικά	0.00%	0.00%	85.83%	10.00%	2.50%	1.67%
Πατάτες	6.67%	2.50%	4.17%	85.00%	1.67%	0.00%
Ρύζι	0.83%	0.00%	0.83%	0.00%	98.33%	0.00%
Λαχανικά	0.00%	0.00%	0.83%	0.83%	0.00%	98.33%

Πίνακας 6.4. Μήτρα Σύγκρισης της 6^{ης} εποχής στο σχήμα 40%-60% του συνόλου δεδομένων

Εποχή 8	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	98.33%	1.67%	0.00%	0.00%	0.00%	0.00%
Κρέας	13.33%	84.17%	0.00%	1.67%	0.83%	0.00%
Ζυμαρικά	0.00%	0.00%	78.33%	15.83%	3.33%	2.50%
Πατάτες	4.17%	4.17%	0.00%	90.83%	0.83%	0.00%
Ρύζι	0.83%	0.00%	0.00%	0.00%	99.17%	0.00%
Λαχανικά	0.00%	0.00%	0.00%	0.83%	0.00%	99.17%

Πίνακας 6.3. Μήτρα Σύγκρισης της 8^{ης} εποχής στο σχήμα 40%-60% του συνόλου δεδομένων

6.1.2.2 Διαχωρισμός του Συνόλου Δεδομένων σε 60%-40%

Στη συνέχεια εξετάζεται η περίπτωση στην οποία το σύνολο εκπαίδευσης αποτελείται από το 60% των εικόνων ενώ το υπόλοιπο 40% αποτελεί το σύνολο ελέγχου. Συγκεκριμένα παρουσιάζεται η ακρίβεια που επιτυγχάνεται στις πρώτες 10 εποχές στον Πίνακα 6.7, με το βέλτιστο αποτέλεσμα να επιτυγχάνεται στην 7^η εποχή. Αναλυτικά παρουσιάζεται η Μήτρα Σύγχυσης της 7^{ης} εποχής στον Πίνακα 6.6 φτάνοντας την ακρίβεια του 95,83%.

Εποχή	Ακρίβεια
1	87.71%
2	92.71%
3	94.58%
4	94.58%
5	95.21%
6	95.42%
7	95.83%
8	95.42%
9	95.21%
10	95.63%

Εποχή 7	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	96.25%	2.50%	0.00%	1.25%	0.00%	0.00%
Κρέας	8.75%	90.00%	0.00%	1.25%	0.00%	0.00%
Ζυμαρικά	0.00%	0.00%	98.75%	0.00%	0.00%	1.25%
Πατάτες	0.00%	2.50%	2.50%	93.75%	1.25%	0.00%
Ρύζι	1.25%	0.00%	0.00%	0.00%	98.75%	0.00%
Λαχανικά	0.00%	0.00%	2.50%	0.00%	0.00%	97.50%

Πίνακας 6.6. Μήτρα Σύγχυσης της 7^{ης} εποχής στο σχήμα 60%-40% του συνόλου δεδομένων

Πίνακας 6.7. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 60%-40% του συνόλου δεδομένων

6.1.2.3 Διαχωρισμός του Συνόλου Δεδομένων σε 80%-20%

Στην τελευταία περίπτωση το σύνολο δεδομένων διακρίνεται σε 80%-20%. Για μία ακόμη φορά η μέγιστη ακρίβεια επιτυγχάνεται στην 7^η εποχή ενώ αναλυτικά η Μήτρα Σύγχυσης αυτής της εποχής φαίνεται στον Πίνακα 6.8

Εποχή	Ακρίβεια
1	90.00%
2	93.33%
3	95.83%
4	95.42%
5	96.25%
6	96.67%
7	97.08%
8	96.67%
9	95.83%
10	96.25%

Εποχή 7	Ψωμί	Κρέας	Ζυμαρικά	Πατάτες	Ρύζι	Λαχανικά
Ψωμί	95.00%	5.00%	0.00%	0.00%	0.00%	0.00%
Κρέας	0.00%	97.50%	0.00%	2.50%	0.00%	0.00%
Ζυμαρικά	0.00%	0.00%	92.50%	7.50%	0.00%	0.00%
Πατάτες	0.00%	2.50%	0.00%	97.50%	0.00%	0.00%
Ρύζι	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
Λαχανικά	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Πίνακας 6.8. Μήτρα Σύγχυσης της 7^{ης} εποχής στο σχήμα 80%-20% του συνόλου δεδομένων

Πίνακας 6.9. Η ακρίβεια που επιτυγχάνεται για κάθε εποχή εκπαίδευσης στο σχήμα 80%-20% του συνόλου δεδομένων

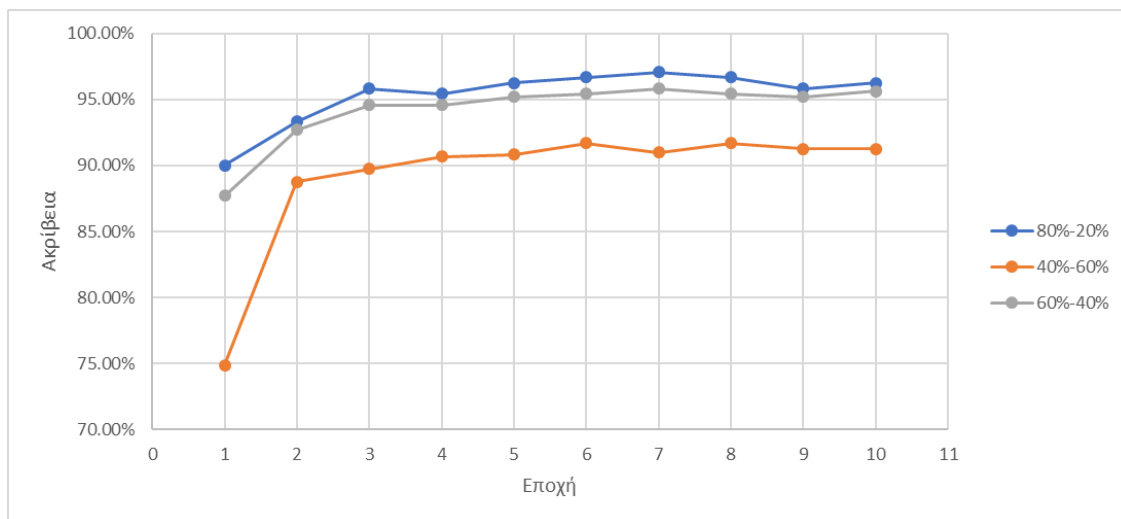
6.1.2.4 Σύγκριση των Ταξινομητών του Fine-Tuning

Στις τρεις παραπάνω περιπτώσεις, αν και τα βέλτιστα αποτελέσματα ακρίβειας που επιτυγχάνονται διαφέρουν, εμφανίζονται σημαντικές ομοιότητες στην καμπύλη που απεικονίζει την μεταβολή της ακρίβειας συναρτήσει του πλήθους των εποχών (Εικόνα 6.4). Συγκεντρωτικά, οι 3 καμπύλες σημειώνουν το μέγιστο τους κοντά στην 7^η εποχή. Ειδικά, στην πρώτη περίπτωση η μέγιστη ακρίβεια πραγματοποιήθηκε στην 6^η και 8^η εποχή, ενώ στις δύο τελευταίες περιπτώσεις η ακρίβεια αυτή πραγματοποιήθηκε στην 7^η εποχή.

Επίσης μπορεί να συναχθεί το συμπέρασμα ότι μέχρι την 6^η με 7^η εποχή εκπαιδεύεται ουσιαστικά το ΣΝΔ, ενώ μετά παρουσιάζεται το φαινόμενο του overfitting. Αυτό συμβαίνει και στις τρεις περιπτώσεις με εξαίρεση την περίπτωση του 80%-20%, κατά την οποία παρουσιάζονται κάποιες αυξομειώσεις στην ακρίβεια κοντά στην 4^η εποχή.

Γενικότερα, φαίνεται ότι όσο μεγαλύτερο είναι το σύνολο δεδομένων τόσο μεγαλύτερη είναι και η ακρίβεια της πρώτης εποχής, παρ' όλα αυτά η σύγκλιση γίνεται σε όλες τις περιπτώσεις στην 7^η εποχή. Επίσης και στις τρεις περιπτώσεις ήδη από την 3^η εποχή έχει επιτευχθεί ακρίβεια πολύ κοντά στην βέλτιστη, αναδεικνύοντας την ταχύτητα σύγκλισης των ΣΝΔ.

Τέλος όσον αφορά τα αποτελέσματα βέλτιστης ακρίβειας που επιτυγχάνονται από τις τρεις διαφορετικές περιπτώσεις, παρατηρείται ότι η καλύτερη ακρίβεια σημειώθηκε στην 7^η εποχή του σχήματος 80%-20%, με τιμή 97,08%. Όπως φαίνεται στην Εικόνα 6.4 σε κάθε εποχή ο διαχωρισμός 80%-20% σημειώνει την υψηλότερη ακρίβεια. Προκύπτει έτσι το συμπέρασμα ότι το μεγαλύτερο πλήθος εικόνων στο σύνολο εκπαίδευσης επιτρέπει υψηλότερα επίπεδα γενίκευσης, άρα και καλύτερα αποτελέσματα.



Εικόνα 6.4. Διαγραμματική Αναπαράσταση της ακρίβειας που επιτυγχάνεται στις πρώτες 10 εποχές στις περιπτώσεις διαχωρισμού του συνόλου δεδομένων σε 40%-60%, 60%-40% και 80%-20%.

6.1.3 Αποτελέσματα

Τα αποτελέσματα από όλες τις τεχνικές ταξινόμησης που χρησιμοποιήθηκαν στο σύνολο δεδομένων αναγράφονται στον Πίνακας 6.10. Όπως φαίνεται, το αποτέλεσμα με την καλύτερη ακρίβεια επιτυγχάνεται με την μέθοδο του Fine-Tuning, διαχωρίζοντας το σύνολο δεδομένων σε 80%-20% κατά την 7^η εποχή της εκπαίδευσης. Γενικότερα, φαίνεται ότι σε όλες τις τεχνικές που χρησιμοποιούνται ΣΝΔ σημειώνονται πολύ υψηλά ποσοστά ακρίβειας. Αν εξαιρέσουμε την περίπτωση του σχήματος 40%-60%, που το σύνολο εκπαίδευσης αποτελείται από μόλις 80 εικόνες ανά κατηγορία και συνολικά από λιγότερες εικόνες από ότι το σύνολο ελέγχου, μπορεί να συναχθεί το συμπέρασμα ότι η τεχνική του Fine-Tuning σημειώνει καλύτερα αποτελέσματα.

Επίσης πολύ σημαντικό είναι να διαπιστωθεί η βελτίωση που πραγματοποιείται στην μέχρι πρότεινως ακρίβεια ταξινόμησης. Συγκεκριμένα, μέσω της μεθόδου του Multiclass SVM που παρουσιάζεται ως βέλτιστος ταξινομητής του συνόλου δεδομένων Sarantea-2016 επιτυγχάνεται ακρίβεια 86,92% με την μέθοδο της διασταυρωμένης επικύρωσης 10 επαναλήψεων (Σαραντέα-Μαγγαλούση, 2016). Συγκρίνοντας το με τα αποτελέσματα του Πίνακας 6.10 παρατηρείται ότι όλες οι προτεινόμενες μέθοδοι επιτυγχάνουν καλύτερη ακρίβεια, ενώ στην περίπτωση του Fine-Tuning 80%-20%, η βελτίωση της ακρίβειας είναι μεγαλύτερη του 10%. Τα αποτελέσματα αυτά επαληθεύουν την αποδοτικότητα των ΣΝΔ στην ταξινόμηση εικόνων και ειδικότερα στον κλάδο της ταξινόμησης εικόνων από φωτογραφικά στιγμιότυπα.

Ο λόγος για τον οποίο γίνεται η παραπάνω διαδικασία είναι για να αποδειχθεί ότι ο ταξινομητής που αναπτύχθηκε στην παρούσα εργασία βελτιώνει την ακρίβεια ενός υπάρχοντος συνόλου δεδομένων. Επίσης το σύνολο δεδομένων Sarantea-2016 λόγω του περιορισμένου πλήθους εικόνων που περιέχει, μόλις 1200, και το γεγονός ότι αυτές ακολουθούν διαχωρισμό σύμφωνα με την περιεκτικότητα τους σε υδατάνθρακες, επιτρέπει την εξαγωγή χρήσιμων συμπερασμάτων σχετικά με την τεχνική ταξινόμησης που θα χρησιμοποιηθεί στο νέο σύνολο δεδομένων NTUA-Food 2017. Στην προκειμένη περίπτωση το συμπέρασμα που εξάγεται είναι ότι τα ΣΝΔ επιτυγχάνουν καλύτερη ακρίβεια συγκριτικά με τις συμβατικές τεχνικές ταξινόμησης και ότι ο βέλτιστος τρόπος αξιοποίησής τους είναι μέσω της μεθόδου του Fine-Tuning.

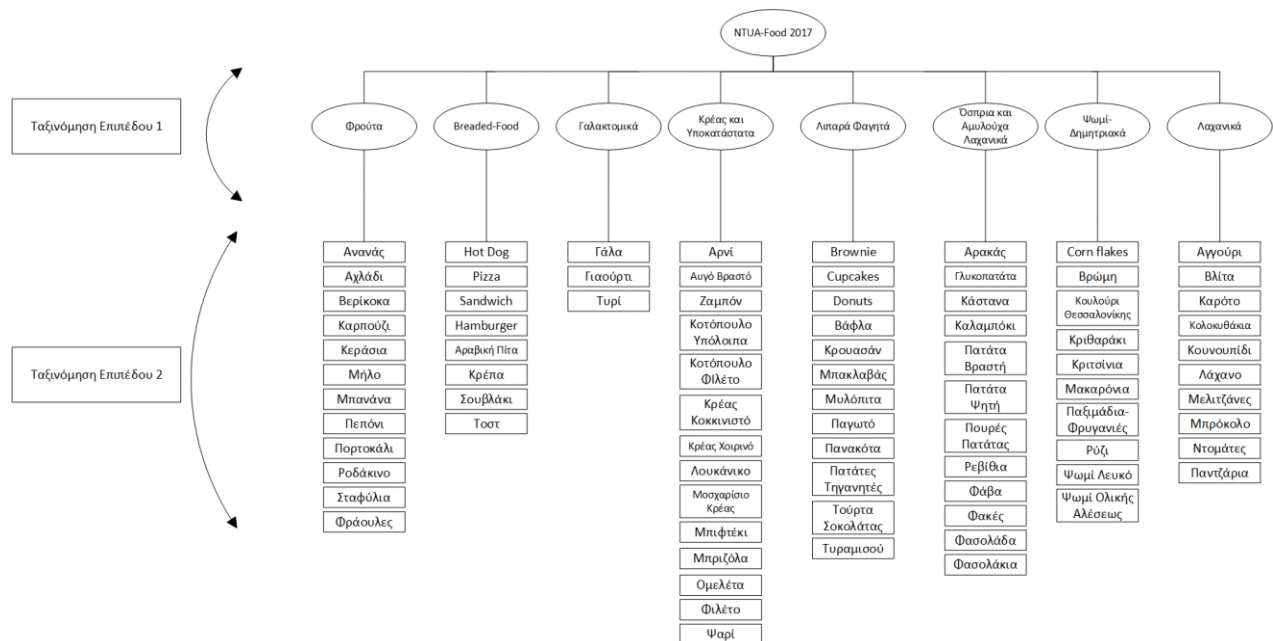
Εποχή	Χαρακτηριστικά	Σύνολο Εκπαίδευσης-Ελέγχου	ΣΝΔ	Μέθοδος	Ακρίβεια
-	4096	-	AlexNet	Εξαγωγή Χαρακτηριστικών	93.96%
-	4096	-	VGGnet	Εξαγωγή Χαρακτηριστικών	94.79%
6, 8	-	40-60%	AlexNet	Fine Tuning	91.67%
7	-	60-40%	AlexNet	Fine Tuning	95.83%
7	-	80-20%	AlexNet	Fine Tuning	97.08%

Πίνακας 6.10. Παρουσίαση αποτελεσμάτων για τις διαφορετικές τεχνικές ταξινόμησης

6.2 Ταξινόμηση στο Νέο Σύνολο Δεδομένων NTUA-Food 2017

Λαμβάνοντας υπόψιν την ανάλυση του συνόλου δεδομένων Sarantea-2016, συμπεραίνεται ότι τα καλύτερα αποτελέσματα προέρχονται από την μέθοδο του Fine-Tuning. Επίσης, καθώς κάθε σύνολο δεδομένων σύμφωνα με το πλήθος, το περιεχόμενο και τον αριθμό των κλάσεων που χρησιμοποιεί εμφανίζει διαφορετική απόκριση, δεν μπορεί να γίνει γενίκευση των παραμέτρων που εξήχθησαν στο σύνολο δεδομένων του Sarantea-2016 για το νέο σύνολο δεδομένων NTUA-Food 2017.

Η ταξινόμηση του νέου συνόλου δεδομένων γίνεται σε δύο επίπεδα. Αρχικά γίνεται ο διαχωρισμός των τροφίμων στις ευρύτερες κατηγορίες Breaded Food, Γαλακτοκομικά, Κρέας και Υποκατάστατα, Λαχανικά, Φαγητά, Λιπαρά Φαγητά, Όσπρια και Αμυλούχα Λαχανικά, Φρούτα και Ψωμί Δημητριακά. Όπως έχει αναφερθεί και στο Υποκεφάλαιο 5.6 η ανάλυση αυτή γίνεται σύμφωνα με την επίδραση που έχει κάθε ομάδα φαγητών στο σάκχαρο του ατόμου με ΣΔ. Έπειτα, γίνεται η ταξινόμηση δευτέρου επιπέδου, κατά την οποία γίνεται ταξινόμηση ανάμεσα στα φαγητά κάθε ομάδας.



Εικόνα 6.5. Ταξινόμηση δύο επιπέδων

6.2.1 Ταξινόμηση Επιπέδου 1

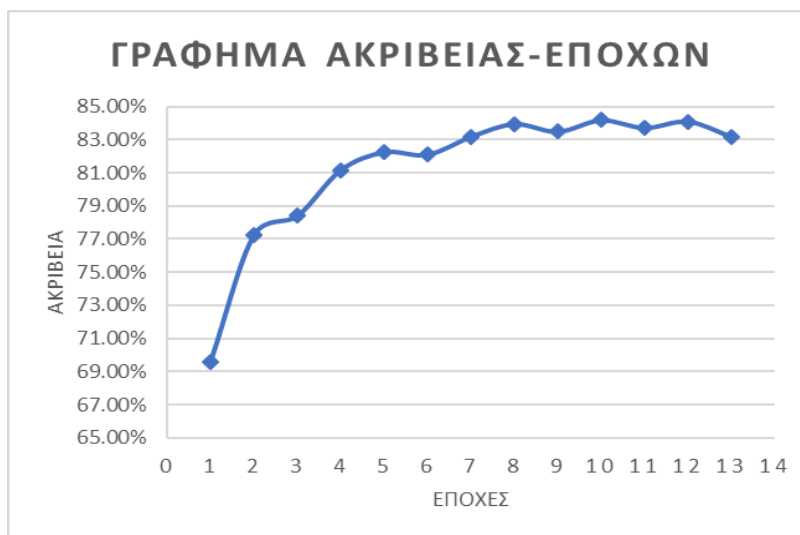
Το πρώτο βήμα στο επίπεδο ταξινόμησης 1, είναι να κρατηθεί ίσο πλήθος φωτογραφιών από κάθε κατηγορία, δίνοντας έτσι ίση βαρύτητα στην προσαρμογή του ταξινομητή σε κάθε μια κλάση κατά την διάρκεια της εκπαίδευσης. Για το λόγο αυτό εντοπίζεται η κατηγορία με το μικρότερο πλήθος εικόνων, που στην συγκεκριμένη περίπτωση είναι τα γαλακτοκομικά προϊόντα με 405 εικόνες, και επιλέγονται τυχαία τόσες εικόνες από τις υπόλοιπες κατηγορίες. Με τον τρόπο αυτό δημιουργείται ένα ομογενές υποσύνολο του αρχικού συνόλου δεδομένων, με ίσο πλήθος φωτογραφιών για κάθε κατηγορία.

Στην συνέχεια ακολουθείται η ίδια διαδικασία που εφαρμόστηκε στην ταξινόμηση με fine-tuning του συνόλου δεδομένων Sarantea-2016. Χρησιμοποιείται το προ-εκπαιδευμένο ΣΝΔ AlexNet, αλλάζοντας το τελευταίο επίπεδο πλήρους σύνδεσης, έτσι ώστε να αποτελείται από 8 νευρώνες, όσες και οι κλάσεις ταξινόμησης. Το σύνολο δεδομένων χωρίζεται σε σύνολο εκπαίδευσης και σύνολο ελέγχου με ποσοστά 60% και 40% αντιστοίχως. Τέλος, το σύστημα εκπαιδεύεται για 13 εποχές, κατά τις οποίες μελετάται η συμπεριφορά του.

Αρχικά είχε οριστεί ως μέγιστο πλήθος εποχών εκπαίδευσης του ΣΝΔ ο αριθμός 10, ο ίδιος αριθμός δηλαδή με το πλήθος εποχών που χρησιμοποιήθηκε κατά την εκπαίδευση στο σύνολο δεδομένων Sarantea-2016. Οι πρώτες 10 εποχές όμως δεν παρήγαγαν σαφή συμπεράσματα για την ακρίβεια του ΣΝΔ, καθώς στην 10^η εποχή επιτεύχθηκε η βέλτιστη ακρίβεια. Αυτός είναι και ο λόγος που η διαδικασία εκπαίδευσης του ΣΝΔ επεκτάθηκε κατά 3 εποχές. Αναλυτικά στην Εικόνα 6.6 παρουσιάζεται η ακρίβεια που σημειώνει το νευρωνικό δίκτυο, κατά τις 13 εποχές εκπαίδευσης του.

Εποχή	Ακρίβεια
1	69.60%
2	77.24%
3	78.40%
4	81.17%
5	82.25%
6	82.10%
7	83.18%
8	83.95%
9	83.49%
10	84.18%
11	83.72%
12	84.10%
13	83.18%

(α)



(β)

Εικόνα 6.6. (α) Πίνακας ακρίβειας των 13 εποχών εκπαίδευσης (β) Το αντίστοιχο γράφημα αναπαράστασης της ακρίβειας συναρτήσει της εποχής

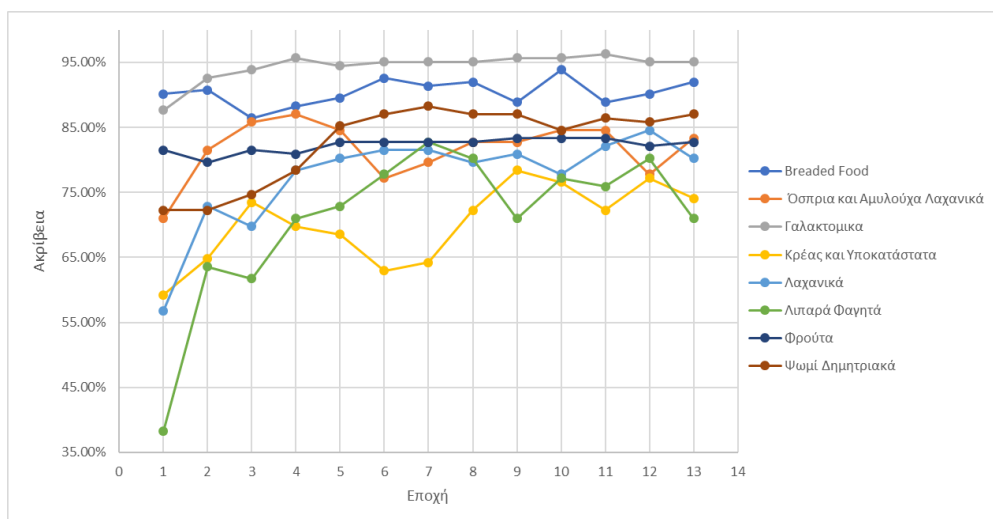
Στην συνέχεια, αξίζει να αναλυθεί η συμπεριφορά του ταξινομητή κατά τις τελευταίες εποχές, ιδίως κατά την 8^η, την 10^η και την 12^η. Στις εποχές αυτές ο ταξινομητής επιτυγχάνει τα καλύτερα ποσοστά ακρίβειας, των οποίων η διαφορά κυμαίνεται σε ένα εύρος τιμών μικρότερο του 0.3%, όπως φαίνεται στην Εικόνα 6.6.

Αναλυτικά, όπως απεικονίζεται τόσο στο Πίνακα 6.11 όσο και στην Εικόνα 6.7, ο ταξινομητής δεν βελτιώνει προοδευτικά τα ποσοστά ακρίβειας κάθε ομάδα φαγητών. Αντιθέτως μεταξύ των εποχών 8, 10 και 12 γίνεται ανταλλαγή (trade-off) μεταξύ των ομάδων φαγητών τις οποίες ο ταξινομητής αναγνωρίζει βέλτιστα. Στον Πίνακα 6.11 με κόκκινο χρώμα έχει σημειωθεί η βέλτιστη ακρίβεια η οποία έχει επιτευχθεί ανά κλάση, με πράσινο χρώμα σημειώνεται η καλύτερη ακρίβεια ανά ομάδα φαγητών που επιτυγχάνεται μεταξύ των εποχών 8 και 10, ενώ τέλος η ίδια σύγκριση γίνεται ανάμεσα στις εποχές 10 και 12 και υπογραμμίζεται με μπλε χρώμα. Είναι εμφανές ότι και στις δύο περιπτώσεις, ο ταξινομητής της εποχής 10 σημειώνει τα καλύτερα αποτελέσματα για τις κατηγορίες Breaded Food, Όσπρια και Αμυλούχα Λαχανικά, Γαλακτοκομικά και Φρούτα, ενώ σημειώνει παρόμοια η χαμηλότερα ποσοστά ακρίβειας στις υπόλοιπες κατηγορίες. Όσον αφορά τους ταξινομητές της εποχής 8 και 12 παρουσιάζουν παρεμφερής συμπεριφορά μεταξύ τους, με τον ταξινομητή της εποχής 8 να σημειώνει καλύτερη ακρίβεια στα φαγητά των Breaded Food και Όσπρια και Αμυλούχα Λαχανικά, ενώ ο ταξινομητής της εποχής 12 να υπερτερεί στο Κρέας και Υποκατάστατα και στα Λαχανικά.

Εποχή	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτομικα	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
1	90.12%	70.99%	87.65%	59.26%	56.79%	38.27%	81.48%	72.22%
2	90.74%	81.48%	92.59%	64.81%	72.84%	63.58%	79.63%	72.22%
3	86.42%	85.80%	93.83%	73.46%	69.75%	61.73%	81.48%	74.69%
4	88.27%	87.04%	95.68%	69.75%	78.40%	70.99%	80.86%	78.40%
5	89.51%	84.57%	94.44%	68.52%	80.25%	72.84%	82.72%	85.19%
6	92.59%	77.16%	95.06%	62.96%	81.48%	77.78%	82.72%	87.04%
7	91.36%	79.63%	95.06%	64.20%	81.48%	82.72%	82.72%	88.27%
8	91.98%	82.72%	95.06%	72.22%	79.63%	80.25%	82.72%	87.04%
9	88.89%	82.72%	95.68%	78.40%	80.86%	70.99%	83.33%	87.04%
10	93.83%	84.57%	95.68%	76.54%	77.78%	77.16%	83.33%	84.57%
11	88.89%	84.57%	96.30%	72.22%	82.10%	75.93%	83.33%	86.42%
12	90.12%	77.78%	95.06%	77.16%	84.57%	80.25%	82.10%	85.80%
13	91.98%	83.33%	95.06%	74.07%	80.25%	70.99%	82.72%	87.04%

Πίνακας 6.11. Παρουσιάζεται το ποσοστό ακρίβειας που επιτυγχάνεται ανά κατηγορία φαγητών ανά εποχή. Με κόκκινο σημειώνεται η βέλτιστη ακρίβεια που επιτυγχάνεται ανά κατηγορία, με πράσινο χρωματίζεται η βέλτιστη ακρίβεια ανά ομάδα μεταξύ των εποχών 8 και 10, ενώ με μπλε σημειώνεται η καλύτερη ακρίβεια ανάμεσα στις εποχές 10 και 12.

Η διάρκεια της εκπαίδευσης του ΣΝΔ AlexNet για 13 εποχές, στο σύνολο δεδομένων NTUA-Food 2017, διήρκεσε περίπου 12 μέρες σε περιβάλλον MATLAB. Η έλλειψη υψηλής χωρητικότητας στην κάρτα γραφικών του υπολογιστή, δεν επέτρεψε την διαδικασία να εκτελεσθεί από την ίδια, γεγονός το οποίο θα βελτιώνει σημαντικά το χρόνο εκπαίδευσης.



Εικόνα 6.7. Διαγραμματική αναπαράσταση του ποσοστού των σωστά αναγνωρισμένων εικόνων ανά εποχή και ανά κατηγορία

6.2.1.1 Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων

Προκειμένου να γίνει μια πιο αντικειμενική σύγκριση των αποτελεσμάτων της ταξινόμησης, η οποία ξεφεύγει από τον απόλυτο αριθμό της ακρίβειας και λαμβάνει υπόψιν τον σκοπό του αλγορίθμου, χρησιμοποιείται η Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων. Η μετρική η οποία χρησιμοποιεί τη Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων, στηρίζεται στις λανθασμένες προβλέψεις κάθε δικτύου, πολλαπλασιάζοντας τες με ένα συντελεστή, ίσο με την απόκλιση του σφάλματος πρόβλεψης που γίνεται ανά φαγητό. Η Μήτρα Αξιολόγησης παρουσιάζεται στην Εικόνα 6.8.

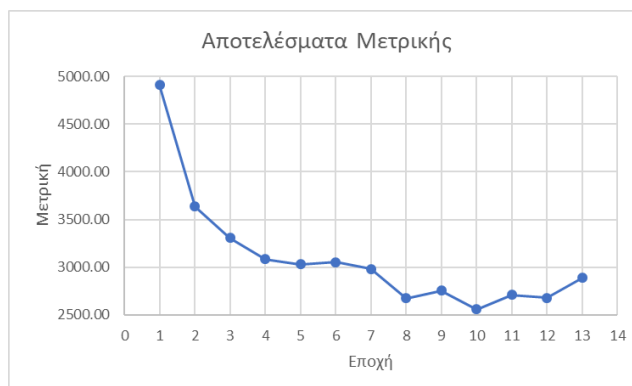
Μήτρα Αξιολόγησης		Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτομικα	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
		30.00	13.75	16.29	8.75	6.07	37.50	15.00	15.00
Breaded Food	30.00	0.00	16.25	13.71	21.25	23.93	7.50	15.00	15.00
Όσπρια και Αμυλούχα Λαχανικά	13.75	16.25	0.00	2.54	5.00	7.68	23.75	1.25	1.25
Γαλακτομικα	16.29	13.71	2.54	0.00	7.54	10.22	21.21	1.29	1.29
Κρέας και Υποκατάστατα	8.75	21.25	5.00	7.54	0.00	2.68	28.75	6.25	6.25
Λαχανικά	6.07	23.93	7.68	10.22	2.68	0.00	31.43	8.93	8.93
Λιπαρά Φαγητά	37.50	7.50	23.75	21.21	28.75	31.43	0.00	22.50	22.50
Φρούτα	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00	0.00
Ψωμί Δημητριακά	15.00	15.00	1.25	1.29	6.25	8.93	22.50	0.00	0.00

Εικόνα 6.8. Μήτρα Αξιολόγησης

Στο Πίνακα 6.12 αναγράφονται τα αποτελέσματα της μετρικής ανά εποχή και αναπαρίστανται στην Εικόνα 6.9. Δεδομένου ότι η μετρική καταγράφει το πλήθος των σφαλμάτων και τα πολλαπλασιάζει με ένα συντελεστή ανάλογο της σοβαρότητας τους, για το λόγο αυτό αναζητάται ο ταξινομητής ο οποίος επιστρέφει το ελάχιστο αποτέλεσμα. Ο Πίνακας 6.12 δείχνει ότι η ελαχιστοποίηση του σφάλματος πραγματοποιείται στην εποχή 10. Ως εκ τούτου, ο ταξινομητής της εποχής 10 όχι μόνο σημειώνει τα καλύτερα αποτελέσματα ακρίβειας αλλά παρουσιάζει και την καλύτερη συμπεριφορά σύμφωνα με το πλαίσιο ανάπτυξης του αλγορίθμου. Αυτό φαίνεται εμφαντικά στην τελευταία στήλη του Πίνακα 6.12, όπου από την εποχή 5 και μετά, το μικρότερο μέσο σφάλμα παρατηρείται την εποχή 10. Ως μέσο σφάλμα ορίζεται ο λόγος του αποτελέσματος της μετρικής ως προς το πλήθος των σφαλμάτων. Το μέσο σφάλμα συνεπώς αποτελεί ένδειξη ανεξάρτητη του πλήθους σφαλμάτων και δείχνει την σοβαρότητα κατά μέσο όρο του σφάλματος κάθε ταξινομητή.

Εποχή	Μετρική	Πλήθος Σφαλμάτων	Μέσο Σφάλμα
1	4912,57	394	12,46894725
2	3634,98	295	12,32324086
3	3305,94	280	11,80961771
4	3085,09	244	12,64188727
5	3028,16	230	13,16360509
6	3052,49	232	13,1581981
7	2981,03	218	13,67525249
8	2671,33	208	12,84244562
9	2751,62	214	12,85985085
10	2556,47	205	12,46893792
11	2709,75	211	12,84310555
12	2676,36	206	12,98802452
13	2887,72	218	13,2472195

Πίνακας 6.12. Στις πρώτες δύο στήλες γίνεται παρουσίαση των αποτελεσμάτων της μετρικής ανά εποχή. Στην τρίτη στήλη παρουσιάζεται το πλήθος σφαλμάτων από τις 1296 συνολικά εικόνες. Στην τέταρτη στήλη φαίνεται το μέσο σφάλμα ανά εποχή, με κόκκινο χρωματίζονται τα κελιά με το υψηλότερο μέσο σφάλμα, ενώ με μπλε αυτά με το μικρότερο.



Εικόνα 6.9. Διαγραμματική αναπαράσταση του αποτελέσματος της μετρικής ανά εποχή

Για να γίνει ακόμα πιο κατανοητή η πρακτική λειτουργία της Μετρικής βασισμένη στην Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων παρατίθεται ο Πίνακας 6.13. Στο πίνακα αυτό εμφανίζεται η μέση απόκλιση υδατανθράκων ανά φαγητό (στήλη Μέση Απόκλιση Υδατανθράκων), η οποία δείχνει κατά μέσο όρο πόσο συνέβαλε στην τελική απόκλιση των υδατανθράκων κάθε εικόνα, δεδομένου ότι απεικονίζεται ένα φαγητό ανά εικόνα. Επιπλέον όπως έχει αναφερθεί στο Κεφάλαιο 3, όταν η απόκλιση πρόβλεψης της

μεταγευματικής κατανάλωσης υδατανθράκων δεν ξεπερνά τα 10γρ. δεν προκαλείται αποσυντονισμός στο μεταγευματικό σάκχαρο (Smart, et al., 2009). Αντιθέτως όταν η πρόβλεψη της περιεχόμενης ποσότητας υδατανθράκων υπερβαίνει τα 20γρ. επιφέρει σοβαρές επιπτώσεις στο άτομο (Smart, King, McElduff, & Collins, 2012). Σύμφωνα με αυτές τις αρχές, υπολογίζεται κατά μέσο όρο το πλήθος φαγητών, τα οποία μπορούν να καταναλωθούν από το άτομο με ΣΔ, προκειμένου να δημιουργηθεί απόκλιση στην πρόβλεψη σακχάρου μέχρι 10 και 20γρ. αντιστοίχως. Πρακτικά, η καλύτερη απόδοση του ταξινομητή της 10^{ης} εποχής υπογραμμίζεται από το γεγονός ότι η πρόβλεψη μέχρι και 10 φαγητών ανά γεύμα, δεν θα προκαλέσει απόκλιση μεγαλύτερη των 20γρ. από την πραγματική, κατά μέσο όρο. Την συμπεριφορά αυτή την υποστηρίζει μόνο ο ταξινομητής της 10^{ης} εποχής.

Εποχή	Μετρική	Μέση Απόκλιση Υδατανθράκων	Πλήθος φαγητών (10γρ.)	Πλήθος φαγητών (20γρ.)
1	4912,57	3,79	2	5
2	3634,98	2,80	3	7
3	3305,94	2,55	3	7
4	3085,09	2,38	4	8
5	3028,16	2,34	4	8
6	3052,49	2,36	4	8
7	2981,03	2,30	4	8
8	2671,33	2,06	4	9
9	2751,62	2,12	4	9
10	2556,47	1,97	5	10
11	2709,75	2,09	4	9
12	2676,36	2,07	4	9
13	2887,72	2,23	4	8

Πίνακας 6.13. Παρουσίαση των αποτελεσμάτων της μετρικής ανά εποχή. Συμπληρωματικά παρατίθενται η μέση απόκλιση των υδατανθράκων ανά φαγητό και το πλήθος φαγητών που κατά μέσο όρο χρειάζονται για απόκλιση μέχρι 10γρ και μέχρι 20γρ. υδατανθράκων αντιστοίχως.

Αναλυτικά τα αποτελέσματα ταξινόμησης που επιτυγχάνονται μετά από την εκπαίδευση του δικτύου AlexNet από το σύνολο δεδομένων NTUA-Food 2017 κατά τη 10^η εποχή απεικονίζονται στο Πίνακα 6.14.

Εποχή 10	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτοκομικά	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
Breaded Food	93.83%	1.23%	0.00%	1.85%	0.00%	1.85%	0.00%	1.23%
Όσπρια και Αμυλούχα Λαχανικά	2.47%	84.57%	0.00%	5.56%	1.85%	1.23%	1.85%	2.47%
Γαλακτοκομικά	0.00%	0.00%	95.68%	0.00%	0.62%	0.62%	0.00%	3.09%
Κρέας και Υποκατάστατα	6.17%	3.70%	3.09%	76.54%	3.70%	5.56%	0.00%	1.23%
Λαχανικά	1.85%	6.79%	3.09%	2.47%	77.78%	0.00%	2.47%	5.56%
Λιπαρά Φαγητά	3.70%	6.17%	0.62%	6.17%	2.47%	77.16%	0.62%	3.09%
Φρούτα	0.00%	4.94%	0.00%	0.00%	9.88%	0.62%	83.33%	1.23%
Ψωμί Δημητριακά	4.94%	3.70%	0.62%	1.85%	2.47%	1.85%	0.00%	84.57%

Πίνακας 6.14. Μήτρα Σύγχυσης της εποχής 10

Τα αποτελέσματα ταξινόμησης που επιτυγχάνονται κατά τη 10^η εποχή παρουσιάζονται και με Μήτρα Σύγχυσης, η οποία περιέχει το πλήθος φωτογραφιών (Πίνακας 6.15). Ο λόγος για τον οποίο παρουσιάζεται σε αυτή τη μορφή η Μήτρα Σύγχυσης είναι, ότι με πίνακες τέτοιας μορφής πολλαπλασιάζεται η Μήτρα Αξιολόγησης Διαφοράς Υδατανθράκων, για να προκύψουν τα αποτελέσματα της Μετρικής. Στη Μήτρα Σύγχυσης του Πίνακα 6.15, κάθε κατηγορία αποτελείται από 162 εικόνες.

Εποχή 10	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτοκομικά	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά
Breaded Food	152	2	0	3	0	3	0	2
Όσπρια και Αμυλούχα Λαχανικά	4	137	0	9	3	2	3	4
Γαλακτοκομικά	0	0	155	0	1	1	0	5
Κρέας και Υποκατάστατα	10	6	5	124	6	9	0	2
Λαχανικά	3	11	5	4	126	0	4	9
Λιπαρά Φαγητά	6	10	1	10	4	125	1	5
Φρούτα	0	8	0	0	16	1	135	2
Ψωμί Δημητριακά	8	6	1	3	4	3	0	137

Πίνακας 6.15. Μήτρα Σύγχυσης της εποχής 10 που αναπαριστά τα αποτελέσματα της ταξινόμησης σε μορφή πλήθους εικόνων και όχι με ποσοστιαία αναπαράσταση

6.2.2 Ανάλυση Σφαλμάτων

Σύμφωνα με τα παραπάνω αποτελέσματα παρατηρείται ότι ο ταξινομητής που δημιουργείται κατά την 10^η εποχή εκπαίδευσης του ΣΝΔ, σημειώνει την βέλτιστη ακρίβεια. Παρακάτω γίνεται μια απόπειρα σκιαγράφησης των σημαντικότερων αδυναμιών που παρουσιάζει ο ταξινομητής αυτός. Η ανάλυση θα βασιστεί στη μήτρα σύγχυσης της 10^{ης} εποχής σε συνδυασμό με την παρουσίαση των μεμονωμένων σφαλμάτων που εμφανίστηκαν στο σύνολο ελέγχου.

Αρχικά ένα σημαντικός αριθμός σφαλμάτων που παρουσιάζεται στην μήτρα σύγχυσης της 10^{ης} εποχής πρόκειται για φωτογραφίες οι οποίες αποτελούν δύσκολα κατανοητές απεικονίσεις του φαγητού ή φωτογραφίες με σύνθετο υπόβαθρο. Αξίζει να τονισθεί στο σημείο αυτό ότι οι εικόνες του συνόλου δεδομένων NTUA-Food 2017, δεν έχουν υποβληθεί στην διαδικασία της κατάτμησης προηγουμένως, περιέχοντας θόρυβο στο υπόβαθρό τους. Επιπρόσθετα, οι δύο εικόνες τύπου gif που εμπεριέχονται στο σύνολο εκπαίδευσης, δεν ταξινομήθηκαν σωστά, αν και το MATLAB μπορεί να επεξεργαστεί τέτοιες εικόνες.

Αναλυτικά, στην κατηγορία των Όσπριων και Αμυλούχων Λαχανικών, εμφανίστηκαν 8 σφάλματα στην ταξινόμηση της πατάτας ψητής. Η αιτία εμφάνισης του σφάλματος αυτού, προκύπτει από το γεγονός ότι οι πατάτες ψητές απεικονίζονται με την φλούδα, σε συνδυασμό με το γεγονός ότι είναι ψητές, με αποτέλεσμα να συγχέονται εύκολα με την ομάδα του κρέατος. Λάθος επίσης εμφανίστηκε σε εικόνα με φασολάκια, η οποία κατηγοριοποιήθηκε στην ομάδα του Κρέατος. Το γεγονός αυτό μπορεί να αποδοθεί στην ύπαρξη της κόκκινης σάλτσας τόσο στα φασολάκια όπως και στο κρέας κοκκινιστό.

Όσον αφορά την ομάδα του Κρέατος και Υποκατάστατα, σημειώνονται 4 λάθη στο ζαμπόν, εκ των οποίων 3 κατηγοριοποιούνται στα Γαλακτοκομικά. Αυτό συμβαίνει επειδή η συσκευασία του ζαμπόν με την συσκευασία των τυριών εμφανίζουν σημαντικές ομοιότητες. Ακόμη ένα σύνηθες σφάλμα, είναι η σύγχυση των εικόνων του λουκάνικου με αυτές του hot dog, σφάλμα λογικό αφού το βασικό στοιχείο της εικόνας και στις δύο περιπτώσεις είναι το λουκάνικο. Συνολικά έγιναν 4 τέτοια σφάλματα. Επιπρόσθετα, όπως ήταν αναμενόμενο, τα αυγά και η ομελέτα που αποτελούν υποκατάστατα του κρέατος προκάλεσαν πολλά σφάλματα ταξινόμησης, συγκεκριμένα 8 η ομελέτα και 1 το αυγό. Το μοναδικό σφάλμα του αυγού

είναι σοβαρό, δεδομένου ότι χρησιμοποιήθηκαν συνολικά 3 εικόνες βραστών αυγών στο σύνολο ελέγχου. Έτσι προκύπτει το συμπέρασμα ότι η ομελέτα, τα βραστά αυγά ακόμα και το ζαμπόν που αποτελούν υποκατάστατα του κρέατος, δυσκολεύουν σημαντικά την διαδικασία της ταξινόμησης στην ομάδα αυτή.

Στην κατηγορία των λαχανικών, το καρότο, το κουνουπίδι και οι μελιτζάνες σημείωσαν υψηλά ποσοστά αποτυχίας, εμφανίζοντας αντιστοίχως 8, 11 και 7 λάθη. Το καρότο αποτελεί δύσκολη ομάδα προς ταξινόμηση τόσο εξαιτίας του διαφορετικού χρώματος συγκριτικά με τα άλλα λαχανικά, καθώς επίσης επειδή μπορεί να εμφανιστεί σε διαφορετικές μορφές όπως ολόκληρο, ψιλοκομμένο και βραστό. Η μελιτζάνα στην πλειονότητα των σφαλμάτων κατηγοριοποιήθηκε στην κατηγορία του κρέατος. Η σύγκριση αυτή, είναι λογική, καθώς η πλειονότητα των εικόνων απεικόνιζαν ψητές μελιτζάνες και εικάζεται ότι οι ραβδώσεις που εμφανίζονται πάνω στα ψητά φαγητά, έχουν εκχωρηθεί από το σύστημα ως χαρακτηριστικό της ομάδας του Κρέατος.

Στην κατηγορία των Λιπαρών Φαγητών 5 σφάλματα ταξινόμησης εμφανίζονται για το κρουασάν, ενώ 2 για το παγωτό. Τα 2 σφάλματα του παγωτού, απεικόνιζαν πράσινες μπάλες παγωτού και προβλέφθηκαν ως λαχανικά από τον ταξινομητή. Μπορεί να προκύψει το συμπέρασμα λοιπόν, ότι το πράσινο χρώμα αποτελεί χαρακτηριστικό που προσδιορίζει τα λαχανικά σύμφωνα με τον αλγόριθμο. Επίσης πολλά λάθη εμφανίστηκαν και για τις πατάτες τηγανητές, που αποτελεί το πιο αμφιλεγόμενο φαγητό της ομάδας, καθώς δεν αποτελεί γλυκό. Συγκεκριμένα από τις 5 εσφαλμένες εικόνες οι 3 κατηγοριοποιήθηκαν στην ομάδα των Οσπρίων και Αμυλούχων λαχανικών, στην οποία βρίσκονται τα φαγητά της βραστής και ψητής πατάτας, όπως και του πουρέ. Η σημαντική απόκλιση των πατατών τηγανητών, εγείρει τον προβληματισμό του κατά πόσο πρέπει να τοποθετηθεί στην ομάδα αυτή, ή θα έπρεπε να προταθεί κάποιο άλλο σχήμα ταξινόμησης.

Στα Φρούτα, ο ταξινομητής σημειώνει 7 λάθη στην περίπτωση του ανανά και 3 στην περίπτωση της μπανάνας. Ο Ανανάς συγχέεται με την ομάδα των Οσπρίων και Αμυλούχων λαχανικών. Η αιτία του σφάλματος αυτού μπορεί να ερμηνευθεί εξαιτίας των ομοιοτήτων που παρουσιάζει ο ανανάς όταν είναι καθαρισμένος με τα αμυλούχα λαχανικά πατάτα και καλαμπόκι. Στην περίπτωση της μπανάνας, τα 3 σφάλματα που εμφανίστηκαν βρίσκονται σε εικόνες, στις οποίες η μπανάνα είναι καθαρισμένη και κομμένη, χάνοντας έτσι το χαρακτηριστικό της σχήμα.

Στην τελευταία κατηγορία Ψωμί-Δημητριακά, η τροφή με την μεγαλύτερη ανακρίβεια είναι τα Corn Flakes, με 6 σφάλματα, στα περισσότερα εκ των οποίων η πρόβλεψη του δικτύου ήταν ότι η τροφή ανήκει στην ομάδα των Οσπρίων και Αμυλούχων Λαχανικών. Επίσης κατηγοριοποιήθηκαν και 5 εικόνες λάθος, οι οποίες ήταν στιγμιότυπα του ψωμιού και θεωρήθηκαν ως Breaded Food, σφάλμα λογικό, καθώς και στις δύο περιπτώσεις επικρατεί το ψωμί στην εικόνα.

	Breaded Food	Όσπρια και Αμυλούχα Λαχανικά	Γαλακτοκομικά	Κρέας και Υποκατάστατα	Λαχανικά	Λιπαρά Φαγητά	Φρούτα	Ψωμί Δημητριακά	Άθροισμα κατά γραμμή
Breaded Food	93.83%	1.23%	0.00%	1.85%	0.00%	1.85%	0.00%	1.23%	6.17%
Όσπρια και Αμυλούχα Λαχανικά	2.47%	84.57%	0.00%	5.56%	1.85%	1.23%	1.85%	2.47%	15.43%
Γαλακτοκομικά	0.00%	0.00%	95.68%	0.00%	0.62%	0.00%	0.00%	3.09%	4.32%
Κρέας και Υποκατάστατα	6.17%	3.70%	3.09%	76.54%	3.70%	5.56%	0.00%	1.23%	23.46%
Λαχανικά	1.85%	6.79%	3.09%	2.47%	77.78%	0.00%	2.47%	5.56%	22.22%
Λιπαρά Φαγητά	3.70%	6.17%	0.62%	6.17%	2.47%	77.16%	0.62%	3.09%	22.84%
Φρούτα	0.00%	4.94%	0.00%	0.00%	9.88%	0.62%	83.33%	1.23%	16.67%
Ψωμί Δημητριακά	4.94%	3.70%	0.62%	1.85%	2.47%	1.85%	0.00%	84.57%	15.43%
Άθροισμα κατά στήλη	19.14%	26.54%	7.41%	17.90%	20.99%	11.73%	4.94%	17.90%	

Πίνακας 6.16. Επανζημένη Μήτρα Σύγκυσης της Εποχής 10. Στο Άθροισμα κατά στήλη και κατά γραμμή δεν εμπεριέχονται τα στοιχεία της κυρίας διαγωνίου

Συνοψίζοντας, παρουσιάστηκαν τα σημαντικότερα σφάλματα και οι αδυναμίες του ταξινομητή. Σε κάποιες περιπτώσεις, όπως τα αυγά και η ομελέτα στην ομάδα του κρέατος και οι πατάτες τηγανητές στην ομάδα των Λιπαρών Φαγητών, τίθεται ζήτημα επαναπροσδιορισμού του σχήματος ταξινόμησης.

Για την ακόμα καλύτερη κατανόηση των σφαλμάτων που γίνονται, μπορεί να βοηθήσει η επανζημένη μήτρα σύγκυσης που φαίνεται στον Πίνακας 6.16. Το άθροισμα κατά γραμμή, εκτός από τα στοιχεία της κυρίας διαγωνίου, δείχνει το ποσοστό των σφαλμάτων που έγινε σε κάθε κατηγορία. Με όρους μήτρας σύγκυσης αποτελεί τα Ψευδώς Αρνητικά κάθε κατηγορίας. Από την άλλη μεριά, το άθροισμα κατά στήλη, χωρίς να λαμβάνονται υπόψιν τα κελιά της κυρίας διαγωνίου, υποδεικνύει το άθροισμα των λανθασμένα κατηγοριοποιημένων εικόνων στην στήλη αυτή (Ψευδώς Θετικά). Συνεπώς, οι κατηγορίες με την μεγαλύτερη ακρίβεια είναι το Breaded Food και τα Γαλακτοκομικά, ενώ την χαμηλότερη ακρίβεια σημειώνουν το Κρέας και Υποκατάστατα και τα Λαχανικά. Επίσης μπορεί να βγει το συμπέρασμα ότι η λιγότερα καλά προσδιορισμένη κλάση, είναι τα Όσπρια και Αμυλούχα Λαχανικά, αφού είναι η κλάση με τα περισσότερα Ψευδώς Θετικά. Αυτό υποδεικνύει ότι ίσως θα μπορούσαν να γίνουν κάποιες τροποποιήσεις στην κλάση αυτή, ως προς τα φαγητά που την αποτελούν.

6.2.3 Ταξινόμηση Επιπέδου 2

Στην συνέχεια υλοποιείται η ταξινόμηση δευτέρου επιπέδου, κατά την οποία γίνεται κατηγοριοποίηση των φαγητών κάθε κλάσης. Η ιδέα αυτή προέκυψε καθώς η δημιουργία ταξινομητή σε 82 διαφορετικές κλάσεις, θα ήταν απαγορευτική με την χρήση SVM, ενώ θα ήταν πολύ χρονοβόρα μέσω της μεθόδου του fine-tuning. Στην δεύτερη περίπτωση θα ήταν απαραίτητο, ένα μεγάλο πλήθος εικόνων για κάθε φαγητό. Πέρα από αυτό, δεδομένου, ότι ο ταξινομητής απευθύνεται σε άτομα με ΣΔ, το πρωταρχικό επίπεδο ταξινόμησης είναι το σημαντικότερο, καθώς ομαδοποιεί φαγητά που η κατανάλωση τους επιφέρει παρόμοιες επιπτώσεις στο σάκχαρο.

Για την υλοποίηση του δεύτερου επιπέδου ταξινόμησης δοκιμάζονται και πάλι οι μέθοδοι της εξαγωγής χαρακτηριστικών από ΣΝΔ και η δημιουργία fine-tuned δικτύου. Σημαντική διαφοροποίηση που παρατηρείται σύμφωνα με το πρώτο στάδιο ταξινόμησης, είναι ότι πλέον έχει μειωθεί σημαντικά ο πλήθος

των εικόνων κάθε φαγητού, κάνοντας την διαδικασία της εκπαίδευσης ακόμα πιο δύσκολη λόγω του μικρού πλήθους παρατηρήσεων. Επιπρόσθετα, δεν είναι δυνατή η εύρεση ίσου πλήθους εικόνων από κάθε φαγητό, με αποτέλεσμα να μένει ανεκμετάλλευτο ένα σημαντικό πλήθος εικόνων. Αυτό συμβαίνει γιατί σε περίπτωση που ο ταξινομητής εκπαιδευθεί με ανόμοιο πλήθος εικόνων για κάθε κατηγορία, δίνεται διαφορετική βαρύτητα σε κάθε ομάδα. Επίσης, η ακρίβεια επηρεάζεται περισσότερο από συγκεκριμένα φαγητά, επειδή αποτελούν την πλειονότητα του συνόλου ελέγχου. Χαρακτηριστικό παράδειγμα, φαίνεται στην κατηγορία 2 της Εικόνα 6.10. Το γεγονός ότι επιτυγχάνεται πολύ καλή ακρίβεια στην κατηγορία αυτή, η οποία περιέχει πιο πολλές εικόνες συγκριτικά με τις υπόλοιπες, αυξάνει σημαντικά την ακρίβεια του ταξινομητή, χωρίς να αποτελεί αυτό αντιπροσωπευτική ένδειξη. Για το λόγο αυτό, είναι απαραίτητο κάθε φαγητό να αποτελείται από ίσο πλήθος εικόνων.

Model 2

1	38	1	1	1				
2		147				2		
3			43	3		3		
4			3	43				
5					27	2	3	
6		2	1		3	41	1	
7		1	2		5	1	28	
8		1				1	1	24
	1	2	3	4	5	6	7	8

Predicted class

Εικόνα 6.10. Μήτρα Σύγχυσης κατηγοριών που δεν αποτελούνται από ίσο πλήθος εικόνων

Αρχικά, δοκιμάζεται το ΣΝΔ AlexNet προσαρμοσμένο (fine-tuned) στις 8 κατηγορίες του Breaded Food. Η κατηγορία με το μικρότερο πλήθος εικόνων είναι το Τοστ, το οποίο περιέχει 27 εικόνες. Έτσι επιλέγονται τυχαία 27 εικόνες για κάθε φαγητό. Ένα ακόμα στοιχείο που δυσχεραίνει την ταξινόμηση με την μέθοδο του fine-tuning, είναι ότι το σύνολο των εικόνων διαχωρίζεται σε σύνολο εκπαίδευσης και σύνολο ελέγχου. Με αποτέλεσμα, χρησιμοποιώντας το σχήμα 60-40% ως σύνολα εκπαίδευσης και ελέγχου, να μένουν μόλις 16 εικόνες για εκπαίδευση του δικτύου και 11 εικόνες για τον έλεγχο του ταξινομητή. Ακόμα και στην περίπτωση του fine-tuned ΣΝΔ, το πλήθος αυτό είναι πολύ μικρό και δεν επιτρέπει καλή γενίκευση. Εκπαιδεύοντας το δίκτυο για 5 εποχές, επιτεύχθηκε ακρίβεια 77,27% και η μήτρα σύγχυσης φαίνεται στον Πίνακα 6.17.

5 Εποχές - Fine Tuning	Hot Dog	Pizza	Sandwich	Hamburger	Αραβική Πίτα	Κρέπα	Σουβλάκι	Τοστ
Hot Dog	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pizza	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sandwich	0.00%	9.09%	72.73%	9.09%	9.09%	0.00%	0.00%	0.00%
hamburger	0.00%	0.00%	36.36%	63.64%	0.00%	0.00%	0.00%	0.00%
Αραβική Πίτα	9.09%	0.00%	0.00%	0.00%	54.55%	0.00%	36.36%	0.00%
Κρέπα	0.00%	0.00%	18.18%	0.00%	9.09%	63.64%	9.09%	0.00%
Σουβλάκι	0.00%	9.09%	9.09%	0.00%	9.09%	0.00%	72.73%	0.00%
Τοστ	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	90.91%

Πίνακας 6.17. Μήτρα Σύγχυσης των φαγητών της κατηγορίας του Breaded Food, το σύνολο ελέγχου αποτελείται από 11 κατηγορίες για κάθε τρόφιμο.

Στην συνέχεια, δοκιμάζεται ταξινόμηση μέσω εξαγωγής χαρακτηριστικών από το προ-εκπαιδευμένο AlexNet, ακολουθούμενο από κάποιον ταξινομητή. Πέρα από την καλύτερη γενίκευση που δημιουργούν οι ταξινομητές όπως το SVM, η μέθοδος k-fold validation επιτρέπει την χρήση όλων των παρατηρήσεων να χρησιμοποιηθούν τόσο ως σύνολο εκπαίδευσης όσο και ως σύνολο ελέγχου.

Κρατώντας και πάλι 27 φωτογραφίες από κάθε φαγητό της ομάδας Breaded Food, δοκιμάζεται η ακρίβεια μεταξύ όλων των παραλλαγών του SVM, που διαθέτει η βιβλιοθήκη του MATLAB R2017a. Συγκεκριμένα ως συναρτήσεις πυρήνα χρησιμοποιούνται η γραμμική, η τετραγωνική και η κυβική καθώς ακόμα και η fine Gaussian, η Medium Gaussian και η Coarse Gaussian. Οι καλύτερες ακρίβειες επιτυγχάνονται κυρίως από το τετραγωνικό SVM (Quadratic SVM), κυβικό SVM (Cubic SVM) και μέσο Γκαουσιανό SVM (Medium Gaussian SVM), με μικρές διαφορές ακρίβειας ταξινόμησης μεταξύ τους. Δοκιμάζοντας αυτούς τους ταξινομητές, συμπεραίνεται ότι η τεχνική one-vs-all δίνει καλύτερα αποτελέσματα συγκριτικά με την one-vs-one. Επίσης η μετατροπή των μεταβλητών εισόδου σε γραμμικά ανεξάρτητες μέσω της μεθόδου του Principal Component Analysis (PCA), μειώνει την ακρίβεια. Συνεπώς επιλέγεται ως ταξινομητής του δεύτερου επιπέδου κυβικό SVM με one-vs-all ταξινόμηση. Όπως φαίνεται από τον Πίνακα 6.18, τα πιο ανταγωνιστικά ποσοστά ακρίβειας επιτυγχάνονται από τον τετραγωνικό και κυβικό πολυωνυμικό πυρήνα καθώς και από το Medium Gaussian, από αυτούς επιλέγεται ο κυβικός.

Ταξινομητής	Μέθοδος ταξινόμησης σε πολλές κλάσεις	PCA	Ακρίβεια
Linear SVM	One-vs-One	Απενεργοποιημένο	78.70%
Quadratic SVM	One-vs-One	Απενεργοποιημένο	79.20%
Cubic SVM	One-vs-One	Απενεργοποιημένο	79.20%
Fine Gaussian SVM	One-vs-One	Απενεργοποιημένο	19.40%
Medium Gaussian SVM	One-vs-One	Απενεργοποιημένο	78.70%
Coarse Gaussian SVM	One-vs-One	Απενεργοποιημένο	58.80%
Linear SVM	One-vs-All	Απενεργοποιημένο	81.00%
Quadratic SVM	One-vs-All	Απενεργοποιημένο	83.80%
Cubic SVM	One-vs-All	Απενεργοποιημένο	84.70%
Medium Gaussian SVM	One-vs-All	Απενεργοποιημένο	81.50%
Cubic SVM	One-vs-All	Ενεργοποιημένο	76.40%
Medium Gaussian SVM	One-vs-All	Ενεργοποιημένο	77.80%

Πίνακας 6.18. Ταξινόμηση με διαφορετικές συναρτήσεις πυρήνα SVM στην ομάδα των Breaded Food

6.2.4 Αποτελέσματα

Σύμφωνα με τα παραπάνω ακολουθείται εξαγωγή χαρακτηριστικών από το προτελευταίο επίπεδο του AlexNet και ο πίνακας που προκύπτει τροφοδοτεί ένα cubic one-vs-all SVM. Τα αποτελέσματα για κάθε ομάδα φαγητών φαίνονται στον Πίνακα 6.19, με την μέση ακρίβεια να αγγίζει το 86%.

Κατηγορία	Ακρίβεια	Πλήθος Εικόνων
Breaded Food	84.70%	27
Φρούτα	91.00%	26
Λιπαρά Φαγητά	81.90%	36
Όσπρια και Αμυλούχα Λαχανικά	84.00%	27
Ψωμί και Δημητριακά	86.30%	35
Γαλακτομικά	93.80%	75
Κρέας και Υποκατάστατα	73.40%	33
Λαχανικά	92.40%	21
Μέση Τιμή Ακρίβειας	85.94%	35

Πίνακας 6.19. Ακρίβεια ταξινόμησης δεύτερου επιπέδου

Από τον Πίνακα 6.19, μπορεί να βγει το συμπέρασμα ότι τα φαγητά στις κατηγορίες των Φρούτων, των Γαλακτοκομικών και των Λαχανικών σημειώνουν πολύ υψηλή ακρίβεια ταξινόμησης, καθώς ξεπερνάει το 90%. Ακολούθως τα Breaded Food, Λιπαρά Φαγητά, Όσπρια-Αμυλούχα Λαχανικά και Ψωμί-Δημητριακά σημειώνουν καλή ακρίβεια η οποία κυμαίνεται άνω του 80%. Τέλος, η μόνη ομάδα στην οποία η ταξινόμηση των φαγητών δεν ξεπερνάει το 80% είναι το Κρέας και Υποκατάστατα. Αναλυτικά από τον Πίνακα 6.20 έως τον Πίνακα 6.27 εμφανίζονται οι μήτρες σύγκρισης κάθε κατηγορίας φαγητών, δείχνοντας τα σφάλματα ταξινόμησης που πραγματοποιήθηκαν.

Συγκεκριμένα η καλύτερη ακρίβεια επιτυγχάνεται στα Γαλακτοκομικά προϊόντα, γεγονός που αποδίδεται στο πολύ μικρό πλήθος φαγητών από το οποίο αποτελείται καθώς και στο μεγάλο πλήθος εικόνων για κάθε είδος φαγητού. Αντίστοιχα η ποικιλία στο χρώμα καθώς και το καλά διαμορφωμένο σχήμα που παρουσιάζουν τα φρούτα και τα λαχανικά εξηγούν την πολύ καλή ακρίβεια που παρουσιάζουν.

Τα λιπαρά φαγητά, ενώ δεν είναι απόλυτα προσδιορισμένη ομάδα, παρουσιάζει αποδεκτή ακρίβεια της τάξεως του 82%. Αξιοσημείωτο είναι ότι το κρουασάν και οι πατάτες τηγανητές που ήταν υπεύθυνες για την πλειονότητα των λαθών στο πρώτο επίπεδο ταξινόμησης, σημειώνουν την υψηλότερη ακρίβεια στο δεύτερο επίπεδο ταξινόμησης. Με αποτέλεσμα να συμπεραίνεται ότι τα φαγητά που εμφανίζουν την πιο αποκλίνουσα συμπεριφορά στο πρώτο επίπεδο ταξινόμησης, να είναι πολύ πιο εύκολο ταξινομήσιμα στο δεύτερο επίπεδο. Αυτό συμβαίνει επειδή εμφανίζουν διαφορετικά χαρακτηριστικά από τα τρόφιμα της ίδιας ομάδας.

Όσον αφορά την ομάδα του Κρέατος και Υποκατάστατα, αποτελεί την ομάδα με εμφανώς την χαμηλότερη ακρίβεια. Το πόρισμα που εξήχθη προηγουμένως επιβεβαιώνεται, καθώς τα βραστά αυγά, η ομελέτα και το ζαμπόν ενώ σημείωσαν την μεγαλύτερη απόκλιση στο πρώτο επίπεδο ταξινόμησης, είχαν την καλύτερη ακρίβεια στο δεύτερο. Οι εικόνες των γενικευμένων κατηγοριών χοιρινό και μοσχάρι κρέας σημειώνουν την μεγαλύτερη απόκλιση και αυτό εξηγείται λόγω της έλλειψης σαφούς προσδιορισμού των κλάσεων. Επίσης και η κατηγορία του ψαριού, αν και διαφέρει θεματικά από τα υπόλοιπα κρέατα φαίνεται να μην διακρίνεται εύκολα από το φιλέτο κοτόπουλο, το χοιρινό και την ομελέτα. Όλα τα παραπάνω αναδεικνύουν την δυσκολία ταξινόμησης της κατηγορίας του κρέατος και την ανάγκη σύστασης ενός άλλου τρόπου διαχωρισμού της.

Αναλυτικά η μήτρα σύγκρισης για κάθε ομάδα φαγητών παρουσιάζεται παρακάτω:

	Hot Dog	Pizza	Sandwich	Hamburger	Αραβική Πίτα	Κρέπα	Σουβλάκι	Τοστ
Hot Dog	85.19%	3.70%	7.41%	0.00%	0.00%	0.00%	0.00%	3.70%
Pizza	0.00%	88.89%	0.00%	0.00%	3.70%	7.41%	0.00%	0.00%
Sandwich	0.00%	0.00%	81.48%	11.11%	0.00%	7.41%	0.00%	0.00%
Hamburger	0.00%	0.00%	3.70%	96.30%	0.00%	0.00%	0.00%	0.00%
Αραβική Πίτα	0.00%	0.00%	0.00%	0.00%	85.19%	7.41%	7.41%	0.00%
Κρέπα	0.00%	3.70%	3.70%	0.00%	14.81%	70.37%	0.00%	7.41%
Σουβλάκι	0.00%	7.41%	7.41%	0.00%	11.11%	0.00%	74.07%	0.00%
Τοστ	0.00%	0.00%	0.00%	0.00%	0.00%	3.70%	0.00%	96.30%

Πίνακας 6.20. Μήτρα Σύγκρισης για τα φαγητά της κατηγορίας Breaded Food

	Ανανάς	Αχλάδι	Βερίκοκα	Καρπούζι	Κεράσια	Μήλο	Μπανάνα	Πεπόνι	Πορτοκάλι	Ροδάκινο	Σταφύλια	Φράουλες
Ανανάς	96.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%	0.00%
Αχλάδι	0.00%	80.77%	0.00%	0.00%	0.00%	7.69%	0.00%	7.69%	0.00%	0.00%	3.85%	0.00%
Βερίκοκα	0.00%	0.00%	80.77%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	15.38%	0.00%	0.00%
Καρπούζι	0.00%	0.00%	0.00%	92.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%
Κεράσια	0.00%	0.00%	0.00%	0.00%	96.15%	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%	0.00%
Μήλο	0.00%	7.69%	0.00%	0.00%	3.85%	88.46%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Μπανάνα	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	92.31%	3.85%	0.00%	0.00%	3.85%	0.00%
Πεπόνι	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	92.31%	0.00%	0.00%	0.00%	0.00%
Πορτοκάλι	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%	0.00%	0.00%	96.15%	0.00%	0.00%	0.00%
Ροδάκινο	0.00%	0.00%	3.85%	0.00%	0.00%	3.85%	7.69%	0.00%	0.00%	84.62%	0.00%	0.00%
Σταφύλια	3.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.15%	0.00%
Φράουλες	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	96.15%

Πίνακας 6.21. Μήτρα Σύγκρισης για τα τρόφιμα της κατηγορίας Φρούτα

	Brownie	Cupcakes	Donuts	Βάφλα	Κρουασάν	Μπακλαβάς	Μυλόπιτα	Παγωτό	Πανακότα	Πατάτες Ήλιανθες	Τούρτα Σοκολάτας	Τυραμισού
Brownie	75.00%	0.00%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.89%	8.33%
Cupcakes	2.78%	69.44%	11.11%	0.00%	0.00%	2.78%	0.00%	5.56%	2.78%	0.00%	2.78%	2.78%
Donuts	2.78%	5.56%	75.00%	2.78%	0.00%	2.78%	0.00%	2.78%	5.56%	0.00%	2.78%	0.00%
Βάφλα	0.00%	0.00%	2.78%	75.00%	0.00%	2.78%	8.33%	0.00%	0.00%	8.33%	0.00%	2.78%
Κρουασάν	0.00%	0.00%	0.00%	2.78%	97.22%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Μπακλαβάς	0.00%	0.00%	0.00%	0.00%	0.00%	88.89%	2.78%	2.78%	0.00%	0.00%	2.78%	2.78%
Μυλόπιτα	2.78%	0.00%	2.78%	8.33%	2.78%	5.56%	77.78%	0.00%	0.00%	0.00%	0.00%	0.00%
Παγωτό	0.00%	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	86.11%	5.56%	0.00%	2.78%	0.00%
Πανακότα	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.78%	2.78%	91.67%	2.78%	0.00%	0.00%
Πατάτες Ήλιανθες	0.00%	0.00%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	97.22%	0.00%	0.00%
Τούρτα Σοκολάτας	13.89%	0.00%	5.56%	0.00%	2.78%	0.00%	2.78%	5.56%	0.00%	0.00%	66.67%	2.78%
Τυραμισού	2.78%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.78%	5.56%	83.33%

Πίνακας 6.22. Μήτρα Σύγκρισης για τα φαγητά της κατηγορίας Λιπαρά Φαγητά

	Αρακάς	Γλυκοπατάτα	Κάστανα	Καλαμπόκι	Πατάτα Βραστή	Πατάτα Ψητή	Πουρές Πατάτας	Ρεβίθια	Φάβα	Φακές	Φασολάδα	Φασολάκια
Αρακάς	88.89%	0.00%	0.00%	0.00%	3.70%	0.00%	0.00%	3.70%	0.00%	3.70%	0.00%	0.00%
Γλυκοπατάτα	0.00%	74.07%	3.70%	0.00%	0.00%	14.81%	0.00%	0.00%	0.00%	0.00%	3.70%	3.70%
Κάστανα	0.00%	11.11%	85.19%	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Καλαμπόκι	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Πατάτα Βραστή	3.70%	0.00%	0.00%	0.00%	92.59%	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%	0.00%
Πατάτα Ψητή	0.00%	7.41%	0.00%	0.00%	0.00%	77.78%	3.70%	7.41%	0.00%	0.00%	0.00%	3.70%
Πουρές Πατάτας	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	88.89%	3.70%	7.41%	0.00%	0.00%	0.00%
Ρεβίθια	14.81%	3.70%	3.70%	0.00%	0.00%	0.00%	0.00%	51.85%	7.41%	0.00%	14.81%	3.70%
Φάβα	0.00%	7.41%	0.00%	0.00%	0.00%	0.00%	7.41%	3.70%	77.78%	3.70%	0.00%	0.00%
Φακές	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	92.59%	7.41%	0.00%
Φασολάδα	3.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.70%	92.59%	0.00%
Φασολάκια	0.00%	3.70%	0.00%	0.00%	0.00%	3.70%	0.00%	0.00%	3.70%	0.00%	3.70%	85.19%

Πίνακας 6.23. Μήτρα Σύγκρισης για τα φαγητά της κατηγορίας Όσπρια και Αμυλούχα Λαχανικά

	Αρνί	Αυγό Βραστό	Ζαμπόν	Πουλα Υπόλευτο	Απότοπουλο Φιλέτο	Κρέας Κοκκινιστό	Κρέας Χοιρινό	Λουκάνικο	Μοσχαρίσιο Κρέας	Μπιφτέκι	Μπριζόλα	Ομελέτα	Φιλέτο	Ψάρι
Αρνί	60.61%	0.00%	0.00%	9.09%	0.00%	0.00%	9.09%	0.00%	6.06%	0.00%	3.03%	0.00%	9.09%	3.03%
Αυγό Βραστό	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Ζαμπόν	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Κοτόπουλο Υπόλευτο	6.06%	0.00%	0.00%	81.82%	0.00%	3.03%	6.06%	0.00%	0.00%	3.03%	0.00%	0.00%	0.00%	0.00%
Κοτόπουλο Φιλέτο	3.03%	0.00%	0.00%	3.03%	66.67%	3.03%	3.03%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	15.15%
Κρέας Κοκκινιστό	0.00%	0.00%	0.00%	3.03%	0.00%	87.88%	0.00%	0.00%	3.03%	0.00%	3.03%	0.00%	0.00%	3.03%
Κρέας Χοιρινό	6.06%	0.00%	3.03%	0.00%	6.06%	6.06%	39.39%	3.03%	21.21%	6.06%	3.03%	0.00%	3.03%	3.03%
Λουκάνικο	0.00%	0.00%	6.06%	0.00%	0.00%	6.06%	0.00%	81.82%	0.00%	0.00%	6.06%	0.00%	0.00%	0.00%
Μοσχαρίσιο Κρέας	3.03%	0.00%	0.00%	3.03%	0.00%	6.06%	18.18%	6.06%	51.52%	0.00%	3.03%	0.00%	9.09%	0.00%
Μπιφτέκι	0.00%	0.00%	0.00%	6.06%	3.03%	3.03%	6.06%	0.00%	3.03%	72.73%	3.03%	0.00%	3.03%	0.00%
Μπριζόλα	3.03%	0.00%	0.00%	0.00%	6.06%	3.03%	3.03%	3.03%	6.06%	3.03%	60.61%	0.00%	12.12%	0.00%
Ομελέτα	0.00%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	0.00%	0.00%	0.00%	0.00%	87.88%	0.00%	9.09%
Φιλέτο	3.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.03%	3.03%	6.06%	3.03%	0.00%	81.82%	0.00%
Ψάρι	6.06%	0.00%	0.00%	0.00%	12.12%	0.00%	12.12%	0.00%	0.00%	0.00%	12.12%	0.00%	0.00%	54.55%

Πίνακας 6.24. Μήτρα Σύγκρισης για τα φαγητά της κατηγορίας Κρέας και Υποκατάστατα

	Γάλα	Γιαούρτι	Τυρί
Γάλα	92.00%	6.67%	1.33%
Γιαούρτι	5.33%	93.33%	1.33%
Τυρί	2.67%	1.33%	96.00%

Πίνακας 6.25. Μήτρα Σύγκρισης για τα τρόφιμα της κατηγορίας Γαλακτοκομικά Προϊόντα

	Αγγούρι	Βλίτα	Καρότο	Κολοκυθάκια	Κουνουπίδι	Λάχανο	Μελιτζάνες	Μπρόκολο	Ντομάτες	Παντζάρια
Αγγούρι	95.24%	0.00%	0.00%	4.76%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Βλίτα	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Καρότο	0.00%	0.00%	90.48%	0.00%	0.00%	9.52%	0.00%	0.00%	0.00%	0.00%
Κολοκυθάκια	4.76%	4.76%	0.00%	80.95%	0.00%	0.00%	4.76%	4.76%	0.00%	0.00%
Κουνουπίδι	0.00%	0.00%	0.00%	0.00%	85.71%	4.76%	4.76%	4.76%	0.00%	0.00%
Λάχανο	0.00%	0.00%	0.00%	0.00%	0.00%	95.24%	4.76%	0.00%	0.00%	0.00%
Μελιτζάνες	0.00%	0.00%	4.76%	0.00%	0.00%	0.00%	90.48%	0.00%	0.00%	4.76%
Μπρόκολο	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
Ντομάτες	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	90.48%	9.52%
Παντζάρια	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.76%	95.24%

Πίνακας 6.26. Μήτρα Σύγκρισης για τα τρόφιμα της κατηγορίας Λαχανικά

	Corn Flake	Βρώμη	Κουλούρι Θεσσαλονίκης	Κριθαράκι	Κριταίνα	Μακαρόνια	Παξιμάδια - Φρυγανιές	Ρύζι	Ψωμί Λευκό	Ψωμί Ολικής Αλέσεως
Corn Flake	85.71%	2.86%	0.00%	2.86%	0.00%	2.86%	0.00%	5.71%	0.00%	0.00%
Βρώμη	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Κουλούρι Θεσσαλονίκης	0.00%	0.00%	97.14%	0.00%	2.86%	0.00%	0.00%	0.00%	0.00%	0.00%
Κριθαράκι	2.86%	0.00%	0.00%	80.00%	0.00%	14.29%	2.86%	0.00%	0.00%	0.00%
Κριταίνα	0.00%	5.71%	2.86%	0.00%	88.57%	0.00%	0.00%	0.00%	0.00%	2.86%
Μακαρόνια	5.71%	0.00%	0.00%	22.86%	0.00%	68.57%	2.86%	0.00%	0.00%	0.00%
Παξιμάδια - Φρυγανιές	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	77.14%	0.00%	17.14%	5.71%
Ρύζι	0.00%	0.00%	0.00%	2.86%	0.00%	0.00%	0.00%	97.14%	0.00%	0.00%
Ψωμί Λευκό	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.71%	0.00%	88.57%	5.71%
Ψωμί Ολικής Αλέσεως	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%	0.00%	80.00%

Πίνακας 6.27. Μήτρα Σύγκρισης για τα τρόφιμα της κατηγορίας Ψωμί-Δημητριακά

7 Συμβολή της Εργασίας και Προτεινόμενη Μελλοντική Έρευνα

7.1 Συμβολή της Εργασίας

Η συμβολή της διπλωματικής εργασίας ορίζεται σε τρία επίπεδα. Αρχικά γίνεται μια αναλυτική ανασκόπηση των σημαντικότερων δημοσιεύσεων των τελευταίων χρόνων στον τομέα της ταξινόμησης εικόνων φαγητού από φωτογραφικά στιγμιότυπα, δημιουργώντας έτσι το κατάλληλο υπόβαθρο για την ανάδειξη των state-of-art μεθόδων. Δευτερευόντως σύμφωνα με την state-of-art τεχνική των ΣΝΔ, δημιουργείται ένας καινοτόμος ταξινομητής δύο επιπέδων. Τρίτον, η ταξινόμηση υλοποιείται στο καινούριο σύνολο δεδομένων NTUA-Food 2017, το οποίο έχει προσαρμοστεί καταλλήλως στις διατροφικές ανάγκες των ατόμων με ΣΔ.

Ο ταξινομητής των δύο επιπέδων αποτελεί καινοτόμο προσέγγιση της ταξινόμησης των φαγητών. Έχουν παρουσιαστεί δημοσιεύσεις που εκτελούν ταξινόμηση τόσο σε ευρύτερες κατηγορίες φαγητών, όσο και στο επίπεδο των φαγητών. Επίσης σε πολλές περιπτώσεις πριν την ταξινόμηση των φαγητών έχει προταθεί η χρήση ταξινομητή, που θα κατηγοριοποιεί την εικόνα σε «φαγητό» ή «όχι φαγητό». Σε καμία περίπτωση όμως η ταξινόμηση των δύο επιπέδων δεν εκτελείται πρώτα σε ομάδες φαγητών και στην συνέχεια στα φαγητά της κάθε ομάδας. Η ταξινόμηση των επιπέδων έχει διττό ρόλο. Αρχικά ακολουθεί το σκεπτικό των ατόμων με ΣΔ, που δεν γνωρίζουν αναλυτικά την περιεκτικότητα κάθε φαγητού σε υδατάνθρακες, αλλά την περιεκτικότητα σε ομάδες φαγητών που προκαλούν παρόμοιες μεταγευματικές μεταβολές του σακχάρου. Επιπρόσθετα τα δύο επίπεδα ταξινόμησης συμβάλλουν στην μείωση του υπολογιστικού φορτίου που θα προέκυπτε από την εκπαίδευση ενός ταξινομητή σε 82 διαφορετικές κλάσεις, ενώ παράλληλα επιδιώκεται βελτίωση του ποσοστού ακρίβειας για κάθε επιμέρους κατηγορία.

Συγκεκριμένα, η δημιουργία ταξινομητή για την κατηγοριοποίηση των εικόνων σε παραπάνω από 80 διαφορετικά φαγητά, αποτελεί πολύ σύνθετο και χρονοβόρο έργο. Χωρίς να είναι σαφής η ακρίβεια που μπορεί να επιτευχθεί, καθώς επίσης το πλήθος των εποχών και εικόνων που είναι απαραίτητο. Η διάκριση της εργασίας αυτής σε δύο βήματα, μειώνει σημαντικά την πολυπλοκότητα του καθενός. Παράλληλα όμως στο πρώτο επίπεδο ταξινόμησης επιτυγχάνονται πολύ καλύτερα αποτελέσματα συγκριτικά με την κατηγοριοποίηση σε 82 φαγητά, χωρίς να χάνεται η βασική πληροφορία της περιεκτικότητας του φαγητού σε υδατάνθρακες, που ενδιαφέρει τα άτομα με ΣΔ. Το δεύτερο επίπεδο ταξινόμησης υλοποιείται είτε για μεγαλύτερη ακρίβεια των φαγητών σε υδατάνθρακες ή για εμφάνιση επιπρόσθετων πληροφοριών όπως οι θερμίδες κάθε φαγητού. Έτσι ουσιαστικά ανάγεται η επίλυση ενός σύνθετου προβλήματος, σε δύο σαφώς απλούστερα, μειώνοντας με αυτό τον τρόπο τόσο την χρονική πολυπλοκότητα αλλά και αυξάνοντας την ακρίβεια στην πληροφορία που εξάγεται.

Ένα ακόμα σημαντικό πλεονέκτημα της διαδικασίας ταξινόμησης που προτείνεται είναι ότι στο πρώτο επίπεδο χρησιμοποιείται ταξινόμηση μέσω fine-tuned ΣΝΔ, ενώ στο δεύτερο επίπεδο το ΣΝΔ λειτουργεί ως εξαγωγέας χαρακτηριστικών. Στο πρώτο επίπεδο, με τον τρόπο αυτό, επιτυγχάνεται η βέλτιστη ακρίβεια. Αντιθέτως στο δεύτερο επίπεδο, λόγω του μικρού πλήθους εικόνων επιτρέπεται η εκμετάλλευση όλων τους, μέσω της cross-validation μεθόδου επικύρωσης, ενώ επιτυγχάνεται και καλύτερη γενίκευση μέσω του SVM. Η εκπαίδευση ΣΝΔ με τόσο μικρό πλήθος εικόνων, είναι πολύ πιθανό να οδηγούσε σε περιορισμένα επίπεδα γενίκευσης. Επιπλέον ο συνδυασμός ΣΝΔ, ακολουθούμενο από SVM, έχει σημαντικά πιο γρήγορη απόκριση από την εκπαίδευση ενός βαθύ δικτύου. Άρα, η διαφοροποίηση της μεθόδου ταξινόμησης στα δύο επίπεδα, συμμορφώνεται στα χαρακτηριστικά και στις ανάγκες του συνόλου δεδομένων που χρησιμοποιείται.

Επιπρόσθετα, σημαντικό ενδιαφέρον παρουσιάζει και η μετρική που αναπτύχθηκε για την αξιολόγηση των αποτελεσμάτων του πρώτου επιπέδου ταξινόμησης. Η μετρική αυτή αναδεικνύει ότι στην ταξινόμηση αυτή, δεν έχει σημασία μόνο το πλήθος των σωστά ταξινομημένων εικόνων, αλλά κάθε λάθος φέρει διαφορετική βαρύτητα σφάλματος. Με αποτέλεσμα μια φωτογραφία που ανήκει στην κατηγορία του Κρέατος και Υποκατάστατα όταν ταξινομείται σαν Breaded Food να έχει μεγαλύτερο συντελεστή σφάλματος από ότι ένα φαγητό που ανήκει στα Όσπρια και κατατάσσεται στα Γαλακτοκομικά. Ο συντελεστής σφάλματος, ή αλλιώς η ποινή του κάθε σφάλματος είναι ανάλογη της διαφοράς των ισοδυνάμων υδατάνθρακα που αντιστοιχεί σε κάθε φαγητό.

Τέλος, μια από τις σημαντικότερες συμβολές της εργασίας αυτής είναι το σύνολο δεδομένων NTUA-Food 2017. Αποτελεί το πρώτο διαθέσιμο σύνολο δεδομένων που έχει δημιουργηθεί σύμφωνα με τις διατροφικές ανάγκες των ανθρώπων με ΣΔ. Παράλληλα, εμπεριέχει κάποια φαγητά πλούσια σε υδατάνθρακες και λίπος, που δεν ενδείκνυνται για άτομα με ΣΔ, προκειμένου να συμπεριληφθούν και μη υγιεινές τροφές και να δημιουργηθεί ένα σύνολο δεδομένων που αντιπροσωπεύει με τον καλύτερο τρόπο τις διατροφικές συνήθειες των ανθρώπων σε όλο τους το φάσμα.

7.2 Προτάσεις για περαιτέρω Έρευνα

Η περαιτέρω έρευνα μπορεί να γίνει σχετικά με την βελτίωση του ταξινομητή ή με τον εμπλουτισμό του συνόλου δεδομένων NTUA-Food 2017.

- Αρχικά θα ήταν καλό να υλοποιηθεί ένα σύνολο δεδομένων με ίσο πλήθος εικόνων από κάθε φαγητό. Στο πλαίσιο της τρέχουσας εργασίας από τις 4273 εικόνες που συλλέχθηκαν, οι 3248 χρησιμοποιήθηκαν στο πρώτο επίπεδο ταξινόμησης, ενώ μόλις 2531 εικόνες

χρησιμοποιήθηκαν για το δεύτερο επίπεδο ταξινόμησης. Προκειμένου να δοθεί ίδια βαρύτητα κατά την διάρκεια της ταξινόμησης δεν αξιοποιείται σημαντικό πλήθος φωτογραφιών. Το σημαντικότερο λόγο για αυτό το πρόβλημα αποτελεί το γεγονός ότι δεν είναι εύκολη η εύρεση καλής ποιότητας φωτογραφίες για κάποιες κατηγορίες τροφών.

- Επιπρόσθετα ίσως το σημαντικότερο μειονέκτημα του συνόλου δεδομένων NTUA – Food 2017, είναι ότι οι εικόνες δεν έχουν υποβληθεί στη διαδικασία της κατάτμησης. Έτσι εμπεριέχεται θόρυβος στην κάθε φωτογραφία, που κανονικά θα έπρεπε να έχει εξαλειφθεί δεδομένου ότι πρώτα ακολουθεί το στάδιο της κατάτμησης. Παρ' όλα αυτά, οι φωτογραφίες του συνόλου δεδομένων έχουν επιλεγεί καταλλήλως έτσι ώστε να μην περιέχουν πολύ θόρυβο.
- Κατά τη διάρκεια της ταξινόμησης και της περιγραφής του συνόλου δεδομένων παρουσιάστηκαν αμφιβολίες σχετικά με το αν κάποια φαγητά ανήκουν σε κάποιες ευρύτερες ομάδες. Αυτό συμβαίνει είτε επειδή τα μακροθρεπτικά τους συστατικά διαφέρουν σημαντικά από όλα τα φαγητά της ομάδας ή επειδή σημειώνουν μεγάλη απόκλιση στο πρώτο επίπεδο ταξινόμησης. Χαρακτηριστικά παραδείγματα είναι η ταξινόμηση της πατάτας τηγανητής στα Λιπαρά Φαγητά, της ομελέτας στο Κρέας και Υποκατάστατα ή ακόμα και ο τρόπος που η ευρύτερη κατηγορία των Ψωμί-Δημητριακά-Όσπρια-Αμυλούχα Λαχανικά διαχωρίστηκε σε δύο. Συνεπώς διατηρώντας τον πυρήνα του συνόλου δεδομένων σταθερό αναζητάται η δημιουργία παραλλαγών του NTUA-Food 2017 που θα επιφέρει καλύτερα αποτελέσματα χωρίς να αποκλίνει σημαντικά από τον κεντρικό σκοπό της εργασίας.
- Όσον αφορά την διαδικασία της ταξινόμησης θεωρείται απαραίτητο να υπολογιστεί η ακρίβεια που θα έχει το δίκτυο, αν αντί για την ταξινόμηση δύο επιπέδων επιτελούνταν ταξινόμηση απευθείας σε 82 διαφορετικά φαγητά, με σκοπό να συγκριθούν τα αποτελέσματα ακρίβειας και να φανεί αν μέσω της μεθόδου που προτάθηκε επιτυγχάνεται βελτιστοποίηση μόνο σε επίπεδο χρόνου ή και σε επίπεδο αποτελεσματικότητας. Επίσης το εύρος της διαφοράς στα επίπεδα ακρίβειας είναι πολύ σημαντικό. Κατά μέσο όρο μπορεί να θεωρηθεί ότι η ακρίβεια ανά φαγητό του παρόντος ταξινομητή είναι $(0,8418 * 0,8594 = 0,7234)72,34\%$, δεδομένου ότι το βέλτιστο σχήμα ταξινόμησης στο πρώτο επίπεδο είχε ακρίβεια 84,18% και η μέση ακρίβεια μεταξύ των διαφόρων ομάδων του δεύτερου επιπέδου είναι 85,94%.
- Επίσης, μέσω της έκδοσης του MATLAB R2017b, υπάρχει η δυνατότητα ταξινόμησης ΣΝΔ χρησιμοποιώντας σύνολο επικύρωσης (validation set), επιτρέποντας έτσι την εύρεση του βέλτιστου πλήθους εποχών εκπαίδευσης.
- Επιπλέον, αν και το AlexNet είναι νικητής του διαγωνισμού ILSVRC 2012, στην πορεία έχουν αναπτυχθεί ισχυρότερα ΣΝΔ και με διαχειρίσιμο βάθος όπως το GoogleNet. Το GoogleNet,

εμπεριέχεται και αυτό στην έκδοση του MATLAB R2017b. Συνεπώς δίνεται η ευκαιρία επίτευξης υψηλότερων επιπέδων ακρίβειας, χρησιμοποιώντας πιο εξελιγμένα ΣΝΔ.

- Το δίκτυο AlexNet, στην μορφή που παρέχεται από το MATLAB, έχει προεκπαιδευθεί σε μια υποκατηγορία 1000 κλάσεων του συνόλου δεδομένων ImageNet. Οι κλάσεις αυτές είναι γενικού σκοπού, με αποτέλεσμα να εξάγονται σημαντικά χαρακτηριστικά αλλά ένα μέρος τους, ιδίως αυτά των υψηλότερων επιπέδων, να μην σχετίζονται με φαγητά. Για τον λόγο αυτό, θα μπορούσε το δίκτυο AlexNet να προεκπαιδευτεί από ένα σύνολο δεδομένων που μια μερίδα των εικόνων θα περιείχε φωτογραφίες φαγητών. Με τον τρόπο αυτό, ήδη από το επίπεδο της προ-εκπαίδευσης θα έχει δημιουργηθεί μια σημαντική κλίση του ταξινομητή στην αναγνώριση φαγητών. Επειδή η προεκπαίδευση είναι προκαθορισμένη στους ταξινομητές που παρέχει το MATLAB, μια εναλλακτική θα ήταν κατά την διάρκεια της εκπαίδευσης το δίκτυο να εκπαιδευτεί από ένα μεγαλύτερο σε πλήθος εικόνων σύνολο δεδομένων, το οποίο θα είχε μη αμελητέους συντελεστές ρυθμού εκπαίδευσης στα επίπεδα συνέλιξης. Ή ακόμα καλύτερα, να τροποποιείται σημαντικά μόνο ο ρυθμός εκπαίδευσης των τελευταίων συνελικτικών επιπέδων, στα οποία γίνεται η σύνθεση υψηλού επιπέδου χαρακτηριστικών.

8 Παραρτήματα

8.1 Παράρτημα Α: Πίνακας Τεχνικών Εξαγωγής Χαρακτηριστικών

Χαρακτηριστικό	Υποκατηγορία
Color Histograms	
	the histogram of the 1024 most dominant food colors (k-means)
	HistRGB
	HistOp
	HistRGnorm
	HistHue
	HistRGBtrans
	HistHSV
Generalized color moments	
Color moment invariants	
SIFT	
	Dense SIFT
	SIFT with sparse coding
Color SIFT variants	
	rgbSIFT
	hsvSIFT
	hueSIFT
	opponentSIFT
	cSIFT
	rgSIFT
	Red-SIFT
	Green-SIFT
	Blue-SIFT
SURF	
	Dense SURF
LBP	
	PRICoLBP
	non uniform invariant Local Binary Pattern on the RGB channels (LBP)
	the 256 values of the local binary pattern (LBP) histogram
(Bag of)Textons	
	Leung-Malik filters bank
	Maximum Response 8 (MR8) filter bank
	Maximum Response 4 (MR4) filter bank
	Schmid filters
	Local Intensity-normalized Colors (LINC) filter banks
Anti-Textons	
Gabor features/filter	
	Gabor-based image decomposition and Fractal Dimension estimation (GFD)
Opponent Gabor Feature	
Local Color Contrast (LCC)	
Chromaticity moments (CM)	
Complex wavelet features(CWT)	
Color and Directivity Descriptor (CEDD)	
HOG	
	RootHoG (square root Histogram of oriented Gradients)
	Dense HoG
Color patch descriptor	
	Color patches - mean and variance values RGB
	L*a*b color values for each superpixel
	RGB-HSV color values
Color statistics	
Gradient Orientation Spatial-Dependence Matrix (GOSDM)	
Entropy-based categorization and Fractal Dimension estimation (EFD)	
Steerable Filters	
GIH (global ingredient histogram)	
Hough lines - Hough transform	
Hu moments	
Color Correlogram	

Haar wavelet	
Tamura	
DAISY descriptor	
Pairwise Features	
	Distance
	Orientation
	Midpoint Category
	Between-pair Category
GIR-STF (Global Ingredient Representation based on statistics of food ingredients collected after semantic segmentation with STF- Semantic Textons Forests)	

Πίνακας 8.1. Αναλυτική παρουσίαση όλων των τεχνικών εξαγωγής χαρακτηριστικών που συναντήθηκαν στην διπλωματική εργασία. Με πράσινο χρώμα σημειώνονται οι πιο γνωστές μέθοδοι εξαγωγής χαρακτηριστικών

8.2 Παράρτημα Β: Πίνακας Σύγκρισης Σημαντικότερων Δημοσιεύσεων

Παραπομπή	Σύνολο Δεδομένων	Εξαγωγή χαρακτηριστικών	Μείωση Χαρακτηριστικών	Ταξινόμητής	Ακρίβεια	Ημερομηνία
(Ciocca, Napoletano, & Schettini, 2016)	UNIMIB2016	LBP Color and Edge Directivity Descriptor(CEDD) Hist (Color Histogram) Gabor Opponent Gabor Features(OG) Local Color Contrast (LCC) Chromaticity Moments (CM) Complex Wavelet Features (CWT) CNN128 CNN4096 Bag of Convolutional Filter Responses (BoCFR)		k-Nearest Neighbor Classifier	82.00%	2016
(Martinel, Piciarelli, Micheloni, & Foresti, 2015)	UNICT-FD889	<u>Texture Filters:</u> Gabor filters Schmid filters Leung-Malik (LM) Maximum Response 8 (MR8) filter bank		Random Forest classification	42.60%	2015
(Christodoulidis, Anthimopoulos, & Mougialakou, 2015)	573 food items categorised into: Pasta, Potatoes, Meat, Breaded Food, Rice, Green Salad and Carrots			CNN with 6 layers combined with voting scheme	84.90%	2015
(Kawano, & Yanai, 2015)	UEC-FOOD100 UEC-FOOD256 Food 101	Color Root HOG Color + HOG DCNN (4096) DCNN-FOOD (6144) FV+DCNN FV+DCNN-FOOD	Fisher Vector	DCNN(ft) DCNN-FOOD(ft) DCNN(ft2) DCNN-FOOD(ft2) Random Forest (RF)	78.77%(UEC-FOOD 100) 67.57%(UEC-FOOD 256) 70.41%(Food-101)	2015
(Farinella, Allegra, & Stanco, 2015)	UNICT-FD889	Bag of Textons PRICoLBP SIFT		Near Duplicate Image Retrieval (NDIR)	67.50%	2015
(Ao, & Ling, 2015)	Food 101	GoogleNet (deep feature extractor)		GoogleNet AlexNet RFDC MLDS	78.11%(Food-101)	2015
(Anthimopoulos, Gianola, Scarnato, Diem, & Mougialakou, 2014)	Diabetes	Color Histograms (HistRGB, HistOp, HistRGnorm, HistHue, and HistRGBtrans) Generalized color moments Color moment invariants SIFT Color SIFT variants	BoF (k-means, hk-mean clustering)	SVM Artificial Neural Network (ANN) Random Forests (RF)	78.00%	2014
(Baxter, 2014)	PFID	RGB Histogram GIH (global ingredient histogram) Pairwise Features		SVM	40.70%(88 Categories) 81.20%(7 Categories)	2014

(Farinella, Moltisanti, & Battiato, 2014)	PFID (Pittsburgh's Fast food Image Dataset)	Textons vs Chance Color D (Distance) M (Midpoint Category) DO (Distancer and Orientation) BoW SIFT GIR SIFT O (Orientation) B (Between Pair Category) OM (Orientation and Midpoint Category)	Bag of Features (Textons)	SVM	31.30% (per food) 79.60% (7 classes)	2014
(Kagaya, Aizawa, & Ogawa, 2014)	Dataset from Food Log (FL), 170,000 images, 10 food categories			CNN compared with: SPM (Spatial Pyramid Method) + Color + SVM GIST + SVM ScSPM (SIFT-BoW)	73.70%	2014
(Bossard, Guillaumin, & Van Gool, 2014)	Food 101	HOG BOW SURF Improved Fisher Vector (IFV)		multiclass SVM Random Forest Classification (RF) Randomized Clustering Forests (RCF) Mid-Level Discriminative Superpixels (MLDS) Convolutional Neural Networks (CNN)	56.40%	2014
(Kawano & Yanai, 2013)	100-category food dataset (100 images per category) 12,905 images totally	HOG Patch Descriptor Color Patch descriptor SURF	Fisher Vector vs BoF	linear SVM (low computational cost) one vs rest strategy MKL-SVM	51.90%	2013
(Bosch, Zhu, Khanna, Boushey, & Delp, 2011)	2 datasets from 2 user studies (totally 39 foods from 179 images)	Local: Color, Entropy, Tamuna, Gabor, SIFT, Haar Wavelet, Steerable, DAISY Global: Color, Predominant, Entropy, Gabor	Bag of Features (BoF)	Global Features: SVM - Radial Basis Function (RBF) kernel Local Features: nearest neighbor	79.20%	2011
(Yang, Chen, Pomerlau, & Sukthankar, 2010)	PFID (Pittsburgh's)	Pairwise Feature: D (Distance) O (Orientation) M (Midpoint) B (Between-Pair) DO (Distance-Orientation) OM (Orientation-Midpoint) Global Ingredient Representation (GIR) vs Baseline Methods Color SIFT	Histogram	SVM (χ^2 kernel)	28.20%	2010

(Hoahi, Joutou, & Yanai, 2010)	85 categories of Japanese foods, with 100 images per category (variation of UEC-FOOD)	SIFT Color Histogram Gabor Features Gradient Histogram	BoF (SIFT)	MKL(Multiple Kernel Learning)-SVM	62.52%	2010
(Joutou, Yanai, 2010)	50 categories of Japanese foods , with 100 images per category (variation of UEC-FOOD)	SIFT Color Histogram Gabor textrue features	BoF (SIFT)	Multiple Kernel Learning (MKL)-SVM	61.34%	2010

Πίνακας 8.2. Παρουσίαση των σημαντικότερων δημοσιεύσεων και των σχημάτων ταξινόμησης που ακολουθούν. Στην στήλη της ακρίβειας αναγράφεται η βέλτιστη ακρίβεια που επιτυγχάνεται σε κάθε δημοσίευση, ενώ με **bold** εμφανίζεται η τεχνική η οποία σημειώνει την ακρίβεια αυτή.

9 Βιβλιογραφία

- Ahmad, Z., Khanna, N., Kerr, D. A., Boushey, C. J., & Delp, E. J. (2014). A mobile phone user interface for image-based dietary assessment. *In Proceedings of the IS&T/SPIE Conference on Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications*.
- Aleksander, & Morton. (1989). *Neural computing architectures: the design of brain-like machines*.
- Anthimopoulos, M. M., Gianola, L., Scarnato, L., Diem, P., & Mougiakakou, S. G. (2014). A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *IEEE Journal of Biomedical and Health Informatics*, 18(4), 1261-1271.
- Anthimopoulos, M., Dehais, J., Diem, P., & Mougiakakou, S. (2013). Segmentation and recognition of multi-food meal images for carbohydrate counting. *In 2013 IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE)* (σσ. 1-4). IEEE.
- Anthimopoulos, M., Dehais, J., Shevchik, S., Ransford, B. H., Duke, D., Diem, P., & Mougiakakou, S. (2015). Computer Vision-Based Carbohydrate Estimation for Type 1 Patients With Diabetes Using Smartphones. *Journal of Diabetes Science and Technology*, 9(3), 507-515.
- Ao, S., & Ling, C. X. (2015). Adapting New Categories for Food Recognition with Deep Representation. *Data Mining Workshop (ICDMW), 2015 IEEE International Conference*. Atlanta City: IEEE.
- Asvestas, P., Matsopoulos, G., & Nikita, K. (1999, March). Estimation of fractal dimension of images using a fixed mass approach. *Pattern Recognition Letters*, 20(3), 347-354.
- Bajwa, I. S., Naweed, M. S., Asif, M. N., & Hyder, S. I. (2009). Feature Based Image Classification by using Principal Component Analysis. *ICGST-GVIP Journal*, 9(2).
- Baxter, J. (2014). *Food Recognition using Ingredient-Level Features*. Massachusetts Institute of Technology.
- Bejibom, O., Joshi, N., Morris, D., Saponas, S., & Khullar, S. (2015). Menu-Match: Restaurant-Specific. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV 2015)* (σσ. 844-851). IEEE.
- Bosch, M., Zhu, F., Khanna, N., Boushey, C. J., & Delp, E. J. (2011). Combining global and local features for food identification in dietary assessment. *Image Processing (ICIP), 2011 18th IEEE International Conference*. doi:10.1109/ICIP.2011.6115809

- Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101 – Mining Discriminative Components with Random Forests. *European Conference on Computer Vision*.
- Britz, D. (2015, November 7). *Understanding Convolutional Neural Networks for NLP*. Ανάκτηση από WILDML: <http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/>
- Caragea, D., Cook, D., & Honavar, V. (2005). Visual Methods for Examining Support Vector Machine Results, with Applications to Gene Expression Data Analysis.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017, May 12). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *ArXiv e-prints*. Ανάκτηση από <https://arxiv.org/pdf/1606.00915.pdf>
- Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R., & Yang, J. (2009). PFID: PITTSBURGH FAST-FOOD IMAGE DATASET. *Conference: Proceedings of the International Conference on Image Processing, ICIP*. Cairo.
- Chen, Yang, Ho, Wang, Liu, Chang, . . . Ouhyoung. (2012). Automatic Chinese food identification and quantity estimation. *SIGGRAPH Asia 2012 Technical Briefs*. ACM New York.
- Christodoulidis, S., Anthimopoulos, M., & Mougiakakou, S. (2015). Food Recognition for Dietary Assessment Using Deep Convolutional Neural Networks. *New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops* (σσ. 458-465). Springer, Cham.
- Ciocca, G., Napoletano, P., & Schettini, R. (2016). Food recognition for dietary monitoring: a new dataset, experiments and results. *IEEE Journal of Biomedical and Health Informatics*, 21(3), 588-598.
- Ciocca, G., Napoletano, P., & Schettini, R. (2017). Food recognition: a new dataset, experiments and results. *IEEE Journal of Biomedical and Health Informatics*, 21(3), 588-598. doi:10.1109/JBHI.2016.2636441
- Cord, M. (2016). Deep learning and weak supervision for image classification.
- Dalakeidi, K., Sarantea, M., & Nikita, K. S. (2017). A Modified All-And-One Classification Algorithm Combined with the Bag-Of-Features Model to Address the Food Recognition Task. *10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017)*, 5, σσ. 284-290. Porto.
- Deshpande, A. (2016, August 24). *The 9 Deep Learning Papers You Need To Know About (Understanding CNNs Part 3)*. Ανάκτηση από <https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>

- Desphande, A. (2016, July 20). *A Beginner's Guide To Understanding Convolutional Neural Networks*. Ανάκτηση από <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>
- Devlin, M. (2007). *BIOXHMEIA Κλινικοί Συσχετισμοί* (Τόμ. ΙΙ). Αθήνα: Π.Χ.Πασχαλίδης Ε.Π.Ε.
- Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2013). DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *CoRR*.
- Duda, R., & Hart, P. (1973). *Pattern Classification and scene analysis*. Wiley. New York.
- Ecole Polytechnique Fédérale de Lausanne (EPFL). (χ.χ.). *Food Image Dataset*. Ανάκτηση από <https://mmsp.epfl.ch/food-image-datasets>
- Farinella, G. M., Allegra, D., & Stanco, F. (2015). A Benchmark Dataset to Study the Representation of Food Images. *International Workshop on Assistive Computer Vision and Robotics (ACVR) 2014*.
- Farinella, G. M., Moltisanti, M., & Battiato, S. (2014). Classifying food images represented as bag of textons. *Image Processing (ICIP), 2014 IEEE International Conference on* (σσ. 5212-5216). IEEE.
- Fergus, R. (2015). *Neural Network*.
- Fioravanti, A., Fico, G., Patón, A. G., Leuteritz, J.-P., Arredondo, A. G., & Waldmeyer, M. T. (2014). Health-Integrated System Paradigm: Diabetes Management. Στο Κ. S. Nikita, & Κ. S. Nikita (Επιμ.), *Handbook of Biomedical Telemetry* (σσ. 623–632). Wiley.
- Foodspotting. (2017). *Find Dishes*. Ανάκτηση από <http://www.foodspotting.com/find/in/The-World>
- Haykin, S. (2010). *Νευρωνικά Δίκτυα και Μηχανική Μάθηση*. (Ε. Γκαγκάτσιου, Μεταφρ.) Αθήνα: Παπασωτηρίου.
- He, Y., Khanna, N., Boushey, C. J., & Delp, E. J. (2012). Specular Highlight Removal for Image-Based Dietary Assessment. *In 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (σσ. 424-428). IEEE.
- He, Y., Khanna, N., Boushey, C. J., & Delp, E. J. (2013). Image segmentation for image-based dietary assessment: A comparative study. *In 2013 International Symposium on Signals, Circuits and Systems (ISSCS)* (σσ. 1-4). IEEE.
- hilda0707. (2014, October). Ανάκτηση από ZHIYAOBME.
- Hoashi, H., Joutou, T., & Yanai, K. (2010). Image Recognition of 85 Food Categories by Feature Fusion. *Multimedia (ISM), 2010 IEEE International Symposium*. doi:10.1109/ISM.2010.51

- International Diabetes Federation. (2015). *IDF Diabetes Atlas*. International Diabetes Federation.
- Joutou, T., & Yanai, K. (2010). A food image recognition system with Multiple Kernel Learning. *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE. doi:10.1109/ICIP.2009.5413400
- Kagaya, H., Aizawa, K., & Ogawa, M. (2014). Food Detection and Recognition Using Convolutional Neural Networks. *Proceedings of the 22nd ACM international conference on Multimedia*, (σσ. 1085-1088).
- Karpathy, A. (2017). *Convolutional Neural Networks*. Ανάκτηση από CS231n Convolutional Neural Networks for Visual Recognition: <http://cs231n.github.io/convolutional-networks/>
- Karpathy, A. (2017). *Neural Networks Part 1: Setting up the Architecture*. Ανάκτηση από CS231n: Convolutional Neural Networks for Visual Recognition: <http://cs231n.github.io/neural-networks-1/>
- Karpathy, A. (2017). *Neural Networks Part 3: Learning and Evaluation*. Ανάκτηση από CS231n: Convolutional Neural Networks for Visual Recognition: <http://cs231n.github.io/neural-networks-3/>
- Kawano, & Yanai. (2015). Food image recognition using deep convolutional network with pre-training and fine-tuning. *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference*. Turin: IEEE. doi:10.1109/ICMEW.2015.7169816
- Kawano, Y., & Yanai, K. (2013). Rapid Mobile Object Recognition Using Fisher Vector. *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference*. IEEE. doi:10.1109/ACPR.2013.39
- Kawano, Y., & Yanai, K. (2014). Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation. Στο *Proc. of ECCV Workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV)*.
- Kawano, Y., & Yanai, K. (2014). Food image recognition with deep convolutional features. *UbiComp 2014*, (σσ. 589-593). doi:10.1145/2638728.2641339
- Kim, E. (2013, September 1). *Everything You Wanted to Know about the Kernel Trick*. Ανάκτηση από eric-kim: [http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html#\[1\]](http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html#[1])
- Kim, E. (2013, September 1). *Everything You Wanted to Know about the Kernel Trick*. Ανάκτηση από eric-kim: [http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html#\[5\]](http://www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html#[5])
- Kreutzer, J., DeLuca, J., & Caplan, B. (2010). *Encyclopedia of Clinical Neuropsychology*. Springer.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional. *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems. 1*, σσ. 1097-1105. Nevada: Curran Associates Inc.
- M., F. G., Allegra, D., Moltisanti, M., Stanco, F., & Battiato, S. (2016). Retrieval and Classification of Food Images. *Computers in Biology and Medicine*, 77, 23-29.
- Martin, C., Kaya, S., & Gunturk, B. (2009). Quantification of food intake using food image analysis. *Conf Proc IEEE Eng Med Biol Soc, 2009*, 6869-6872.
- Martinel, N., Piciarelli, C., Micheloni, C., & Foresti, G. L. (2015). On filter Banks of texture features for mobile food classification. *Proceedings of the 9th International Conference on Distributed Smart Cameras*, (σσ. 14-19). Seville.
- MathWorks. (χ.χ.). *Image Classification with Bag of Visual Words*. Ανάκτηση από <https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html>
- MathWorks. (χ.χ.). *Measuring Planar Objects with a Calibrated Camera*. Ανάκτηση από <https://www.mathworks.com/help/vision/examples/measuring-planar-objects-with-a-calibrated-camera.html>
- Matsuda, Y., Hoashi, H., & Yanai, K. (2012). Recognition of Multiple-Food Images by Detecting Candidate Regions. *Multimedia and Expo (ICME), 2012 IEEE International Conference*. IEEE. doi:10.1109/ICME.2012.157
- Merler, M., Wu, H., Uceda-Sosa, R., Nguyen, Q.-B., & Smith, J. R. (2016). Snap, Eat, RepEat: A Food Recognition Engine for Dietary Logging. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (σσ. 31-40). Amsterdam: ACM.
- Miyazaki, T., De Silva, G. C., & Aizawa, K. (2011). Image-based Calorie Content Estimation for Dietary Assessment. *Multimedia (ISM), 2011 IEEE International Symposium*. IEEE.
- Mougiakakou, S., Valavanis, I. K., Nikita, K., & Kelekis, D. (2003). Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. 2*. Cancun: IEEE.
- Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., . . . Murphy, K. (2015). Im2Calories: towards an automated mobile vision food diary. *International Conference on Computer Vision (ICCV)*. Ανάκτηση από <http://www.cv->

foundation.org/openaccess/content_iccv_2015/papers/Meyers_Im2Calories_Towards_an_ICCV_2015_paper.pdf

- Neel. (2015, March 3). *Training Neural Networks with Genetic Algorithms*. Ανάκτηση από <https://blog.abhranil.net/2015/03/03/training-neural-networks-with-genetic-algorithms/>
- Neelamegam, P., Abirami, S., Vishnu Priya, K., & Rubalya Valantina, S. (2013). Analysis of rice granules using image processing and neural network. *In IEEE Conference on Information & Communication Technologies (ICT)* (σσ. 879–884). IEEE.
- Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.
- Nikita, K. S. (2014). *Handbook of Biomedical Telemetry*. Wiley.
- Normalization in Neural Networks*. (2016, August 3). Ανάκτηση από <http://yeephycho.github.io/2016/08/03/Normalizations-in-neural-networks/>
- Noronha, J., Hysen, E., Z. H., & Gajos, K. Z. (2011). PlateMate: Crowdsourcing Nutrition Analysis from Food. *ACM 2011*, (σσ. 1-12).
- Ophardt, C. E. (2003). *VIRTUAL CHEMBOOK*. Ανάκτηση από Glycogenesis, Glycogenolysis and Gluconeogenesis: <http://chemistry.elmhurst.edu/vchembook/604glycogenesis.html>
- Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Networks, Learning and Transferring Mid-level Image Representations Using Convolutional Neural. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference*. IEEE.
- Papageorgiou, C., Liappas, I., Asvestas, P., Vasios, C., Matsopoulos, G., Nikolaou, C., . . . Rabavilas, A. (2001). Abnormal P600 in heroin addicts with prolonged abstinence elicited during a working memory test. *Neuroreport*, *12*, 1773-1778.
- Patel, H. N., Jain, R. K., & Joshi, M. V. (2012). Automatic Segmentation and Yield Measurement of. *International Journal of Computer Applications*, *45*(7), 19-24.
- Pouladzadeh, P., Shirmohammadi, S., & Al-Maghrabi, R. (2014, February 25). Measuring Calorie and Nutrition From Food Image. *IEEE Transactions on Instrumentation and Measurement*, *63*(8), 1947-1956.
- Pouladzadeh, P., Shirmohammadi, S., & Yassine, A. (2014). Using Graph Cut Segmentation for Food. *In IEEE International Symposium on Medical Measurements and*, (σσ. 1–6).

- Pouladzadeh, P., Yassine, A., & Shirmohammadi, S. (2015). FooDD: Food Detection Dataset for Calorie Measurement Using Food Images. *International Conference on Image Analysis and Processing 2015*, (σσ. 441-448).
- Puri, M., Zhu, Z., Yu, Q., Divakaran, A., & Sawhney, H. (2009). Recognition and Volume Estimation of Food Intake using a Mobile Device. *Applications of Computer Vision (WACV), 2009 Workshop*. Snowbird: IEEE.
- Raschka, S. (2015, March 24). *Single-Layer Neural Networks and Gradient Descent*. Ανάκτηση από http://sebastianraschka.com/Articles/2015_singlelayer_neurons.html
- Raschka, S. (χ.χ.). *Machine Learning FAQ*. Ανάκτηση από <https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html>
- Rhyner, D., Loher, H., Dehais, J., Anthimopoulos, M., Shevchik, S., Botwey, R., . . . Mougiakakou, S. (2016, May 11). Carbohydrate Estimation by a Mobile Phone-Based System Versus Self-Estimations of Individuals With Type 1 Diabetes Mellitus: A Comparative Study. *J Med Internet Res*, 18(5), 101.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *CoRR*.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3), 211-252. doi:10.1007/s11263-015-0816-y
- Satya, M. (2016, December 6). *Histogram of Oriented Gradients*. Ανάκτηση από LearnOpenCV: <https://www.learnopencv.com/histogram-of-oriented-gradients/>
- Schlag, I. (2016, April 5). *IMPORTANT ILSVRC ACHIEVEMENTS FROM 2012-2015*. Ανάκτηση από <http://ischlag.github.io/2016/04/05/important-ILSVRC-achievements/>
- Schneider, J. (1997). *Cross Validation*. Ανάκτηση από <https://www.cs.cmu.edu/~schneide/tut5/node42.html>
- Simonyan, K., & Zisserman, A. (2015). VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. *International Conference on Learning Representations (ICLR)*.
- Smart, C., King, B., McElduff, P., & Collins, C. (2012, Jul). In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia. *Diabet Med*, 29(4), 21-24.

- Smart, C., Ross, K., Edge, J., Collins, C., Colyvas, K., & King, B. (2009, March). Children and adolescents on intensive insulin therapy maintain postprandial glycaemic control without precise carbohydrate counting. *Diabet Med.*, 26(3), 279-285.
- Souza, C. (2010, March 17). *Kernel Functions for Machine Learning Applications*. Ανάκτηση από crsouza: <http://crsouza.com/2010/03/17/kernel-functions-for-machine-learning-applications/>
- Stanford. (χ.χ.). *Neural Networks*. Ανάκτηση από <https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Neuron/index.html>
- Sun, D., & Du, C. (2004). Segmentation of complex food images by stick growing and merging algorithm. *Journal of Food Engineering*, 61(1), 17-26.
- Sun, D.-W., & Du, C.-J. (2004). Segmentation of complex food images by stick growing and merging algorithm. *Journal of Food Engineering*, 61(1), 17-26.
- Wikipedia. (χ.χ.). *Support Vector Machine*. Ανάκτηση από wikipedia: https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_max_sep_hyperplane_with_margin.png
- Wikipedia, c. (2017). *Scale-invariant feature transform*. Ανάκτηση από https://en.wikipedia.org/wiki/Scale-invariant_feature_transform
- World Health Organization. (2013). *Global Health Observatory data repository*. Ανάκτηση από <http://apps.who.int/gho/data/node.home>
- Xu, C., He, Y., Khanna, N., Boushey, C. J., & Delp, E. J. (2013). Model-based food volume estimation using 3D pose. In *20th IEEE International Conference on Image Processing (ICIP)*, (σσ. 2534–2538).
- Yanai, K., & Kawano, Y. (2014). Twitter Food Photo Mining and Analysis for One Hundred Kinds of Foods. *Advances in Multimedia Information Processing – PCM 2014* (σσ. 22-32). Springer, Cham.
- Yang, S., Chen, M., Pomerleau, D., & Sukthankar, R. (2010). Food recognition using statistics of pairwise local features. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference*.
- YPSOMED. (2017). *Type 1 vs. Type 2*. Ανάκτηση από Diabetescare: <http://www.mylife-diabetescare.com/mylife-diabetes-knowledge-type1-type2.html>
- Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., & Sawhney, H. (2015). "Snap-n-Eat": Food Recognition and Nutrition Estimation on a Smartphone. *Diabetes Science and Tecnology*, 9(3), 525-533.

- Zhu, F., Bosch, M., Khanna, N., Boushey, C. J., & Delp, E. J. (2015). Multiple Hypotheses Image Segmentation and Classification With Application to Dietary Assessment. *IEEE Journal of Biomedical and Health Informatics*, 19(1), 377 - 388.
- Zhu, F., Mariappan, A., Boushey, C., Kerr, D., Lutes, K., Ebert, D., & Delp, E. (2008, March). Technology-Assisted Dietary Assessment. *Proc SPIE Int Soc Opt Eng*, 6814, 681411.
- Βρότσος, Ι., & Καρούσης, Ι. (2016). *Περιοδοντολογία Εμφυτευματολογία*. Αθήνα: ΒΗΤΑ Ιατρικές Εκδόσεις ΜΕΠΕ.
- ΓΝΑ Ελληνικός Ερυθρός Σταυρός. (2005, Ιούνιος 28). *Επιπλοκές του σακχαρώδη διαβήτη*. Ανάκτηση από iatronet: <http://www.iatronet.gr/ygeia/endokrinologia/article/1017/epiplokes-toy-sakxarwdi-diaviti.html>
- Ελληνική Διαβητολογική Εταιρεία. (2013). *Οδηγός Διατροφής για τη Ρύθμιση του Διαβήτη*. Ιατρικές Εκδόσεις "ΖΗΤΑ".
- ΕΛΛΗΝΙΚΗ ΕΝΔΟΚΡΙΝΟΛΟΓΙΚΗ ΕΤΑΙΡΕΙΑ. (2017). *Σακχαρώδης Διαβήτης*. Ανάκτηση από endo: <http://www.endo.gr/%CF%83%CE%B1%CE%BA%CF%87%CE%B1%CF%81%CF%8E%CE%B4%CE%B7%CF%82-%CE%B4%CE%B9%CE%B1%CE%B2%CE%AE%CF%84%CE%B7%CF%82/>
- Σαραντέα-Μαγγαλούση, Μ. (2016). *Ανάπτυξη Μοντέλου Εκτίμησης της Περιεχόμενης Ποσότητας Υδατανθράκων στα Λαμβανόμενα Γεύματα από Φωτογραφικά Στιγμιότυπα*. Αθήνα.