



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Μέθοδοι Υπολογιστικής Νοημοσύνης για
Εφαρμογές σε Περιβάλλοντα Διάχυτης
Νοημοσύνης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΣΠΥΡΙΔΩΝΟΣ Β. ΓΕΩΡΓΑΚΟΠΟΥΛΟΥ

Επιβλέποντες: Βασίλειος Π. Πλαγιανάκος
Ηλίας Μαγκλογιάννης
Ιωάννης Αναγνωστόπουλος

Λαμία, Ιανουάριος 2019



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Μέθοδοι Υπολογιστικής Νοημοσύνης για
Εφαρμογές σε Περιβάλλοντα Διάχυτης
Νοημοσύνης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΣΠΥΡΙΔΩΝΟΣ Β. ΓΕΩΡΓΑΚΟΠΟΥΛΟΥ

Επιβλέποντες: Βασίλειος Π. Πλαγιανάκος
Ηλίας Μαγκλογιάννης
Ιωάννης Αναγνωστόπουλος

Λαμία, Ιανουάριος 2019



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Copyright ©–All rights reserved Σπυρίδων Β. Γεωργακόπουλος, 2019.

Με την επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό 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.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Υπεύθυνη Δήλωση

Βεβαιώνω ότι είμαι συγγραφέας αυτής της διδακτορικής διατριβής, και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην διατριβή. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης, βεβαιώνω ότι αυτή η διδακτορική διατριβή προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας.

(Υπογραφή)

.....

Σπυρίδων Β. Γεωργακόπουλος

Περίληψη

Η παρούσα διατριβή ασχολείται με προβλήματα του τομέα της Περιβάλλουσας Νοημοσύνης θεμελιώνοντας λύσεις, μελετώντας και αναπτύσσοντας καινοτόμες μεθοδολογίες του τομέα της Υπολογιστικής Νοημοσύνης, για την ενσωμάτωση “ευφύιας” και “αντίληψης” στις συσκευές σε σχέση με τον περιβάλλοντα χώρο τους. Η παρούσα διατριβή οργανώνεται ως εξής:

Στο Κεφάλαιο 1 παρέχουμε τους βασικούς άξονες της Περιβάλλουσας Νοημοσύνης. Παρουσιάζουμε ανοιχτά προβλήματα με τα οποία ασχολείται η Περιβάλλουσα Νοημοσύνη συνοψίζοντας, τις κατηγορίες προβλημάτων που θα ασχοληθούμε και εμείς στην παρούσα διδακτορική διατριβή.

Στο Κεφάλαιο 2 περιγράφουμε βασικές μεθοδολογίες που θα χρησιμοποιηθούν στη συνέχεια των κεφαλαίων. Αφιερώνεται κυρίως στη δομή και λειτουργία των Νευρωνικών Δικτύων Συνελίξεων, στα προβλήματα στα οποία εφαρμόζονται ενώ γίνεται και μια ιστορική αναδρομή στο πρόγονό τους, τα Τεχνητά Νευρωνικά Δίκτυα.

Στα επόμενα δύο κεφάλαια μελετάται το πρόβλημα της αυτόματης αναγνώρισης της στάσης του ανθρώπου, μέσω ευρυγώνιας κάμερας τοποθετημένης στην οροφή ενός δωματίου. Πιο συγκεκριμένα στο Κεφάλαιο 3 αναπτύσσεται ένας νέος ολικός περιγραφέας εξαγωγής χαρακτηριστικών σε εικόνες ευρυγώνιας κάμερας. Ο περιγραφέας αυτός εμφανίζει ανοχή στις αλλοιώσεις του αντικειμένου στο συγκεκριμένο τύπο κάμερας.

Στο Κεφάλαιο 4 συνεχίζουμε την μελέτη του προβλήματος αναγνώρισης της στάσης του ανθρώπου, προτείνοντας μια υβριδική αρχιτεκτονική Νευρωνικών Δικτύων Συνελίξεων με επαυξημένη πληροφορία. Το δίκτυο ενισχύθηκε με την επιπλέον πληροφορία των ολικών περιγραφέων που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, ενώ εφαρμόστηκε η τεχνική μεταφοράς γνώσης για βελτίωση των τελικών αποτελεσμάτων.

Στο Κεφάλαιο 5 μελετάται ένα πρόβλημα της Περιβάλλουσας Νοημοσύνης που σχετίζεται με τα Βοηθητικά Περιβάλλοντα και συγκεκριμένα με τον έγκαιρο και έγκυρο εντοπισμό της πτώσης ενός ατόμου. Για το πρόβλημα αυτό προτείνεται μια μεθοδολογία χαμηλής υπολογιστικής ισχύος, κάνοντας ανάλυση πολυδιάστατων σημάτων φορητών ή φορετών συσκευών.

Το Κεφάλαιο 6 πραγματεύεται το πρόβλημα αυτόματου εντοπισμού παθογενών καταστάσεων του γαστρεντερικού σωλήνα, κάνοντας χρήση ασύρματης ενδοσκοπικής κάψουλας. Προτείνεται μια ολιστική μεθοδολογία ασθενούς εκπαίδευσης βασισμένη στα Νευρωνικά Δίκτυα Συνελίξεων για τον χαρακτηρισμό και εντοπισμό παθολογικών καταστάσεων πάνω σε εικόνες.

Η παρουσίαση του ερευνητικού έργου αυτής της διατριβής ολοκληρώνεται με το Κεφάλαιο 7, όπου προτάθηκε και μελετήθηκε εκτενώς μια μέθοδος εκπαίδευσης Νευρωνικών

Δικτύων Συνελίξεων με αυτόματη προσαρμογή του ρυθμού εκμάθησης. Η μέθοδος που προτείνουμε μπορεί και προσαρμόζει τον αρχικό ρυθμό εκμάθησης σε ένα σταθερό μικρό εύρος τιμών ανεξάρτητα της αρχικής τιμής του.

Η παρουσίαση της διατριβής ολοκληρώνεται με το Κεφάλαιο 8 όπου εμπεριέχονται τα γενικά συμπεράσματα της διατριβής.

Synopsis

This thesis investigates Ambient Intelligence problems establishing novel solutions and innovative Computational Intelligence algorithms, where the “intelligent” and the “perception” are embraced and integrated to smart devices in order to understand the real world environment.

In the first two Chapters we provide the basic Ambient Intelligence definitions and open problems of the field, in which we will deal with in this thesis. Additionally, we discuss the evolution of the classic Artificial Neural Networks (ANN), the Convolutional Neural Networks (CNN), their theoretical background and well known network architectures.

In Chapters 3 and 4, the problem of automatic human pose recognition via omnidirectional fisheye camera is investigated. More specifically, in Chapter 3 a novel feature descriptor is proposed, based on Zernike moments invariant descriptor, adapted on the fisheye camera’s lens by using the geodesic distance. The proposed descriptor is more robust than the classic Zernike on images where have been acquired by the imperfect segmentation of real videos. The proposed Geodesically-Corrected Zernike Moments (GCZM) are used on synthetic as well as real human silhouettes to recognize their poses.

In Chapter 4 the investigation of the aforementioned problem is continued using a GCZM enhanced CNN architecture. The model is trained with synthetic human silhouettes poses enhanced with image features based on Zernike moments, which are modified for fisheye images. Then, using the Transfer Learning technique the model is fine-tuned with a sample of real human silhouettes. Results show that the proposed enhance CNN architecture, combined with Transfer Learning, improves pose classification accuracy for both the synthetic and the real silhouette images.

In Chapter 5, a framework that detects changes on human activity is proposed, designed for devices with low computational power. The proposed methodology is better fitted for high dimensional data streams of multiple device sensors, applying a low computational Incremental Principal Components Analysis method in order to convert the stream to one-dimension. Subsequently, the Cumulative Sum algorithms are detecting changes on them.

In Chapter 6, a methodology for automatic detection and localization of gastrointestinal (GI) anomalies in endoscopic video frame sequences using wireless endoscopy capsule is presented. A generic Neural Network architecture was trained solely with semantically annotated images, indicating whether they contain anomalies or not. While, using the

information that after the train process captured inside their structure, the methodology is capable to localize the abnormalities on video frames.

In Chapter 7, a method for adapting the learning rate of all the CNN weights is proposed. The method uses the first-order information of the gradient vectors of three consecutive iterations during the training phase to adapt the learning rate making the algorithm to converge regardless the initial selectiob of the learning rate value.

στους Δασκάλους μου και στους γονείς μου

Ευχαριστίες

Το τέλος αυτής της διδακτορικής διατριβής σηματοδοτεί την ολοκλήρωση του διδακτορικού κύκλου σπουδών μου και αποτελεί το έναυσμα ενός καινούριου ερευνητικού “ταξιδιού”. Μέσα στο διάστημα αυτό είχα την τύχη να γνωρίσω σπουδαίους και αξιόλογους, όχι μόνο επιστήμονες αλλά κυρίως ανθρώπους.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, καθηγητή κ. Βασίλειο Π. Πλαγιανάκο για την αμέριστη βόηθειά του στην ολοκλήρωση της παρούσας διδακτορικής διατριβής. Η συμβολή του ήταν καθοριστική για την επίτευξη της διατριβής, παρέχοντας ουσιαστική καθοδήγηση και ηθική στήριξη για την αντιμετώπιση των δυσκολιών που προέκυπταν στον “δρόμο”, που είχαμε μαζί χαράξει. Αποτελεί ένα λαμπρό παράδειγμα επιστήμονα και ανθρώπου και τον ευχαριστώ πραγματικά για όλα.

Στο ταξίδι αυτό υπάρχουν και άλλοι σπουδαίοι καθηγητές και επιστήμονες που με βοήθησαν καταλυτικά έτσι ώστε να φτάσω στο τέλος του. Ο αναπληρωτής καθηγητής κ. Ηλίας Μαγκλογιάννης του Πανεπιστημίου Πειραιά ο οποίος ήταν και είναι πάντα δίπλα μου, από την αρχή του ταξιδιού να με παρακινεί και να με συμβουλεύει καθοριστικά σε όλα τα βήματα που έκανα. Ο επίκουρος καθηγητής κ. Κωνσταντίνος Δελήμπασης και ο καθηγητής κ. Δημήτριος Ιακωβίδης του Πανεπιστημίου Θεσσαλίας με τις εμπνευσμένες ιδέες τους μου έδειξαν ερευνητικά μονοπάτια, καθώς και το τρόπο ώστε να τα ακολουθήσω. Ο εξαίρετος νέος επιστήμονας και φίλος δρ. Σωτήριος Τασουλής που με βοήθησε σε όλο αυτό το ταξίδι, αντιμετωπίζοντας μαζί τις δυσκολίες που προέκυπταν. Φυσικά θα ήθελα να ευχαριστήσω όλα τα μέλη της επταμελούς επιτροπής, για τον χρόνο που αφιέρωσαν να μελετήσουν ενδελεχώς το κείμενο της παρούσας διατριβής ώστε με τις καίριες παρατηρήσεις τους να βελτιωθεί συνολικά.

Τέλος, δεν θα μπορούσα να ξεχάσω έναν άνθρωπο του οποίου οι συμβουλές τοποθετούνται χρονικά από το δεύτερο έτος των προπτυχιακών σπουδών μου και φτάνουν μέχρι σήμερα. Η βοήθεια του είναι ανεκτίμητη τόσο στον επιστημονικό τομέα όσο και στον προσωπικό. Χωρίς την παρουσία του αυτή, η διατριβή αυτή δεν θα είχε ξεκινήσει ποτέ ενώ δεν θα υπήρχε ούτε σαν σχέση. Έτσι λοιπόν ευχαριστώ αυτόν τον λαμπρό επιστήμονα δρ. Μιχαήλ Γ. Επιτροπάκη. Κυρίως όμως τον ευχαριστώ για τη φιλία του, γιατί αυτό είναι που αξίζει περισσότερο.

Περιεχόμενα

Περιεχόμενα	xi
Κατάλογος Σχημάτων	xvi
Κατάλογος Πινάκων	xix
I Εισαγωγή και Βασικές Έννοιες	1
1 Εισαγωγή	3
1.1 Περιβάλλουσα Νοημοσύνη	3
1.1.1 Παράγοντες της Περιβάλλουσας Νοημοσύνης	5
1.2 Εικονικά Σενάρια του Χώρου της ΠΝ	6
1.3 Προβλήματα του Πραγματικού Κόσμου στο Χώρο της ΠΝ	7
1.3.1 Έξυπνα Σπίτια	8
1.3.2 Οικογενειακή Φροντίδα Ηλικιωμένων και Βοηθητικά Περιβάλλοντα	9
1.3.3 Ηλεκτρονική Ιατρική Φροντίδα	10
2 Νευρωνικά Δίκτυα Συνελίξεων	11
2.1 Εισαγωγή	11
2.2 Ο Τελεστής Συνέλιξης	13
2.3 Βασικές Αρχές των ΤΝΔΣ	15
II Εντοπισμός της Ανθρώπινης Στάσης μέσω Στατικής Κάμερας και Αισθητήρων Έξυπνων Συσκευών	19
3 Εντοπισμός της Ανθρώπινης Στάσης από Ευρυγώνια Κάμερα	21
3.1 Εισαγωγή	21
3.2 Εξαγωγή Χαρακτηριστικών σε Εικόνες από Ευρυγώνια Κάμερα	22
3.2.1 Ροπές Zernike	24
3.2.2 Βαθμονόμηση Ευρυγώνιας Κάμερας	25
3.2.3 Προσαρμογή των Ροπών Zernike σε Ευρυγώνια Κάμερα	26
3.2.4 Μελέτη της Γεωδαιτικής Διόρθωσης των Ροπών Zernike	29

3.3	Πειραματική Θεμελίωση των Διορθωμένων Ροπών Zernike σε Συνθετικά και Πραγματικά Δεδομένα	31
3.3.1	Συνθετικές Σιλουέτες	31
3.3.2	Πραγματικές Σιλουέτες	34
3.4	Πειραματικά Αποτελέσματα	34
3.4.1	Ταξινόμηση των Πέντε Όρθιων Ανθρώπινων Στάσεων	36
3.4.2	Ταξινόμηση Δύο Γενικευμένων Ανθρώπινων Στάσεων	38
3.5	Συμπεράσματα	40
4	Επαυξημένα Νευρωνικά Δίκτυα Συνελίξεων για τον Εντοπισμό της Ανθρώπινης Στάσης	41
4.1	Εισαγωγή	42
4.2	Επαυξημένα Νευρωνικά Δίκτυα Συνελίξεων	42
4.2.1	Αρχιτεκτονική του Επαυξημένου Νευρωνικού Δικτύου Συνελίξεων	43
4.2.2	Περιγραφή Καθημερινών Ανθρώπινων Στάσεων	44
4.3	Πειραματική Επιβεβαίωση των ΕΝΔΣ	45
4.3.1	Ταξινόμηση των Τριών Ανθρώπινων Στάσεων σε Συνθετικά Δεδομένα	46
4.3.2	Εντοπισμός Ανθρώπινης Στάσης σε Συνεχές Βίντεο	47
4.4	Συμπεράσματα	50
5	Μέθοδος Χαμηλής Υπολογιστικής Ισχύος για τον Εντοπισμό της Ανθρώπινης Πτώσης	51
5.1	Εισαγωγή	52
5.2	Μέθοδος Εντοπισμού Αλλαγών σε Σήματα Μεγάλης Διάστασης	53
5.2.1	Διαδοχική Ανάλυση Κυρίων Συνιστωσών	54
5.2.2	Αλγόριθμος Συσσωρευτικού Αθροίσματος	55
5.3	Πειραματικά Αποτελέσματα	57
5.3.1	Πρόβλημα Εντοπισμού Ανθρώπινης Πτώσης σε Τεχνητά Υψηλής Διάστασης Δεδομένα	57
5.3.2	Εφαρμογή σε Δεδομένα τριών Διαστάσεων	59
5.4	Συμπεράσματα	64
III	Εντοπισμός Παθολογικών Καταστάσεων σε Ιατρικές Εικόνες	67
6	Αναγνώριση και Εντοπισμός Παθολογικών Καταστάσεων στο Γαστρεντερικό Σωλήνα	69
6.1	Εισαγωγή	69
6.1.1	Βιβλιογραφική αναφορά Συστημάτων Ιατρικών Αποφάσεων του Γαστρεντερικού Σωλήνα	71

6.2	Μέθοδος Αναγνώρισης και Εντοπισμού Παθολογικών Καταστάσεων στο Γαστρεντερικό Σωλήνα	72
6.2.1	Ασθενές Νευρωνικό Δίκτυο Συνελίξεων για Εικόνες Γαστρεντερικού Σωλήνα	73
6.2.2	Εντοπισμός Σημαντικών Σημείων στις Εικόνες Γαστρεντερικού Σωλήνα	74
6.2.3	Μέθοδος Ενοποίησης Επαναληπτικής Ομαδοποίησης	76
6.3	Προσομοιώσεις και Αποτελέσματα	77
6.3.1	Σύνολα Δεδομένων Εικόνων του Γαστρεντερικού Σωλήνα	77
6.3.2	Ανάλυση Παραμέτρων Εκπαίδευσης του ΑΝΔΣ	78
6.3.3	Σύγκριση ΑΝΔΣ με ΤΝΔΣ Δεδομένων Αυστηρού Προσδιορισμού . .	80
6.3.4	Σύγκριση Διαφορετικών μεθοδολογιών ΑΝΔΣ	81
6.3.5	Εύρεση Σημαντικών Σημείων και Αξιολόγηση των Μεθόδων	83
6.3.6	Περαιτέρω Εξέταση την Μεθοδολογίας σε Βίντεο και Εικόνες με Πολύποδες	87
6.4	Συμπεράσματα	90
IV	Μέθοδοι Μη Μονότονοι Εκπαίδευσης ΤΝΔΣ	93
7	Εκπαίδευση ΤΝΔΣ με Μεταβλητό Ρυθμό Εκπαίδευσης	95
7.1	Εισαγωγή	95
7.2	Αλγόριθμοι Εκπαίδευσης Μεταβλητού Ρυθμού Εκμάθησης	97
7.2.1	Αλγόριθμος	98
7.3	Προσομοιώσεις και Αποτελέσματα	99
7.4	Συμπεράσματα	104
V	Συμπεράσματα - Βιβλιογραφία	105
8	Συμπεράσματα Διατριβής	107

Κατάλογος Σχημάτων

1.1	Γραφική απεικόνιση των όρων που συντελούν την έννοια Περιβάλλουσα Νοημοσύνη.	4
2.1	Σχηματική αναπαράσταση συνέλιξης σε διδιάστατα δεδομένα.	14
2.2	Αναπαράσταση της αλληλεπίδρασης ενός κρυφού νευρώνα (F_3) με τα δεδομένα εισόδου (X). Παρατηρούμε ότι στην περίπτωση των ΤΝΔ (πάνω) ο νευρώνας F_3 αλληλεπιδρά με όλα τα στοιχεία της εισόδου, οπότε απαιτούνται πέντε βάρη να εκπαιδευτούν. Στην περίπτωση των ΤΝΔΣ (κάτω) με συνάρτηση πυρήνα διάστασης 3 το στοιχείο του μητρώο χαρακτηριστικών F_3 αλληλεπιδρά μόνο με τρία στοιχεία του συνόλου δεδομένων X	16
3.1	Λήψη όρθιας στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (α). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας, στην οποία εφαρμόζεται η μέθοδος εξαγωγής χαρακτηριστικών (β).	23
3.2	Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας.	24
3.3	Σχηματική αναπαράσταση της απεικόνισης μιας εικόνας σε σφαιρική ευρυγώνια κάμερα.	26
3.4	Απεικόνιση των τιμών του αξιμούθιου (κέντρο) και της ανύψωσης (δεξιά) μετά από την ευθυγράμμιση της ευρυγώνιας κάμερας που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά).	26
3.5	Αναπαράσταση των σημείων P_0 και P_1 στην μοναδιαία σφαίρα και ο υπολογισμός της γεωδαιτικής απόστασης και γωνίας τους.	27
3.6	Στιγμιότυπα της ευρυγώνιας κάμερας που απεικονίζουν ένα αντικείμενο σε διαφορετικές θέσεις, στροφές και απόσταση από την κάμερα (α). Τμηματοποίηση των αντίστοιχων στιγμιότυπων, αφαιρώντας το παρασκήνιο διατηρώντας το σχήμα του αντικειμένου (β).	29
3.7	Η τιμή της πρώτης κεντρικής διαφοράς στα 450 στιγμιότυπα του σχήματος του αντικειμένου των ΓΔΡΖ έναντι των κλασικών ροπών Zernike (α). Ο λόγος των διαφορών: των τρίτων προς τις δεύτερες (β) και των δεύτερων προς τις πρώτες (γ), των ΓΔΡΖ έναντι των κλασικών ροπών Zernike.	30
3.8	Τα συνθετικά μοντέλα των πέντε όρθιων ανθρώπινων στάσεων.	31

3.9	Γραφική απεικόνιση των συνθετικών μοντέλων στο πραγματικό δωμάτιο της ευρυγώνιας κάμερας (α) και (β). Τοποθέτηση των συνθετικών μοντέλων στον πραγματικό χώρο με απόσταση 0,5 μέτρα και σταθερό προσανατολισμό για την δημιουργία του συνόλου εκπαίδευσης (γ). Υπο-εικόνα μιας σιλουέτας των συνθετικής μοντέλων για την εφαρμογή των μεθόδων εξαγωγής ολικών χαρακτηριστικών.	32
3.10	Λήψη της στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας (δεξιά).	34
3.11	Τα ποσοστά επιτυχίας των ΓΔΡΖ και ΚΡΖ στις συνθετικές σιλουέτες, καθώς απομακρύνονται από το κέντρο της ευρυγώνιας κάμερας.	38
4.1	Λήψη όρθιας στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας (κέντρο). Υπο-εικόνα της ανθρώπινης σιλουέτας που δέχεται ως είσοδο το ΤΝΔΣ (δεξιά).	43
4.2	Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας.	43
4.3	Η βασική αρχιτεκτονική ΤΝΔΣ για τον αναγνώριση της ανθρώπινης στάσης μέσω της εικόνας σιλουέτας του ανθρώπου.	44
4.4	Η προτεινόμενη αρχιτεκτονική ΕΝΔΣ. Τα ΓΔΡΖ ολικά χαρακτηριστικά από μια εικόνα ανθρώπινης σιλουέτας εισέρχονται στο πλήρως διασυνδεδεμένο επίπεδο νευρώνων, σε συνδυασμό με τα τοπικά χαρακτηριστικά των αλληπάλληλων συνελίξεων.	45
4.5	Το δωμάτιο στο οποίο είναι τοποθετημένη και έχει ευθυγραμμιστεί η ευρυγώνια κάμερα. Οι παράμετροι από την ευθυγράμμιση της χρησιμοποιήθηκαν για το δωμάτιο προσομοίωσης των συνθετικών μοντέλων ανθρώπινης στάσης.	46
4.6	Τα μοντέλα των τριών ανθρώπινων στάσεων. Στις εικόνες α_1 , α_2 παρουσιάζεται η όρθια στάση, στις εικόνες β_1 , β_2 παρουσιάζεται η καθιστή στάση και στις εικόνες γ_1 , γ_2 παρουσιάζεται η πτώση στο πάτωμα.	47
5.1	Τα δεδομένα εκπαίδευσης μετασχηματισμένα στο 2-διάστατο χώρο μέσω της μεθόδου ΑΚΣ (αριστερά). Τα δεδομένα εκπαίδευσης στο 1-διάστατο χώρο μέσω της μεθόδου ΑΚΣ. Τα διαφορετικά χρώματα υποδηλώνουν τις τέσσερις καταστάσεις του συνόλου δεδομένων.	58
5.2	Τα δεδομένα ελέγχου μετασχηματισμένα στον 1-διάστατο χώρο του ΔΑΚΣΣ. Τα κόκκινα σημεία του γραφήματος υποδηλώνουν την στιγμή ενεργοποίησης του ΣΥΣΑΘ.	58
5.3	Παράδειγμα της συνάρτησης απόφασης του ΣΥΣΑΘ στο σύνολο ελέγχου. Οι πράσινες κάθετες γραμμές υποδηλώνουν την χρονική περίοδο της πραγματικής κατάστασης πτώσης ενώ το κόκκινο σημείο δείχνει την ενεργοποίηση του ΣΥΣΑΘ.	59

5.4	Η χρονική ακολουθία των τριών αξόνων του επιταχυνσιόμετρου των δεδομένα χαμηλής διάστασης.	60
5.5	Χρονοσειρά των μετασχηματισμένων δεδομένων των τριών διαστάσεων επιταχυνσιόμετρου σε 1-διάστατο μέσω του ΔΑΚΣΣ (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.	62
5.6	Χρονοσειρά των μετασχηματισμένων δεδομένων των τριών διαστάσεων επιταχυνσιόμετρου σε 1-διάστατο μέσω του ΑΚΣ (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.	62
5.7	Χρονοσειρά του δεδομένων του άξονα X του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.	63
5.8	Χρονοσειρά του δεδομένων του άξονα Y του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος αποτυγχάνει στον εντοπισμό της ανθρώπινης πτώσης.	63
5.9	Χρονοσειρά του δεδομένων του άξονα Z του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.	64
6.1	Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας χαρακτηρισμού και εντοπισμού παθογενών καταστάσεων σε εικόνες από το γαστρεντερικό σωλήνα.	73
6.2	Εισαγωγή ολόκληρης της έγχρωμης εικόνας στο επίπεδο των συνελιξίων ενός ΝΔΣ (αριστερά). Εισαγωγή τμημάτων (υποσύνολα) από την αρχική έγχρωμη εικόνας στο επίπεδο των συνελιξίων ενός ΝΔΣ (δεξιά).	74
6.3	Το ποσοστό επιτυχίας του μοντέλου ΑΝΔΣ με τον αλγόριθμο εκπαίδευσης ΣΑΚ σε διαφορετικού πλήθους σύνολα εκπαίδευσης.	80
6.4	Τα σημαντικά σημεία που παράγονται από τον αλγόριθμο ΕΣΣΒΕ (αριστερή στήλη) και τον CSD (δεξιά στήλη). Εικόνες από το Σ1 (α - β). Εικόνες από το Σ2 (γ - δ).	85

- 6.5 Τέσσερις εικόνες που εμφανίζουν παθογένια και η περιοχή υποδεικνύεται από τους ειδικούς ιατρούς (αριστερή στήλη). Οι ΕΜ των αντίστοιχων εικόνων και με τα σημεία του αλγορίθμου ΕΣΣΒΕ (κεντρική στήλη). Φιλτράρισμα των σημαντικών σημείων με την χρήση του ΜΕΕΟ (δεξιά στήλη). Εικόνες από το Σ1 (α - β). Εικόνες από το Σ2 (γ - δ). 86
- 6.6 Στιγμιότυπα εικόνων με παθογενείς περιοχές, όπως υποδεικνύονται από τους ειδικούς ιατρούς με το μαύρο βέλος, από το βίντεο KID: αγγειεκτασία (α), λεμφαγγειεκτασία (β), οζώδεις λεμφοειδείς υπερπλασία (γ). Στις εικόνες της δεύτερης γραμμής (δ) εμφανίζονται τρία διαδοχικά στιγμιότυπα με αγγειεκτασία όπου η προτεινόμενη μεθοδολογία έχει εντοπίσει και υποδεικνύει την περιοχή παθογένειας. 87
- 6.7 Εικόνες που απεικονίζουν παθογενείς περιοχές από τα τέσσερα διαφορετικά είδη συνόλων δεδομένων που χρησιμοποιήθηκαν στο κεφάλαιο αυτό. CVC-CLININC (α), ETIS-LARIB (β), MICCAI (γ) και KID (δ). 89
- 7.1 Σχηματική απεικόνιση της μεθόδου εύρεσης τοπικού ελαχίστου συνάρτησης στις δύο διαστάσεις. Στη περίπτωση που έχουμε μικρή τιμή ρυθμού εκμάθησης, το συνημίτονο της γωνίας των διανυσμάτων των βαρών των δύο τελευταίων επαναλήψεων είναι θετικό (αριστερά). Όταν η τιμή του ρυθμού εκμάθησης είναι μεγάλη και αποκλίνει από το τοπικό ελάχιστο, το συνημίτονο της γωνίας των διανυσμάτων των βαρών των δύο τελευταίων επαναλήψεων είναι αρνητικό (δεξιά). 99
- 7.2 Το ποσοστό επιτυχίας της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης. 101
- 7.3 Η προσαρμογή του ρυθμού εκμάθησης μέσω του προτεινόμενου αλγορίθμου χρησιμοποιώντας διαφορετικές αρχικές τιμές. 102
- 7.4 Το σφάλμα εκπαίδευσης της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης. 103

Κατάλογος Πινάκων

3.1	Το σύνολο δεδομένων για τις πέντε συνθετικές σιλουέτες όρθιας ανθρώπινης στάσης.	33
3.2	Το σύνολο δεδομένων συνθετικών σιλουετών των δύο γενικευμένων ανθρώπινων στάσεων.	33
3.3	Οι τιμές των παραμέτρων n και m της εξίσωσης 3.2 με το αντίστοιχο πλήθος χαρακτηριστικών. Το αθροιστικό πλήθος χαρακτηριστικών είναι το συνολικό πλήθος χαρακτηριστικών (πραγματικών και μιγαδικών) για όλους τους έγκυρους συνδυασμούς των παραμέτρων (n, m) για το τρέχον n κάθε φορά εξαιρώντας τα μιγαδικά συζηγή.	35
3.4	Το ποσοστό επιτυχίας των ΚΡΖ και ΓΔΡΖ στις συνθετικές σιλουέτες χρησιμοποιώντας διαφορετικά μέτρα των παραγόμενων χαρακτηριστικών τους. . . .	36
3.5	Το μητρώο σύγκρισης του ΚΡΖ στις συνθετικές σιλουέτες όρθιας στάσης. . .	36
3.6	Το μητρώο σύγκρισης του ΓΔΡΖ στις συνθετικές σιλουέτες όρθιας στάσης. .	36
3.7	Το μητρώο σύγκρισης του ΚΡΖ στις πραγματικές ανθρώπινες σιλουέτες όρθιας στάσης.	37
3.8	Το μητρώο σύγκρισης του ΓΔΡΖ στις πραγματικές ανθρώπινες σιλουέτες όρθιας στάσης.	37
3.9	Τα ποσοστά επιτυχίας των τριών μεθόδων στις συνθετικές και πραγματικές σιλουέτες ανθρώπινων όρθιων στάσεων.	37
3.10	Το μητρώο σύγκρισης του ΚΡΖ στις συνθετικές σιλουέτες των δύο γενικευμένων στάσεων.	39
3.11	Το μητρώο σύγκρισης του ΓΔΡΖ στις συνθετικές σιλουέτες των δύο γενικευμένων στάσεων.	39
3.12	Το μητρώο σύγκρισης του ΚΡΖ στις πραγματικές ανθρώπινες σιλουέτες γενικευμένων στάσεων.	39
3.13	Το μητρώο σύγκρισης του ΚΡΖ στις πραγματικές ανθρώπινες σιλουέτες γενικευμένων στάσεων.	39
3.14	Τα ποσοστά επιτυχίας των τριών μεθόδων στις συνθετικές και πραγματικές σιλουέτες ανθρώπινων γενικευμένων στάσεων.	39
4.1	Το πλήθος των συνθετικών σιλουετών για τις τρεις ανθρώπινες στάσεις. . . .	45

4.2	Τα ποσοστά επιτυχίας των τριών μεθόδων στα δεδομένα ελέγχου συνθετικών σιλουετών.	47
4.3	Το μητρώο σύγχυσης του ΕΝΔΣ στις συνθετικές σιλουέτες.	48
4.4	Το μητρώο σύγχυσης του ΤΝΔΣ στις συνθετικές σιλουέτες.	48
4.5	Το μητρώο σύγχυσης του ΓΔΡΖ στις συνθετικές σιλουέτες.	49
4.6	Τα ποσοστά επιτυχίας των εκπαιδευμένων μοντέλων με συνθετικές σιλουέτες, για την αναγνώριση των πραγματικών ανθρώπινων σιλουετών.	49
4.7	Τα ποσοστά επιτυχίας των εκπαιδευμένων μοντέλων με πραγματικές ανθρώπινες σιλουέτες του βίντεο 1, για την αναγνώριση των πραγματικών ανθρώπινων σιλουετών. Στα ΕΝΔΣ και ΤΝΔΣ έχει χρησιμοποιηθεί η μέθοδος μεταφοράς γνώσης από τα αρχικά εκπαιδευμένα μοντέλα με συνθετικές σιλουέτες, ενώ η μέθοδος του Poppe έχει εκπαιδευτεί εξολοκλήρου με τις πραγματικές σιλουέτες του βίντεο 1.	49
4.8	Το μητρώο σύγχυσης του ΤΝΔΣ στις πραγματικές σιλουέτες ανθρώπων από τα Βίντεο 2 και 3.	50
4.9	Το μητρώο σύγχυσης του ΕΝΔΣ στις πραγματικές σιλουέτες ανθρώπων από τα Βίντεο 2 και 3.	50
5.1	Χρήση της προτεινόμενης μεθόδου για τον εντοπισμό των καταστάσεων πτώσης για διαφορετικές τιμές της παραμέτρου h	59
5.2	Πειραματικά αποτελέσματα της προτεινόμενης μεθόδου στο πρόβλημα αναγνώρισης πτώσης του ανθρώπου και σύγκριση με τις μεθόδους [125], για διαφορετικές τιμές της παραμέτρου h	61
6.1	Απόδοση του ΑΝΔΣ κάνοντας χρήση διαφορετικών αλγορίθμων εκπαίδευσης στα Σύνολα 1 και 2B (Σ1 - Σ2B)	79
6.2	Οι p -τιμές του ελέγχου σημαντικότητας Wilcoxon στα αποτελέσματα των διαφορετικών αλγορίθμων εκπαίδευσης του ΑΝΔΣ στα Σύνολα 1 και 2B (Σ1 - Σ2B)	79
6.3	Επιδόσεις ΤΝΔΣ αυστηρής επισήμανσης στα Σ1 και Σ2B.	81
6.4	Επιδόσεις προτεινόμενου μοντέλου ΑΝΔΣ και ΤΝΔΣ αυστηρής επισήμανσης στα Σ1 και Σ2B σε ολόκληρη την εξεταζόμενη εικόνα.	82
6.5	Επιδόσεις μοντέλων ΑΝΔΣ στα Σ1 και Σ2.	82
6.6	Τα αποτελέσματα των αλγορίθμων εύρεσης σημαντικών σημείων στα Σ1 και Σ2. Για το κάθε σύνολο δεδομένων παρουσιάζεται το ελάχιστο και μέγιστο πλήθος σημείων που παράγει ο αλγόριθμος σε μια εικόνα του συνόλου.	84
6.7	Απόδοση της προτεινόμενης μεθοδολογίας έναντι των μεθοδολογιών που προτάθηκαν για τον διαγωνισμό ‘Automatic Polyp Detection in Colonoscopy Videos’ στα σύνολα δεδομένων ETIS-LARIB και CVC-CLINIC. Καθώς το σύνολο δεδομένων CVC-CLINIC χρησιμοποιήθηκε από τις άλλες μεθοδολογίες ως σύνολο εκπαίδευσης δεν έχουμε δυνατότητα παρουσίασης απόδοσης πέραν την προτεινόμενης μεθόδου η οποία έχει εκπαιδευτεί από το Σ3.	89

- 7.1 Η μέση τιμή επιτυχίας και η διασπορά της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης, για το πρόβλημα CIFAR-10. 102

Μέρος Ι

Εισαγωγή και Βασικές Έννοιες

Κεφάλαιο 1

Εισαγωγή

Η Τεχνητή Νοημοσύνη και η συνεχής εξέλιξή της έχει φέρει επανάσταση στην καθημερινότητα των ανθρώπων. Η εποχή μας διανύει την περίοδο της τέταρτης βιομηχανικής επανάστασης (Industry 4.0), όπου τεχνολογικά επιτεύγματα συνδέονται και ενσωματώνονται στις κοινωνίες, ακόμα και στον ίδιο τον άνθρωπο¹. Η τέταρτη βιομηχανική επανάσταση χαρακτηρίζεται από την ένωση διαφορετικών υπολογιστικών στοιχείων τα οποία μπορούν και ελέγχουν φυσικές οντότητες και αναφέρονται με τον ορισμό Κυβερνο-Φυσικά Συστήματα (ΚφΣ) (Cyber-Physical Systems). Τα ΚφΣ συνδέουν τους ενσωματωμένους υπολογιστές, τα δίκτυα παρακολούθησης και τις φυσικές διαδικασίες με έναν αμφίδρομο τρόπο επικοινωνίας, έτσι ώστε οι φυσικές διαδικασίες να επηρεάζουν τις υπολογιστικές και αντιστρόφως. Βασικό στοιχείο των κυβερνο-φυσικών συστημάτων είναι η αξιοποίηση πρωτοκόλλων επικοινωνίας ταχυτήτων υψηλής απόδοσης (4G) με καινοτόμα τεχνολογικά επιτεύγματα διαφορετικών επιστημονικών κλάδων και κυρίως της ρομποτικής, της Τεχνητής Νοημοσύνης (ΤΝ), της Περιβάλλουσας Νοημοσύνης (ΠΝ), του Διαδικτύου των Αντικειμένων (ΔτΑ) (Internet of Things, IoT), των αυτόνομων μετακινούμενων οχημάτων, των ιατρικών συστημάτων ελέγχου, κ.α. [76]. Έτσι τα ΚφΣ θα μπορούσαν να χαρακτηριστούν ως εξέλιξη της ΠΝ, καθώς μοιράζονται κάποιες κοινές βασικές αρχιτεκτονικές. Όμως, τα ΚφΣ παρουσιάζουν μεγαλύτερη ενσωμάτωση των φυσικών και υπολογιστικών στοιχείων [103].

1.1 Περιβάλλουσα Νοημοσύνη

Στα τέλη της δεκαετίας του 1990, η Συμβουλευτική Ομάδα Κοινωνίας των Πληροφοριών και Τεχνολογίας της Ευρωπαϊκής Επιτροπής (ΣΟΚΗΠΤΕΕ) (European Commission's Information Society and Technology Advisory Group, ISTAG) όρισε ως Περιβάλλουσα Νοημοσύνη (Ambient Intelligence), τα περιβάλλοντα τα οποία ενσωματώνουν αισθητήρες και έξυπνα συστήματα. Τα περιβάλλοντα αυτά έχουν τις ακόλουθες ιδιότητες:

- η ικανότητα αναγνώρισης της παρουσίας ατόμων στο χώρο,
- αναγνώριση της ταυτότητας του ατόμου,

¹<https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/>

- η ικανότητα αναγνώρισης διαφορετικών πλαισίων/εννοιών (π.χ. καιρός, κίνηση, νέα),
- αναγνώριση δραστηριοτήτων,
- η ικανότητα προσαρμογής στις μεταβαλλόμενες ανάγκες των ατόμων.

Σκοπός της ΠΝ είναι να μπορεί να παρέχει εξατομικευμένες υπηρεσίες αυτόματα, σύμφωνα τις ανάγκες των ατόμων που δρουν στον περιβάλλοντα χώρο. Η ΣΟΚΠΤΕΕ δεν καθόρισε αυστηρά τις προδιαγραφές για την ΠΝ, καθώς αναγνώρισε την πολυπλοκότητα και την ταχεία εξέλιξη των τεχνολογιών που εμπεριέχονται σε αυτήν. Όρισε όμως μια ολιστική προσέγγιση, προσδιορίζοντας τις ερευνητικές και τεχνολογικές περιοχές που πρέπει να συνδυαστούν για την υλοποίηση της ΠΝ, τόσο στο επίπεδο της τεχνολογία, όσο και στο επίπεδο της κοινωνίας και των επιχειρήσεων.



Σχήμα 1.1: Γραφική απεικόνιση των όρων που συντελούν την έννοια Περιβάλλουσα Νοημοσύνη.

Αν προσπαθήσουμε να χωρίσουμε την έννοια ΠΝ στους όρους που τη συνθέτουν, έχουμε σαν αποτέλεσμα τις έννοιες “Περιβάλλουσα” και “Νοημοσύνη”. Προχωρώντας σε μια ανάλυση των όρων αυτών θα αποκτήσουμε μια καλύτερη αντίληψη της ΠΝ. Στην Εικόνα 1.1 βλέπουμε τα συστατικά που περιλαμβάνουν οι παραπάνω έννοιες.

Η έννοια “Περιβάλλουσα” υποδηλώνει το μέρος του υλικού και του συστήματος γενικότερα περιλαμβάνοντας τους αισθητήρες, επεξεργαστές, επικοινωνίες και το διαχειριστικό λογισμικό.

Αξίζει να σημειωθεί ότι τα τελευταία χρόνια υπάρχει τεράστια πρόοδος στους τομείς των έξυπνων τηλεφώνων (smartphones), φορετών έξυπνων συσκευών (wearables), καθώς και στο ΔτΑ. Ενώ στο άμεσο μέλλον, υπάρχει η πεποίθηση ότι οι τεχνολογίες αυτές μπορούν ουσιαστικά να ενσωματωθούν μεταξύ τους.

Η έννοια “Νοημοσύνη” αναφέρεται στο λογισμικό και καθορίζει την επιτυχία ή την αποτυχία του τομέα της ΠΝ. Η έννοια της “Νοημοσύνης” σχετίζεται με την Τεχνητή Νοημοσύνη (ΤΝ) και αποτελεί έναν σημαντικό τομέα της επιστήμης των υπολογιστών. Σκοπός της ΤΝ είναι να δώσει την ικανότητα προσομοίωσης των διαδικασιών ανθρώπινης νοημοσύνης στις μηχανές και τα συστήματα Πληροφορικής. Αυτές οι διαδικασίες περιλαμβάνουν τη μάθηση (απόκτηση πληροφοριών και δημιουργία κανόνων για τη ορθή χρήση των πληροφοριών αυτών), τη συλλογιστική (χρήση κανόνων για την επίτευξη προσεγγιστικών ή απόλυτων συμπερασμάτων) και την αυτοδιόρθωση [111].

Έξυπνα συστήματα με την βοήθεια κανόνων (Rule-based Expert Systems) εμφανίστηκαν στα πρώτα στάδια της ΤΝ, με σκοπό την εισαγωγή “ευφυίας” στις μηχανές [55]. Πάνω σε αυτά τα συστήματα στηρίζονται και οι συσκευές του ΔτΑ, εισάγοντας έτσι τη βασική μορφή νοημοσύνης στις συσκευές αυτές. Ωστόσο, η ΠΝ εστιάζει σε βαθύτερα επίπεδα νοημοσύνης των μηχανές, για τον λόγο αυτό ο τομέας ΔτΑ αποτελεί τμήμα της ΠΝ, ενσωματώνοντας τομείς της ΤΝ που μελετούν τον τρόπο που ο ίδιος ο άνθρωπος τείνει να μαθαίνει και να αποκτά νοητικές ικανότητες.

Η Υπολογιστική Νοημοσύνη αποτελεί έναν τομέα της ΤΝ, που ορίζεται ως το σύνολο των υπολογιστικών διαδικασιών εκείνων, όπου εξομοιώνοντας μηχανισμούς της φύσης έχουν σκοπό τη λύση σύνθετων προβλημάτων του πραγματικού κόσμου [21, 119]. Οι μέθοδοι που χρησιμοποιεί η Υπολογιστική Νοημοσύνη τείνουν να προσομοιώνουν τον τρόπο σκέψης και τις γνωστικές διαδικασίες του ανθρώπου. *Το κομμάτι αυτό θα αποτελέσει το ενδιαφέρον της παρούσας διδακτορικής διατριβής, καθώς θα εξετάσουμε συγκεκριμένα προβλήματα του τομέα της ΠΝ θεμελιώνοντας λύσεις, μελετώντας και αναπτύσσοντας καινοτόμες μεθοδολογίες του τομέα της Υπολογιστικής Νοημοσύνης.*

1.1.1 Παράγοντες της Περιβάλλουσας Νοημοσύνης

Η ΣΟΚΠΤΕΕ πρότεινε μια ολιστική άποψη για την προσέγγιση της ανάπτυξης της ΠΝ με τον άνθρωπο στο κέντρο της: ‘Η ΠΕ πρέπει να καθοδηγείται από τις ανθρωπιστικές ανησυχίες και όχι από τα τεχνολογικά επιτεύγματα, ενώ θα πρέπει να ελέγχεται από τους ανθρώπους’. Βασικοί παράγοντες που επηρεάζουν την υιοθέτηση της ΠΝ είναι οι εξής:

Ευχρηστία: Η αλληλεπίδραση του ατόμου με τη ΠΝ πρέπει να είναι διακριτική. Το άτομο στον περιβάλλοντα χώρο που υπάρχει η ΠΝ θα πρέπει να βιώνει το αποτέλεσμα χωρίς την ανάγκη ειδικών ρυθμίσεων σε αυτό.

Τεχνική σκοπιμότητα: Η ΠΝ θα πρέπει να λειτουργεί αξιόπιστα, μέσα στους περιορισμούς των τεχνολογιών αιχμής. Πρέπει να ληφθούν υπόψη παράγοντες όπως η ακρίβεια, η ικανότητα και τα μέτρα ασφάλειας για όλα τα εξαρτήματα υλικού και λογισμικού.

Ιδιωτικότητα: Η δυνατότητα μοναδικής αναγνώρισης και παρακολούθησης των ανθρώπων, διασφαλίζοντας τις πολιτικές απορρήτου σε δημόσιο και ιδιωτικό χώρο.

Κοινωνικές και οικονομικές επιπτώσεις: Η ΠΝ θα πρέπει να είναι σε θέση να ενισχύει τις κοινωνικές αλληλεπιδράσεις όλων των συμμετεχόντων στο περιβάλλον. Εξίσου σημαντικό είναι ο οικονομικός αντίκτυπος της ΠΝ, καθώς θα αλλάξει τις λειτουργίες εργασίας πολλών ανθρώπων.

Η ΠΝ πρέπει να κερδίσει την εμπιστοσύνη των χρηστών πριν υιοθετηθεί ευρέως. Ο Andrew Ng, πρώην επικεφαλής επιστήμονας της Baidu, δήλωσε ότι η ΤΝ είναι η νέα ηλεκτρική ενέργεια. Για να κερδίσει την εμπιστοσύνη, η ΠΝ πρέπει να τοποθετήσει τη διαχείριση κινδύνων, την ασφάλεια των δεδομένων και την αξιοπιστία του συστήματος ως κορυφαίες προτεραιότητές της.

1.2 Εικονικά Σενάρια του Χώρου της ΠΝ

Όραμα της ΠΝ είναι η ολοκλήρωση της κοινωνίας της πληροφορίας, όπου τα κανονικά περιβάλλοντα εργασίας και διαβίωσης περιβάλλονται από ενσωματωμένες έξυπνες συσκευές που μπορούν να ενοποιηθούν πλήρως και να επικοινωνούν μέσω κατάλληλων διεπαφών. Τέτοιες συσκευές, εξειδικευμένες με μία ή περισσότερες δυνατότητες, προορίζονται να συνεργάζονται μέσα σε μια υποδομή έξυπνων συστημάτων, να παρέχουν ένα πλήθος υπηρεσιών που στοχεύουν στη βελτίωση της ποιότητας ζωής και της ασφάλειας του καθημερινού περιβάλλοντος διαβίωσης, ταξιδιού και εργασίας του ατόμου.

Η Ευρωπαϊκή Επιτροπή πρότεινε σενάρια που σχετίζονται με την ΠΝ [35, 36], με σκοπό να θέσει και με τη χρήση παραδειγμάτων, το ερευνητικό πλαίσιο στο οποίο θα κινηθεί η έρευνα στον τομέα αυτό. Στη συνέχεια παρουσιάζονται δύο από τα προτεινόμενα σενάρια της Ευρωπαϊκής Επιτροπής:

Σενάριο 1: Ο Δημήτρης κατά την διάρκεια του διαλείμματός του από την δουλειά θέλει να μην τον διαταράσσει κανένας. Φοράει στα ρούχα ή το σώμα του μια συσκευή με φωνητική ενεργοποίηση ενός ψηφιακού βοηθού, με το όνομα “Προσωπικός-Βοηθός (Π-Β)”. Ο Π-Β είναι μια συσκευή μάθησης, η οποία μαθαίνει τον Δημήτρη και το περιβάλλον του, ενώ προσφέρει λειτουργίες επικοινωνίας, επεξεργασίας και λήψης αποφάσεων. Κατά τη διάρκεια του διαλείμματος, ο Π-Β απαντά στις εισερχόμενες κλήσεις και τα μηνύματα ηλεκτρονικού ταχυδρομείου του Δημήτρη. Αυτό το επιτυγχάνει χρησιμοποιώντας τη φυσική γλώσσα του Δημήτρη κάνοντας αναπαραγωγή ακόμα και της φωνής και της προφοράς του. Στη συνέχεια, ο Π-Β λαμβάνει μια κλήση από τη σύζυγο του Δημήτρη, αναγνωρίζει ότι είναι επείγουσα και τη μεταδίδει στο Δημήτρη. Ταυτόχρονα, δέχεται ένα μήνυμα από το αντίστοιχο Π-Β ενός ηλικιωμένου, που βρίσκεται κοντινή απόσταση. Αυτό το άτομο έχει φύγει από το σπίτι χωρίς το φάρμακό του και θα ήθελε να μάθει πού μπορεί βρει το συγκεκριμένο φάρμακο. Ο Δημήτρης πάσχει από ένα παρόμοιο πρόβλημα υγείας με τον ηλικιωμένο και χρησιμοποιεί το ίδιο φάρμακο. Ο Π-Β επεξεργάζεται την εισερχόμενη αίτηση και αποφασίζει να μην αποκαλύψει την ταυτότητα του Δημήτρη ούτε να προσφέρει άμεση βοήθεια, αλλά να παράσχει στον Π-Β

του ηλικιωμένου έναν κατάλογο με τα πλησιέστερα φαρμακεία και πιθανή επαφή με μια ομάδα αυτοβοήθειας.

Σενάριο 2: Η Κάρμεν ετοιμάζεται να πάει στη δουλειά της. Ζητάει από την ΠΝ, με φωνητική εντολή, να βρει κάποιον με τον οποίο μπορεί να μοιραστεί την διαδρομή για την δουλειά της, σε μισή ώρα. Στη συνέχεια προγραμματίζει το δείπνο που πρόκειται να κάνει εκείνο το βράδυ. Θέλει να φτιάξει ένα γλυκό και το ηλεκτρονικό της ψυγείο δείχνει στην οθόνη του μια συνταγή επισημαίνοντας τα συστατικά που λείπουν. Η Κάρμεν ολοκληρώνει την λίστα αγορών της μέσω της οθόνης του ψυγείου και ζητά να παραδοθεί στο πλησιέστερο σημείο διανομής της γειτονιάς της. Όλα τα προϊόντα είναι έξυπνα με ετικέτα, έτσι ώστε να μπορεί να ελέγχει την πρόοδο των εικονικών αγορών από οποιαδήποτε συσκευή οπουδήποτε και αν βρίσκεται και να κάνει απαιτούμενες αλλαγές. Η Κάρμεν πάει στην εργασία της, με ένα αυτοκίνητο με δυναμικό σύστημα οδήγησης και ενσωματωμένο σύστημα ελέγχου της κυκλοφορίας, το οποίο ρυθμίζει δυναμικά τα όρια ταχύτητας ανάλογα με τη συμφόρηση και τα επίπεδα ρύπανσης. Όταν επιστρέφει στο σπίτι, η ΠΝ την καλωσορίζει και της προτείνει ότι την επόμενη μέρα θα πρέπει να δουλέψει από το σπίτι, καθώς έχει προγραμματιστεί μια μεγάλη διαδήλωση στο κέντρο της πόλης.

1.3 Προβλήματα του Πραγματικού Κόσμου στο Χώρο της ΠΝ

Εκτός των εικονικών σεναρίων της ΠΝ, που έθεσαν το όραμα στο χώρο αυτό, εμφανίστηκαν πραγματικά, ανοιχτά, προβλήματα με μεγάλο ερευνητικό ενδιαφέρον. Τα προβλήματα αυτά, συνδέονται με τις παρακάτω κύριες κατηγορίες εφαρμογής της ΠΝ:

Έξυπνα Σπίτια Ένα περιβάλλον ενός “Έξυπνου Σπιτιού” είναι ένα σπίτι εξοπλισμένο με αισθητήρες διαφόρων τύπων για την παρακολούθηση δραστηριοτήτων και κίνησης του ατόμου, για την παρακολούθηση καταστάσεων κινδύνου, όπως η πυρκαγιά ή συναγερμούς καπνού.

Οικογενειακή Φροντίδα Ηλικιωμένων και Βοηθητικά Περιβάλλοντα Ένας τομέας της ΠΝ που έχει σκοπό την υποστήριξη της ανεξάρτητης διαβίωσης των ηλικιωμένων και βελτίωση της ποιότητας ζωής τους.

Ηλεκτρονική Ιατρική Φροντίδα Ένα περιβάλλον της ΠΝ όπου η εξέταση και η θεραπεία των ασθενών θα μπορεί να γίνει απομακρυσμένα από ειδικούς ιατρούς, με την βοήθεια αισθητήρων και ρομπότ.

Εμπόριο και Συστήματα Σύστασης Περιβάλλοντα της ΠΝ με συσκευές που ελέγχονται από ‘έξυπνα’ λογισμικά, που αντιλαμβάνονται την παρουσία πελατών σύμφωνα με την ταυτότητα και τα προφίλ τους προτείνοντάς τους τα κατάλληλα προϊόντα.

Στην παρούσα διατριβή θα ασχοληθούμε με προβλήματα των τριών πρώτων κατηγοριών της ΠΝ, εξετάζοντας το τμήμα του προβλήματος που σχετίζεται με την δημιουργία “ευφυίας” και

αντίληψης των συσκευών με τον περιβάλλοντα χώρο τους. Στη συνέχεια, θα αναφερθούμε στις κατηγορίες αυτές κάνοντας μια βιβλιογραφική ανασκόπηση.

1.3.1 Έξυπνα Σπίτια

Ένα όραμα της ΠΝ στοχεύει στο περιβάλλον του σπιτιού, δημιουργώντας το “Έξυπνο Σπίτι”. Πολλά ονόματα έχουν χρησιμοποιηθεί για να περιγράψουν σπίτια εξοπλισμένα με διάχυτη τεχνολογία για την παροχή υπηρεσιών ΠΝ στους κατοίκους. “Έξυπνο Σπίτι” μπορεί να είναι ο πιο δημοφιλής όρος, ενώ άλλοι όροι περιλαμβάνουν γνωστικές κατοικίες, έξυπνες κατοικίες, ολοκληρωμένα περιβάλλοντα, ζωντανά - διαδραστικά σπίτια/περιβάλλοντα. Καινοτόμα τεχνολογικά προϊόντα έχουν δημιουργήσει την επιθυμία για βελτίωση της ποιότητας του χρόνου που δαπανάται στο σπίτι. Αυτό αποτελεί ένα από τα κίνητρα ανάπτυξης της ΠΝ στο σπίτι. Άλλοι παράγοντες για την ανάπτυξη της ΠΝ στο χώρο του σπιτιού είναι οι τεχνολογικές εξελίξεις και προσδοκίες, καθώς και η αυξανόμενη τάση σε έναν τρόπο ζωής όπου τα όρια ανάμεσα στο σπίτι, την εργασία και τους χώρους ανάπαυσης και ψυχαγωγίας γίνονται δυσδιάκριτα.

Η ΠΝ στο έξυπνο σπίτι ενσωματώνει τη γνώση των αντικειμένων που υπάρχουν σε αυτό, για βελτίωση του καθημερινού τρόπου ζωής. Ορισμένες ιδέες και λειτουργίες αυτή τη στιγμή είναι μακρινά οράματα, όπως ο αυτο-ελεγχόμενος και αυτο-ζωγραφιζόμενος τοίχος, ο φωτισμός, ή έπιπλα τα οποία μπορούν να αναγνωρίζουν τη διάθεση του ατόμου. Σήμερα με την βοήθεια του ΔΤΑ, οι άνθρωποι μπορούν να ελέγχουν τα φώτα ή τη θέρμανση, χρησιμοποιώντας διακόπτες και χειριστήρια [105]. Οι ανάγκες των χρηστών έχουν αναλυθεί σε βάθος και παρέχονται προσωποποιημένες λύσεις και οι εφαρμογές αυτές υπάρχουν ήδη στην αγορά. Τέτοιου είδους όμως λειτουργίες αυτοματισμού δεν ενσωματώνουν κάποια νοημοσύνη.

Στο έξυπνο σπίτι το σύστημα αυτοματισμού θα πρέπει να αναγνωρίζει τον κάτοικο και να προσαρμόζει τις λειτουργίες του σύμφωνα με τις γνωστές προτιμήσεις του ατόμου, λαμβάνοντας υπόψη τις διαθέσεις του καθώς και κάποιες εξωτερικές παραμέτρους. Παραδείγματος χάρη, χώροι που μαθαίνουν τις προτιμήσεις των ατόμων σχετικά με το πότε θέλουν να ανοίξουν ή να κλείσουν τα παράθυρα ή την αυτόματη ρύθμιση του επιπέδου του φωτισμού ή της θέρμανσης [121]. Ο καθρέφτης του μπάνιου μπορεί να υπενθυμίσει στο άτομο που καθρεφτίζεται, τα φάρμακα που πρέπει να πάρει, ενώ το στερεοφωνικό του αυτοκινήτου μπορεί να συντονιστεί στον ίδιο ραδιοφωνικό σταθμό που άκουγε το άτομο κατά τη διάρκεια του πρωινού του [38].

Το περιβάλλον έξυπνου σπιτιού είναι εξοπλισμένο με αισθητήρες και διακόπτες διαφόρων τύπων για την παρακολούθηση των δραστηριοτήτων και της κίνησης του ατόμου, με σκοπό την έγκαιρη και έγκυρη ενημέρωσή του κατάσταση κινδύνου, όπως πυρκαγιά ή συναγερμός καπνού. Γενικά, στο έξυπνο σπίτι συνδυάζονται συνήθως τρία βασικά στοιχεία, ένα σύνολο αισθητήρων, ένα σύνολο διακοπών για τον έλεγχο των αισθητήρων και εξειδικευμένος εξοπλισμός, όπως υπολογιστικές εγκαταστάσεις οι οποίες συνδέουν τους αισθητήρες και τους διακόπτες μέσω “ευφυών” λογισμικών.

Οι αισθητήρες μπορούν για παράδειγμα, να ανιχνεύσουν αν κάποιος κάθεται σε ένα κάθι-

σμα, ή αν κάποιος που φορούσε ένα ρούχο με ετικέτα αναγνώρισης, να αναγνωρίσουν αν πέρασε κάτω από μια πόρτα που ενώνει δύο δωμάτια. Ωστόσο, οι πληροφορίες αυτές απαιτούν μια μορφή περαιτέρω επεξεργασίας των δεδομένων, για παράδειγμα σχετικά με τον εντοπισμό της θέσης του ατόμου [109] ή τη δραστηριότητά του καθώς το άτομο κάθεται στο κάθισμα [43, 44]. Στις εργασίες [8, 10] έχει προταθεί η χρήση μικρής παρέμβασης στο χώρο του σπιτιού με σκοπό την αποτροπή καταστροφών.

Πολλές εργασίες γύρω από το έξυπνο σπίτι επικεντρώνονται στα υλικά και τους αισθητήρες, όμως υπάρχουν αρκετές εργασίες, στις οποίες οι συγγραφείς υποστηρίζουν ότι οι τεχνικές TN μπορούν να βοηθήσουν την εξέλιξη των έξυπνων σπιτιών, ενσωματώνοντας τις καθιερωμένες και σύγχρονες τεχνικές μηχανικής μάθησης, στην επεξεργασία των πληροφοριών που παρέχονται από τις συσκευές και τους αισθητήρες [9, 18]. Χρησιμοποιώντας κατάλληλες μεθόδους της TN και με την επεξεργασία των παραγόμενων δεδομένων, μπορεί να επιτευχθεί εκμάθηση των συνηθειών των χρηστών [24], παρέχοντας συμβουλές σχετικά με το αν μια τροποποίηση της τυπικής τους καθημερινής δραστηριότητας είναι αποδεκτή από τους ίδιους τους χρήστες [11], συνδυάζοντας χρονικά και πιθανοτικά δεδομένα για την δημιουργία μοντέλων ισχυρών αποφάσεων [9], κτλ.

1.3.2 Οικογενειακή Φροντίδα Ηλικιωμένων και Βοηθητικά Περιβάλλοντα

Ένας άλλος τομέας της ΠΝ που έχει προσελκύσει μεγάλο ερευνητικό ενδιαφέρον, είναι η υποστήριξη της ανεξάρτητης διαβίωσης των ηλικιωμένων. Η καλύτερη υγειονομική περίθαλψη και η βελτίωση της ποιότητας ζωής των ανθρώπων έχουν προκαλέσει αύξηση της μακροζωίας. Το 2002 εκτιμήθηκε ότι το 2004 στην Ευρώπη οι ηλικιωμένοι αποτελούσαν το 12-17% του συνολικού πληθυσμού [4]. Στο Ηνωμένο Βασίλειο ο πληθυσμός άνω των 65 ετών είναι τώρα σε ποσοστό 15%, έναντι 11% το 1951 και 5% το 1911². Ο αριθμός των ατόμων άνω των 60 ετών είναι μεγαλύτερος από τον αριθμό των ατόμων ηλικίας 16 ετών. Πολλές μελέτες έχουν επισημάνει μια ευρέως συνεχιζόμενη αύξηση του πληθυσμού των ηλικιωμένων. Στις ΗΠΑ η ηλικιακή ομάδα των ανθρώπων άνω των 65 ετών είναι η ταχύτερα αναπτυσσόμενη ηλικιακή ομάδα σε αριθμό και μέχρι το 2020 αναμένεται να αντιπροσωπεύουν το 1/6 του πληθυσμού [25].

Σύμφωνα με το [5] ‘...το οικονομικό κόστος υποστήριξης ενός ηλικιωμένου ατόμου είναι πολύ μεγαλύτερο από το αντίστοιχο κόστος υποστήριξης ενός παιδιού, συγκεκριμένα είναι ο λόγος πέντε τρίτων. Μεγάλο μέρος αυτού του πρόσθετου κόστους προέρχεται από το υψηλότερο κόστος για την υγεία...’. Επιπλέον, πολλά ηλικιωμένα άτομα (κυρίως του Δυτικού κόσμου) ζουν σε ξεχωριστά καταλύματα και όχι μαζί με τα παιδιά τους. Αυτή η τάση αυξάνει το κόστος υποστήριξης του ηλικιωμένου, ειδικά όταν προστίθενται επιπλέον υγειονομικές υπηρεσίες, όπως η εξειδικευμένη υγειονομική περίθαλψη. Στο ίδιο άρθρο εκτιμάται ότι μέχρι το 2050 ο παγκόσμιος πληθυσμός ηλικιωμένων (άνω των 60 ετών) θα είναι μεγαλύτερος από τον πληθυσμό των παιδιών (0-14 ετών). Ο ΟΗΕ χρησιμοποιεί τον όρο Λόγος Δυνητικός Υποστήριξης (ΛΔΥ) των ηλικιωμένων, που είναι το πλήθος των ατόμων ηλικίας 15-64 ετών

²Centre for Economic Policy Research, <http://www.cepr.org/pubs/bulletin/meets/416.htm>.

ανά άτομο ηλικίας 65 ετών και άνω. Μεταξύ 1950 και το 2002 ο ΛΔΥ μειώθηκε από 12 σε 9 και αναμένεται να πέσει στα 4 μέχρι τα μέσα αιώνα.

Έτσι, κύριος στόχος της έρευνας της ΠΝ στον τομέα φροντίδα των ηλικιωμένων και βοηθητικά περιβάλλοντων είναι η υποστήριξη των ηλικιωμένων ατόμων ώστε να ζούν ανεξάρτητα. Ιδιαίτερα άτομα που μπορεί να βρίσκονται σε πρώιμα στάδια μιας γνωστικής ή σωματικής βλάβης αλλά είναι σε θέση να ζήσουν στο σπίτι τους, όμως χρειάζονται μια συνεχή παρακολούθηση για λόγους ασφάλειας. Οι ακόλουθες εργασίες [93, 12, 112] απεικονίζουν ένα ευρύ φάσμα δραστηριοτήτων σε αυτό τον τομέα, με συστήματα που αξιοποιούν τεχνολογίες αιχμής, σε ένα θεωρητικό επίπεδο, καθώς και μελέτες στοχευμένες σε ομάδες ατόμων.

1.3.3 Ηλεκτρονική Ιατρική Φροντίδα

Ο τομέας της ηλεκτρονικής ιατρικής φροντίδας αποτελεί έναν συνεχώς εξελισσόμενο τομέα της πληροφορικής και της ιατρικής επιστήμης. Ιατροί μπορούν να βλέπουν απομακρυσμένα ακτινογραφίες μέσω κατάλληλων λογισμικών ή να εκτελούν χειρουργικές επεμβάσεις μέσω ρομπότ [108]. Η εξέλιξη του αποτελεί έναν κύριο κλάδο της ΠΝ, με την ενσωμάτωση ευφύων λογισμικών στους αισθητήρες βιολογικών και ιατρικών σημάτων, για τον αυτόματο εντοπισμό ασθενειών με σκοπό την υπόδειξή τους στον ιατρό.

Καθώς οι ιατρικοί και βιολογικοί αισθητήρες για τη μέτρηση διαφόρων αναπνευστικών, βιοχημικών (π.χ. επίπεδα γλυκόζης) ή βιοϊατρικών σημάτων (π.χ. ηλεκτροκαρδιογράφημα) εξελίσσονται και μπορούν να είναι φορητοί, αποδεδεσμεύουν το άτομο από την φυσική του παρουσία στα εξειδικευμένα ιατρικά κέντρα για την διεξαγωγή των αντίστοιχων εξετάσεων. Η διασύνδεση των συσκευών μέσω ασύρματης επικοινωνίας και χρήση ευφύων λογισμικών, μπορούν να διεξάγουν μια πρώτη διάγνωση κατάστασης του ατόμου ενημερώνοντας τον προσωπικό ιατρό του με αυτόματο τρόπο. Στην εργασία [53] αναπτύχθηκε ένα ευφύες φορητό ηλεκτροκαρδιογράφημα για την έγκαιρη ανίχνευση καρδιακών παθήσεων. Το σύστημα αυτό, παρακολουθεί και ανιχνεύει ανωμαλίες και παράγει διαφορετικά επίπεδα συναγερμού, ενημερώνοντας κατάλληλα παρόχους ιατρικής φροντίδας μέσω ασύρματης επικοινωνίας.

Έξυπνα συστήματα υποστήριξης αποφάσεων [39, 40, 129] έχουν χρησιμοποιηθεί ευρέως στον τομέα της ιατρικής περίθαλψης για την παροχή βοήθειας στους ιατρούς και άλλους επαγγελματίες υγείας, με σκοπό την υπόδειξη παθολογικών καταστάσεων του ατόμου για την λήψη της τελικής τους απόφασης, μέσω της ανάλυσης των δεδομένων του ασθενούς [90, 104].

Κεφάλαιο 2

Νευρωνικά Δίκτυα Συνελίξεων

Όπως αναφέραμε στο Κεφάλαιο 1, η Υπολογιστική Νοημοσύνη βασίζεται στους μηχανισμούς της φύσης για τη δημιουργία μεθόδων και αλγορίθμων, που σκοπό έχουν την επίλυση προβλημάτων του πραγματικού κόσμου, προσομοιώνοντας πολλές φορές τις γνωστικές διαδικασίες του ανθρώπου. Κύριοι πυλώνες της Υπολογιστικής Νοημοσύνης είναι τα *Νευρωνικά Δίκτυα*, τα *Ασαφή Συστήματα* και οι *Εξελικτικοί Αλγόριθμοι*. Ωστόσο καθώς αποτελεί ένα συνεχώς εξελισσόμενο ερευνητικό πεδίο, πολλές νέες μέθοδοι και αλγόριθμοι εμπνευσμένοι από τους μηχανισμούς της φύσης έχουν παρουσιαστεί και εδραιωθεί, συνθέτοντας ένα πολυδιάστατο σύνολο μεθόδων που σκοπό έχουν την ενσωμάτωση ευφυίας στο υπολογιστικό σύστημα για την επίλυση προβλημάτων.

Η Υπολογιστική Νοημοσύνη διαδραματίζει σημαντικό ρόλο στην ανάπτυξη έξυπνων συστημάτων αλλά και γνωστικών αναπτυξιακών συστημάτων. Τα τελευταία χρόνια υπήρξε μια “έκρηξη”, σε ερευνητικό επίπεδο, γύρω από τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) μέσω της Βαθιάς Εκμάθησης (Deep Learning) και συγκεκριμένα με τα Τεχνητά Νευρωνικά Δίκτυα Συνελίξεων (ΤΝΔΣ). Σήμερα, η Βαθιά Εκμάθηση αποτελεί βασική μέθοδο της ΤΝ, ενώ μερικά από τα πιο επιτυχημένα συστήματα ΤΝ βασίζονται στα ΤΝΔΣ.

2.1 Εισαγωγή

Τα ΤΝΔΣ είναι μια μέθοδος της Μηχανικής Μάθησης που χρησιμοποιείται σε πολλούς και διαφορετικούς ερευνητικούς τομείς (ταξινόμηση εικόνων [80], αναγνώριση αντικειμένων [52], αναγνώριση ανωμαλιών σε ιατρικές εικόνες [46], αναγνώριση κειμένων [50], κ.α.) βρίσκοντας εφαρμογή και στη βιομηχανία, με σκοπό την επίλυση προβλημάτων του πραγματικού κόσμου με την βοήθεια έξυπνων μηχανών [42, 81]. Τα ΤΝΔΣ αποτελούν την εξέλιξη των κλασικών Τεχνητών Νευρωνικών Δικτύων [60], χρησιμοποιώντας αντί των τυπικών τεχνητών νευρώνων, δομές και τελεστές περισσότερο σύνθετους με βασικότερο τελεστή την πράξη της συνελίξεως.

Κύριο χαρακτηριστικό των ΤΝΔΣ είναι η ικανότητά τους να επεξεργάζονται δεδομένα όχι μόνο διανυσματικής μορφής, δηλαδή δεδομένα τοποθετημένα σειριακά, μιας μορφής η οποία μπορεί να ερμηνευτεί τοπολογικά ως ένα μονοδιάστατο πλέγμα, όπου τα δεδομένα τοποθετούνται το ένα δίπλα στο άλλο, αλλά και μορφές μητρώου ή δισδιάστατου πλέγματος. Ένα

παράδειγμα δεδομένων μονοδιάστατου πλέγματος είναι τα δεδομένα χρονοσειρών, ενώ η δομή που απαρτίζεται από τα εικονοστοιχεία μιας εικόνας αποτελεί ένα διδιάστατο πλέγμα. Καθώς ο βασικός τελεστής που εφαρμόζεται στα δεδομένων των ΤΝΔΣ είναι η πράξη της συνέλιξης, η οποία εφαρμόζεται διαδοχικά σε μικρές γειτονίες των δεδομένων, τα ΤΝΔΣ ενσωματώνουν μια βασική ιδιότητα να διατηρούν την τοπική πληροφορία που εμπεριέχεται στα δεδομένα που επεξεργάζονται [17].

Η επεξεργασία των δεδομένων στα ΤΝΔΣ ακολουθεί την φιλοσοφία των ΤΝΔ και οργανώνεται σε πολλαπλά επίπεδα στρωμάτων, συνδυάζοντας τους τυπικούς νευρώνες των ΤΝΔ με μια νέα δομή, τα μητρώα χαρακτηριστικών (feature maps). Τα μητρώα χαρακτηριστικών των διαδοχικών επιπέδων συνδέονται μεταξύ τους με ομάδες βαρών. Τα ΤΝΔΣ θα μπορούσαν να χαρακτηριστούν ως ένας μαζικά παράλληλος επεξεργαστής, με την έμφυτη ιδιότητα να δημιουργούν και να απομνημονεύουν εμπειρική γνώση την οποία θα μπορούν να την χρησιμοποιήσουν στο μέλλον. Ο τρόπος απόκτησης της εμπειρικής γνώσης προσομοιώνεται με τον τρόπο απόκτησης της γνώσης από τον ίδιο τον άνθρωπο μέσω ενός μηχανισμού μάθησης, που στην Υπολογιστική Νοημοσύνη είναι ο αλγόριθμος εκπαίδευσης. Κάνοντας χρήση δεδομένων με γνωστό εκ των προτέρων το επιθυμητό αποτέλεσμα και με την χρήση του αλγόριθμου εκπαίδευσης, το ΤΝΔΣ αποκτά την εμπειρία να μπορεί να αντιληφθεί άλλα άγνωστα σε αυτό δεδομένα μέσω της γνώσης που έχει λάβει.

Η πληροφορία στα ΤΝΔΣ κωδικοποιείται ως μια μαθηματική συνάρτηση, αποτελούμενη από έναν μεγάλο αριθμό ανεξάρτητων στοιχείων-μεταβλητών. Η εκπαίδευση των ΤΝΔΣ στη συνέχεια επιτυγχάνεται μέσω μιας κατηγορίας μεθόδων, την μέθοδο εκπαίδευσης με επίβλεψη. Στην κατηγορία αυτή, η εκπαίδευση του ΤΝΔΣ γίνεται με την επίβλεψη ενός “δασκάλου”, όπου το αποτέλεσμα της εξόδου του ΤΝΔΣ συγκρίνεται με το επιθυμητό, γνωστό εκ των προτέρων αποτέλεσμα των δεδομένων-προτύπων που εισάγονται σε αυτό και υπολογίζεται η διαφορά τους. Η διαφορά αυτή αποτελεί το σφάλμα εκπαίδευσης του δικτύου. Στη συνέχεια τα βάρη του ΤΝΔΣ μεταβάλλονται με τέτοιο τρόπο έτσι ώστε στην επόμενη επανάληψη του αλγορίθμου εκπαίδευσης, η τιμή του σφάλματος να μειωθεί.

Έστω η συνάρτηση σφάλματος (αντικειμενική συνάρτηση) E που ορίζεται ως το άθροισμα, όλων των προτύπων εισόδου, των τετραγώνων των διαφορών της εξόδου του ΤΝΔΣ με την επιθυμητή τιμή ανά πρότυπο,

$$E(w) = \frac{1}{n} \sum_{p=1}^P \sum_{j=1}^{N_L} (o_{j,p}^L - y_{j,p})^2, \quad (2.1)$$

όπου P είναι ο συνολικός αριθμός των προτύπων, $y_{j,p}^L$ η έξοδος του j νευρώνα/μητρώου χαρακτηριστικών που ανήκει στο L επίπεδο (αναλόγως αν βρίσκεται στο επίπεδο των συνελίξεων ή στο επίπεδο των τυπικών νευρώνων), N_L ο αριθμός των νευρώνων στο επίπεδο εξόδου και $o_{j,p}$ η επιθυμητή έξοδος του j νευρώνα στο πρότυπο p . Το ζητούμενο είναι η ελαχιστοποίηση της συνάρτησης σφάλματος E , η εύρεση δηλαδή ενός διανύσματος $w^* = (w_1^*, w_2^*, \dots, w_n^*) \in \mathbb{R}^n$ έτσι ώστε:

$$w^* = \min_{w \in \mathbb{R}^n} E(w). \quad (2.2)$$

Η αναγωγή του προβλήματος εκπαίδευσης σε πρόβλημα εύρεσης του βέλτιστου μιας συνάρτησης ελαχίστων τετραγώνων, έχει πολλά πλεονεκτήματα καθώς υπάρχουν πολλές μέθοδοι εύρεσης ελαχίστου, ενώ επιπλέον γνωρίζουμε το ολικό ελάχιστο της συνάρτησης.

Η εκπαίδευση του ΤΝΔΣ επιτυγχάνεται μέσω της ελαχιστοποίησης της συνάρτησης σφάλματος με την διαδοχική μεταβολή των βαρών του δικτύου. Το διάνυσμα μεταβολής των βαρών δείχνει την κατεύθυνση του διανύσματος στο χώρο των βαρών που δημιουργείται από τα πρότυπα εκπαίδευσης, με σκοπό την ελαχιστοποίηση της συνάρτησης E . Σύμφωνα με την μέθοδο απότομης κλίσης [16] τα βάρη μεταβάλλονται με τον ακόλουθο τύπο:

$$w^{k+1} = w^k + \delta w^k, k = 0, 1, \dots, \quad (2.3)$$

όπου w^{k+1} είναι το νέο διάνυσμα των βαρών, w^k το τρέχον διάνυσμα των βαρών, ενώ δw^k είναι το διάνυσμα μεταβολής. Το διάνυσμα μεταβολής υπολογίζεται χρησιμοποιώντας την πληροφορία της πρώτης τάξης παραγώγων της συνάρτησης των βαρών. Συγκεκριμένα, το διάνυσμα μεταβολής με την μέθοδο απότομης κλίσης ορίζεται ως:

$$\nabla E(w^k) = \left(\frac{\partial E}{\partial w^{k,1}}, \dots, \frac{\partial E}{\partial w^{k,n}} \right), \quad (2.4)$$

πολλαπλασιασμένο με μια παράμετρο η , που ονομάζεται ρυθμός εκμάθησης. Έτσι τα βάρη ανανεώνονται σύμφωνα με επόμενο τύπο:

$$w^{k+1} = w^k - \eta \nabla E(w^k). \quad (2.5)$$

2.2 Ο Τελεστής Συνέλιξης

Όπως αναφέραμε στην προηγούμενη ενότητα, βασικός τελεστής των ΤΝΔΣ είναι ο τελεστής της “Συνέλιξης”. Η συνέλιξη είναι μια μαθηματική πράξη μεταξύ δύο συναρτήσεων (f, g) που παράγει μια τρίτη συνάρτηση (s), η οποία εκφράζει την αλλαγή που υφίσταται η πρώτη συνάρτηση από τη δεύτερη. Η πράξη αυτή εφαρμόζεται σε σήματα είτε μονοδιάστατα είτε μεγαλύτερης διάστασης και αναπαρίσταται με τον τελεστή “*”. Έτσι έχουμε την πράξη της συνέλιξης όπως ορίζεται στη συνέχεια,

$$s(t) = (f * g)(t) = \int f(a)g(t-a)da, \quad (2.6)$$

για συνεχείς συναρτήσεις. Ωστόσο, στα ΤΝΔΣ οι πράξεις σχετίζονται με διακριτές τιμές δεδομένων. Έτσι η εξίσωση 2.6 μετασχηματίζεται σε

$$s(t) = (f * g)(t) = \sum_a f(a)g(t-a), \quad (2.7)$$

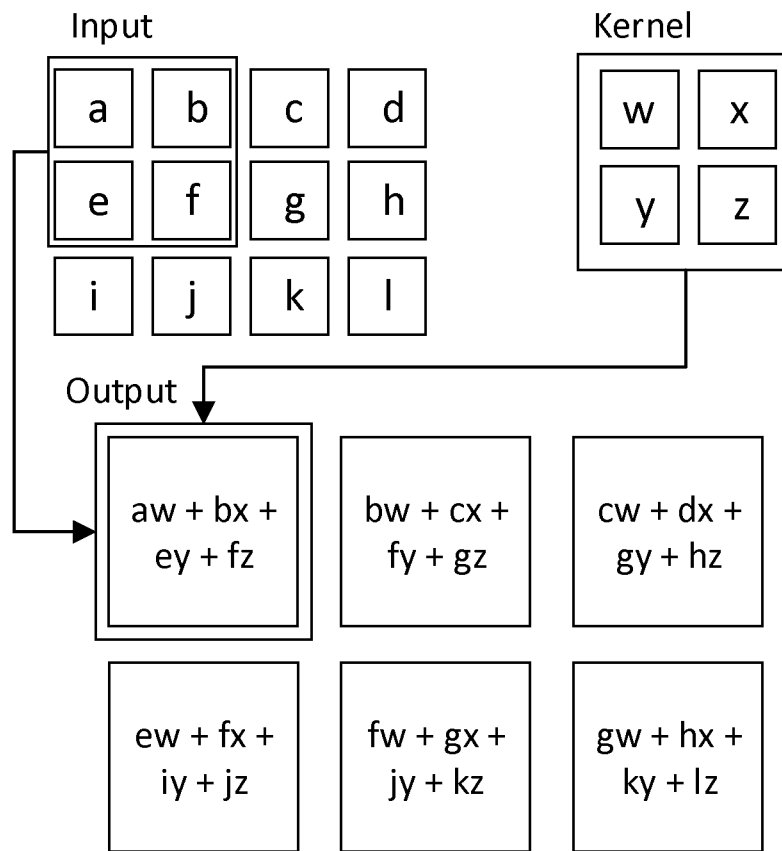
όπου a το πλήθος των διακριτών τιμών και t το στιγμιότυπο του μονοδιάστατου σήματος στο χρόνο. Η εξίσωση 2.7 χαρακτηρίζεται ως διακριτή συνέλιξη.

Η συνάρτηση f ορίζεται ως η είσοδος του τελεστή συνέλιξης, ενώ η συνάρτηση g ως το φίλτρο (συνάρτηση πυρήνα) που επιδρά στην είσοδο του τελεστή. Τέλος το αποτέλεσμα της συνέλιξης ορίζεται ως το μητρώο χαρακτηριστικών.

Σε περιπτώσεις όπου το σήμα που αναλύεται δεν είναι μονοδιάστατο αλλά διδιάστατο, όπως για παράδειγμα το σήμα εικόνων, τότε το φίλτρο που εφαρμόζεται στην είσοδο του τελεστή είναι διάστασης δύο. Ο τελεστής συνέλιξης μετασχηματίζεται σε

$$s(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n), \quad (2.8)$$

όπου I είναι η εικόνα εισόδου με (i, j) η θέση των εικονοστοιχείων, K η διδιάστατη συνάρτηση φίλτρου. Στην Εικόνα 2.1, παρουσιάζεται ένα παράδειγμα συνέλιξης σε ένα διδιάστατο σήμα.



Σχήμα 2.1: Σχηματική αναπαράσταση συνέλιξης σε διδιάστατα δεδομένα.

Εξετάζοντας την πράξη της συνέλιξης σε διακριτά δεδομένα, η πράξη αυτή μπορεί να αναπαρασταθεί ως ένας πολλαπλασιασμός μητρώων. Έστω η πράξη της διακριτής συνέλιξης σε μονοδιάστατα δεδομένα, το μητρώο εισόδου μπορεί να μετασχηματιστεί σε ένα μητρώο Toeplitz, όπου κάθε στήλη είναι ίδια με την προηγούμενη μετατοπισμένη κατά ένα στοιχείο συμπληρώνοντας με μηδενικά τα υπόλοιπα. Η συνέλιξη με διδιάστατα δεδομένα μπορεί να επιτευχθεί κάνοντας χρήση μιας ειδικής μορφής μητρώου Toeplitz, όπου κάθε στοιχείο του μητρώου είναι ένα υπομητρώο Toeplitz. Καθώς το φίλτρο είναι πολύ μικρότερο από το μητρώο

εισόδου, πολλά από τα στοιχεία του μητρώου Toeplitz είναι μηδενικά, έτσι έχουμε να κάνουμε υπολογιστικές πράξεις με αραιά μητρώα.

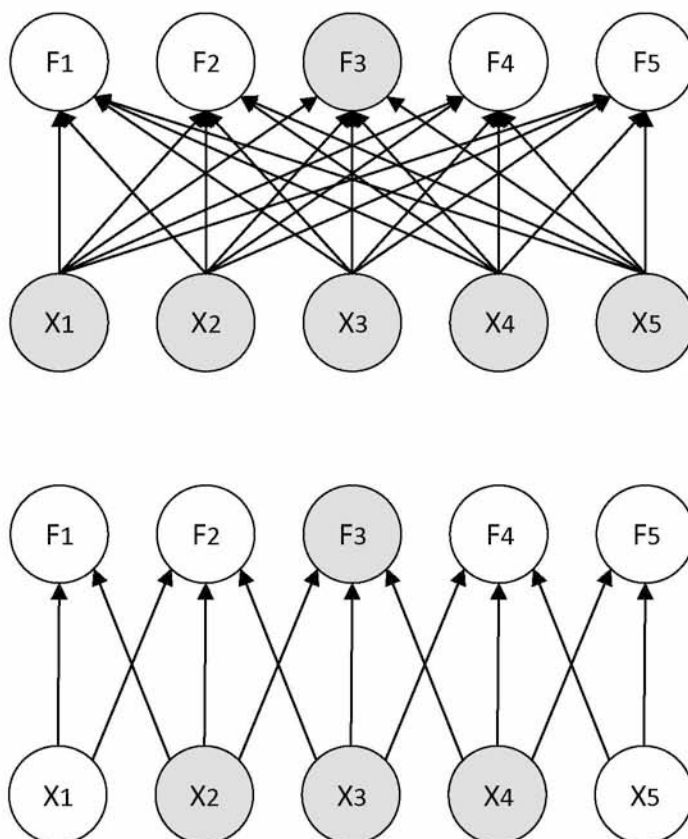
2.3 Βασικές Αρχές των ΤΝΔΣ

Τα ΤΝΔΣ ενσωματώνουν και αξιοποιούν τρεις βασικές αρχές με σκοπό την δημιουργία ενός μοντέλου μηχανικής μάθησης ικανού για την επίλυση προβλημάτων του πραγματικού κόσμου: τις αραιές συνδέσεις, την χρήση κοινών παραμέτρων και συναρτήσεις με την ιδιότητα της ισοδυναμίας.

Τα τυπικά ΤΝΔ δέχονται ως είσοδο δεδομένα μονοδιάστατου πλέγματος όπου κάθε στοιχείου του συνδέεται με όλους τους νευρώνες που πρώτου επιπέδου νευρώνων. Οι συνδέσεις αυτές αναφέρονται ως βάρη του δικτύου. Στα ΤΝΔΣ τα οποία δέχονται ως είσοδο όχι μόνο δεδομένα μονοδιάστατου πλέγματος αλλά και διδιάστατου, όπως είναι τα εικονοστοιχεία μιας εικόνας, τα στοιχεία εισόδου τους δεν συνδέονται ένα προς ένα με τους αντίστοιχους νευρώνες του επόμενου επιπέδου, όπου στα ΤΝΔΣ ονομάζονται μητρώα χαρακτηριστικών. Αντίθετα, χρησιμοποιούνται ομάδες βαρών που απαρτίζουν τη συνάρτηση πυρήνα όπου συνδέουν τα δεδομένα εισόδου με κάθε μητρώο χαρακτηριστικών του επόμενου επιπέδου. Τέλος, εφαρμόζοντας την πράξη της συνέλιξης μεταξύ των δεδομένων εισόδου και κάθε πυρήνα παράγονται τα μητρώα χαρακτηριστικών. Αυτό το χαρακτηριστικό των ΤΝΔΣ ονομάζεται αραιές συνδέσεις των βαρών του δικτύου και είναι ιδιαίτερα σημαντικό καθώς στα ΤΝΔΣ τα δεδομένα εισόδου είναι πολλές χιλιάδες ή εκατοντάδες στοιχεία και με τον χαρακτηριστικό αυτό επιτυγχάνεται μια σημαντική μείωση στις απαιτήσεις του μοντέλου ως προς τη μνήμη αλλά και στο υπολογιστικό κόστος συγκρινόμενα με τα ΤΝΔ. Συγκεκριμένα θεωρώντας ότι έχουμε m πλήθος συνδέσεων και n πλήθος νευρώνων ή μητρώων χαρακτηριστικών, τότε απαιτείται ο πολλαπλασιασμός ενός μητρώου παραμέτρων $m \times n$, έτσι η τάξη του αλγόριθμου εκτέλεσης θα είναι $O(m \times n)$. Αν όμως περιορίσουμε το πλήθος των συνδέσεων που μπορεί να έχει μια έξοδος σε k , κάνοντας χρήση αραιών συνδέσεων θα απαιτούσε τάξη εκτέλεσης $O(k \times n)$, όπου η τιμή της παραμέτρου $k \ll m$.

Ένα άλλο χαρακτηριστικό που αξιοποιούν τα ΤΝΔΣ και συνδέεται άμεσα με το προηγούμενο χαρακτηριστικό τους είναι η χρήση κοινών παραμέτρων. Στα ΤΝΔ κάθε στοιχείο του μητρώου των βαρών χρησιμοποιείται μια και μόνο μια φορά για τον υπολογισμό των δεδομένων του επιπέδου εξόδου. Συγκεκριμένα καθώς κάθε βάρος χρησιμοποιείται από ένα στοιχείο των δεδομένων εισόδου, τα βάρη αυτά θα μπορούσαν να χαρακτηριστούν ως αυστηρά προσδιορισμένα βάρη ως προς τα στοιχεία του επιπέδου εισόδου. Στα ΤΝΔΣ κάθε στοιχείο της συνάρτησης πυρήνα αλληλεπιδρά με όλα τα στοιχεία των δεδομένων εισόδου. Έτσι χρησιμοποιούνται κοινά σύνολα παραμέτρων, όπου αντί να “μαθαίνουν” ξεχωριστά για το κάθε τμήμα του μητρώου δεδομένων εισόδου, εκπαιδεύονται και “μαθαίνουν” ένα κοινό σύνολο. Αυτό το χαρακτηριστικό δεν επηρεάζει τον χρόνο εκτέλεσής του ο οποίος παραμένει τάξης $O(k \times n)$, μειώνει όμως το κόστος αποθήκευσης του μοντέλου σε k παραμέτρους. Όπως αναφέρεται στην προηγούμενη παράγραφο, η παράμετρος k είναι τάξης μεγέθους σημαντικά μικρότερης από την παράμετρο m . Έτσι η πράξη της συνέλιξης είναι σημαντικά αποδοτικότερη ως προς

τις απαιτήσεις σε μνήμη. Στη εικόνα 2.2 φαίνεται ένα παράδειγμα χρήσης κοινών παραμέτρων των ΤΝΔΣ έναντι των παραμέτρων ενός τυπικού ΤΝΔ.



Σχήμα 2.2: Αναπαράσταση της αλληλεπίδρασης ενός κρυφού νευρώνα (F_3) με τα δεδομένα εισόδου (X). Παρατηρούμε ότι στην περίπτωση των ΤΝΔ (πάνω) ο νευρώνας F_3 αλληλεπιδρά με όλα τα στοιχεία της εισόδου, οπότε απαιτούνται πέντε βάρη να εκπαιδευτούν. Στην περίπτωση των ΤΝΔΣ (κάτω) με συνάρτηση πυρήνα διάστασης 3 το στοιχείο του μητρώου χαρακτηριστικών F_3 αλληλεπιδρά μόνο με τρία στοιχεία του συνόλου δεδομένων X .

Τέλος, μια επιπλέον ιδιότητα που έχει η συνάρτηση συνέλιξης την οποία αξιοποιούν τα ΤΝΔΣ, είναι η δυνατότητα τα δεδομένα πριν και μετά την εφαρμογή της συνέλιξης να παραμένουν ισοδύναμα μετά από όμοια εφαρμογή μετατόπισης - μετακίνησης των στοιχείων τους. Μια συνάρτηση f είναι ισοδύναμη με μια συνάρτηση g αν $f(g(x)) = g(f(x))$. Θεωρώντας ως παράδειγμα μια (συνάρτηση) εικόνα I και $I' = g(I)$ μια εικόνα μετασχηματισμένη από την συνάρτηση g , όπου $I'(x, y) = I(x - 1, y)$, μια μετατόπιση του κάθε εικονοστοιχείου κατά μια μονάδα δεξιότερα. Αν εφαρμόσουμε τον τελεστή συνέλιξης στην εικόνα I , το αποτέλεσμα θα είναι το ίδιο με το αν εφαρμόσουμε τον τελεστή συνέλιξης στην I' και στη συνέχεια προβούμε στο μετασχηματισμό g . Η ιδιότητα αυτή είναι εξαιρετικά χρήσιμη στα ΤΝΔΣ καθώς μπορούν να αναγνωρίζουν συγκεκριμένα χαρακτηριστικά, που μπορεί να έχουν εντοπίσει σε ένα τμήμα

των δεδομένων που επεξεργάζονται, σε διαφορετικές θέσεις στα δεδομένα αυτά. Για παράδειγμα ο εντοπισμός ακμών σε μια εικόνα μπορεί να επιτευχθεί από τον ίδιο πυρήνα σε άλλη εικόνα όπου τα εικονοστοιχεία που απεικονίζουν μια ακμή έχουν υποστεί μετατόπιση.

Μέρος II

Εντοπισμός της Ανθρώπινης Στάσης μέσω Στατικής Κάμερας και Αισθητήρων Έξυπνων Συσκευών

Κεφάλαιο 3

Εντοπισμός της Ανθρώπινης Στάσης από Ευρυγώνια Κάμερα

Στο Κεφάλαιο αυτό θα παρουσιάσουμε μια καινούρια μέθοδο εξαγωγής χαρακτηριστικών από εικόνες ευρυγώνιας κάμερας, για τον αυτόματο εντοπισμό της ανθρώπινης στάσης. Σκοπός του κεφαλαίου είναι η θεμελίωση της μεθόδου εξαγωγής χαρακτηριστικών η οποία βασίζεται στις ροπές Zernike και η επίλυση του προβλήματος εντοπισμού της ανθρώπινης στάσης, με τη βοήθεια χαρακτηριστικών που εξάγονται χρησιμοποιώντας αυστηρά και μόνο ως πληροφορία την σιλουέτα του ανθρώπου.

Η προτεινόμενη μέθοδος εξαγωγής χαρακτηριστικών ονομάζεται “Γεωδαιτική Διόρθωση των Ροπών Zernike” (ΓΔΡΖ). Βασίζεται στις “κλασικές Ροπές Zernike” (ΚΡΖ) και στην αντικατάσταση της Ευκλείδειας απόστασης που χρησιμοποιούν ως μετρική με τη Γεωδαιτική απόσταση. Αρχικά, χρησιμοποιώντας μια ευρυγώνια κάμερα θα εντοπίσουμε συγκεκριμένες στάσεις του ανθρώπου χρησιμοποιώντας μόνο την πληροφορία της σιλουέτας του. Για να το επιτύχουμε, αρχικά θα δημιουργήσουμε ένα συνθετικό σύνολο δεδομένων εκπαίδευσης, τοποθετώντάς το τεχνητά στο χώρο που βρίσκεται η κάμερα. Στη συνέχεια, θα εξάγουμε χαρακτηριστικά από το σύνολο εκπαίδευσης με την μέθοδο ΓΔΡΖ και θα εκπαιδεύσουμε έναν ταξινομητή. Τέλος, θα ελέγξουμε το μοντέλο μας σε εικόνες από ελήφθησαν από την πραγματική ευρυγώνια κάμερα.

Η προτεινόμενη μέθοδος εξόρυξης χαρακτηριστικών, βελτιώνει τα αποτελέσματα των κλασικών ροπών Zernike σε εικόνες από ευρυγώνια κάμερα, για τον εντοπισμό της ανθρώπινης στάσης όπως παρουσιάζονται στις εργασίες [27, 29].

3.1 Εισαγωγή

Ο αυτόματος εντοπισμός του ανθρώπου και κατ’ επέκταση η χρήση τεχνικών Μηχανικής Μάθησης για την αναγνώριση της δραστηριότητας του ανθρώπου, παρουσιάζει μεγάλο ερευνητικό ενδιαφέρον τα τελευταία χρόνια. Αυτό φαίνεται τόσο από την ερευνητική δραστηριότητα στον τομέα αυτόν όσο και από την εμπορική τάση της εποχής για παραγωγή συσκευών, συ-

στημάτων¹ και εφαρμογών που θα μπορούν να εντοπίσουν, πιθανή πτώση ενός ανθρώπου (ιδίως ηλικιωμένων [125]) ή την δραστηριότητα ενός ανθρώπου κατά την διάρκεια της ημέρας (παραδείγματος χάρη τον ορθότερο υπολογισμό των θερμίδων που έχει καταναλώσει εντός της ημέρας [120, 141]). Το πρόβλημα αναγνώρισης της ανθρώπινης δραστηριότητας αποτελεί ένα ανοιχτό πρόβλημα και αντικείμενο μελέτης μεγάλων ερευνητικών έργων σύμφωνα με τα οποία ο άνθρωπος μελετάται ολιστικά [85].

Η επίλυση του προβλήματος αναγνώρισης της ανθρώπινης δραστηριότητας, εντάσσεται στη γενικότερη κατηγορία επίλυσης προβλημάτων αναγνώρισης προτύπων της υπολογιστικής όρασης. Διαδεδομένες προσεγγίσεις σε αυτού του τύπου προβλήματα, είναι η χρήση μεθόδων εξαγωγής χαρακτηριστικών από τις εικόνες [41, 63, 102, 115, 126]. Μια κατηγορία μεθόδων εξαγωγής χαρακτηριστικών και συγκεκριμένα ολικών χαρακτηριστικών [59] είναι ο υπολογισμός των ροπών στις εικόνες [91]. Βασική ιδιότητα των χαρακτηριστικών αυτών είναι ότι παραμένουν αναλλοίωτα στις στροφές που μπορεί το αντικείμενο της εικόνας να υποστεί, καθώς και στην αλλαγή της θέσης του μέσα στην εικόνα.

Η χρήση κάμερας για την αναγνώριση της ανθρώπινης δραστηριότητας είναι ένας κλάδος που συνεχώς εξελίσσεται. Η μεταβλητότητα που παρουσιάζει το πρόβλημα ως προς τον τύπο και το πλήθος των καμερών που θα χρησιμοποιήσουμε, την τοποθέτησή τους στο χώρο, αλλά και ο ίδιος ο χώρος στον οποίο θέλουμε να εντοπίσουμε την δραστηριότητα αυτή, κάνει μη δυνατή την ύπαρξη μοναδικής λύσης του προβλήματος.

Σε ένα τυπικό δωμάτιο υπάρχουν αντικείμενα τα οποία διακόπτουν το οπτικό πεδίο μιας τυπικής κάμερας (ως προς τον τύπο του φακού της) με γωνία θέασης περίπου 90° , που είναι τοποθετημένη στον ορίζοντα του δωματίου. Οπότε ο εντοπισμός της δραστηριότητας του ανθρώπου γίνεται δύσκολος και η χρήση περισσότερων της μιας κάμερα στον ίδιο χώρο γίνεται επιτακτική. Αυτό έχει ως αποτέλεσμα την αύξηση του οικονομικού κόστους του συστήματος καμερών όσο και των υπολογιστικών πόρων, καθώς θα πρέπει να γίνονται ταυτόχρονοι υπολογισμοί στα στιγμιότυπα όλων των καμερών. Έτσι, θεωρούμε ότι μια καλή λύση για την αντιμετώπιση των εμποδίων μέσα σε ένα δωμάτιο/βιομηχανικό χώρο εργασίας, είναι η χρήση ευρυγώνιων φακών με δυνατότητα θέασης 180° , τοποθετημένες στην οροφή του δωματίου.

Στις κάμερες αυτού του τύπου, η απόσταση μεταξύ δύο γειτονικών εικονοστοιχείων δεν μπορεί να οριστεί με όρους συντεταγμένων επί του επιπέδου της εικόνας, καθώς η εικόνα δεν λαμβάνεται μέσω απλής επίπεδης προβολής, αλλά χρησιμοποιώντας ένα σφαιρικό στοιχείο με οπτικό πεδίο 180° .

3.2 Εξαγωγή Χαρακτηριστικών σε Εικόνες από Ευρυγώνια Κάμερα

Στην ενότητα αυτή παρουσιάζεται η μέθοδος εξαγωγής χαρακτηριστικών από δυαδικές εικόνες, κάνοντας χρήση των ροπών Zernike (ή κλασικές ροπές Zernike) (KPZ) και η επέκταση της μεθόδου σε εικόνες που παράγονται από κάμερες ευρυγώνιας λήψης. Σκοπός είναι η χρη-

¹<https://www.lifeline.philips.com/medical-alert-systems/fall-detection.html>

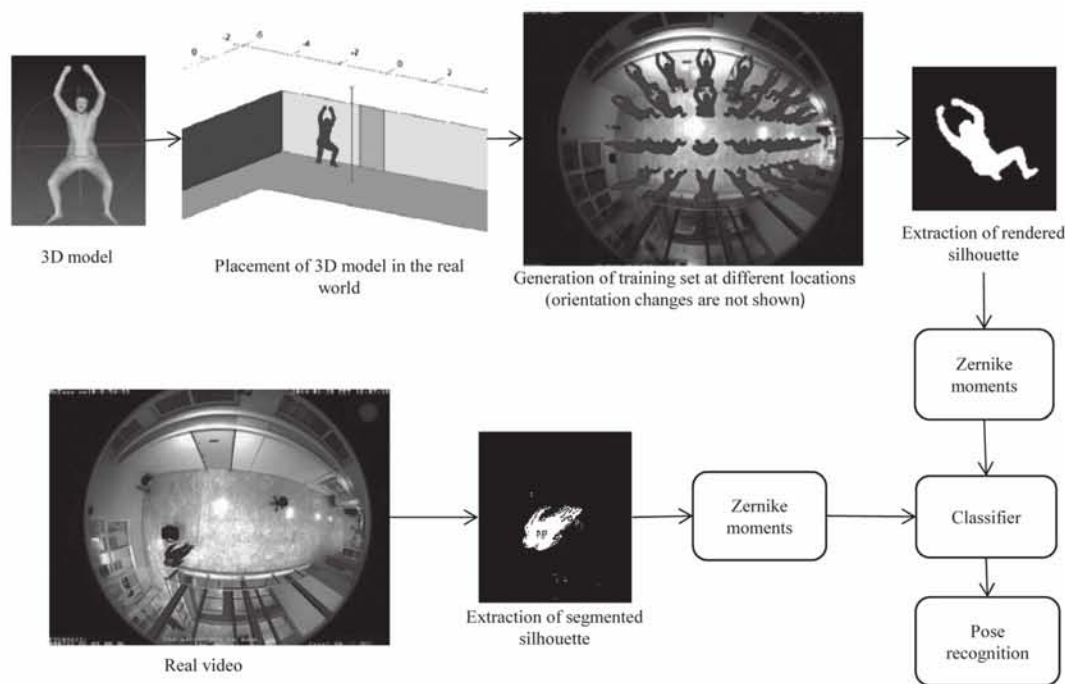


Σχήμα 3.1: Λήψη όρθιας στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (α). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας, στην οποία εφαρμόζεται η μέθοδος εξαγωγής χαρακτηριστικών (β).

σιμοποίηση των χαρακτηριστικών που παράγουν οι KPZ για τον εντοπισμό της ανθρώπινης στάσης μέσω μια στατικής κάμερας. Ένα παράδειγμα ενός ανθρώπου σε όρθια στάση φαίνεται στην Εικόνα 3.1-α, ενώ στην Εικόνα 3.1-β διακρίνεται η σιλουέτα που εξήχθη με την βοήθεια ενός αλγόριθμου τμηματοποίησης ([31]) της εικόνας με αφαίρεση του παρασκήνιου.

Για τον χαρακτηρισμό της στάσης του ανθρώπου θα χρησιμοποιηθεί μια ευρυγώνια κάμερα τοποθετημένη στην οροφή ενός δωματίου, έτσι ώστε να επιτευχθεί το μέγιστο δυνατό της λήψης ολόκληρου του ανθρώπου, χωρίς παρεμβολές του οπτικού πεδίου της κάμερας από τυχόν αντικείμενα που ίσως υπάρχουν στο χώρο. Λόγω των ιδιαίτερων χαρακτηριστικών αυτού του τύπου κάμερας (180° οπτικό πεδίο) θα γίνει προσαρμογή των KPZ χρησιμοποιώντας την γεωδαιτική απόσταση. Στη συνέχεια, κάνοντας χρήση ενός ταξινομητή ο οποίος θα εκπαιδευτεί με εικόνες από διάφορες στάσεις του ανθρώπου, θα εντοπίζει σε πραγματικό χρόνο τις στάσεις του από την ευρυγώνια κάμερα.

Ωστόσο για την εκπαίδευση του ταξινομητή απαιτείται ένα επαρκώς μεγάλο σύνολο δεδομένων με συγκεκριμένες στάσεις του ανθρώπου, το οποίο δεν είναι διαθέσιμο. Για την αντιμετώπιση του προβλήματος αυτού, θα χρησιμοποιηθούν διάφορες στάσεις τριδιάστατων ανθρώπινων μοντέλων, που θα προσαρμοστούν στο χώρο που υπάρχει η κάμερά μας σε διάφορες θέσεις και προσανατολισμούς. Στη συνέχεια εξάγοντας τις σιλουέτες των συνθετικών αυτών μοντέλων και παράγοντας χαρακτηριστικά μέσω των ΓΔΡΖ από αυτές, θα εκπαιδεύσουμε έναν ταξινομητή. Το παραγόμενο εκπαιδευμένο μοντέλο από τον ταξινομητή θα χρησιμοποιηθεί για τον χαρακτηρισμό των στάσεων του ανθρώπου σε ένα πραγματικό βίντεο, αφού πρώτα χρησιμοποιηθεί ένας αλγόριθμος τμηματοποίησης για την εξαγωγή τις σιλουέτας του ανθρώπου, στην οποία θα εφαρμοστεί η ΓΔΡΖ για την παραγωγή χαρακτηριστικών. Στην Εικόνα 3.2 παρουσιάζεται σχηματικά η μεθοδολογία εντοπισμού της ανθρώπινης στάσης. Η προτεινόμενη



Σχήμα 3.2: Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας.

μεθοδολογία εξαγωγής χαρακτηριστικών θα συγκριθεί με της κλασικές ροπές Zernike, καθώς και με μια άλλη μέθοδο ολικών χαρακτηριστικών, τα Ιστογράμματα της Προσανατολισμένης Κλίσης (Histogram of oriented Gradients HoG) [1, 72].

3.2.1 Ροπές Zernike

Οι ροπές Zernike αποτελούν έναν ολικό περιγραφέα χαρακτηριστικών μιας εικόνας, με την ιδιότητα τα χαρακτηριστικά αυτά να παραμένουν αναλλοίωτα στις μεταβολές της (π.χ. στροφές, αλλαγές κλίμακας μετακίνηση του αντικειμένου, κτλ.). Οι ροπές Zernike παράγονται από την ακόλουθη οικογένεια πολυωνύμων $R_{n,m}(r)$ που είναι ανα δύο ορθογώνια στο μοναδιαίο κύκλο και εξαρτώνται στο πλήθος από την τάξη n , m του πολυωνύμου:

$$R_{n,m}(r) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n+|m|}{2} + s\right)!} r^{n-2s}. \quad (3.1)$$

Η μεταβλητή r είναι το μήκος του διανύσματος από το κέντρο του μοναδιαίου κύκλου προς το κάθε εικονοστοιχείο, με εύρος τιμών στο διάστημα $[0, 1]$, n είναι ένας θετικός ακέραιος και m ένας ακέραιος με $n - |m|$ να είναι πάντα άρτιος και $|m| \leq n$. Η ορθοκανονικότητα διατηρείται με τον ακόλουθο τύπο:

$$\int_0^1 R_{n,m}(r) R_{k,m}(r) dr = \frac{1}{2(n+1)} \delta_{n,k}. \quad (3.2)$$

Το πολυώνυμο Zernike, ορίζεται σύμφωνα με τον τύπο:

$$V_{n,m}(r, \gamma) = R_{n,m}(r) e^{jm\gamma}, \quad (3.3)$$

όπου γ ορίζεται η γωνία που σχηματίζει ο άξονας x' του Καρτεσιανού επιπέδου με το διάνυσμα r . Έτσι έχουμε τον τύπο για τις ροπές Zernike για μια συνάρτηση $f(r, \gamma)$ που ορίζεται ως εξής:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^1 \int_0^{2\pi} f(r, \gamma) V_{n,m}(r, \gamma) r dr d\gamma. \quad (3.4)$$

Η εξίσωση 3.4 ωστόσο χρησιμοποιείται και σε διακριτές εικόνες. Αν A ένα υποσύνολο του \mathbb{N}^2 με πλάτος w και βαρύκεντρο (κεντρικό εικονοστοιχείο της εικόνας) το σημείο (x_0, y_0) , με $A = [x_0 - w, x_0 + w] \times [y_0 - w, y_0 + w]$ η ροπή Zernike της εικόνας f στο A υπολογίζεται και κάθε τιμή των παραμέτρων n, m με τον ακόλουθο τύπο:

$$Z_{nm}(x_0, y_0) = \frac{n+1}{\pi} \sum_{(x,y) \in A} f(r(x, y), \gamma(x, y)) V_{n,m}(r(x, y), \gamma(x, y)), \quad (3.5)$$

όπου οι $r(x, y), \gamma(x, y)$ αντιστοιχούν στο εικονοστοιχείο (x, y) και παράγονται με το μετασχηματισμό τους από Καρτεσιανές συντεταγμένες σε πολικές ως προς το (x_0, y_0) .

3.2.2 Βαθμονόμηση Ευρυγώνιας Κάμερας

Πριν προχωρήσουμε όμως στην ανάλυση της διόρθωσης της μεθόδου ροπών Zernike για εικόνες που παράγονται από ευρυγώνιες κάμερες, για λόγους πληρότητας θα παρουσιάσουμε το πρόβλημα που εμφανίζεται σε αυτού του τύπου κάμερες/φακούς. Αυτό δεν είναι άλλο από την αλλοίωση της πραγματικής απόστασης που υφίστανται δύο εικονοστοιχεία γειτονικών περιοχών πάνω στην εικόνα, καθώς και ότι όσο απομακρυνόμαστε από το κέντρο του φακού μιας ευρυγώνιας κάμερας η αλλοίωση αυτή γίνεται μεγαλύτερη.

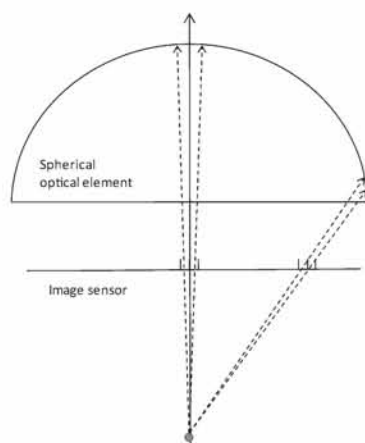
Η αλλοίωση αυτή εξηγείται λεπτομερώς στην εργασία [31] και απεικονίζεται επίσης στην Εικόνα 3.3 όπου χρησιμοποιείται μια αναπαράσταση του ευρυγώνιου φακού στις δύο διαστάσεις. Τα σημεία του πραγματικού κόσμου στο οπτικό στοιχείο απεικονίζονται στον αισθητήρα του φακού μέσω της διακεκομμένης γραμμής που συνδέει τον πραγματικό κόσμο (το σημείο στο οπτικό στοιχείο) με το κέντρο του οπτικού στοιχείου.

Στην Εικόνα 3.3, είναι ξεκάθαρο ότι γειτονικά εικονοστοιχεία που εμφανίζονται σε διαφορετικές θέσεις στον αισθητήρα εικόνας, η απόστασή τους είναι σαφώς διαφορετική όταν μετρείται στον αισθητήρα και στο σφαιρικό οπτικό στοιχείο.

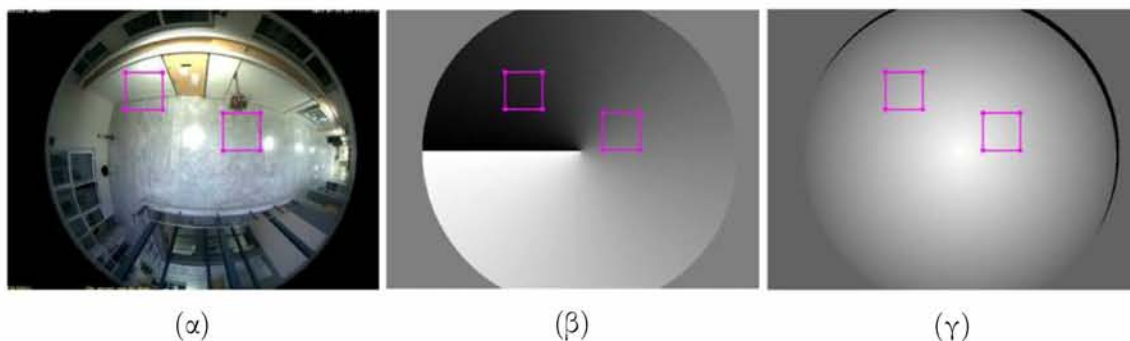
Ως εκ τούτου, προκειμένου να παραχθούν ακριβή αποτελέσματα, οι διάφοροι αλγόριθμοι επεξεργασίας εικόνας που βασίζονται στον υπολογισμό των αποστάσεων μεταξύ των εικονοστοιχείων, πρέπει να υπολογίσουν εκ νέου την απόσταση για τις ευρυγωνικές εικόνες. Έτσι, χρησιμοποιούμε για την ευθυγράμμιση της ευρυγώνιας κάμερας, την μέθοδο που προτείνεται στην εργασία [31], όπου υπολογίζει το διάνυσμα (θ, ϕ) (που αντιστοιχεί στο αζιμουθίο και την ανύψωση αντίστοιχα) στο σφαιρικό οπτικό στοιχείο για οποιοδήποτε εικονοστοιχείο (x, y) της εικόνας:

$$(\theta, \phi) = M(x, y). \quad (3.6)$$

Οι τιμές (θ, ϕ) της κάμερας αφού υπολογιστούν για κάθε εικονοστοιχείο του οπτικού της πεδίου αποθηκεύονται ώστε να μπορούν να χρησιμοποιηθούν για τον γρήγορο υπολογισμό της διόρθωσης των αποστάσεων (Εικόνα 3.4).



Σχήμα 3.3: Σχηματική αναπαράσταση της απεικόνισης μιας εικόνας σε σφαιρική ευρυγώνια κάμερα.



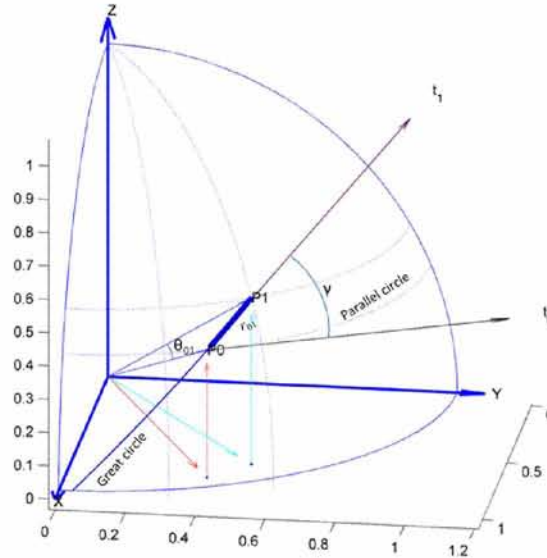
Σχήμα 3.4: Απεικόνιση των τιμών του αξιμούθιου (κέντρο) και της ανύψωσης (δεξιά) μετά από την ευθυγράμμιση της ευρυγώνιας κάμερας που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά).

3.2.3 Προσαρμογή των Ροπών Zernike σε Ευρυγώνια Κάμερα

Όπως αναφέραμε στην Παράγραφο 3.1 το πρόβλημα που καλούμαστε να αντιμετωπίσουμε είναι ο χαρακτηρισμός της ανθρώπινης στάσης μέσα από στατικές εικόνες ευρυγώνιας κάμερας. Σκοπός μας είναι να εξάγουμε κατάλληλα και αντιπροσωπευτικά χαρακτηριστικά από την στάση του ανθρώπου, έτσι ώστε να μπορέσουμε να εκπαιδεύσουμε ένα κατάλληλο ταξινομητή. Βασιζόμενοι στην μέθοδο εξαγωγής χαρακτηριστικών με ροπές Zernike, τροποποιημένες κατάλληλα στο συγκεκριμένο είδος κάμερας που βοηθάει στην επίλυση του προβλήματός μας.

Στην Εικόνα 3.1 παρουσιάζεται μια εικόνα από την ευρυγώνια κάμερα και την μεγάλη παραμόρφωση που προκαλείται σε αυτή όταν απομακρυνόμαστε από το κέντρο του οπτικού στοιχείου. Κάνοντας χρήση γνωστού αλγόριθμου τμηματοποίησης [31] στα στιγμιότυπα του βίντεο (από το οποίο αποτελείται η εικόνα), μπορούμε και εξάγουμε μια υπο-εικόνα συγκεκριμένου μεγέθους, που περιέχει την σιλουέτα του ανθρώπου σε δυαδική απεικόνιση. Για τον υπολογισμό των ροπών Zernike, σύμφωνα με την εξίσωση 3.5, απαιτείται η απόσταση ($r(x, y)$) και η γωνία ($\gamma(x, y)$) που ορίζεται από τα εφαπτομενικά διανύσματα t_0 του κύκλου

P_0 και το εφαπτομενικό διάνυσμα του κύκλου που ορίζεται από τον κύκλο P_1 (Εικόνα 3.5), για κάθε εικονοστοιχείο της τμηματοποιημένης εικόνας ως προς το τρέχων εικονοστοιχείο. Ας υποθέσουμε ότι θέλουμε να υπολογίσουμε τις ροπές Zernike για την εικόνα 3.1(β). Αν P_0 είναι το τρέχων εικονοστοιχείο, πρέπει να υπολογίσουμε την απόσταση και τη γωνία για κάθε εικονοστοιχείο P_1 της εικόνας.



Σχήμα 3.5: Αναπαράσταση των σημείων P_0 και P_1 στην μοναδιαία σφαίρα και ο υπολογισμός της γεωδαιτικής απόστασης και γωνίας τους.

Ωστόσο, καθώς η εικόνα από μια ευρυγώνια κάμερα απεικονίζεται ως ένα σφαιρικό στοιχείο, για τον υπολογισμό της απόστασης και της γωνίας θα χρησιμοποιήσουμε τις σφαιρικές συντεταγμένες. Ο προ-υπολογισμένος πίνακας αναφοράς που προέκυψε από την βαθμονόμηση της εικόνας (αξιμύθιο και ανύψωση - παρουσιάζονται στις Εικόνες 3.4-β και 3.4-γ) και αναφέρει σε κάθε εικονοστοιχείο τις τιμές τις αξιμυθιακής γωνίας και γωνίας ανύψωσης, χρησιμοποιείτε ώστε να υπολογίσουμε τις σφαιρικές συντεταγμένες (θ_0, ϕ_0) , (θ_1, ϕ_1) στο μοναδιαίο κύκλο των σημείων P_0 και P_1 , αντίστοιχα. Η γεωδαιτική απόσταση μεταξύ δύο σημείων σε μια σφαίρα είναι το μήκος του τόξου που ορίζουν τα σημεία αυτά και ανήκει στο κύκλο που διέρχεται από δύο σημεία με κέντρο το κέντρο της σφαίρας. Ο κύκλος αυτός ορίζεται ως ο “μεγάλος κύκλος”. Έστω v_0 και v_1 τα διανύσματα θέσης των P_0 και P_1 αντίστοιχα, από το κέντρο της σφαίρας όπως παρουσιάζονται στην Εικόνα 3.5. Η απόσταση r_{01} για τα σημεία P_0 και P_1 , στην μοναδιαία σφαίρα, υπολογίζεται ως το μήκος του τόξου των σημείων αυτών (θεωρώντας ότι το \cos^{-1} επιστρέφει τιμές σε ακτίνια) με τον εξής τύπο:

$$v_0 \cdot v_1 = (\cos\theta_0 \cos\phi_0, \sin\theta_0 \cos\phi_0, \sin\phi_0) \cdot (\cos\theta_1 \cos\phi_1, \sin\theta_1 \cos\phi_1, \sin\phi_1) = \cos\theta_0 \cos\phi_0 \cos\theta_1 \cos\phi_1 + \sin\theta_0 \cos\phi_0 \sin\theta_1 \cos\phi_1 + \sin\phi_0 \sin\phi_1, \quad (3.7)$$

$$r_{01} = \cos^{-1}(v_0 \cdot v_1). \quad (3.8)$$

Επιπλέον, εκτός από την απόσταση r , για τον υπολογισμό των ροπών Zernike απαιτείται και η γωνία γ για κάθε εικονοστοιχείο, σύμφωνα με την εξίσωση 3.5. Θεωρώντας ότι ο παράλληλος κύκλος που διέρχεται από το σημείο P_0 και ο μεγάλος κύκλος που διέρχεται από τα σημεία P_0 και P_1 και με τη γωνία γ να ορίζεται ως η γωνία μεταξύ των εφαπτομενικών διανυσμάτων t_0 και t_1 των δύο κύκλων στο σημείο P_0 (Εικόνα 3.5), το εφαπτομενικό διάνυσμα t_0 του παράλληλου κύκλου από το P_0 υπολογίζεται ως:

$$t_0 = \begin{cases} (-v_{0y}, v_{0x}, 0) / \sqrt{v_{0x}^2 + v_{0y}^2}, v_{0y} < 0 \\ (v_{0y}, -v_{0x}, 0) / \sqrt{v_{0x}^2 + v_{0y}^2}, v_{0y} \geq 0. \end{cases} \quad (3.9)$$

Έστω τώρα B ένα μοναδιαίο διάνυσμα κάθετο στο επίπεδο που ορίζεται από το v_0, v_1 . Το επίπεδο αυτό περιέχει το εφαπτομενικό διάνυσμα t_1 , ορίζοντας έτσι την επόμενη εξίσωση:

$$v_0 \times t_1 = v_0 \times v_1 = B = (B_x, B_y, B_z), \quad (3.10)$$

όπου το διάνυσμα t_1 είναι επιπλέον και κάθετο με το διάνυσμα v_0 :

$$v_0 \cdot t_1 = 0. \quad (3.11)$$

Συνδυάζοντας τις 3.10, 3.11 παράγεται η παρακάτω μορφή:

$$\begin{bmatrix} 0 & -u_{0z} & u_{0y} & 1 \\ u_{0z} & 0 & -u_{0x} & 1 \\ -u_{0y} & u_{0x} & 0 & 1 \\ u_{0x} & -u_{0y} & u_{0z} & 0 \end{bmatrix} \begin{bmatrix} t_{1x} \\ t_{1y} \\ t_{1z} \\ 0 \end{bmatrix} = \begin{bmatrix} B_x \\ B_y \\ B_z \\ 0 \end{bmatrix}. \quad (3.12)$$

Η παραπάνω εξίσωση μπορεί να γενικευτεί για N -σημεία / διαστάσεις από αυτά που παράγονται από το συγκεκριμένο πρόβλημα. Έτσι έχουμε την γενικευμένη εξίσωση:

$$\begin{bmatrix} 0 & -u_{0z} & u_{0y} & 1 \\ u_{0z} & 0 & -u_{0x} & 1 \\ -u_{0y} & u_{0x} & 0 & 1 \\ u_{0x} & -u_{0y} & u_{0z} & 0 \end{bmatrix} \begin{bmatrix} t_{1x}^1 & t_{1x}^2 & \dots & t_{1x}^N \\ t_{1y}^1 & t_{1y}^2 & \dots & t_{1y}^N \\ t_{1z}^1 & t_{1z}^2 & \dots & t_{1z}^N \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} B_x^1 & B_x^2 & \dots & B_x^N \\ B_y^1 & B_y^2 & \dots & B_y^N \\ B_z^1 & B_z^2 & \dots & B_z^N \\ 0 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{4 \times N}, \quad (3.13)$$

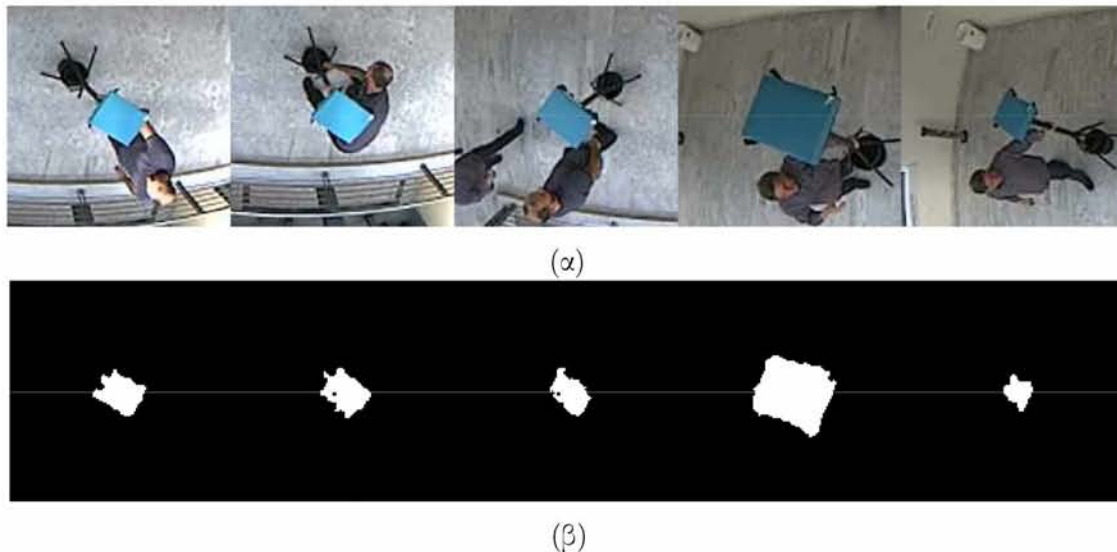
όπου οι υπερδείκτες υποδηλώνουν το δείκτη της διάστασης του σημείου. Η γωνία γ υπολογίζεται ως το εσωτερικό γινόμενο μεταξύ των διανυσμάτων t_0 και t_1 , ενώ το πρόσημό του ορίζεται από το πρόσημο του εξωτερικού γινομένου $B_z = (t_0 \times t_1)_z$ στον Z -άξονα. Έτσι η γωνία γ ορίζεται ως εξής:

$$\gamma = \text{sign}((t_0 \times t_1)_z) \cdot \cos^{-1} \left(\frac{t_0 \cdot t_1}{|t_0||t_1|} \right). \quad (3.14)$$

Η απόσταση και η γωνία (r, γ) της εξίσωσης 3.3 αντικαθίσταται αντίστοιχα με την απόσταση της εξίσωσης 3.8 και τη γωνία της εξίσωσης 3.14. Η νέα απόσταση ορίζεται ως ‘γεωδαιτική’ και οι ροπές Zernike ορίζονται ως ‘γεωδαιτικές ροπές Zernike’, ενώ η μέθοδος εξαγωγής χαρακτηριστικών που βασίζεται σε αυτή την διόρθωση ονομάζεται ‘Γεωδαιτική Διόρθωση των Ροπών Zernike’ (ΓΔΡΖ).

3.2.4 Μελέτη της Γεωδαιτικής Διόρθωσης των Ροπών Zernike

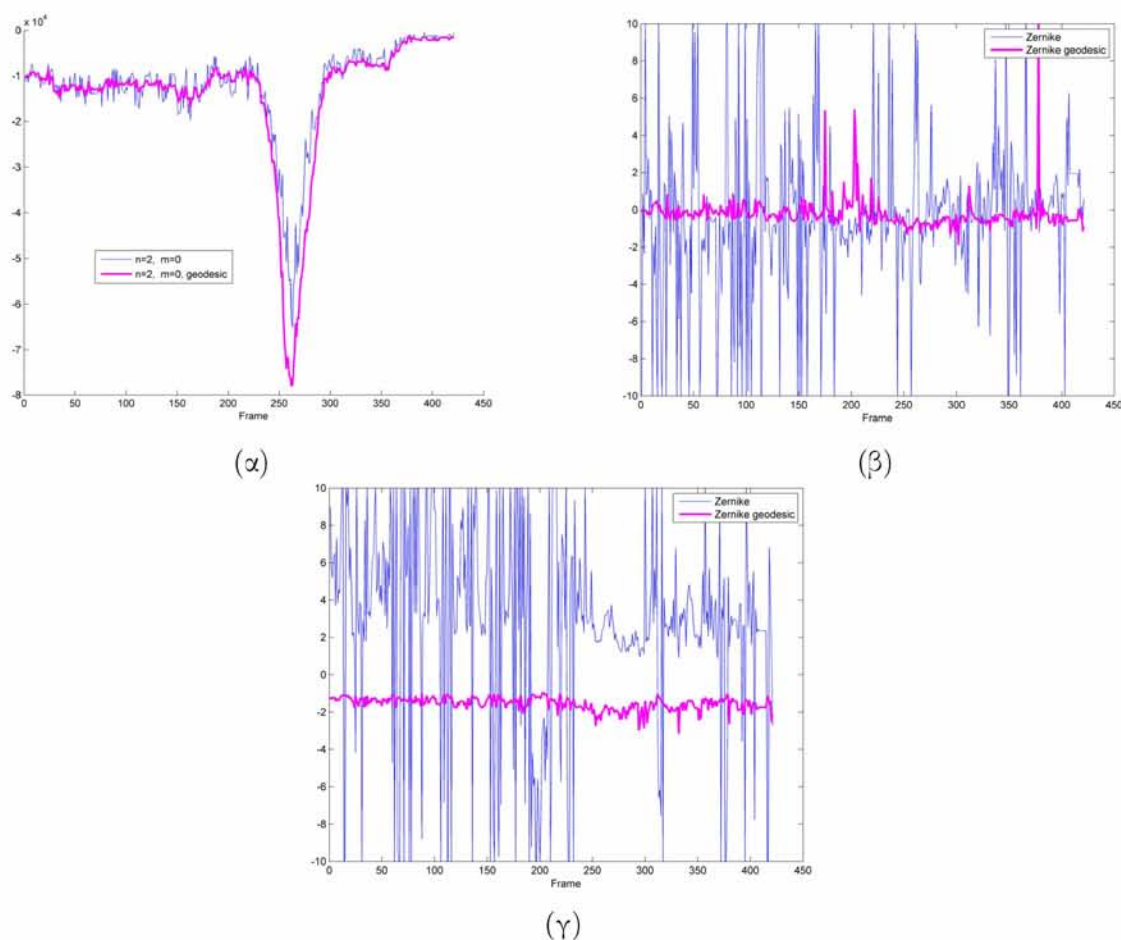
Στη συνέχεια, βασιζόμενοι στην εργασία [29] μελετήσαμε πειραματικά τις προτεινόμενες γεωδαιτικές ροπές Zernike σε σύγκριση με την ροπές Zernike χωρίς την διόρθωση. Για τον σκοπό χρησιμοποιήσαμε ένα αντικείμενο παραλληλόγραμμου σχήματος με σταθερές και γνωστές τιμές στο χρωματικό χώρο RGB, το οποίο το καταγράψαμε με την ευρυγώνια κάμερα σε διαφορετικές θέσεις και αποστάσεις (Εικόνα 3.6-α). Με τον τρόπο αυτό παρήχθησαν 450 στιγμιότυπα και στη συνέχεια απομονώσαμε το αντικείμενο αυτό (ως μια δυαδική εικόνα), από την υπόλοιπη σκηνή που κατέγραψε η κάμερα, θέτοντας ένα κατώφλι στις τιμές του RGB χώρου, διατηρώντας έτσι μόνο το αντικείμενο (Εικόνα 3.6-β).



Σχήμα 3.6: Στιγμιότυπα της ευρυγώνιας κάμερας που απεικονίζουν ένα αντικείμενο σε διαφορετικές θέσεις, στροφές και απόσταση από την κάμερα (α). Τμηματοποίηση των αντίστοιχων στιγμιότυπων, αφαιρώντας το παρασκήνιο διατηρώντας το σχήμα του αντικειμένου (β).

Όπως μπορούμε να διακρίνουμε στην Εικόνα 3.6-β, το αντικείμενο μετά τον διαχωρισμό του από την υπόλοιπη σκηνή έχει υποστεί αλλοιώσεις χαλώντας την συνεχή δομή παραλληλόγραμμου που θα έπρεπε να έχει σύμφωνα με την έγχρωμη εικόνα του. Σκοπός μας είναι πειραματικά, να ελέγξουμε την ικανότητα των διορθωμένων ροπών Zernike όπως αναφέρεται στην εργασία [29], κατά πόσο διατηρούν σταθερότερες τιμές, καθώς και το χαρακτηριστικό να παραμένουν αναλλοίωτες ως προς τις στροφές που μπορεί να κάνει το αντικείμενο αυτό. Οι τιμές των ροπών σε πραγματικές συνθήκες όπου η δομή του αντικειμένου παραμένει σταθερή, μεταξύ των διαδοχικών στιγμιότυπων του ίδιου αντικειμένου, παραμένουν πανομοιότυπες καθώς ο ροπές είναι ένας αναλλοίωτος περιγραφέας ωστόσο εδώ έχουμε το σφάλμα που εμφανίζεται στο αντικείμενο λόγω της εφαρμογής του αλγορίθμου τμηματοποίησης.

Για τον λόγο αυτό επιλέγουμε να συγκρίνουμε τις πρώτες τρεις κεντρικές ροπές Zernike ($Z_{2,0}$, $Z_{4,0}$, $Z_{6,0}$, για τιμές $n = 2, 4, 6$ και $m = 0$) με και χωρίς διόρθωση στο αντικείμενο της δυαδικής εικόνας. Όπως φαίνεται στην Εικόνα 3.7-α, οι τιμές των πρώτων διορθωμένων γεω-



Σχήμα 3.7: Η τιμή της πρώτης κεντρικής διαφοράς στα 450 στιγμιότυπα του σχήματος του αντικειμένου των ΓΔΡΖ έναντι των κλασικών ροπών Zernike (α). Ο λόγος των διαφορών: των τρίτων προς τις δεύτερες (β) και των δεύτερων προς τις πρώτες (γ), των ΓΔΡΖ έναντι των κλασικών ροπών Zernike.

δαιτικών ροπών Zernike, κεντρικής τάξης, είναι εμφανώς σταθερότερες μεταξύ των διαδοχικών στιγμιότυπων σε σύγκριση με της κλασικές ροπές Zernike, καθώς οι τιμές των διορθωμένων γεωδαιτικών ροπών μεταξύ των διαδοχικών στιγμιότυπων είναι κοντινότερες και δεν παρουσιάζουν απότομες αλλαγές όπως οι κλασικές ροπές. Η μεγάλη τάση προς τις αρνητικές τιμές είναι λόγω της δραστηρικής αλλαγής του σχήματος του αντικειμένου στα συγκεκριμένα στιγμιότυπα. Μπορούμε να παρατηρήσουμε ότι οι τιμές των διορθωμένων ροπών Zernike συνεχίζουν να είναι σταθερές σε ομαλές μεταβολές της θέσης του αντικειμένου. Υπολογίζοντας τον λόγο των γειτονικών κεντρικών ροπών Zernike $Z_{4,0}/Z_{2,0}$ και $Z_{6,0}/Z_{4,0}$ όπως εμφανίζονται αντίστοιχα στις Εικόνες 3.7-β και 3.7-γ, βλέπουμε ότι ο λόγος των διορθωμένων ροπών Zernike παρουσιάζει μια ανοχή στις ξαφνικές αλλαγές του σχήματος του αντικειμένου σε σχέση με τις μη διορθωμένες ροπές Zernike.

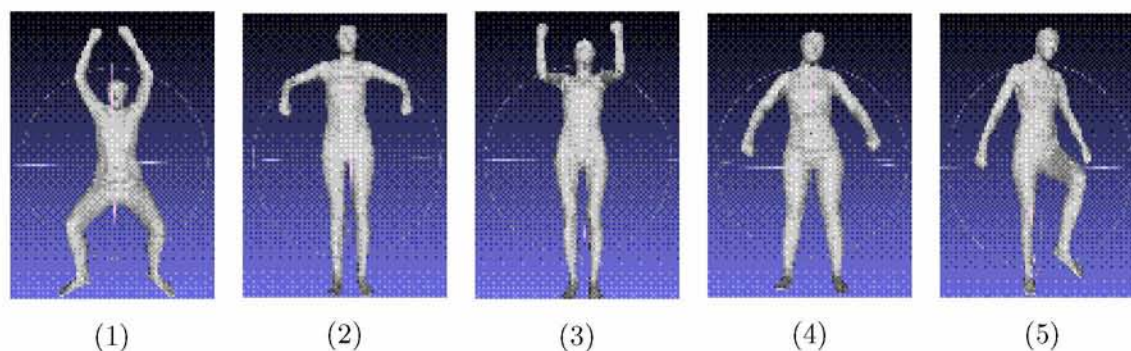
3.3 Πειραματική Θεμελίωση των Διορθωμένων Ροπών Zernike σε Συνθετικά και Πραγματικά Δεδομένα

Στην ενότητα αυτή θα εξετάσουμε τις διορθωμένες γεωδαιτικές ροπές Zernike έναντι των κλασικών ροπών Zernike για την αναγνώριση της ανθρώπινης στάσης μέσω της σιλουέτας του ανθρώπου. Οι εικόνες αυτές λαμβάνονται από μια ευρυγώνια κάμερα η οποία βρίσκεται τοποθετημένη στην οροφή ενός δωματίου. Η θέση αυτή και κατ' επέκταση ο συγκεκριμένος τύπος κάμερας επιλέχθηκε ώστε να αποφύγουμε παρεμβολές από αντικείμενα που μπορεί να υπάρχουν μέσα στο χώρο που κινείται ο άνθρωπος, καθώς μια τυπικού φακού κάμερα δεν θα μπορούσε να απεικονίζει πάντα ολόκληρη την ανθρώπινη σιλουέτα.

Χρησιμοποιήσαμε για τα πειράματά μας συνθετικές σιλουέτες ανθρώπινης στάσης από τρισδιάστατα μοντέλα ανθρώπου, τοποθετημένα στο χώρο ενός πραγματικού δωματίου στο οποίο υπάρχει μια ευρυγώνια κάμερα στην οροφή του, αλλά και πραγματικά δεδομένα με τις αντίστοιχες ανθρώπινες στάσεις στο δωμάτιο αυτό. Ο λόγος που χρησιμοποιήσαμε τα συνθετικά δεδομένα είναι διττός. Αρχικά για να ελέγξουμε τα δεδομένα σε σιλουέτες οι οποίες μπορούν να παραχθούν με ακρίβεια (καθώς είναι συνθετικά τρισδιάστατα μοντέλα), καθώς η δημιουργία ανθρώπινης σιλουέτας από πραγματικές εικόνες εμπεριέχει σφάλμα λόγω του αλγορίθμου τμηματοποίησης για τον εντοπισμού και διαχωρισμού του ανθρώπου από την υπόλοιπη σκηνή της εικόνας. Δεύτερον, για την εκπαίδευση ένας απλού μοντέλου ταξινόμησης χρειαζόμαστε ένα επαρκές σύνολο δεδομένων εκπαίδευσης και ελέγχου.

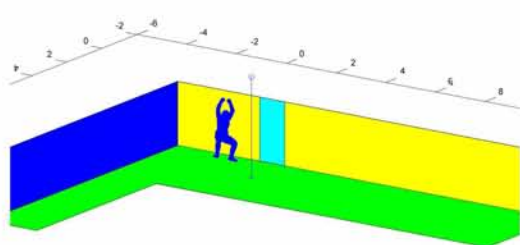
3.3.1 Συνθετικές Σιλουέτες

Για την δημιουργία των συνθετικών σιλουετών ανθρώπινης στάσης, θα χρησιμοποιήσουμε την προσομοίωση της ευρυγώνιας κάμερας που έχει δημιουργηθεί μέσω του μοντέλου ευθυγράμμισης της [31] όπως παρουσιάστηκε στην υπο-ενότητα 3.2.2, προσομοιώνοντας έτσι τον χώρο του δωματίου. Στην προσομοίωση αυτή θα τοποθετηθούν τα τρισδιάστατα μοντέλα ανθρώπων [9,14] που παρουσιάζονται στην Εικόνα 3.8 σε διαφορετικές θέσεις. Κάθε στάση αναπαρίσταται από ένα διαφορετικό πλήθος τρισδιάστατων μοντέλων/ανθρώπων. Συγκεκριμένα έχουμε 19 διαφορετικά μοντέλα για την στάση 1, 13 για την στάση 3, 18 για την στάση 3, 16 για την στάση 4 και 16 για την στάση 5.



Σχήμα 3.8: Τα συνθετικά μοντέλα των πέντε όρθιων ανθρώπινων στάσεων.

Τα συγκεκριμένα πέντε μοντέλα επιλέχθηκαν λόγω της φυσικότητας των συγκεκριμένων στάσεων που απεικονίζουν. Επιπλέον, οι στάσεις αυτές δεν εμφανίζουν μεταξύ τους πολύ μεγάλες διαφοροποιήσεις σε σχέση με τις υπόλοιπες διαθέσιμες στάσεις του συνόλου δεδομένων [57]. Αυτό αυξάνει τον βαθμό δυσκολίας του προβλήματος εντοπισμού ανθρώπινης στάσης, συνυπολογίζοντας το γεγονός ότι η ευρυγώνια κάμερα (με οπτικό πεδίο 180°) λόγω της απόστασής της από το πάτωμα και τις διαφορετικές θέσεις που μπορεί να βρίσκεται το αντικείμενο στο χώρο (ιδιαίτερα όσο πλησιάζει το κέντρο του οπτικού πεδίου), η σιλουέτα του ανθρώπου αρχίζει και γίνεται δυσδιάκριτη, όπως φαίνεται στην Εικόνα 3.9-(γ).



(α)



(β)



(γ)



(δ)

Σχήμα 3.9: Γραφική απεικόνιση των συνθετικών μοντέλων στο πραγματικό δωμάτιο της ευρυγώνιας κάμερας (α) και (β). Τοποθέτηση των συνθετικών μοντέλων στον πραγματικό χώρο με απόσταση 0,5 μέτρα και σταθερό προσανατολισμό για την δημιουργία του συνόλου εκπαίδευσης (γ). Υπο-εικόνα μιας σιλουέτας των συνθετικής μοντέλων για την εφαρμογή των μεθόδων εξαγωγής ολικών χαρακτηριστικών.

Για την δημιουργία του συνόλου δεδομένων κάθε τριδιάστατο μοντέλο από τα πέντε προαναφερθέντα, τοποθετήθηκε σε διαφορετικές θέσεις του πραγματικού δωματίου και με διάφορους προσανατολισμούς (Εικόνα 3.9-(γ)). Ένα παράδειγμα σιλουέτας παρουσιάζεται στην Εικόνα 3.9-(δ), ενώ στην Εικόνα 3.9-(α) βλέπουμε μια γραφική αναπαράσταση της ανθρώπινης στάσης 3 στο χώρο. Για την παραγωγή όλων των σιλουέτων μέσω του προσομοιωτή, χρησιμοποιήσαμε ένα πλέγμα 13×8 διαφορετικών θέσεων με απόσταση 0,5 μέτρο μεταξύ

τους, εξαιρώντας τις θέσεις αυτές που βρίσκονται στο κυκλικό χώρο που σχηματίζεται από το κέντρο του οπτικού στοιχείου με ακτίνα 1 μέτρο, καθώς σε αυτό το χώρο η ανθρώπινη σιλουέτα είναι ακριβώς ίδια σε όλες τις στάσεις. Σε κάθε θέση που τοποθετήθηκε το μοντέλο στο χώρο παρήχθη μια σιλουέτα σε διαφορετικό προσανατολισμό, κάνοντας στροφή κάθε $\pi/5$ rad στον κάθετο άξονά του (Z-άξονα). Έτσι, για κάθε μοντέλο/άνθρωπο κάθε διαφορετικής στάσης παράγονται 920 σιλουέτες $((13 \times 8 - 12) \times 10)$. Το πλήθος του συνόλου δεδομένων για κάθε στάση υπολογίζεται ως το γινόμενο του πλήθους των διαφορετικών μοντέλων/ανθρώπων για κάθε στάση πολλαπλασιασμένο με το 920. Έτσι, για τις πέντε στάσεις έχουμε ένα σύνολο 75440 σιλουέτες που παρήχθησαν από τα 82 διαφορετικά μοντέλα/ανθρώπους. Τέλος, σε κάθε εικόνα σιλουέτας εφαρμόζουμε τις διορθωμένες και μη ροπές Zernike τάξης $n = 9$ εξάγοντας 58 χαρακτηριστικά (29 μιγαδικά χαρακτηριστικά).

Στάση	Πλήθος Εικόνων	Χαρακτηριστικά
1	17480	58
2	11960	58
3	16560	58
4	14720	58
5	14720	58
Σύνολο	75440	

Πίνακας 3.1: Το σύνολο δεδομένων για τις πέντε συνθετικές σιλουέτες όρθιας ανθρώπινης στάσης.

Το σύνολο σιλουέτων ανθρώπινης στάσης που περιγράψαμε, παρουσιάζεται στον Πίνακα 3.1 και αποτελείται από το πλήθος των σιλουετών που παρήχθησαν για κάθε διαφορετική στάση. Οι στάσεις αυτές όπως είδαμε στην Εικόνα 3.8 είναι όρθιες στάσεις του ανθρώπου. Οι ίδιες στάσεις μπορούν τεχνητά να μετασχηματιστούν σε στάση πτώσης του ανθρώπου στο έδαφος. Αυτό επιτυγχάνεται με την αντικατάσταση των συντεταγμένων του Z-άξονα με του Y-άξονα και του Y-άξονα με του X-άξονα, αντίστοιχα. Έτσι, δημιουργείται ένα ακόμα σύνολο δεδομένων δύο κλάσεων πλήθους 2×75440 όπως παρουσιάζεται στον Πίνακα 3.2.

Στάση	Πλήθος Εικόνων	Χαρακτηριστικά
Όρθιος	75440	58
Πεσμένος	150880	58
Σύνολο	226320	

Πίνακας 3.2: Το σύνολο δεδομένων συνθετικών σιλουετών των δύο γενικευμένων ανθρώπινων στάσεων.

3.3.2 Πραγματικές Σιλουέτες

Στη συνέχεια, για να ελέγξουμε τις διορθωμένες ροπές Zernike και την προτεινόμενη μεθοδολογία κατ' επέκταση σε ρεαλιστικές συνθήκες, δημιουργήσαμε 4 μικρά βίντεο στα οποία ένας άνθρωπος προσομοιώνει τις πέντε όρθιες στάσεις που έχουν περιγραφεί νωρίτερα καθώς και τη στάση του πεσμένου ανθρώπου στο πάτωμα. Τα βίντεο παρήχθησαν σε εσωτερικό χώρο (δωμάτιο) χρησιμοποιώντας μια ευρυγώνια κάμερα τοποθετημένη στην οροφή του δωματίου. Τα στιγμιότυπα των βίντεο χαρακτηρίστηκαν ως προς την στάση που απεικονίζουν, έτσι ώστε να έχουμε ένα ολοκληρωμένο σύνολο δεδομένων για τον έλεγχο της μεθοδολογίας μας.

Οι ροπές Zernike (διορθωμένες και μη) εφαρμόζονται σε δυαδικές εικόνες. Για τον λόγο αυτό εξάγουμε την σιλουέτα του ανθρώπου στις διάφορες στάσεις που εμφανίζεται στο βίντεο, χρησιμοποιώντας τον αλγόριθμο [31], ο οποίος επιτυγχάνει την εξαγωγή της σιλουέτας με σχετικά μικρό βαθμό σφάλματος. Στην Εικόνα 3.10 παρουσιάζεται ένα δείγμα από ένα στιγμιότυπο όπως καταγράφεται από την ευρυγώνια κάμερα (αριστερά) και η εξαγωγή της σιλουέτας του ανθρώπου (δεξιά).



Σχήμα 3.10: Λήψη της στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας (δεξιά).

3.4 Πειραματικά Αποτελέσματα

Στην ενότητα αυτή θα προχωρήσουμε στην πειραματική μελέτη των διορθωμένων ροπών Zernike έναντι των κλασικών ροπών Zernike για την εύρεση της ανθρώπινης στάσης από ευρυγώνια κάμερα κάνοντας χρήση μόνο της σιλουέτας. Για τον υπολογισμό των χαρακτηριστικών KPZ των εικόνων γίνεται χρήση του πολυωνύμου V_{nm} με τον κλασικό ορισμό της απόστασης και γωνίας, ενώ για τον υπολογισμό των χαρακτηριστικών ΔΓΡΖ η απόσταση και η γωνία ορίζεται σύμφωνα με τις Εξισώσεις 3.8 και 3.14, αντίστοιχα. Συγκεκριμένα, υπολογίζονται όλες οι ροπές που παράγονται για τις τιμές $n = 1, 2, \dots, 9$ και όλες οι πιθανές τιμές του $m \geq 0$ για κάθε ένα από τα n , όπως παρουσιάζονται στον Πίνακα 3.3. Έτσι, ένα σύνολο 58 χαρακτηριστικών εξάγεται για κάθε εικόνα, δηλαδή από το πραγματικό και φανταστικό μέρος των 29

μιγαδικών ροπών που παράγονται από το πολυώνυμο V_{nm} . Οι διορθωμένες γεωδαιτικές ροπές Zernike, αλλά και οι κλασικές ροπές Zernike, εφαρμόζονται σε υποεικόνες που περιέχουν ανθρωπίνες σιλουέτες διάστασης 64×64 .

Στη συνέχεια, εκπαιδεύσαμε έναν ταξινομητή χρησιμοποιώντας το 40% του συνόλου των συνθετικών σιλουετών, ενώ οι υπόλοιπες (60%) χρησιμοποιήθηκαν για έλεγχο του μοντέλου ως προς την ακρίβεια αναγνώρισης της στάσης. Ο ταξινομητής k -Κοντινότερος Γείτονας (k -ΚΓ) με τιμή $k = 3$ επιλέχθη λόγω της απλότητάς του, καθώς θελήσαμε να ελέγξουμε την ικανότητα των προτεινόμενων χαρακτηριστικών για την επίλυση του προβλήματος και όχι τον συνδυασμό τους με σύνθετους ταξινομητές που χρησιμοποιούνται για την επίλυση γενικότερων δύσκολων προβλημάτων.

n	Πλήθος	Αθροιστικό	Πλήθος	m									
	Μιγαδικών										χαρακτηριστικών		
	KPZ	με $m \geq 0$											
1	2	1		-1	1								
2	3	3		-2	0	2							
3	4	5		-3	-1	-1	3						
4	5	8		-4	-2	0	2	4					
5	6	11		-5	-3	-1	1	3	5				
6	7	15		-6	-4	-2	0	2	4	6			
7	8	19		-7	-5	-3	-1	1	3	5	7		
8	9	24		-8	-6	-4	-2	0	2	4	6	8	
9	10	29		-9	-7	-5	-3	-1	1	3	5	7	9

Πίνακας 3.3: Οι τιμές των παραμέτρων n και m της εξίσωσης 3.2 με το αντίστοιχο πλήθος χαρακτηριστικών. Το αθροιστικό πλήθος χαρακτηριστικών είναι το συνολικό πλήθος χαρακτηριστικών (πραγματικών και μιγαδικών) για όλους τους έγκυρους συνδυασμούς των παραμέτρων (n, m) για το τρέχον n κάθε φορά εξαιρώντας τα μιγαδικά συζυγή.

Στον Πίνακα 3.4 παρουσιάζονται τα αποτελέσματα της ακρίβειας του εκπαιδευμένου μοντέλου k -ΚΓ, κάνοντας χρήση των χαρακτηριστικών από τις διορθωμένες γεωδαιτικές ροπές Zernike έναντι των κλασικών. Καθώς οι ροπές Zernike παράγουν μιγαδικά χαρακτηριστικά, για λόγους πληρότητας έγινε έλεγχος στην αποτελεσματικότητα του μέτρου και της φάσης του κάθε μιγαδικού αριθμού, αλλά και τον συνδυασμό τους ως χαρακτηριστικά για την εκπαίδευση του ταξινομητή. Φαίνεται στον Πίνακα 3.4 ότι η χρήση των τιμών του πραγματικού και φανταστικού μέρους των μιγαδικών αριθμών για την εκπαίδευση του μοντέλου εμφανίζει μεγαλύτερη ακρίβεια στα δεδομένα ελέγχου, ενώ επιπλέον παρατηρείται μεγαλύτερη ακρίβεια των διορθωμένων ροπών Zernike έναντι των κλασικών με μια βελτίωση της τάξης του 4%. Στην συνέχεια των πειραματικών ελέγχων, θα χρησιμοποιούμε ως χαρακτηριστικά των ροπών Zernike τις τιμές από το πραγματικό και φανταστικό μέρος των μιγαδικών χαρακτηριστικών.

Χαρακτηριστικά	KPZ	ΓΔPZ
Πραγματικό και Φανταστικό μέρος	91, 15%	95, 31%
Μέτρο	68, 76%	60, 99%
Φάση	57, 08%	63, 56%
Μέτρο και Φάση	73, 38%	70, 62%

Πίνακας 3.4: Το ποσοστό επιτυχίας των KPZ και ΓΔPZ στις συνθετικές σιλουέτες χρησιμοποιώντας διαφορετικά μέτρα των παραγόμενων χαρακτηριστικών τους.

3.4.1 Ταξινόμηση των Πέντε Όρθιων Ανθρώπινων Στάσεων

Συνεχίζοντας την μελέτη των πειραματικών αποτελεσμάτων και εξετάζοντας ενδελεχώς την ακρίβεια του μοντέλου στις συνθετικές σιλουέτες όρθιας στάσης, παρουσιάζονται τα μητρώα σύγχυσης στους Πίνακες 3.5 και 3.6, των ροπών Zernike με τις τιμές από το πραγματικό και φανταστικό μέρος τους, για τις κλασικές και γεωδαιτικές ροπές Zernike, αντίστοιχα. Οι στήλες του μητρώου παρουσιάζουν την πρόβλεψη του ταξινομητή για την κάθε στάση (κλάση), ενώ οι γραμμές αντιστοιχούν στις πραγματικές στάσεις.

		Πρόβλεψη				
	Στάση	1	2	3	4	5
Πραγματική	1	9424	189	496	246	135
Τιμή	2	155	6520	117	279	105
	3	141	156	9315	196	128
	4	139	273	128	8092	200
	5	126	214	175	410	7907

Πίνακας 3.5: Το μητρώο σύγχυσης του KPZ στις συνθετικές σιλουέτες όρθιας στάσης.

		Πρόβλεψη				
	Στάση	1	2	3	4	5
Πραγματική	1	12057	116	495	187	39
Τιμή	2	75	8484	28	228	55
	3	57	75	11928	124	57
	4	55	220	42	10418	75
	5	39	148	101	389	12057

Πίνακας 3.6: Το μητρώο σύγχυσης του ΓΔPZ στις συνθετικές σιλουέτες όρθιας στάσης.

Στην Εικόνα 3.11 βλέπουμε την απόδοση του ταξινομητή στο συνθετικό σύνολο ελέγχου και τη μεταβολή του σφάλματός του καθώς οι σιλουέτες απομακρύνονται από το κέντρο του

		Πρόβλεψη				
	Στάση	1	2	3	4	5
Πραγματική	1	0	180	0	396	144
Τιμή	2	0	720	0	0	0
	3	0	270	0	0	0
	4	0	0	54	486	0
	5	0	0	0	207	0

Πίνακας 3.7: Το μητρώο σύγχυσης του KPZ στις πραγματικές ανθρώπινες σιλουέτες όρθιας στάσης.

		Πρόβλεψη				
	Στάση	1	2	3	4	5
Πραγματική	1	495	54	81	90	0
Τιμή	2	0	585	90	18	27
	3	0	0	171	63	36
	4	0	27	0	405	108
	5	0	0	0	9	198

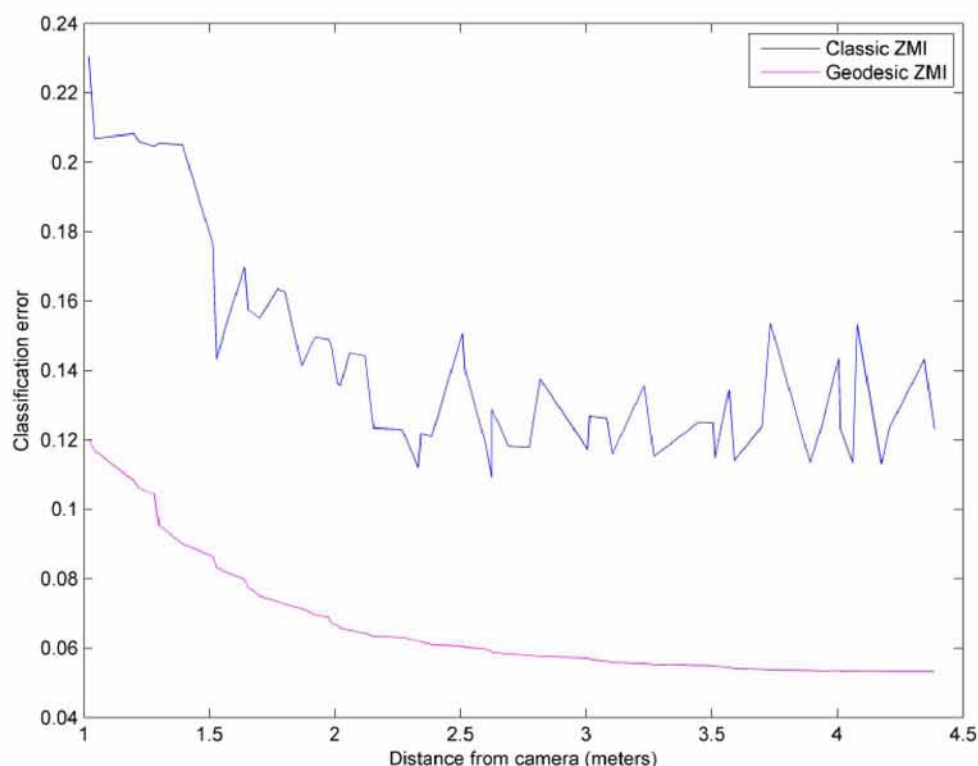
Πίνακας 3.8: Το μητρώο σύγχυσης του ΓΔΡΖ στις πραγματικές ανθρώπινες σιλουέτες όρθιας στάσης.

Τύπος Δεδομένων	Μέθοδος	Ακρίβεια
Συνθετικά	KPZ	91, 14%
Συνδετικά	HoG	77, 26%
Συνδετικά	ΓΔΡΖ	95, 31%
Πραγματικά	KPZ	49, 08%
Πραγματικά	HoG	50, 15%
Πραγματικά	ΓΔΡΖ	75, 45%

Πίνακας 3.9: Τα ποσοστά επιτυχίας των τριών μεθόδων στις συνθετικές και πραγματικές σιλουέτες ανθρώπινων όρθιων στάσεων.

οπτικού στοιχείου της κάμερας. Όπως μπορούμε να παρατηρήσουμε το σφάλμα των γεωδαιτικών ροπών Zernike είναι χαμηλότερο από το αντίστοιχο των κλασικών σε κάθε μονάδα μέτρου. Επιπλέον, το σφάλμα μειώνεται σταθερά και εμφανίζει μικρότερες διακυμάνσεις.

Στη συνέχεια, ελέγχουμε την απόδοση των εκπαιδευμένων μοντέλων σε δεδομένα από συνθετικές σιλουέτες με πραγματικές όρθιες στάσεις από τα στιγμιότυπα των τεσσάρων βίντεο που περιγράψαμε νωρίτερα. Ο λόγος που χρησιμοποιούμε τα υπάρχοντα εκπαιδευμένα μοντέλα είναι το μικρό πλήθος πραγματικών εικόνων που έχουμε στη διάθεσή μας για τις πέντε όρθιες στάσεις, αφού πρέπει να εκπαιδεύσουμε ένα μοντέλο εξ' ολοκλήρου με αυτές. Ο χαρακτηρισμός της στάσης του ανθρώπου που απεικονίζεται στο κάθε στιγμιότυπο δεν μπορεί να γίνει με



Σχήμα 3.11: Τα ποσοστά επιτυχίας των ΓΔΡΖ και ΚΡΖ στις συνθετικές σιλουέτες, καθώς απομακρύνονται από το κέντρο της ευρυγώνιας κάμερας.

αυτόματο τρόπο, έτσι δεν μπορούμε να έχουμε αρκούντως μεγάλο σύνολο εκπαίδευσης και ελέγχου.

Στους Πίνακες 3.7 και 3.8 παρουσιάζονται τα μητρώα σύγχυσης από τις πραγματικές σιλουέτες των τεσσάρων βίντεο ως προς τη στάση που απεικονίζουν, χρησιμοποιώντας τις κλασικές και διορθωμένες ροπές Zernike, αντίστοιχα. Στον Πίνακα 3.9 παρουσιάζεται μια συνοπτική απεικόνιση της απόδοσης των ροπών Zernike με και χωρίς διόρθωση, καθώς και της μεθόδου εξαγωγής χαρακτηριστικών HoG, τόσο για το σύνολο ελέγχου των συνθετικών σιλουετών όσο και των πραγματικών. Αυτό που μπορούμε να παρατηρήσουμε είναι ότι το εκπαιδευμένο μοντέλο με τις διορθωμένες ροπές Zernike στις συνθετικές εικόνες, μπορεί να ταξινομήσει με πολύ μεγαλύτερη ακρίβεια σιλουέτες πραγματικών ανθρώπων σε μια από τις πέντε όρθιες στάσεις σε σχέση με τις άλλες μεθοδολογίες. Επιπλέον, ενώ η μέθοδος HoG στις συνθετικές σιλουέτες έχει χαμηλότερα απόδοση έναντι των ΚΡΖ, στις πραγματικές σιλουέτες εμφανίζουν να έχουν την ίδια απόδοση.

3.4.2 Ταξινόμηση Δύο Γενικευμένων Ανθρώπινων Στάσεων

Κλείνοντας την πειραματική μελέτη των ΓΔΡΖ θα εξετάσουμε την απόδοσή τους σε δύο γενικευμένες στάσεις, την όρθια στάση και την στάση σε κατάσταση πτώσης. Στους Πίνα-

Πρόβλεψη			
	Στάση	Όρθια	Πτώση
Πραγματική	Όρθια	73599	836
Τιμή	Πτώση	385	60972

Πίνακας 3.10: Το μητρώο σύγχυσης του KPZ στις συνθετικές σιλουέτες των δύο γενικευμένων στάσεων.

Πρόβλεψη			
	Στάση	Όρθια	Πτώση
Πραγματική	Όρθια	73770	665
Τιμή	Πτώση	254	61103

Πίνακας 3.11: Το μητρώο σύγχυσης του ΓΔΡΖ στις συνθετικές σιλουέτες των δύο γενικευμένων στάσεων.

Πρόβλεψη			
	Στάση	Όρθια	Πτώση
Πραγματική	Όρθια	209	105
Τιμή	Πτώση	204	489

Πίνακας 3.12: Το μητρώο σύγχυσης του KPZ στις πραγματικές ανθρώπινες σιλουέτες γενικευμένων στάσεων.

Πρόβλεψη			
	Στάση	Όρθια	Πτώση
Πραγματική	Όρθια	278	36
Τιμή	Πτώση	148	545

Πίνακας 3.13: Το μητρώο σύγχυσης του KPZ στις πραγματικές ανθρώπινες σιλουέτες γενικευμένων στάσεων.

Τύπος Δεοδομένων	Μέθοδος	Ακρίβεια
Συνθετικά	KPZ	99, 10%
Συνθετικά	HoG	77, 26%
Συνθετικά	ΓΔΡΖ	99, 32%
Πραγματικά	KPZ	69, 31%
Πραγματικά	HoG	50, 15%
Πραγματικά	ΓΔΡΖ	81, 72%

Πίνακας 3.14: Τα ποσοστά επιτυχίας των τριών μεθόδων στις συνθετικές και πραγματικές σιλουέτες ανθρώπινων γενικευμένων στάσεων.

κες 3.10 και 3.11 παρουσιάζονται τα μητρώα σύγχυσης για τις συνθετικές σιλουέτες των δύο γενικευμένων στάσεων. Παρατηρούμε ότι οι ΓΔΡΖ χαρακτηρίζουν σωστά μεγαλύτερο πλήθος συνθετικών σιλουετών σε σχέση με τα ΚΡΖ.

Επιπλέον, για την μελέτη των γενικευμένων στάσεων παρήχθησαν ακόμα δύο βίντεο που προσομοιώνουν τις δύο στάσεις αυτές. Οι σιλουέτες του ανθρώπου εξήχθησαν όπως έχει περιγραφεί προηγουμένως και χρησιμοποιήθηκαν οι μέθοδοι εξαγωγής χαρακτηριστικών ΚΡΖ, ΓΔΡΖ και HoG. Οι Πίνακες 3.12 και 3.13 παρουσιάζουν τα μητρώα σύγχυσης των χαρακτηριστικών ΚΡΖ και ΓΔΡΖ, αντίστοιχα. Στο Πίνακα 3.14 έχουμε συνοπτικά τα αποτελέσματα της ακρίβειας των τριών μεθόδων τόσο στο σύνολο ελέγχου από συνθετικές σιλουέτες όσο και από τις πραγματικές σιλουέτες. Βλέπουμε ότι ο γεωδαιτικός ορισμός των ροπών Zernike παρουσιάζει υψηλότερα ποσοστά ακρίβειας σε σχέση με τις άλλες μεθόδους.

3.5 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάστηκε ένας νέος ολικός περιγραφέας χαρακτηριστικών για εικόνες που λαμβάνονται μέσω ευρυγώνιας κάμερας, χρησιμοποιώντας την γεωδαιτική απόσταση. Οι περισσότεροι περιγραφείς χαρακτηριστικών είναι σχεδιασμένοι για κάμερες με τυπική γωνία θέασης (90°), χρησιμοποιώντας ως μετρική απόσταση την Ευκλείδεια απόσταση. Ο προτεινόμενος περιγραφέας είναι κατάλληλος για ευρυγώνιες κάμερες με σφαιρικό οπτικό στοιχείο και γωνία θέασης 180° , εφαρμόζοντας την γεωδαιτική απόσταση ως μετρική μεταξύ δύο εικονοστοιχείων πάνω στο σφαιρικό αυτό στοιχείο.

Τα πειραματικά αποτελέσματα στο πρόβλημα της αναγνώρισης της στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, έδειξαν σταθερή υπεροχή την προτεινόμενης μεθόδου έναντι παραδοσιακών μεθόδων εξαγωγής ολικών χαρακτηριστικών. Συγκεκριμένα, η προτεινόμενη γεωδαιτική διόρθωση ροπών Zernike παρουσίασε καλύτερα ποσοστά επιτυχίας στον έλεγχο με τα συνθετικά μοντέλα ανθρώπινων στάσεων (4% διαφορά από την δεύτερη μέθοδο), αλλά η πραγματική υπεροχή φάνηκε στα πραγματικά δεδομένα ανθρώπινης στάσης (25% διαφορά από την δεύτερη μέθοδο).

Επιπλέον, ο περιγραφέας γεωδαιτικής διόρθωσης ροπών Zernike αποδεικνύεται σθεναρός στις αλλοιώσεις του σχήματος του αντικειμένου από στροφές και μετατόπιση σε εικόνα ευρυγώνιας κάμερας, έναντι των κλασικών ροπών Zernike μέσω της πειραματικής μελέτης. Επιπλέον, η μεταβολή του σφάλματος ταξινόμησης των στάσεων καθώς ο άνθρωπος απομακρύνεται από το κέντρο της κάμερας διατηρείται σταθερή.

Κεφάλαιο 4

Επαυξημένα Νευρωνικά Δίκτυα Συνελίξεων για τον Εντοπισμό της Ανθρώπινης Στάσης

Στο Κεφάλαιο αυτό θα μελετήσουμε ένα μοντέλο Τεχνητού Νευρωνικού Δικτύου Συνελίξεων (ΤΝΔΣ) για τον εντοπισμό της ανθρώπινης στάσης σε εικόνες που λαμβάνονται από μία ευρυγώνια κάμερα. Για την ενίσχυση της απόδοσης του ΝΔΣ θα εισάγουμε εκτός της τυπικής εικόνας, επιπλέον πληροφορία από την εικόνα αυτή σε ένα ενδιάμεσο επίπεδό του, την οποία δεν μπορεί εύκολα το ίδιο το δίκτυο να την παράξει μέσω των επάλληλων συνελίξεων. Το μοντέλο αυτό θα το ονομάζουμε “Επαυξημένο Νευρωνικό Δίκτυο Συνελίξεων” (ΕΝΔΣ) και ως επιπλέον πληροφορία θα χρησιμοποιήσουμε τα ολικά χαρακτηριστικά εικόνας που παράγονται μέσω της μεθόδου ΓΔΡΖ, που φάνηκε μέσω της πειραματικής διαδικασίας του Κεφάλαιου 3 να είναι ιδιαίτερος αποδοτικά, καθώς είναι σχεδιασμένα για εικόνες του συγκεκριμένου τύπου κάμερας.

Τα ΤΝΔΣ χρησιμοποιούν σύνθετες δομές από διαδοχικά φίλτρα συνελίξεων στα ενδιάμεσα επίπεδά τους, παράγοντας τα κατάλληλα εκείνα τοπικά χαρακτηριστικά για το πρόβλημα που αντιμετωπίζουν κάθε φορά. Όμως υπάρχουν χαρακτηριστικά που παράγονται από ολικούς περιγραφείς χαρακτηριστικών (όπως οι ροπές Zernike) που δεν μπορούν να εξαχθούν μέσω συνελίξεων, έτσι η πληροφορία αυτή δεν μπορεί να παραχθεί από το ίδιο το δίκτυο. Σε προβλήματα όπως αυτό που εξετάζουμε φαίνεται ότι η πληροφορία αυτή είναι σημαντική, καθώς το παραγόμενο μοντέλο βελτιώνει τα αποτελέσματα (της αναγνώρισης της ανθρώπινης στάσης) σημαντικά όπως φαίνεται και στις εργασίες [51, 78].

Τέλος, το μοντέλο ΕΝΔΣ, αφού εκπαιδεύτηκε από συνθετικές σιλουέτες ανθρώπινης στάσης όπως τα μοντέλα του Κεφάλαιου 3, υιοθετώντας την τεχνική μεταφοράς γνώσης, επανεκπαιδεύτηκε με ένα μικρό σύνολο από πραγματικές ανθρώπινες σιλουέτες, πετυχαίνοντας επιπλέον βελτίωση της απόδοσής του.

4.1 Εισαγωγή

Τα ΤΝΔΣ έχουν χαρακτηριστεί ως η κυρίαρχη μεθοδολογία σε πολλούς τομείς της Μηχανικής Μάθησης, όπως στην αναγνώριση προτύπων [70], τον εντοπισμό αντικειμένων [97], την αναγνώριση αντικειμένων σε βάσεις δεδομένων μεγάλης κλίμακας [32], τον εντοπισμό παθολογικών καταστάσεων σε ιατρικές εικόνες [19, 83, 122] κ.α. Η επιτυχία των ΤΝΔΣ έγκειται στο γεγονός ότι δημιουργούν στο εσωτερικό της δομής του δικτύου τους τα χαρακτηριστικά εκείνα που χρειάζονται για την επίλυση του συγκεκριμένου προβλήματος. Βασιζόμενα στην ικανότητα τους να αξιοποιούν τις τοπικές ιδιότητες της μετάφρασης ενός σήματος [17], επιτρέπουν την δημιουργία αλληλεξαρτήσεων μεταξύ μεγάλων γειτονικών περιοχών του σήματος σε μικρότερες γειτονικές περιοχές.

Βασικό μειονέκτημα των ΤΝΔΣ είναι η ανάγκη μεγάλων συνόλων δεδομένων για την εκπαίδευσή τους. Μια ευρέως διαδεδομένη τεχνική τόσο σε δεδομένα εικόνας όσο και ήχου, είναι η τεχνητή αύξηση των δεδομένων εκπαίδευσης με χρήση στροφών και μετατόπισης του αντικειμένου ενδιαφέροντος [26]. Σε προβλήματα όμως που ο προσανατολισμός του αντικειμένου που θέλουμε να αναγνωρίσουμε έχει σημασία, όπως η αναγνώριση της στάσης του ανθρώπου από ευρυγώνια κάμερα που παρουσιάστηκε στο Κεφάλαιο 3, η τεχνική αυτή δεν μπορεί εύκολα να εφαρμοστεί.

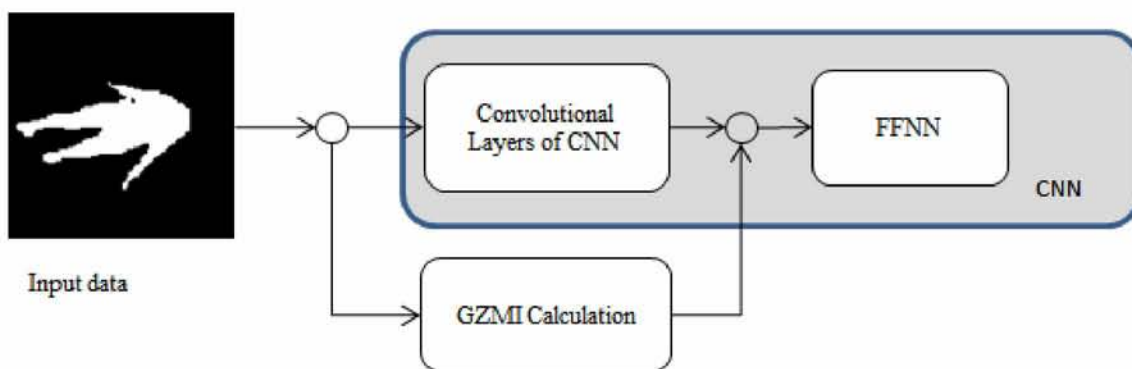
Στη ενότητα αυτή, για την αντιμετώπιση του προβλήματος του Κεφάλαιο 3 κάνοντας χρήση ΤΝΔΣ, προτείνεται η δημιουργία ενός συνόλου εκπαίδευσης με στάσης του ανθρώπου από τριδιάστατα μοντέλα (συνθετικές ανθρώπινες στάσης) και στη συνέχεια η εφαρμογή της τεχνικής μεταφοράς γνώσης, επανεκπαιδύοντας το ΤΝΔΣ με λίγες εικόνες από πραγματικές ανθρώπινες στάσεις.

4.2 Επαυξημένα Νευρωνικά Δίκτυα Συνελίξεων

Συνεχίζοντας τη μελέτη του προβλήματος που αναπτύξαμε στο Κεφάλαιο 3, θα προσπαθήσουμε να βελτιώσουμε τις επιδόσεις στην αναγνώριση της στάσης τους ανθρώπου μέσω της σιλουέτας του από ευρυγώνια κάμερα τοποθετημένη στην οροφή ενός δωματίου. Συνδυάζοντας την ικανότητα των ΤΝΔΣ στον τομέα της ταξινόμησης μαζί με τα ολικά χαρακτηριστικά που παράγουν οι ΓΔΡΖ θα δημιουργήσουμε ένα μοντέλο ικανό να αναγνωρίζει καθημερινές στάσεις του ανθρώπου από μια στατική εικόνα ευρυγώνιας κάμερας. Στην Εικόνα 4.2 παρουσιάζεται συνοπτικά η μεθοδολογία που θα ακολουθήσουμε. Το ΤΝΔΣ δέχεται ως είσοδο την εικόνα της σιλουέτας ενώ παράλληλα εξάγονται από αυτή τα χαρακτηριστικά των ΓΔΡΖ, όπου εισάγονται στο επίπεδο του τυπικού κρυφού επιπέδου από νευρώνες του ΤΝΔΣ. Το ΤΝΔΣ εκπαιδεύεται ολιστικά κάνοντας χρήση του αλγορίθμου οπισθοδρομικής διάδοσης του σφάλματος (backpropagation) [110].



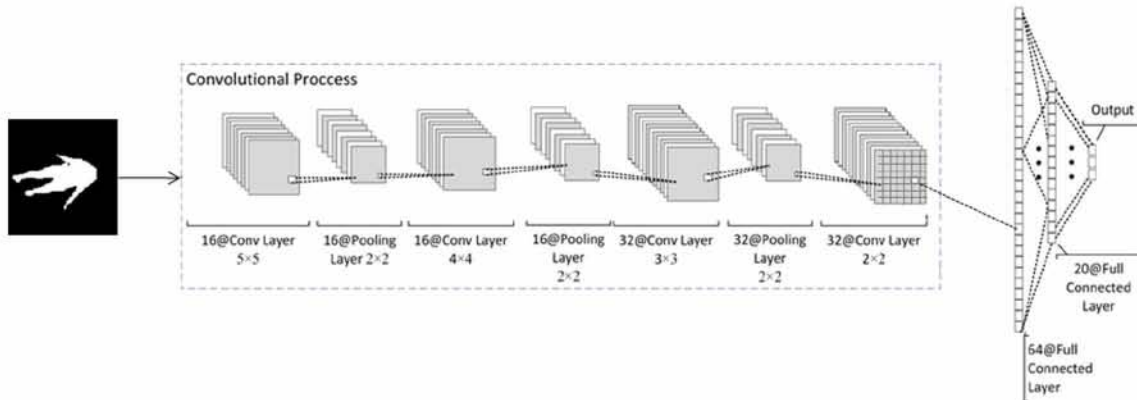
Σχήμα 4.1: Λήψη όρθιας στάσης του ανθρώπου μέσω ευρυγώνιας κάμερας, που είναι τοποθετημένη στην οροφή του δωματίου (αριστερά). Τμηματοποίηση της εικόνας και εξαγωγή της ανθρώπινης σιλουέτας (κέντρο). Υπο-εικόνα της ανθρώπινης σιλουέτας που δέχεται ως είσοδο το ΤΝΔΣ (δεξιά).



Σχήμα 4.2: Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας.

4.2.1 Αρχιτεκτονική του Επευξημένου Νευρωνικού Δικτύου Συνελίξεων

Το ΤΝΔΣ που θα χρησιμοποιήσουμε αποτελείται από τέσσερα Επίπεδα Συνελίξεων (ΕΣ) τα οποία θα αναφέρονται ως Συνεχή Διαδικασία Συνελίξεων (ΣΔΣ). Τα πρώτα τρία επίπεδα συνελίξεων ακολουθούνται από ένα επίπεδο μέγιστης υποδειγματοληψίας, ενώ το τέταρτο επίπεδο συνελίξεων από ένα πλήρως διασυνδεδεμένο feedforward δίκτυο αποτελούμενο από δύο κρυφά επίπεδα 4.3. Τα επίπεδα των συνελίξεων αποτελούνται από 16, 16, 32 και 32 φίλτρα. Το μέγεθος των φίλτρων σε κάθε ΕΣ επιλέχθη να είναι 5×5 , 4×4 , 3×3 και 2×2 , αντίστοιχα. Τα επίπεδα υποδειγματοληψίας μέγιστης τιμής έχουν ίδιο πλήθος φίλτρων με τον αριθμό των φίλτρων του επιπέδου συνελίξεων που ακολουθούν. Συγκεκριμένα, αποτελούνται από 16, 16 και 32 μητρώα χαρακτηριστικών, ενώ το μέγεθός τους είναι 2×2 . Τέλος, το τμήμα του τυπικού επιπέδου από κρυφούς νευρώνες του ΤΝΔΣ αποτελείται από ένα κρυφό επίπεδο με 64 νευρώνες, ακολουθείται από ένα κρυφό επίπεδο 20 νευρώνων και τέλος από ένα επίπεδο εξόδου με 3 νευρώνες (τόσοι όσες και οι κλάσεις/στάσεις του προβλήματός μας). Το επίπεδο της εξόδου ενσωματώνει τον γενικευμένο τύπο της συνάρτησης λογιστική παλινδρόμησης, όπου αναπαριστά το αποτέλεσμα του ΤΝΔΣ ως άθροισμα πιθανοτήτων για την κάθε διαφορετική κλάση του. Το πλήθος των επιπέδων και των φίλτρων (όπως και το μέγεθος



Σχήμα 4.3: Η βασική αρχιτεκτονική ΤΝΔΣ για τον αναγνώριση της ανθρώπινης στάσης μέσω της εικόνας σιλουέτας του ανθρώπου.

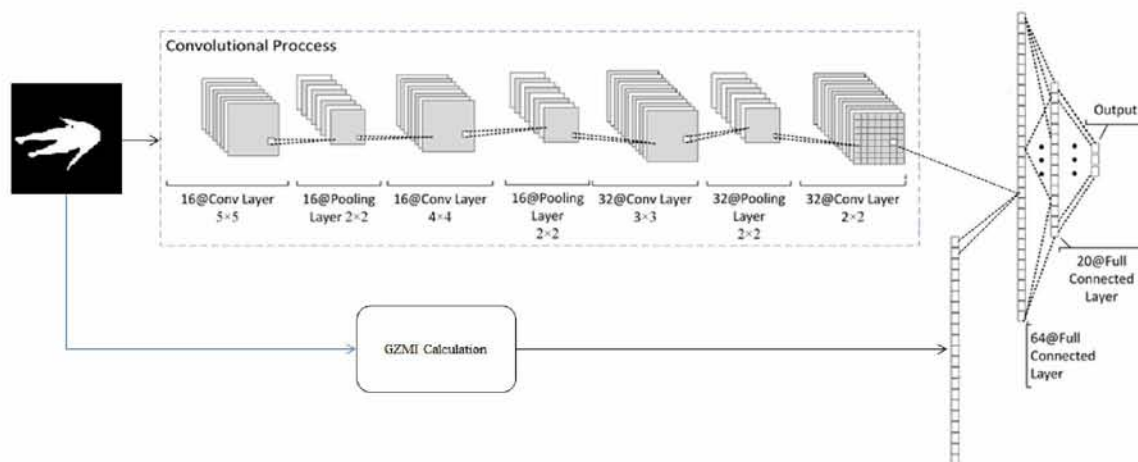
τους) προσδιορίστηκε μετά από πολλές δοκιμές και ελέγχους, έχοντας πάντα υπόψη μας ότι τα πρώτα επίπεδα ενός ΤΝΔΣ τείνουν να μαθαίνουν κλασικά φίλτρα από την θεωρία επεξεργασίας εικόνας, ενώ τα βαθύτερα εστιάζουν σε λεπτομέρειες του προβλήματος [19]. Στην Εικόνα 4.3 παρουσιάζεται σχηματικά η αρχιτεκτονική που περιγράψαμε.

Στη συνέχεια προκειμένου να βοηθήσουμε το μοντέλο στην επιτυχή ταξινόμηση των πραγματικών ανθρώπινων στάσεων δημιουργήσαμε μια ενισχυμένη έκδοση του τυπικού ΤΝΔΣ που περιγράψαμε στη προηγούμενη ενότητα. Συγκεκριμένα, προτείνουμε μια αρχιτεκτονική η οποία θα συνδυάζει τα χαρακτηριστικά που εξάγονται από την κάθε εικόνα εισόδου στο δίκτυο και παράγονται στο τέλος του ΣΔΣ με τα χαρακτηριστικά που εξάγονται από την εικόνα αυτή μέσω των ΓΔΡΖ. Τα χαρακτηριστικά αυτά έρχονται ως είσοδος στον πρώτο κρυφό νευρώνα το ΤΝΔΣ που περιγράψαμε νωρίτερα.

Ο υπολογισμός των παραγώγων που χρειάζονται κατά την διαδικασία της εκπαίδευσης του δικτύου σε συνδυασμό με την καινούρια είσοδο των ΓΔΡΖ χαρακτηριστικών, επιτυγχάνεται με την βοήθεια της βιβλιοθήκης CAFFE [71]. Η εκπαίδευση του δικτύου γίνεται με τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος, καθώς η αρχιτεκτονική μπορεί να περιγραφεί ως ένας ακυκλικός κατευθυνόμενος γράφος. Οι μερικές παράγωγοι υπολογίζονται μέσω της συνάρτησης σφάλματος και τον κανόνα της αλυσίδας. Στην Εικόνα 4.4 παρουσιάζεται σχηματικά η αρχιτεκτονική του ‘Επαυξημένου Νευρωνικού Δικτύου Συνελίξεων’ (ΕΝΔΣ).

4.2.2 Περιγραφή Καθημερινών Ανθρώπινων Στάσεων

Για να ελέγξουμε πειραματικά το ΕΝΔΣ θα δημιουργήσουμε ένα σύνολο δεδομένων από καθημερινές ανθρώπινες στάσεις μέσω τρισδιάστατων μοντέλων: τη στάση του “όρθιου” ανθρώπου, του “καθιστού” και τη στάση της “πτώσης” του ανθρώπου στο πάτωμα. Σε σύγκριση με το προηγούμενο σύνολο στάσεων αυτές είναι στάσεις που μπορεί ο άνθρωπος να κάνει στην καθημερινή του δραστηριότητα. Ο έλεγχος που θα κάνουμε στο πραγματικό βίντεο θα είναι συνεχής ταξινόμηση του κάθε στιγμιότυπου και όχι μόνο αυτών που περιέχουν μια από τις συγκεκριμένες στάσεις. Οι στάσεις των τριδιάστατων μοντέλων που χρησιμοποιήθηκαν πα-



Σχήμα 4.4: Η προτεινόμενη αρχιτεκτονική ΕΝΔΣ. Τα ΓΔΡΖ ολικά χαρακτηριστικά από μια εικόνα ανθρώπινης σιλουέτας εισέρχονται στο πλήρως διασυνδεδεμένο επίπεδο νευρώνων, σε συνδυασμό με τα τοπικά χαρακτηριστικά των αλληπάλληλων συνελίξεων.

ρουσιάζονται στην Εικόνα 4.6, όπου τα $\alpha_1 - \alpha_2$ απεικονίζουν την 'όρθια' στάση, τα $\beta_1 - \beta_2$ την 'καθιστή' και τέλος τα $\gamma_1 - \gamma_2$ την 'πτώση'. Για κάθε μια από τις στάσεις χρησιμοποιήθηκαν οχτώ, δέκα και έντεκα διαφορετικά μοντέλα αντίστοιχα [2, 3, 58].

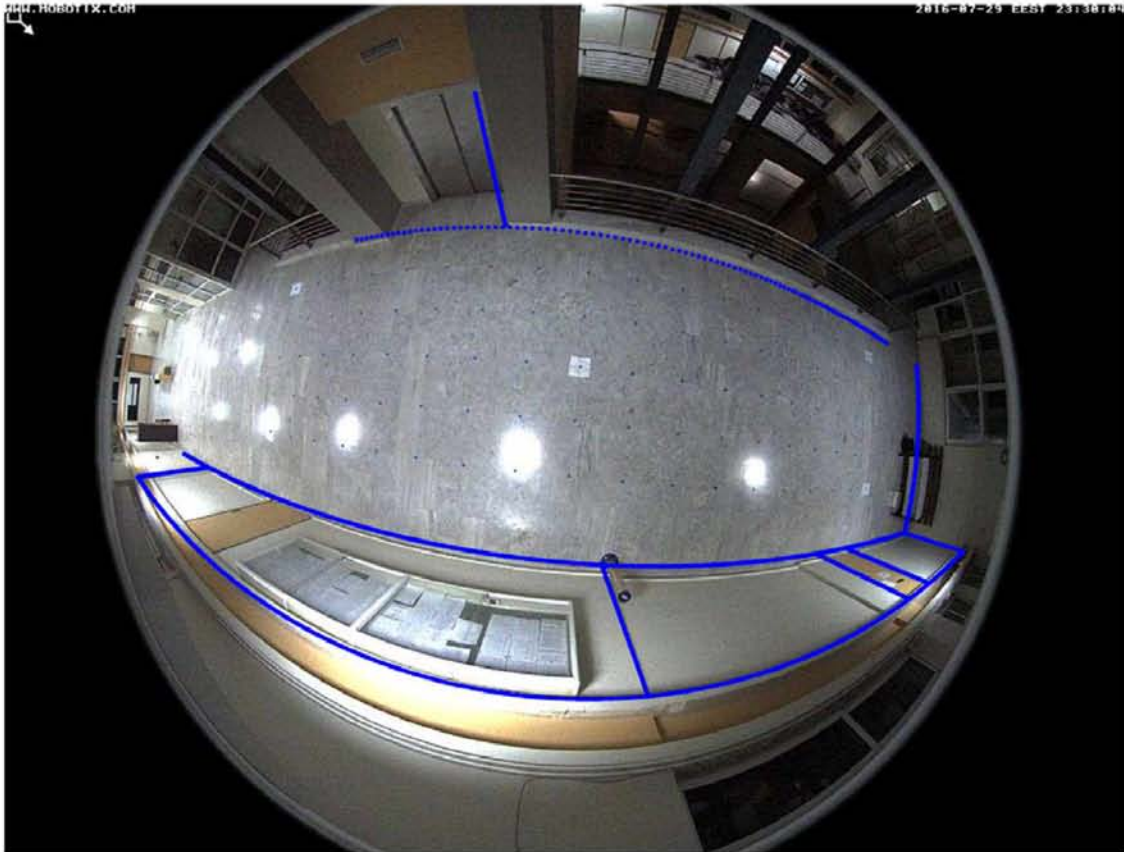
Για τις ανάγκες των καινούριων πειραμάτων τοποθετήσαμε την ευρυγώνια κάμερα σε κεντρικότερο σημείο της οροφής του δωματίου (Εικόνα 4.5) και επαναευθυγραμμίσαμε την κάμερα στο χώρο όπως περιγράψαμε στην Ενότητα 3.2.2.

4.3 Πειραματική Επιβεβαίωση των ΕΝΔΣ

Στη συνέχεια θα προχωρήσουμε στη μελέτη των ΕΝΔΣ για την επίλυση του προβλήματος της ανθρώπινης στάσης μέσω της σιλουέτας του ανθρώπου από ευρυγώνια κάμερα. Αρχικά, δημιουργήσαμε ένα συνθετικό σύνολο από σιλουέτες προσομοιώνοντας το πραγματικό δωμάτιο όπως περιγράψαμε στην Ενότητα 3.3.1. Το πλήθος των σιλουετών για τις τρεις στάσεις είναι 42, 262 και περιγράφεται αναλυτικά στον Πίνακα 4.1. Το 50% από τις σιλουέτες θα χρησιμοποιηθεί για την εκπαίδευση του ΕΝΔΣ και το υπόλοιπο για έλεγχο.

Στάση	Πλήθος Εικόνων
Όρθια	15, 916
Καθιστή	14, 674
Πτώση	11, 672
Σύνολο	42, 262

Πίνακας 4.1: Το πλήθος των συνθετικών σιλουετών για τις τρεις ανθρώπινες στάσεις.

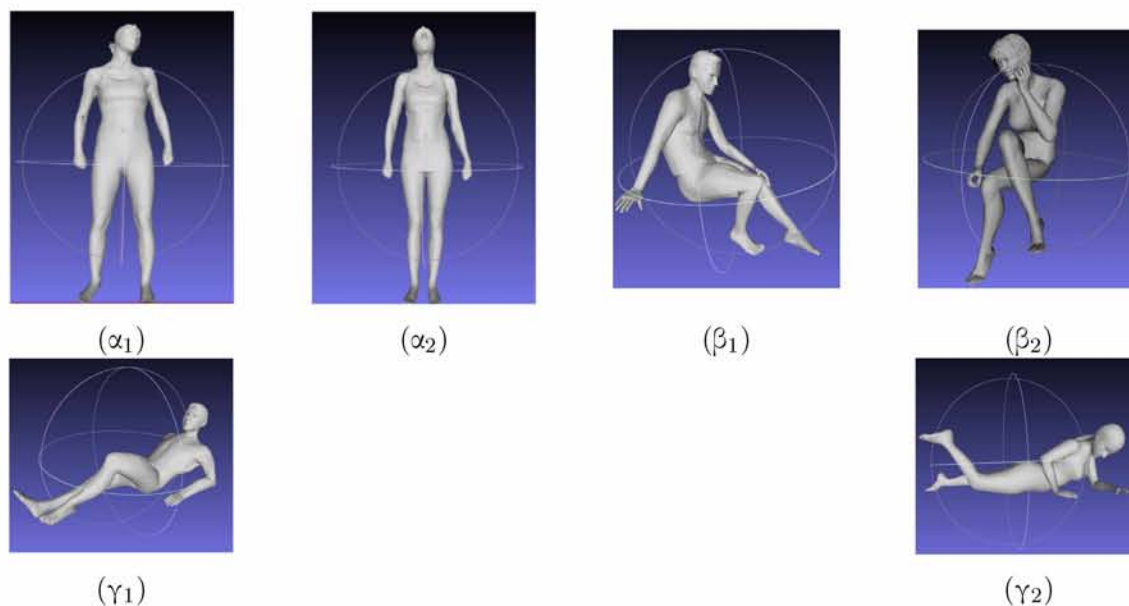


Σχήμα 4.5: Το δωμάτιο στο οποίο είναι τοποθετημένη και έχει ευθυγραμμιστεί η ευρυγώνια κάμερα. Οι παράμετροι από την ευθυγράμμιση της χρησιμοποιήθηκαν για το δωμάτιο προσομοίωσης των συνθετικών μοντέλων ανθρώπινης στάσης.

4.3.1 Ταξινόμηση των Τριών Ανθρώπινων Στάσεων σε Συνθετικά Δεδομένα

Αρχικά, θα ελέγξουμε την απόδοση του ΕΝΔΣ στις σιλουέτες ελέγχου έναντι των κλασικών ΤΝΔΣ και του περιγραφέα ΓΔΡΖ σε συνδιασμό με τον αλγόριθμο ταξινόμησης k -ΚΓ με τιμή $k = 3$. Το ΕΝΔΣ θα εκπαιδευτεί χρησιμοποιώντας το 50% των συνθετικών σιλουετών και κάνοντας χρήση ψευδο-ομάδων προτύπων με μέγεθος 50 εικόνες για κάθε επανάληψη του αλγορίθμου, με διαστάσεις 113×113 ως δεδομένα εισόδου και 20,000 επαναλήψεις. Οι τιμές του ρυθμού εκμάθησης και ορμής είναι 0,01 και 0,9, αντίστοιχα. Το ίδιο μοντέλο θα χρησιμοποιηθεί και στα κλασικά ΤΝΔΣ. Επιπλέον, ένα σύνολο 58 χαρακτηριστικών εξάγονται για κάθε εικόνα μέσω του ΓΔΡΖ που αποτελούνται από το πραγματικό και φανταστικό μέρος των 29 μιγαδικών ροπών που παράγονται από το πολυώνυμο V_{nm} το οποίο θα εισάγεται στο ΕΝΔΣ.

Στον Πίνακα 4.2 παρουσιάζεται η απόδοση των εκπαιδευμένων μοντέλων στο σύνολο ελέγχου. Όπως μπορούμε να παρατηρήσουμε τα ΕΝΔΣ παρουσιάζουν λιγότερο από 1% βελτίωση σε σχέση με τα υπόλοιπα μοντέλα ενώ τα κλασικά ΤΝΔΣ εμφανίζουν ίδιο ποσοστό επιτυχίας με το μοντέλο των ΓΔΡΖ. Στους Πίνακες 4.3, 4.4 και 4.5 παρουσιάζονται τα μητρώα



Σχήμα 4.6: Τα μοντέλα των τριών ανθρώπινων στάσεων. Στις εικόνες α_1 , α_2 παρουσιάζεται η όρθια στάση, στις εικόνες β_1 , β_2 παρουσιάζεται η καθιστή στάση και στις εικόνες γ_1 , γ_2 παρουσιάζεται η πτώση στο πάτωμα.

	Απόδοση (%)
ΕΝΔΣ	98,5
ΤΝΔΣ	97,7
ΓΔΡΖ	97,2

Πίνακας 4.2: Τα ποσοστά επιτυχίας των τριών μεθόδων στα δεδομένα ελέγχου συνθετικών σιλουετών.

σύγχυσης των ΕΝΔΣ, ΤΝΔΣ και ΓΔΡΖ αντίστοιχα.

4.3.2 Εντοπισμός Ανθρώπινης Στάσης σε Συνεχές Βίντεο

Ωστόσο, το πραγματικό ενδιαφέρον υπάρχει στον εντοπισμό της ανθρώπινης δραστηριότητας σε πραγματικές συνθήκες και όχι σε συνθετικά δεδομένα. Για τον λόγο αυτό δημιουργήθηκαν τρία βίντεο κάνοντας χρήση της ευρυγώνιας κάμερας τοποθετημένης στην οροφή του δωματίου ώστε να ελέγξουμε την απόδοση των εκπαιδευμένων μοντέλων, με τις συνθετικές σιλουέτες. Στα βίντεο ο άνθρωπος εμφανίζεται να εκτελεί τρεις καθημερινές στάσεις, την όρθια στάση, την καθιστή στάση σε μια καρέκλα και την στάση της πτώσης στο έδαφος. Κάθε στιγμιότυπο των βίντεο χαρακτηρίστηκε με χειροκίνητο τρόπο ως προς τη στάση του ανθρώπου. Επειδή όμως το πλήθος των στιγμιότυπων δεν είναι επαρκές για την εκπαίδευση των μοντέλων, που αναφέρονται στις προηγούμενες παραγράφους, θα χρησιμοποιήσουμε τα ήδη εκπαιδευμένα μοντέλα με τις συνθετικές σιλουέτες για να αναγνωρίσουμε την στάση του πραγματικού ανθρώπου. Οι σιλουέτες των πραγματικών ανθρώπων εξάγονται από τα βίντεο

		Πρόβλεψη			
		Στάση	Όρθια	Καθιστή	Πτώση
Πραγματική	Όρθια	7851	33	62	
Τιμή	Καθιστή	44	7254	60	
	Πτώση	63	50	5714	

Πίνακας 4.3: Το μητρώο σύγχυσης του ΕΝΔΣ στις συνθετικές σιλουέτες.

		Πρόβλεψη			
		Στάση	Όρθια	Καθιστή	Πτώση
Πραγματική	Όρθια	7825	63	81	
Τιμή	Καθιστή	39	7175	105	
	Πτώση	94	99	5650	

Πίνακας 4.4: Το μητρώο σύγχυσης του ΤΝΔΣ στις συνθετικές σιλουέτες.

με την βοήθεια ενός αλγορίθμου τμηματοποίησης όπως αναφέρεται και στην Παράγραφο 3.4.2.

Στον Πίνακα 4.6 παρουσιάζεται το ποσοστό επιτυχίας του κάθε μοντέλου στο καθένα από τα τρία βίντεο. Πριν προχωρήσει η ανάλυση, πρέπει να σημειωθεί ότι κάθε στιγμιότυπο των βίντεο χαρακτηρίστηκε από τα μοντέλα καθώς περιέχουν μια από τις τρεις στάσεις. Ακόμα και η στάση της μετάβασης από την μια στάση στην επόμενη, έχει παραμείνει στο βίντεο και έχει χαρακτηριστεί και αυτή από τα μοντέλα, με πιθανά λανθασμένο τρόπο, καθώς δεν έχουν εκπαιδευτεί σε αυτές. Βλέπουμε στον Πίνακα 4.6 των αποτελεσμάτων ότι τα ΤΝΔΣ αποτυγχάνουν να εντοπίσουν την στάση του πραγματικού ανθρώπου, καθώς οι σιλουέτες τους εμφανίζουν παραμόρφωση λόγω του αλγορίθμου τμηματοποίησης που χρησιμοποιήσαμε για την εξαγωγή τους. Το μοντέλο των ΓΔΡΖ εμφανίζει καλύτερα αποτελέσματα από τα ΤΝΔΣ καθώς όπως είχαμε δει εμφανίζει ανοχή στις μικρές παραμορφώσεις του σχήματος των αντικειμένων. Καθώς τα ΕΝΔΣ ενσωματώνουν την πληροφορία των ΓΔΡΖ έχουν υψηλότερα ποσοστά επιτυχίας σε σχέση με τα κλασικά ΤΝΔΣ.

Το σφάλμα που παράγεται από τον αλγόριθμο τμηματοποίησης στις σιλουέτες δεν μπορούμε να το αποφύγουμε. Για να το αντιμετωπίσουμε, θα χρησιμοποιήσουμε την τεχνική μεταφοράς γνώσης των ΤΝΔΣ από ένα εκπαιδευμένο μοντέλο για την δημιουργία ενός καινούριου μοντέλου. Συγκεκριμένα, θα χρησιμοποιήσουμε το εκπαιδευμένο μοντέλο από τις συνθετικές σιλουέτες ως βάση-αρχικοποίησης των βαρών ενός καινούριου μοντέλου, όπου θα το εκπαιδεύσουμε με ένα μικρό σύνολο από σιλουέτες με στάσεις πραγματικού ανθρώπου. Συγκεκριμένα, το μοντέλο αυτό θα εκπαιδευτεί από τα στιγμιότυπα του βίντεο 1 που είναι 1315 για 50 επαναλήψεις. Τα υπόλοιπα χαρακτηριστικά του δικτύου παραμένουν σταθερά. Αυτή η διαδικασία εφαρμόζεται τόσο τα ΕΝΔΣ όσο και στα κλασικά ΤΝΔΣ. Επιπλέον, για τον έλεγχο της απόδοσης του προτεινόμενου μοντέλου θα συγκριθεί με μια άλλη γνωστή καθιερωμένη μέθοδο για τον εντοπισμό στάσεων του ανθρώπου από στατική κάμερα, την μέθοδο του Poppe [101].

		Πρόβλεψη			
		Στάση	Όρθια	Καθιστή	Πτώση
Πραγματική	Όρθια	7785	83	101	
Τιμή	Καθιστή	59	7135	125	
	Πτώση	114	119	5610	

Πίνακας 4.5: Το μητρώο σύγχυσης του ΓΔΡΖ στις συνθετικές σιλουέτες.

Απόδοση (%)			
	Βίντεο 1	Βίντεο 2	Βίντεο 3
ΕΝΔΣ	36,6	50,9	55,3
ΤΝΔΣ	36,4	43,1	47,8
ΓΔΡΖ	51,3	63,5	72,4

Πίνακας 4.6: Τα ποσοστά επιτυχίας των εκπαιδευμένων μοντέλων με συνθετικές σιλουέτες, για την αναγνώριση των πραγματικών ανθρώπινων σιλουετών.

Στον Πίνακα 4.7 παρουσιάζεται το ποσοστό επιτυχίας των μοντέλων στα δύο βίντεο καθώς το βίντεο 1 έχει χρησιμοποιηθεί για την εκπαίδευση του μοντέλου. Βλέπουμε ότι το ΕΝΔΣ με την χρήση της μεταφοράς γνώσης εμφανίζει σημαντική βελτίωση σε σχέση χωρίς τη μεταφορά γνώσης, ενώ συνολικά έχει καλύτερα αποτελέσματα και από τα κλασικά ΤΝΔΣ αλλά και από την μέθοδο του Poppe. Τα μητρώα σύγχυσης των μεθόδων ΕΝΔΣ και ΤΝΔΣ στις στάσεις των ανθρώπινων σιλουετών των Βίντεο 2 και 3 παρουσιάζονται στους Πίνακες 4.8 και 4.9, αντίστοιχα.

Απόδοση (%)		
	Βίντεο 2	Βίντεο 3
ΕΝΔΣ	73,2	80,1
ΤΝΔΣ	65,5	77,8
Poppe	47,3	74,5

Πίνακας 4.7: Τα ποσοστά επιτυχίας των εκπαιδευμένων μοντέλων με πραγματικές ανθρώπινες σιλουέτες του βίντεο 1, για την αναγνώριση των πραγματικών ανθρώπινων σιλουετών. Στα ΕΝΔΣ και ΤΝΔΣ έχει χρησιμοποιηθεί η μέθοδος μεταφοράς γνώσης από τα αρχικά εκπαιδευμένα μοντέλα με συνθετικές σιλουέτες, ενώ η μέθοδος του Poppe έχει εκπαιδευτεί εξολοκλήρου με τις πραγματικές σιλουέτες του βίντεο 1.

	Στάση	Πρόβλεψη		
		Όρθια	Καθιστή	Πτώση
Πραγματική Τιμή	Όρθια	90	157	54
	Καθιστή	207	495	248
	Πτώση	135	211	1900

Πίνακας 4.8: Το μητρώο σύγχυσης του ΤΝΔΣ στις πραγματικές σιλουέτες ανθρώπων από τα Βίντεο 2 και 3.

	Στάση	Πρόβλεψη		
		Όρθια	Καθιστή	Πτώση
Πραγματική Τιμή	Όρθια	212	147	54
	Καθιστή	143	500	146
	Πτώση	77	216	1963

Πίνακας 4.9: Το μητρώο σύγχυσης του ΕΝΔΣ στις πραγματικές σιλουέτες ανθρώπων από τα Βίντεο 2 και 3.

4.4 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια νέα αρχιτεκτονική νευρωνικών δικτύων συνελίξεων επαυξημένης πληροφορίας, όπου το μοντέλο εκπαιδεύεται όχι μόνο με τα τοπικά χαρακτηριστικά που δημιουργεί μέσω των συνελίξεών του, αλλά και με επιπλέον πληροφορία που δεν μπορεί το ίδιο να δημιουργήσει. Το μοντέλο αυτό χρησιμοποιήθηκε για την αναγνώριση των καθημερινών ανθρώπινων στάσεων από ευρυγώνια κάμερα τοποθετημένη στην οροφή του δωματίου, με σκοπό την καταγραφή και παρακολούθηση της δραστηριότητας του ανθρώπου για την αντιμετώπιση επειγουσών καταστάσεων (π.χ. η πτώση ενός ηλικιωμένου ανθρώπου) κ.α.

Η προτεινόμενη μέθοδος συνδύασε την πληροφορία που παράγει η μέθοδος των ολικών περιγραφών διορθωμένων γεωδαιτικών ροπών Zernike, καθώς επιλέχθηκε ως καταλληλότερη για το συγκεκριμένο πρόβλημα αφού πρώτα μελετήθηκε και συγκρίθηκε με άλλους ολικούς περιγραφείς, με τα νευρωνικά δίκτυα συνελίξεων. Το προτεινόμενο μοντέλο κατάφερε επιτυχή αναγνώριση της ανθρώπινης δραστηριότητας σε συνεχόμενα βίντεο με μέση απόδοση 76%. Πρέπει να τονισθεί ότι το μοντέλο σε πρώτη φάση εκπαιδεύτηκε με συνθετικές σιλουέτες της ανθρώπινης δραστηριότητας και στη συνέχεια, αξιοποιώντας την μέθοδο μεταφοράς γνώσης επανεκπαιδεύτηκε με λίγες πραγματικές ανθρώπινες σιλουέτες.

Το προτεινόμενο μοντέλο συγκρίθηκε με άλλες μεθόδους (νευρωνικά δίκτυα συνελίξεων, μέθοδο διορθωμένων γεωδαιτικών ροπών Zernike, μέθοδο του Porre) και τα αποτελέσματα έδειξαν σαφή υπεροχή του στο πρόβλημα αυτό.

Κεφάλαιο 5

Μέθοδος Χαμηλής Υπολογιστικής Ισχύος για τον Εντοπισμό της Ανθρώπινης Πτώσης

Στο κεφάλαιο αυτό θα ασχοληθούμε με τον εντοπισμό της ανθρώπινης πτώσης κάνοντας χρήση έξυπνων φορητών συσκευών. Σκοπός του κεφαλαίου είναι η δημιουργία ενός υπολογιστικού μοντέλου χαμηλής πολυπλοκότητας, το οποίο θα μπορεί να λειτουργήσει σε ασθενείς υπολογιστικά συσκευές, όπου λαμβάνοντας σήματα από τους αισθητήρες των συσκευών αυτών να μπορεί να προβλέψει τυχόν πτώση του ανθρώπου.

Το υπολογιστικό σύστημα για την ανάλυση των σημάτων βασίζεται στη χρήση δύο χαμηλού υπολογιστικού κόστους μεθόδων. Αρχικά μελετήσαμε τη μέθοδο διαδοχικής ανάλυσης κυρίων συνιστωσών, που σκοπό έχει τη μείωση της διάστασης του συνόλου των σημάτων που λαμβάνουμε σε μια διάσταση. Στη συνέχεια εφαρμόσαμε την ανάλυση του μονοδιάστατου μετασχηματισμένου σήματος μέσω της τεχνικής στατιστικού ελέγχου Συσσωρευτικού Αθροίσματος (ΣΥΣΑΘ). Η προτεινόμενη προσέγγιση δίνει την δυνατότητα χρήσης σημάτων πολύ μεγάλης διάστασης, συγκριτικά με άλλες μεθόδους εντοπισμού της ανθρώπινης πτώσης.

Το προτεινόμενο σύστημα [48, 49], έχει εφαρμοστεί και εξεταστεί σε ένα σύνολο δεδομένων μεγάλης διάστασης, αποτελούμενο από διαφορετικές ανθρώπινες δραστηριότητες, που παρήχθησαν μέσω αισθητήρων ενός έξυπνου κινητού τηλεφώνου, με πολλαπλές αλλαγές της δραστηριότητας του ανθρώπου [48]. Επιπλέον, εξετάστηκε σε ένα σύνολο δεδομένων μικρότερης διάστασης σημάτων, ενός έξυπνου κινητού τηλεφώνου, για τον εντοπισμό της πτώσης του ανθρώπου [49]. Το προτεινόμενο σύστημα έδειξε ότι παράγει ακριβή αποτελέσματα με πολύ μικρή εμφάνιση λανθασμένων ενδείξεων.

5.1 Εισαγωγή

Καθώς οι συσκευές συνδέονται μεταξύ τους και αντιλαμβάνονται τον περιβάλλοντα χώρο με την βοήθεια ενός πλήθους αισθητήρων που διαθέτουν, γίνονται όλο και περισσότερο ανθρωποκεντρικές. Φορητές και φορετές συσκευές συνδέονται με τον άνθρωπο αντλώντας πληροφορίες-δεδομένα από αυτόν, αξιοποιώντας τα δεδομένα που συλλέγουν είτε σε κάποιον απομακρυσμένο εξυπηρετητή είτε από τις ίδιες τις συσκευές. Σκοπός της περιβάλλουσας νοημοσύνης είναι να εφοδιάσει τις συσκευές αυτές με το κατάλληλο λογισμικό, το οποίο θα μπορεί να λύνει καθημερινά και ζωτικής σημασίας προβλήματα του ανθρώπου.

Ένα τέτοιο πρόβλημα, είναι ο εντοπισμός της ανθρώπινης πτώσης του ανθρώπου, με σκοπό την έγκυρη και έγκαιρη ειδοποίηση κάποιου οικείου προσώπου ή ειδικού προσωπικού. Όπως έχει παρατηρηθεί σε πολλές έρευνες [88, 96, 133], περίπου το 33% των ηλικιωμένων ανθρώπων πάνω από την ηλικία των 65 χρόνων και το 50% των ηλικιωμένων άνω των 85 βιώνουν μια πτώση κάθε χρόνο. Ένα τέτοιο ατύχημα πολλές φορές μπορεί να επιφέρει από ελαφρείς τραυματισμούς μέχρι και πολύ σοβαρούς για την ζωή τους. Μελέτες επισημαίνουν ότι το 50% των ηλικιωμένων που έχουν υποστεί κάταγμα του ισχίου δεν μπορούν να ζήσουν ανεξάρτητοι, ενώ ένα ποσοστό 25% έως 33% αυτών, αναλόγως της ηλικιακής ομάδας στην οποία ανήκουν, κινδυνεύει να πεθάνει μετά από 6 μήνες έως έναν χρόνο.

Η δυνατότητα να μπορεί ο άνθρωπος να ζει και να κινείται ανεξάρτητος, χωρίς την συνεχή επιτήρηση κάποιου τρίτου είναι ένα δικαίωμα όλων, πολλοί όμως ηλικιωμένοι χάνουν το δικαίωμα αυτό καθώς η επιτήρησή τους γίνεται αναγκαστική. Η διακριτική επίβλεψή τους χωρίς τον περιορισμό της ιδιωτικότητάς τους και τον περιορισμό της ανεξαρτησίας τους είναι επιβεβλημένη και αποτελεί έναν καινούριο τομέα στο σύστημα φροντίδας των ανθρώπων και αυτόματου εντοπισμού επειγόντων περιστατικών.

Πολλά ολοκληρωμένα συστήματα έχουν παρουσιαστεί στη βιβλιογραφία [33, 34, 77, 89], που σκοπό έχουν το εντοπισμό ενός επείγοντος περιστατικού και στη συνέχεια ενημέρωση ενός οικείου προσώπου για βοήθεια. Μέθοδοι που μπορούν να εντοπίσουν την πτώση ενός ανθρώπου με την χρήση κάμερας, όπως παρουσιάσαμε στα Κεφάλαια 3 και 4, δίνουν την ευκολία της διακριτικής επιτήρησης, χωρίς να επεμβαίνουν στον ίδιο τον άνθρωπο αλλάζοντάς του τον καθημερινό τρόπο ζωής του. Αν όμως βρεθεί εκτός εμβέλειας της κάμερας και συμβεί ένα επείγον περιστατικό το σύστημα δεν θα μπορέσει να το εντοπίσει. Έτσι υπάρχει η τάση της αξιοποίησης των φορητών ή φορετών συσκευών, καθώς τέτοιες συσκευές χρησιμοποιεί ο άνθρωπος στην καθημερινότητά του.

Το επιταχυνσιόμετρο είναι ένας αισθητήρας καταγραφής κίνησης που ενσωματώνεται σε πολλές συσκευές. Έχει παρατηρηθεί ότι κατά την πτώση του ανθρώπου συμβαίνει μια ταχεία επιτάχυνση πραγματοποιώντας αλλαγές στους τρεις άξονες του επιταχυνσιόμετρου, τους άξονες X, Y και Z. Με μια κατάλληλη επεξεργασία των δεδομένων αυτών μπορεί να επιτευχθεί εντοπισμός της πτώσης. Ένα άλλο μεγάλο πλεονέκτημα των αισθητήρων κίνησης είναι το χαμηλό κόστος και το εξαιρετικά μικρό μέγεθός τους, επιτρέποντας έτσι την ενσωμάτωσή τους σε φορητές και φορετές συσκευές (π.χ. σε ένα ρολόι καρπού ή κινητό τηλέφωνο, σε ένα πουκάμισο, κ.λ.π.). Αυτό το γεγονός κάνει τους αισθητήρες κίνησης ιδανικούς για συνεχείς

καταστάσεις παρακολούθησης, σε οποιοδήποτε τύπο περιβάλλοντος (εσωτερικό ή εξωτερικό περιβάλλον), ανεξάρτητα από τη δραστηριότητα του χρήστη.

Ένα βασικό μειονέκτημα των αισθητήρων κίνησης (επιταχυνσιομέτρων) είναι ο λανθασμένος εντοπισμός επείγοντος περιστατικού (π.χ. όταν ο χρήστης εκτελεί ταχείες κινήσεις, τρέξιμο, κατέβασμα σκάλας, κ.λ.π.). Πολλές εργασίες [84, 125] εντοπισμού πτώσης του ανθρώπου βασίζονται στη χρήση δεδομένων αισθητήρων κίνησης, μετασχηματίζοντας το τριδιάστατο σήμα που παράγουν οι αισθητήρες (δηλ. τους X , Y και Z άξονες) σε μονοδιάστατο με χρήση του τύπου 5.1 και στη συνέχεια με μεθόδους καταωφλιού υπολογίζουν ένα κατάλληλο κατώφλι ώστε να εντοπίζουν την πτώση:

$$sig = \sqrt{(X^2 + Y^2 + Z^2)}. \quad (5.1)$$

Το πρόβλημα όμως σε αυτές τις μεθόδους είναι ότι καθώς τα όρια του καταωφλιού έχουν υπολογιστεί σε ένα συγκεκριμένο σύνολο δεδομένων, δεν μπορούν να εφαρμοστούν με την ίδια ακρίβεια σε δεδομένα που παράγονται από διαφορετικές συσκευές, αφού πιθανά διαφέρουν λόγω σφάλματος ή διαφοροποίησης του τρόπου συλλογής από τη συσκευή. Ένα δεύτερο σημαντικό μειονέκτημά τους είναι η δυσκολία εύρεσης του κατώφλιού που μπορεί να ξεχωρίσει τα δεδομένα. Έτσι, έχουν προταθεί μέθοδοι αυτόματου εντοπισμού της πτώσης, χρησιμοποιώντας στατιστικές μεθόδους εντοπισμού αλλαγών σε σήματα ροών δεδομένων [125]. Ωστόσο αυτές οι μέθοδοι εμφανίζουν περιορισμούς αφού μπορούν να αξιοποιήσουν πεπερασμένο αριθμό διάστασης των σημάτων, καθώς αν αυξήσουμε τα δεδομένα σε περισσότερες από τρεις διαστάσεις δεν υπάρχει κάποιος κατάλληλος μετασχηματισμός τους σε μια διάσταση όπως απαιτούν.

Αυτό το πρόβλημα ερχόμαστε να αντιμετωπίσουμε συνδυάζοντας κατάλληλα γνωστές μεθόδους μείωσης της διάστασης των δεδομένων και εντοπισμού αλλαγών σε μονοδιάστατα σήματα. Το προτεινόμενο σύστημα [48, 49], έχει εφαρμοστεί και εξεταστεί σε ένα σύνολο δεδομένων με διαφορετικές δραστηριότητες ανθρώπων, χρησιμοποιώντας αισθητήρες έξυπνου κινητού τηλεφώνου μεγάλης διάστασης με πολλαπλές αλλαγές της ανθρώπινης δραστηριότητας, όπως και σε ένα σύνολο δεδομένων από σήματα επιταχυνσιομέτρου (τριών διαστάσεων) ενός έξυπνου κινητού τηλεφώνου με πτώση ανθρώπου. Το προτεινόμενο σύστημα έδειξε ότι παράγει ακριβή αποτελέσματα με πολύ μικρή εμφάνιση λανθασμένων ενδείξεων.

5.2 Μέθοδος Εντοπισμού Αλλαγών σε Σήματα Μεγάλης Διάστασης

Στην ενότητα αυτή θα παρουσιάσουμε μια μεθοδολογία εντοπισμού αλλαγών σε σήματα μεγάλης διάστασης αξιοποιώντας την Ανάλυση Κυρίων Συνιστωσών (ΑΚΣ) για τον μετασχηματισμό των πολυδιάστατων σημάτων σε μια διάσταση και στην συνέχεια χρησιμοποιώντας την τεχνική Συσσωρευτικού Αθροίσματος (ΣΥΣΑΘ) θα εντοπίσουμε τυχόν αλλαγές που μπορεί να εμφανίζονται σε αυτό. Σκοπός της συγκεκριμένης μεθοδολογίας είναι να μπορεί να εφαρμοστεί σε πραγματικά δεδομένα που λαμβάνονται από φορητές συσκευές και ολόκληρη

η επεξεργασία να γίνεται εξ' ολοκλήρου σε αυτές. Έτσι θα πρέπει να μεθοδολογία να είναι χαμηλού υπολογιστικού κόστους.

5.2.1 Διαδοχική Ανάλυση Κυρίων Συνιστωσών

Η Ανάλυση Κυρίων Συνιστωσών είναι μια στατιστική μέθοδος μείωσης της διάστασης των δεδομένων όπου μετασχηματίζει τα αρχικά δεδομένα μέσω μια ομάδας μη γραμμικών συσχετιζόμενων μεταβλητών, οι οποίες καλούνται κύριες συνιστώσες. Βασικό χαρακτηριστικό των κύριων συνιστωσών είναι η αντιστοιχία τους με γραμμικούς διανυσματικούς υπόχωρους που εμφανίζουν τη μέγιστη δυνατή διακύμανση μεταξύ των μεταβλητών των αρχικών δεδομένων. Επιλέγοντας τους υπόχωρους αυτούς με τη μεγαλύτερη διασπορά, τα αρχικά δεδομένα μετασχηματίζονται σε νέα μικρότερης διάστασης, διατηρώντας το μεγαλύτερο μέρος της πληροφορίας τους. Το μέγεθος της νέας διάστασης των δεδομένων εξαρτάται από το πλήθος των κυρίων συνιστωσών που θα χρησιμοποιηθούν (και των διανυσματικών υπόχωρων κατ' επέκταση).

Ωστόσο, ο υπολογισμός των κυρίων συνιστωσών βασίζεται στην χρήση του υπολογισμού των μητρώων συνδιασπορών, όπου το υπολογιστικό κόστος είναι ιδιαίτερα μεγάλο, κυρίως σε συσκευές χαμηλής υπολογιστικής ισχύος. Σε περιπτώσεις όπου τα δεδομένα έρχονται με συχνότητα δειγματοληψίας 50hz, ο συνεχής υπολογισμός του μητρώου συνδιασπορών και ο μετασχηματισμός των δεδομένων στις καινούριες κύριες συνιστώσες είναι απαγορευτικός. Στο συγκεκριμένο πρόβλημα για να επιτύχουμε χαμηλό υπολογιστικό κόστος στην ανάλυση των σημάτων, το πολυδιάστατο αρχικό σήμα θα μετασχηματιστεί σε μονοδιάστατο. Έτσι ιδιαίτερη σημασία έχει ο γρήγορος υπολογισμός της πρώτης κύριας συνιστώσας και όχι όλων των κυρίων συνιστωσών του κάθε νέου δείγματος.

Για τον λόγο αυτό έχουν αναπτυχθεί παραλλαγές της αρχικής μεθόδου κυρίων συνιστωσών, που ονομάζονται μέθοδοι Διαδοχικής Ανάλυσης Κυρίων Συνιστωσών (ΔΑΚΣ), για προβλήματα όπου τα δεδομένα έρχονται με μεγάλο ρυθμό και θα πρέπει η ανάλυσή τους να γίνεται άμεσα. Σκοπός των μεθόδων αυτών είναι η ανεξαρτητοποίησή τους από την χρήση του μητρώου συνδιασπορών, με τον υπολογισμό των κυρίων συνιστωσών να επιτυγχάνεται μέσω προσεγγιστικών μεθόδων [94, 95, 113]. Ένα πρόβλημά που παρουσιάζουν οι συγκεκριμένες μέθοδοι είναι το μεγάλο σφάλμα υπολογισμού των κυρίων συνιστωσών όταν τα αρχικά δεδομένα είναι πολύ μεγάλης διάστασης.

Αυτό το πρόβλημα έρχεται να αντιμετωπίσει η μέθοδος Διαδοχικής Ανάλυσης Κυρίων Συνιστωσών χωρίς Συνδιασπορές (ΔΑΚΣΣ) Candid Covariance-free IPCA (CCIPCA) [132], η οποία βασίζεται στις εργασίες των Oja και Karhunen [95] όπως και Sanger [113], ανανεώνοντας το διάνυσμα κυρίων συνιστωσών με κάθε είσοδο ενός νέου δείγματος δεδομένων, χωρίς να κάνει χρήση υπολογισμού του μητρώου συνδιασπορών. Για λόγους πληρότητας θα δώσουμε μια μικρή περιγραφή της μεθόδου για τον υπολογισμό της πρώτης κύριας συνιστώσας u_1 .

Έστω ένα d_n , $n = 1, 2, \dots$ τα a -διάστατα διανύσματα των σημάτων που έρχονται στην n χρονική στιγμή και η κάθε διάσταση αντιστοιχεί στο σήμα ενός αισθητήρα. Χωρίς σφάλμα της

γενικότητας, μπορούμε να θεωρήσουμε ότι η μέση τιμή του διάνυσματος d_n είναι μηδέν, καθώς μπορεί να προσεγγιστεί με διαδοχικό τρόπο. Έτσι, στο n -βήμα η πρώτη κύρια συνιστώσα u_1^n (που μας ενδιαφέρει για το πρόβλημα μας) προσεγγίζεται με την εξίσωση

$$u_1^n = \frac{n-1-l}{n} u_1^{n-1} + \frac{1+l}{n} d_n d_n^T \frac{u_1^{n-1}}{\|u_1^{n-1}\|},$$

όπου $(n-1)/n$ είναι ένας συντελεστής βαρύτητας για την τελευταία προσέγγιση ενώ το $1/n$ είναι ένας συντελεστής βαρύτητας για τα καινούρια δεδομένα. Τέλος, το διάνυσμα d_n μετασχηματίζεται σε μονοδιάστατο διάνυσμα y_n μέσω του τύπου

$$y_n = u_1^n d_n.$$

Η παράμετρος l ονομάζεται παράμετρος λήθης, δίνοντας την ικανότητα στην μέθοδο να δίνει ιδιαίτερη βαρύτητα για τον υπολογισμό του καινούριου χώρου των κυρίων συνιστωσών, στα τελευταία δεδομένα που επεξεργάζεται. Αυτό βοηθάει την μέθοδο να μην είναι μεροληπτική στα αρχικά δεδομένα και να προσαρμόζεται στις αλλαγές που εμφανίζουν τα καινούρια δεδομένα. Η μαθηματική απόδειξη της σύγκλισης της μεθόδου ΔΑΚΣΣ υπάρχει στην εργασία [142].

Αφού τα αρχικά πολυδιάστατα σήματα μετατρέπονται σε μονοδιάστατα θα πρέπει να χρησιμοποιήσουμε έναν αποδεδειγμένα χαμηλής πολυπλοκότητας αλγόριθμο για την εύρεση αλλαγών σε αυτά. Για τον σκοπό αυτό επιλέχθηκε ο αλγόριθμος ΣΥΣΑΘ και παρουσιάζεται στη συνέχεια.

5.2.2 Αλγόριθμος Συσσωρευτικού Αθροίσματος

Ο αλγόριθμος ΣΥΣΑΘ είναι ένας στατιστικός αλγόριθμος εύρεσης αλλαγών σε μονοδιάστατα δεδομένα και μπορεί να χρησιμοποιηθεί είτε σε στατικά δεδομένα είτε σε ροές δεδομένων [98]. Για το πρόβλημα του εντοπισμού ανθρώπινης πτώσης, όπου τα δεδομένα εμφανίζονται ως ροές, θα εφαρμόσουμε την αντίστοιχη έκδοση του αλγορίθμου, που θα περιγράψουμε στη συνέχεια.

Θεωρώντας μια ακολουθία από ανεξάρτητες τυχαίες μεταβλητές y_n οι οποίες αντιστοιχούν σε ένα σήμα σε διακριτές χρονικές στιγμές n , παραγόμενες από μια συνάρτηση πυκνότητας πιθανότητας $p_\theta(y)$ η οποία εξαρτάται μόνο από τις τιμές της παραμέτρου θ . Για να εντοπίσουμε μια αλλαγή στο σήμα σε κάποια χρονική στιγμή t_0 , η παράμετρος θ θα πρέπει να αλλάξει από την αρχική της τιμή, έστω θ αρχικοποιείται ως θ_0 , με θ_1 όπου $\theta_0 \neq \theta_1$.

Συγκεκριμένα, για το πρόβλημα εντοπισμού της ανθρώπινης δραστηριότητας που σκοπεύουμε να εφαρμόσουμε την μέθοδό μας, θα θεωρούμε τις ακόλουθες υποθέσεις για την παράμετρο θ :

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &= \theta_1. \end{aligned} \tag{5.2}$$

Όταν ικανοποιείται η υπόθεση H_0 τότε το σήμα δεν εμφανίζει καμία μεταβολή και η κατάσταση της ανθρώπινης δραστηριότητας δεν αλλάζει, ενώ όταν ικανοποιείται η υπόθεση H_1

τότε συμβαίνει μεταβολή στην ανθρώπινη κατάσταση. Έστω,

$$S_n = \sum_{i=1}^n s_i, \text{ όπου } s_i = \ln \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} \quad (5.3)$$

ο λόγος της λογαριθμικής πιθανότητας των παρατηρήσεων y_i στο y_n , όπου n αντιπροσωπεύει την τρέχουσα χρονική στιγμή, ενώ το s_i αναφέρεται ως στατιστικό μέτρο. Θεωρώντας ότι τα δεδομένα ακολουθούν την κανονική κατανομή, με μ να αντιστοιχεί στη μέση τιμή και σ στη διασπορά, η συνάρτηση πυκνότητας πιθανότητας ορίζεται ως:

$$p_{\theta}(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp^{-(y-\theta)^2/2\sigma^2}, \quad (5.4)$$

ενώ το στατιστικό μέτρο s_i ορίζεται ως:

$$s_i = \frac{\theta_1 - \theta_0}{\sigma^2} (y_i - \frac{\theta_0 + \theta_1}{2}). \quad (5.5)$$

Για να επιτευχθεί ο εντοπισμός μιας αλλαγής, σε κάθε χρονική στιγμή, εφαρμόζεται ο παρακάτω κανόνας:

$$g_n = S_n - \mu_n \geq h, \text{ όπου } \mu_n = \min_{1 \leq j \leq n} S_j. \quad (5.6)$$

Ως κριτήριο τερματισμού ορίζουμε το t_a

$$t_a = \min \{n : g_n \geq h\}, \quad (5.7)$$

όπου μπορεί να μετασχηματιστεί στη μορφή

$$t_a = \min \{n : S_n \geq \mu_n + h\}. \quad (5.8)$$

Ο παραπάνω κανόνας συγκρίνει το συσσωρευτικό άθροισμα S_n με το κατώφλι $\mu_n + h$, καθώς το μ_n διαιρεί την πληροφορία της μέσης τιμής από τις προηγούμενες παρατηρήσεις, ενώ η παράμετρος h πρέπει να ορισθεί. Η παράμετρος h είναι κρίσιμη για την απόφαση του αλγορίθμου και την εύρεση αλλαγής στο σήμα.

Για τον προσδιορισμό των παραμέτρων για τον ορισμό των υποθέσεων 5.2 αλλά και τον προσδιορισμό της u_1^0 της μεθόδου ΔΑΣΣ χρησιμοποιείται ένα σύνολο εκπαίδευσης από παρατηρήσεις μεγάλης διάστασης χωρισμένες σε δύο κλάσεις (παρατηρήσεις που εμφανίζουν ανθρώπινη πτώση και παρατηρήσεις απλής ανθρώπινης δραστηριότητας).

Αρχικά, υπολογίζουμε την πρώτη κύρια συνιστώσα u_1^0 χρησιμοποιώντας ολόκληρο το δείγμα εκπαίδευσης με την κλασική μέθοδο ΑΚΣ. Στη συνέχεια, μετασχηματίζουμε σε μονοδιάστατο σήμα τα δεδομένα μεγάλης διάστασης του συνόλου εκπαίδευσης και υπολογίζουμε την μέση τιμή των παρατηρήσεων αυτών που αντιστοιχούν σε ανθρώπινη πτώση, αντιστοιχίζοντας με αυτή την τιμή το θ_1 , ενώ αντίστοιχα η μέση τιμή των μετασχηματισμένων παρατηρήσεων απλής ανθρώπινης δραστηριότητας υποδηλώνει την τιμή της παραμέτρου θ_0 .

Στη συνέχεια, καθώς ροές παρατηρήσεων εμφανίζονται, εφαρμόζεται η μέθοδος ΔΑΣΣ ώστε διαδοχικά να μειώνει την διάσταση των παρατηρήσεων σε μία διάσταση. Για τον εντοπισμό αλλαγής και κατ' επέκταση εντοπισμό της ανθρώπινης πτώσης στα μετασχηματισμένα δεδομένα χρησιμοποιείται ο κανόνας απόφασης g_n .

5.3 Πειραματικά Αποτελέσματα

Στην παράγραφο αυτή θα αναφερθούμε στα πειραματικά αποτελέσματα της μεθόδου σε δύο διαφορετικά σύνολα δεδομένων μεγάλης διάστασης για τον εντοπισμό την ανθρώπινης πτώσης. Τα σύνολα δεδομένων προέρχονται από διαφορετικούς αισθητήρες φορητών έξυπνων συσκευών.

5.3.1 Πρόβλημα Εντοπισμού Ανθρώπινης Πτώσης σε Τεχνητά Υψηλής Διάστασης Δεδομένα

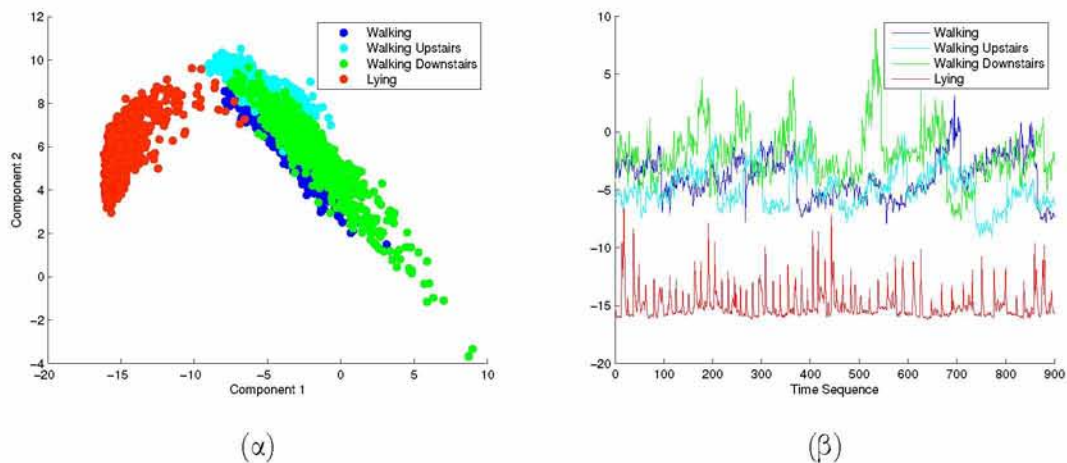
Στην υποενότητα αυτή θα χρησιμοποιήσουμε το σύνολο δεδομένων “Human Activities and Postural Transitions¹” (HAPT) [106] για τον εντοπισμό πτώσης έναντι διαφορετικών καταστάσεων ανθρώπινης βάρδισης. Ο σκοπός που χρησιμοποιείται το συγκεκριμένο σύνολο δεδομένων είναι η εδραίωση της προτεινόμενης μεθόδου, αφού τα δεδομένα του συνόλου δεδομένων είναι υψηλής διάστασης. Στο σύνολο αυτό έχουν καταγραφεί πολλές διαφορετικές δραστηριότητες του ανθρώπου και μπορούν να διακριθούν γενικά σε ενεργείς και μη-ενεργείς. Εμείς θα ασχοληθούμε μόνο με τις εξής καταστάσεις, την βάρδιση του ανθρώπου στο επίπεδο όπως και το ανέβασμα και κατέβασμα στη σκάλα χωρίς όμως ο άνθρωπος να τρέχει. Αυτές τις καταστάσεις της θεωρούμε ως μία κατάσταση, ενώ η ως δεύτερη κατάσταση θεωρείται όταν ο άνθρωπος κείται στο πάτωμα, που αντιστοιχεί στην πτώση και θεωρείται ότι αποτελεί μια αλλαγή στην κατάσταση του σήματός μας.

Το σύνολο HAPT είναι διακεκομμένο σε σύνολο εκπαίδευσης και ελέγχου, οπότε για να έχουμε συγκρίσιμα αποτελέσματα με μελλοντικές μεθόδους θα διατηρούμε την διαίρεση αυτή.

Αρχικά, θα οπτικοποιήσουμε τα δεδομένα του συνόλου εκπαίδευσης ώστε να εξετάσουμε κατά πόσο οι κλάσεις είναι διαχωρίσιμες μεταξύ τους μετά την εφαρμογή της μεθόδου ΑΚΣ. Εφαρμόζοντας της μέθοδο ΑΚΣ μετασχηματίζουμε το σύνολο εκπαίδευσης, κάνοντας χρήση των δύο κύριων συνιστωσών σε δύο διαστάσεις (Εικόνα 5.1-α) και κάνοντας χρήση της πρώτης κύριας συνιστώσας σε μονοδιάστατο σήμα (Εικόνα 5.1-β). Στην Εικόνα 5.1-β, φαίνεται ξεκάθαρα ότι οι παρατηρήσεις που ανήκουν στην ενεργή κατάσταση μπορούν να ομαδοποιηθούν και να διαχωριστούν από την κατάσταση στην οποία ο άνθρωπος είναι στο πάτωμα. Ωστόσο, είναι εμφανής η παρουσία ακραίων τιμών μέσα στις παρατηρήσεις αυτές, οι οποίες πρέπει να εντοπισθούν και ένας απλός κανόνας κατωφλιού δεν θα μπορούσε να τις εξαρέσει.

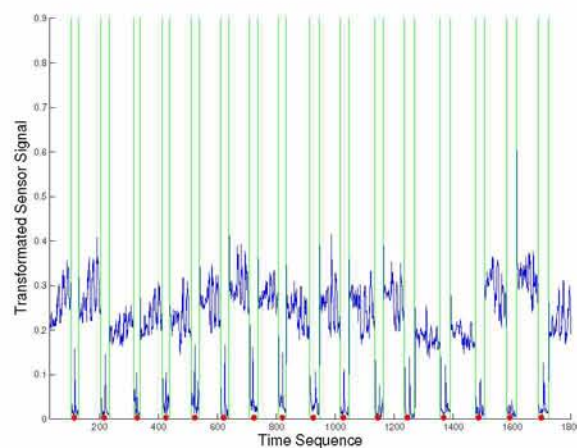
Έτσι, θα εφαρμόσουμε την προτεινόμενη μεθοδολογία χρησιμοποιώντας τις εξής παραμέτρους, για τον αλγόριθμο ΔΑΚΣΣ θα ορίσουμε ως τιμή της παραμέτρου αμνησίας $l = 2$ όπως προτείνουν οι συγγραφείς της μεθόδου [132], ενώ για τον αλγόριθμο ΣΥΣΑΘ η παράμετρος $h = 0.2$. Επιπλέον, θα προχωρήσουμε σε μια διερεύνηση της παραμέτρου h χρησιμοποιώντας διαφορετικές τιμές. Η τιμές των παραμέτρων θ_0 και θ_1 έχουν ορισθεί ως η μέση τιμή των δύο καταστάσεων, από το μετασχηματισμένο σύνολο εκπαίδευσης με την βοήθεια της πρώτης κύριας συνιστώσας. Το αποτέλεσμα της μεθόδου στο σύνολο ελέγχου, οι παρατηρήσεις του οποίου εμφανίζονται ως ροές δεδομένων, παρουσιάζεται στην Εικόνα 5.2. Στην εικόνα υποδηλώνονται, με δύο διαδοχικές κάθετες πράσινες γραμμές ότι τη δεδομένη

¹<https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones>



Σχήμα 5.1: Τα δεδομένα εκπαίδευσης μετασχηματισμένα στο 2-διάστατο χώρο μέσω της μεθόδου ΑΚΣ (αριστερά). Τα δεδομένα εκπαίδευσης στο 1-διάστατο χώρο μέσω της μεθόδου ΑΚΣ. Τα διαφορετικά χρώματα υποδηλώνουν τις τέσσερις καταστάσεις του συνόλου δεδομένων.

χρονική στιγμή ο άνθρωπος ήταν πεσμένος στο πάτωμα, ενώ το κόκκινο σημείο παρουσιάζει την χρονική στιγμή εκείνη που η μέθοδος εντόπισε αλλαγή της κατάστασης του ανθρώπου. Όπως μπορούμε να παρατηρήσουμε η μέθοδος εντοπίζει όλες τις πτώσεις του ανθρώπου, ενώ αξίζει να παρατηρήσουμε ότι ο αλγόριθμος ΣΥΣΑΘ εντοπίζει την κάθε επόμενη κατάσταση πτώσης χωρίς να χρειάζεται να τον επανεκκινήσουμε. Στην Εικόνα 5.3, παρουσιάζουμε μια χρονική περίοδο της συνάρτησης απόφασης (S_n) που υπολογίζει ο ΣΥΣΑΘ και όπως φαίνεται η συνάρτηση αυτή μηδενίζεται πολύ γρήγορα έως ότου ξαναεμφανιστεί η επόμενη κατάσταση πτώσης.

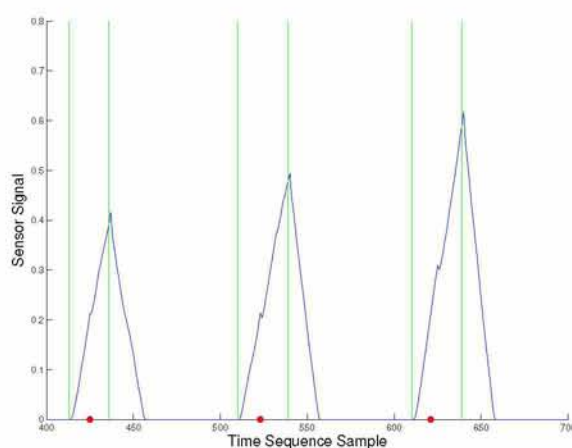


Σχήμα 5.2: Τα δεδομένα ελέγχου μετασχηματισμένα στον 1-διάστατο χώρο του ΔΑΚΣΣ. Τα κόκκινα σημεία του γραφήματος υποδηλώνουν την στιγμή ενεργοποίησης του ΣΥΣΑΘ.

Τέλος, για να εξετάσουμε την αντοχή της μέθοδο ΣΥΣΑΘ, εξετάσαμε διαφορετικές τιμές της παραμέτρου h . Στον Πίνακα 5.1, παρουσιάζουμε το λόγο του επιτυχούς εντοπισμού των καταστάσεων πτώσης προς το πραγματικό πλήθος αυτών των καταστάσεων. Παρατηρούμε ότι η απόδοση του ΣΥΣΑΘ επηρεάζεται αρνητικά και δεν εντοπίζει όλες τις πτώσεις για τιμές μεγαλύτερες του 0.4.

	Παράμετρος h					
	0.02	0.1	0.2	0.3	0.4	0.5
Εντοπισμός της κατάστασης πτώσης	16/16	16/16	16/16	16/16	16/16	11/16

Πίνακας 5.1: Χρήση της προτεινόμενης μεθόδου για τον εντοπισμό των καταστάσεων πτώσης για διαφορετικές τιμές της παραμέτρου h .

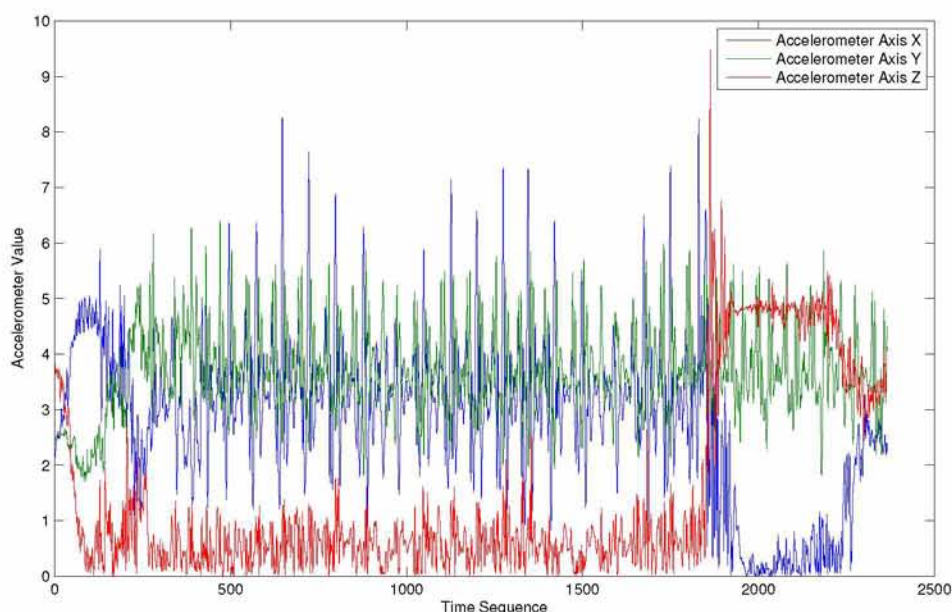


Σχήμα 5.3: Παράδειγμα της συνάρτησης απόφασης του ΣΥΣΑΘ στο σύνολο ελέγχου. Οι πράσινες κάθετες γραμμές υποδηλώνουν την χρονική περίοδο της πραγματικής κατάστασης πτώσης ενώ το κόκκινο σημείο δείχνει την ενεργοποίηση του ΣΥΣΑΘ.

5.3.2 Εφαρμογή σε Δεδομένα τριών Διαστάσεων

Αφού εξετάσαμε την μέθοδο σε ένα σύνολο δεδομένων υψηλής διάστασης επιτυγχάνοντας τον εντοπισμό όλων των πτώσεων του συνόλου έλεγχου, στη συνέχεια θα εξετάσουμε την μέθοδο σε ένα σύνολο δεδομένων τριών διαστάσεων, που παράγεται μέσω ενός επιταχυνσιόμετρου φορητής συσκευής. Η προτεινόμενη μεθοδολογία θα συγκριθεί με μια καθιερωμένη μέθοδο για τον εντοπισμό πτώσεων, που χρησιμοποιεί τον αλγόριθμο ΣΥΣΑΘ ξεχωριστά σε κάθε μια διάσταση του σήματος [125]. Η μέθοδος [125] έχει επιτύχει πολύ υψηλή απόδοση σε σχέση με άλλες μεθόδους για τον εντοπισμό την ανθρώπινης δραστηριότητας χρησιμοποιώντας

κανόνες καταφωλιών σε δεδομένα από επιταχυνσιόμετρα.



Σχήμα 5.4: Η χρονική ακολουθία των τριών αξόνων του επιταχυνσιόμετρου των δεδομένα χαμηλής διάστασης.

Το σύνολο δεδομένων δημιουργήθηκε χρησιμοποιώντας ένα έξυπνο κινητό τηλέφωνο το οποίο τοποθετήθηκε σε τυχαία θέση στην τσέπη του εθελοντή. Ο εθελοντής είτε θα περπατούσε, είτε θα έπεφτε στο πάτωμα, ενώ η συσκευή εγγραφής του επιταχυνσιόμετρου κατέγραφε μια παρατήρηση κάθε 12ms, το οποίο αποτελεί για εμάς ένα στιγμιότυπο. Αφού δημιουργήσαμε ένα σύνολο εκπαίδευσης, θα ελέγξουμε την μεθοδολογία μας στο δείγμα έλεγχου που παρουσιάζεται στην Εικόνα 5.4, όπου υπάρχει μια πτώση κατά το στιγμιότυπο $t = 1800$. Σκοπός μας είναι να ελέγξουμε την προτεινόμενη μεθοδολογία αν μπορεί να εντοπίσει την πτώση και σε πόσο χρόνο μετά από αυτή, στη συνέχεια θα ελέγξουμε την μεθοδολογία [125] αν μπορεί να εντοπίσει την πτώση σε οποιοδήποτε άξονα και αν εφαρμοστεί. Τέλος, θα μετρήσουμε το πλήθος των σφαλμάτων που κάνει, δηλαδή κατά πόσο εντοπίζει μια πτώση η οποία δεν υφίσταται. Ο τελευταίος έλεγχος είναι πολύ σημαντικός για το πρόβλημα που αντιμετωπίζουμε, καθώς ο σκοπός της μεθοδολογίας είναι ο ακριβής εντοπισμός της ανθρώπινης πτώσης έτσι ώστε στην συνέχεια η έξυπνη συσκευή να καλέσει για βοήθεια.

Χρησιμοποιώντας το σύνολο εκπαίδευσης που αναφέραμε στην προηγούμενη παράγραφο, και μετασχηματίζοντάς το με την μέθοδο ΑΚΣ σε μονοδιάστατο σήμα, υπολογίζουμε τις παραμέτρους θ_0 και θ_1 έτσι ώστε να ορίσουμε πότε υπάρχει αλλαγή κατάστασης και κατά συνέπεια πτώση του ανθρώπου. Αντίστοιχα για την μέθοδο [125] οι παράμετροι θ_0 και θ_1 της κάθε διάστασης υπολογίστηκαν από το μονοδιάστατο σήμα των δεδομένων του συνόλου εκπαίδευσης της συγκεκριμένης διάστασης. Επιπλέον, για λόγους πληρότητας εξετάσαμε το σύνολο έλεγχου με την μέθοδο ΑΚΣ, όπου η παρατήρηση του κάθε στιγμιότυπου μετασχηματιζόταν

σε μονοδιάστατο σήμα χρησιμοποιώντας ολόκληρο το δείγμα του συνόλου ελέγχου που είχε εμφανιστεί μέχρι εκείνη την στιγμή.

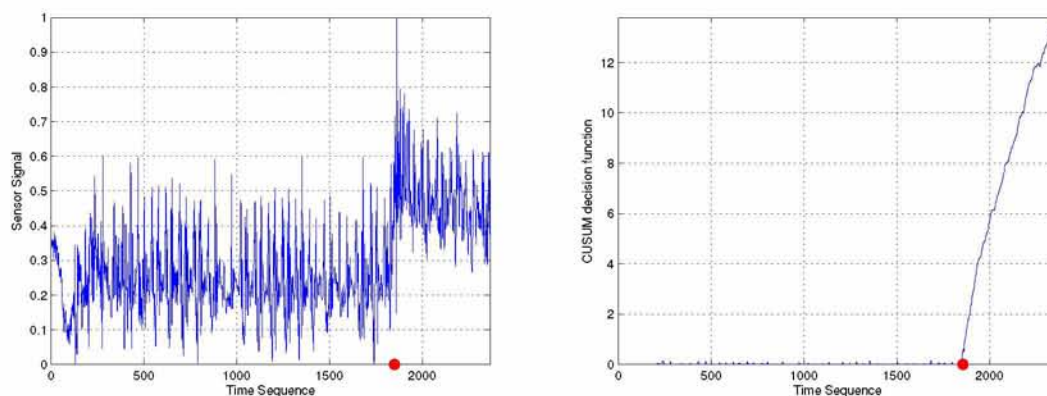
Έτσι έχουμε για τις διαφορετικές μεθόδους που ορίσαμε νωρίτερα στον Πίνακα 5.2, όπου παρουσιάζονται τα αποτελέσματα των μεθόδων, δηλ. κατά πόσο βρίσκουν την πτώση και μετά από πόσα στιγμιότυπα, καθώς επίσης και το πλήθος των λανθασμένων εντοπισμών πτώσης.

h	0.010	0.050	0.100	0.150	0.200	0.500	1.000	1.500	2.000
Άξονας X [125]									
Λανθασμένη Ένδειξη	112	27	9	5	1	1	0	0	0
Καθυστέρηση (στιγμιότυπο)	6	7	28	39	39	54	78	91	101
Άξονας Ψ [125]									
Λανθασμένη Ένδειξη	-	-	-	-	-	-	-	-	-
Καθυστέρηση (στιγμιότυπο)	-	-	-	-	-	-	-	-	-
Άξονας Ζ [125]									
Λανθασμένη Ένδειξη	21	13	9	6	5	2	1	0	0
Καθυστέρηση (στιγμιότυπο)	7	8	8	9	9	15	17	22	26
ΑΚΣ									
Λανθασμένη Ένδειξη	130	47	18	10	3	0	0	0	0
Καθυστέρηση (στιγμιότυπο)	7	40	44	88	92	103	122	141	162
Προτεινόμενη Μέθοδος									
Λανθασμένη Ένδειξη	88	29	11	2	0	0	0	0	0
Καθυστέρηση (στιγμιότυπο)	1	1	3	4	5	7	18	31	45

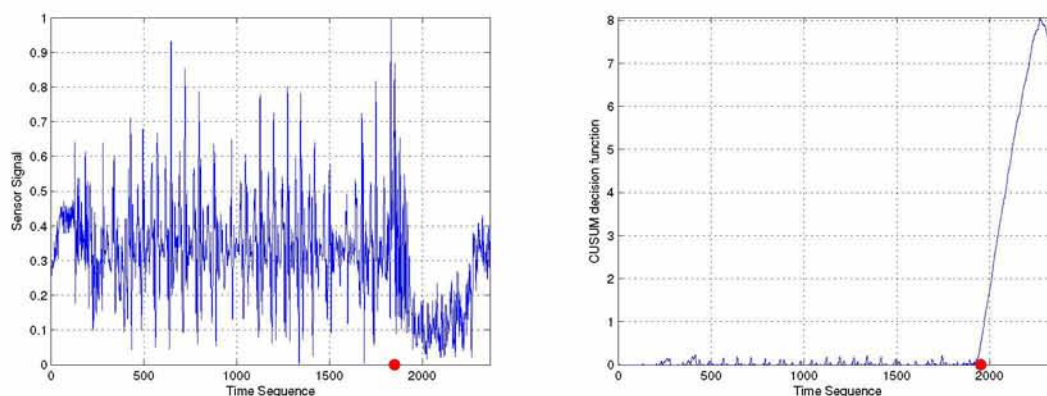
Πίνακας 5.2: Πειραματικά αποτελέσματα της προτεινόμενης μεθόδου στο πρόβλημα αναγνώρισης πτώσης του ανθρώπου και σύγκριση με τις μεθόδους [125], για διαφορετικές τιμές της παραμέτρου h .

Όπως, μπορούμε να παρατηρήσουμε η χρήση του συνδυασμού μείωσης της διάστασης των δεδομένων και εφαρμογή της μεθόδου ΣΥΣΑΘ μπορεί να εντοπίσει πάντα την πτώση, και με κατάλληλη επιλογή της τιμής h να επιτύχει να μηδενίσει τις λανθασμένες ενδείξεις για πτώση του ανθρώπου. Αντίθετα, η χρήση του ΣΥΣΑΘ σε κάθε διάσταση ξεχωριστά, εμφανίζει το πρόβλημα επιλογής της κατάλληλης διάστασης του επιταχυνσιόμετρου, αφού στο συγκεκριμένο πείραμα δεν κατάφερε να εντοπίσει αλλαγή στον άξονα Υ του επιταχυνσιόμετρου. Επιπλέον, η χρήση της μεθόδου ΔΑΚΣΣ σε σύγκριση με την χρήση της απλής ΑΚΣ φαίνεται ότι έχει πολλαπλά οφέλη, καθώς εμφανίζει μικρότερο χρόνο καθυστέρησης εντοπισμού της πτώσης. Τέλος, χρησιμοποιεί πολύ λιγότερη μνήμη για τον υπολογισμό της κύριας συνιστώσας, σε σχέση με τη κλασική προσέγγιση της ΑΚΣ που απαιτεί την ύπαρξη (διατήρηση) ολόκληρου του δείγματος που έχει εμφανιστεί μέχρι την δεδομένη χρονική στιγμή.

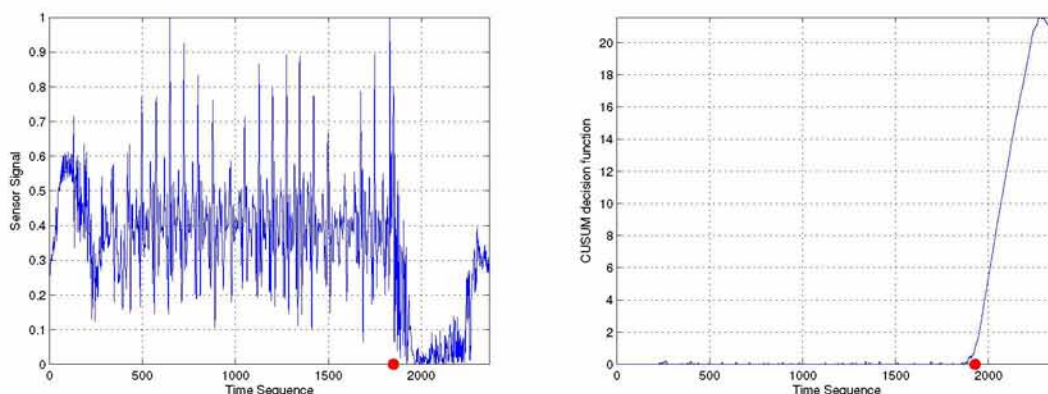
Στις Εικόνες 5.5 και 5.6 παρουσιάζεται το μετασχηματισμένο σήμα των τριών διαστάσεων συνόλου ελέγχου σε μια διάσταση χρησιμοποιώντας την κλασική προσέγγιση ΑΚΣ καθώς και την ΔΑΚΣΣ, όπως και συνάρτηση απόφασης της μεθόδου ΣΥΣΑΘ όπου το κόκκινη σημείο υποδηλώνει το στιγμιότυπο όπου εντόπισε την πτώση. Στις Εικόνες 5.7 έως 5.9 παρουσι-



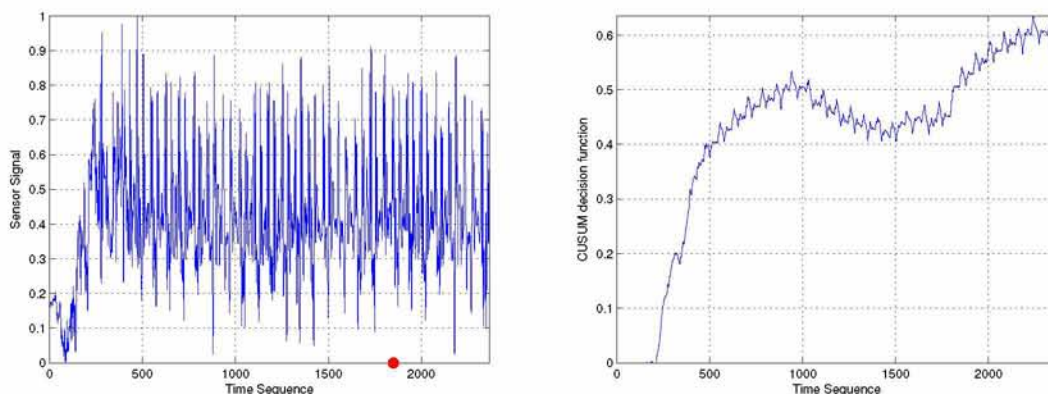
Σχήμα 5.5: Χρονοσειρά των μετασχηματισμένων δεδομένων των τριών διαστάσεων επιταχυνσιόμετρου σε 1-διάστατο μέσω του ΔΑΚΣΣ (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.



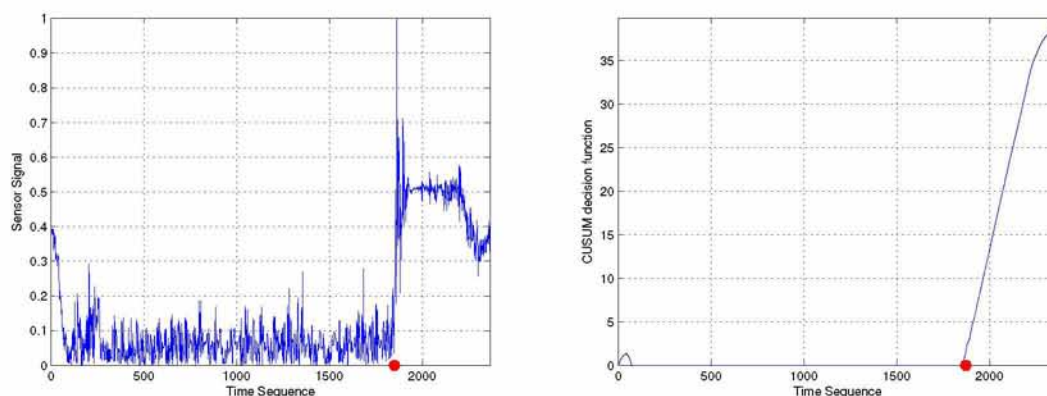
Σχήμα 5.6: Χρονοσειρά των μετασχηματισμένων δεδομένων των τριών διαστάσεων επιταχυνσιόμετρου σε 1-διάστατο μέσω του ΑΚΣ (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.



Σχήμα 5.7: Χρονοσειρά του δεδομένων του άξονα X του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.



Σχήμα 5.8: Χρονοσειρά του δεδομένων του άξονα Y του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος αποτυγχάνει στον εντοπισμό της ανθρώπινης πτώσης.



Σχήμα 5.9: Χρονοσειρά του δεδομένων του άξονα Z του επιταχυνσιόμετρου (αριστερά) και η αντίστοιχη συνάρτηση απόφασης του ΣΥΣΑΘ (δεξιά). Το κόκκινο σημείο υποδεικνύει την πραγματική χρονική στιγμή της πτώσης του ανθρώπου (αριστερά) και την στιγμή ενεργοποίησης του ΣΥΣΑΘ (δεξιά). Η μέθοδος επιτυγχάνει τον εντοπισμό της ανθρώπινης πτώσης.

άζεται το σήμα της κάθε διάστασης του επιταχυνσιόμετρου ξεχωριστά όπως και η συνάρτηση απόφασης της μεθόδου ΣΥΣΑΘ, με το κόκκινο σημείο να υποδηλώνει το στιγμιότυπο που εντοπίστηκε η πτώση. Όπως, μπορούμε να παρατηρήσουμε στην Εικόνα 5.8 θα γίνεται εντοπισμός της πτώσης αφού η συνάρτηση απόφασης λαμβάνει πολύ μεγάλες τιμές από την αρχή χωρίς να μηδενίζει σε καμία χρονική στιγμή τις τιμές της.

5.4 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια μεθοδολογία χαμηλής πολυπλοκότητας για την ενσωμάτωση σε υπολογιστικές συσκευές χαμηλής ισχύος, με σκοπό την πρόβλεψη τυχόν πτώσης του ανθρώπου. Σκοπός του κεφαλαίου είναι η δημιουργία ενός γενικευμένου υπολογιστικού μοντέλου το οποίο δεν θα έχει περιορισμό στο πλήθος των αισθητήρων που θα χρησιμοποιεί (καθώς η ενσωμάτωση περισσότερης ασυσχέτιστης πληροφορία μπορεί να δώσει καλύτερη ακρίβεια στα αποτελέσματα), ενώ θα διατηρεί την ιδιότητά του να δύναται να ενσωματωθεί σε χαμηλής ισχύος συσκευές.

Αυτό επιτυγχάνεται, συνδυάζοντας την χαμηλής πολυπλοκότητας μέθοδο διαδοχικών κυρίων συνιστωσών, μετασχηματίζοντας ένα πολυδιάστατο σήμα σε μονοδιάστατο. Στη συνέχεια, ο αλγόριθμος διαγραμμάτων έλεγχου συσσωρευτικού αθροίσματος, επιλέγοντας κατάλληλες τιμές των παραμέτρων μέσω ενός συνόλου εκπαίδευσης, εντοπίζει αλλαγές στο μονοδιάστατο αυτό σήμα.

Η προτεινόμενη μεθοδολογία εξετάστηκε σε ένα ανοιχτό σύνολο δεδομένων μεγάλης διάστασης από αισθητήρες έξυπνου κινητού τηλεφώνου, για τον εντοπισμό αλλαγής της ανθρώπινης δραστηριότητας. Η προτεινόμενη μεθοδολογία κατάφερε να εντοπίσει κάθε αλλαγή της ανθρώπινης δραστηριότητας. Στη συνέχεια, εξετάσαμε την προτεινόμενη μεθοδολογία σε

δεδομένα τριών διαστάσεων, που ελήφθησαν μέσω του επιταχυνσιόμετρου ενός έξυπνου κινητού τηλεφώνου για τον εντοπισμό της πτώσης του ανθρώπου. Η μεθοδολογία συγκρίθηκε με γνωστή σύγχρονη μεθοδολογία εντοπισμού πτώσεων σε μονοδιάστατα σήματα και φάνηκε να υπερτερεί, καθώς δεν απαιτεί την επιλογή του σήματος στο οποίο θα εφαρμοστεί, αλλά συνδυάζει την πληροφορία όλων των σημάτων σε ένα καινούριο σήμα ως γραμμικό μετασχηματισμό των προηγούμενων.

Μέρος ΙΙΙ

Εντοπισμός Παθολογικών Καταστάσεων σε Ιατρικές Εικόνες

Κεφάλαιο 6

Αναγνώριση και Εντοπισμός Παθολογικών Καταστάσεων στο Γαστρεντερικό Σωλήνα

Στο κεφάλαιο αυτό θα ασχοληθούμε με την αναγνώριση και εντοπισμό μια σειράς διαφορετικών παθολογικών καταστάσεων, που μπορεί να εμφανιστούν στο γαστρεντερικό σύστημα και συγκεκριμένα στο γαστρεντερικό σωλήνα του ανθρώπου, μέσω ευφυών αλγορίθμων της μηχανικής μάθησης. Σκοπός του κεφαλαίου είναι η δημιουργία ενός υπολογιστικού μοντέλου, το οποίο χρησιμοποιώντας εικόνες από το γαστρεντερικό σωλήνα να μπορεί επικουρικά με τον γενικό ιατρό, να αναγνωρίζει την ύπαρξη μια παθολογικής κατάστασης και να υποδεικνύει την περιοχή αυτή μέσα στο γαστρεντερικό σωλήνα.

Το υπολογιστικό σύστημα, για την αναγνώριση της ύπαρξης παθολογικής κατάστασης στην εικόνα του γαστρεντερικού σωλήνα βασίζεται στα ΤΝΔΣ. Η εκπαίδευση του δικτύου γίνεται από εικόνες του γαστρεντερικού σωλήνα χαρακτηρισμένες ως προς το κατά πόσον υπάρχει σε αυτές κάποια παθολογική κατάσταση, “ανωμαλία”, ή όχι. Στη συνέχεια, αξιοποιώντας την πληροφορία που παράγεται στα ενδιάμεσα επίπεδα του δικτύου επιτυγχάνεται ο εντοπισμός της παθολογικής περιοχής. Η νέα πτυχή της προσέγγισης αυτής είναι ότι η μέχρι τώρα μεθοδολογίες εντοπισμού των παθολογικών περιοχών, με μεθόδους μηχανικής μάθησης, απαιτούν εκπαίδευση με αυστηρά χαρακτηρισμένες εικόνες ως προς τις παθογενείς περιοχές.

Το προτεινόμενο σύστημα [45, 64] έχει εφαρμοστεί, εξεταστεί και συγκριθεί με άλλες μεθόδους αναγνώρισης και εντοπισμού παθολογικών καταστάσεων σε εικόνες του γαστρεντερικού συστήματος. Τα αποτελέσματα δείχνουν ότι το προτεινόμενο σύστημα επιτυγχάνει να χαρακτηρίσει και να εντοπίσει παθογενείς καταστάσεις σε πολλά διαφορετικά σύνολα δεδομένων με εικόνες του γαστρεντερικού σωλήνα.

6.1 Εισαγωγή

Η διασύνδεση των συσκευών και αισθητήρων έχει φέρει επανάσταση στον καθημερινό τρόπο ζωής. Καθώς οι συσκευές γίνονται όλο και μικρότερες σε διαστάσεις μπορούν και

ενσωματώνονται στον χώρο, χωρίς να διαταράσσουν και να μεταβάλλουν τις συνήθειες του ανθρώπου. Ο κανόνας αυτός ισχύει και στις ιατρικές συσκευές αφού γίνονται όλο και μικρότερες, ενσωματώνοντας δυνατότητες ασύρματης επικοινωνίας και μεγάλης αυτονομίας. Έτσι οι άλλοτε μεγάλες ενδοσκοπικές συσκευές, που χρησιμοποιούνταν για την παρακολούθηση και καταγραφή του γαστρεντερικού σωλήνα, σήμερα μπορούν να αντικατασταθούν από ασύρματες κάψουλες πολύ μικρού μεγέθους.

Οι κάψουλες αυτές έχουν μέγεθος περίπου 11 × 27 χιλ. και ενσωματώνουν έγχρωμη κάμερα με φακό μικρής εστιακής απόστασης, καθώς επίσης διαθέτουν σύστημα φωτισμού. Η προώθηση της κάψουλας στον γαστρεντερικό σωλήνα επιτυγχάνεται δια του φυσιολογικού εντερικού περισταλτισμού και αποβάλλεται μέσω της φυσιολογικής οδού. Η διάρκεια λειτουργίας τους υπολογίζεται στις 8 – 9 ώρες, όπου μέσω του ασύρματης λειτουργίας που διαθέτουν επιτυγχάνουν την καταγραφή σε μορφή εικόνας και βίντεο του γαστρεντερικού σωλήνα.

Γίνεται εύκολα αντιληπτό ότι αυτή η νέα τάση πραγμάτων στον χώρο του διαγνωστικού ελέγχου του γαστρεντερικού συστήματος, κάνοντας χρήση συσκευών εφοδιασμένων με κάμερες μικρού μεγέθους και ασύρματων λειτουργιών, γίνεται κομμάτι των διασυνδεδεμένων συσκευών και της περιβάλλουσας νοημοσύνης στο χώρο της ιατρικής φροντίδας. Ωστόσο, οι συσκευές δεν μπορούν να διαθέτουν νοημοσύνη χωρίς το κατάλληλο λογισμικό. Έτσι στον τομέα του διαγνωστικού ελέγχου μέσα από εικόνες και βίντεο, για την υποβοήθηση του ιατρικού προσωπικού για την λήψη αποφάσεων ως προς την παθολογική κατάσταση των ασθενών, αναπτύσσονται και παρουσιάζονται έξυπνα “Συστήματα Ιατρικών Αποφάσεων με χρήση Εικόνας” (ΣΙΑΕ), κατάλληλα για την αναγνώριση και τον εντοπισμό διαφορετικών παθήσεων.

Για την παθολογία του γαστρεντερικού σωλήνα, η χρήση ΣΙΑΕ που υποδεικνύουν παθολογικές καταστάσεις αποτελεί ένα σημαντικό εργαλείο στα χέρια των ιατρών, καθώς η εξέταση του βίντεο ενός ασθενούς απαιτεί πολύ χρόνο πλήρους συγκέντρωσης σε αυτό, με τον κίνδυνο πολλές φορές να συμβούν λάθη από τους ίδιους τους ιατρούς. Η ανάπτυξη κατάλληλων έξυπνων ΣΙΑΕ για το γαστρεντερικό σωλήνα παρουσιάζει μεγάλο ενδιαφέρον καθώς είναι ένα δύσκολο πρόβλημα, αφού οι εικόνες που παράγονται κατά την διαδικασία της γαστρεντερικής ενδοσκόπησης είναι αρκετές δεκάδες ή και χιλιάδες.

Τα σύγχρονα συστήματα αυτόματου εντοπισμού παθογενενών καταστάσεων βασίζονται στη χρήση μεγάλων συνόλων δεδομένων από εικόνες που εμφανίζουν διαφορετικές παθολογικές καταστάσεις. Αυτά τα σύνολα δεδομένων χρησιμοποιούνται για να εκπαιδύσουν και να παράξουν υπολογιστικά μοντέλα που επιλύουν το συγκεκριμένο πρόβλημα. Ωστόσο, το σύνολο των ΣΙΑΕ απαιτεί τα σύνολα εκπαίδευσης να είναι αυστηρώς προσδιορισμένα ως προς την περιοχή της παθογένειας που παρουσιάζουν. Αυτή η διαδικασία είναι χρονοβόρα, απαιτητική, με σημαντικό κόστος καθώς εξειδικευμένοι ιατροί πρέπει να χαρακτηρίσουν τις εικόνες ως προς το κάθε εικονοστοιχείο, με τον κίνδυνο πάντα να συμβούν μικρά λάθη. Αυτές τις μεθόδους θα τις ονομάσουμε “Μεθόδους Αυστηρού Χαρακτηρισμού” (ΜΑυΧ). Στον αντίποδα, μέθοδοι που χρησιμοποιούν σύνολα εκπαίδευσης όπου οι εικόνες έχουν ασθενώς χαρακτηριστεί ως προς το περιεχόμενό τους, δηλαδή αν εμφανίζεται κάποια παθολογική κατάσταση ή όχι, θα τις ονομάζουμε ‘Μεθόδους Ασθενούς Χαρακτηρισμού’ (ΜΑσΧ), ενώ τα σύνολα δεδομένων, ως ασθενή χαρακτηρισμένα. Είναι εύκολο να αντιληφθούμε ότι η παραγωγή ασθενούς

χαρακτηρισμού συνόλων δεδομένων είναι πολύ ευκολότερη σε σχέση με αυτή των αυστηρώς χαρακτηρισμένων. Οι μέθοδοι της ΜΑσΧ σε σχέση με τις ΜΑυΧ χρησιμοποιούν λιγότερη αρχική πληροφορία για την παραγωγή του μοντέλου, αφού τα δεδομένα εκπαίδευσης δεν περιέχουν χαρακτηρισμό ανα εικονοστοιχείο, οπότε τα μοντέλα είναι δυσκολότερο να παράγουν υψηλά ποσοστά επιτυχίας στα αποτελέσματά τους συγκριτικά με τις ΜΑυΧ.

Το προτεινόμενο σύστημα ανήκει στην κατηγορία ΜΑσΧ και σκοπό έχει τον εντοπισμό παθογενών καταστάσεων από εικόνες του γαστρεντερικού σωλήνα που έχουν αποκτηθεί μέσω Ασύρματης Ενδοσκοπικής Κάψουλας (ΑΕΚ). Το σύστημα έχει εκπαιδευτεί με ένα σύνολο δεδομένων ασθενούς χαρακτηρισμού, τόσο για την αναγνώριση παθολογιών στις εικόνες αυτές όσο και εντοπισμό των παθολογιών αυτών πάνω στην εικόνα.

6.1.1 Βιβλιογραφική αναφορά Συστημάτων Ιατρικών Αποφάσεων του Γαστρεντερικού Σωλήνα

Στη βιβλιογραφία, τα πρώτα συστήματα ΣΙΑΕ αυτόματου εντοπισμού παθολογικών καταστάσεων του γαστρεντερικού σωλήνα σε ακολουθίες βίντεο εμφανίστηκαν πρώτη φορά στις αρχές του 2000 [74]. Όλο και περισσότερα συστήματα άρχισαν να παρουσιάζονται με σκοπό να μειώσουν το πλήθος των οργανικών βλαβών που μπορεί να μην εντοπιστούν κατά την διάρκεια μιας γαστρεντερικής ενδοσκόπησης [82]. Τα συστήματα επικεντρώνονται στο να αναγνωρίσουν μόνο έναν τύπο ή κάποιους λίγους τύπους γαστρεντερικής παθολογίας [66], όπως πολύποδες [14, 69, 87, 107, 124, 123, 140] ή έλκος και πολύποδες [73], καρκίνο του οισοφάγου [130], κοιλιοκάκι [135], φλεγμονές του εντέρου [45, 131] και αιμορραγίες στο γαστρεντερικό σωλήνα [67, 68, 137]. Πρόσφατες δημοσιεύσεις παρουσιάζουν συστήματα που τείνουν να είναι γενικότερου τύπου, αναγνωρίζοντας ένα μεγάλο σύνολο διαφορετικών παθολογικών καταστάσεων. Τα συστήματα αυτά βασίζονται σε μεθόδους όπως τις Μηχανές Υποστήριξης Διανυσμάτων με Χαλαρή εις Βάθος Κωδικοποίηση (Deep Sparse Support Vector Machine, DSSVM) και τμηματοποίησης με την βοήθεια υπερ-εικονοστοιχείων (Superpixel Segmentation) [22, 23], μεθόδους που βασίζονται στις εύρεση σημαντικών χρωματικών περιοχών σε μια εικόνα συνδυάζοντας τις Μηχανές Υποστήριξης Διανυσμάτων [65], και τέλος σε μεθόδους που κάνουν χρήση των ΤΝΔΣ [117, 143].

Τα ΤΝΔΣ, όπως έχουμε αναφέρει και στην εισαγωγή έχουν χρησιμοποιηθεί για την επίλυση προβλημάτων με εικόνες αντικειμένων του πραγματικού κόσμου, έχουν όμως μεγάλη αποδοχή και στον χώρο της ιατρικής εικόνας. Στη πλειονότητα των περιπτώσεων των ΤΝΔΣ η εκπαίδευση γίνεται χρησιμοποιώντας υποσύνολα των αρχικών εικόνων [14, 107, 124, 123, 117, 143] ενώ λίγες είναι οι περιπτώσεις όπου εκπαιδεύονται με ολόκληρη της εικόνα [140, 45]. Η χρήση υποσυνόλων των εικόνων εφαρμόζεται είτε για αναγνώριση της εικόνας είτε για τον εντοπισμό ανωμαλιών πάνω σε αυτή. Η εκπαίδευση γίνεται συνήθως με τυχαία δειγματοληψία περιοχών της εικόνας όπου είναι γνωστό αν απεικονίζεται ή όχι κάποια παθογενής κατάσταση. Αυτό όμως προϋποθέτει ότι έχει γίνει αυστηρός χαρακτηρισμός των εικόνων.

Επειδή όμως ο αυστηρός χαρακτηρισμός των εικόνων είναι μια επίπονη διαδικασία όπως αναφέραμε νωρίτερα, η χρήση μεθόδων ΤΝΔΣ που κάνουν χρήση συνόλων δεδομένων α-

σθενούς χαρακτηρισμού έχουν σημαντική αξία και θα τις ορίσουμε ως “Ασθενή Νευρωνικά Δίκτυα Συνελίξεων” (ΑΝΔΣ). Στη εργασία [116] χρησιμοποιήθηκε ένα μοντέλο ΑΝΔΣ για αναγνώριση παθογενιών σε εικόνες γαστροσκοπίου από κάψουλα χρησιμοποιώντας εκτός των καναλιών RGB, την παραγόμενη πληροφορία από τους Εσοιανούς και Λαπλασιανούς μετασχηματισμούς των εικόνων. Ενώ ένα σύνθετο μοντέλο από μια αλληλουχία ΤΝΔΣ χρησιμοποιήθηκε για την αναγνώριση οργάνων του γαστρεντερικού σωλήνα [20]. Μια αρχιτεκτονική πανομοιότυπης μορφής με αυτή που χρησιμοποιήσαμε στην εργασία [45], αντικαθιστώντας το δεύτερο επίπεδο του πλήρους διασυνδεδεμένων επιπέδου νευρώνων με Μηχανές Υποστήριξης Διανυσμάτων για τον εντοπισμό αιμορραγίας στο γαστρεντερικό σωλήνα παρουσιάστηκε στην εργασία [67]. Μια ενδιαφέρουσα προσέγγιση ΑΝΔΣ [140] για την αναγνώριση πολύποδων στηρίχθηκε στην εξαγωγή χαρακτηριστικών από ένα προ-εκπαιδευμένο ΤΝΔΣ. Το δίκτυο αυτό έχει προ-εκπαιδευτεί όχι σε ιατρικές εικόνες αλλά σε εικόνες του πραγματικού κόσμου [32]. Το μοντέλο αυτό εξάγει για κάθε εικόνα ένα διάνυσμα των χαρακτηριστικών της, ενώ στη συνέχεια μια Μηχανή Υποστήριξης Διανυσμάτων εκπαιδεύεται με τα διανύσματα αυτά για να πραγματοποιήσει την ταξινόμηση των εικόνων.

Μια άλλη κατηγορία μεθόδων είναι οι Ασθενείς Μέθοδοι Επιβλεπόμενης Μάθησης (Α-MEM) για το πρόβλημα της αναγνώρισης παθογενειών σε εικόνες του γαστρεντερικού σωλήνα, όπου που χρησιμοποιούν την μέθοδο εξαγωγής χαρακτηριστικών “Ομάδων εικονικών Λέξεων” (ΟεΛ), (Bag of visual Words). Οι εργασίες που στηρίζονται στην εξαγωγή χαρακτηριστικών από τις ΟεΛ δημιουργούν το “λεξιλόγιο” από πληροφορίες όπως, το ιστόγραμμα διαφορετικών χρωματικών χώρων για τον εντοπισμό αιμορραγίας [137], του χρωματικού χώρου του CIE-Lab [131] για την αναγνώριση φλεγμονών του εντέρου ή τον συνδυασμό Χαρακτηριστικών Αμετάβλητων στις Διαστάσεις της εικόνας (ΧΑΔ) (Scale-Invariant Feature Transforma SIFT) με ένα ιστόγραμμα τοπικού δυαδικού προτύπου (complete Local Binary Pattern histogram) για τον εντοπισμό πολύποδων [136].

Το βασικό μειονέκτημα των ΑΜΕΜ σε σύγκριση με τις μεθόδους υποσυνόλων είναι η αδυναμία εντοπισμού της παθογένειας λόγω της έλλειψης πληροφορίας κατά την διάρκεια εκπαίδευσης με την περιοχή της παθογένειας πάνω στην εικόνα. Υβριδικές προσεγγίσεις ήμι-Ασθενών Μεθόδων Επιβλεπόμενης Μάθησης έχουν προταθεί για τον λόγο αυτό όπου χρησιμοποιούν έναν συνδυασμό ασθενών χαρακτηρισμένων συνόλων δεδομένων μαζί με υποσύνολα εικόνων για την τμηματοποίηση της παθογένειας πάνω στην εικόνα [97, 99].

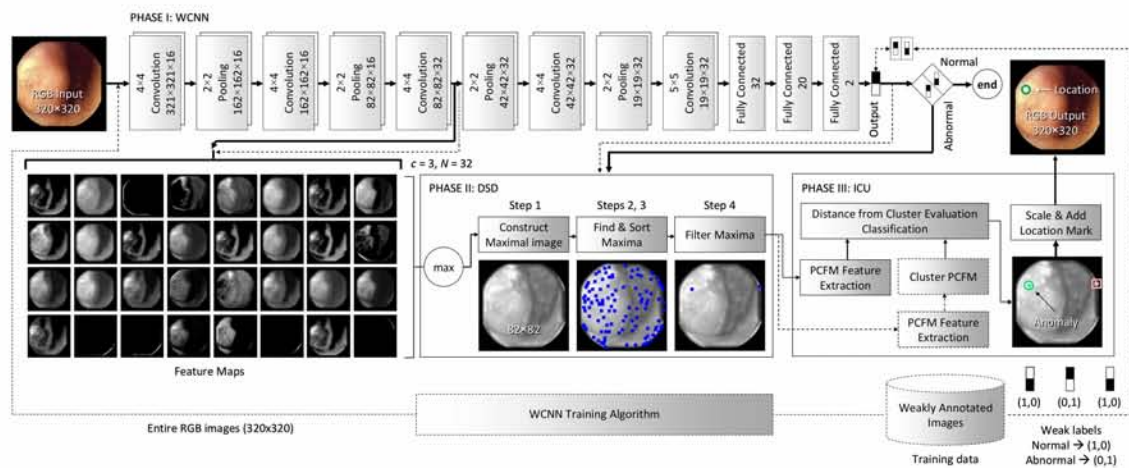
6.2 Μέθοδος Αναγνώρισης και Εντοπισμού Παθολογικών Καταστάσεων στο Γαστρεντερικό Σωλήνα

Όπως έγινε αντιληπτό στην προηγούμενη ενότητα, η χρήση ολοκληρωμένων συστημάτων αναγνώρισης και εντοπισμού παθολογικών καταστάσεων στο γαστρεντερικό σωλήνα με χρήση συνόλων δεδομένων εξ’ ολοκλήρου με ασθενή χαρακτηρισμό είναι επιτακτική. Η προτεινόμενη μεθοδολογία σκοπεύει να καλύψει αυτή την ανάγκη βασιζόμενη στην εκπαίδευση ενός ΑΝΔΣ με ένα σύνολο εκπαίδευσης ασθενές χαρακτηρισμένο, για την αναγνώριση πολλών διαφορετικών μορφών παθογόνων καταστάσεων του γαστρεντερικού σωλήνα και εντοπισμού τους,

κάνοντας χρήση της εσωτερικής πληροφορίας που διαθέτει το υπάρχον εκπαιδευμένο δίκτυο. Η μεθοδολογία διακρίνεται σε τρεις φάσεις:

- Φάση I: Εκπαίδευση ενός ΑΝΔΣ για τον εντοπισμό των εικόνων του γαστρεντερικού σωλήνα που εμφανίζουν παθολογική κατάσταση μέσα από ένα σύνολο πολλών διαφορετικών τύπων παθογένειας.
- Φάση II: Εντοπισμός σημαντικών σημείων πάνω στην εικόνα που υποδεικνύουν μια παθολογική κατάσταση σε αυτή, χρησιμοποιώντας την πληροφορία που υπάρχει σε ένα βαθύ επίπεδο του δικτύου.
- Φάση III: Εξέταση των παραγόμενων σημαντικών σημείων και κατηγοριοποίησή τους ως σημεία παθογούς ή φυσιολογικής κατάστασης.

Στην Εικόνα 6.1 παρουσιάζεται η προτεινόμενη μεθοδολογία σχηματικά. Στη συνέχεια θα δώσουμε αναλυτική περιγραφή της κάθε φάσης ξεχωριστά.



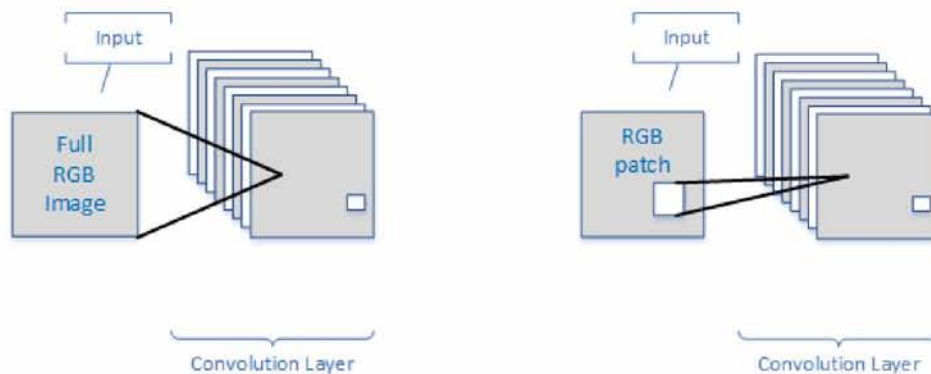
Σχήμα 6.1: Σχηματική αναπαράσταση της προτεινόμενης μεθοδολογίας χαρακτηρισμού και εντοπισμού παθογώνων καταστάσεων σε εικόνες από το γαστρεντερικό σωλήνα.

6.2.1 Ασθενές Νευρωνικό Δίκτυο Συνελίξεων για Εικόνες Γαστρεντερικού Σωλήνα

Η πρώτη φάση της μεθοδολογίας αποτελείται από ένα ΑΝΔΣ όπου εκπαιδεύεται με εικόνες του γαστρεντερικού σωλήνα χωρίς αυστηρό χαρακτηρισμό. Συγκεκριμένα, κάθε εικόνα του συνόλου εκπαίδευσης χαρακτηρίζεται ως προς το αν εμφανίζεται οποιαδήποτε παθογένεια του γαστρεντερικού σωλήνα ή όχι και χρησιμοποιείται για να εκπαιδεύσει το ΤΝΔΣ. Έτσι το δίκτυο είναι ένας δυαδικός ταξινομητής.

Ως είσοδο στο δίκτυο χρησιμοποιούμε έγχρωμες εικόνες διαστάσεων 320×320 ενώ εικόνες μεγαλύτερης διάστασης διαμορφώνονται ώστε να ταιριάζουν με την απαίτηση του δικτύου ως είσοδο. Βασικό χαρακτηριστικό του δικτύου είναι ότι χρησιμοποιεί ολόκληρη την εικόνα για επεξεργασία σε σχέση με άλλες αρχιτεκτονικές ΤΝΔΣ που χρησιμοποιούν υποσύνολα

των αρχικών εικόνων, Εικόνων 6.2. Το ΤΝΔΣ αποτελείται από πέντε επίπεδα συνελίξεων (F) όπου τα τέσσερα πρώτα συνδυάζονται με ένα επίπεδο μέγιστης υποδειγματοληψίας ενώ το πέμπτο ακολουθείται από δύο πλήρως διασυνδεδεμένα επίπεδα νευρώνων και το επίπεδο εξόδου. Τα πρώτα τέσσερα επίπεδα συνελίξεων αποτελούνται από πυρήνες διάστασης 4×4 και εφαρμόζονται ανά δύο εικονοστοιχεία, ενώ το πέμπτο επίπεδο συνελίξεων αποτελείται από πυρήνες διάστασης 5×5 και εφαρμόζεται ανα ένα εικονοστοιχείο, επιπλέον τοποθετείται περιμετρικά από κάθε μητρώο χαρακτηριστικών μια σειρά μηδενικών εικονοστοιχείων. Τα επίπεδα υποδειγματοληψίας δημιουργούνται από πυρήνες διάστασης 2×2 και εφαρμόζονται ανα δύο εικονοστοιχεία. Τα πρώτα δύο επίπεδα συνελίξεων παράγουν 16 μητρώα χαρακτηριστικών ενώ τα υπόλοιπα επίπεδα συνελίξεων 32 μητρώα χαρακτηριστικών. Τέλος, το πρώτο επίπεδο πλήρως διασυνδεδεμένων νευρώνων αποτελείται από 32 νευρώνες ενώ το δεύτερο από 20. Στο αποτέλεσμα κάθε επιπέδου συνελίξεων εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU, ενώ στα πλήρως διασυνδεδεμένα επίπεδα η σιγμοειδής συνάρτηση ενεργοποίησης. Στο επίπεδο εξόδου εφαρμόζεται η συνάρτηση παλινδρόμησης softmax.



Σχήμα 6.2: Εισαγωγή ολόκληρης της έγχρωμης εικόνας στο επίπεδο των συνελίξεων ενός ΝΔΣ (αριστερά). Εισαγωγή τμημάτων (υποσύνολα) από την αρχική έγχρωμη εικόνας στο επίπεδο των συνελίξεων ενός ΝΔΣ (δεξιά).

6.2.2 Εντοπισμός Σημαντικών Σημείων στις Εικόνες Γαστρεντερικού Σωλήνα

Η δεύτερη φάση της μεθοδολογίας αφορά τον εντοπισμό σημείων στις εικόνες πάνω στις περιοχές που απεικονίζονται παθολογικές καταστάσεις, αφού το δίκτυο της φάσης I την έχει χαρακτηρίσει ως παθολογική εικόνα. Αυτό επιτυγχάνεται χρησιμοποιώντας τα μητρώα χαρακτηριστικών που παράγονται από το c επίπεδο του εκπαιδευμένου ΑΝΔΣ κατά την διάρκεια ανάλυσης μια εικόνας. Συγκεκριμένα, από το c επίπεδο συνελίξεων του ΑΝΔΣ, δημιουργούμε μια εικόνα από τα μητρώα χαρακτηριστικών του ως την προβολή της μέγιστης τιμής όλων των μητρώων χαρακτηριστικών $F_j^c, j = 1, 2, \dots, N$, όπου N το πλήθος των μητρώων χαρακτηριστικών του επιπέδου c . Η παραγόμενη εικόνα ονομάζεται Εικόνα Μεγίστων (EM) M^c (Maximal Image) και έχει διάσταση ίση με τη διάσταση των μητρώων χαρακτηριστικών του c επιπέδου, ενώ κάθε εικονοστοιχείο της εικόνας μεγίστων στη χωρική θέση (x, λ) αποτελεί την μέγιστη

τιμή της αντίστοιχης θέσης όλων των μητρώων χαρακτηριστικών του επιπέδου c ,

$$M^c(k, l) = \max\{F_j^c(k, l) | j = 1, 2, \dots, N\}. \quad (6.1)$$

Η παραγωγή της εικόνας στηρίχθηκε στην εργασία [97], σύμφωνα με την οποία τα ΤΝΔΣ στα βαθύτερα επίπεδά τους δημιουργούν μητρώα χαρακτηριστικών τα οποία τείνουν να τονίζουν εκείνες τις περιοχές της εικόνας που χρησιμοποιούνται εν κατακλείδι από το δίκτυο για την ορθή τους ταξινόμηση. Άρα, όσο μεγαλύτερη απόδοση έχει το δίκτυο ως προς την ταξινόμηση των εικόνων, τα μητρώα χαρακτηριστικών θα τονίζουν τις περιοχές αυτές που τις κάνουν να διαφέρουν και τις ομαδοποιούν στην ίδια κλάση, στη συγκεκριμένη περίπτωση τις περιοχές παθογενών καταστάσεων. Έτσι τα μέγιστα των EM αντιστοιχούν στα σημαντικά σημεία των αρχικών εικόνων και υποδεικνύουν περιοχή παθογένειας. Για τον εντοπισμό των σημαντικών σημείων δημιουργήσαμε τον Αλγόριθμο 1 Εύρεσης Σημαντικών Σημείων από Βαθύ Επίπεδο (ΕΣΣΒΕ).

Algorithm 1: Deep Saliency Detection (DSD)

- 1 Construct a maximal image M^c from a deeper WCNN convolutional layer c using Eq. 6.1;
 - 2 Find the local maxima in M^c using a maximum filter and add them in a list L ;
 - 3 Sort L in a descending order of intensities;
 - 4 **for** each element l of L **do**
 - 5 Visit each element k of M^c in the 8-connected neighborhood of l in M^c ;
 - 6 Initialize list Υ by adding l ;
 - 7 **if** $valueOf(k) \in (valueOf(l) - t, valueOf(l)]$ **then**
 - 8 $k' = k$, mark k' as “candidate”, and recursively **do**:
 - 9 **if** $valueOf(k') = valueOf(l)$ **then**
 - 10 | add k' as a “valid maximum” in Υ
 - 11 **end**
 - 12 Visit and mark as “candidate” each 8-connected neighbor k of “candidate” k' until all k have $valueOf(k) < valueOf(l) - t$;
 - 13 **if** k exists in L **then**
 - 14 | remove k from L
 - 15 **end**
 - 16 **end**
 - 17 Produce a salient point by calculating the geometric center of all elements in Υ ;
 - 18 **end**
-

Ο αλγόριθμος στο πρώτο βήμα του υπολογίζει την EM M^c . Στο δεύτερο βήμα βρίσκει όλα τα τοπικά μέγιστα της εικόνας M^c τα οποία θα τα ονομάζουμε l , και τα αποθηκεύει σε μία λίστα L , ταξινομώντας τα σημεία της λίστας σε φθίνουσα σειρά ως προς την τιμή τους

($valueOf(l)$). Υπενθυμίζουμε ότι τα σημεία l είναι εικονοστοιχεία με τιμές από $0...255$. Στο τέταρτο βήμα του αλγορίθμου 1 υπολογίζονται τα μέγιστα της εικόνας M^c που υπερτερούν από τα γειτονικά εικονοστοιχεία κατά t . Η παράμετρος t του αλγορίθμου ορίζεται ως η ανοχή του αλγορίθμου στις διαφοροποιήσεις των γειτονικών εικοστοιχείων σύμφωνα με την τεχνική της υπερχειλήσης [114, 127], μειώνοντας τα πολλά αρχικά μέγιστα. Αυτό επιτυγχάνεται επισκεπτόμενος όλα τα σημεία l της λίστας L , χαρακτηρίζοντας τα σημεία l ως “υποψήφια” (*candidate*) αν η τιμή τους είναι μεταξύ των $valueOf(l) - t$ και $valueOf(l)$. Αυτή η διαδικασία γίνεται σε όλα τα “υποψήφια” σημεία αναδρομικά ωσότου δεν θα υπάρχουν άλλα γειτονικά σημεία που οι τιμές να ανήκουν στο παραπάνω διάστημα τιμών. Αν τώρα κάποιο στοιχείο της λίστας L περάσει από την διαδικασία της επίσκεψης ως γειτονικό, τότε αφαιρείται από την λίστα L ως αδύναμο μέγιστο και δεν θα εξεταστεί αργότερα. Από τα σημεία αυτά τα οποία έχουν χαρακτηριστεί ως “υποψήφια”, αυτά που έχουν τιμή ίση με την τιμή του μεγίστου l της συγκεκριμένης επανάληψης, χαρακτηρίζονται ως “ικανά μέγιστα” και τοποθετούνται στην λίστα Γ . Τα σημαντικά σημεία που παράγει ως έξοδο ο αλγόριθμος παράγονται από τον γεωμετρικό μέσο των “ικανών μεγίστων” της λίστας Γ , σε κάθε επανάληψη του βήματος τέσσερα.

6.2.3 Μέθοδος Ενοποίησης Επαναληπτικής Ομαδοποίησης

Η τελευταία φάση της προτεινόμενης μεθοδολογίας έχει σκοπό τον προσδιορισμό των σημαντικών σημείων που έχουν παραχθεί από την φάση II και υποδεικνύουν παθογενείς περιοχές. Ο αλγόριθμος ΕΣΣΒΕ παράγει ένα σύνολο σημείων που πιθανά ανήκουν σε παθογενή περιοχή της εικόνας αλλά δεν ενσωματώνει καμία πρότερη γνώση στο μηχανισμό του. Έτσι χρειάζεται ένα επιπλέον βήμα (η φάση III) όπου τα παραγόμενα σημεία θα ελεγχθούν και θα ταξινομηθούν ως προς το αν πραγματικά είναι σημεία παθογενούς περιοχής ή όχι.

Για την ταξινόμηση των σημείων θα χρησιμοποιήσουμε την Μέθοδο Ενοποίησης Επαναληπτικής Ομαδοποίησης (MEEO) [64] που βασίζεται στον αλγόριθμο ομαδοποίησης k -μέσων για την ταξινόμηση των σημαντικών σημείων του ΕΣΣΒΕ. Ο MEEO αποτελεί μια μέθοδος ασθενούς επιβλεπόμενης μάθησης, καθώς κατά την διάρκεια της εκπαίδευσης χρησιμοποιεί εικόνες με και χωρίς παθογένεια και ομαδοποιεί τα σημαντικά σημεία τους. Η μέθοδος κατά την εκπαίδευσή της, μέσω μια επαναληπτικής διαδικασίας ενοποιεί τις ομάδες των σημαντικών σημείων σε μια παθογενή εικόνα, έτσι ώστε να είναι όμοιες (μέσω συγκεκριμένων μέτρων ομοιότητας), με τις ομάδες σημαντικών σημείων των φυσιολογικών εικόνων, παράγοντας έναν “ενοποιημένο χώρο απόφασης”. Κατά την διάρκεια του ελέγχου, μια εικόνα η οποία έχει χαρακτηριστεί ως παθογενής από το μοντέλο ΑΝΔΣ της Φάσης I, τα σημαντικά σημεία που έχουν παραχθεί μέσω του αλγορίθμου της Φάσης II, ταξινομούνται ως πιθανά παθογενή ή όχι με την βοήθεια του k -ΚΓ ως προς μια ομάδα ανήκουν στον ενοποιημένο χώρο των ομάδων που έχει δημιουργήσει κατά την εκπαίδευσή του ο MEEO.

Η MEEO δέχεται ως είσοδο μια αναπαράσταση των σημαντικών σημείων ως διάνυσμα χαρακτηριστικών 32 θέσεων, όσα είναι τα μητρώα χαρακτηριστικών του επιπέδου $c = 3$ που χρησιμοποιείται για την παραγωγή τους από τον αλγόριθμο ΕΣΣΒΕ. Οι τιμές του διανύσματος είναι η τιμές που έχει το σημαντικό σημείο στη συγκεκριμένη θέση στα μητρώα χαρακτηριστι-

κών που συντελούν στην δημιουργία της EM.

6.3 Προσομοιώσεις και Αποτελέσματα

Στην ενότητα αυτή θα παρουσιάσουμε συγκριτικά αποτελέσματα της προτεινόμενης μεθοδολογίας έναντι άλλων μεθοδολογιών που παρουσιάστηκαν στην ενότητα 6.1.1. Οι μέθοδοι έχουν δοκιμαστεί σε διαφορετικά σύνολα δεδομένων εικόνων με παθήσεις του γαστρεντερικού σωλήνα.

6.3.1 Σύνολα Δεδομένων Εικόνων του Γαστρεντερικού Σωλήνα

Αρχικά θα αναφέρουμε τα σύνολα δεδομένων που χρησιμοποιήσαμε για να εξετάσουμε τις μεθοδολογίες.

MICCAI Gastroscopy Challenge

Το πρώτο σύνολο δεδομένων αποτελείται από εικόνες που έχουν παραχθεί από ένα τυπικό γαστροσκόπιο και έχουν χρησιμοποιηθεί σε έναν διαγωνισμό¹ στο συνέδριο MICCAI του 2015 [23]. Το πρόβλημα ήταν η εύρεση των εικόνων που παρουσιάζουν μια παθολογική κατάσταση.

Το σύνολο δεδομένων δημιουργήθηκε ύστερα από επιλογή μεταξύ 10,000 εικόνων που παρήχθησαν από 544 υγιείς εθελοντές και 519 εθελοντές με διάφορες βλάβες του γαστρεντερικού σωλήνα όπως καρκίνος, αιμορραγία, έλκος κ.α. Η ανάλυση των εικόνων αρχικά ήταν 768×576 . Από τις αρχικές εικόνες για λόγους προστασίας των προσωπικών δεδομένων αφαιρέθηκε το τμήμα αυτό που περιείχε ευαίσθητα προσωπικά δεδομένα και η τελική διάσταση των εικόνων έγινε 489×409 εικονοστοιχεία.

Για τον διαγωνισμό του MICCAI επιλέχθηκε ένα υποσύνολο των αρχικών 10,000 εικόνων με τις πιο αντιπροσωπευτικές και δημιουργήθηκαν δύο ισορροπημένα σύνολα δεδομένων ως προς τις φυσιολογικές και μη. Το ένα σύνολο δεδομένων χρησιμοποιείται ως σύνολο εκπαίδευσης και αποτελείται από 205 φυσιολογικές εικόνες και 260 εικόνες με παθολογική κατάσταση, ενώ το δεύτερο ως σύνολο ελέγχου με 104 φυσιολογικές και 129 μη. Από εδώ και στο εξής θα αναφερόμαστε σε αυτό το σύνολο ως Σύνολο 1 (Σ1).

KID Dataset

Το δεύτερο σύνολο δεδομένων το οποίο θα το ονομάζουμε Σύνολο² 2 (Σ2) και με το οποίο θα ασχοληθούμε, αποτελείται από εικόνες ενδοσκοπικής κάψουλας. Συγκεκριμένα έχει χρησιμοποιηθεί η κάψουλα MicroCam με ανάλυση διάσταση 360×360 καταγράφοντας ολόκληρο το γαστρεντερικό σωλήνα [79]. Περιλαμβάνει 303 εικόνες που παρουσιάζουν αγγειεκτασία και αίμα στο λεπτό έντερο, 44 εικόνες πολυπόδων και 227 εικόνες με φλεγμονή του εντέρου ενώ,

¹Dataset1: <http://endovissub-abnormal.grand-challenge.org/>

²Dataset 2: <http://is-innovation.eu/kid>

διαθέτει 1,778 φυσιολογικές εικόνες από τον οισοφάγο, το στομάχι, το λεπτό έντερο και το παχύ έντερο.

Ακολουθώντας την διαδικασία διαχωρισμού του συνόλου $\Sigma 1$, για να μπορέσουμε να εξάγουμε ανάλογα συμπεράσματα αλλά και να χρησιμοποιήσουμε τυπικές μετρικές αξιολόγησης των αποτελεσμάτων, θα χρησιμοποιήσουμε ένα υποσύνολο εικόνων του $\Sigma 2$ ώστε να δημιουργήσουμε ισορροπημένα σύνολα εκπαίδευσης και ελέγχου για τις μεθοδολογίες μας. Η επιλογή των εικόνων που θα χρησιμοποιηθούν για τα δύο υποσύνολα έγινε με τυχαία δειγματοληψία χωρίς επανατοποθέτηση και κατασκευάσαμε ένα σύνολο εκπαίδευσης με 429 εικόνες φυσιολογικές και 423 εικόνες με διάφορες παθολογικές καταστάσεις, καθώς επίσης ένα σύνολο ελέγχου με 172 φυσιολογικές εικόνες και 172 μη φυσιολογικές. Το ισορροπημένο σύνολο δεδομένων χαρακτηρίζεται ως Σύνολο 2B ($\Sigma 2B$).

6.3.2 Ανάλυση Παραμέτρων Εκπαίδευσης του ANΔΣ

Σε αυτή την υποενότητα θα εξετάσουμε την απόδοση του μοντέλου ANΔΣ της φάσης I στα Σύνολα 1 και 2B χρησιμοποιώντας διαφορετικούς αλγορίθμους εκπαίδευσης, την μέθοδο της στοχαστικής απότομης κλίσης [16] με συντελεστή ορμής (ΣΑΚΟ, momentum Stochastic Gradient Decent, MSGD) και χωρίς (ΣΑΚ, Stochastic Gradient Decent, SGD) καθώς και τον αλγόριθμο εκπαίδευσης ADAM [13].

Για την αξιολόγηση των αλγορίθμων επαναλάβουμε τα πειράματα 100 φορές και στα δύο σύνολα δεδομένων και τους αξιολογήσαμε ως προς την ακρίβεια των αποτελεσμάτων. Χρησιμοποιήσαμε το τυπικό αυτό μέτρο αξιολόγησης γιατί τα σύνολα ελέγχου και των δύο συνόλων δεδομένων είναι ισορροπημένα ως προς το πλήθος των δειγμάτων σε κάθε κλάση. Ο συντελεστής εκμάθησης που χρησιμοποιήθηκε ήταν 10^{-3} ενώ για τον αριθμό επαναλήψεων ορίσαμε τις 100 'εποχές' δηλαδή 100 πλήρη περάσματα ολόκληρου του συνόλου εκπαίδευσης πριν ολοκληρωθεί η εκπαίδευση του δικτύου. Επιπλέον, για την εκπαίδευση των αλγορίθμων σε κάθε επανάληψη χρησιμοποιήσαμε την μέθοδο των ψευδο-ομάδων εκπαίδευσης με 50 εικόνες ανα ψευδο-ομάδα. Η τιμή της ορμής ορίστηκε ως 0,9, ενώ για τις παραμέτρους του αλγορίθμου ADAM χρησιμοποιήθηκαν οι προτεινόμενες τιμές από τους συγγραφείς [13].

Τα αποτελέσματα των μεθόδων στα Σύνολα 1 και 2B παρουσιάζονται στον Πίνακα 6.1 όπου m αντιστοιχεί στην μέση απόδοση του αλγορίθμου στις 100 επαναλήψεις του πειράματος σ στην διασπορά ενώ M στη μέγιστη τιμή που εμφανίστηκε σε ένα από τα 100 πειράματα. Η χρήση του στοχαστικού αλγορίθμου μέγιστης κλίσης χωρίς την χρήση ορμής παρουσίασε την μέγιστη απόδοση σε σχέση με τους άλλους αλγορίθμους εκπαίδευσης και στα δύο σύνολα δεδομένων.

Επιπλέον, για την παραγωγή ασφαλών συμπερασμάτων διεξάγονται μια σειρά από ελέγχους σημαντικότητας στα αποτελέσματα του Πίνακα 6.1. Αρχικά, χρησιμοποιήθηκε ο στατιστικός έλεγχος Shapiro–Wilk [118] με σκοπό να ελεγχθεί αν τα δεδομένα προέρχονται από μια κανονική κατανομή. Η μηδενική υπόθεση θεωρεί ότι τα δεδομένα προέρχονται από κανονική κατανομή όπου απορρίπτεται στο 5% επίπεδο σημαντικότητας. Στην συνέχεια, διεξάγεται ένας μη-παραμετρικός έλεγχος Friedman του οποίου η μηδενική υπόθεση απορρίπτεται στο

Αλγόριθμος	μ	σ	Max	μ	σ	Max
	$\Sigma 1$	$\Sigma 1$	$\Sigma 1$	$\Sigma 2B$	$\Sigma 2B$	$\Sigma 2B$
ΣΑΚ	0.836	0.048	0.909	0.874	0.007	0.892
ΣΑΚΟ	0.801	0.027	0.867	0.869	0.012	0.898
ADAM	0.820	0.030	0.884	0.863	0.017	0.895

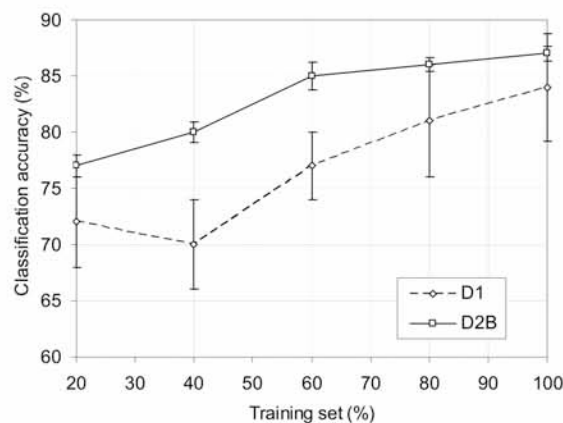
Πίνακας 6.1: Απόδοση του ANΔΣ κάνοντας χρήση διαφορετικών αλγορίθμων εκπαίδευσης στα Σύνολα 1 και 2B ($\Sigma 1$ - $\Sigma 2B$)

5% επίπεδο σημαντικότητας, δείχνοντας έτσι ότι υπάρχουν διαφορές μεταξύ των αποτελεσμάτων των διαφορετικών αλγορίθμων εκπαίδευσης. Τέλος, εφαρμόζεται ο μη-παραμετρικός έλεγχος σημαντικότητας του Wilcoxon [134] σε κάθε ζευγάρι μεθόδων. Η μηδενική υπόθεση σε κάθε έλεγχο σημαντικότητας υποδεικνύει ότι τα δείγματα είναι ανεξάρτητα από πανομοιότυπες συνεχείς κατανομές με ίσες διάμεσους. Στον Πίνακα 6.2 παρουσιάζονται η p-τιμές για κάθε ζευγάρι αλγορίθμων εκπαίδευσης στα δύο σύνολα δεδομένων. Οι p-τιμές που είναι μεγαλύτερες από 5×10^{-2} δεν απορρίπτονται την μηδενική υπόθεση και κατ' επέκταση η διαφορά των αποτελεσμάτων δεν είναι σημαντικές. Στη συγκεκριμένη περίπτωση βλέπουμε ότι η μηδενική υπόθεση απορρίπτεται πάντα, άρα οι διαφορές των αποτελεσμάτων είναι στατιστικά σημαντικές και ο στοχαστικός αλγόριθμος της απότομης κλίσης χωρίς την παράμετρο της ορμής έχει στατιστικά σημαντικά καλύτερη απόδοση από τους υπολοίπους.

	ΣΑΚ		ΣΑΚΟ		ADAM	
	$\Sigma 1$	$\Sigma 1$	$\Sigma 1$	$\Sigma 2B$	$\Sigma 2B$	$\Sigma 2B$
ΣΑΚ	-	-	0.0000	0.0301	0.0057	0.0381
ΣΑΚΟ	-	-	-	-	0.0019	0.0000
ADAM	-	-	-	-	-	-

Πίνακας 6.2: Οι p-τιμές του ελέγχου σημαντικότητας Wilcoxon στα αποτελέσματα των διαφορετικών αλγορίθμων εκπαίδευσης του ANΔΣ στα Σύνολα 1 και 2B ($\Sigma 1$ - $\Sigma 2B$)

Ολοκληρώνοντας τον έλεγχο των αλγορίθμων εκπαίδευσης, ελέγχεται η απόδοση του ΣΑΚ σε εκπαίδευση με μικρότερο σύνολο δεδομένων. Έτσι το αρχικό σύνολο εκπαίδευσης του Σύνολου 1 και 2B υποδειγματοληπτείται σε υποσύνολα της τάξης του 20% έως 80% του αρχικού πλήθους δειγμάτων-εικόνων. Η δειγματοληψία έγινε χρησιμοποιώντας ένα σχήμα ομοιόμορφης κανονικής κατανομής όπου διαιρείται η ισοδυναμία μεταξύ των δύο κλάσεων. Στην Εικόνα 6.3 παρουσιάζεται η απόδοση της ταξινόμησης ως προς το μέγεθος του συνόλου εκπαίδευσης, με την απόδοση του αλγορίθμου εκπαίδευσης να μειώνεται σημαντικά όταν το μέγεθος του συνόλου εκπαίδευσης μειώνεται κάτω από το 60% του αρχικού μεγέθους.



Σχήμα 6.3: Το ποσοστό επιτυχίας του μοντέλου ΑΝΔΣ με τον αλγόριθμο εκπαίδευσης ΣΑΚ σε διαφορετικού πλήθους σύνολα εκπαίδευσης.

6.3.3 Σύγκριση ΑΝΔΣ με ΤΝΔΣ Δεδομένων Αυστηρού Προσδιορισμού

Στην υποενότητα αυτή θα συγκρίνουμε την προτεινόμενη μεθοδολογία από την Φάση I με μεθόδους ΤΝΔΣ που κάνουν χρήση αυστηρής επισήμανσης στις εικόνες. Θα χρησιμοποιήσουμε την βασική εκδοχή των ΤΝΔΣ όπου θα εξάγουμε υποσύνολα εικόνων που είτε θα περιέχουν εξολοκλήρου περιοχές χωρίς να εμφανίζουν καμία παθογενή κατάσταση είτε περιοχές πλήρεις παθογένειας, ονομάζοντας τη μέθοδο αυτή “βασικά Τεχνητά Νευρωνικά Δίκτυα Συνελίξεων” (β -ΤΝΔΣ). Η δεύτερη μεθοδολογία βασίζεται στη γνωστή εργασία των Sekuboyina [117], όπου οι εικόνες μετασχηματίζονται στο χρωματικό χώρο CIE-Lab ενώ χρησιμοποιείται ένα παράθυρο διάστασης 36×36 εικονοστοιχείων για την επιλογή των υπο-εικόνων.

Για την αξιολόγηση των μεθόδων δημιουργήθηκαν ισοκατανομημένα σύνολα δεδομένων ως προς τις κλάσεις, τόσο για το επίπεδο του συνόλου εκπαίδευσης όσο και για το σύνολο ελέγχου. Συγκεκριμένα, το Σ1 αποτελείται από 6,000 υπο-εικόνες για κάθε κλάση ενώ το Σ2 από 4,500 υπο-εικόνες για την κάθε κλάση. Από αυτές τα 2/3 των εικόνων χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων ενώ οι υπόλοιπες για τον έλεγχο τους.

Πριν προχωρήσουμε με την παρουσίαση των αποτελεσμάτων θα περιγράψουμε την αρχιτεκτονική των μοντέλων. Τα β -ΤΝΔΣ χρησιμοποιήθηκαν με δύο διαφορετικού μεγέθους υπο-εικόνες ως είσοδο αλλά με την ίδια βασική αρχιτεκτονική. Η αρχιτεκτονική αποτελείται από τρία επίπεδα συνελίξεων (F) όπου τα δύο πρώτα συνδυάζονται με ένα επίπεδο μέγιστης υποδειγματοληψίας, ενώ το τρίτο ακολουθείται από δύο πλήρως διασυνδεδεμένα επίπεδα νευρώνων και το επίπεδο εξόδου. Τα πρώτα δύο επίπεδα συνελίξεων αποτελούνται από 16 μητρώα χαρακτηριστικών με πυρήνες διάστασης 5×5 , ενώ συνδυάζονται με 16 επίπεδα μέγιστης υποδειγματοληψίας. Το τρίτο επίπεδο συνελίξεων αποτελείται 32 μητρώα χαρακτηριστικών και πυρήνες διάστασης 5×5 , ενώ μετά από κάθε επίπεδο συνελίξεων εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU. Τέλος, το πρώτο επίπεδο πλήρων διασυνδεδεμένων νευρώνων αποτελείται από 32 νευρώνες ενώ το δεύτερο από 20. Για μπορέσουμε να χρησιμοποιήσουμε τις δύο

διαφορετικού μεγέθους υπο-εικόνες ως είσοδο, χρησιμοποιήσαμε διαφορετικές τιμές περιθωρίων και βήματος στα δεδομένα εισόδου που εφαρμόζονται οι πυρήνες για την συνέλιξη και την παραγωγή του πρώτου μητρώου χαρακτηριστικών. Συγκεκριμένα, ορίζουμε ως β -TN Δ Σ_1 το δίκτυο που χρησιμοποιεί ως είσοδο υπο-εικόνες διάστασης 64×64 με τιμές περιθωρίου 1 και βήμα 1 ενώ ως β -TN Δ Σ_2 ορίζουμε το δίκτυο που χρησιμοποιεί ως είσοδο υπο-εικόνες διάστασης 32×32 με τιμές περιθωρίου 2 και βήμα 3. Για την μεθοδολογία των Sekuboyina χρησιμοποιείται μια αρχιτεκτονική με τρία επίπεδα συνέλιξεων με διάσταση πυρήνων 3×3 και τριών μητρώων χαρακτηριστικών στο κάθε επίπεδο. Το δεύτερο και το τρίτο επίπεδο συνέλιξεων ακολουθείται από ένα επίπεδο μέγιστης υποδειγματοληψίας, ενώ το δίκτυο συμπληρώνεται από έναν συνδυασμό πλήρους διασυνδεδεμένου δικτύου νευρώνων $75 - 10 - 2$ με δύο νευρώνες ως έξοδο.

Τα αποτελέσματα των μεθόδων παρουσιάζονται στον πίνακα 6.3 όπου η απόδοση των β -TN Δ Σ εμφανίζουν μεγαλύτερη επιτυχία και στα δύο σύνολα δεδομένων έναντι της μεθόδου των Sekuboyina, με την μέθοδο β -TN Δ Σ_1 να έχει την μέγιστη επιτυχία. Ωστόσο όμως, ο πίνακας 6.3 εξετάζει την επιτυχία των μεθόδων σε επίπεδο υπο-εικόνων και όχι σε ολόκληρη την εικόνα. Για να έχουμε μια αντιπροσωπευτική σύγκριση με τις μεθόδους AN Δ Σ οι οποίες αποφαίνονται σε ολόκληρη την εικόνα ως προς το αν παρουσιάζει παθολογεία ή όχι ορίζουμε τον εξής κανόνα: μια εικόνα χαρακτηρίζεται ως φυσιολογική όταν όλες οι υπο-εικόνες που την απαρτίζουν χαρακτηρίζονται από την μέθοδο ως φυσιολογικές, ενώ παθολογής αν έστω και μια υπο-εικόνα χαρακτηρίζεται από την μέθοδο ως παθολογής.

	β -TN Δ Σ_1		β -TN Δ Σ_2		Sekuboyina	
	$\Sigma 1$	$\Sigma 2B$	$\Sigma 1$	$\Sigma 2B$	$\Sigma 1$	$\Sigma 2B$
Επιτυχία	0.839	0.846	0.821	0.830	0.804	0.824
Ευαισθησία	0.872	0.908	0.852	0.840	0.817	0.840
Ακρίβεια	0.807	0.770	0.788	0.821	0.793	0.804

Πίνακας 6.3: Επιδόσεις TN Δ Σ αυστηρής επισήμανσης στα $\Sigma 1$ και $\Sigma 2B$.

Στον πίνακα 6.4 παρουσιάζονται τα αποτελέσματα των μεθόδων για τον τελικό χαρακτηρισμό των εικόνων ως προς το αν εμφανίζουν παθολογεία ή όχι, χρησιμοποιώντας τον προηγούμενο κανόνα. Χρησιμοποιώντας την αρχιτεκτονική β -TN Δ Σ_1 , μεταξύ των δύο β -TN Δ Σ αρχιτεκτονικών, καθώς πέτυχε υψηλότερη απόδοση στην ταξινόμηση των υπο-εικόνων και τη μεθοδολογία των Sekuboyina βλέπουμε ότι η β -TN Δ Σ_1 συνεχίζει να έχει υψηλή επιτυχία και σε επίπεδο χαρακτηρισμού ολόκληρης της εικόνας. Ωστόσο, η προτεινόμενη μεθοδολογία ασθενούς TN Δ Σ από την Φάση I εμφανίζει τα μεγαλύτερα ποσοστά επιτυχίας σε σύγκριση με τις άλλες δύο μεθόδους TN Δ Σ αυστηρού χαρακτηρισμού.

6.3.4 Σύγκριση Διαφορετικών μεθοδολογιών AN Δ Σ

Στην υποενοότητα αυτή θα συγκρίνουμε την προτεινόμενη μεθοδολογία από την Φάση I με άλλες μεθόδους Ασθενούς Εκπαίδευσης που κάνουν χρήση TN Δ Σ και συγκεκριμένα με

	ΑΝΔΣ		β-TNΔΣ ₁		Sekuboyina	
	Σ1	Σ2B	Σ1	Σ2B	Σ1	Σ2B
Επιτυχία	0.909	0.892	0.717	0.541	0.619	0.517
Ευαισθησία	0.930	0.924	0.961	0.478	0.957	0.952
Ακρίβεια	0.885	0.858	0.413	0.610	0.251	0.091

Πίνακας 6.4: Επιδόσεις προτεινόμενου μοντέλου ΑΝΔΣ και TNΔΣ αυστηρής επισημάνσης στα Σ1 και Σ2B σε ολόκληρη την εξεταζόμενη εικόνα.

την μεθοδολογία των Zhang [140], Jia και Meng [67]. Ακόμα, η προτεινόμενη μεθοδολογία θα συγκριθεί και με μεθόδους που βασίζονται σε ΟεΛ, συγκεκριμένα την μεθοδολογία των Yuan [136] και Vasilakaki [131], όπου αφού εξάγουν χαρακτηριστικά από τις εικόνες, κάνουν χρήση μηχανών υποστήριξης διανυσμάτων για την εκπαίδευση και ταξινόμησή τους.

Στην υποενότητα 6.3.2 όπου εξετάσαμε την προτεινόμενη μεθοδολογία ΑΝΔΣ είχαμε χρησιμοποιήσει το Σ2B ώστε να έχουμε ισορροπημένο σύνολο δεδομένων για την επιλογή των κατάλληλων παραμέτρων. Τώρα όμως όλες οι μέθοδοι θα εξεταστούν στο πλήρες σύνολο δεδομένων Σ2 και Σ1. Η αξιολόγηση των μεθόδων θα γίνει με διασταυρωμένη επικύρωση 10-θέσεων σε ολόκληρα τα σύνολα δεδομένων, έτσι ώστε να επιτύχουμε μια αμερόληπτη αξιολόγηση των μεθόδων ως προς την επιλογή του συνόλου εκπαίδευσης και ελέγχου. Όλες οι εικόνες του συνόλου δεδομένων θα διαμεριστούν τυχαία σε 10 τμήματα όπου τα 9 θα χρησιμοποιηθούν για εκπαίδευση ενώ το δέκατο για έλεγχο μέχρις ότου όλα τα τμήματα να χρησιμοποιηθούν στην διαδικασία ελέγχου. Για την αξιολόγηση των μεθόδων θα χρησιμοποιήσουμε εκτός των προηγούμενων μετρικών και τις καμπύλες Receiver Operation Characteristics (ROC) καθώς και την περιοχή ενδιαφέροντος κάτω από τις καμπύλες ROC (AUC), αφού τα δεδομένα του συνόλου ελέγχου θα είναι πολύ πιθανόν μη ισορροπημένα μεταξύ των δύο κλάσεων και η μετρική αυτή έχει ανοχή σε αυτού του τύπου δεδομένα [6].

	ΑΝΔΣ		Zhang		Jia και Meng		Yuan		Vasilakakis	
	Σ1	Σ2	Σ1	Σ2	Σ1	Σ2	Σ1	Σ2	Σ1	Σ2
AUC	0.963	0.814	0.951	0.773	0.902	0.705	0.940	0.709	0.946	0.802
Επιτυχία	0.899	0.775	0.851	0.760	0.827	0.690	0.867	0.696	0.892	0.768
Ευαισθησία	0.907	0.362	0.930	0.537	0.806	0.602	0.876	0.432	0.911	0.454
Ακρίβεια	0.882	0.913	0.779	0.836	0.857	0.785	0.854	0.820	0.872	0.886

Πίνακας 6.5: Επιδόσεις μοντέλων ΑΝΔΣ στα Σ1 και Σ2.

Στον πίνακα 6.5, παρουσιάζονται τα αποτελέσματα των μεθόδων. Συνολικά βλέπουμε ότι η προτεινόμενη μεθοδολογία ΑΝΔΣ παρουσιάζει την υψηλότερη απόδοση ως προς τα μέτρα AUC και Επιτυχίας σε σχέση με τις υπόλοιπες μεθόδους. Η προτεινόμενη μεθοδολογία υπερσχύει σημαντικά έναντι της μεθοδολογίας του Yuan στο Σ2 ενώ είναι ελαφρώς καλύτερη έναντι του Zhang στο Σ1 και της μεθοδολογίας του Vasilakakis για το Σ2.

Ωστόσο, σημαντικό πλεονέκτημα των μεθοδολογιών που βασίζονται στα TNΔΣ, έναντι

των μεθόδων ΟεΛ ότι δεν εξαρτώνται από άλλη μέθοδο για την παραγωγή των χαρακτηριστικών, αλλά παράγουν οι ίδιες οι μέθοδοι τα χαρακτηριστικά αυτά που διαχωρίζουν βέλτιστα τις δύο κατηγορίες. Επιπλέον, θα θέλαμε να τονίσουμε ότι η μέθοδος DSSVM [23] που εξετάσασε το Σ1 εμφάνισε απόδοση ως προς το μέτρο AUC ίση με 0.898, τιμή χαμηλότερη από αυτή της προτεινόμενης μεθοδολογίας.

Τέλος, για να ενισχύσουμε την απόδοση της προτεινόμενης μεθοδολογίας έναντι των υπολοίπων, δημιουργήσαμε ένα ακόμα σύνολο δεδομένων το οποίο το ονομάζουμε Σύνολο 3 (Σ3) και αποτελείται από την ένωση των Σ1 και Σ2. Το Σ3 αποτελείται από 3,050 εικόνες (698 εικόνες από το Σ1 και 2,352 εικόνες από το Σ2) οι οποίες διαφέρουν μεταξύ τους τόσο στην αρχική ανάλυση όσο και στο εύρος του οπτικού πεδίου με το οποίο απεικονίζεται ο γαστρεντερικός σωλήνας. Το σύνολο αυτό το αξιολογήσαμε όπως και τα προηγούμενα κάνοντας χρήση της διασταυρωμένης επικύρωσης 10-θέσεων. Η προτεινόμενη μεθοδολογία παρουσίασε την υψηλότερη τιμή ως προς την μετρική AUC φτάνοντας το 0.861. Η μεθοδολογία των Jia και Meng πέτυχε 0.846, των Zhang 0.803, των Yuan 0.681 και των Vasilakakis 0.840.

6.3.5 Εύρεση Σημαντικών Σημείων και Αξιολόγηση των Μεθόδων

Εκτός της αναγνώρισης των εικόνων αυτών που εμφανίζουν μια παθολογική κατάσταση, ιδιαίτερη σημασία έχει ο εντοπισμός των περιοχών παθολογείας κάνοντας χρήση σημείων-δεικτών, τα οποία επισημαίνουν μια πιθανή παθολογία πάνω σε μια περιοχή της εικόνας. Συνεχίζοντας από την προηγούμενη ενότητα, στις εικόνες που έχουν χαρακτηριστεί από τη Φάση I της μεθοδολογίας μας ως παθολογικές, θα εξαχθούν (σημαντικά) σημεία τα οποία θα υποδεικνύουν την περιοχή της παθολογείας. Στην Φάση II της προτεινόμενης μεθοδολογίας, χρησιμοποιείται ο αλγόριθμος ΕΣΣΒΕ για την εύρεση των σημαντικών σημείων που εμφανίζουν παθολογία στις εικόνες γαστρεντερικού σωλήνα. Για τον αλγόριθμο απαιτείται ένα εκπαιδευμένο ΑΝΔΣ έτσι ώστε να γίνει εξαγωγή της πληροφορίας για την εύρεση των σημαντικών σημείων από ένα βαθύτερο επίπεδο του. Συγκεκριμένα, μέσα από πειραματική μελέτη διαλέξαμε το τρίτο επίπεδο συνελίξεων ($c=3$) για τον σκοπό αυτό, επιλέγοντας το μοντέλο ΑΝΔΣ με τη μεγαλύτερη τιμή AUC για το κάθε διαφορετικό σύνολο δεδομένων που παρουσιάστηκε στην προηγούμενη ενότητα. Η επιλογή αυτή μπορεί να δικαιολογηθεί για δύο λόγους πρώτον, τα βαθύτερα επίπεδα των ΤΝΔΣ εμφανίζουν (κωδικοποιούν) τοπικά χαρακτηριστικά των εικόνων σε μεγαλύτερη ακρίβεια σε σχέση με τα μητρώα των πρώτων επιπέδων, σύμφωνα και με τα συμπεράσματα από την εργασία [139], ενώ επιπλέον, το μέγεθος των μητρώων χαρακτηριστικών πέραν του επιπέδου τρία γίνεται πολύ μικρό (συγκεκριμένα $\leq 42 \times 42$), κάνοντας την εύρεση παθολογειών σε τόσο μικρές περιοχές μη αποτελεσματικό καθώς θα πρέπει να κάνουμε αναγωγή της θέσης της ανωμαλίας από αυτή την περιοχή στην πραγματική περιοχή της εικόνας. Η αβεβαιότητα που θα προκύψει από την θέση της παθολογείας στο πολύ μικρών διαστάσεων μητρώο χαρακτηριστικών στις πραγματικές διαστάσεις την εικόνας (320×320), ενισχύουν την επιλογή των μητρώων χαρακτηριστικών του τρίτου επιπέδου. Ακόμα, βασιζόμενοι στην εργασία [56] χρησιμοποιήθηκε μια ποσοτικοποίηση της συσχέτισης μεταξύ των επιπέδων του ΑΝΔΣ, έτσι ώστε να θεμελιώσουμε την επιλογή του συγκεκριμένου επιπέδου. Για τα Σ1

και Σ2 υπολογίσαμε την συσχέτιση κατά Pearson μεταξύ των τιμών των εικονοστοιχείων των EM που παράγονται από τα διαφορετικά επίπεδα συνελίξεων, με τις τιμές των εικονοστοιχείων των περιοχών παθογένειας των αντίστοιχων εικόνων. Για τον υπολογισμό αυτό οι διαστάσεις των αρχικών εικόνων τροποποιήθηκαν έτσι ώστε να έχουν ίδιο μέγεθος με τη EM του κάθε επιπέδου. Οι τιμές συσχέτισης μεταξύ των EM των πέντε επιπέδων συνελίξεων από το πρώτο προς το πέμπτο επίπεδο ήταν 0, 14, 0, 05, 0, 16 0, 13 και 0, 04, αντίστοιχα. Η μεγαλύτερη τιμή παρατηρήθηκε στο τρίτο επίπεδο συνελίξεων θεμελιώνοντας την επιλογή του συγκεκριμένου επιπέδου.

	ΕΣΣΒΕ		CSD [65]	
	Σ1	Σ2	Σ1	Σ2
Σημεία στην Παθογενή Περιοχή	1 – 12	1 – 7	2 – 288	1 – 478
Συνολικά Σημεία	1 – 18	1 – 10	195 – 349	548 – 683

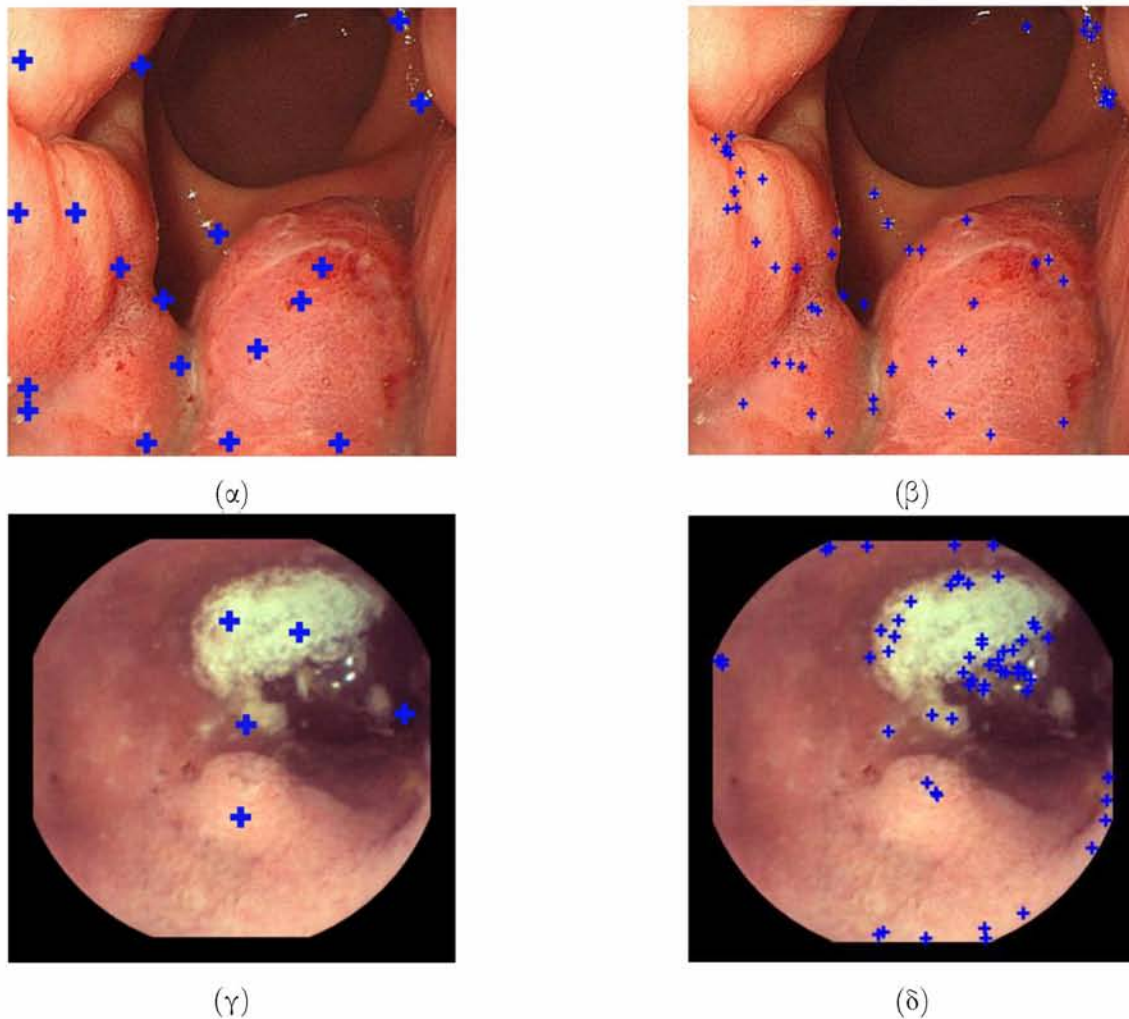
Πίνακας 6.6: Τα αποτελέσματα των αλγορίθμων εύρεσης σημαντικών σημείων στα Σ1 και Σ2. Για το κάθε σύνολο δεδομένων παρουσιάζεται το ελάχιστο και μέγιστο πλήθος σημείων που παράγει ο αλγόριθμος σε μια εικόνα του συνόλου.

Συνεχίζοντας, εξετάζεται η ικανότητα του ΕΣΣΒΕ στην εύρεση σημαντικών σημείων, ελέγχοντάς τον σε κάθε σύνολο ελέγχου της διασταυρωμένης επικύρωσης 10-θέσεων της Ενότητας 6.3.4 των εικόνων, που ορθά ο ANΔΣ χαρακτήρισε ως παθογενείς. Ο προτεινόμενος αλγόριθμος συγκρίνεται με τον αλγόριθμο Color Saliency Detector (CSD) [65]. Η παράμετρος ανοχής του ΕΣΣΒΕ (t) ορίζεται ευρετικά ως 120 ενώ οι παράμετροι του αλγορίθμου CSD σύμφωνα με τους συγγραφείς του [65].

Στον Πίνακα 6.6 παρουσιάζονται τα αποτελέσματα των μεθόδων εύρεσης σημαντικών σημείων για κάθε εικόνα. Συγκεκριμένα, παρουσιάζεται το ελάχιστο και μέγιστο πλήθος σημείων που μπορεί να εμφανιστεί σε μια εικόνα στα διαφορετικά σύνολα δεδομένων στις περιοχές που έχουν σημειωθεί από τους ειδικούς ως παθογενή. Ακόμα, παρουσιάζεται το ελάχιστο και μέγιστο των σημείων που παράγει ο κάθε αλγόριθμος συνολικά σε ολόκληρη την εικόνα, δηλαδή σημεία που έχουν εμφανιστεί και εκτός της παθογενούς περιοχής. Είναι εμφανές ότι ο αλγόριθμος ΕΣΣΒΕ παράγει σημαντικά λιγότερα σημεία σε σχέση με τον CSD. Για παράδειγμα στο Σ1 ο ΕΣΣΒΕ παράγει το πολύ 12 σημεία εντός της παθογενούς περιοχής ενώ ο CSD 288. Η διαφορά αυτή είναι μεγαλύτερη στο Σ2. Πρέπει να τονίσουμε ότι για το συγκεκριμένο πρόβλημα δύο παράγοντες είναι πολύ σημαντικοί, ο εντοπισμός των περιοχών που εμφανίζεται η παθογένεια έτσι ώστε να δώσει προσοχή ο ειδικός ιατρός αλλά και το πλήθος των σημείων, καθώς δεν πρέπει να είναι πάρα πολλά και σε λάθος περιοχές, ώστε να καθυστερούν τον ειδικό με τον έλεγχο όλων των περιοχών που υποδεικνύονται από το σύστημα.

Στην Εικόνα 6.4 γίνεται μια οπτική παρουσίαση των σημείων που παράγει ο προτεινόμενος αλγόριθμος έναντι του CSD σε μία εικόνα από το κάθε σύνολο δεδομένων. Στις Εικόνες 6.4-α και 6.4-β βλέπουμε πως ο προτεινόμενος αλγόριθμος παράγει ένα σημαντικά μικρότερο πλήθος σημείων έναντι του πλήθους των σημείων του CSD που παρουσιάζονται στις άλλες δύο εικόνες.

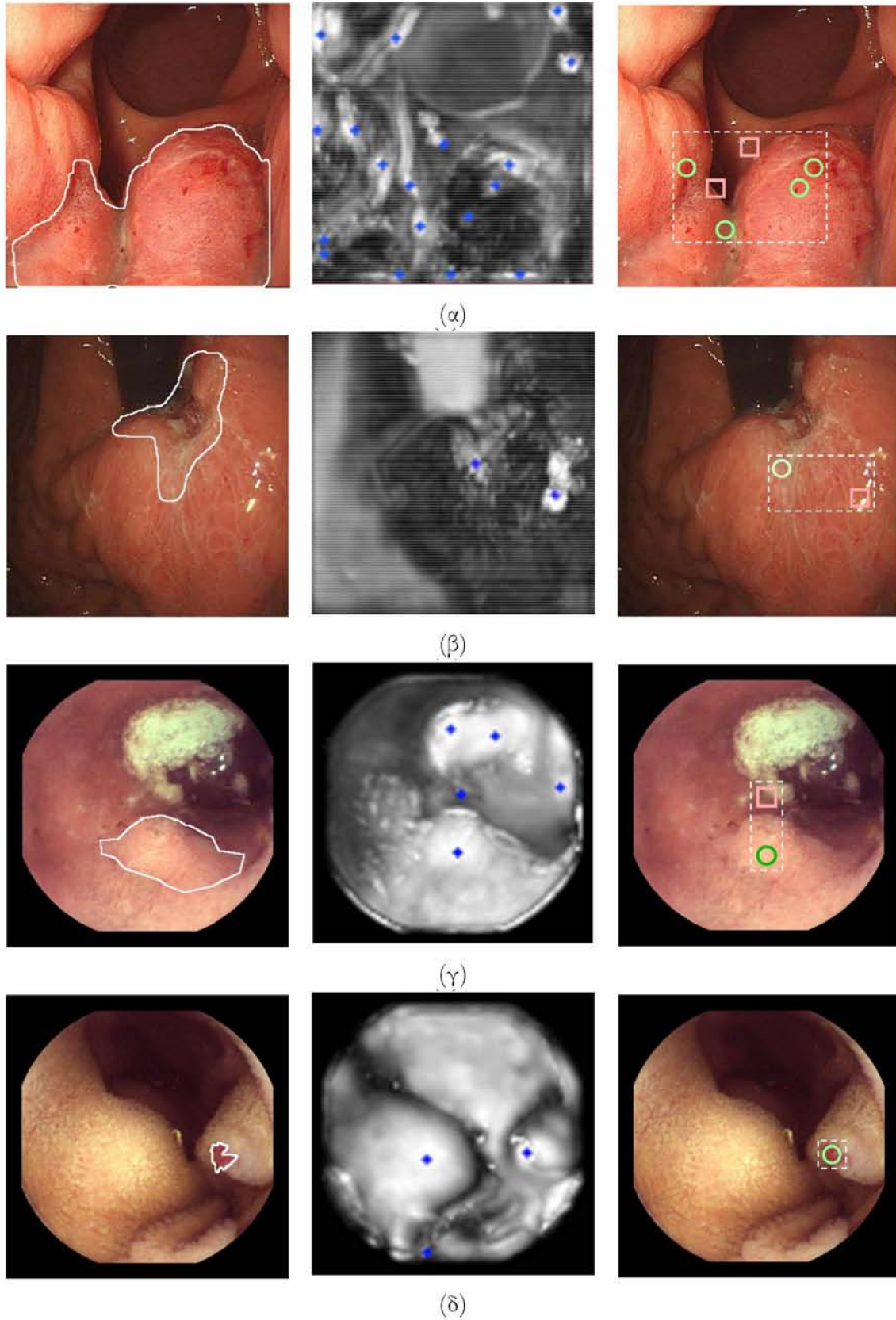
Στην Εικόνα 6.5 βλέπουμε στην κεντρική στήλη τις EM που αντιστοιχούν στις εικόνες εν-



Σχήμα 6.4: Τα σημαντικά σημεία που παράγονται από τον αλγόριθμο ΕΣΣΒΕ (αριστερή στήλη) και τον CSD (δεξιά στήλη). Εικόνες από το Σ1 (α - β). Εικόνες από το Σ2 (γ - δ).

δοσκοπίου που εμφανίζονται στη πρώτη στήλη, στις οποίες εγγράφεται η περιοχή παθογένειας από τους ειδικούς. Στις EM εμφανίζονται τα σημεία που παράγει ο αλγόριθμος ΕΣΣΒΕ, συγκεκριμένα η Εικόνα 6.5-α είναι η εικόνα από το Σ1, στην οποία παράγονται τα περισσότερα σημαντικά σημεία (18) από τον ΕΣΣΒΕ όπως παρουσιάζεται στο Πίνακα 6.6.

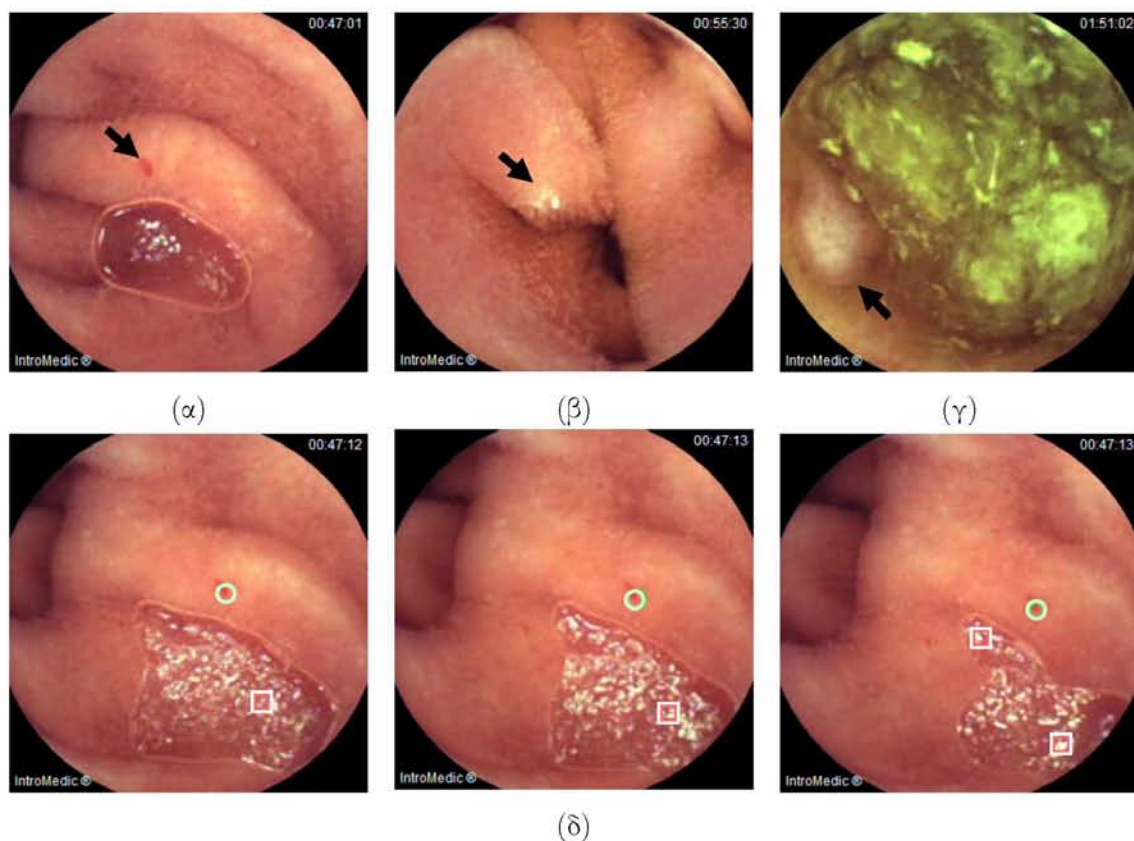
Στη συνέχεια εφαρμόζεται η ΜΕΕΟ της Φάση III της προτεινόμενης μεθοδολογίας για την βελτίωση των παραγόμενων σημαντικών σημείων του ΕΣΣΒΕ. Στην Εικόνα 6.5 στη δεξιά στήλη βλέπουμε τα σημαντικά σημεία που απομένουν μετά το φιλτράρισμα που γίνεται από τη ΜΕΕΟ. Με πράσινο κύκλο παρουσιάζονται τα σημεία που σωστά χαρακτηρίστηκαν ως σημαντικά σημεία και υποδεικνύουν παθογενή περιοχή ενώ με κόκκινο τετράγωνο τα λανθασμένα σημεία καθώς η περιοχή αυτή είναι φυσιολογική. Αξίζει να αναφερθεί ότι τουλάχιστον ένα σημείο σε κάθε χαρακτηρισμένη ως παθογενή εικόνα μετά το φιλτράρισμα μέσω του ΜΕΕΟ υποδεικνύει σωστά μια παθογενή περιοχή της εικόνας.



Σχήμα 6.5: Τέσσερις εικόνες που εμφανίζουν παθολογία και η περιοχή υποδεικνύεται από τους ειδικούς ιατρούς (αριστερή στήλη). Οι EM των αντίστοιχων εικόνων και με τα σημεία του αλγορίθμου ΕΣΣΒΕ (κεντρική στήλη). Φιλτράρισμα των σημαντικών σημείων με την χρήση του ΜΕΕΟ (δεξιά στήλη). Εικόνες από το Σ1 (α - β). Εικόνες από το Σ2 (γ - δ).

6.3.6 Περαιτέρω Εξέταση την Μεθοδολογίας σε Βίντεο και Εικόνες με Πολύποδες

Καθώς εξετάστηκε η προτεινόμενη μεθοδολογία σε μια σειρά από σύνολα δεδομένων με εικόνες του γαστρεντερικού σωλήνα, δημιουργήθηκε ένα εκπαιδευμένο μοντέλο το οποίο είχε υψηλότερη απόδοση σε σχέση με τις συγκρινόμενες μεθοδολογίες. Θελήσαμε να δημιουργήσουμε ένα ολιστικό μοντέλο για τον εντοπισμό εικόνων που εμφανίζουν παθολογικές καταστάσεις στο γαστρεντερικό σωλήνα, υποδεικνύοντας τις περιοχές αυτές στους ειδικούς χωρίς το μοντέλο να χρειάζεται εκπαίδευση κάθε φορά στο διαφορετικό τύπο εικόνων ως προς τη συσκευή λήψης τους. Για τον σκοπό αυτό χρησιμοποιήθηκε ένα εκπαιδευμένο μοντέλο ANΔΣ στο Σ3, το οποίο αποτελείται από τα Σ1 και Σ2 και ελέγχθηκε στο χαρακτηρισμό των στιγμιότυπων ενός βίντεο AEK του συνόλου δεδομένων KID [79] και των εικόνων κολonosκόπησης του CVC-CLINIC και ETIS-LARIB [15].



Σχήμα 6.6: Στιγμιότυπα εικόνων με παθολογικές περιοχές, όπως υποδεικνύονται από τους ειδικούς ιατρούς με το μαύρο βέλος, από το βίντεο KID: αγγειεκτασία (α), λεμφαγγειεκτασία (β), οξώδεις λεμφοειδείς υπερπλασία (γ). Στις εικόνες της δεύτερης γραμμής (δ) εμφανίζονται τρία διαδοχικά στιγμιότυπα με αγγειεκτασία όπου η προτεινόμενη μεθοδολογία έχει εντοπίσει και υποδεικνύει την περιοχή παθολογείας.

Το βίντεο KID αποτελείται από 2, 8 ώρες συνεχούς λήψης του γαστρεντερικού σωλήνα, με ρυθμό λήψης 3 καρέ το δευτερόλεπτο μέσω της κάμερας MiroCam CE. Έχει χαρακτηριστεί α-

πό 2 ειδικούς και περιλαμβάνει στιγμιότυπα με αγγειεκτασίες Εικόνα 6.6-α, λεμφοαγγειεκτασίες Εικόνα 6.6-β και οζώδεις λεμφοειδείς υπερπλασίες Εικόνα 6.6-γ. Η απόδοση της προτεινόμενης μεθοδολογίας ως προς το μέτρο AUC ήταν 0,886 για τον σωστό χαρακτηρισμό των στιγμιότυπων, ενώ 0,769 για τον σωστό χαρακτηρισμό των σημαντικών σημείων.

Καθώς τώρα γίνεται έλεγχος της μεθοδολογίας σε επίπεδο βίντεο, πρέπει να λάβουμε υπόψη ότι μια παθογενής περιοχή δεν θα είναι ένα απομονωμένο στιγμιότυπο αλλά σίγουρα θα αποτυπώνεται σε μια αλληλουχία στιγμιότυπων. Άρα πρέπει η μεθοδολογία να μπορεί να εντοπίζει τουλάχιστον ένα στιγμιότυπο από τις αλληλουχίες αυτές. Το βίντεο KID αποτελείται από 18 ακολουθίες και σε όλες αυτές η προτεινόμενη μεθοδολογία βρίσκει τουλάχιστον ένα στιγμιότυπο. Για να ποσοτικοποιήσουμε την απόδοση της μεθοδολογίας ως προς τον χαρακτηρισμό των στιγμιότυπων και τον εντοπισμό σημαντικών σημείων, θα εφαρμόσουμε την μετρική “Διαδοχικών Στιγμιότυπων” που προτείνεται στην Εργασία [15]. Η μετρική αυτή υπολογίζεται ως το ποσοστό των ανα δύο συνεχόμενων (ζευγαριών) στιγμιότυπων που έχουν χαρακτηριστεί σωστά στην κάθε ακολουθία συνεχόμενων στιγμιότυπων. Έτσι, ο μέσος όρος Διαδοχικών Στιγμιότυπων μεταξύ όλων των ακολουθιών που περιέχει το βίντεο KID είναι 50,4% με διασπορά 32,4%. Επιπλέον, ορίζεται και η μετρική “Εντός Ακολουθίας” [15] για τον προσδιορισμό της σταθερότητας της μεθόδου, η Εντός Ακολουθίας μέση τιμή της προτεινόμενης μεθοδολογίας ως προς τις ακολουθίες είναι 65,9% με διασπορά 28,1%. Η μετρική αυτή ορίζεται ως το ποσοστό των σωστά χαρακτηρισμένων στιγμιότυπων σε κάθε ακολουθία. Στην Εικόνα 6.6-δ παρουσιάζεται ένα παράδειγμα τριών συνεχόμενων στιγμιότυπων που εμπεριέχουν μια παθογενή περιοχή.

Τελειώνοντας την πειραματική μελέτη, η προτεινόμενη μεθοδολογία θα εξεταστεί σε δύο ακόμα σύνολα δεδομένων που είναι στην πραγματικότητα τα σύνολα δεδομένων του διαγωνισμού “Automatic Polyp Detection in Colonoscopy Videos”³. Το σύνολο δεδομένων CVC-CLINIC αποτελείται από 31 ακολουθίες 31 διαφορετικών τύπων πολύποδων χωρίς καμία φυσιολογική εικόνα (δηλ. εικόνα χωρίς πολύποδες), ενώ συνολικά διαθέτει 612 εικόνες. Το σύνολο δεδομένων ETIS-LARIB αποτελείται από 196 εικόνες και συγκεκριμένα χωρίζεται σε 34 ακολουθίες στις οποίες εμφανίζονται 44 διαφορετικοί τύποι πολύποδων, ενώ και αυτό το σύνολο δεδομένων δεν περιέχει φυσιολογικές εικόνες. Το σύνολο δεδομένων CVC-CLINIC είχε δοθεί ως σύνολο εκπαίδευσης για τις ανάγκες του διαγωνισμού ενώ το ETIS-LARIB ήταν το σύνολο ελέγχου των μεθοδολογιών. Καθώς η προτεινόμενη μεθοδολογία είναι εκπαιδευμένη ήδη με το Σ3, δεν θα χρησιμοποιήσουμε το CVC-CLINIC για εκπαίδευση.

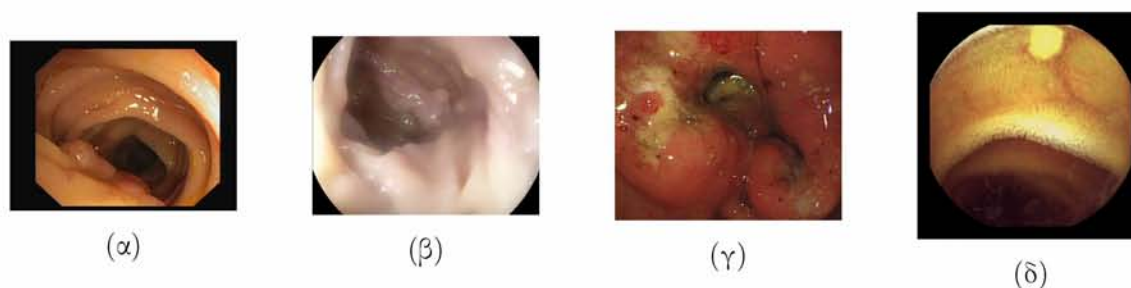
Για λόγους σύγκρισης της προτεινόμενης μεθοδολογίας με τις αντίστοιχες μεθόδους που προτάθηκαν στον διαγωνισμό θα ακολουθήσουμε τις μετρικές που ορίστηκαν στην Εργασία [15]. Συγκεκριμένα, θα χρησιμοποιήσουμε το Αληθές Θετικό (ΑΘ), Λανθασμένα Θετικό (ΛΘ), Ψευδές Θετικό (ΨΘ) και τα μέτρα $F1$ και $F2$. Η προτεινόμενη μεθοδολογία ακολουθώντας την πειραματική μελέτη της Εργασίας [15], θα αξιολογηθεί ως προς την ικανότητά της να μπορεί να χαρακτηρίζει ορθά μια εικόνα ως παθογενή, αν δηλαδή περιέχει κάποια παθογενή κατάσταση και στην συνέχεια αν τα σημαντικά σημεία σε αυτή είναι ορθά (εντός ή εκτός της παθογενούς περιοχής). Συγκεκριμένα οι μετρικές που θα χρησιμοποιήσουμε ορίζονται ως εξής

³<https://polyp.grand-challenge.org/site/Polyp/>

(Εργασία [15]): ένα ΑΘ αντιστοιχεί σε μια περιοχή-πολύποδα αρκεί να υπάρχει ένα σημαντικό σημείο εντός αυτής. Κάθε σημαντικό σημείο εκτός της παθολογούς περιοχής λογίζεται ως ΛΘ ενώ η απουσία εντοπισμού μια εικόνας που περιέχει παθολογεία χαρακτηρίζεται ως ΨΘ. Πρέπει να τονιστεί ότι καθώς δεν υπάρχουν φυσιολογικές εικόνες χωρίς πολύποδες δεν υφίσταται η έννοια των Αληθώς Αρνητικών (ΑΑ).

Μέθοδος	ΑΘ	ΛΘ	ΨΘ	Ακρίβεια	Ανάκληση	F1	F2
ETIS-LARIB							
1 ^η Θέση [15]	144	55	64	72,3	69,5	70,7	69,8
2 ^η Θέση [15]	131	57	77	69,7	63	66,1	64,2
Προτεινόμενη	94	69	114	57,7	45,2	50,7	47,2
3 ^η Θέση [15]	110	226	98	32,7	52,8	40,4	47,1
CVC-CLINIC							
Προτεινόμενη	284	223	362	56	44	49,3	45,9

Πίνακας 6.7: Απόδοση της προτεινόμενης μεθοδολογίας έναντι των μεθοδολογιών που προτάθηκαν για τον διαγωνισμό ‘Automatic Polyp Detection in Colonoscopy Videos’ στα σύνολα δεδομένων ETIS-LARIB και CVC-CLINIC. Καθώς το σύνολο δεδομένων CVC-CLINIC χρησιμοποιήθηκε από τις άλλες μεθοδολογίες ως σύνολο εκπαίδευσης δεν έχουμε δυνατότητα παρουσίασης απόδοσης πέραν την προτεινόμενης μεθόδου η οποία έχει εκπαιδευτεί από το Σ3.



Σχήμα 6.7: Εικόνες που απεικονίζουν παθολογικές περιοχές από τα τέσσερα διαφορετικά είδη συνόλων δεδομένων που χρησιμοποιήθηκαν στο κεφάλαιο αυτό. CVC-CLINIC (α), ETIS-LARIB (β), MICCAI (γ) και KID (δ).

Για την προτεινόμενη μεθοδολογία οι εικόνες πρέπει να μετασχηματιστούν στις διαστάσεις 320×320 σύμφωνα με το μοντέλο ΑΝΔΣ της μεθοδολογίας (Εικόνα 6.1). Τα αποτελέσματα της απόδοσης της προτεινόμενης μεθοδολογίας παρουσιάζονται στον Πίνακα 6.7 και για τα δύο σύνολα δεδομένων. Η προτεινόμενη μεθοδολογία κατατάσσεται τρίτη στη σειρά απόδοσης σε σχέση με τις 7 μεθοδολογίες που παρουσιάζονται στην Εργασία [15] ως προς τις μετρικές Ακρίβειας, F1 και F2. Πρέπει να τονιστεί ότι οι υπόλοιπες μεθοδολογίες έχουν εκπαιδευτεί στο CVC-CLINIC σύνολο δεδομένων που αποτελείται μόνο από πολύποδες, ενώ χρησιμοποιούν τον αυστηρό χαρακτηρισμό που δίνουν οι ειδικοί και όχι τον ασθενή χαρακτηρισμό με

τον οποίο έχει εκπαιδευτεί το προτεινόμενο μοντέλο. Το μοντέλο της προτεινόμενης μεθοδολογίας είναι εκπαιδευμένο σε πολλές και διαφορετικές εικόνες του γαστρεντερικού σωλήνα, ενώ ο τύπος της κάμερας αλλά και το οπτικό πεδίο των εικόνων αυτών διαφέρει (Εικόνα 6.7). Επιπλέον, οι μεθοδολογίες της [15] είναι στοχευμένες στον εντοπισμό των περιοχών των παθογενών καταστάσεων σε εικόνες που εμφανίζουν πολύποδα. Η προτεινόμενη μεθοδολογία αντίθετα, χαρακτηρίζει πρώτα τις εικόνες ως προς το αν παρουσιάζουν κάποια παθολογία του γαστρεντερικού συστήματος (Φάση I) και στη συνέχεια εντοπίζει τις περιοχές αυτές (Φάση II και III). Έτσι, στο σύνολο δεδομένων ETIS-LARIB χαρακτηρίζει σωστά τις 191 (97,5%) εικόνες ότι ορθώς περιέχουν κάποια παθολογία ενώ στο CVC-CLINIC τις 578 (94,5%). Η προτεινόμενη μεθοδολογία είναι γενικότερη και μπορεί να χρησιμοποιηθεί σε διαφορετικού τύπου εικόνες του γαστρεντερικού σωλήνα με μεγάλη ακρίβεια χωρίς επιπλέον εκπαίδευση.

6.4 Συμπεράσματα

Στο κεφάλαιο αυτό προτάθηκε μια νέα μεθοδολογία ασθενούς εκπαίδευσης για τον χαρακτηρισμό εικόνων του γαστρεντερικού σωλήνα ως προς αν εμφανίζουν ή όχι τυχόν παθογενή κατάσταση και στη συνέχεια προσδιορισμό της παθογενούς περιοχής. Ολόκληρη η μεθοδολογία υλοποιείται χωρίς την χρήση αυστηρής υπόδειξης από τους ειδικούς ιατρούς της παθογενούς περιοχής αλλά μόνο από την ‘ασθενή’ πληροφορία για το αν εμφανίζεται τυχόν παθολογία στην εικόνα ή όχι. Έτσι, συγκριτικά με άλλες μεθόδους που απαιτούν πλήρη πληροφορία για τον εντοπισμό των παθογενών περιοχών, παρουσιάζει στις περισσότερες των περιπτώσεων καλύτερα αποτελέσματα ή ισοδύναμα, συνεισφέροντας σημαντικά ακόμα στον οικονομοτεχνικό τομέα, λαμβάνοντας υπόψη το κόστος υλοποίησης που έχουν οι μέθοδοι πλήρους πληροφορίας [74], [117].

Η μέθοδος μπορεί να διαιρεθεί σε τρεις φάσεις. Πυλώνας της μεθόδου είναι το εκπαιδευμένο μοντέλο Τεχνητών Νευρωνικών Δικτύων Συνελίξεων (Φάση I) που δημιουργήθηκε από ασθενείς χαρακτηρισμένες εικόνες του γαστρεντερικού σωλήνα που εμφανίζουν πολλές διαφορετικές παθογενείς καταστάσεις. Μέσω της πειραματικής διαδικασίας και συγκρινόμενο με άλλες μεθόδους αναφοράς, που είτε χρησιμοποιούσαν ασθενούς είτε πλήρους χαρακτηρισμού σύνολο δεδομένων, το προτεινόμενο μοντέλο είχε καλύτερη απόδοση και ποσοστά επιτυχίας στον χαρακτηρισμό των εικόνων.

Στη συνέχεια, μέσω του αλγορίθμου Εύρεσης Σημαντικών Σημείων από Βαθύ Επίπεδο (Φάση II) και της μεθόδου φιλτραρίσματος σημείων (Φάση III) υποδεικνύεται η παθογενής περιοχή μέσω σημείων πάνω στην εικόνα. Σημαντικό στοιχείο και σε αυτό το κομμάτι είναι η έλλειψη οποιασδήποτε γνώσης των πραγματικά παθογενών και φυσιολογικών περιοχών που εμφανίζουν οι γαστρεντερικές εικόνες. Τα αποτελέσματα δείχνουν σαφή υπεροχή της προτεινόμενης μεθόδου έναντι άλλων μεθόδων παραγωγής σημείων.

Τέλος, είναι σημαντική η ικανότητα του μοντέλου να δρα ως ένα ολιστικό και αυτόνομο μοντέλο αποδεδειγμένο από τον τύπο και τα χαρακτηριστικά της κάμερας που λαμβάνει ως εικόνες. Είδαμε να κατατάσσεται τρίτο μεταξύ των 7 μεθοδολογιών που είχαν σχεδιαστεί ειδικά για να λύσουν ένα πρόβλημα πολύποδων σε έναν ανοιχτό διαγωνισμό, χωρίς να χρησιμοποιεί

το σύνολο εκπαίδευσης του διαγωνισμού όπως οι άλλες μέθοδοι, αλλά έχοντας εκπαιδευτεί με άλλα σύνολα δεδομένων με πολύ περισσότερες και διαφορετικές παθογενείς καταστάσεις.

Μέρος IV

Μέθοδοι Μη Μονότονοι Εκπαίδευσης ΤΝΔΣ

Κεφάλαιο 7

Εκπαίδευση ΤΝΔΣ με Μεταβλητό Ρυθμό Εκπαίδευσης

Στο κεφάλαιο αυτό θα ασχοληθούμε με την εκπαίδευση των ΤΝΔΣ χρησιμοποιώντας μεταβλητό ρυθμό εκμάθησης. Η χρήση κανόνων που προσαρμόζουν τον ρυθμό εκμάθησης κατά την διάρκεια εκπαίδευσης του δικτύου, επιτυγχάνουν καλύτερη γενίκευση του εκπαιδευμένου μοντέλου ιδιαίτερα σε περιπτώσεις με μεγάλο σύνολο δεδομένων εκπαίδευσης, ή μεγάλο πλήθος διαφορετικών κλάσεων τις οποίες το μοντέλο εκπαιδεύεται να μάθει.

Στη συνέχεια θα παρουσιάσουμε μια μέθοδο αυτόματης προσαρμογής ενός κοινού ρυθμού εκμάθησης για τα βάρη όλων των επιπέδων του δικτύου. Η προτεινόμενη μέθοδος βασίζεται στην μέθοδο [7] και στην εργασία [100], αποτελώντας μια τροποποίηση της γνωστής μεθόδου βελτιστοποίησης, της Στοχαστικής μεθόδου Απότομης Κλίσης (ΣΑΚ) (Stochastic Gradient Descent) [16]. Η προτεινόμενη προσέγγιση χρησιμοποιεί την πληροφορία της κλίσης προηγούμενων επαναλήψεων αλλά ενσωματώνει και έναν κανόνα επανεκκίνησης της τιμής του ρυθμού εκμάθησης σε περιπτώσεις αρνητικών τιμών.

Ο προτεινόμενος αλγόριθμος εφαρμόστηκε και εξετάστηκε σε γνωστό πρόβλημα που χρησιμοποιείται ευρέως από την επιστημονική κοινότητα ως σημείο αναφοράς [128] ενώ συγκρίθηκε με καθιερωμένες στο χώρο μεθόδους μεταβλητού ρυθμού εκμάθησης για την εκπαίδευση ΤΝΔΣ.

7.1 Εισαγωγή

Στο ερευνητικό πεδίο των ΤΝΔΣ, τα σύνολα δεδομένων όπου χρησιμοποιούνται για την εκπαίδευσή τους είναι συνήθως μεγάλου όγκου, έτσι η χρήση αλγορίθμων εκπαίδευσης ανά πρότυπο είναι επιτακτική. Ωστόσο η χρήση αλγορίθμων εκπαίδευσης αυστηρά ανα πρότυπο εισόδου είναι μη αποδοτική τόσο στη γενίκευση του παραγόμενου μοντέλου όσο και στο χρόνο εκπαίδευσής του, έτσι υιοθετείται η χρήση ψευδο-ομάδων προτύπων εισόδου μεταβλητού μεγέθους κατά την διαδικασία της εκπαίδευσης [54].

Για την εκπαίδευση των ΤΝΔΣ με ψευδο-ομάδες προτύπων εισόδου η στοχαστική μέθοδος απότομης κλίσης είναι ευρέως διαδεδομένη παράγοντας μοντέλα με πολύ υψηλά ποσοστά

γενίκευσης στα δεδομένα ελέγχου. Βασική παράμετρος της μεθόδου είναι ο ρυθμός εκμάθησης, όμως δεν έχει καθιερωθεί ένας καθολικός κανόνας επιλογής της βέλτιστης τιμής ρυθμού εκμάθησης για όλα τα διαφορετικά προβλήματα. Έχουν παρατηρηθεί περιπτώσεις με διαφορετικές τιμές ρυθμού εκμάθησης στις οποίες ο αλγόριθμος δεν συγκλίνει καθόλου είτε συγκλίνει σε κάποιο όχι καλό τοπικό ελάχιστο, με αποτέλεσμα η γενίκευση του μοντέλου να είναι εξαιρετικά χαμηλή.

Το μέγεθος των ψευδο-ομάδων προτύπων εισόδου στην εκπαίδευση των ΤΝΔΣ έχει καθοριστικό ρόλο. Η σύγκλιση σε κάποιο όχι καλό τοπικό ελάχιστο οφείλεται στο γεγονός ότι μεγάλοι μεγέθους ψευδο-ομάδες προτύπων εισόδου τείνουν να συγκλίνουν σε στενές περιοχές της συνάρτησης εκπαίδευσης με απότομη κλίση [75]. Αυτές η περιοχές χαρακτηρίζονται από υψηλές θετικές ιδιοτιμές της δεύτερης παραγώγου ($\nabla^2 f(x)$) της συνάρτησης σφάλματος του ΤΝΔΣ (f) και τείνουν να έχουν χαμηλές τιμές γενίκευσης. Σε αντίθεση με την χρήση μικρότερου μεγέθους ψευδο-ομάδων προτύπων εισόδου, η σύγκλιση της μεθόδου τείνει σε περιοχές σχετικά επίπεδες με χαμηλές ιδιοτιμές της δεύτερης παραγώγου της συνάρτησης σφάλματος f .

Η χρήση ρυθμού εκμάθησης μεταβλητού μεγέθους κατά την διάρκεια εκπαίδευση του ΤΝΔΣ πειραματικά έχει δείξει ότι βοηθάει στη λύση του προβλήματος. Έχουν προταθεί 2 στρατηγικές μεταβολής του ρυθμού εκμάθησης:

- η στρατηγική σταδιακής μείωσης
- η στρατηγική σταδιακής αύξησης (warmup strategy).

Ο ρυθμός εκμάθησης αφού αρχικοποιηθεί μεταβάλλεται σταδιακά κατά την διάρκεια εκπαίδευσης του δικτύου, ύστερα από έναν προεπιλεγμένο αριθμό επαναλήψεων κάθε φορά [62]. Παραλλαγές των δύο στρατηγικών έχουν προταθεί στη βιβλιογραφία για καλύτερη απόδοση σε διαφορετικά προβλήματα [61, 52]. Προτείνεται η χρήση ενός χαμηλού ρυθμού εκμάθησης κατά τις πρώτες επαναλήψεις της εκπαίδευσης του δικτύου ενώ για την συνέχεια της εκπαίδευσης του δικτύου, επαναφορά του ρυθμού εκμάθησης σε μεγαλύτερη τιμή [52]. Ωστόσο, με αυτή την τεχνική παρατηρείται πολλές φορές η απότομη αλλαγή του σφάλματος εκπαίδευσης του δικτύου. Μια άλλη παραλλαγή είναι η σταδιακή αύξηση ρυθμού εκμάθησης από μια μικρή αρχική τιμή σε μεγαλύτερη, μέσα σε ένα συγκεκριμένο εύρος επαναλήψεων και στη συνέχεια επιστροφή στις αρχική τιμή του.

Η επιλογή όμως τις αρχικής τιμής του ρυθμού εκμάθησης έχει σημαντική και καθοριστική αξία στις στρατηγικές που θα εφαρμοστούν για την εκπαίδευση του δικτύου. Στην εκπαίδευση ενός δικτύου με μεγάλο μεγέθους ψευδο-ομάδες προτύπων εισόδου (> 1000) η χρήση μια τυπικής τιμής ρυθμού εκπαίδευσης ($\eta = 0.1$) δεν είναι κατάλληλη ώστε να συγκλίνει ο αλγόριθμος εκπαίδευσης. Η υιοθέτηση μεγαλύτερης τιμής φαίνεται ότι βοηθάει σημαντικά, καθώς έχει φανεί πειραματικά ότι υπάρχει μια γραμμική σχέση του ρυθμού εκμάθησης με το πλήθος των προτύπων που χρησιμοποιούνται [54].

Μια στρατηγική αντιμετώπισης του προβλήματος εκπαίδευσης των ΤΝΔΣ με χρήση μεταβλητού ρυθμού εκμάθησης είναι η ανάπτυξη αλγορίθμων εκπαίδευσης όπου ενσωματώνουν την μεταβολή του ρυθμού εκμάθησης στην βασική τους λειτουργία [138, 37, 13, 92, 86, 100].

7.2 Αλγόριθμοι Εκπαίδευσης Μεταβλητού Ρυθμού Εκμάθησης

Η επιλογή του κατάλληλου ρυθμού εκμάθησης είναι ιδιαίτερα σημαντική για την εκπαίδευση του δικτύου και την σύγκλιση του στοχαστικού αλγορίθμου απότομης κλίσης. Όταν η τιμή του ρυθμού εκμάθησης είναι πολύ μικρή, η εκπαίδευση του μοντέλου γίνεται πολύ αργά και με δεδομένο ότι τα ΤΝΔΣ έχουν μεγάλους χρόνους εκπαίδευσης η επιλογή αυτή είναι αποτρεπτική. Αντίθετα, για πολύ μεγάλες τιμές του ρυθμού εκμάθησης μπορεί ο αλγόριθμος εκπαίδευσης να μην συγκλίνει ποτέ.

Στη βιβλιογραφία έχουν προταθεί πολλοί διαφορετικοί αλγόριθμοι προσαρμογής του ρυθμού εκμάθησης για την εκπαίδευση των ΤΝΔΣ. Ο αλγόριθμος AdaDelta [138] προσαρμόζει δυναμικά την τιμή του ρυθμού εκμάθησης κάνοντας χρήση μόνο της πληροφορίας του διανύσματος της κλίσης των βαρών του (gradient vector). Μια άλλη δημοφιλής μέθοδος βελτιστοποίησης είναι ο AdaGrad [37] ο οποίος χρησιμοποιεί την πληροφορία της κλίσης των βαρών από τις προηγούμενες επαναλήψεις. Συγκεκριμένα, προσαρμόζει το ρυθμό εκμάθησης στο χρόνο, κάνοντας χρήση του αθροίσματος των διανυσμάτων της κλίσης των βαρών από τις προηγούμενες επαναλήψεις. Η τιμή του ρυθμού εκμάθησης (Εξίσωση 7.1) είναι αντιστρόφως ανάλογη της τετραγωνικής ρίζας του αθροίσματος των τετραγώνων διανυσμάτων της κλίσης των βαρών της κάθε προηγούμενης επανάληψης.

$$w_{i+1} = w_i - \frac{\eta_0 \cdot \nabla E(w_i)}{e + \sqrt{\sum_{k=1}^i (\nabla E(w_k))^2}} \quad (7.1)$$

Η παράμετρος η_0 αντιστοιχεί στην αρχική τιμή του ρυθμού εκμάθησης, η παράμετρος $e = 10^{-7}$ αποτελεί έναν πολύ μικρό θετικό αριθμό ώστε να αποφευχθεί η διαίρεση με το μηδέν. Ωστόσο, λόγω του τετραγώνου στον παρανομαστή του κλάματος ο παρανομαστής είναι πάντα θετικός. Αυτό οδηγεί στη συνεχή μείωση του ρυθμού εκμάθησης έτσι μετά από πολλές επαναλήψεις τα βάρη σταματούν να 'μαθαίνουν'.

Η δημοφιλέστερη μέθοδος προσαρμογής του ρυθμού εκμάθησης όπου έχει καλύτερα αποτελέσματα σε σύγκριση με τις προηγούμενες μεθόδους είναι η μέθοδος ADAM [13]. Βασίζεται στην προσέγγιση των ροπών χαμηλότερης τάξης (Εξισώσεις 7.2, 7.3). Συγκεκριμένα για τον υπολογισμό της καινούριου ρυθμού εκμάθησης, χρησιμοποιείται ένας κινητός μέσος της ροπής πρώτης τάξης (Εξίσωση 7.2) και της ροπής δεύτερης τάξης (Εξίσωση 7.3) του διανύσματος κλίσης των βαρών. Οι τιμές των βαρών ανανεώνονται με τη βοήθεια της εξίσωσης 7.4.

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \cdot \nabla E(w_{i-1}) \quad (7.2)$$

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) \cdot (\nabla E(w_{i-1}))^2 \quad (7.3)$$

$$w_{i+1} = w_i - \eta_0 \frac{m_i / (1 - \beta_1)}{v_i / (1 - \beta_2)} \quad (7.4)$$

Η παράμετρος η_0 αντιστοιχεί στην αρχική τιμή του ρυθμού εκμάθησης ενώ οι παράμετροι β_1 και β_2 είναι μεταπαράμετροι της μεθόδου με τιμές πολύ κοντά στο ένα. Βασικό μειονέκτημα της μεθόδου είναι η εξάρτησή της από την επιλογή της αρχικής τιμής του ρυθμού εκμάθησης.

7.2.1 Αλγόριθμος

Η εξάρτηση των μεθόδων εκπαίδευσης από την αρχική τιμή του ρυθμού εκμάθησης αποτελεί ένα από τα σοβαρότερα προβλήματα στην εκπαίδευση των ΤΝΔΣ. Η στρατηγική δοκιμής-και-επιλογής διαφορετικών αρχικών τιμών είναι πολλές φορές μια χρονοβόρα διαδικασία, ιδιαίτερα όταν η εκπαίδευση ενός δικτύου απαιτεί αρκετές ημέρες ακόμα και σε συστήματα υψηλών επιδόσεων. Στηριζόμενοι στις εργασίες [7, 100], προτείνουμε έναν αλγόριθμο εκπαίδευσης των ΤΝΔΣ με υψηλή ανοχή στις διαφορετικές αρχικές τιμές του ρυθμού εκμάθησης.

Ο προτεινόμενος αλγόριθμος [47] βασίζεται στον αλγόριθμο [7] στον οποίο χρησιμοποιείται η πληροφορία της κλίσης του διανύσματος των βαρών της τελευταίας και της προηγούμενης επανάληψης. Συγκεκριμένα, σύμφωνα με την εργασία [7] όταν η διεύθυνση του διανύσματος των βαρών δύο συνεχόμενων επαναλήψεων τείνουν να είναι πανομοιότυπες τότε ο ρυθμός εκμάθησης αυξάνεται, σε αντίθετη περίπτωση όταν οι διευθύνσεις διαφέρουν σημαντικά τότε ο ρυθμός εκμάθησης μειώνεται. Για τον υπολογισμό αυτό υπολογίζουμε το συνημίτονο της γωνίας μεταξύ των δύο διανυσμάτων αφού πρώτα κανονικοποιηθούν με μέτρο μονάδα. Προσεγγιστικά η εύρεση του συνημιτόνου της γωνίας επιτυγχάνεται με την χρήση του εσωτερικού γινομένου ($\langle \cdot, \cdot \rangle$) των διανυσμάτων των βαρών. Αξίζει να σημειωθεί ότι το εσωτερικό γινόμενο των ορθοκανονικών διανυσμάτων είναι μηδέν, ενώ των παράλληλων διανυσμάτων είναι ένα. Έτσι ο ρυθμός εκμάθησης προσαρμόζεται με τον ακόλουθο τύπο:

$$n_i = n_{i-1} + \gamma \langle \nabla E_{i-1}(w_{i-1}), \nabla E_i(w_i) \rangle, \quad (7.5)$$

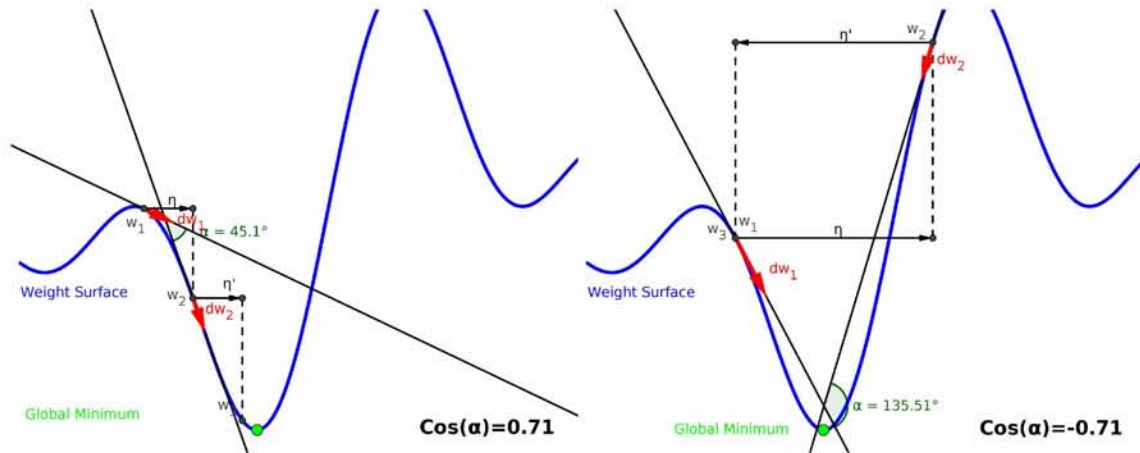
όπου γ είναι μια παράμετρος ελέγχου του ρυθμού εκμάθησης. Στην εικόνα 7.1 ακολουθούν δύο παραδείγματα εύρεσης του ολικού ελαχίστου της συνάρτησης των βαρών για μικρές (αριστερά) και μεγάλες (δεξιά) τιμές του ρυθμού εκμάθησης, καθώς και η εύρεση της γωνίας των διανυσμάτων των βαρών.

Ο προτεινόμενος αλγόριθμος βασίζεται στον αλγόριθμο μεταβλητού ρυθμού εκμάθησης για την εκπαίδευση ΤΝΔ [86, 100]. Ο προτεινόμενος αλγόριθμος κάνει χρήση της διεύθυνσης των διανυσμάτων των βαρών των τελευταίων τριών επαναλήψεων. Ο ρυθμός εκμάθησης προσαρμόζεται λαμβάνοντας υπόψη το εσωτερικό γινόμενο των δύο τελευταίων και των δύο προηγούμενων διανυσμάτων των βαρών. Έτσι ο ρυθμός εκμάθησης προσαρμόζεται με τον ακόλουθο τύπο:

$$\begin{aligned} n_i = n_{i-1} &+ \gamma_1 \langle \nabla E_{i-1}(w_{i-1}), \nabla E_i(w_i) \rangle + \\ &+ \gamma_2 \langle \nabla E_{i-2}(w_{i-2}), \nabla E_{i-1}(w_{i-1}) \rangle, \end{aligned} \quad (7.6)$$

όπου γ_1 και γ_2 είναι παράμετροι ελέγχου του ρυθμού εκμάθησης και ονομάζονται παράμετροι μετα-ρυθμού εκμάθησης. Για τις τιμές των παραμέτρων του μετα-ρυθμού εκμάθησης προτείνεται να ισχύει $\gamma_2 \ll \gamma_1$ και $\gamma_1 < 1$ έτσι ώστε η πληροφορία του διανύσματος των βαρών από τις προηγούμενες επαναλήψεις να έχει χαμηλότερη επίδραση στη μεταβολή του ρυθμού εκμάθησης σε σχέση με αυτή των δύο τελευταίων. Επιπλέον, για να αποτρέψουμε την απόκλιση της μεθόδου σε περίπτωση που ο ρυθμός εκμάθησης αποκτήσει αρνητικές τιμές, εισάγαμε τον ακόλουθο κανόνα:

$$n_i = \begin{cases} n_0 \cdot \gamma_3, & \text{if } n_i < 0 \\ n_i, & \text{if } n_i \geq 0 \end{cases} \quad (7.7)$$



Σχήμα 7.1: Σχηματική απεικόνιση της μεθόδου εύρεσης τοπικού ελαχίστου συνάρτησης στις δύο διαστάσεις. Στη περίπτωση που έχουμε μικρή τιμή ρυθμού εκμάθησης, το συνημίτονο της γωνίας των διανυσμάτων των βαρών των δύο τελευταίων επαναλήψεων είναι θετικό (αριστερά). Όταν η τιμή του ρυθμού εκμάθησης είναι μεγάλη και αποκλίνει από το τοπικό ελάχιστο, το συνημίτονο της γωνίας των διανυσμάτων των βαρών των δύο τελευταίων επαναλήψεων είναι αρνητικό (δεξιά).

όπου η_0 είναι ο αρχικός ρυθμός εκμάθησης ενώ η_3 συντελεστής μείωσης του αρχικού ρυθμού εκμάθησης. Η χρήση του συντελεστή μείωσης είναι σημαντική αφού ο κανόνας (7.7) αυτός ενεργοποιείται συνήθως μετά από πολλές επαναλήψεις της μεθόδου όπου τείνει να προσεγγίσει ένα καλό τοπικό ελάχιστο. Η χρήση του αρχικού ρυθμού εκμάθησης χωρίς τον συντελεστή μείωσης, στην περίπτωση μεγάλης αρχικής τιμής, αποτρέπει την συνέχιση της προσέγγισης στο τοπικό ελάχιστο απαιτώντας αρκετές επαναλήψεις ώστε να συγκλίνει ξανά. Μια αλγοριθμική περιγραφή του προτεινόμενου αλγορίθμου δίνεται στον Αλγόριθμο 2.

7.3 Προσομοιώσεις και Αποτελέσματα

Στην ενότητα αυτή θα παρουσιάσουμε τα αποτελέσματα του αλγορίθμου σε γνωστό πρόβλημα εκπαίδευσης ΤΝΔΣ. Το Αλγόριθμος 2 έχει αξιολογηθεί και συγκριθεί με άλλους γνωστούς αλγορίθμους εκπαίδευσης ΤΝΔΣ για διαφορετικές αρχικές τιμές του ρυθμού εκμάθησης.

Συγκεκριμένα, αξιολογήσαμε τον προτεινόμενο αλγόριθμο στο πρόβλημα αναγνώρισης αντικειμένων σε μικρές εικόνες CIFAR-10 [128]. Το πρόβλημα αυτό αποτελείται από 60,000 μικρές έγχρωμες εικόνες διάστασης 32×32 εικονοστοιχείων και περιέχει 10 διαφορετικές κλάσης με 6,000 εικόνες να αντιστοιχούν στην κάθε κλάση. Το πρόβλημα αυτό έχει διαχωριστεί σε 50,000 εικόνες οι οποίες χρησιμοποιούνται για την εκπαίδευση του δικτύου ενώ οι υπόλοιπες 10,000 εικόνες για τον έλεγχο της απόδοσης του εκπαιδευμένου δικτύου. Για λόγους επαναληψιμότητας και εύκολου ελέγχου των πειραμάτων από την επιστημονική κοινότητα χρησιμοποιήθηκε μια γνωστή και ευρέως διαδεδομένη αρχιτεκτονική σχεδίαση του ΤΝΔΣ. Συγκεκριμένα, το ΤΝΔΣ αποτελείται από δύο επίπεδα συνελίξεων με 32 μητρώα χαρακτηριστικών και μέγεθος πυρήνα διάστασης 5×5 , ενώ το κάθε επίπεδο ακολουθείται από

Algorithm 2: The Stochastic Gradient Descent with Adaptive Learning Rate Algorithm (AdLR)

```

1 Input:  $w_0, n_0, \gamma_1, \gamma_2, \gamma_3, iter$  ;
2  $i = 0$ ;
3 while  $i < iter$  do
4    $i = i + 1$ ;
5   Calculate  $E(w_i)$  and  $\nabla E_i(w_i)$ ;
6   Update the weights
7    $w_{i+1} = w_i - n_i \nabla E_i(w_i)$ ;
8   Update learning rate
9    $n_i = n_{i-1} + \gamma_1 \langle \nabla E_{i-1}(w_{i-1}), \nabla E_i(w_i) \rangle +$ 
       $+ \gamma_2 \langle \nabla E_{i-2}(w_{i-2}), \nabla E_{i-1}(w_{i-1}) \rangle$ 
10  if  $n_i < 0$  then
11     $n_i = n_0 \cdot \gamma_3$ 
12  end
13 end
14 Output:  $w_{i+1}$ 

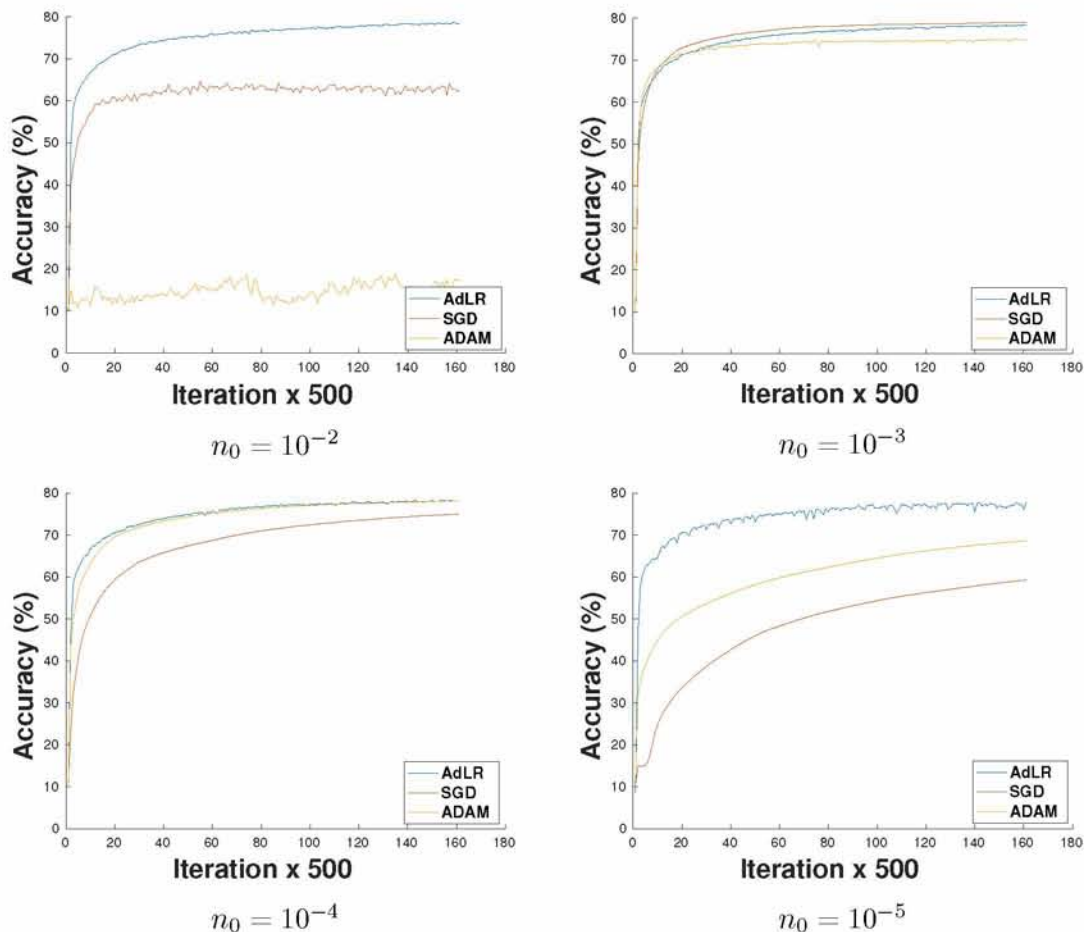
```

ένα επίπεδο μέγιστης δειγματοληψίας διάστασης 3×3 . Στη συνέχεια, ακολουθεί ένα ακόμα επίπεδο συνελίξεων με 64 μητρώα χαρακτηριστικών και μέγεθος πυρήνα διάστασης 5×5 και ένα επίπεδο μέγιστης δειγματοληψίας διάστασης 3×3 . Τέλος, ακολουθεί ένα πλήρως διασυνδεδεμένο επίπεδο απλών νευρώνων μεγέθους 10 όπου η συνάρτηση παλινδρόμησης softmax εφαρμόζεται σε αυτό. Μετά από κάθε επίπεδο συνελίξεων εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU.

Ο προτεινόμενος Αλγόριθμος 2 συγκρίθηκε με τον κλασικό στοχαστικό αλγόριθμο της απότομης κλίσης (ΣΑΚ) καθώς και με τον αλγόριθμο εκπαίδευσης ΤΝΔΣ μεταβλητού ρυθμού εκπαίδευσης ADAM. Οι τρεις αλγόριθμοι εξετάστηκαν στην εκπαίδευση της προαναφερθείσας αρχιτεκτονικής ΤΝΔΣ για την επίλυση του προβλήματος CIFAR-10 χρησιμοποιώντας διαφορετικές αρχικές τιμές για τον ρυθμό εκμάθησης $\eta_0 = \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. Το μέγιστο πλήθος επαναλήψεων και για τους τρεις αλγορίθμους ορίστηκε η τιμή 80,000 ενώ το πλήθος των προτύπων ανα ψευδο-ομάδα ως 100 καθώς και ο συντελεστής momentum ως 0.9. Για τον αλγόριθμο ADAM οι τιμές των παραμέτρων ορίζονται σύμφωνα με τους συγγραφείς [13] του. Τέλος για τον προτεινόμενο αλγόριθμο οι τιμές των παραμέτρων γ_1 και γ_2 ορίζονται ως 10^{-2} και 10^{-3} αντίστοιχα ενώ για η παράμετρος $\gamma_3 = 10^{-2}$. Κάθε πείραμα επαναλήφθηκε 100 φορές ώστε να παραχθούν ασφαλή συμπεράσματα διεξάγοντας ένα μη-παραμετρικό τεστ σημαντικότητας Wilcoxon [134] στο ποσοστό του 5%.

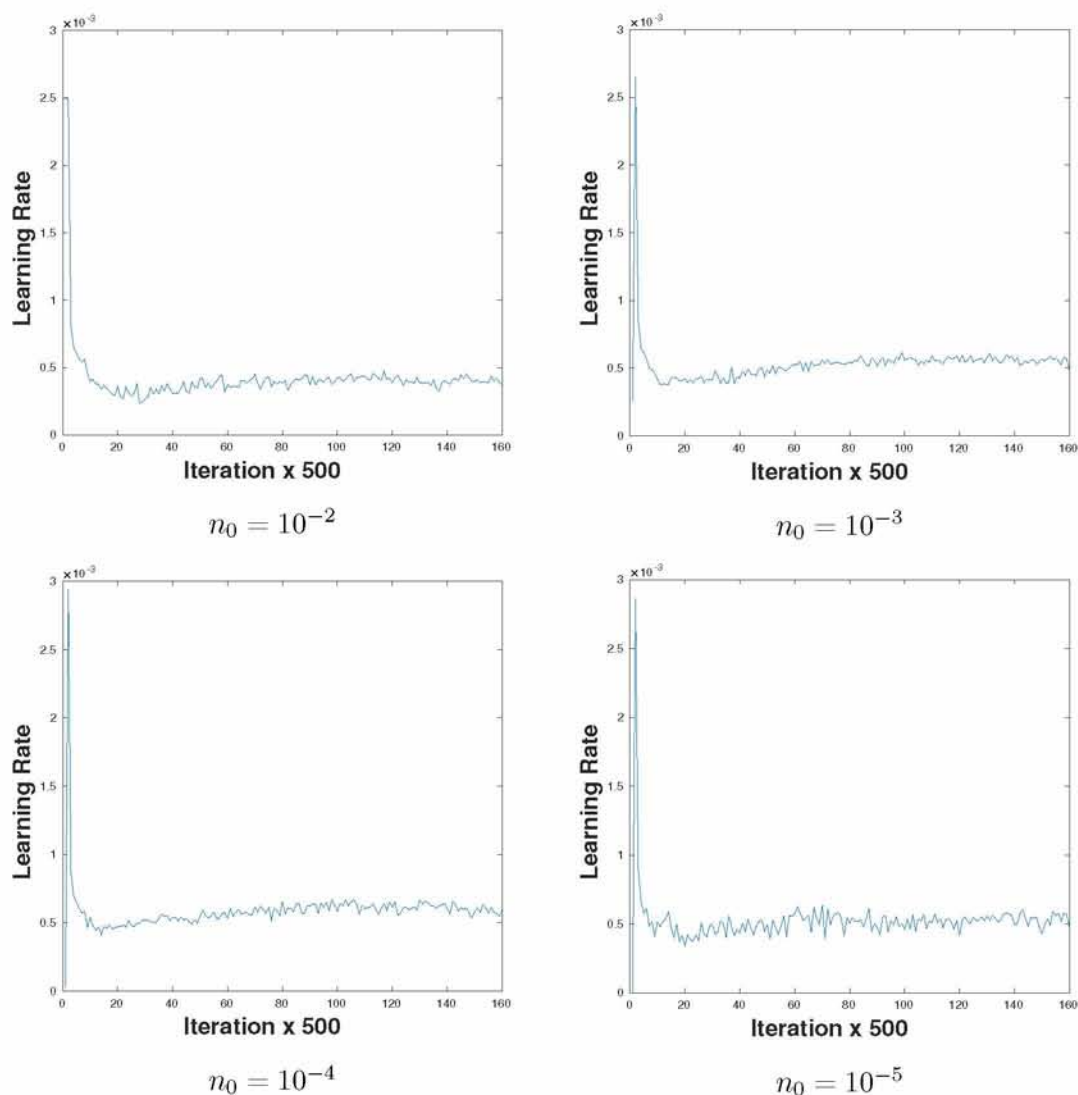
Στην Εικόνα 7.2 παρουσιάζεται η μέση απόδοση του κάθε αλγορίθμου στο σύνολο ελέγχου για τις τέσσερις διαφορετικές αρχικές τιμές του ρυθμού εκμάθησης, καθώς και η εξέλιξη της απόδοσής τους κατά την διάρκεια της εκπαίδευσης. Η μέση απόδοση του προτεινόμενου αλγορίθμου είναι καλύτερη ή ισοδύναμη με την απόδοση του αλγορίθμου ΣΑΚ καθώς και

του αλγορίθμου ADAM, ανεξαρτήτως της αρχικής τιμής του ρυθμού εκμάθησης. Όπως ήταν αναμενόμενο η απόδοση του αλγορίθμου ΣΑΚ επηρεάζεται σημαντικά από την επιλογή του ρυθμού εκμάθησης. Ο αλγόριθμος ADAM ενώ προσαρμόζει την διεύθυνση εύρεσης του ελαχίστου η επιλογή του αρχικού ρυθμού εκμάθησης είναι κρίσιμη όπως αναφέρεται και στην εργασία [13].



Σχήμα 7.2: Το ποσοστό επιτυχίας της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης.

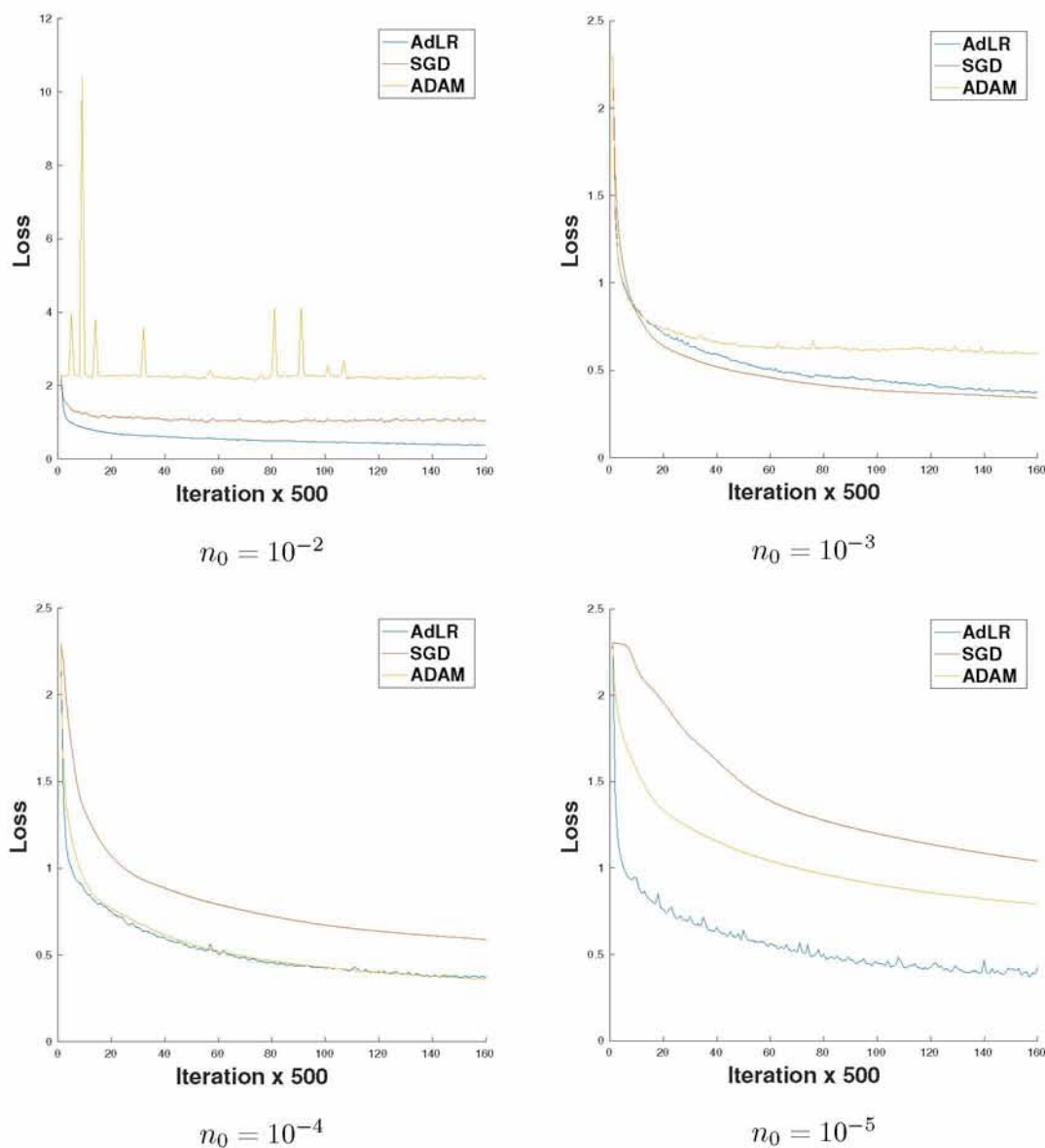
Στον Πίνακα 7.1 παρουσιάζονται οι εξής στατιστικές μετρικές: η μέση τιμή (μ) της απόδοσης για το κάθε πείραμα και η διασπορά (σ), καθώς και τα αποτελέσματα του τεστ Wilcoxon του προτεινόμενου αλγορίθμου προς τους αλγόριθμους ΣΑΚ και ADAM με τον συμβολισμό (\cdot/\cdot). Συγκεκριμένα, η μηδενική υπόθεση σε κάθε έλεγχο σημαντικότητας υποδεικνύει ότι τα δείγματα είναι ανεξάρτητα από πανομοιότυπες συνεχείς κατανομές με ίσες διαμέσους. Έτσι, με '+' επισημαίνονται οι περιπτώσεις, κατά τις οποίες η μηδενική υπόθεση απορρίπτεται στο 5% επίπεδο σημαντικότητας και ο αντίστοιχος προτεινόμενος αλγόριθμος παρουσιάζει ανώτερη απόδοση. Με '-' επισημαίνονται οι περιπτώσεις, κατά τις οποίες η μηδενική υπόθεση απορρίπτεται στο 5% επίπεδο σημαντικότητας και ο αντίστοιχος προτεινόμενος αλγόριθμος



Σχήμα 7.3: Η προσαρμογή του ρυθμού εκμάθησης μέσω του προτεινόμενου αλγορίθμου χρησιμοποιώντας διαφορετικές αρχικές τιμές.

	AdLR		ΣΑΚ		ADAM	
	μ (%)	σ	μ (%)	σ	μ (%)	σ
$n_0 = 10^{-2}$	78.4 (+/+)	0.7	62.3	1.9	17.2	8.3
$n_0 = 10^{-3}$	78.4 (=/+)	0.5	78.9	0.6	74.7	1.0
$n_0 = 10^{-4}$	78.3 (+/=)	0.7	75.0	0.3	78.1	0.5
$n_0 = 10^{-5}$	78.0 (+/+)	0.5	59.3	0.4	68.7	0.5

Πίνακας 7.1: Η μέση τιμή επιτυχίας και η διασπορά της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης, για το πρόβλημα CIFAR-10.



Σχήμα 7.4: Το σφάλμα εκπαίδευσης της προτεινόμενης μεθόδου έναντι των μεθόδων ADAM και ΣΑΚ με σταθερό ρυθμό εκμάθησης χρησιμοποιώντας διαφορετικές αρχικές τιμές εκμάθησης.

παρουσιάζει κατώτερη απόδοση, ενώ με '=' επισημαίνονται οι περιπτώσεις, κατά τις οποίες οι επιδόσεις δεν παρουσιάζουν στατιστικά σημαντική διαφορά.

Ο προτεινόμενος αλγόριθμος προσαρμόζει το ρυθμό εκμάθησης σε κάθε επανάληψη, στην Εικόνα 7.3 παρουσιάζεται η μέση τιμή του ρυθμού εκμάθησης κατά την διάρκεια εκπαίδευσης του ΤΝΔΣ για διαφορετικές αρχικές τιμές. Φαίνεται ότι ανεξάρτητα από την επιλογή της αρχικής τιμής ο ρυθμός εκμάθησης προσαρμόζεται με παρόμοιο τρόπο κάθε φορά. Ενώ, αντίστοιχα φαίνεται στην Εικόνα 7.4 η παρόμοια προσαρμογή της τιμής του σφάλματος του μοντέλου κατά την διάρκεια εκπαίδευσής του.

7.4 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε ένας νέος αλγόριθμος εκπαίδευσης Νευρωνικών Δικτύων Συνελίξεων μεταβλητού ρυθμού εκμάθησης. Οι αλγόριθμοι εκπαίδευσης των Τεχνητών Νευρωνικών Δικτύων Συνελίξεων ακόμα και αλγόριθμοι σημεία αναφοράς που μεταβάλλουν τον ρυθμό εκμάθησης κατά την διαδικασία της εκπαίδευσής τους [13], φαίνεται να επηρεάζονται από την επιλογή της αρχικής τιμής του ρυθμού εκμάθησης. Η συνήθης διαδικασία των επαναλαμβανόμενων δοκιμών, τιμών του ρυθμού εκμάθησης είναι ιδιαίτερα δαπανηρή σε επίπεδο χρόνου και ενέργειας, χρησιμοποιώντας ακόμα και ισχυρές μονάδες επεξεργασίας γραφικών (που χρησιμοποιούνται στα Τεχνητά Νευρωνικά Δίκτυα Συνελίξεων).

Στην ενότητα αυτή προτάθηκε ένας αλγόριθμος μεταβλητού ρυθμού εκμάθησης αξιοποιώντας την πληροφορία της κλίσης των βαρών από τις τελευταίες τρεις επαναλήψεις κατά τη διαδικασία εκπαίδευσής του. Ο αλγόριθμος αξιολογήθηκε σε ένα γνωστό σύνολο δεδομένων και συγκρίθηκε με γνωστούς αλγορίθμους σταθερού ρυθμού αλλά και μεταβλητού ρυθμού εκμάθησης, μελετώντας διαφορετικές αρχικές τιμές ρυθμού εκμάθησης. Τα αριθμητικά αποτελέσματα έδειξαν ότι προτεινόμενος αλγόριθμος διατηρεί σταθερά ποσοστά επιτυχίας ανεξάρτητα της αρχικής επιλογής, παρουσιάζοντας ισοδύναμα ή καλύτερα ποσοστά επιτυχίας σε σχέση με τους υπόλοιπους.

Αξίζει να σημειώσουμε ότι ο αλγόριθμος αυτός μπορεί να προταθεί ως ένα μοντέλο επιλογής μιας καλής αρχικής τιμής του ρυθμού εκμάθησης για άλλους αλγορίθμους, καθώς προσαρμόζει πάντα την τιμή του ρυθμού εκμάθησης σε μια σταθερή τιμή, ικανή έτσι ώστε το μοντέλο να συγκλίνει σε μια καλή λύση.

Μέρος V

Συμπεράσματα - Βιβλιογραφία

Κεφάλαιο 8

Συμπεράσματα Διατριβής

Στο κεφάλαιο αυτό ανακεφαλαιώνουμε τα προβλήματα του πραγματικού κόσμου που μελετήσαμε κάνοντας χρήση συσκευών με δυνατότητες διασύνδεσης, καθώς και οι υπολογιστικές μέθοδοι που προτείναμε για την ενσωμάτωση 'ευφυίας' στις συσκευές αυτές, με σκοπό την αυτόματη αντιμετώπιση των συγκεκριμένων προβλημάτων.

Η συνεισφορά της παρούσας διδακτορικής διατριβής επικεντρώνεται στη θεωρητική θεμελίωση της επίλυσης προβλημάτων του τομέα της περιβάλλουσας νοημοσύνης και συγκεκριμένα την αναγνώριση βιολογικών δραστηριοτήτων και πρόδρομης ιατρικής διάγνωσης και παρακολούθησης του ασθενούς από γενικούς ιατρούς. Μέσα από την μελέτη των υπολογιστικών μεθόδων προτάθηκαν και αναπτύχθηκαν μεθοδολογίες και αλγόριθμοι στοχευμένοι για την αντιμετώπιση των συγκεκριμένων προβλημάτων όπως και τεχνικές που βρίσκουν εφαρμογή σε γενικού τύπου προβλήματα.

Ένα αντικείμενο της περιβάλλουσας νοημοσύνης είναι η ανάπτυξη ανθρωποκεντρικών υπολογιστών οι οποίοι να μπορούν να κατανοήσουν τη δραστηριότητα του ανθρώπου στο χώρο που κινείται, χωρίς ή με μικρή επέμβαση στο καθημερινό τρόπο ζωής τους. Στόχος τους είναι η αναγνώριση της δραστηριότητάς του και κατ' επέκταση ο εντοπισμός επειγόντων περιστατικών. Κάνοντας χρήση δικτυακών καμερών που είναι τοποθετημένες στο χώρο, είτε ασύρματων φορητών συσκευών καταγραφής κίνησης που είναι τοποθετημένες στον άνθρωπο, αναπτύξαμε μια σειρά μεθοδολογιών προσαρμοσμένων στις συσκευές για τον αυτόματο εντοπισμό της δραστηριότητας του ανθρώπου (Κεφάλαια 3, 4 και 5).

Τα Κεφάλαια 3 και 4, σχετίζονται με την αναγνώριση της στάσης του ανθρώπου, κάνοντας χρήση της σιλουέτας του, χωρίς επέμβαση στο καθημερινό τρόπο ζωής του. Αυτό επιτυγχάνεται κάνοντας μια διακριτική παρέμβαση στο χώρο του ανθρώπου (όχι στον ίδιο τον άνθρωπο), τοποθετώντας μια δικτυακή κάμερα και στη συνέχεια κάνοντας χρήση τεχνικών υπολογιστικής όρασης. Για την αποφυγή επικαλύψεων του οπτικού πεδίου της κάμερας και μείωση του κόστους κάνοντας χρήση πολλαπλών καμερών, χρησιμοποιείται μια ευρυγώνια κάμερα τοποθετημένη στην οροφή του χώρου. Για την αναγνώριση της στάσης του ανθρώπου θα χρησιμοποιήσουμε έναν περιγραφέα χαρακτηριστικών σε συνδυασμό με έναν αλγόριθμο ταξινόμησης. Το πρόβλημα που δημιουργείται λόγω της ευρυγώνιας κάμερας είναι ότι η πλειοψηφία των περιγραφέντων χαρακτηριστικών χρησιμοποιούν την Ευκλείδεια απόσταση ως μετρική, που στη

συγκεκριμένη περίπτωση δεν είναι η βέλτιστη λόγω του σφαιρικού οπτικού αντικειμένου που χρησιμοποιείται από την ευρυγώνια κάμερα.

Έτσι στο Κεφάλαιο 3 προτείνεται ένας περιγραφέας χαρακτηριστικών που βασίζεται στη χρήση της γεωδαιτικής απόστασης. Συγκεκριμένα επιλέγοντας την μέθοδο ροπών Zernike λόγω της ιδιότητάς τους να παράγουν ολικά χαρακτηριστικά αναλλοίωτα στις αλλαγές θέσης του αντικειμένου, θα την τροποποιήσουμε κατάλληλα κάνοντας χρήση της γεωδαιτικής μετρικής. Ο προτεινόμενος περιγραφέας χαρακτηριστικών (ΓΔΡΖ) μελετήθηκε και εξετάστηκε σε πραγματικές εικόνες της ευρυγώνιας κάμερας, αφού πρώτα τμηματοποιήθηκαν με την βοήθεια καθιερωμένου αλγορίθμου τμηματοποίησης. Ο ΓΔΡΖ φάνηκε να παρουσιάζει ανοχή σε ξαφνικές παραμορφώσεις της ευρυγώνιας εικόνας, λόγω του αλγορίθμου τμηματοποίησης έναντι των κλασικών ροπών Zernike. Η χρήση των τιμών του πραγματικού και φανταστικού μέρους των μιγαδικών χαρακτηριστικών που παράγει, έδειξε καλύτερα αποτελέσματα στο τμήμα της αναγνώρισης αντικειμένων έναντι του μέτρου και της φάσης τους αλλά και τον συνδυασμό αυτών.

Στη συνέχεια στο επόμενο Κεφάλαιο ολοκληρώσαμε την μελέτη του προβλήματος αναγνώρισης της ανθρώπινης στάσης, χρησιμοποιώντας μια υβριδική αρχιτεκτονική ΝΔΣ, το ΕΝΔΣ που συνδυάζει την τοπική πληροφορία που δημιουργεί στα ενδιάμεσα επίπεδά του καθώς εκπαιδεύεται με μια εικόνα, με την πληροφορία των ολικών περιγραφέων των ΓΔΡΖ. Για την εκπαίδευση του ΕΝΔΣ καθώς απαιτεί μεγάλο πλήθος δεδομένων, δημιουργήσαμε ένα συνθετικό σύνολο δεδομένων από τρισδιάστατα ανθρώπινα μοντέλα σε τρεις γενικές ανθρώπινες στάσεις εκπαιδεύοντας το ΕΝΔΣ με αυτό. Στη συνέχεια χρησιμοποιώντας λίγα στιγμιότυπα από πραγματικές ανθρώπινες σιλουέτες και εφαρμόζοντας την τεχνική μεταφοράς γνώσης, εκπαιδεύσαμε για λίγες επαναλήψεις το εκπαιδευμένο μοντέλο με συνθετικές σιλουέτες. Ελέγχοντας το τελικό εκπαιδευμένο μοντέλο με συνεχή στιγμιότυπα ενός βίντεο με πραγματικές ανθρώπινες στάσεις, πετύχαμε πολύ μεγαλύτερα ποσοστά επιτυχίας έναντι των τυπικών ΝΔΣ αλλά και άλλων γνωστών μεθόδων αναγνώρισης δραστηριότητας από στατική κάμερα.

Στο Κεφάλαιο 5 μελετήσαμε τον εντοπισμό της δραστηριότητας του ανθρώπου σε εξωτερικούς χώρους όπου η χρήση κάμερας δεν είναι δυνατή. Σκοπός είναι η αξιοποίηση των φορετών συσκευών (κινητού τηλεφώνου) και αισθητήρων (επιταχυνσιόμετρου κ.α.) που διαθέτουν για τον εντοπισμό πιθανής πτώσης του ανθρώπου, ως κύριος μηχανισμός μια κατηγορίας προβλημάτων της περιβάλλουσας νοημοσύνης, του εντοπισμού επειγόντων περιστατικών. Συγκεκριμένα, αναπτύχθηκε μια μεθοδολογία χαμηλού υπολογιστικού κόστους αξιοποίησης δεδομένων μεγάλης διάστασης, με σκοπό την έγκαιρη και έγκυρη αναγνώριση πιθανής πτώσης του ανθρώπου. Η μεθοδολογία αυτή αξιοποιεί την ανάλυση διαδοχικών κυρίων συνιστωσών χωρίς τον υπολογισμό του μητρώου συνδιασπορών για τη μετατροπή πολυδιάστατων σημάτων σε μονοδιάστατα σήματα. Στη συνέχεια αξιοποιεί τον αλγόριθμο συσσωρευτικού αθροίσματος για τον εντοπισμό αλλαγών στο μονοδιάστατο σήμα αυτό. Η μεθοδολογία φάνηκε να εντοπίζει γρήγορα και με ακρίβεια αλλαγές στο μετασχηματισμένο μονοδιάστατο σήμα από την κίνηση του ανθρώπου. Βασίζεται σε αλγορίθμους χαμηλού υπολογιστικού κόστους που μπορούν να εφαρμοστούν σε χαμηλής ισχύος συσκευές. Ενώ, η πολύ σημαντική συνεισφορά

του είναι η χρήση σημάτων από πολλές πηγές και όχι μόνο από σήματα τριών διαστάσεων που χρησιμοποιούν οι καθιερωμένες μέθοδοι.

Στο Κεφάλαιο 6 παρουσιάστηκε το πρόβλημα αυτόματου εντοπισμού παθήσεων του γαστρεντερικού σωλήνα κάνοντας χρήση ασύρματης ενδοσκοπικής κάμερας. Προτάθηκε μια ολιστική μεθοδολογία χαρακτηρισμού μιας γαστρεντερικής εικόνας ως προς το αν εμφανίζει τυχόν παθολογία του γαστρεντερικού σωλήνα καθώς και επισήμανση των περιοχών αυτών. Η μεθοδολογία στηρίχθηκε στην εκπαίδευση ΑΝΔΣ με δεδομένα ασθενώς χαρακτηρισμένα, επιτυγχάνοντας μεγάλη ακρίβεια τόσο στον χαρακτηρισμό της εικόνας όσο και για στον εντοπισμό της περιοχής της παθολογίας. Για τον εντοπισμό των περιοχών έγινε χρήση σημαντικών σημείων χωρίς καμία πρότερη γνώση των περιοχών αυτών, αλλά αξιοποιώντας την πληροφορία που δημιουργήσαν τα ΑΝΔΣ κατά την φάση της εκπαίδευσής τους. Η μεθοδολογία συγκρίθηκε με άλλες γνωστές μεθόδους και σύνολα δεδομένων καταφέροντας πολύ υψηλά ποσοστά επιτυχίας σε όλες τις μετρικές αξιολόγησης.

Η παρουσίαση του ερευνητικού μέρους αυτής της διατριβής ολοκληρώνεται με το Κεφάλαιο 7 στο οποίο παρουσιάστηκε μια μέθοδος προσαρμογής του ρυθμού εκμάθησης των ΝΔΣ. Συγκεκριμένα, προτάθηκε η προσαρμογή του ρυθμού εκμάθησης, αξιοποιώντας την πληροφορία της κλίσης των βαρών των τριών τελευταίων επαναλήψεων του αλγορίθμου εκπαίδευσης ενός ΝΔΣ. Τα πειραματικά αποτελέσματα της προτεινόμενης μεθόδου έδειξαν σταθερή σύγκλιση του μοντέλου, χρησιμοποιώντας ένα μεγάλο εύρος αρχικών τιμών του ρυθμού εκμάθησης. Συγκριτικά, με άλλες καθιερωμένες μεθόδους προσαρμογής του ρυθμού εκμάθησης η προτεινόμενη μεθοδολογία παρουσίασε ισοδύναμα ή καλύτερα ποσοστά επιτυχίας σε καθιερωμένο ανοιχτό σύνολο δεδομένων.

Βιβλιογραφία

- [1] <http://www.mathworks.com/matlabcentral/fileexchange/>
- [2] <http://resources.mpi-inf.mpg.de/scandb/>
- [3] <http://archive3d.net/?tag=People>
- [4] UN 2002. Population aging. un department of economic social affairs, population division. <http://www.un.org/esa/population/publications/ageing/graph.pdf>
- [5] UN 2004. World population to 2300. un department of economic social affairs, population division, united nations, new york. <http://www.un.org/esa/population/publications/longrange2/worldpop2300final.pdf>
- [6] D. Alemayehu and K. H. Zou. Applications of roc analysis in medical research: recent developments and future directions. *Academic radiology*, 19(12):1457–1464
- [7] Luis B. Almeida, Thibault Langlois, Jose D. Amaral and Alexander Plakhov. Parameter adaptation in stochastic optimization. In *In On-line Learning in Neural Networks*, pages 111–134. Cambridge University Press, 1998.
- [8] Juan C. Augusto and Chris D. Nugent. The use of temporal reasoning and management of complex events in smart homes. In *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'04*, pages 778–782, Amsterdam, The Netherlands, The Netherlands, 2004. IOS Press.
- [9] Juan C. Augusto and Chris D. Nugent. *Smart Homes Can Be Smarter*, pages 1–15. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [10] Juan Carlos Augusto, Jun Liu, Paul McCullagh, Hui Wang and Jian Bo Yang. Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a smart home. *International Journal of Computational Intelligence Systems*, 1(4):361–378, 2008.
- [11] S. Bahadori, A. Cesta, G. Grisetti, L. Iocchi, R. Leone, D. Nardi, A. Oddi, F. Pecora and R. Rasconi. Robocare: Pervasive intelligence for the domestic care of the elderly

- [12] Shahram Bahadori, Amedeo Cesta, Luca Iocchi, G. R. Leone, Daniele Nardi, Federico Pecora, Riccardo Rasconi and L. Scozzafava. Towards ambient intelligence for the domestic care of the elderly. In *Ambient Intelligence: A Novel Paradigm*, 2004.
- [13] Jimmy Ba and Diederik Kingma. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.
- [14] J. Bernal, N. Tajkbaksh, F. J. Sánchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Córdova, C. Sánchez-Montes, S. R. Gurudu, G. Fernández-Esparrach, X. Dray, J. Liang and A. Histace. Comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge. *IEEE Transactions on Medical Imaging*, 36(6):1231–1249, 2017.
- [15] J. Bernal, N. Tajkbaksh, F. J. Sánchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Córdova, C. Sánchez-Montes, S. R. Gurudu, G. Fernández-Esparrach, X. Dray, J. Liang and A. Histace. Comparative validation of polyp detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision challenge. *IEEE Transactions on Medical Imaging*, 36(6):1231–1249, 2017.
- [16] L. Bottou. *On-line Learning in Neural Networks*. Cambridge University Press, New York, NY, USA, 1998.
- [17] Joan Bruna, Wojciech Zaremba, Arthur Szlam and Yann LeCun. Spectral networks and locally connected networks on graphs. *CoRR*, abs/1312.6203, 2013.
- [18] A. Cesta, G. Cortellessa, F. Pecora and R. Rasconi. Exploiting Scheduling Techniques to Monitor the Execution of Domestic Activities. *Intelligenza Artificiale (Italian Journal of Artificial Intelligence)*, 2005.
- [19] Jie Zhi Cheng, Dong Ni, Yi Hong Chou, Jing Qin, Chui Mei Tiu, Yeun Chung Chang, Chiun Sheng Huang, Dinggang Shen and Chung Ming Chen. Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific Reports*, 6, 2016.
- [20] Honghan Chen, Xiao Wu, Gan Tao and Qiang Peng. Automatic content understanding with cascaded spatial–temporal deep framework for capsule endoscopy videos. *Neurocomputing*, 229:77 – 87, 2017.
- [21] M. . Chen. A notion for machine learning: knowledge developability. In *[1988] Proceedings of the Twenty-First Annual Hawaii International Conference on System Sciences. Volume III: Decision Support and Knowledge Based Systems Track*, volume 3, pages 274–279, 1988.

- [22] Yang Cong, Shuai Wang, Baojie Fan, Yunsheng Yang and Haibin Yu. Udfs: Un-supervised deep sparse feature selection. *Neurocomputing*, 196(Συμπλημεντ ^):150 – 158, 2016.
- [23] Yang Cong, Shuai Wang, Ji Liu, Jun Cao, Yunsheng Yang and Jiebo Luo. Deep sparse feature selection for computer aided endoscopy diagnosis. *Pattern Recognition*, 48(3):907 – 917, 2015.
- [24] Diane J. Cook, Michael Youngblood and Sajal K. Das. *A Multi-agent Approach to Controlling a Smart Environment*, pages 165–182. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [25] J. M. Corchado, J. Bajo and A. Abraham. Gerami: Improving healthcare delivery in geriatric residences. *IEEE Intelligent Systems*, 23(2):19–25, 2008.
- [26] X. Cui, V. Goel and B. Kingsbury. Data augmentation for deep convolutional neural network acoustic modeling. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4545–4549, 2015.
- [27] K.K. Delibasis, Spiros Georgakopoulos, Vassilis Plagianakos and Ilias Maglogiannis. Calculation of complex zernike moments with geodesic correction for pose recognition in omni-directional images. In *Artificial Intelligence Applications and Innovations* Lazaros Iliadis, Ilias Maglogiannis and Harris Papadopoulos, editors, volume 436 in *IFIP Advances in Information and Communication Technology*, pages 375–384. Springer Berlin Heidelberg, 2014.
- [28] K.K. Delibasis, V.P. Plagianakos and I. Maglogiannis. Refinement of human silhouette segmentation in omni-directional indoor videos. *Computer Vision and Image Understanding*, 128:65 – 83, 2014.
- [29] Konstantinos K. Delibasis, Spiros V. Georgakopoulos, Konstantina Kottari, Vassilis P. Plagianakos and Ilias Maglogiannis. Geodesically-corrected zernike descriptors for pose recognition in omni-directional images. *Integrated Computer-Aided Engineering*, 23(2):185–199, 2016.
- [30] Kostas Delibasis, Vassilis P. Plagianakos and Ilias Maglogiannis. Refinement of human silhouette segmentation in omni-directional indoor videos. *Computer Vision and Image Understanding*, 128:65–83, 2014.
- [31] Kostas Delibasis, Vassilis P. Plagianakos and Ilias Maglogiannis. Refinement of human silhouette segmentation in omni-directional indoor videos. *Computer Vision and Image Understanding*, 128:65–83, 2014.
- [32] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

- [33] C. Doukas, L. Athanasiou, K. Fakos and I. Maglogiannis. Advanced sound and distress speech expression classification for human status awareness in assistive environments. *The Journal on Information Technology in Healthcare*, 7(2):111–117, 2009.
- [34] Charalampos Doukas and Ilias Maglogiannis. Advanced classification and rules-based evaluation of motion, visual and biosignal data for patient fall incident detection. *International Journal on Artificial Intelligence Tools*, 19(2):175–191, 2010.
- [35] K. DUCATEL, M. BOGDANOWICZ, F. SCAPOLO, J. LEIJTEN and J. C. BURGELMAN. Scenarios for ambient intelligence in 2010. ist advisory group final report, european commission, ec. brussels., 2001.
- [36] K. DUCATEL, M. BOGDANOWICZ, F. SCAPOLO, J. LEIJTEN and J. C. BURGELMAN. Ambient intelligence : From vision to reality. ist advisory group draft report, european commission., 2003.
- [37] John Duchi, Elad Hazan and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, 2011.
- [38] José L. Encarnação and Thomas Kirste. *Ambient Intelligence: Towards Smart Appliance Ensembles*, pages 261–270. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [39] S. Eom and E. Kim. A survey of decision support system applications (1995–2001). *Journal of the Operational Research Society*, 57(11):1264–1278, 2006.
- [40] S B Eom, S M Lee, E B Kim and C. Somarajan. A survey of decision support system applications (1988–1994). *Journal of the Operational Research Society*, 49(2):109–120, 1998.
- [41] J. Flusser, B. Zitova and T. Suk. Invariant-based registration of rotated and blurred images. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*, volume 2, pages 1262–1264 ολ.2, 1999.
- [42] Kunihiro Fukushima and Sei Miyake. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, 15(6):455 – 469, 1982.
- [43] Christopher W. Geib. Problems with intent recognition for elder care. In *Proceedings of the AAI Workshop on Automation as Caregiver*, 2002.
- [44] Christopher W. Geib and Robert P. Goldman. Probabilistic plan recognition for hostile agents. In *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference*, pages 580–584. AAAI Press, 2001.

- [45] S. V. Georgakopoulos, D. K. Iakovidis, M. Vasilakakis, V. P. Plagianakos and A. Koulaouzidis. Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 510–514, 2016.
- [46] S. V. Georgakopoulos, K. Kottari, K. Delibasis, V. P. Plagianakos and I. Maglogiannis. Improving the performance of convolutional neural network for skin image classification using the response of image analysis filters. *Neural Computing and Applications*, 2018.
- [47] S. V. Georgakopoulos and V. P. Plagianakos. A novel adaptive learning rate algorithm for convolutional neural network training. In *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings*, pages 327–336, Cham, 2017. Springer International Publishing.
- [48] S. V. Georgakopoulos, S. K. Tasoulis and V. P. Plagianakos. Efficient change detection for high dimensional data streams. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2219–2222, 2015.
- [49] Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Ilias Maglogiannis and Vassilis P. Plagianakos. On-line fall detection via mobile accelerometer data. In *proceeding of the 11th International Conference on Artificial Intelligence Applications and Innovations, (IFIP AIAI 2015)*, 2015, το αππεαρ.
- [50] Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Aristidis G. Vrahatis and Vassilis P. Plagianakos. Convolutional neural networks for toxic comment classification. *CoRR*, abs/1802.09957, 2018.
- [51] S.V. Georgakopoulos, K. Kottari, K. Delibasis, V.P. Plagianakos and I. Maglogiannis. Pose recognition using convolutional neural networks on omni-directional images. *Neurocomputing*, 280:23 – 31, 2018.
- [52] Ross B. Girshick. Fast R-CNN. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 1440–1448, 2015.
- [53] Flore Gouaux, Lucas Simon-Chautemps, Jocelyne Fayn, Stefano Adami, Mohammad Arzi, Deodato Assanelli, MC Forlini, Cesare Malossi, Andrés Martínez, Joël Placide, GL Ziliani and Paul Rubel. Ambient intelligence and pervasive systems for the monitoring of citizens at cardiac risk : New solutions from the epi-medics project. *Comput. Cardiol.*, 29:289–292, 2002.
- [54] Priya Goyal, Piotr Dollár, Ross B. Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia and Kaiming He. Accurate, large minibatch SGD: training imagenet in 1 hour. *CoRR*, abs/1706.02677, 2017.

- [55] Crina Grosan and Ajith Abraham. *Rule-Based Expert Systems*, pages 149–185. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [56] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, 2003.
- [57] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen and H. Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1823–1830, 2010.
- [58] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen and H. Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1823–1830, 2010.
- [59] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly. *Image Features Detection, Description and Matching*, pages 11–45. Springer International Publishing, Cham, 2016.
- [60] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2ηδη έκδοση, 1998.
- [61] Kaiming He, Georgia Gkioxari, Piotr Dollár and Ross B. Girshick. Mask r-cnn. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.
- [62] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [63] Ming Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.
- [64] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis and V. P. Plagianakos. Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification. *IEEE Transactions on Medical Imaging*, 37(10):2196–2210, 2018.
- [65] D. K. Iakovidis and A. Koulaouzidis. Automatic lesion detection in wireless capsule endoscopy: A simple solution for a complex problem. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2236–2240, 2014.
- [66] Dimitris K. Iakovidis and Anastasios Koulaouzidis. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nature Reviews Gastroenterology and Hepatology*, 12:172 – 186, 2015.

- [67] X. Jia and M. Q. H. Meng. A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 639–642, 2016.
- [68] X. Jia and M. Q. H. Meng. A study on automated segmentation of blood regions in wireless capsule endoscopy images using fully convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 179–182, 2017.
- [69] Y. Jia. Polyps auto-detection in wireless capsule endoscopy images using improved method based on image segmentation. In *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1631–1636, 2015.
- [70] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 675–678, New York, NY, USA, 2014. ACM.
- [71] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [72] O. L. Junior, D. Delgado, V. Goncalves and U. Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6, 2009.
- [73] A. Karargyris and N. Bourbakis. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE Transactions on Biomedical Engineering*, 58(10):2777–2786, 2011.
- [74] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras and M. Tzivras. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3):141–152, 2003.
- [75] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *ICLR*, 2017.
- [76] S. K. Khaitan and J. D. McCalley. Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal*, 9(2):350–365, 2015.
- [77] Shadi Khawandi, Bassam Daya and Pierre Chauvet. Real time recognition of elderly daily activity using fuzzy logic through fusion of motion and location data. *International Journal of Computer Applications*, 54(3):55–60, 2012.

- [78] Spiros V. kopoulos, Konstantina Kottari, Kostas Delibasis, Vassilis P. Plagianakos and Ilias Maglogiannis. Convolutional neural networks for pose recognition in binary omni-directional images. In *Artificial Intelligence Applications and Innovations - 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016, Thessaloniki, Greece, September 16-18, 2016, Proceedings*, pages 106–116, 2016.
- [79] Anastasios Koulaouzidis, Dimitris K. Iakovidis, Diana E. Yung, Emanuele Rondonotti, Uri Kopylov, J N Plevris, Ervin Toth, Abraham Rami Eliakim, Gabrielle Wurm Johansson, Wojciech Marlicz, Georgios Mavrogenis, Artur Nemeth, Henrik Thorlacius and Gian Eugenio Tontini. Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. In *Endoscopy international open*, volume 5, pages E477–E483
- [80] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, editors, pages 1097–1105. Curran Associates, Inc., 2012.
- [81] Yann LeCun, Patrick Haffner, Léon Bottou and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*, pages 319–, London, UK, UK, 1999. Springer-Verlag.
- [82] M. Liedlgruber and A. Uhl. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. *IEEE Reviews in Biomedical Engineering*, 4:73–88, 2011.
- [83] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng and M. Chen. Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, pages 844–848, 2014.
- [84] I. Maglogiannis, C. Ioannou and P. Tsanakas. Fall detection and activity identification using wearable and hand-held devices. *Integrated Computer Aided Engineering*, 23(2):161–172, 2016.
- [85] Ilias Maglogiannis, Kostas Delibasis, Dimitrios Kosmopoulos, Theodosios Goudas and Charalampos Doukas. Activity recognition in assistive environments: The sthenos approach. In *Universal Access in Human-Computer Interaction. Aging and Assistive Environments*Constantine Stephanidis and Margherita Antona, editors, pages 525–536, Cham, 2014. Springer International Publishing.
- [86] G.D. Magoulas, V.P. Plagianakos and M.N. Vrahatis. Adaptive stepsize algorithms for on-line training of neural networks. *Nonlinear Analysis: Theory, Methods & Applications*, 47(5):3425 – 3430, 2001.

- [87] A. V. Mamonov, I. N. Figueiredo, P. N. Figueiredo and Y. H. Richard Tsai. Automated polyp detection in colon capsule endoscopy. *IEEE Transactions on Medical Imaging*, 33(7):1488–1502, 2014.
- [88] I. Melzer, N. Benjuya and J. Kaplanski. Postural stability in the elderly: a comparison between fallers and non-fallers. *Age Ageing*, 33:602–607, 2004.
- [89] S.G. Miaou, Pei Hsu Sung and Chia Yuan Huang. A customized human fall detection system using omni-camera images and personal information. In *Distributed Diagnosis and Home Healthcare*, pages 39–42. IEEE Computer Society, 2006.
- [90] Romano MJ and Stafford RS. Electronic health records and clinical decision support systems: Impact on national ambulatory care quality. *Archives of Internal Medicine*, 171(10):897–903, 2011.
- [91] R. Mukundan, S. H. Ong and P. A. Lee. Image analysis by tchebichef moments. *IEEE Transactions on Image Processing*, 10(9):1357–1364, 2001.
- [92] Yurii Nesterov. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.
- [93] Marketta Niemelä, Rafael Gonzalez Fuentetaja, Eija Kaasinen and Jorge Lorenzo Gallardo. Supporting independent living of the elderly with mobile-centric ambient intelligence: User evaluation of three scenarios. In *Ambient Intelligence* Bernt Schiele, Anind K. Dey, Hans Gellersen, Borisde Ruyter, Manfred Tscheligi, Reiner Wichert, Emile Aarts and Alejandro Buchmann, editors, pages 91–107, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [94] Erkki Oja. *Subspace methods of pattern recognition*. Letchworth, Hertfordshire, England : Research Studies Press ; New York : Wiley, 1983.
- [95] Erkki Oja and Juha Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 106:69–84, 1985.
- [96] Terence Ong, Wei Tan, Lindsay Marshall and Opinder Sahota. The relationship between socioeconomic status and fracture in a fracture clinic setting: Data from the nottingham fracture liaison service. *Injury*, in press.
- [97] M. Oquab, L. Bottou, I. Laptev and J. Sivic. Is object localization for free? - weakly-supervised learning with convolutional neural networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 685–694, 2015.
- [98] E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1/2):100–115, 1954.

- [99] G. Papandreou, L. Chen, K. P. Murphy and A. L. Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1742–1750, 2015.
- [100] V. P. Plagianakos, G. D. Magoulas and M. N. Vrahatis. Global learning rate adaptation in on-line neural network training. In *Proceedings of the Second International ICSC Symposium on Neural Computation (NC'2000), Berlin, Germany,*, 2000.
- [101] R. Poppe. Evaluating Example-based Pose Estimation: Experiments on the HumanEva Sets. In *CVPR 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2)*,, 2007.
- [102] D. Psaltis and Y. S. Abu-Mostafa. Recognitive aspects of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:698–706, 1984.
- [103] Ciprian Radu Rad, Olimpiu HancuandIoana Alexandra Takacs and Gheorghe Olteanu. Smart monitoring of potato crop: A cyber-physical system architecture model in the field of precision agriculture. *Agriculture and Agricultural Science Procedia*, 6:73–79, 2015.
- [104] Mohammad Raza Perwez, Naveed Ahmad, Muhammad Sajid Javaid and Muhammad Ehsanul Haq. A critical analysis on efficacy of clinical decision support systems in health care domain. In *Manufacturing Science and Technology, ICMST2011*, volume 383 in *Advanced Materials Research*, pages 4043–4050. Trans Tech Publications, 2012.
- [105] K. Rentto, A. Vaatanen and R. Lappalainen. Wireless wellness monitor ii, 2002.
- [106] Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra, Davide Anguita, Joan Cabe-stany and Andreu Català. Human activity and motion disorder recognition: towards smarter interactive cognitive environments. In *21st European Symposium on Artificial Neural Networks, ESANN 2013, Bruges, Belgium, April 24-26, 2013*, 2013.
- [107] E. Ribeiro, A. Uhl and M. Häfner. Colonic polyp classification with convolutional neural networks. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 253–258, 2016.
- [108] Giuseppe Riva. Ambient intelligence in health care. *CyberPsychology & Behavior*, 6(3):295–300, 2003.
- [109] M. D. Rodriguez, J. Favela, E. A. Martinez and M. A. Munoz. Location-aware access to hospital information and services. *IEEE Transactions on Information Technology in Biomedicine*, 8(4):448–455, 2004.

- [110] D. E. Rumelhart, G. E. Hinton and R. J. Williams. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. *Ήαπτερ Λεαρνινγ Ιντερναλ Ρεπρεσεντατιονς βψ Ερρορ Προπαγατιον*, pages 318–362. MIT Press, Cambridge, MA, USA, 1986.
- [111] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3ρδη εδιτιον, 2009.
- [112] H.G. Sandalidis and I. Maglogiannis. Chapter 6.2 - channel models for on-body communications. In *Wearable Sensors* Edward Sazonov and Michael R. Neuman, editors, pages 453 – 473. Academic Press, Oxford, 2014.
- [113] Terence D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.
- [114] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak and Albert Cardona.
- [115] M. Schlemmer, M. Heringer, F. Morr, I. Hotz, M. Hering-Bertram, C. Garth, W. Kollmann, B. Hamann and H. Hagen. Moment invariants for the analysis of 2d flow fields. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1743–1750, 2007.
- [116] Santi Segu, Michal Drozdal, Guillem Pascual, Petia Radeva, Carolina Malagelada, Fernando Azpiroz and Jordi Vitri. Generic feature learning for wireless capsule endoscopy analysis. *Comput. Biol. Med.*, 79(ϕ):163–172, 2016.
- [117] A. K. Sekuboyina, S. T. Devarakonda and C. S. Seelamantula. A convolutional neural network approach for abnormality detection in wireless capsule endoscopy. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 1057–1060, 2017.
- [118] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 3(52), 1965.
- [119] Nazmul Siddique and Hojjat Adeli. *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*. John Wiley & Sons, 2013.
- [120] Barbara Korouvi Seljak Simon Mezgec. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. In *Nutrients*, 2017.
- [121] E. Simon. High tech buildings use sunlight, sea water to save energy, 2004.

- [122] Heung Il Suk, Seong Whan Lee and Dinggang Shen. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569 – 582, 2014.
- [123] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.
- [124] Nima Tajbakhsh, Gurudu Suryakanth R. and Jianming Liang. volume 2015-Θυλψ, pages 79–83. IEEE Computer Society, 2015.
- [125] S. K. Tasoulis, C. N. Doukas, V. P. Plagianakos and I. Maglogiannis. Statistical data mining of streaming motion data for activity and fall recognition in assistive environments. *Neurocomputing*, 107:87–96, 2013.
- [126] Michael Reed Teague. Image analysis via the general theory of moments*. *J. Opt. Soc. Am.*, 70(8):920–930, 1980.
- [127] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Elsevier/Academic Press,, 2008.
- [128] Antonio Torralba, Rob Fergus and William T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(11):1958–1970, 2008.
- [129] E. Turban. Implementing decision support systems: a survey. In *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, volume 4, pages 2540–2545 ολ.4, 1996.
- [130] F. Van Der Sommen, S. Zinger, E. J. Schoon and P. H. N. De With. Supportive automatic annotation of early esophageal cancer using local gabor and color features. *Neurocomput.*, 144:92–106, 2014.
- [131] Michael Vasilakakis, Dimitrios K. Iakovidis, Evaggelos Spyrou and Anastasios Koulaouzidis. *Weakly-Supervised Lesion Detection in Video Capsule Endoscopy Based on a Bag-of-Colour Features Model*, pages 96–103. Springer International Publishing, Cham, 2017.
- [132] Juyang Weng, Yilu Zhang and Weyshuan Hwang. Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25:1034–1040, 2003.
- [133] J. West, J. Hippisley-Cox, C. A. Coupland, G. M. Price, L. M. Groom, D. Kendrick and E. Webber. Do rates of hospital admission for falls and hip fracture in elderly people vary by socio-economic status? *Public Health*, 118:576–581, 2004.

- [134] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.
- [135] G. Wimmer, A. Vécsei and A. Uhl. Cnn transfer learning for the automated diagnosis of celiac disease. In *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2016.
- [136] Y. Yuan, B. Li and M. Q. . Meng. Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. *IEEE Transactions on Automation Science and Engineering*, 13(2):529–535, 2016.
- [137] Y. Yuan, B. Li and M. Q. H. Meng. Bleeding frame and region detection in the wireless capsule endoscopy video. *IEEE Journal of Biomedical and Health Informatics*, 20(2):624–630, 2016.
- [138] Matthew D. Zeiler. Adadelta: An adaptive learning rate method. *CoRR*, abs/1212.5701, 2012.
- [139] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer Vision – ECCV 2014* David Fleet, Tomas Pajdla, Bernt Schiele and Tinne Tuytelaars, editors, pages 818–833, Cham, 2014. Springer International Publishing.
- [140] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. W. Lau and C. C. Y. Poon. Automatic detection and classification of colorectal polyps by transferring low-level cnn features from nonmedical domain. *IEEE Journal of Biomedical and Health Informatics*, 21(1):41–47, 2017.
- [141] W. Zhang, D. Zhao, W. Gong, Z. Li, Q. Lu and S. Yang. Food image recognition with convolutional neural networks. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pages 690–693, 2015.
- [142] Yilu Zhang and Juyang Weng. Convergence analysis of complementary candid incremental principal component analysis. *Comput. Sci. Eng.*, Michigan State Univ., East, 2001.
- [143] Rongsheng Zhu, Rong Zhang and Dixiu Xue. Lesion detection of endoscopy images based on convolutional neural network features. *2015 8th International Congress on Image and Signal Processing (CISP)*, pages 372–376, 2015.

Κατάλογος Δημοσιεύσεων Υποψηφίου

Δημοσιεύσεις σε Επιστημονικά Περιοδικά με Σύστημα Κριτών:

1. S. V. Georgakopoulos, K. Kottari, K. Delibasis, V. P. Plagianakos, and I. Maglogiannis. Improving the performance of convolutional neural network for skin image classification using the response of image analysis filters. *Neural Computing and Applications*, pages 1–18, 2018. (in press)
2. D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, and V. P. Plagianakos. Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification. *IEEE Transactions on Medical Imaging*, 37(10):2196–2210, Oct 2018.
3. S.V. Georgakopoulos, K. Kottari, K. Delibasis, V.P. Plagianakos, and I. Maglogiannis. Pose recognition using convolutional neural networks on omni-directional images. *Neurocomputing*, 280:23 – 31, 2018.
4. Konstantinos K. Delibasis, Spiros V. Georgakopoulos, Konstantina Kottari, Vassilis P. Plagianakos, and Ilias Maglogiannis. Geodesically-corrected zernike descriptors for pose recognition in omni-directional images. *Integrated Computer-Aided Engineering*, 23(2):185–199, 2016.
5. I. Maglogiannis, S.V. Georgakopoulos, S.K. Tasoulis, and V.P. Plagianakos. A software tool for the automatic detection and quantification of fibrotic tissues in microscopy images. *Information Sciences*, 308:125 – 139, 2015.

Δημοσιεύσεις σε Διεθνή Επιστημονικά Συνέδρια με Σύστημα Κριτών:

1. S. V. Georgakopoulos and V. P. Plagianakos. Efficient learning rate adaptation for convolutional neural network training. In *2019 International Joint Conference on Neural Networks (IJCNN)*, July 2019 (to appear).
2. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Aristidis G. Vrahatis, and Vassilis P. Plagianakos. Convolutional neural networks for twitter text toxicity analysis. In Luca Oneto, Nicolò Navarin, Alessandro Sperduti, and Davide Anguita, editors, *Recent Advances in Big Data and Deep Learning*, pages 370–379, Cham, 2020. Springer International Publishing.
3. S. K. Tasoulis, A. G. Vrahatis, S. V. Georgakopoulos, and V. P. Plagianakos. Biomedical data ensemble classification using random projections. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 166–172, Dec 2018.

4. S. K. Tasoulis, A. G. Vrahatis, S. V. Georgakopoulos, and V. P. Plagianakos. Visualizing high-dimensional single-cell rna-sequencing data through multiple random projections. In 2018 IEEE International Conference on Big Data (Big Data), pages 5448–5450, Dec 2018.
5. S. K. Tasoulis, A. G. Vrahatis, S. V. Georgakopoulos, and V. P. Plagianakos. Real time sentiment change detection of twitter data streams. In 2018 Innovations in Intelligent Systems and Applications (INISTA), pages 1–6, July 2018.
6. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Aristidis G. Vrahatis, and Vassilis P. Plagianakos. Convolutional neural networks for toxic comment classification. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, SETN '18, pages 35:1–35:6, New York, NY, USA, 2018. ACM.
7. K. Delibasis, Ilias Maglogiannis, S. Georgakopoulos, K. Kottari, and V. Plagianakos. Assessing image analysis filters as augmented input to convolutional neural networks for image classification. In Artificial Neural Networks and Machine Learning – International Conference on Artificial Neural Networks (ICANN) 2018, pages 188–196, Cham, 2018. Springer International Publishing.
8. S. V. Georgakopoulos, K. Kottari, K. Delibasis, V. P. Plagianakos, and I. Maglogiannis. Detection of malignant melanomas in dermoscopic images using convolutional neural network with transfer learning. In Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, pages 404–414, Cham, 2017. Springer International Publishing.
9. S. V. Georgakopoulos and V. P. Plagianakos. A novel adaptive learning rate algorithm for convolutional neural network training. In Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, pages 327–336, Cham, 2017. Springer International Publishing.
10. S. V. Georgakopoulos, D. K. Iakovidis, M. Vasilakakis, V. P. Plagianakos, and A. Koulaouzidis. Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions. In 2016 IEEE International Conference on Imaging Systems and Techniques (IST), pages 510–514, Oct 2016.
11. Spiros V. Georgakopoulos, Konstantina Kottari, Kostas Delibasis, Vassilis P. Plagianakos, and Ilias Maglogiannis. Convolutional neural networks for pose recognition in binary omnidirectional images. In Artificial Intelligence Applications and Innovations - 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016, Thessaloniki, Greece, September 16-18, 2016, Proceedings, pages 106–116, 2016.
12. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, and Vassilis P. Plagianakos. Efficient change detection for high dimensional data streams. In 2015 IEEE International Conference on Big Data, Big Data 2015, Santa Clara, CA, USA, October 29 - November 1, 2015, pages 2219–2222, 2015.
13. S.V. Georgakopoulos, S.K. Tasoulis, I. Maglogiannis, and V.P. Plagianakos. On-line fall detection via mobile accelerometer data. In Richard Chbeir, Yannis Manolopoulos, Ilias Maglogiannis, and Reda Alhajj, editors, Artificial Intelligence Applications and Innovations, volume 458 of IFIP Advances in Information and Communication Technology, pages 103–112. Springer International Publishing, 2015.

14. K.K. Delibasis, Spiros Georgakopoulos, Vassilis Plagianakos, and Ilias Maglogiannis. Calculation of complex zernike moments with geodesic correction for pose recognition in omnidirectional images. In Lazaros Iliadis, Ilias Maglogiannis, and Harris Papadopoulos, editors, *Artificial Intelligence Applications and Innovations*, volume 436 of *IFIP Advances in Information and Communication Technology*, pages 375–384. Springer Berlin Heidelberg, 2014.
15. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Vassilis P. Plagianakos, and Ilias Maglogiannis. Artificial neural networks and principal components analysis for detection of idiopathic pulmonary fibrosis in microscopy images. In Lazaros Iliadis, Harris Papadopoulos, and Chrisina Jayne, editors, *Engineering Applications of Neural Networks*, volume 383 of *Communications in Computer and Information Science*, pages 292–301. Springer Berlin Heidelberg, 2013.

