ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΡΟΗ ΛΟΓΙΣΜΙΚΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Αναγνώριση Ανθρώπινης Δραστηριότητας με χρήση Βαθέων Αναδρομικών Νευρωνικών Δικτύων σε Πολυτροπικά Δεδομένα Αισθητήρων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΒΑΡΙΟΖΙΔΗ ΙΩΑΝΝΗ

Επιβλέπων: Αθανάσιος Βουλόδημος

Επίκουρος Καθηγητής

Αθήνα , Μάρτιος 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΡΟΗ ΛΟΓΙΣΜΙΚΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Αναγνώριση Ανθρώπινης Δραστηριότητας με χρήση Βαθέων Αναδρομικών Νευρωνικών Δικτύων σε Πολυτροπικά Δεδομένα Αισθητήρων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΒΑΡΙΟΖΙΔΗ ΙΩΑΝΝΗ

Επιβλέπων: Αθανάσιος Βουλόδημος

Επίκουρος Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 2η Μαρτίου του 2021.

Athanasios Voulodimos Digitally signed by Athanasios Voulodimos Date: 2021.03.11 12:00:21 +02'00'	Nikolaos Vasilas Digitally signed by Nikolaos Vasilas Date: 2021.03.11 18:36:36 +02'00'	Anastasios Kesidis Digitally signed by Anastasios Kesidis Date: 2021.03.11 15:57:16 +02'00'
……..	……..	……..
Αθανάσιος Βουλόδημος	Νικόλαος Βασιλάς	Αναστάσιος Κεσίδης
Επίκουρος Καθηγητής	Καθηγητής	Αναπληρωτής Καθηγητής

Αθήνα , Μάρτιος 2021

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Βαριοζίδης Ιωάννης του Ευκλείδη, με αριθμό μητρώου cs141065 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

(*Υπογραφή*)

……..

Ιωάννης Βαριοζίδης

Περίληψη

Η αναγνώριση της ανθρώπινης δραστηριότητας με χρήση της τεχνητής νοημοσύνης θα μπορούσε να χαρακτηριστεί ιδιαίτερα δημοφιλής την τελευταία δεκαετία, καθώς η ευρεία χρήση της εδραιώνεται σε ολοένα και περισσότερους τομείς της καθημερινής ζωής και μάλιστα σε ορισμένους πλέον θεωρείται αναγκαία. Αυτό το γεγονός φαντάζει λογικό, αφού μεταβαίνοντας σε έναν πιο αυτοματοποιημένο και έξυπνο κόσμο η ανάγκη για δημιουργία ενός έξυπνου συστήματος, το οποίο να είναι σε θέση να αναγνωρίζει ανθρώπινες δραστηριότητες είναι επιτακτική. Μερικά πεδία τα οποία εμπλέκουν την αναγνώριση ανθρώπινης δραστηριότητας αποτελούν τα αυτοματοποιημένα συστήματα παρακολούθησης για εφαρμογές υποβοήθησης διαβίωσης και εν γένει υγειονομικού χαρακτήρα καθώς και τα συστήματα ασφαλείας. Η διεθνής βιβλιογραφία περιέχει πληθώρα τέτοιων συστημάτων, ωστόσο η πλειοψηφία αυτών βασίζεται σε οπτικά δεδομένα από κάμερες ενώ λιγότερα είναι τα συστήματα που βασίζονται σε πολυτροπικά δεδομένα που συλλέγονται από διαφορετικούς τύπους αισθητήρων και συνδυάζονται με ευφυείς μεθόδους.

Με την παρούσα διπλωματική εργασία επιχειρείται να εξεταστεί η εφαρμογή των μεθόδων βαθιάς μάθησης για την αναγνώριση ανθρώπινης δραστηριότητας με χρήση αναδρομικών νευρωνικών δικτύων σε πολυτροπικά δεδομένα αισθητήρων. Τα σύνολα δεδομένων τα οποία χρησιμοποιούνται είναι δύο. Ειδικότερα, το πρώτο αφορά δεδομένα, τα οποία έχουν αποκτηθεί από τους αισθητήρες έξυπνων κινητών τηλεφώνων (smartphones), ενώ το δεύτερο αφορά αλληλουχίες εικονοσειρών (βίντεο), οι οποίες έχουν αποκτηθεί μέσω οπτικών μέσων (κάμερα). Τα δύο σύνολα δεδομένων παρουσιάζουν ανθρώπινες καταστάσεις κίνησης με στόχο πάντα την επίτευξη φυσικότητας αναφορικά με τις ενέργειες και δραστηριότητες των εθελοντών.

Τα μοντέλα τα οποία αναπτύχθηκαν στα πλαίσια της διπλωματικής εργασίας αποτελούνται από αναδρομικά νευρωνικά δίκτυα και σε ορισμένες περιπτώσεις εμπλουτισμένα με στοιχεία συνελικτικών νευρωνικών δικτύων. Πιο συγκεκριμένα, αναπτύχθηκαν μοντέλα βασισμένα σε κλασικά αναδρομικά νευρωνικά δίκτυα (RNNs), σε δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory – LSTM) και σε αναδρομικά δίκτυα μονάδων με πύλες (Gated Recurrent Units – GRU), τα οποία επιτρέπουν τη σύγκριση μεταξύ των δικτύων και των μεθόδων, καταλήγοντας με αυτόν τον τρόπο σε συμπεράσματα αναφορικά με την αποτελεσματικότητα αυτών στο πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας.

Λέξεις Κλειδιά: Βαθιά Μάθηση, Αναδρομικά Νευρωνικά Δίκτυα, Όραση υπολογιστών, Αναγνώριση Προτύπων, Αναγνώριση Ανθρώπινης Δραστηριότητας

Abstract

The Human Activity Recognition based in artificial intelligence could be characterized as the state-of-art in the last decade, as its widespread use is consolidating in more and more areas of daily life. Since moving to a more automated and intelligent world the need to create an intelligent system that is able to recognize human activities is required. Some of the areas that involve those automated systems are health care monitoring systems and security systems. The international bibliography contains a plethora of such systems; however, the absence of systems is found, which are based on multimodal data and combine information.

The purpose of this thesis, focuses to examine the application of deep learning methods for the human activity recognition with using recurrent neural networks in multimodal sensor data. The data sets that are used are two. The first one contains data, which have been collected from the sensors of smartphones, while the second one contains sequences of images (video) that have been collected through optical media (camera). Both of the two datasets simulate human movement always based on achieving naturalness in the action perform by volunteers.

The models developed in the thesis are based mainly on recurrent neural networks combined in some cases with convolutional. More specifically, it has used Simple Recurrent Networks (RNNs), Long Short-Term Memory Networks or LSTMs and Gated Recurrent Units, GRU. Upon completion of the implementation of these networks results are obtained, which allow the comparison between the networks and the methods, thus reaching the conclusion regarding their effectiveness in the problem of recognizing human activity.

Keywords: Deep Learning, Recurrent Neural Networks, Computer Vision, Pattern Recognition, Human Activity Recognition

Πίνακας Περιεχομένων

Περιεχόμενα Εικόνων

Περιεχόμενα Πινάκων

Κεφάλαιο 1 – Εισαγωγή

Η αυτοματοποιημένη αναγνώριση ανθρώπινης δραστηριότητας παίζει πολύ σημαντικό ρόλο πλέον σε πολλά πεδία της επιστήμης. Σε γενικές γραμμές, ένα τέτοιο σύστημα χρησιμοποιείται σε διαφορετικές τεχνολογίες με σκοπό πάντα να παρακολουθεί και να αναγνωρίζει τις καθημερινές συνήθειες των ανθρώπων με αυτοματοποιημένο τρόπο. Αναλυτικότερα, υπάρχουν πολλοί τρόποι για να εισαχθούν αυτές οι δραστηριότητες στο σύστημα, οι πιο γνωστές είναι μέσω αισθητήρων ή μέσω εικονοσειρών (βίντεο). Με αυτόν τον τρόπο σε μια πιο απλοποιημένη προσέγγιση, όπου το βίντεο ή οι πληροφορίες των αισθητήρων παρέχονται σε τμήματα τα οποία περιγράφουν την περίοδο μιας κίνησης, το σύστημα καλείται να κατηγοριοποιήσει το συγκεκριμένο τμήμα σε μια συγκεκριμένη κατηγορία κινήσεων.

Τα συστήματα αυτοματοποιημένης ανθρώπινης αναγνώρισης βρίσκουν χρήση σε πληθώρα εφαρμογών. Χαρακτηριστικό παράδειγμα αποτελεί η παρακολούθηση δημόσιων χώρων για την ανάλυση συμπεριφοράς του πλήθους με στόχο τον εντοπισμό ύποπτων κινήσεων. Επιπλέον, κάποια από τα συστήματα αυτά μπορούν να προσφέρουν συνεχής παρακολούθηση, με αποτέλεσμα πολλές φορές να είναι σε θέση να προσφέρουν υπηρεσίας φροντίδας, να αποτρέψουν σωματικές βλάβες, αλλά και να αποτελέσουν βοηθητικό εργαλείο για την υγειονομική βοήθεια ενός ασθενή, με στόχο πάντα να δημιουργήσει έναν πιο έξυπνο περιβάλλον για τους ανθρώπους.

Το ενδιαφέρον το οποίο έχει προκαλέσει η αναγνώριση ανθρώπινης δραστηριότητας ιδιαίτερα στο πλαίσιο της αλληλεπίδρασης ανθρώπου-μηχανής, έχει αυξήσει κατά πολύ την πολυπλοκότητα του προβλήματος. Αυτό το γεγονός έχει ως αποτέλεσμα διάφορες μετατοπίσεις του υποκειμένου, αλλαγές στην φωτεινότητα, διαφορετικές γωνίες λήψης των πληροφοριών. Αυτά αποτελούν μερικά από τα βασικά προβλήματα, τα οποία αντιμετωπίζει σήμερα το πεδίο. Επιπροσθέτως, δεν μπορεί να αγνοηθεί και η συμπεριφορά των ατόμων, καθώς και ο τρόπος με τον οποίο εκτελείται μια κίνηση. Γίνεται κατανοητό επομένως ότι όλα τα παραπάνω επηρεάζονται σε μεγάλο βαθμό από το είδος και τον στόχο, τον οποίο θέλει να πετύχει το εκάστοτε σύστημα. (Γουδέλη, 2018)

1.1 Περιγραφή Προβλήματος

Η παρούσα διπλωματική εργασία εστιάζει στο πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας με την χρήση μεθόδων βαθιάς μάθησης (deep learning) και πιο συγκεκριμένα με αναδρομικά νευρωνικά δίκτυα. Στόχο έχει την αποτελεσματική αυτοματοποιημένη κατηγοριοποίηση των

ανθρώπινων ενεργειών από σύνολα δεδομένων, τα οποία έχουν προέλθει από αισθητήρες και οπτικά μέσα. Πλέον το πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας έχει γίνει από τα πιο μοντέρνα και ο τομέας έχει γνωρίσει τεράστια ανάπτυξη, γεγονός το οποίο οφείλεται στην ραγδαία ανάπτυξη της τεχνολογίας και ειδικότερα της τεχνητής νοημοσύνης. Ένα σύστημα ανθρώπινης αναγνώρισης χωρίζεται σε δύο βασικά μέρη. Το πρώτο αποτελεί το μέρος αυτό με το οποίο συλλέγονται δεδομένα από τις δραστηριότητες, τις οποίες πραγματοποιεί ο άνθρωπος. Οι πιο γνωστοί τρόποι να γίνει αυτό είναι μέσω αισθητήρων ή μέσω οπτικών μέσων. Παραδείγματος χάρη τα σύγχρονα κινητά τηλέφωνα κατέχουν μια πληθώρα τέτοιων αισθητήρων, καθιστώντας με αυτόν τον τρόπο πολύ πιο εύκολη τη συλλογή αυτή. Το δεύτερο μέρος αποτελεί αυτό της ανάλυσης των δεδομένων. Σε αυτό το σημείο η τεχνητή νοημοσύνη παίζει καθοριστικό ρόλο, μιας και έχει την δυνατότητα ανάλυσης μεγάλο όγκου δεδομένων σε πραγματικό χρόνο πολλές φορές με μικρό υπολογιστικό κόστος. Τόσο η εξέλιξη αλγορίθμων της μηχανικής μάθησης, όσο και η ραγδαία ανάπτυξη του hardware, με την πρόοδο των καρτών γραφικών καθιστούν το πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας αποδοτικά επιλύσιμο.

1.2 Δομή Διπλωματικής Εργασίας

Η παρούσα διπλωματική δομείται στα παρακάτω κεφάλαια:

- **Κεφάλαιο 2 – Αναγνώριση Ανθρώπινης Δραστηριότητας με Μηχανική Μάθηση:** Το παρόν κεφάλαιο επικεντρώνεται στο θεωρητικό και τεχνολογικό υπόβαθρο πίσω από τις πρώτες προσεγγίσεις της ανθρώπινης δραστηριότητας με χρήση της μηχανική μάθησης. Πιο αναλυτικά, στην αρχή θα αναφερθούν οι διάφορες προσεγγίσεις, οι οποίες έχουν γίνει με τον καιρό στο πεδίο της ανθρώπινης δραστηριότητας. Στην συνέχεια γίνεται αναφορά στο γνωστικό υπόβαθρο της μηχανικής μάθησης και πως αυτή λειτουργεί. Στο τελευταίο μέρος του κεφαλαίου αναλύεται πως η μηχανική μάθηση συνδυάζεται με το πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας.
- **Κεφάλαιο 3 – Μέθοδοι Βαθιάς Μάθησης για την Αναγνώριση Δραστηριότητας:** Στο κεφάλαιο αυτό εξετάζεται το γνωστικό, καθώς και το τεχνολογικό υπόβαθρο πίσω από τη βαθιά μάθηση -υποκατηγορία της μηχανικής μάθησης- μιας και οι πειραματικές προσεγγίσεις της διπλωματικής εργασίας αυτής βασίζονται σε αυτήν. Ειδικότερα, θα γίνει εκτενή αναφορά στους τύπους δικτύων που θα χρησιμοποιηθούν στο πλαίσιο της εργασίας αυτής αλλά και πως συνδυάζονται με τα σύνολα δεδομένων ανθρώπινης δραστηριότητας.

- **Κεφάλαιο 4 – Πειραματική Αξιολόγηση:** Στο εν λόγω κεφάλαιο θα παρουσιαστούν και θα αξιολογηθούν τα αποτελέσματα των υλοποιημένων δικτύων και τεχνικών για τα δύο διαφορετικά σύνολα δεδομένων.
- **Κεφάλαιο 5 – Συμπεράσματα - Επίλογος**
- **Βιβλιογραφία**

Κεφάλαιο 2 – Αναγνώριση Ανθρώπινης Δραστηριότητας με Μηχανική Μάθηση

2.1 Εισαγωγή

Τα τελευταία χρόνια, το πεδίο της Αναγνώρισης Ανθρώπινης Δραστηριότητας (**H**uman **A**ctivity **R**ecognition) έχει γίνει ένα από τα πιο μοντέρνα ερευνητικά θέματα, λόγω της μεγάλης διαθεσιμότητας αισθητήρων και επιταχυνσιόμετρων, του χαμηλού κόστους και της λιγότερης κατανάλωσης ενέργειας , της ζωντανής ροής δεδομένων, αλλά και λόγω της μεγάλη πρόοδού που υφίστανται οι τομείς της Όρασης Υπολογιστών, της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης (Charmi, Jatna, & Nishant, 2019). Η ανάγκη επίσης για σχεδιασμό έξυπνων λύσεων σε οικιακά περιβάλλοντα, η ραγδαία αύξηση της τρίτης ηλικίας σε συνδυασμό με την ιατρική φροντίδα, η οποία απαιτείται, καθώς και τα υψηλά ποσοστά εγκληματικότητας, έχουν ως αποτέλεσμα η επιστημονική κοινότητα και κατά επέκταση οι κοινωνίες να αναζητούν τρόπους λύσης.

2.2 Συμβολή Αναγνώρισης Ανθρώπινης Δραστηριότητας

Η αναγνώριση δραστηριότητας πρόκειται για ένα πολύ σημαντικό και διεπιστημονικό πεδίο έρευνας, αφού δεν αποτελείται μόνο από την πληροφορική, αλλά και από την αλληλεπίδραση ανθρώπου - υπολογιστή, καθώς και την ψυχολογία με την κοινωνιολογία. Γίνεται αντιληπτό, επομένως, ότι προσελκύει αυξανόμενο ενδιαφέρον από ερευνητές διαφόρων τομέων.

Πρωταρχικός σκοπός στην Αναγνώριση Ανθρώπινης Δραστηριότητας αποτελεί η ταυτοποίηση κοινών διαφόρων ανθρωπίνων δραστηριοτήτων παρατηρώντας την συμπεριφορά των ανθρώπων και τα χαρακτηριστικά του περιβάλλοντος τους, όπως το περπάτημα, το τρέξιμο, το μαγείρεμα, την οδήγηση, το άνοιγμα μιας πόρτας. Τα δεδομένα αυτά μπορούν να συλλεχθούν με τη βοήθεια διαφόρων φορητών τεχνολογικών συσκευών, όπως είναι οι αισθητήρες, οι κάμερες, smartphones ή και σε συνδυασμό αυτών. Μάλιστα, λόγω της ταχείας εξέλιξης των smartphones και των wearables με ενσωματωμένους αισθητήρες, γυροσκόπιο, επιταχυνσιόμετρο, GPS και σε συνδυασμό με το χαμηλό κόστος τους, πλέον η χρήση αυτών των συστημάτων δεν περιορίζεται σε μόνο εσωτερικούς ή ελεγχόμενους χώρους, σε σύγκριση όπως γινόταν παλαιότερα με την χρήση μόνο κάμερας.

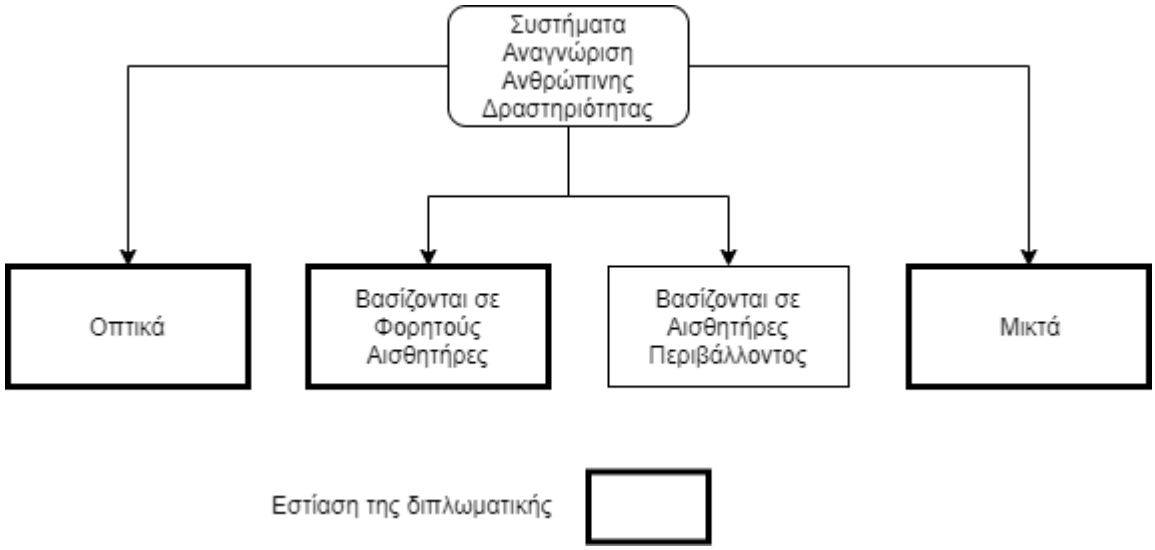
Σχετικά με την ύπαρξη συστημάτων αναγνώρισης δραστηριότητας σε ένα οικιακό περιβάλλον, δύναται πλέον η παρακολούθηση των δραστηριοτήτων των χρηστών για μεγάλο χρονικό διάστημα,

προκειμένου να υπενθυμίζουν στους χρήστες να κάνουν ή να ολοκληρώσουν κάποιες καθημερινές δραστηριότητες, λόγου χάρη να πάρουν κάποιο φάρμακο, να τους βοηθήσουν να θυμηθούν σημαντικές πληροφορίες ακόμα και να τους ενθαρρύνουν-προειδοποιήσουν να ενεργούν με μεγαλύτερη ασφάλεια σε συγκεκριμένες δραστηριότητες. Στα πλαίσια ενός νοσοκομειακού περιβάλλοντος τα συστήματα αναγνώρισης δραστηριότητας, μπορούν να υπενθυμίζουν σε έναν γιατρό ή σε μια νοσοκόμα να προβεί στις απαραίτητες ενέργειες για κάποιον ασθενή. Τέτοια συστήματα δύνανται να λειτουργήσουν ακόμα και σε εφαρμογές παρακολούθησης ασφαλείας, ώστε να προβλέψουν την πρόθεση και το κίνητρο μεταξύ ανθρώπων που αλληλεπιδρούν.

Οι άνθρωποι είναι σε θέση να κατανοήσουν και να ερμηνεύσουν τις δραστηριότητες των ανθρώπων γύρω τους. Η ικανότητα, όμως, αυτή ενώ φαίνεται απλή και φυσιολογική για τους καθημερινούς ανθρώπους, στην πραγματικότητα είναι πολύ πιο περίπλοκη και αποτελεί απόρροια ενός συνδυασμού της αίσθησης, της εμπειρίας και των συμπερασμάτων. Αυτό γίνεται πιο κατανοητό με την αναφορά ενός παραδείγματος. Ειδικότερα, είναι 4:00 μ.μ. το απόγευμα μιας συνηθισμένης ημέρας. Ένα κορίτσι βλέπει τον πατέρα της να στέκεται στο δωμάτιο του ακριβώς δίπλα στο γραφείο του κρατώντας ένα ποτήρι νερό στο χέρι. Μέσω της προηγούμενης εμπειρίας της και της γνώσης της σχετικά με το ιατρικό ιστορικό του πατέρα της μπορεί να συμπεράνει ότι ο πατέρας της λαμβάνει την ημερήσια δόση του φαρμάκου του. Ωστόσο, η αναγνώριση αυτής καθ’ αυτής της κίνησης θα ήταν πολύ μεγάλη πρόκληση για ένα σύστημα αναγνώρισης, αφού θα χρειαζόταν ένα μεγάλο αριθμό αισθητήρων. Επομένως γίνεται αντιληπτό ότι η εμπειρία του παρελθόντος και η εφαρμογή αυτής ως γνώση για το παρόν αντιπροσωπεύει τη+ μεγάλη πρόκληση που έχουν να αντιμετωπίσουν οι μηχανές. (Adil, 2011)

2.3 Προσεγγίσεις στην Αναγνώριση Ανθρώπινης Δραστηριότητας

Το πρώτο βήμα προς την επίτευξη του στόχου της αναγνώρισης των δραστηριοτήτων της καθημερινής ζωής είναι να εξοπλίσει το σύστημα με την «αίσθηση» όπως αναφέρθηκε και παραπάνω. Για να επιτευχθεί αυτό χρησιμοποιούνται κυρίως τέσσερις προσεγγίσεις με συστήματα, τα οποία βασίζονται σε βίντεο (οπτικά), σε αισθητήρες περιβάλλοντος, σε φορητούς αισθητήρες καθώς και στο συνδυασμό αυτών (μικτά) (σχ. 2.1).



Εικόνα 2.3-1: Προσεγγίσεις για την αναγνώριση ανθρώπινης δραστηριότητας.

Οπτικά Συστήματα: Αυτά τα συστήματα χρησιμοποιούν μία ή περισσότερες συνήθως κάμερες, οι οποίες τοποθετούνται σε διαφορετικές οπτικές γωνίες ή σε διαφορετικά σημεία του επιβλεπόμενου χώρου, για να παρακολουθούν και να αναγνωρίζουν την φυσική δραστηριότητα. Τα δεδομένα τα οποία συλλέγονται παρουσιάζονται στο σύστημα, είτε ως ροές δεδομένων σε πραγματικό χρόνο, είτε ως βίντεο, αφού πρώτα έχουν αποθηκευτεί. Τα εν λόγω δεδομένα αποτελούν μία σειρά ακολουθιακών εικόνων (frames), η οποία αναπαριστά την δραστηριότητα. Συνήθως πρόκειται είτε για έγχρωμες εικόνες (RGB) με απεικόνιση τριών διαστάσεων R_i, G_i, B_i και κατά συνέπεια τριών καναλιών, είτε για εικόνες απόστασης (Depth Images), οι οποίες χρησιμοποιούν αισθητήρες βάθους. Αυτό έχει ως συνέπεια να μετρούνται οι αποστάσεις από τα σημεία ενδιαφέροντος με απεικόνιση ενός καναλιού.

Ένα χαρακτηριστικό των οπτικών δεδομένων είναι ότι λόγω του μεγάλου όγκου, του οποίου δύναται να καταλαμβάνουν, επιβαρύνουν με αυτόν τον τρόπο τη διαδικασία της αναγνώρισης, από άποψη υπολογιστικής ισχύς και κατά επέκταση χρονικού κόστους. Για αυτό το λόγο, πολλές φορές υπόκεινται σε προ-επεξεργασία, όπου γίνεται συμπίεση του μεγέθους τους, ενώ ταυτόχρονα με την χρήση διαφόρων τεχνικών γίνεται η εξαγωγή κάποιων χαρακτηριστικών τους, πιο εκτεταμένη αναφορά θα γίνει στο Κεφάλαιο 3.

Επιπρόσθετα, αυτή η προσέγγιση είναι πιο αποτελεσματική συνήθως σε εργαστηριακό περιβάλλον, συγκεκριμένα σε ειδικά εσωτερικούς διαμορφωμένους χώρους, καθώς εκεί υπάρχει σταθερός φωτισμός και συγκεκριμένες δραστηριότητες, οι οποίες επιτελούνται. Αντιθέτως, σε ένα οικιακό περιβάλλον θα είναι λιγότερο ακριβές, λόγω ότι ο φωτισμός μεταβάλλεται, το σπίτι μπορεί να είναι ακατάστατο και στο περιβάλλον όπου βρίσκεται το σύστημα είναι πιθανό να υπάρχουν πολλές διαφορετικές δραστηριότητες, οι οποίες πραγματοποιούνται ταυτόχρονα με αποτέλεσμα να δημιουργείται

«σύγχυση» στο σύστημα. Με άλλα λόγια το σύστημα έρχεται αντιμέτωπο με ένα πολυπαραμετρικό και περίπλοκο περιβάλλον.

Συστήματα που βασίζονται σε Αισθητήρες Περιβάλλοντος: Τέτοια συστήματα αναπτύσσονται για την παρακολούθηση της αλληλεπίδρασης μεταξύ των χρηστών και του οικιακού τους περιβάλλοντος (Van Kasteren, Noylas, Englebienne, & Krose, 2008) (Tapia, Intille, & Larson, 2004). Αυτός ο στόχος επιτυγχάνεται με την ύπαρξη ενός αριθμού ειδικά διαμορφωμένων αισθητήρων περιβάλλοντος (δυαδικών κατάστασης on-off), οι οποίοι τοποθετούνται στο περιβάλλοντα χώρο. Τα δεδομένα τα οποία συλλέγονται από αυτούς τους αισθητήρες δύαται να χρησιμοποιηθούν για την έξυπνη προσαρμογή του οικιακού περιβάλλοντος στις ανάγκες του κάθε χρήστη. Τα συστήματα αυτά παρακολουθούν παθητικά τους χρήστες όλη την ημέρα, κάθε μέρα, χωρίς να απαιτείται καμία ενέργεια από τον χρήστη. Τοποθετούνται σε όλο το σπίτι, και έχουν λιγότερους περιορισμούς αναφορικά με το μέγεθος, το βάρος και την ισχύ σε σύγκριση με άλλους τύπους αισθητήρων, απλοποιώντας έτσι στο σύνολο τον σχεδιασμό του συστήματος. Ωστόσο, τέτοια συστήματα εξαρτώνται αποκλειστικά από την υποδομή του εσωτερικού της εκάστοτε οικίας και δεν μπορούν να παρακολουθούν εξωτερικά συμβάντα. Αξίζει να σημειωθεί ότι πολλές φορές παρουσιάζουν δυσκολίες στη διάκριση μεταξύ διαφορετικών χρηστών του σπιτιού.

Συστήματα που βασίζονται σε Φορητούς Αισθητήρες: Τέτοια συστήματα έχουν σχεδιαστεί για να φοριούνται κατά τη διάρκεια της καθημερινής δραστηριότητας για τη συνεχή μέτρηση βιολογικών δεδομένων, λόγου χάρη τα έξυπνα ρολόγια, bands, ανεξάρτητα της τοποθεσίας που βρίσκεται ο χρήστης και ως εκ τούτου αποτελεί μια εναλλακτική λύση για την αναγνώριση των καθημερινών ανθρώπινων δραστηριοτήτων, ιδίως των σωματικών δραστηριοτήτων. Οι σωματικές δραστηριότητες απαιτούν επαναλαμβανόμενη κίνηση του ανθρώπινου σώματος και αξίζει να αναφερθεί ότι περιορίζονται σε μεγάλο βαθμό, από τη δομή του σώματος. Χαρακτηριστικά παραδείγματα αποτελούν το περπάτημα, το τρέξιμο, και η σωματική άσκηση. Γίνεται επομένως κατανοητό ότι οι φορητοί αισθητήρες είναι κατάλληλοι για τη συλλογή δεδομένων σχετικά με την καθημερινή φυσική κατάσταση για μεγάλο χρονικό διάστημα, καθώς μπορούν να ενσωματωθούν σε ρούχα (Noury, et al., 2004), κοσμήματα ή να φορεθούν ως φορητές συσκευές. Δεδομένου ότι συνδέονται άμεσα με τους χρήστες και τους παρακολουθούν είναι ανεξάρτητοι του περιβάλλοντος, του οποίου βρίσκονται και είναι σε θέση να μετρούν παραμέτρους που οι δύο προηγούμενες κατηγορίες δε μπορούν. Μια σειρά τέτοιων αισθητήρων είναι γωνιόμετρα, επιταχυνσιόμετρα, γυροσκόπια, βηματόμετρα και ακτόμετρα. Ειδικότερα, τα επιταχυνσιόμετρα προσφέρουν ποικίλα πλεονεκτήματα στην παρακολούθηση ανθρώπινης κίνησης. Η απόκρισή τους τόσο στη συχνότητα όσο και στην ένταση της κίνησης τους κάνει ανώτερα από τα ακτόμετρα ή τα βηματόμετρα, τα οποία εξασθενούν από την κρούση ή την κλίση. Ορισμένοι τύποι επιταχυνσιόμετρων μπορούν να μετρήσουν τόσο την κλίση όσο και την κίνηση, και επομένως θεωρούνται ανώτερα από τους υπόλοιπους

αισθητήρες κίνησης, οι οποίοι δεν είναι σε θέση να μετρήσουν στατικά χαρακτηριστικά. Μάλιστα, ο συνδυασμός του μεγέθους τους, αλλά και του χαμηλού κόστους τα καθιστά μοναδικά. Πλέον τα περισσότερα smartphone περιέχουν συνδυασμό των αισθητήρων αυτών.

Μικτά Συστήματα: Σε αυτή την κατηγορία όπως γίνεται κατανοητό και από το όνομα ανήκουν τα συστήματα, τα οποία συνδυάζουν όλες τις προηγούμενες κατηγορίες. Η λογική του συνδυασμού αυτού βασίζεται στην αλληλοσυμπλήρωση των πλεονεκτημάτων και μειονεκτημάτων της κάθε κατηγορίας. Αξιοσημείωτο είναι ότι πρόκειται για την πιο αποδοτική κατηγορία, αλλά και συγχρόνως αυτή με την μεγαλύτερη πρόκληση.

2.4 Αναγνώριση Ανθρώπινης Δραστηριότητας & Μηχανική Μάθηση

2.4.1 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί ένα πεδίο της επιστήμης υπολογιστών, το οποίο εστιάζει στη δημιουργία αλγορίθμων, οι οποίοι αποκτούν γνώση βάσει δεδομένων και εξάγουν αποφάσεις ή προβλέψεις βάσει αυτών, χωρίς την ανάγκη επιπλέον προγραμματισμού. Με άλλα λόγια μια διεργασία της μηχανικής μάθησης στοχεύει στον εντοπισμό (για μάθηση) μιας συνάρτησης $f: X \rightarrow Y$, όπου X είναι η είσοδος, η οποία τροφοδοτείται με δεδομένα και Y η αποτελεί η έξοδος με τις πιθανές προβλέψεις. Οι συναρτήσεις f επιλέγονται και προσαρμόζονται κάθε φορά ανάλογα με το τύπο του αλγόριθμου εκμάθησης που χρησιμοποιείται. Ο (Mitchell, 1997) ορίζει τη «μάθηση» ως εξής: ‘’Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία E σε σχέση με μία κατηγορία εργασιών T και μία μετρική απόδοσης P, αν η απόδοσή του σε εργασίες της T, όπως μετριούνται από την P, βελτιώνονται με την εμπειρία E’’. Η μετρική απόδοσης P αναφέρεται στο κατά πόσο καλά λειτουργεί ο αλγόριθμος εκμάθησης. Για προβλήματα κατηγοριοποίησης, επιλέγεται η ακρίβεια του συστήματος, συνήθως αντί για μετρική απόδοσης, όπου ακρίβεια ορίζεται ως η αναλογία για την οποία το σύστημα παράγει σωστά την έξοδο. Όσον αφορά την εμπειρία E οι αλγόριθμοι αυτοί την αποκτούν μέσω των συνόλων δεδομένων (Datasets). Συνήθως αυτά τα σύνολα δεδομένων περιέχουν ένα σύνολο παραδειγμάτων που χρησιμοποιούνται για την δοκιμή και εκπαίδευση του εκάστοτε αλγόριθμου εκμάθησης.

2.4.2 Τύποι Μηχανικής Μάθησης

Η μηχανική μάθηση στην ερευνητική πλευρά των πραγμάτων, μπορεί να εξεταστεί εις βάθος είτε από την θεωρητική ειτε από την μαθηματική πλευρά. Επομένως υπάρχουν πολλοί τρόποι για να την περιγραφή και κατηγοριοποίηση της , αλλά σε μεγάλο βαθμό υπάρχουν τέσσερις μεγάλες κατηγορίες:

- ***Επιβλεπόμενη Μάθηση (Supervised Learning):*** Η επιβλεπόμενη μάθηση αποτελεί ένα από τους πιο δημοφιλείς και εύκολους να κατανοηθεί τύπους μάθησης. Ο αλγόριθμος εκμάθησης δέχεται σαν δεδομένα στην είσοδο του, παραδείγματα δεδομένων και τις επιθυμητές εξόδους αυτών (ετικέτες), με στόχο να μάθει έναν γενικό κανόνα, ο οποίος να αντιστοιχεί τις εισόδους με τα αποτελέσματα. Για παράδειγμα, είναι ένα σύνολο εικόνων με σκύλους και γάτες δίνοντας στον αλγόριθμο την κάθε εικόνα μαζί με την ετικέτα που την αντιπροσωπεύει (σκύλος ή γάτα), ο αλγόριθμος θα εκπαιδευτεί πάνω σε αυτό και στο τέλος όταν θα δίνονται διαφορετικές εικόνες θα είναι δυνατό να προβλέπει τι είδους ετικέτα βρίσκεται στην εικόνα.
- ***Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning):*** Η μη επιβλεπόμενη μάθηση πρόκειται για το αντίθετο της επιβλεπόμενης. Ουσιαστικά ο αλγόριθμος εκμάθησης δέχεται σαν δεδομένα πάλι παραδείγματα δεδομένων, αλλά χωρίς ετικέτες αυτή την φορά. Αντιθέτως, με διάφορα εργαλεία έχει ως στόχο να ανακαλύψει μόνος του τις συσχετίσεις μεταξύ των δεδομένων.
- ***Ημι-Επιβλεπόμενη Μάθηση (Semi Supervised Learning):*** Είναι ένας συνδυασμός επιβλεπόμενης και μη επιβλεπόμενης μάθησης. Ο αλγόριθμος μαθαίνει από δεδομένα τα οποία περιλαμβάνουν δεδομένα με ετικέτα ή χωρίς, αλλά σε μεγάλο βαθμό χωρίς ετικέτα.
- ***Ενισχυτική Μάθηση (Reinforcement Learning):*** Η ενισχυτική μάθηση είναι αρκετά διαφορετική από τις δύο προαναφερθείσες. Ο αλγόριθμος εκμάθησης λειτουργεί χωρίς καθοδήγηση και μαθαίνει από τα λάθη του, λόγω διάφορων δυναμικών αλλαγών, των οποίων δέχεται μετά από κάθε προσπάθεια.

Μια άλλη κατηγοριοποίηση των προβλημάτων μηχανικής μάθησης προκύπτει από το επιθυμητό αποτέλεσμα το οποίο είναι αναγκαίο να επιτευχθεί. Πιο συγκεκριμένα:

Ταξινόμηση (Classification): Συνήθως χρησιμοποιείται στην επιβλεπόμενη μάθηση και ένα χαρακτηριστικό παράδειγμα αποτελεί όταν φιλτράρετε ένα email ως ‘’spam’’ ή ως ‘’no spam’’. Με άλλα λόγια, τα δεδομένα εισόδου είναι τα emails και το αποτέλεσμα προς επίτευξη χωρίζεται σε δύο κατηγορίες,

οι οποίες ονομάζονται κλάσεις τις "spam" και "no spam". Έτσι, ο αλγόριθμος εκμάθησης στην προκειμένη περίπτωση πρέπει να κατασκευάσει ένα μοντέλο, το οποίο θα αντιστοιχεί τα email σε αυτές τις κλάσεις.

Παλινδρόμηση (Regression): Η διαφορά με τα προβλήματα της ταξινόμησης είναι ότι αναφέρεται σε συνεχή δεδομένα και όχι διακριτά.

Συσταδοποίηση (Clustering): Θεωρείται πρόβλημα της Μη επιβλεπόμενης μάθησης. Τα δεδομένα σε αυτή την περίπτωση διαχωρίζονται σε ομάδες, ανάλογα με τις ιδιότητες τους, χωρίς να είναι γνωστές εκ των προτέρων.

2.4.3 Μηχανική Μάθηση στην Αναγνώριση Ανθρώπινης Δραστηριότητας

Το πρόβλημα της Αναγνώρισης Ανθρώπινης Δραστηριότητας αποτελεί ένα πρόβλημα Ταξινόμησης πολλαπλών κλάσεων και κατά επέκταση ανήκει στην κατηγορία της επιβλεπόμενης μάθησης. Όπως αναφέρθηκε και προηγουμένως, ο αλγόριθμος εκμάθησης στο εν λόγω πρόβλημα λαμβάνει σαν είσοδο δεδομένα με τις αντίστοιχες ετικέτες τους, είτε από αισθητήρες, είτε από οπτικά μέσα και καλείται να δημιουργήσει ένα μοντέλο, το οποίο θα κατηγοριοποιεί τα δεδομένα αυτά σε αντίστοιχες κλάσεις δραστηριότητας (τρέξιμο, περπάτημα, μαγείρεμα). Στόχος είναι ο αλγόριθμος, καταληκτικά, μέσω της εκπαίδευσης της οποίας έχει δεχθεί να είναι σε θέση να συσχετίζει καινούρια δεδομένα (που δεν έχει καμία εμπειρία πάνω σε αυτά) με τις αντίστοιχες σωστές κλάσεις τους.

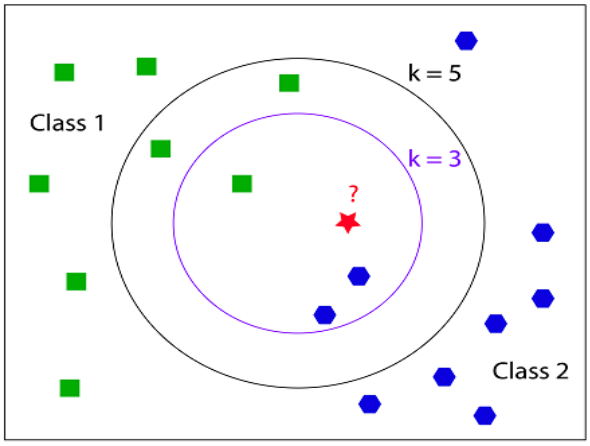
Το σύνολο δεδομένων στο οποίο εκπαιδεύεται ο αλγόριθμος παίζει πολύ σημαντικό ρόλο στη μετέπειτα αποδοτικότητα και αποτελεσματικότητα του μοντέλου. Η ετερογένεια των δεδομένων είναι από τους πιο σημαντικούς παράγοντες που μπορούν να επηρεάσουν ένα μοντέλο, καθώς ένα φτωχό σύνολο δεδομένων δεν επιτρέπει στον αλγόριθμο να εκπαιδευτεί σωστά. Αντίθετα, αν το σύνολο των δεδομένων είναι σχετικά μικρό υπάρχει το πρόβλημα της υπερ-εξειδίκευσης, με αποτέλεσμα ο αλγόριθμος να απομνημονεύει τους συσχετισμούς και να αδυνατεί να γενικευτεί σε καινούρια δεδομένα. Για τη διαδικασία της εκπαίδευσης το σύνολο δεδομένων χωρίζεται σε δύο υποσύνολα (σύνολο εκπαίδευσης και σύνολο δοκιμών) με στόχο την καλύτερη αποτελεσματικότητα του αλγορίθμου, εκτενέστερη αναφορά θα γίνει στο **Κεφάλαιο 3**. Παρακάτω θα γίνει αναφορά στις μεθόδους, οι οποίες χρησιμοποιούνται στη μηχανική μάθηση για την ταξινόμηση στο πρόβλημα της Αναγνώριση Ανθρώπινης Δραστηριότητας:

k-NN (k- Nearest Neighbors)

Ο ταξινομητής k-NN ή αλλιώς k πλησιέστερων γειτόνων, είναι ένας από τους πιο γνωστούς και απλούς αλγόριθμους ταξινόμησης. Βασίζεται ότι παρόμοια στοιχεία βρίσκονται κοντά. Πιο συγκεκριμένα εντοπίζει τα δεδομένα που έχουν παρόμοια χαρακτηριστικά μεταξύ τους (βάση κάποιας μετρικής απόστασης) στο δοθέν σύνολο δεδομένων και τα ομαδοποιεί στην αντίστοιχη κλάση. Κάθε δείγμα του συνόλου που εξετάζει ο αλγόριθμος ταξινομείται βάσει των k πλησιέστερων δειγμάτων. Κάθε ένα από τα γειτονικά δείγματα ανήκει σε μια κλάση, η οποία έχει οριστεί, είτε από την αρχή, είτε κατά την διάρκεια εκτέλεσης του αλγορίθμου, έτσι το νέο στοιχείο ταξινομείται βάσει του γειτονικού του. Επομένως, γίνεται κατανοητό πως τρία σημαντικά στοιχεία χρειάζονται για αυτόν τον αλγόριθμο, το πρώτο είναι το σύνολο δεδομένων να είναι με κλάσεις που περιέχουν ετικέτες, το δεύτερο είναι η σωστή μετρική απόστασης που θα χρησιμοποιηθεί και τρίτον ο αριθμός τον k πλησιέστερων γειτόνων. Η πιο συχνή μετρική, η οποία υπολογίζει την απόσταση μεταξύ δείγματος και γειτόνων συνήθως είναι η Ευκλείδεια που δίνεται και από το γενικό τύπο που παρουσιάζεται στην παρακάτω εξίσωση.

$$d(X,Y) = \sqrt{\sum_{i=1}^{m} |x_i - y_i|^2}$$

Υπάρχουν και άλλες μετρικές απόστασης, όπως οι αποστάσεις Hamming, Manhattan, Cambera. Όσον αφορά την επιλογή των k πλησιέστερων γειτόνων δε θα πρέπει να είναι αριθμητικά ούτε πολύ μεγάλη, διότι ενώ θα υπάρχει μεγαλύτερη ακρίβεια θα υπάρχει και μεγάλό υπολογιστικό κόστος, αλλά ούτε και πολύ μικρή μιας και η αποτελεσματικότητα του αλγορίθμου θα είναι χαμηλότερη.



Εικόνα 2.4-1: Παράδειγμα εφαρμογής του αλγόριθμου εκμάθησης k-NN σε δεδομένα δύο χαρακτηριστικών

Όταν ένα νέο δείγμα εισέλθει στην λίστα με τους k πλησιέστερους γείτονες ταξινομείται βάσει της κλάσης πλειοψηφίας, στην οποία ανήκουν τα γειτονικά δείγματα, και βασίζεται στην εξίσωση 2:

$$Y' = argmax \sum_{(x,y)\in D_2} I(v = y_i)$$

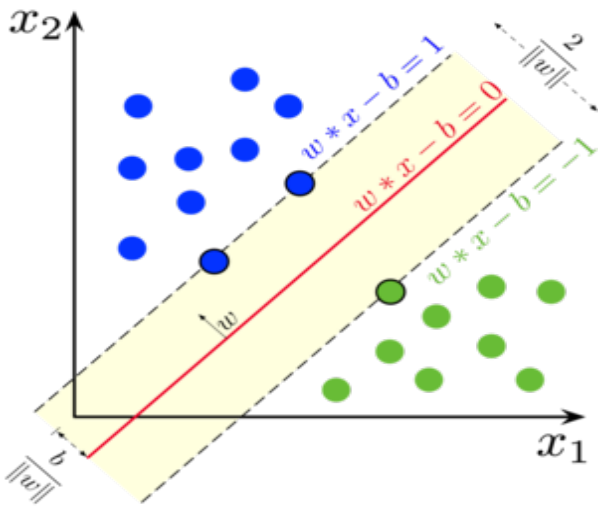
Όπου v είναι η ετικέτα της κλάσης , y_i αναπαριστά το i-οστό πλησιέστερο γείτονα του v, και το I είναι η χαρακτηριστική εξίσωση που επιστρέφει μια τιμή για έγκυρο όρισμα ή μηδέν για μη έγκυρο (Wu, et al., 2007).

Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

Οι μηχανές διανυσμάτων υποστήριξης βασίζονται στην αρχή ελαχιστοποίησης ρίσκου (Vapnik, 1995) από την θεωρία υπολογισμού. Η ιδέα αυτή βασίζεται στην εύρεση μιας υπόθεσης h για την οποία υπάρχει το χαμηλότερο δυνατό πραγματικό σφάλμα. Το h η αλλιώς Hyperplane είναι ένας διαχωριστής, ο οποίος χωρίζει τις κατηγορίες των δεδομένων. Εντούτοις, το πραγματικό σφάλμα είναι η πιθανότητα του h να έχει σφάλμα σε ένα τυχαίο παράδειγμα που ο αλγόριθμος δεν έχει ξανά δει. Με αυτόν τον τρόπο, μπορεί να χρησιμοποιηθεί ένα ανώτατο όριο, το οποίο θα συνδέει το πραγματικό σφάλμα της υπόθεσης h με την υπόθεση h του συνόλου δεδομένων εκπαίδευσης. Επομένως, οι μηχανές διανυσμάτων υποστήριξης καλούνται να βρουν αυτό το h το οποίο να ελαχιστοποιεί το όριο μεταξύ των κλάσεων αποδοτικά, (εξίσωση 3) ώστε όταν θα εισαχθεί ένα νέο στοιχείο, ο αλγόριθμος να μπορεί να το κατηγοριοποιήσει στην σωστή κλάση.

$$L_p = \frac{1}{2}||\vec{w}|| - \sum_{i=1}^{t} a_i * \gamma_i(\vec{w} * \vec{x_i} + b) + \sum_{i=1}^{t} a_i$$

Όπου το t είναι το σύνολο των δεδομένων εκπαίδευσης, a_i οι πολλαπλασιαστές Lagrangian και το L_p αποτελεί παράδειγμα του Lagrangian. Τα διανύσματα $\vec{w}$ και η σταθερά b χαρακτηρίζουν το hyperplane.



Εικόνα 2.4-2: Παράδειγμα εφαρμογής του αλγόριθμου εκμάθησης SVM σε δεδομένα δύο χαρακτηριστικών.

Τα δεδομένα εκπαίδευσης δίνονται ως ζεύγη (x,y) όπου $\vec{x}$ το χαρακτηριστικό διάνυσμα και y η δυαδική τιμή (-1 η 1) που υποδεικνύει σε ποια κλάση ανήκει το δείγμα. Το υπερ-επίπεδο που χωρίζει τα δεδομένα δίνεται από την εξίσωση 4:

$$w * x - b = 0$$

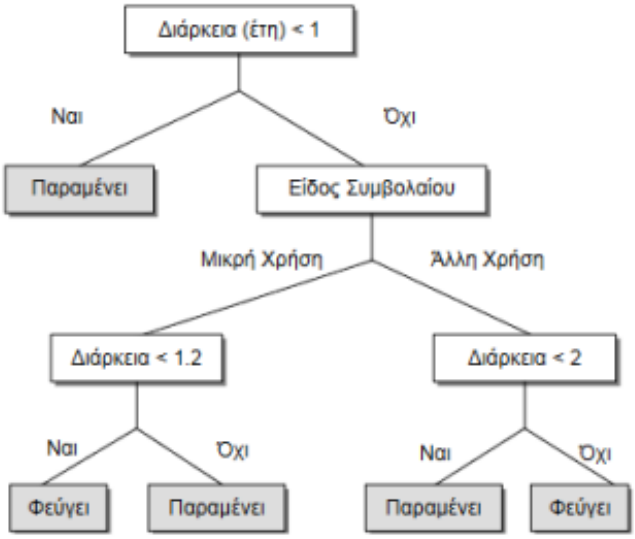
Δέντρα Αποφάσεων (Decision Trees)

Οι αλγόριθμοι δέντρων αποφάσεων, όπως υποδεικνύει και η ονομασία τους, περιλαμβάνουν δέντρα, τα οποία κατηγοριοποιούν τα δεδομένα βάσει κάποιων χαρακτηριστικών. Κάθε κόμβος σε ένα δέντρο αποφάσεων απεικονίζει ένα χαρακτηριστικό, το οποίο πρέπει να ταξινομηθεί, αντίστοιχα κάθε κλάδος του δέντρου απεικονίζει τις τιμές, οι οποίες λαμβάνονται υπόψη από το δέντρο και κάθε φύλο αναπαριστά μια κλάση. Επιπλέον, αναφορικά με τη μέτρηση της ποιότητας ενός διαχωρισμού και κατά επέκταση πότε ο αλγόριθμος θα πρέπει να σταματήσει να φτιάχνει και άλλες διακλαδώσεις ορίζονται διάφορες μετρικές. Οι πιο γνωστές είναι η μετρική Gini και η εντροπία που δίνονται στις παρακάτω εξισώσεις 5,6:

$$Gini\ (E) = 1 - \sum_{j=1}^{c} p_j^2$$

$$Entropy\ (E) = -\sum_{j=1}^{c} p_j \log_2 p_i$$

Όσον αφορά τη διαδικασία κατασκευής του αλγόριθμου, υπολογίζεται αρχικά η εντροπία του συνόλου δεδομένων για κάθε χαρακτηριστικό. Το σύνολο δεδομένων, αποτελώντας τον ριζικό κόμβο, διαιρείται σε υποσύνολα, τα οποία αποτελούν τα διάδοχα πεδία. Ο διαχωρισμός αυτός βασίζεται στην εντροπία των χαρακτηριστικών. Δημιουργείται δηλαδή στο δέντρο ένας κόμβος, ο οποίος έχει ως κριτήριο τα προηγούμενα χαρακτηριστικά. Αυτή η διαδικασία επαναλαμβάνεται σε κάθε παράγωγο υποσύνολο με αναδρομικό τρόπο (recursive partioning) για όλα τα υποσύνολα.



Εικόνα 2.4-3: Παράδειγμα εφαρμογής του δένδρου απόφασης (Αγαπητός, 2018).

Naive Bayes

Οι ταξινομητές Bayes είναι μια συλλογή αλγορίθμων που βασίζονται στο θεώρημα Bayes. Στην συγκεκριμένη περίπτωση οι τιμές κάθε χαρακτηριστικού κατανέμονται σύμφωνα με την Gaussian κατανομή. Ο στόχος είναι να βρεθεί η υπό συνθήκη πιθανότητά για το συμβάν C_k μεταξύ ενός συνόλου πιθανών αποτελεσμάτων τάξης $C = \{c_1, c_2, \dots, c_k\}$. Ο κανόνας του Bayes διατυπώνεται ως:

$$P(C_k|x) = P(c_k) * \frac{P(x|c_k)}{P(x)} = p(C_k, x_1, \dots, x_n)$$

όπου $X = \{x_1, x_2, \dots, x_n\}$ το διάνυσμα χαρακτηριστικών, η πιθανότητα p(x) που αποτελεί σταθερά, δεδομένου πως οι τιμές του x είναι γνωστές ο παρονομαστής αγνοείτε.

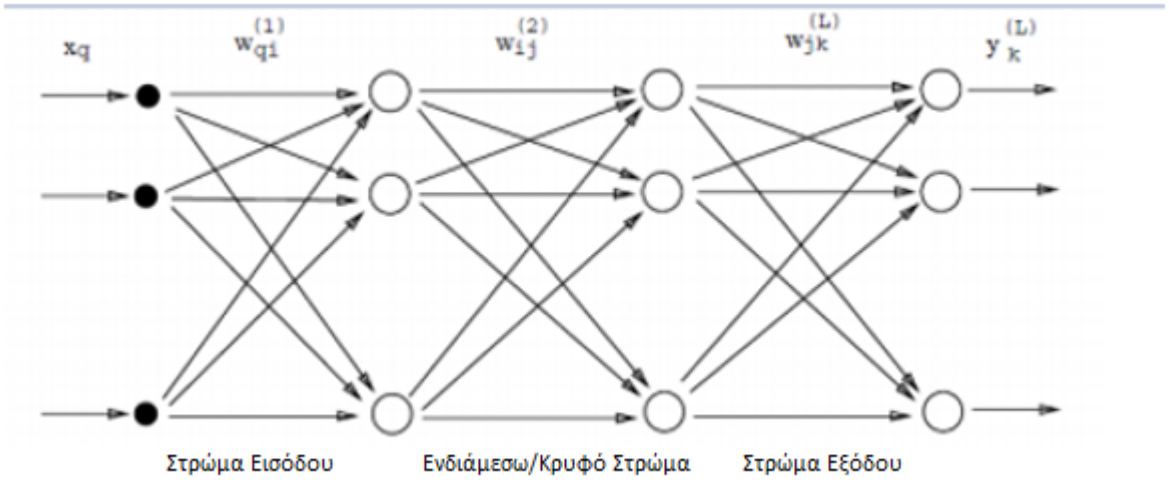
Κάνοντας χρήση του κανόνα της αλυσίδας για την υπό συνθήκη πιθανότητα και χρησιμοποιώντας έναν κανόνα απόφασης. Συνήθως κανόνας είναι αυτός της μέγιστης υπολογισμένης πιθανότητας. Προκύπτει ο τελικός τύπος:

$$P(C_k|x_1, \dots, x_n) \propto P(c_k \prod_{i=1}^{n} P(x_i|C_k)$$

$$y(x) = argmax\left(P\left(c_j \prod_{i=1}^{n} P(x_i|C_j\right)\right)$$

Τεχνητά Νευρωνικά Δίκτυα (Neural Networks)

Τα νευρωνικά δίκτυα είναι εμπνευσμένα από τους βιολογικούς ανθρώπινους νευρώνες. Μάλιστα, τα τελευταία χρόνια τα εν λόγω δίκτυα έχουν αποκτήσει ευρεία αναγνώριση στον τομέα της μηχανικής μάθησης ως ένας από τους πιο αποτελεσματικούς αλγόριθμους εκμάθησης -ξεπερνώντας πολλούς και από τους προαναφερθέντες- σε διάφορες εφαρμογές. Αναλυτικότερα, όπως και στον άνθρωπο, το νευρωνικό δίκτυο αποτελεί μια αρχιτεκτονική, η οποία αποτελείται από χιλιάδες μονάδες που ονομάζονται νευρώνες. Αυτές οι αρχιτεκτονικές συνήθως αποτελούνται από τρία διαφορετικά επίπεδα (layers): Το επίπεδο εισόδου περιέχει το διάνυσμα των χαρακτηριστικών εισόδου. Το στρώμα εξόδου αποτελείται από την απόκριση του νευρωνικού δικτύου και περιέχει τις κλάσεις που κατηγοριοποιείται το πρόβλημα. Τέλος, το ενδιάμεσο στρώμα περιέχει τους νευρώνες, οι οποίοι αλληλοσυνδέονται μεταξύ τους ταυτόχρονα και με την είσοδο και την έξοδο. Ένα παράδειγμα νευρωνικού δικτύου απεικονίζεται στο παρακάτω σχήμα:



Εικόνα 2.4-4: Παράδειγμα αρχιτεκτονικής νευρωνικού δικτύου.

Επιπροσθέτως, τα τεχνητά νευρωνικά δίκτυα εκτός από την αρχιτεκτονική τους, διέπονται και από άλλα δύο χαρακτηριστικά: Τις συναρτήσεις ενεργοποίησης και το βάρος των συνδέσεων εισόδου. Όσον αφορά το πρώτο χαρακτηριστικό ως συνάρτηση ενεργοποίησης ορίζεται η συνάρτηση αυτή που θα δοθεί σε έναν νευρώνα ώστε να επεξεργασθεί τα δεδομένα και να παράγει το αντίστοιχο αποτέλεσμα. Οι πιο συχνές συναρτήσεις είναι οι σιγμοειδείς και οι Relu. Το δεύτερο χαρακτηριστικό θεωρείται μια βασική παράμετρος, η οποία εντοπίζεται στο ενδιάμεσο στρώμα μεταξύ των νευρώνων και πολλαπλασιάζεται με την είσοδο κάθε νευρώνα και η έξοδος δίνεται σαν είσοδος στον επόμενο. Επιπλέον, τα βάρη αυτά ανάλογα με το επιθυμητό αποτέλεσμα αλλάζουν σε κάθε εκτέλεση του αλγορίθμου, για αυτό άλλωστε παίζουν και μεγάλο ρόλο στην απόδοση του μοντέλου. Η λειτουργία ενός νευρώνα δύναται να περιγραφή με την ακόλουθη εξίσωση:

$$y = f_a(\sum_{j=1}^{n} w_j * o_j + b)$$

Όπου o_j είναι η έξοδος του j-οστού προηγούμενου νευρώνα, ενώ με w_j αναπαρίσταται το συνοπτικό βάρος. Το πλήθος των νευρώνων είναι n, ενώ υπάρχει και ο σταθερός όρος b που ονομάζεται bias. Η έξοδος y δίνεται σαν είσοδος στη συνάρτηση ενεργοποίησης.

2.5 Προκλήσεις στην Αναγνώριση Ανθρώπινης Δραστηριότητας

Οι προκλήσεις στο πεδίο της αναγνώρισης ανθρώπινης δραστηριότητας είναι πολλές και πολύ παραγοντικές. Πιο συγκεκριμένα, οι περισσότερες από τις εν λόγω προκλήσεις εντοπίζονται σε παρόμοια προβλήματα μηχανικής μάθησης και αναγνώρισης μοτίβων, αλλά υπάρχουν και κάποιες μοναδικές. Πέρα των πρακτικών ζητημάτων που αφορούν τον αριθμό, τη θέση και την φύση των οπτικών μέσων και αισθητήρων, υπάρχουν επίσης πολλά άλλα ζητήματα, τα οποία επηρεάζουν άμεσα την απόδοση ενός συστήματος αναγνώρισης. Παρακάτω θα αναφερθούν οι παράγοντες οι οποίοι μπορούν να συμβάλουν στην πολυπλοκότητα ενός τέτοιου συστήματος.

2.5.1 Περιπλοκότητα των Δραστηριοτήτων

Στον τομέα της μηχανικής μάθησης, ειδικότερα στο πεδίο της αναγνώρισης ανθρώπινης δραστηριότητας η αξιολόγηση των αλγορίθμων εκμάθησης γίνεται με βάση την πολυπλοκότητα των δραστηριοτήτων που αναγνωρίζουν. Οι δραστηριότητες μπορεί να ποικίλουν και να εξαρτώνται από διαφορετικούς παράγοντες, όπως τον αριθμό αυτών των δραστηριοτήτων, το είδος, καθώς και την περιπλοκότητα των δεδομένων εκπαίδευσης που έχουν συλλεχθεί για αυτές.

Αριθμός Δραστηριοτήτων: Οι άνθρωποι στην καθημερινή ζωή ‘’εκτελούν’’ μεγάλο αριθμό διαφορετικών δραστηριοτήτων. Επομένως, ένα σύστημα αναγνώρισης ανθρώπινης δραστηριότητας οφείλει να είναι σε θέση να αναγνωρίζει ένα μεγάλο εύρος αυτών των δραστηριοτήτων. Συνήθως, η αναγνώριση μικρού συνόλου δραστηριοτήτων είναι πολύ ευκολότερη διαδικασία για ένα τέτοιο σύστημα. Αυτό οφείλεται στο γεγονός ότι όσο ο αριθμός των δραστηριοτήτων αυξάνεται, τόσο ο ταξινομητής πρέπει να κάνει διάκριση μεταξύ του μεγαλύτερου συνόλου δραστηριοτήτων και κατά επέκταση να χρειάζεται μεγαλύτερη υπολογιστική ισχύ, χρόνο αλλά και να είναι πιο ‘’ευαίσθητος΄΄ σε λάθη.

Είδος Δραστηριοτήτων: Οι δραστηριότητες οι οποίες είναι στατικές, όπως χειρονομίες ή μία ακίνητη στάση είναι ευκολότερο να αναγνωριστούν σε σύγκριση με δυναμικές δραστηριότητες, όπως αποτελεί το τρέξιμο ή το περπάτημα. Αξίζει να αναφερθεί ότι οι στάσεις, οι οποίες είναι παρόμοιες, όπως το να στέκεται κάποιος με το να κάθεται είναι εξίσου περίπλοκες να αναγνωρισθούν αφού αλληλεπικαλύπτονται σε σημαντικό βαθμό, αναφορικά με τον χώρο των χαρακτηριστικών. Ακόμη, ότι αφορά τις εν λόγω δραστηριότητες, λόγου χάρη το ‘’περπάτημα σε έναν διάδρομο’’, με το ‘’ανέβασμα ή το κατέβασμα μιας σκάλας’’ είναι επίσης δύσκολο να διακριθούν, καθώς υπάρχει μεγάλη ομοιότητα στα μοτίβα κίνησης τους.

Έτσι γίνεται κατανοητό πως η αναγνώριση μεγάλου αριθμού δραστηριοτήτων, οι οποίες διαφέρουν μεταξύ τους σε μεγάλο βαθμό όσο και παρόμοιες σε χαρακτηριστικά ταυτόχρονα, έχει ως αποτέλεσμα το πρόβλημα της αναγνώρισης να καθίσταται ακόμη πιο δύσκολο. Σε τέτοιες περιπτώσεις, μπορεί να υπάρχει υψηλή ομοιότητα μεταξύ των δραστηριοτήτων, αλλά αυτές να μην είναι ομοιόμορφες σε ολόκληρο το σύνολο δεδομένων. Με άλλα λόγια, ένα υποσύνολο δραστηριοτήτων δύναται να έχει μεγάλή ομοιότητα μεταξύ των δραστηριοτήτων του, αλλά να είναι πολύ διαφορετικό από ένα άλλο υποσύνολο. Για παράδειγμα το ‘’να κάθεται κάποιος’’ με το ‘’να στέκεται’’ είναι παρόμοιες δραστηριότητες τελείως διαφορετικές από το περπάτημα.

Σύνολα Δεδομένων για τις δραστηριότητες: Τα δεδομένα τα οποία συλλέγονται για την εκπαίδευση ενός τέτοιου συστήματος μπορούν να συλλέγονται, είτε σε συνθήκες εργαστηρίου, είτε σε πραγματικές συνθήκες. Τα εργαστηριακά δεδομένα είθισται να συλλέγονται χρησιμοποιώντας αυστηρό πρωτόκολλο. Πιο συγκεκριμένα, οι δραστηριότητες εκτελούνται με την ίδια ταχύτητα από τα άτομα και με περιορισμένους τρόπους, ενώ κατά τη διάρκεια των πραγματικών συνθηκών τα άτομα ενδέχεται να συμπεριφέρονται διαφορετικά και με λιγότερους περιορισμούς. Μακροπρόθεσμα συνήθως συνθήκες εκτός εργαστηρίου σημαίνει Μη Επιβλεπόμενη μάθηση και κατά επέκταση λιγότερο ελεγχόμενη και μη καθορισμένη κατάσταση από τον προγραμματιστή. Αυτό το γεγονός έχει ως αποτέλεσμα να υπάρχουν πολλές προκλήσεις. Οι πιο σημαντικές από αυτές περιλαμβάνουν:

- Υπό αυτές τις συνθήκες, αν οι ετικέτες των δεδομένων εκπαίδευσης δεν είναι καθορισμένες από τον προγραμματιστή, θα έχει ως αποτέλεσμα το σύστημα να πρέπει να βρει μόνο του τις ετικέτες μέσω διάφορων τρόπων clustering και κατά επέκταση μπορεί να δημιουργηθεί θέμα αναξιοπιστίας και τελικά να υποβαθμιστεί ολόκληρη η ακρίβεια του συστήματος.
- Δεν υπάρχει τυπικός τρόπος για να εκτελεσθεί μια δραστηριότητα. Για παράδειγμα ένα άτομο μπορεί να έχει ξαπλώσει στον καναπέ με αποτέλεσμα να μην μπορεί να κατηγοριοποιηθεί αυτή η δραστηριότητα ότι το άτομο στέκεται ή είναι καθιστό.

  Επομένως, στις δραστηριότητες όπου τα δεδομένα εκπαίδευσης έχουν συλλεχθεί σε εργαστηριακές συνθήκες είναι συνήθως πιο εύκολα στην αναγνώριση σε σύγκριση από τις πραγματικές συνθήκες.

2.5.2 Απαιτήσεις Δεδομένων Εκπαίδευσης

Τα συστήματα αναγνώρισης ανθρώπινης δραστηριότητας και κατά επέκταση οι αλγόριθμοι εκμάθησης βασίζονται και αξιολογούνται βάσει τον τύπο καθώς και την ποσότητα δεδομένων εκπαίδευσης που χρειάζονται.

Ανεξάρτητη Αναγνώριση Δραστηριότητας: Ιδανικά, ένα αλγόριθμος εκμάθησης αναγνώρισης ανθρώπινης δραστηριότητας θα εκπαιδευόταν σε ένα συγκεκριμένο σύνολο δεδομένων δραστηριοτήτων και στη συνέχεια θα αναγνώριζε δραστηριότητες, χωρίς όμως να είχε προηγούμενη εμπειρία σε αυτές. Εντούτοις, όπως έχει αποδειχθεί από προηγούμενες δουλείες, σύμφωνα με τους (Bao & Intille, 2004) , υποδηλώνεται έντονα πώς η ανεξάρτητη αναγνώριση των δραστηριοτήτων είναι δύσκολο να επιτευχθεί, ιδιαίτερα στην περίπτωση πολλών διαφορετικών δραστηριοτήτων, λόγω της υψηλής μεταβλητότητας στον τρόπο που οι άνθρωποι ‘’εκτελούν’’ αυτές.

Ποσότητα Δεδομένων Εκπαίδευσης: Η αναγνώριση ανθρώπινης δραστηριότητας υποδηλώνει πως τα δεδομένα, τα οποία απαιτείται, αποδίδουν καλύτερα όταν είναι του ίδιου μεγέθους ,παρότι μπορεί να προέρχονται από διαφορετικές πηγές. Σε περίπτωση που ο αριθμός των δεδομένων είναι λιγοστός ή αντίθετα με πάρα πολλές διαφορετικές δραστηριότητες, μπορεί να οδηγήσει σε χρονοβόρα και μη αποδοτικά αποτελέσματα.

2.5.3 Απαιτήσεις Αισθητήρων & Οπτικών Μέσων

Ο αριθμός των οπτικών μέσων και αισθητήρων που χρησιμοποιούνται, ο τύπος τους αλλά και η θέση που τοποθετούνται είτε στο σώμα είτε στο περιβάλλον παίζουν κομβικό ρόλο στην πολυπλοκότητα και στην απόδοση των αλγορίθμων αναγνώρισης.

Αριθμός Αισθητήρων: Τα συστήματα αναγνώρισης ανθρώπινης δραστηριότητας που χρησιμοποιούν μικρό αριθμό αισθητήρων συνήθως είναι πιο βολικά και πιο κοντά στις πραγματικές συνθήκες. Όμως, δεδομένου αυτού του μικρού αριθμού, το σύστημα δεν είναι σε θέση να έχει υψηλή ακρίβεια στις αναγνωρίσεις, αφού τα σήματα, τα οποία αποστέλλονται ως δεδομένα σε αυτό είναι λιγοστά.

Τοποθεσία Αισθητήρων/Οπτικών Μέσων: Οι αισθητήρες συνήθως συνδέονται με διαφορετικά μέρη του ανθρώπινου σώματος είτε του περιβάλλοντος που βρίσκονται τα άτομα με στόχο την συλλογή δεδομένων. Όσον αφορά τους αισθητήρες που τοποθετούνται πάνω στο ανθρώπινο σώμα, είναι αποδεκτοί για μικρό χρονικό διάστημα, καθώς η μακροχρόνια παρακολούθηση δύναται να εμποδίζει το άτομο από

την ικανότητα άσκησης και να το αναγκάζει σε επαναλαμβανόμενα μοτίβα, λόγω του περιορισμού κίνησης που μπορεί να του επιβάλει ο αισθητήρας. Επιπλέον, ένα σημαντικό θέμα, το οποίο αφορά τα οπτικά μέσα και τους αισθητήρες περιβάλλοντος αποτελεί η σωστή τοποθέτηση αυτών στα σημεία ενδιαφέροντος, ώστε να συλλέγονται σωστά τα δεδομένα και όχι ασήμαντες πληροφορίες.

Επομένως, ιδανικό θα ήταν να υπήρχε ένα σύστημα, το οποίο θα παρακολουθούσε τα άτομα με αισθητήρες που θα μπορούσαν να είναι βολικοί για αυτούς που συγχρόνως όμως να είναι ακριβείς και να μπορούν να αντλούν την απαιτούμενη πληροφορία. Μια τέτοια δοκιμή έχει ξεκινήσει να γίνεται με την χρήση των smartphone και των smart ρολογιών, αλλά ακόμα δεν έχει την ίδια αποτελεσματικότητα που έχουν οι αισθητήρες που τοποθετούνται πάνω στο σώμα.

2.5.4 Περιορισμοί σε πραγματικό χρόνο

Οι αλγόριθμοι αναγνώρισης δραστηριότητας, ειδικά αυτοί οι οποίοι λειτουργούν σε φορητές συσκευές, πρέπει να είναι συγχρόνως αρκετά γρήγοροι, αποδοτικοί, καθώς και ελαφριοί στο να αναγνωρίζουν δραστηριότητες σε πραγματικό χρόνο, χρησιμοποιώντας όσο το δυνατόν λιγότερους πόρους, όπως μνήμη και υπολογιστική ισχύ. Με άλλα λόγια, τα συστήματα αυτά θα πρέπει να χρησιμοποιούν ένα μικτό αισθητήρα, οποίος πραγματοποιεί τη διαδικασία αναγνώρισης. Τα συστήματα τα οποία χρησιμοποιούν πολλούς αισθητήρες διαφορετικού τύπου αυξάνουν τον χρόνο επεξεργασίας και την πολυπλοκότητα του συστήματος σημαντικά.

Πιο αναλυτικά, οι περισσότερες προσεγγίσεις προκειμένου να ταξινομήσουν την εκάστοτε δραστηριότητα λαμβάνουν ως δεδομένα τις εξόδους από διαφορετικούς αισθητήρες και οπτικά μέσα και στην συνέχεια επεξεργάζονται κάθε κανάλι εισόδου ξεχωριστά και τελικά τα ενοποιήσουν όλα για να μπορέσει ο αλγόριθμος εκμάθησης να εκπαιδευτεί. Όπως γίνεται κατανοητό, η διαδικασία αυτή απαιτεί αρκετό χρόνο και υπολογιστική ισχύ, ειδικά αν γίνεται λόγος για πραγματικό χρόνο όπου οι απαιτήσεις είναι πολύ περισσότερες.

Μια πιο ρεαλιστική λύση αποτελεί η ανάπτυξη τέτοιων μοντέλων, τα οποία να μην απαιτούν σε μεγάλο βαθμό πολλούς υπολογιστικούς πόρους. Η έρευνα έχει δείξει ότι σημαντικό ρόλο στο εν λόγω πρόβλημα παίζει η βαθιά μάθηση και η χρήση νευρωνικών δικτύων, (όπως θα γίνει αναφορά στο κεφάλαιο 3) συγκεκριμένα κάποια από αυτά είναι τα συνελικτικά ή τα αναδρομικά δίκτυα, τα οποία δείχνουν να αποδίδουν καλύτερα από τις συμβατικές μεθόδους μηχανικής μάθησης. Τα μοντέλα αυτά έχουν μεγάλες υπολογιστικές απαιτήσεις, αλλά η ακρίβεια και η αποδοτικότητας τους είναι πρωτόγνωρη. Επομένως,

γίνεται αντιληπτό ότι υπάρχει η ανάγκη ανάπτυξης μεθόδων για την ελαχιστοποίηση αυτών των μοντέλων, κάτι το οποίο δεν είναι απαίτηση μόνο του προβλήματος της ανθρώπινης δραστηριότητας, αλλά όλων των επιμέρων πεδίων.

Κεφάλαιο 3 – Μέθοδοι Βαθιάς Μάθησης για την Αναγνώριση Δραστηριότητας

3.1 Εισαγωγή

Το ανθρώπινο είδος ονομάζεται homo sapiens – άνθρωπος ο σοφός – επειδή οι νοητικές ικανότητες παίζουν βασικό ρόλο. Για χιλιάδες χρόνια γίνονταν προσπάθειες να γίνει κατανοητό, το πώς σκεπτόμαστε· δηλαδή, πώς μια χούφτα ύλης μπορεί να αντιλαμβάνεται, να κατανοεί, να προβλέπει και να χειρίζεται έναν κόσμο πολύ μεγαλύτερο και πολύ πιο πολύπλοκο από τον εαυτό της. Το πεδίο της τεχνητής νοημοσύνης (artificial intelligence), ή εν συντομία ΤΝ, πηγαίνει πιο πέρα: Επιχειρεί όχι μόνο να κατανοήσει, αλλά και να κατασκευάσει νοήμονες οντότητες.

Μια υποκατηγορία της τεχνητής νοημοσύνης, όπως αναφέρθηκε και στο κεφάλαιο 2, αποτελεί το πεδίο της μηχανικής μάθησης. Στην εν λόγω υποκατηγορία, οι αλγόριθμοι μέσα από στατιστικές μεθόδους προσπαθούν να βελτιώσουν τις επιδόσεις τους, ώστε να λύσουν κάποιο πρόβλημα. Χαρακτηριστικό του πεδίου αυτού είναι ότι οι αλγόριθμοι δύνανται να μαθαίνουν από τα δεδομένα που τους δίνονται και η συμπεριφορά τους δεν ορίζεται ρητά από τον προγραμματιστή.

Η Βαθιά μάθηση ή αλλιώς Deep Learning αποτελεί μια υποκατηγορία της Μηχανικής Μάθησης που μελετά στατικά μοντέλα, τα οποία ονομάζονται βαθιά νευρωνικά δίκτυα. Στόχος των μοντέλων αυτών είναι μαθαίνοντας σύνθετες και ιεραρχικές αναπαραστάσεις ακατέργαστων δεδομένων να λύνουν ένα πρόβλημα. Τα πρώτα μοντέλα τα οποία παρουσιάστηκαν στο πεδίο αυτό ήταν κατά τη δεκαετία του 1940 με πιο γνωστό το μοντέλο Perceptron. Αυτός ο αλγόριθμος παραμένει μέχρι σήμερα η βάση για τη δημιουργία νέων μοντέλων. Η εξέλιξη του σήμερα αποτελούν τα λεγόμενα Συνελικτικά νευρωνικά δίκτυα ή αλλιώς Convolution Neural Networks τα οποία έχουν σχεδιαστεί κυρίως για την αναγνώριση προτύπων σε εικόνες. Από το 2006 και μετά έχει ξεκινήσει η σύγχρονη εποχή της βαθιάς μάθησης κατά την οποία δημιουργούνται νέες πιο σύνθετες αρχιτεκτονικές μοντέλων. Από τις ανακαλύψεις στην αναγνώριση και επεξεργασία φυσικής γλώσσας το 2011 μέχρι τα αυτοοδηγούμενα αυτοκίνητα το 2016. Μάλιστα, τα τελευταία χρόνια το Deep learning έχει τεράστια επίδραση στο πεδίο της μηχανικής μάθησης και της όρασης υπολογιστών, καθώς επιτυγχάνει επιδόσεις άνευ προηγουμένου σε προβλήματα, όπως είναι η αναγνώριση ηλικίας, η κατηγοριοποίηση και τμηματοποίηση εικόνων, η ανίχνευση και εντοπισμός αντικειμένων, η αναγνώριση ανθρώπινης δραστηριότητας (Bakalos et al, 2019). Αξίζει να σημειωθεί σε

αυτό το σημείο πως η εν λόγω πρόοδος οφείλεται στην ραγδαία αύξηση των υπολογιστικών και προγραμματιστικών πόρων αλλά και στην μεγάλη διαθεσιμότητα δεδομένων, η οποία υπάρχει στην κοινότητα. Παρόλο όμως των τεχνολογικών εξελίξεων αυτών η πολυπλοκότητα των συστημάτων είναι αντιστρόφως ανάλογη, με αποτέλεσμα το πεδίο αυτό να έχει πολλά ‘’μαύρα κουτιά’’ που ακόμα η επιστημονική κοινότητα δεν μπορεί να εξηγήσει την λειτουργία τους επακριβώς. Έτσι το μεγαλύτερο μέρος της έρευνας σε αυτόν τον τομέα πραγματοποιείται σε πειραματικό στάδιο και με τη μεθοδολογία ‘’trial and error’’.

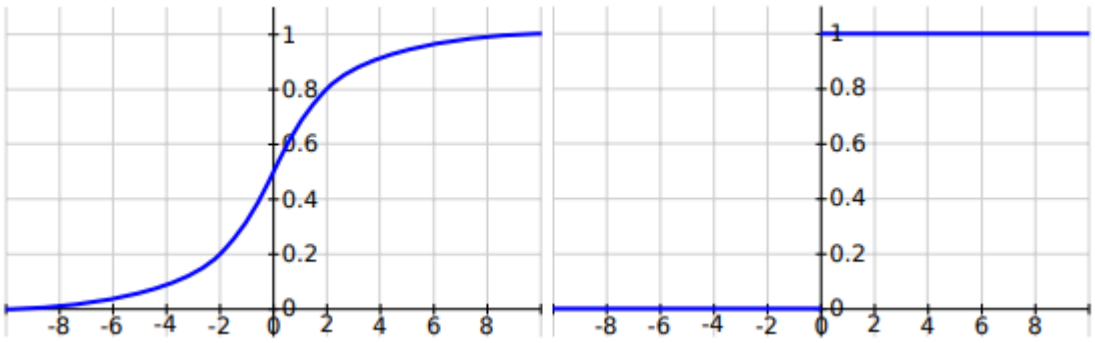
Στο κεφάλαιο 2.4.3 έγινε μια σύντομη αναφορά στα νευρωνικά δίκτυα και δόθηκε μια περιγραφή για τον τρόπο τον οποίο λειτουργούν. Ωστόσο όπως αναφέρθηκε και προηγουμένως υπάρχουν πολλά είδη και πολλές αρχιτεκτονικές, στη συνέχεια θα αναλυθούν τα πιο σημαντικά δίκτυα αλλά και χαρακτηριστικά αυτών που βρίσκουν εφαρμογή σε αυτήν την διπλωματική.

3.2 Νευρώνας Perceptron

Το πιο γνωστό και απλό νευρωνικό δίκτυο αποτελεί ο λεγόμενος νευρώνας Perceptron. Πιο συγκεκριμένα, το δίκτυο στην συγκεκριμένη περίπτωση αποτελείται από έναν και μόνο νευρώνα που ικανοποιεί την εξίσωση 3.1**:**

$$y = f_a(\sum_{j=1}^{n} w_j * o_j + b), \qquad \textit{(Εξίσωση 3.2-1)}$$

Στο παρακάτω διάγραμμα φαίνονται οι έξοδοι των δύο πιο συχνών συναρτήσεων ενεργοποίησης f_a: η σιγμοειδής και η βηματική. Στην μια περίπτωση η σιγμοειδείς συνάρτηση παράγει μια συνεχή έξοδο, ενώ αντίθετα η βηματική παράγει δυαδική. Συνήθως -μόνος του- ο νευρώνας Perceptron χρησιμοποιεί βηματική.



Εικόνα 3.2-1: Γράφημα Σιγμοειδούς και βηματικής συνάρτησης (Vojt, 2016).

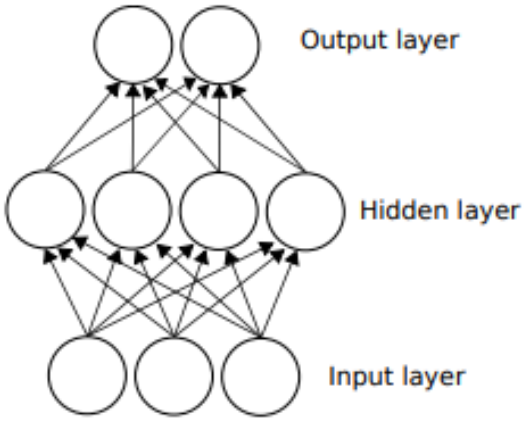
Τα νευρωνικά δίκτυα τα οποία αποτελούνται από έναν απλό νευρώνα Perceptron, όπως γίνεται κατανοητό, δεν βρίσκονται σε θέση να επεξεργαστούν περίπλοκα προβλήματα. Υπάρχει, ωστόσο, δυνατότητα να συνδυαστούν σε ένα δίκτυο πολλαπλών επιπέδων (Multilayer Perceptron) για πιο περίπλοκα προβλήματα. Όπως εξετάζεται στην ενότητα 3.3.1 η οποία αφορά την εκπαίδευση ενός MLP δικτύου, θα γίνει κατανοητό πως οι συνεχείς συναρτήσεις ενεργοποίησης είναι πιο κατάλληλες για εκπαίδευση βάσει κλίσης (gradient), με έναν απλό υπολογισμό της παραγώγου, με απώτερο στόχο τον σωστό ορισμό του συναπτικού βάρους.

3.3 Δίκτυα Πολλαπλών Στρωμάτων

Τα δίκτυα πολλαπλών στρωμάτων (MLP) ή αλλιώς τα δίκτυα εμπρόσθιας προώθησης αποτελούνται από πολλαπλά αλληλοσυνδεόμενα στρώματα νευρώνων Perceptron. Τα επίπεδα αυτά βρίσκονται σε σειρά το ένα με το άλλο, όπου κάθε νευρώνας από το ένα στρώμα συνδέεται με κάθε νευρώνα από το επόμενο στρώμα. Ο σκοπός πίσω από τον σχεδιασμό τέτοιων δικτύων είναι η αντιμετώπιση περίπλοκων προβλημάτων, τα οποία απαιτούν πολλούς υπολογισμούς. Στην περίπτωση των MLP δικτύων συνήθως η βηματική συνάρτηση δεν είναι κατάλληλη, έτσι κατά επέκταση, λόγω της συνέχειας και της μεγαλύτερής ευελιξίας που χρειάζεται, χρησιμοποιείται η σιγμοειδής. Η επιλογή συνάρτησης ενεργοποίησης στα MLP δίκτυα πρέπει να διέπεται από μη-γραμμικότητα μιας και η έξοδος πρέπει να είναι ανεξάρτητη γραμμικά από την δεδομένη είσοδο.

Οι νευρώνες Perceptron σε ένα πολυστρωματικό τέτοιο δίκτυο είναι διατεταγμένη σε αριθμό $k \geq 2$. Πιο συγκεκριμένα, έστω ένα δίκτυο M με στρώματα K. Το σύνολο των νευρώνων C χωρίζεται σε ίσα ασύνδετα υποσύνολα που ονομάζονται στρώματα. $L_1, \dots, L_k$. Με πιο μαθηματικό τρόπο, $\forall\, i, j : 1 \leq i, j \leq$

$\kappa\left(L_i \neq \emptyset \wedge L_i \cap L_j \neq \emptyset\right) \Rightarrow i = j$, όπου το L_1 είναι το στρώμα εισόδου , $L_2, \dots L_{k-1}$ τα ενδιάμεσα κρυφά στρώματα , L_k το στρώμα εξόδου και κάθε νευρώνας του στρώματος L_i συνδέεται με κάθε νευρώνα του στρώματος $L_i + 1$.



Εικόνα 3.3-1: Δίκτυο πολλαπλών στρωμάτων αποτελούμενο από τρία επίπεδα (Vojt, 2016).

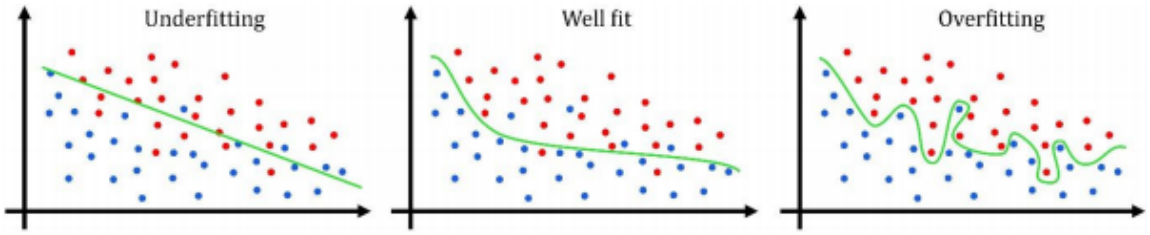
Η έξοδος του δικτύου υπολογίζεται διαδοχικά από στρώμα σε στρώμα. Ξεκινώντας από το στρώμα εισόδου εκχωρώντας σαν είσοδο $\overrightarrow{y^0} = \vec{x}$. Στην συνέχεια για κάθε L_i δίνεται σαν είσοδος $\vec{x}^i = \overrightarrow{y^{\iota-1}}$. Τα βάρη και η συνάρτηση ενεργοποίησης δηλώνεται από την αρχή στο δίκτυο, επομένως η έξοδος κάθε στρώματος εξαρτάται μόνο από την έξοδο του προηγούμενου στρώματος. Η τελική έξοδος του δικτύου παράγεται ως $\vec{y}^k$ στο στρώμα εξόδου L_k.

3.3.1 Διαδικασία Εκπαίδευσης

Η ικανότητα μάθησης είναι βασικό χαρακτηριστικό των αλγορίθμων της βαθιάς μάθησης και κατά επέκταση και των νευρωνικών δικτύων. Στόχος της διαδικασίας αυτής είναι να βρεθούν οι βέλτιστες παράμετροι (και δομή) του δικτύου, ώστε να είναι σε θέση να λύση περίπλοκα προβλήματα. Όπως γίνεται κατανοητό, πριν το δίκτυο ξεκινήσει να λειτουργεί πρέπει να δοθούν εξ ορισμού κάποιες παράμετροι σε αυτό. Οι αρχικές αυτές τιμές συνήθως επιλέγονται τυχαία, ωστόσο με την χρήση κάποιας ευρετικής μπορεί να επιτευχθούν πιο γρήγορα και βέλτιστα αποτελέσματα. Όπως εξετάστηκε στις προηγούμενες ενότητες η διαδικασία της εκπαίδευσης απαιτεί δεδομένα εκπαίδευσης, τα οποία δίνονται σαν είσοδο στο δίκτυο. Πρόκειται για επαναληπτική διαδικασία, κατά την οποία η κάθε έξοδος παράγεται από κάθε είσοδο του

συνόλου δεδομένων εκπαίδευσης, αναλύονται και το δίκτυο προσαρμόζεται ανάλογα για να παράγει καλύτερα αποτελέσματα. Περάτωση της διαδικασίας θεωρείται όταν το δίκτυο επιτύχει την επιθυμητή απόδοση πάνω στα δεδομένα εκπαίδευσης. Ωστόσο υπάρχουν διάφορες μετρικές που μπορούν να χρησιμοποιηθούν για την αξιολόγηση του μοντέλου , παρακάτω θα ορισθεί το μέσω τετραγωνικό σφάλμα.

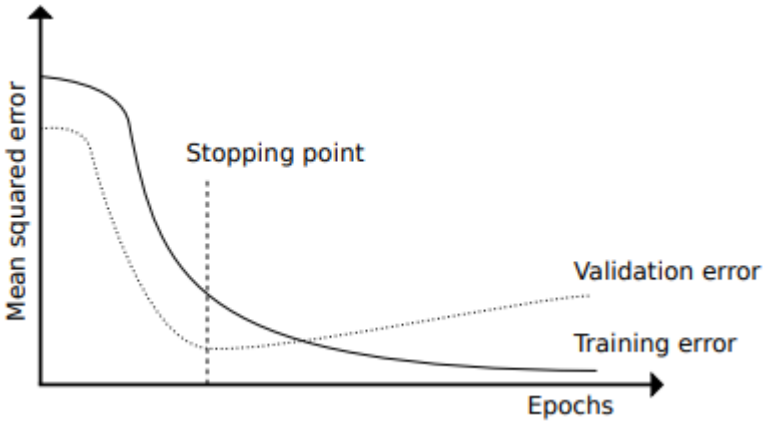
Συχνό πρόβλημα που αντιμετωπίζεται κατά την διαδικασία της εκπαίδευσης και το δίκτυο μπορεί να είναι ανεπαρκές ως προς τις δυνατότητες γενίκευσης είναι το λεγόμενο Overfitting. Ο όρος γενίκευση αφορά το πόσο καλά εφαρμόζονται οι έννοιες που έχει μάθει το μοντέλο σε ένα σύνολο δεδομένων που βλέπει για πρώτη φορά. Το Overfitting συμβαίνει όταν το μοντέλο εκπαιδεύεται σ' ένα σύνολο δεδομένων εκπαίδευσης πολύ μικρό που δεν καλύπτει ομοιόμορφα τις περιπτώσεις και τους τύπους ενός προβλήματος. Σε αυτήν την περίπτωση η διαδικασία εκπαίδευσης προσαρμόζει στο δίκτυο να εκπαιδευτεί σε τυχαία χαρακτηριστικά που υπάρχουν στο σύνολο δεδομένων. Η γενίκευση αυτή συνήθως παρατηρείται , όταν η απόδοση του δικτύου αυξάνεται ραγδαία στα δεδομένα εκπαίδευσης και μειώνεται στα δεδομένα που βλέπει για πρώτη φορά. Ένας άλλος λόγος γενίκευσης είναι ότι το δίκτυο αδυνατεί να μοντελοποιήσει τα δεδομένα εκπαίδευσης και κατά συνέπεια δε μπορεί να τα γενικεύσει με αποτέλεσμα να οδηγείτε σε underfitting.



Εικόνα 3.3-2: Γραφική Απεικόνιση underfitting και overfitting.

Για να αντιμετωπιστεί αυτό το πρόβλημα της γενίκευσης, το σύνολο δεδομένων (με ετικέτα) χωρίζεται σε δεδομένα εκπαίδευσης και δεδομένα επικύρωσης. Ο κύριος λόγος για τον οποίο χρησιμοποιείται το δεύτερο είναι ότι δείχνει τον ρυθμό σφάλματος στα δεδομένα ανεξάρτητα των δεδομένων εκπαίδευσης. Μια μελέτη από τον Guyon προτείνει ότι η βέλτιστη αναλογία μεταξύ των δεδομένων εκπαίδευσης και των δεδομένων επικύρωσης εξαρτάται από τον αριθμό των κλάσεων αλλά και την περιπλοκότητα των χαρακτηριστικών τους. (Guyon, 1997). Ωστόσο λόγω ότι η εκτίμηση της περιπλοκότητας των χαρακτηριστικών είναι αρκετά δύσκολη συνήθως προτείνονται οι αναλογίες 80:20. Όπου 80% από το σύνολο δεδομένων αφορά τα δεδομένα εκπαίδευσης και το 20% αφορά τα δεδομένα επικύρωσης. Αυτό δεν αναιρεί όμως ότι για κάποιο πρόβλημα η βέλτιστη αναλογία μπορεί να είναι διαφορετική.

Ένα ακόμα πολύ σημαντικό χαρακτηριστικό για την σωστή εκπαίδευση ενός πολυστρωματικού δικτύου είναι πότε η διαδικασία εκπαίδευσης πρέπει να σταματήσει. Αυτό συμβαίνει συνήθως όταν τα σφάλματα στα δεδομένα επικύρωσης εκμηδενιστούν (σχήμα 3.3.3).



Εικόνα 3.3-3: Γράφημα που συγκρίνει την εξέλιξη του σφάλματος εκπαίδευσης σε σχέση με το σφάλμα επικύρωσης.

Πιο συγκεκριμένα, για να οριστεί η διαδικασία εκπαίδευσης πιο τυπικά και με περισσότερες λεπτομέρειες, ας ορίσουμε P ένα σύνολο δεδομένων με ετικέτες $(\vec{x}^p, \vec{d}^p)$, όπου $\vec{x}^p$ το διάνυσμα εισόδου, και $\vec{d}^p$ το επιθυμητό διάνυσμα εξόδου, με $1 \leq p \leq P$. Δεδομένου της τρέχουσα κατάστασης του δικτύου η είσοδος $\vec{x}^p$ αποδίδεται στην έξοδο $\vec{y}^p$. Όπου $\vec{y}^p$ είναι επιθυμητό να είναι όσο πιο δυνατό κοντά στην επιθυμητή έξοδο $\vec{d}^p$. Έτσι ορίζεται το τετραγωνικό σφάλμα για κάθε νευρώνα ως η διαφορά της πραγματικής εξόδου από την επιθυμητή:.

$$\vec{e}_j^p = \vec{y}_j^p - \vec{d}_j^p$$

Γενικεύοντας τον τύπο για όλο το νευρωνικό δίκτυο προκύπτει:

$$E_p = \frac{1}{m_k} \sum_{j=1}^{m_k} (\vec{y}_j^p - \vec{d}_j^p),$$

Όπου m_k είναι ο συνολικός αριθμός των νευρώνων στο στρώμα εξόδου. Στην περίπτωση που η τιμές εξόδου είναι ίδιες με τις επιθυμητές τότε το τετραγωνικό σφάλμα είναι μηδενικό. Με άλλα λόγια

$$\forall j : E_p = 0 \Leftrightarrow \vec{e}_j^p = 0 \Leftrightarrow \vec{y}_j^p = \vec{d}_j^p$$

Επίσης, πολλές φορές είναι χρήσιμο να υπολογίζεται το μέσο σφάλμα για την αξιολόγηση του δικτύου που προκύπτει από τον παρακάτω γενικευμένο τύπο:

$$E_{avg} = \frac{1}{P}\sum_{p=1}^{P} E_p$$

Κατά τη διάρκεια της εκπαίδευσης, για κάθε διασυνδεδεμένο ζεύγος νευρώνων (i, j), ισχύει ότι i είναι ο νευρώνας στο στρώμα l, j ένας νευρώνα στο στρώμα $l + 1$ και $w_{i,j}$ το βάρος σύνδεσης τους για να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα τους E_{avg}. Εφόσον η συνάρτηση ενεργοποίησης είναι ολοκληρωτικά διαφοροποιήσιμη, η E_{avg} είναι επίσης διαφοροποιήσιμη. Συνεπώς, κατά την ρύθμιση του βάρους $w_{i,j}$ του νευρώνα j που βρίσκεται στο στρώμα k, ενδιαφερόμαστε για την μερική παράγωγο:

$$\frac{\partial E_{avg}}{\partial w_{i,j}} = \frac{1}{P}\frac{\partial}{\partial w_{i,j}}\sum_{p=1}^{P} E_p = \frac{1}{P}\sum_{p=1}^{P}\frac{\partial E_p}{\partial w_{i,j}}, \qquad \textit{(Εξίσωση 3.3-1)}$$

Για να μπορέσει να προσαρμοστεί το δίκτυο μετά από κάθε μοτίβο εισόδου, θα πρέπει να υπολογίζεται η παράγωγος για κάθε δοθέν μοτίβο p και του αντίστοιχου τετραγωνικού σφάλματος E_p. Έτσι με την εφαρμογή του κανόνα αλυσίδας και συνδυάζοντας την εξίσωση 3.2 προκύπτει:

$$\frac{\partial E}{\partial w_{i,j}} = (y_j - d_j) f'(\xi_j) y_i, \qquad \textit{(Εξίσωση 3.3-2)}$$

Για ευκολία, ο Haykin ορίζει τη λεγόμενη τοπική κλίση δ_j για τον νευρώνα j στο επίπεδο εξόδου, ως η ακόλουθη σχέση (Haykin, 1999):

$$\delta_j = \frac{\partial E}{\partial \xi_j} = \frac{\partial E}{(\partial y_j \partial \xi_j)} = (y_j - d_j) f'(\xi_j), \qquad \textit{(Εξίσωση 3.3-3)}$$

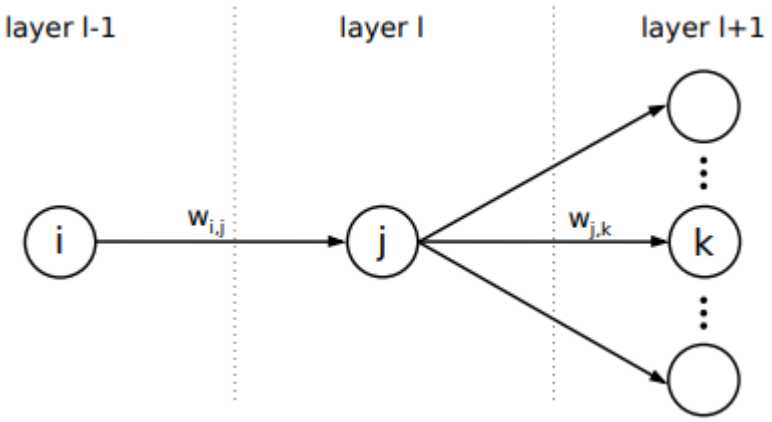
Έτσι η εξίσωση 3.3.2 γίνεται:

$$\frac{\partial E}{\partial w_{i,j}} = \delta_j y_i, \qquad \textit{(Εξίσωση 3.3-4)}$$

Έπειτα βάσει της εξίσωσης 3.3.4 είναι δυνατό να υπολογιστεί η κλήση της συνάρτησης σφάλματος για καθένα από τα δοθέντα μοτίβα p. Με αυτόν τον τρόπο προσαρμόζεται το βάρος από την αντίθετη κατεύθυνση της κλίσης. Μέχρι στιγμής, όμως, η έξοδος για κάθε είσοδο είναι πολύ ασταθής, προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα χρησιμοποιείται μια παράμετρος εκμάθησης $0 < \eta < 1$ και έτσι εν τέλει το βάρος υπολογίζεται από τον παρακάτω τύπο:

$$\Delta w_{i,j} = -\eta \frac{\partial E}{\partial w_{(i,j)}} = -\eta \delta_j y_i, \qquad \textit{(Εξίσωση 3.3-5)}$$

Αξίζει να σημειωθεί σε αυτό το σημείο πως η ρύθμιση του βάρους $\Delta w_{i,j}$ ισχύει για τους νευρώνες του στρώματος εξόδου. Οι προσαρμογές στα βάρη στο ενδιάμεσο στρώμα είναι πιο περίπλοκες. Χαρακτηριστικό παράδειγμα αποτελούν οι τρείς νευρώνες i, j και k. Όλοι ακολουθούν μεταξύ τους την ίδια διαδρομή κατά μήκος των στρωμάτων $l-1, l$ και $l+1$ αντίστοιχα, όπως φαίνεται στο παρακάτω σχήμα 3.3.4. Η προσαρμογή του βάρος $w_{i,j}$ θα πρέπει να γίνει πολύ προσεκτικά,καθώς εκτός από το γεγονός ότι επηρεάζεται η έξοδος του κάθε νευρώνα επηρεάζονται και όλες οι έξοδοι σε όλα τα ακόλουθα στρώματα l.



Εικόνα 3.3-4: Αναπαράσταση που δείχνει πως μια αλλαγή στο βάρος w,j του νευρώνα μέσα στο κρυφό επίπεδο l-1 επηρεάζει το βάρος w_(j,k) του νευρώνα στο επίπεδο l.

Παρόλο που η εφαρμογή της εξίσωσης 3.3.2 εφαρμόζεται στα ενδιάμεσα στρώματα, είναι αναγκαίο να ορισθεί ξανά η τοπική κλίση. Στην εξίσωση 3.3.3 χρησιμοποιείται η επιθυμητή έξοδος d_j για να υπολογιστεί το $\frac{\partial E}{\partial y_j}$. Στα ενδιάμεσα στρώματα, όπως γίνεται κατανοητό, δεν υπάρχει επιθυμητή έξοδος, επομένως χρειάζεται να γίνει ένα βήμα πίσω και να ορισθεί η δ_j:

$$\delta_j = \frac{\partial E}{\partial y_j} f'(\xi_j), \qquad \textit{(Εξίσωση 3.3-6)}$$

Σε αυτό το σημείο χρειάζεται να ορισθεί ξανά η $\frac{\partial E}{\partial y_j}$ για τα ενδιάμεσα στρώματα. Έτσι δεδομένου των νευρώνων i, j και k για κάθε στρώμα όπως φαίνεται και στο σχήμα 3.3.4 προκύπτει:

$$\frac{\partial E}{\partial \xi_j} = \sum_{k=1}^{m_{l+1}} \delta_k w_{j,k}, \qquad \textit{(Εξίσωση 3.3-7)}$$

Από τις εξισώσεις 3.3.6 και 3.3.5 προκύπτει:

$$\delta_j = \left(\sum_{k=1}^{m_{l+1}} \delta_k w_{j,k}\right) f'(\xi_j), \qquad \textit{(Εξίσωση 3.3-8)}$$

Καταλήγοντας, η εξίσωση 3.8 υποδεικνύει την τοπική κλίση του νευρώνα k στο στρώμα $l+1$, μπορεί να υπολογιστεί η τοπική κλίση του νευρώνα j στο στρώμα l. Αυτό το γεγονός επιτρέπει στο δίκτυο να λειτουργεί αναδρομικά, ώστε να προσαρμόζει σωστά τα βάρη (προς τα πίσω) σε όλα τα στρώματα. Επομένως, συνοπτικά η προσαρμογή του βάρους που ισχύει για όλα τα στρώματα μπορεί να γραφεί:

$$\Delta \mathrm{w}_{\mathrm{i},j} \;=\; -\eta \frac{\partial \mathrm{E}}{\partial \mathrm{w}_{(\mathrm{i,j})}} = \; -\eta \delta_{\mathrm{j}} \mathrm{y}_{\mathrm{i}}$$

Όπου,

$$\forall j \; \sigma το \; \sigma τρώμα \; l \; < \; L \; , \qquad \delta_j = \left(\sum_{k=1}^{m_{l+1}} \delta_k w_{j,k}\right) f'(\xi_j)$$

$$\forall j \; \sigma το \; \sigma τρώμα \; \; L \; , \qquad \delta_j = \left(y_j - d_j\right) f'(\xi_j)$$

3.3.2 Αλγόριθμος Μετάδοσης Προς τα Πίσω

Ο αλγόριθμος μετάδοσης προς τα πίσω ή αλλιώς backpropagation έχει ως στόχο την σωστή προσαρμογή των συνοπτικών βαρών, ώστε το δίκτυο να παράγει την επιθυμητή έξοδο. Με άλλα λόγια το αποτέλεσμα αυτού του αλγόριθμου αποτελεί ένα νευρωνικό δίκτυο, το οποίο έχει διαμορφωθεί με τέτοιον τρόπο που να ελαχιστοποιεί την συνάρτηση σφάλματος. Είναι σημαντικό να προστεθεί ότι ο όρος «μετάδοση προς τα πίσω» χρησιμοποιείται στη βιβλιογραφία των νευρωνικών υπολογιστικών συστημάτων, για να υποδηλώσει μια ποικιλία διαφορετικών πραγμάτων.

Η εκπαίδευση η οποία αρμόζει να πραγματοποιηθεί στο δίκτυο απαιτεί ένα σύνολο δεδομένων με ετικέτες και κατά επέκταση να είναι εποπτευόμενη. Πριν ξεκινήσει ο αλγόριθμος να εκτελείται, είναι αναγκαίο να αρχικοποιηθούν τα συναπτικά βάρη με κάποιες τιμές. Η αρχικοποίηση δεν αποτελεί μέρος του αλγορίθμου, καθώς μπορεί να υπάρχουν διαφορετικές προσεγγίσεις, με την πιο κοινή να είναι η τυχαία. Ύστερα ο αλγόριθμος είναι έτοιμος να εκτελεσθεί.

Κάθε διάνυσμα εισόδου με την αντίστοιχη ετικέτα του $(\overrightarrow{x_p}, \overrightarrow{d_p},)$ επεξεργάζεται διαδοχικά σε δύο φάσεις. Στην πρώτη φάση ή αλλιώς στην φάση της προώθησης το διάνυσμα εισόδου τοποθετείται στην είσοδο του δικτύου, $\vec{y}^0 = \overrightarrow{x_p}$. Στην συνέχεια το δίκτυο υπολογίζει την έξοδο $\vec{y}^L$. Στόχος της φάσης προώθησης αποτελεί ο υπολογισμός της εξόδου για το δοθέν διάνυσμα εισόδου, χωρίς το δίκτυο να έχει προσαρμοστεί καθόλου. Έπειτα, η δεύτερη φάση μετάδοσης προς τα πίσω ξεκινάει, με σκοπό να προσαρμόσει τα βάρη του δικτύου έτσι ώστε να επιτύχει καλύτερη απόδοση πάνω στα δεδομένα εισαγωγής.

Η διαδικασία αυτή της μάθησης πραγματοποιείται για πολλές εποχές. Σε κάθε εποχή, το δίκτυο τροφοδοτείται και επεξεργάζεται το σύνολο δεδομένων εκπαίδευσης. Η διάρκεια μιας εποχής εξαρτάται άμεσα από το μέγεθος και την αρχιτεκτονική του δικτύου, καθώς και από το μέγεθος των δεδομένων εκπαίδευσης. Επομένως, γίνεται κατανοητό πως η σημαντική «απόφαση» του αλγορίθμου είναι πότε θα πρέπει να σταματήσει η διαδικασία της εκπαίδευσης. Αξίζει να προστεθεί σε αυτό το σημείο, ότι μετά από κάθε εποχή το σφάλμα στο σύνολο δεδομένων επικυρώνεται από το σύνολο δεδομένων επικύρωσης. Μόλις αυτό το σφάλμα αρχίζει να αυξάνεται, αυτό σημαίνει ότι έχει επιτευχθεί ένα ελάχιστο σφάλμα, τότε συνήθως η διαδικασία σταματάει.

Η διαδικασία μετάδοσης με συναπτικά βάρη προς τα πίσω είναι η ακόλουθη:

Θεωρείται ως **Είσοδο**:

- Σύνολο Δεδομένων εκπαίδευσης με τις ετικέτες τους $(\overrightarrow{x^p}, \overrightarrow{d^p},)$, $1 \leq p \leq P$

- Σύνολο Δεδομένων επικύρωσης με τις ετικέτες τους $(\overrightarrow{v^q}, \overrightarrow{d^q},)$, $1 \leq q \leq Q$
- Συνάρτηση Ενεργοποίησης f με την παράγωγο της $f'(\xi)$
  - Για παράδειγμα , η σιγμοειδής: $f(\xi) = 1/(1 + e^{-\xi})$, $f'(\xi) = f(\xi)(1 - f(\xi)$
- Παράμετρος εκμάθησης $\eta \in (0, 1)$
- Δίκτυο εμπρόσθιας προώθησης M με τυχαία αρχικοποιήσεις στα συνοπτικά βάρη

Επιθυμητή **Έξοδος**:

- Εκπαιδευμένο νευρωνικό δίκτυο εμπρόσθιας προώθησης M

Αλγόριθμος:

1. Θέτουμε $\vec{y}^0 = \infty$
2. Ξεκίνημα νέας εποχής
3. Για κάθε $p \in \{1, \dots, P\}$ που αναπαρίσταται στο σύνολο δεδομένων $(\overrightarrow{x^p}, \overrightarrow{d^p},)$

   - *Φάση προώθησης*
     - Για $l = 1,2 \dots, k$ υπολόγισε την έξοδο του δικτύου M για κάθε $\vec{x}^p$ με τον ακόλουθο τρόπο:

$$\forall j \in L_\iota \, \upsilon\pi o\lambda \acute{o}\gamma\iota\sigma\varepsilon \, \xi_j = \sum_{i=1}^{n} w_i x_i + \Theta \;\; \kappa\alpha\iota \, y_j = f(\xi)$$

   - *Φάση μετάδοσης προς τα πίσω*
     - $\forall i \in \; L_\kappa - 1, j \in L_\kappa \, \upsilon\pi o\lambda \acute{o}\gamma\iota\sigma\varepsilon \, \delta_j = \left(y_j \; - \; d_j^p \; \right) f'\left(\xi_j\right) \kappa\alpha\iota \, \Delta w_{i,j} = -\eta \delta_j y_i$
     - $\Gamma\iota\alpha \; l = \kappa - 1, \dots, 1$

$$\forall i \in L_{l-1}, j \; \in \; L_l \; \upsilon\pi o\lambda \acute{o}\gamma\iota\sigma\varepsilon \; \delta_j \; = \left(\sum_{k=1}^{m_{l+1}} \delta_k w_{j,k} \right) f'\left(\xi_j\right), \qquad \Delta wi, j \; = \; -\eta \delta_j y_i$$

     - $\Gamma\iota\alpha \; l = 1,2, \dots, \kappa \; \forall (i, j) \in L_{l-1} \; \times L_l \; \pi\rho o\sigma \acute{a}\rho\mu o\sigma\varepsilon \, w_{i,j} \; \sigma \acute{\upsilon}\mu\varphi\omega\nu\alpha \; \mu\varepsilon \; \tau o \; \Delta w_{i,j}$
     - $\Upsilon\pi o\lambda \acute{o}\gamma\iota\sigma\varepsilon \; \tau o \; \sigma\varphi \acute{a}\lambda\mu\alpha \; \gamma\iota\alpha \; \tau o \; \delta\iota \acute{a}\nu\upsilon\sigma\mu\alpha \; p$: $E_p = 1/m_k \sum_{j=1}^{m_k} \left(y_j^p - d_j^p\right)^2$

4. Θέσε $E_{prev} = E_{avg} \; \kappa\alpha\iota \; \upsilon\pi o\lambda \acute{o}\gamma\iota\sigma\varepsilon \; \tau o \; \nu \acute{e}o \; E_{avg} = 1/P \sum_{p=1}^{P} E_p$ για το σύνολο δεδομένων επικύρωσης

5. Αν $E_{avg} < E_{prev} \; \pi \acute{\eta} \gamma\alpha\iota\nu\varepsilon \; \sigma\tau o \; \beta \acute{\eta}\mu\alpha \; 2$

6. Τέλος διαδικασίας

3.4 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional neural networks η CNNs) αποτελούν από τα σημαντικότερα είδη νευρωνικών δικτύων, καθώς έχουν μεγάλη αποτελεσματικότητα σε εφαρμογές που έχουν να κάνουν με αναγνώριση εικόνας και μοτίβων. Πρόκειται για πολυστρωματικά δίκτυα εμπρόσθιας προώθησης ειδικά σχεδιασμένα για να αναγνωρίζουν δισδιάστατα χαρακτηριστικά. Η αρχιτεκτονική τους έχει εμπνευστή από την μελέτη των (Hubel & Wiesel, 1968) για την νευροβιολογική επεξεργασία σήματος του οπτικού φλοιού της γάτας.

Ένα συνελικτικό δίκτυο αναγνωρίζει τα δισδιάστατα χαρακτηριστικά σε εικόνες κυρίως, επομένως η αρχιτεκτονική του βασίζεται σε μια δισδιάστατη ορθογώνια εικόνα, η οποία αποτελείται από εικονοστοιχεία (pixels). Κάθε εικονοστοιχείο φέρει πληροφορίες χρώματος. Το χρώμα μπορεί να αναπαρασταθεί από πολλά κανάλια (π.χ. για της έγχρωμες (RGB) εικόνες τα κανάλια είναι τρία, ένα για κάθε χρώμα κόκκινο, πράσινο και μπλε). Στη συνέχεια για λόγους απλότητας θα εξετασθούν εικόνες με ένα μόνο κανάλι (αποχρώσεις του γκρι).

Αναφορικά με τους νευρώνες σε ένα συνελικτικό δίκτυο εξετάζουν κάθε φορά ένα μικρό μέρος της εικόνας, το οποίο ονομάζεται υπό εικόνα (sub-image). Έπειτα από αυτές τις υπό εικόνες εξάγονται χαρακτηριστικά, τα οποία είναι απαραίτητα για τη διαδικασία της αναγνώρισης του δικτύου. Τα εν λόγω χαρακτηριστικά για παράδειγμα μπορεί να είναι κάποια γραμμή στην εικόνα, ένας κύκλος ή μια γωνία. Στη συνέχεια αυτά τα χαρακτηριστικά συλλέγονται από έναν χάρτη χαρακτηριστικών (feature map) και ο συνδυασμός όλων αυτών τελικά οδηγεί στην ταξινόμηση της εικόνας.

Υπάρχουν δύο είδη επιπέδων στα συνελικτικά δίκτυα που αποτελούνται από τους χάρτες χαρακτηριστικών. Το πρώτο είδος αποτελείται από το συνελικτικό επίπεδο (convolutional layer), στο οποίο πραγματοποιείται η εξαγωγή και η αναγνώριση των χαρακτηριστικών. Συνήθως το επίπεδο αυτό, αποτελείται από πολλούς χάρτες χαρακτηριστικών που καθ’ ένας αντιστοιχεί σε συγκεκριμένο χαρακτηριστικό. Το δεύτερο είδος επιπέδου ονομάζεται επίπεδο δειγματοληψίας (subsampling layer) και ακολουθεί πάντα μετά το συνελικτικό επίπεδο. Ειδικότερα, αποτελείται από τον ίδιο αριθμό χαρτών χαρακτηριστικών, όπου ο κάθε χάρτης από το συνελικτικό επίπεδο χρησιμοποιείται ως είσοδος στον αντίστοιχο χάρτη στο επίπεδο δειγματοληψίας. Ανάλογα την αρχιτεκτονική του δικτύου και το βάθος του, η εναλλαγή των συνελικτικών και των δειγματοληπτικών επιπέδων υπάρχει μέχρι να υπάρξει το τελευταίο

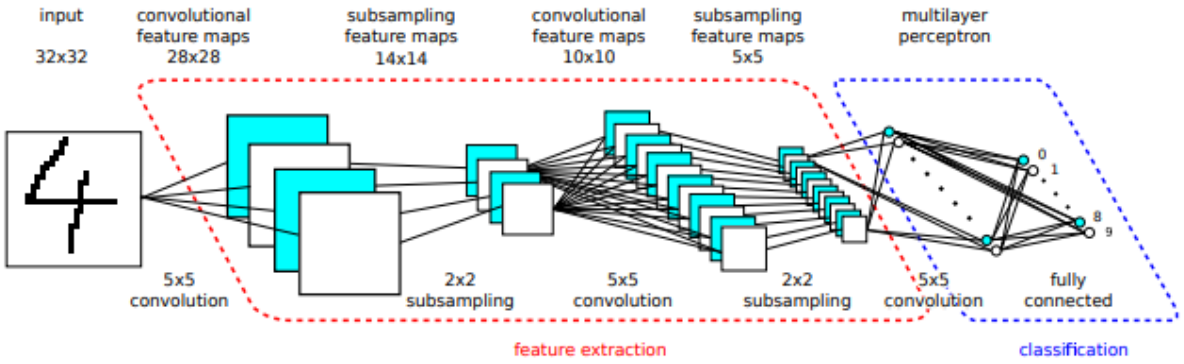
δειγματοληπτικό. Μετά από το τελευταίο δειγματοληπτικό επίπεδο μπορούν να υπάρχουν και άλλα πλήρως συνδεδεμένα επίπεδα -με διαφορετικές λειτουργίες από τα προηγούμενα- μέχρι το στρώμα εξόδου. (LeCun, Bottou, Bengio, & Haffner, 1998)

3.4.1 Αρχιτεκτονική Νευρωνικών Συνελικτικών Δικτύων

Για να εξηγηθεί η αρχιτεκτονική ενός συνελικτικού δικτύου, είναι αναγκαίο να εξετασθούν τα δεδομένα εικόνας στο στρώμα εισόδου. Εξετάζοντας σε υψηλό επίπεδο την αρχιτεκτονική τέτοιων δικτύων, θα μπορούσε να υποστηριχθεί πως χωρίζεται σε δύο σημαντικά μέρη, με το εκάστοτε μέρος να έχει σχεδιαστεί για διαφορετικό σκοπό. Το πρώτο μέρος αφορά την εξαγωγή χαρακτηριστικών από τις εικόνες, μέσω των συνελικτικών και δειγματοληπτικών επιπέδων, ενώ το δεύτερο μέρος αφορά την ταξινόμηση αυτών των χαρακτηριστικών. Στο συγκεκριμένο μέρος θα χρησιμοποιηθεί το πολυστρωματικό δίκτυο.

Η εξαγωγή χαρακτηριστικών από μια εικόνα αποτελεί τη βασική λειτουργία των συνελικτικών νευρωνικών δικτύων και ταυτοχρόνως βασική διαφορά σε σχέση με τα υπόλοιπα είδη δικτύων, παρακάτω θα εξηγηθεί ο τρόπος λειτουργίας των συνελικτικών δικτύων. Κάθε εικόνα εισόδου στο δίκτυο απαρτίζεται από εικονοστοιχεία, τα οποία με τη σειρά τους αποτελούν εισόδους των νευρώνων, ομαδοποιημένων σε χάρτες χαρακτηριστικών, στο πρώτο συνελικτικό επίπεδο. Οι νευρώνες στον χάρτη χαρακτηριστικών είναι οργανωμένη σε δισδιάστατη μορφή και όλοι οι νευρώνες, οι οποίοι βρίσκονται στον ίδιο χάρτη μοιράζονται τα βάρη τους. Αυτό το γεγονός επιτρέπει τη βελτιστοποίηση της αρχιτεκτονικής του δικτύου, ώστε να χρειάζεται λιγότερη μνήμη, αλλά συγχρόνως να έχει καλύτερη απόδοση. Με αυτόν τον τρόπο κάθε νευρώνας σε δεδομένο χάρτη χαρακτηριστικών αναμένεται να αναγνωρίσει το ίδιο χαρακτηριστικό. Πιο συγκεκριμένα, το χαρακτηριστικό αναγνωρίζεται από έναν συνδυασμό βαρών, το οποίο ουσιαστικά φιλτράρει τις εισόδους του νευρώνα. Μοιράζοντας το ίδιο βάρος σε όλους του νευρώνες του συγκεκριμένου χάρτη χαρακτηριστικών διασφαλίζεται ότι χρησιμοποιείται το ίδιο φιλτράρισμα για κάθε εικονοστοιχείο.

Έστω ότι F_1 το σύνολο των χαρτών χαρακτηριστικών στο στρώμα l. Όλοι οι χάρτες έχουν το ίδιο μέγεθος και αναπαρίστανται ως $m_l \times n_l$. Επιπλέον, $\forall\, 0 \leq \mathrm{i} < \mathrm{m}_l\,, 0 \leq \mathrm{j} < \mathrm{n}$ το $y_{i,j}^{\varphi,l}$ η έξοδος του νευρώνα (i, j, φ, l), στην θέση (i, j) στο χάρτη χαρακτηριστικών φ στο επίπεδο l.



Εικόνα 3.4-1: Παράδειγμα αρχιτεκτονικής ενός συνελικτικού δικτύου που έχει σχεδιαστεί για την κατηγοριοποίηση χειρόγραφων ψηφίων.

Θεωρώντας μια εικόνα στην είσοδο ως έξοδο του στρώματος με μηδενικό δείκτη και με μέγεθος $m_0 \times n_0$. Το επίπεδο εισόδου μπορεί να ληφθεί υπόψιν ως επίπεδο δειγματοληψίας με έναν και μοναδικό χάρτη χαρακτηριστικών. Στη συνέχεια το επίπεδο εισόδου ακολουθείται από εναλλαγές μεταξύ του συνελικτικού επιπέδου και του δειγματοληπτικού. Όπου οι νευρώνες του χάρτη χαρακτηριστικών φ στο συνελικτικό επίπεδο l παίρνουν σαν είσοδο το σύνολο των χαρακτηριστικών χαρτών $F'_{\varphi,l}$, που έχει προέλθει από το προηγούμενο δειγματοληπτικό επίπεδο ($0 \neq F'_{\varphi,l} \subset F_{l-1}$). Στην συνέχεια το συνελικτικό επίπεδο ακολουθείτε από το δειγματοληπτικό με στόχο να μειώσει το μέγεθος του χάρτη χαρακτηριστικών, μέχρι η διαδικασία να φτάσει στο τελευταίο δειγματοληπτικό επίπεδο πραγματοποιείτε και η περάτωση της.

Σε αυτό το σημείο ξεκινάει το μέρος της ταξινόμησης. Η ταξινόμηση πραγματοποιείται, όπως φαίνεται και στο σχήμα 3.4.1, με ένα πλήρως συνδεδεμένο πολυστρωματικό δίκτυο, το οποίο ταξινομεί τις εισόδους του σύμφωνα με τα εξαγόμενα χαρακτηριστικά από τα προηγούμενα επίπεδα.

3.4.2 Συνελικτικό Επίπεδο

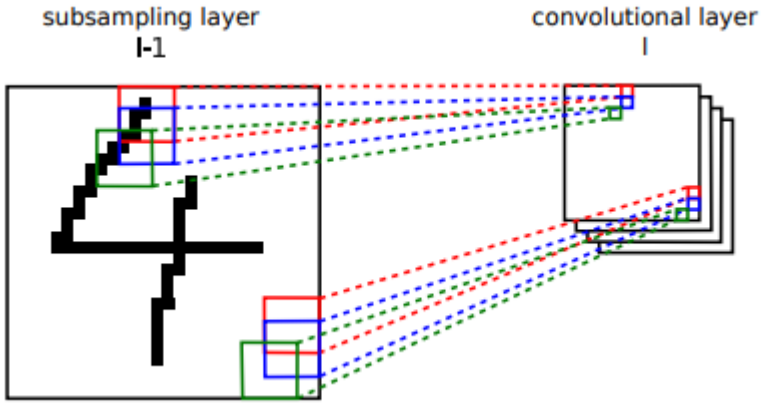
Ο σκοπός των συνελικτικών επιπέδων, όπως παρουσιάστηκε και στην ενότητα 3.4.1, είναι η εξαγωγή χαρακτηριστικών από τα δεδομένα εισόδου, τα οποία δίνονται στο δίκτυο. Αναλυτικότερα, το επίπεδο αυτό αποτελείται από πολλαπλούς χάρτες χαρακτηριστικών, όπου ο καθένας αναγνωρίζει

συγκεκριμένο χαρακτηριστικό. Η αναγνώριση των χαρακτηριστικών αυτών μπορεί να θεωρηθεί ως ένα φιλτράρισμα της υπό εικόνας, το οποίο επιτυγχάνεται με την αναπροσαρμογή των συνοπτικών βαρών.

Ο αριθμός των χαρτών χαρακτηριστικών σε ένα συνελικτικό δίκτυο αποτελεί έναν από τους πιο σημαντικούς παράγοντες, οι οποίοι πρέπει να ληφθούν υπόψιν για την αρχιτεκτονική του δικτύου. Ο βέλτιστος αριθμός εξαρτάται από την φύση του προβλήματος, καθώς και των δεδομένων που είναι διαθέσιμα και προκύπτει από πειραματική διαδικασία. Αξίζει να προστεθεί ότι κατά κανόνα, οι πιο σύνθετες εικόνες αναγνωρίζονται με το να χρησιμοποιούνται περισσότερα χαρακτηριστικά, ενώ αντίθετα οι πιο απλές με λιγότερα χαρακτηριστικά.

Ας υποθέσουμε ένα συνελικτικό επίπεδο l με ένα σύνολο χαρτών χαρακτηριστικών F_l. Αυτό το επίπεδο προηγείται πάντα από ένα δειγματοληπτικό επίπεδο με σύνολο χαρτών χαρακτηριστικών F_{l-1}. Για λόγους απλότητας και διάκρισης μεταξύ των επιπέδων, θα γίνεται αναφορά σε ένα σύνολο χαρτών χαρακτηριστικών του συνελικτικού επιπέδου ως συνελικτικός χάρτης και αντίστοιχα για σύνολο χαρτών χαρακτηριστικών ενός δειγματοληπτικού επιπέδου ως δειγματοληπτικός χάρτης.

Όλοι οι χάρτες χαρακτηριστικών στο ίδιο επίπεδο έχουν το ίδιο μέγεθος. Το μέγεθος αυτό καθορίζεται από το μέγεθος των δειγματοληπτικών χαρτών στο προηγούμενο επίπεδο, από την παράμετρο επιπέδου r_c^l αλλά και από την παράμετρο επικάλυψης s_c^l. Κάθε νευρώνας στο συνελικτικό χάρτη παίρνει την είσοδο του από $\left(r_c^l\right)^2$ νευρώνες που αντιστοιχούν στους δειγματοληπτικούς του προηγούμενου επιπέδου. Έτσι αυτοί οι νευρώνες εισόδου δημιουργούν ένα τετράγωνο διάστασης $r_c^l \times r_c^l$. Όσον αφορά την παράμετρο επικάλυψης $s_c^l \geq 1$ καθορίζει την τοπολογία των νευρώνων και πως αυτοί χωρίζουν τα τετράγωνα μεταξύ τους (σχ. 3.4.2**).**



Εικόνα 3.4-2: Αναπαράσταση της λειτουργίας της συνέλιξης στο χάρτη χαρακτηριστικών μεγέθους 28 x 28 νευρώνων.

Όπως φαίνεται και στο παραπάνω σχήμα οι γειτονικοί νευρώνες σε ένα συνελικτικό χάρτη έχουν γειτονικά πεδία που αλληλεπικαλύπτονται ακριβώς ανά $(r_c^l - s_c^l)$ νευρώνες. Αν οι διαστάσεις του δειγματοληπτικού χάρτη είναι (m_{l-1}, n_{l-1}) , τότε οι συνελικτικοί χάρτες οι οποίοι έπονται θα πρέπει να έχουν τις εξής διαστάσεις:

$$(m_l, n_l) = (\frac{m_{l-1} - r_c^l + 1}{s_c^l}, \frac{n_{l-1} - r_c^l + 1}{s_c^l})$$

Έτσι κάθε νευρώνας (i, j, φ, l) συνδέεται με όλους τους προηγούμενους νευρώνες με αυτόν τον τρόπο $(is_c^l + \Delta i, js_c^l + \Delta j, \varphi', l-1)$, όπου $\varphi' \in F_{\varphi,l}$ και $0 \leq \Delta i, \Delta j < r_c^l$ και με το ίδιο συνοπτικό βάρος $w_{\Delta i, \Delta j}^{\varphi, \varphi', l}$. Σε αυτό το σημείο, μπορεί να υπολογιστεί ο αριθμός των βαρών που απαιτούνται για το συνελικτικό στρώμα l. Ανεξάρτητα από τις διαστάσεις των συνελικτικών χαρτών, ο αριθμός των συνελικτικών βαρών δίνεται από την παρακάτω εξίσωση:

$$|W_i| = |F_i| * (r_c')^2$$

Άρα η έξοδος $y_{i,j}^{\varphi,l}$ και το $\xi_{i,j}^{\varphi,l}$ του νευρώνα (i, j, φ, l) υπολογίζεται από την εξίσωση:

$$y_{i,j}^{\varphi,l} = \xi_{i,j}^{\varphi,l} = \sum_{\varphi' \in F_{\varphi,l-1}} \sum_{\Delta i=0}^{r_c^l - 1} \sum_{\Delta j=0}^{r_c^l - 1} y\_(is_c^l + \Delta \mathrm{i}, \mathrm{js_c^l} + \Delta j^{w_{\Delta i, \Delta j}^{\varphi, \varphi', l}})$$

Δεδομένου των παραπάνω ορισμών, υπάρχουν τρεις σημαντικές παράμετροι που πρέπει να επιλεχθούν με προσοχή, ώστε το δίκτυο να είναι αποδοτικό και αποτελεσματικό.

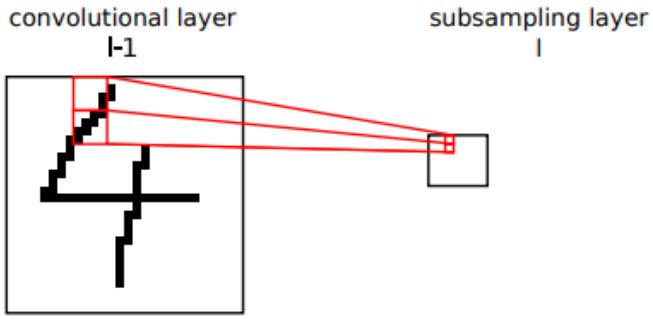
- $|F_l|$: Ο αριθμός των συνελικτικών χαρτών που αντιστοιχούν στους δειγματοληπτικούς
- r_c^l: Παράμετρος μεγέθους πεδίου
- s_c^l: Παράμετρος επικάλυψης

Όπως προαναφέρθηκε, ο αριθμός των συνελικτικών χαρτών εξαρτάται από την φύση και την πολυπλοκότητα των δεδομένων. Όσον αφορά τη βέλτιστη τιμή πεδίου, εξαρτάται επίσης από το πρόβλημα. Συνήθως, πρόκειται για μικρή τιμή, προκειμένου το δίκτυο να μπορεί να αναγνωρίζει με περισσότερη λεπτομέρεια μικρά τοπικά χαρακτηριστικά.

Τέλος, η παράμετρος επικάλυψης συνήθως ισούται με το ένα. Η αύξηση της σπάνια έχει νόημα, αφού τα πεδία μεταξύ τους θα βρίσκονται πολύ μακριά και το δίκτυο μπορεί να χάσει σημαντικά χαρακτηριστικά.

3.4.3 Επίπεδο Δειγματοληψίας

Το επίπεδο δειγματοληψίας βρίσκεται, είτε στο στρώμα εισόδου στην αρχή του δικτύου, είτε ακολουθεί μετά από το συνελικτικό επίπεδο. Σκοπός του επίπεδου αυτού, είναι να μειώσει τα μεγέθη των χαρτών χαρακτηριστικών και κατά επέκταση να απλοποιήσει τη διαδικασία αναγνώρισης των χαρακτηριστικών.



Εικόνα 3.4-3: Απεικόνιση της λειτουργίας δειγματοληψίας στον χάρτη εισαγωγής μεγέθους 28 x 28 νευρώνων.

Κάθε συνελικτικός χάρτης στο στρώμα $l-1$ συνδέεται με τον αντίστοιχο χάρτη δειγματοληψίας στο στρώμα l. Τα νέα πεδία που δημιουργούνται κατά της δειγματοληψίας εξαρτώνται από δύο καθοριστικές παραμέτρους r_x^l και r_y^l, που συνήθως δεν αλληλεπικαλύπτονται. Μια τέτοια διαδικασία, όπως γίνεται κατανοητό, μειώνει το μέγεθος των χαρτών κατά r_x^l και r_y^l για κάθε διάσταση τους. Όπως φαίνεται και στο παραπάνω σχήμα 3.4.3, έτσι το νέο μέγεθος ενός χάρτη χαρακτηριστικών ύστερα από την διαδικασία της δειγματοληψίας θα είναι της μορφής:

$$m_l = \frac{m_{l-1}}{r_x^l} \; και \; n_l = \frac{n_{l-1}}{r_y^l},$$

Επομένως, με αυτόν τον ορισμό γίνεται κατανοητό πως, κάθε νευρώνας από το συνελικτικό επίπεδο συνδέεται ακριβώς με έναν από το επίπεδο δειγματοληψίας. Αυτό το γεγονός έχει ως αποτέλεσμα, να οδηγεί τον σχεδιασμό αυτού του δικτύου στην ακόλουθη εξίσωση πολυπλοκότητας χρόνου κατά τον υπολογισμό της εξόδου από το επίπεδο δειγματοληψίας:

$$O(|F_{l-1}| \cdot m_{l-1} \cdot n_{l-1}),$$

Γενικώς, οι παράμετροι στο δειγματοληπτικό επίπεδο μπορούν να επιλέγουν από ένα εύρος $1 \dots m_{l-1}$ για το r_x^l και $1 \dots n_{l-1}$ για το r_y^l. Οι συνήθεις τιμές για αυτές τις δύο παραμέτρους είναι το δύο

($r_x^l = r_y^l = 2$). Εάν τα νέα πεδία τα οποία δημιουργούνται πρέπει να είναι εκτός των ορίων των χαρτών χαρακτηριστικών, τότε οι είσοδοι τους θεωρείται ότι έχουν την τιμή μηδέν. Με άλλα λόγια, είναι σαν συμπληρώνεται ο χάρτης χαρακτηριστικών με νευρώνες, οι οποίοι έχουν μηδενικό δυναμικό.

Οι νευρώνες από τους χάρτες δειγματοληψίας λαμβάνουν τις εισροές από τα πεδία. Στη συνέχεια πολλαπλές είσοδοι $r_x^l \times r_y^l$ συνδυάζονται σε μια συγκεκριμένη τιμή που δηλώνεται ως δυναμικό των νευρώνων. Ο πιο συχνός τρόπος για να συνδυαστούν οι είσοδοι είναι, είτε με την μέθοδο εύρεσης μέσης τιμής (εξίσωση 3.4.1), είτε με την εύρεση μέγιστης τιμής (εξίσωση 3.4.2). Έτσι έχοντας έναν χάρτη χαρακτηριστικών φ στο στρώμα l, η έξοδος του χάρτη πολλαπλασιάζεται στη συνέχεια με το συντελεστή εκπαίδευσης $a^{\varphi,l}$, προσθέτοντας σε αυτό το bias $b^{\varphi,l}$, και στο τέλος επεξεργάζεται από την συνάρτηση ενεργοποίησης f (εξίσωση 3.4.3).

$$\xi_{i,j}^{\varphi,l} = \frac{1}{r_x^l r_y^l} \sum_{\Delta i=0}^{r_c^l-1} \sum_{\Delta j=0}^{r_c^l-1} y_{ir_x^l+\Delta i, jr_y^l+\Delta j} \,, \qquad \textit{(Εξίσωση 3.4-1)}$$

$$\max_{\substack{\Delta i \in (0, r_x^l-1) \\ \Delta j \in (0, r_y^l-1)}} (y_{ir_x^l+dx, jr_y^l+dy}{}^{\varphi,l-1}), \qquad \textit{(Εξίσωση 3.4-2)}$$

$$y_{i,j}^{\varphi,l} = F(a^{\varphi,l}\xi^{\varphi,l} + b^{\varphi,l}, \qquad \textit{(Εξίσωση 3.4-3)}$$

3.5 Αναδρομικά Νευρωνικά Δίκτυα

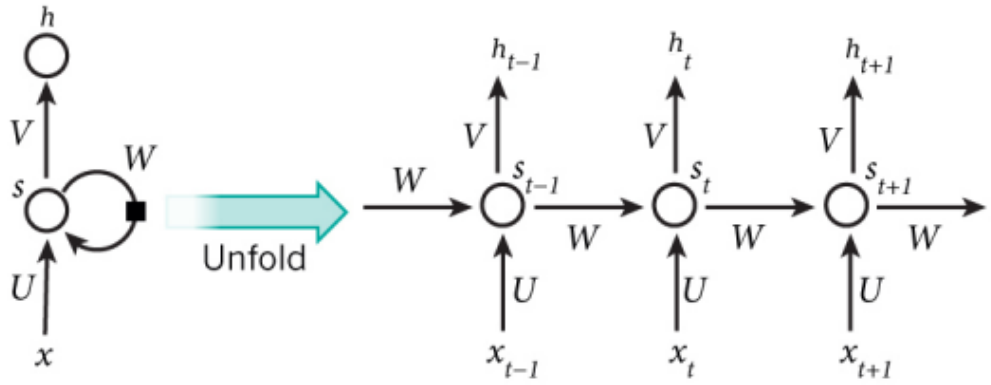
Μια σημαντική υπόθεση αναφορικά με τα προηγούμενα δίκτυα για τα οποία έγινε λόγος αποτελεί το γεγονός ότι τα δεδομένα εισόδου μεταξύ τους ήταν ανεξάρτητα. Αυτή η υπόθεση, ωστόσο δεν ισχύει για δεδομένα, τα οποία αφορούν την ομιλία, τη γλώσσα, τις χρονοσειρές, τα βίντεο, σε αυτήν την περίπτωση όλα τα δεδομένα μεταξύ τους εμφανίζουν μια εξάρτηση με την πάροδο του χρόνου. Με άλλα λόγια η κάθε ακολουθία δεδομένων εξαρτάται από την προηγούμενη και την επόμενη, με αποτέλεσμα τα κλασικά νευρωνικά δίκτυα ή ακόμα και τα συνελικτικά δεν μπορούν να ανταπεξέλθουν με μεγάλη αποτελεσματικότητα. Μια προσέγγιση που είχε γίνει από τους (Frank, Davey, & Hunt, 2001), επρόκειτο για έναν μηχανισμό, ο οποίος χώριζε σε -σταθερού αριθμού- διαδοχικά δείγματα δεδομένων και τα αντιμετώπιζε ως ένα ενιαίο σημείο δεδομένων, παρόμοιο με ένα σταθερού μεγέθους κινούμενο παράθυρο

στο σύνολο δεδομένων. Όπως απεδείχθη αργότερα η προσέγγιση αυτή επειδή εξαρτιόταν από το σωστό μέγεθος του παραθύρου, σε σύνολα δεδομένων που ήταν πολύ μεγάλα δεν ήταν καθόλου αποδοτική. Αυτό είχε ως αποτέλεσμα, οι ερευνητές να στρέψουν την προσοχή τους σε διαφορετικές αρχιτεκτονικές με στόχο να επεξεργαστούν ακόμα πιο αποτελεσματικά ακολουθιακά δεδομένα.

Τα αναδρομικά νευρωνικά δίκτυα ή αλλιώς Recurrent Neural Networks (RNN) επεξεργάζονται μια ακολουθία εισόδου με ένα στοιχείο κάθε φορά και διατηρούν ένα κρυφό διάνυσμα, το οποίο λειτουργεί ως μνήμη για την προηγούμενη πληροφορία. Μαθαίνουν επιλεκτικά σχετικές πληροφορίες από τις ακολουθίες που τους επιτρέπουν να καταλαβαίνουν τις εξαρτήσεις μεταξύ των δεδομένων στο πέρασμα του χρόνου. Με αυτόν τον τρόπο, έχοντας ως δεδομένο τις τρέχουσες πληροφορίες, τις προηγούμενες, καθώς και την συσχέτιση αυτών, είναι σε θέση τα δίκτυα αυτά να κάνουν μελλοντικές προβλέψεις και να αντιμετωπίσουν διάφορα προβλήματα που απαιτούν διαδοχική επεξεργασία δεδομένων. Συχνές εφαρμογές των επαναληπτικών νευρωνικών δικτύων είναι σε εφαρμογές επεξεργασίας γλώσσας και αναγνώρισης ομιλίας, σε αναγνώριση ανθρώπινης δραστηριότητας από αισθητήρες και βίντεο αλλά ακόμα και σε παραγωγή μουσικής.

3.5.1 Αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων

Τα αναδρομικά νευρωνικά δίκτυα αποτελούν ένα ειδικό τύπο των κλασικών νευρωνικών δικτύων απλά κατάλληλα για την επεξεργασία διαδοχικών δεδομένων. Το κύριο χαρακτηριστικό τους είναι η κρυφή μνήμη, η οποία διατηρούν για τα προηγούμενα στοιχεία της ακολουθίας. Το πιο απλό αναδρομικό νευρωνικό δίκτυο φαίνεται στο σχήμα 3.5.1.



Εικόνα 3.5-1: Αναπαράσταση ενός τυπικού RNN. Η αριστερή πλευρά του σχήματος είναι ένα τυπικό RNN. Ενώ στην δεξιά πλευρά βρίσκεται το ίδιο δίκτυο και το πως αυτό έχει "ξετυλιχθεί" κατά την πάροδο του χρόνου. (Singh, 2017)

Όπως φαίνεται από το παραπάνω σχήμα ένα τέτοιο δίκτυο έχει μια σύνδεση ανατροφοδότησης, η οποία συνδέει τους νευρώνες μεταξύ τους στον χρόνο. Στην χρονική στιγμή t (timestep) το δίκτυο λαμβάνει ως είσοδο το ακολουθιακό στοιχείο x_t και την προηγούμενη κρυφή κατάσταση από την προηγούμενη χρονική στιγμή s_{t-1}. Στη συνέχεια, η κρυφή κατάσταση ενημερώνεται σε s_t και τέλος η έξοδος h_t του δικτύου υπολογίζεται. Με αυτόν τρόπο η τρέχουσα έξοδος h_t εξαρτάται από όλες τις προηγούμενες εισόδους x'_t ($όπου\ t' \leq t$). U είναι ο πίνακας βαρών μεταξύ των ενδιάμεσων και εισόδου στρωμάτων με παρόμοιο τρόπο που υπάρχουν και στα συνελικτικά δίκτυα. W είναι τα βάρη που χρησιμοποιούνται μεταξύ των ενδιάμεσων στρωμάτων και V τα βάρη μεταξύ τον ενδιάμεσων στρωμάτων και του στρώματος εξόδου. Παρακάτω αναφέρονται οι δύο εξισώσεις, οι οποίες κάνουν τους υπολογισμούς για κάθε χρονική στιγμή:

$$s_t = \sigma(Ux_t + Ws_{t-1} + b_s$$

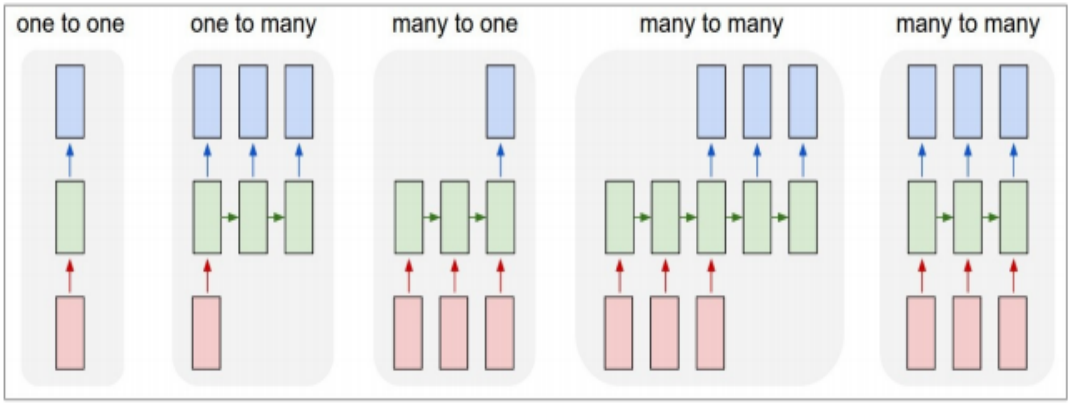
$$h_t = softmax(Vs_t + b_h)$$

Όπου SoftMax αναπαριστά την συνάρτηση ενεργοποίησης, η οποία χρησιμοποιείται συχνά στο στρώμα εξόδου για προβλήματα κατηγοριοποίησης. Κύρια λειτουργία της είναι να εξασφαλίσει πως όλες οι έξοδοι είναι μεταξύ 0 και 1 και το άθροισμα τους είναι 1. Η παρακάτω εξίσωση καθορίζει το SoftMax για ένα πρόβλημα κατηγοριοποίησης με κλάσεις K.

$$y_k = e^{a_k}/(\sum_{k'=1}^{K} 1^{e}\text{^}(a'_k))\ για\ k = 1, \dots, K$$

Ένα τυπικό Επαναληπτικό Νευρωνικό Δίκτυο (σχ. 3.5.1**)** είναι το ίδιο βαθύ με ένα κλασικό νευρωνικό δίκτυο, αναλογίζοντας τον τρόπο που λειτουργεί. Όπως φαίνεται στη δεξιά πλευρά του σχήματος το δίκτυο επεκτείνεται (unfold) με το χρόνο, έτσι μπορεί να θεωρηθεί σαν ένα βαθύ δίκτυο με τον αριθμό των στρωμάτων να είναι ίσος με τον αριθμό των χρονικών στιγμών, των οποίων υπάρχουν σε μια ακολουθία εισόδου. Επιπροσθέτως, ένα τέτοιο δίκτυο μπορεί να διαχειριστεί και ακολουθίες εισόδου μη σταθερού μήκους . Αξίζει να προστεθεί ότι χρησιμοποιούνται τα ίδια βάρη σε κάθε χρονική στιγμή. Σε κάθε χρονική στιγμή λαμβάνεται νέα είσοδος, λόγω όμως του τρόπου, του οποίου ενημερώνεται η κρυφή κατάσταση s_t , οι πληροφορίες σε ένα τέτοιο δίκτυο μπορούν να ρέουν σε αυθαίρετο αριθμό χρονικών στιγμών, επιτρέποντας στο δίκτυο να διατηρήσει μια μνήμη όλων των προηγούμενων πληροφοριών.

Ανάλογα με το εκάστοτε πρόβλημα, υπάρχουν διαφορετικές αρχιτεκτονικές για αντίστοιχες πιθανές εισόδους και αντίστοιχες απαιτούμενες εξόδους. Στο παρακάτω σχήμα φαίνονται οι διαφορετικοί τύποι αρχιτεκτονικών σε αντίθεση με ένα κλασικό νευρωνικό δίκτυο, όπου κάθε ορθογώνιο αναπαριστά ένα διάνυσμα, το οποίο περιέχει πολλά χαρακτηριστικά ή νευρώνες. Τα χρώματα δείχνουν τον τύπο του διανύσματος ή αντίστοιχα του στρώματος: με κόκκινο αναπαρίσταται το στρώμα εισόδου, με πράσινο τα ενδιάμεσα στρώματα και με μπλε τα στρώματα εξόδου.



Εικόνα 3.5-2: Αναπαράσταση διαφορετικών τύπων RNNs (Woditsch, 2017)

- ***Ένα προς ένα:*** Αναπαριστά το κλασικό νευρωνικό δίκτυο με σταθερό αριθμό εισόδου και εξόδου.

- ***Ένα προς πολλά:*** Αναπαριστά ένα επαναληπτικό νευρωνικό δίκτυο με είσοδο ένα διάνυσμα χαρακτηριστικών και παράγει μια ακολουθιακή έξοδο. Για παράδειγμα, ένα μοντέλο που παίρνει ως είσοδο μια εικόνα και παράγει λεζάντες για αυτή την εικόνα με πολλές λέξεις.
- ***Πολλά προς ένα:*** Αναπαριστά ένα επαναληπτικό νευρωνικό δίκτυο με είσοδο ακολουθιακά δεδομένα (πολλαπλά διανύσματα) και παράγει μια έξοδο, λόγου χάρη μια ακολουθία από δορυφορικές εικόνες που χρησιμοποιούνται για να κατηγοριοποιηθούν φυτά στο χρόνο.
- ***Πολλά προς πολλά:*** Αναπαριστά ένα επαναληπτικό νευρωνικό δίκτυο με είσοδο ακολουθιακά δεδομένα και υπολογίζει ακολουθιακές εξόδους, παραδείγματος χάρη μια ακολουθία από λέξεις σε μια γλώσσα που πρέπει να μεταφρασθεί σε μια άλλη.
- ***Πολλά προς πολλά (συγχρονισμένα):*** Αναπαριστά ένα επαναληπτικό νευρωνικό δίκτυο με είσοδο ακολουθιακά δεδομένα και υπολογίζει συγχρονισμένες ακολουθιακές εξόδους, η κατηγοριοποίηση βίντεο σε αντιστοιχίες συγκεκριμένων frame, αποτελεί ένα χαρακτηριστικό παράδειγμα.

3.5.2 Εκπαίδευση Αναδρομικών Νευρωνικών Δικτύων

Η εκπαίδευση των επαναληπτικών νευρωνικών δικτύων επιτυγχάνεται επεκτείνοντας το δίκτυο σε βάθος χρόνου και δημιουργώντας ένα αντίγραφο του μοντέλου σε κάθε χρονική στιγμή. Όπως γίνεται αντιληπτό από το σχήμα 3.5.1, ένα τέτοιο επεκτεινόμενο δίκτυο αντιμετωπίζεται σαν ένα πολυστρωματικό νευρωνικό δίκτυο και κατά επέκταση η διαδικασία της εκπαίδευσης μοιάζει πολύ με την διαδικασία της μετάδοσης προς τα πίσω, η μόνη διαφορά αποτελεί το γεγονός ότι πραγματοποιείται σε βάθος του χρόνου. Επομένως η διαδικασία αυτή, σύμφωνα με τον (Werbos, 1990) ονομάζεται πλέον << μετάδοση προς τα πίσω σε βάθος χρόνου» ή για λόγους συντομίας BPTT (Back-Propagation Through Time).

Στην ιδανική περίπτωση, τα αναδρομικά νευρωνικά δίκτυα για να μάθουν εξαρτήσεις μεγάλων ακολουθιών μπορούν να χρησιμοποιήσουν τον αλγόριθμο BBPTT. Ο εν λόγω αλγόριθμος θα πρέπει να είναι σε θέση να μάθει και να παραμετροποιεί τα βάρη σωστά, ώστε να τοποθετούν την σωστή πληροφορία στην μνήμη. Στην πράξη, όμως, η διαδικασία αυτή είναι δύσκολη. Στην πραγματικότητα, τα τυπικά RNN αποδίδουν άσχημα, ακόμα και όταν οι έξοδοι και οι σχετικές είσοδοι χωρίζονται σε μικρές χρονικές στιγμές. Έτσι πλέον είναι ευρέως γνωστό ότι τα τυπικά RNN δε μπορούν να εκπαιδευτούν αποδοτικά για να μάθουν εξαρτήσεις με μεγάλα χρονικά διαστήματα (Bengio, Simard, & Frasconi, 1994) (Hochreiter, Bengio, Frasconi, & Schmidhuber, 2001). Η εκπαίδευση ενός RNN με BPTT απαιτεί τη μετάδοση προς τα πίσω του σφάλματος σε βάθος των χρονικών στιγμών. Θεωρώντας το κλασικό RNN, το

επαναλαμβανόμενο άκρο όπως αναφέρθηκε και προηγουμένως έχει το ίδιο βάρος. Έτσι, το σφάλμα δημιουργείται με τον πολλαπλασιασμό της κλίσης σφάλματος με την ίδια τιμή ξανά και ξανά κάθε φορά. Αυτό έχει ως αποτέλεσμα, οι κλίσεις να γίνονται πολύ μεγάλες ή να πλησιάζουν το μηδέν. Αυτές οι δύο καταστάσεις αναφέρονται στη βιβλιογραφία ως 'έκρηξη κλίσης (exploding gradient)'και 'εξαφάνιση κλίσης (vanishing gradient)' αντίστοιχα. Σε αυτές τις περιπτώσεις, το μοντέλο εκμάθησης δεν συγκλίνει καθόλου στα επιθυμητά αποτελέσματα ή ο χρόνος εκτέλεσης να είναι υπερβολικός. Οι καταστάσεις αυτές εξαρτώνται κυρίως από το μέγεθος του βάρους του επαναλαμβανόμενου άκρου, αλλά και από την συνάρτηση ενεργοποίησης που χρησιμοποιείται. Εάν το μέγεθος του βάρους είναι μικρότερο από το 1 και χρησιμοποιείται σιγμοειδής συνάρτηση ενεργοποίησης, είναι πιθανότερο να εμφανιστεί η περίπτωση της εξαφάνισης κλίσης. Αντίθετα, αν το βάρος είναι μεγαλύτερο του 1 και χρησιμοποιείται ReLU συνάρτηση ενεργοποίησης, τότε είναι πιο πιθανό να υπάρξει η περίπτωση της έκρηξης κλίσης. (Pascanu, Mikolov, & Bengio, 2013)

Έχουν προταθεί αρκετές προσεγγίσεις με την πάροδο του χρόνου για την αντιμετώπιση του προβλήματος της εκπαίδευσης σε μεγάλες ακολουθιακές εξαρτήσεις των επαναληπτικών νευρωνικών δικτύων. Οι εν λόγω προσεγγίσεις περιλαμβάνουν τροποποιήσεις στη διαδικασία της εκπαίδευσης, καθώς και νέες αρχιτεκτονικές. Στην προσέγγιση (Pascanu, Mikolov, & Bengio, 2013) προτάθηκε ένας τρόπος να μειώνεται η κλίση αν περνάει κάποιο προκαθορισμένο κατώφλι. Αυτή η στρατηγική ονομάστηκε "αποκοπή κλίσης (gradient clipping)'' και αποδείχτηκε αρκετά αποτελεσματικό σε προβλήματα, τα οποία αφορούσαν εκρήξεις κλίσης. Όσον αφορά την αντιμετώπιση εξαφάνισης κλίσης οι (Pascanu, Mikolov, & Bengio, 2013) εισάγουν έναν όρο ποινών παρόμοιο με τις ποινές κανονικοποίησης $L1, L2$ οι οποίες χρησιμοποιούνται για την αποφυγή υπερφόρτωσης στα κλασικά νευρωνικά δίκτυα. Ωστόσο η χρήση ενός τέτοιου περιορισμού μπορεί να κάνει πιο πιθανή την εμφάνιση της έκρηξης κλίσης. Η αρχιτεκτονική των δικτύων βραχυπρόθεσμης μακράς μνήμης (LSTM) που θα εξετασθούν στην ενότητα 3.6 εισήχθη από τους (Hochreiter & Schmidhuber, Long Short-Term Memory, 1997), είχε ως σκοπό να αντιμετωπίσει το πρόβλημα της εξαφάνισης της κλίσης. Τα δίκτυα LSTM έχουν αποδειχθεί πολύ πιο αποτελεσματικά στην εκμάθηση μεγάλων ακολουθιακών εξαρτήσεων σε σύγκριση με τα τυπικά RNN και αποτελούν την πιο δημοφιλή παραλλαγή τους.

3.6 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (long short-term memory networks η LSTM) αποτελούν ένα είδος επαναληπτικών δικτύων, τα οποία δημιουργήθηκαν προκειμένου να αντιμετωπίσουν τα προβλήματα, τα οποία αφορούν τις κλίσεις που τείνουν στο μηδέν (vanishing gradient), και ως ένα βαθμό το πετυχαίνουν. Πιο συγκεκριμένα, αντικαθιστά έναν συνηθισμένο νευρώνα με μια πιο σύνθετη αρχιτεκτονική, η οποία ονομάζεται LSTM μονάδα η μπλοκ. Μια τέτοια μονάδα αποτελείται από απλούς κόμβους, η οποία είναι συνδεδεμένη με συγκεκριμένο τρόπο. Τα κύρια χαρακτηριστικά τα οποία εισάχθηκαν από τους (Hochreiter & Schmidhuber, Long Short-Term Memory, 1997) είναι:

- ***Carousel Σταθερού Σφάλματος (Constant Error Carousel η CEC):*** Κεντρική μονάδα που περιέχει μια επαναλαμβανόμενη σύνδεση με μοναδιαίο βάρος. Η επαναλαμβανόμενη σύνδεση αντιπροσωπεύει έναν βρόχο ανατροφοδότησης με χρονική στιγμή ίση με ένα. Η ενεργοποίηση του σφάλματος είναι εσωτερική διεργασία, η οποία δρα ως μνήμη για τις προηγούμενες πληροφορίες.
- ***Πύλη Εισόδου (Input Gate):*** Πολλαπλασιαστική μονάδα η οποία προστατεύει τις αποθηκευμένες πληροφορίες του CEC από πιθανές άσχετες εισόδους.
- ***Πύλη Εξόδου (Output Gate):*** Πολλαπλασιαστική μονάδα η οποία προστατεύει τις άλλες μονάδες από πιθανές ‘παρεμβολές’ από πληροφορίες που είναι αποθηκευμένες στο CEC.

Όπως γίνεται κατανοητό η πύλη εισόδου και εξόδου ελέγχει την πρόσβαση στο CEC. Κατά τη διάρκεια της εκπαίδευσης του μοντέλου, η πύλη εισόδου μαθαίνει πότε να αφήσει νέες πληροφορίες στο CEC. Όσο η πύλη εισόδου έχει τιμή μηδέν, δεν επιτρέπονται πληροφορίες μέσα. Ομοίως, με την πύλη εξόδου, μαθαίνει και αυτή κατά τη διάρκεια της εκπαίδευσης πότε να επιτρέπει πληροφορίες στο CEC. Όταν και οι δύο πύλες είναι κλειστές -τιμές κοντά στο μηδέν- η πληροφορία «παγιδεύεται» μέσα στη μνήμη της μονάδας. Αυτό φέρει ως αποτέλεσμα να επιτρέπεται στα σήματα σφάλματος να μεταφέρονται (με την βοήθεια του επαναλαμβανομένου άκρου με βάρος) σε όλο το δίκτυο χωρίς να αντιμετωπίζουν προβλήματα κλίσης εξαφάνισης.

Τα κλασικά LSTM έχουν καλύτερη απόδοση από τα RNN στην εκπαίδευση μεγάλου εύρους εξαρτήσεων σε ακολουθιακά δεδομένα. Ωστόσο, από τους (Gers & Cummins, 2000) εντοπίστηκε ένα μειονέκτημα. Ειδικότερα, σε μεγάλες συνεχείς ακολουθίες εισόδου, χωρίς να έχει οριστεί ρητά σημείο έναρξης και σημείο λήξης, το δίκτυο LSTM θα αυξηθεί απεριόριστα με αποτέλεσμα στο τέλος να γίνει ασταθές. Με αυτόν τον τρόπο, το μοντέλο δε θα είναι σε θέση να εκπαιδευτεί σωστά. Η κατάσταση αυτή δε θα επαναφερθεί, εκτός και αν οριστούν σημεία έναρξης και λήξης. Σε ιδανικά πλαίσια, θα άρμοζε το

LSTM να μάθει να επαναφέρει τα περιεχόμενα της μονάδας μνήμης, αφού ολοκληρωθεί η επεξεργασία μιας ακολουθίας και πριν ξεκινήσει να επεξεργάζεται μια νέα. Προκειμένου να επιλυθεί αυτό το ζήτημα, παρουσιάστηκε μια νέα αρχιτεκτονική με τις λεγόμενες ‘’ξεχασμένες’’ πύλες (forget gates) (Gers & Cummins, 2000). Οι ξεχασμένες πύλες ουσιαστικά μαθαίνουν να επαναφέρουν τη μνήμη μόλις τελειώσει μια ακολουθία και πριν ξεκινήσει μια νέα, στην παρακάτω ενότητα θα αναλυθεί περισσότερο η αρχιτεκτονική αυτή.

3.6.1 Μονάδα Μακράς Βραχυπρόθεσμης μνήμης με Forget Πύλες

Η αρχιτεκτονική μια μονάδας LSTM με ξεχασμένες πύλες φαίνεται στο παρακάτω σχήμα. Τα κύρια χαρακτηριστικά αυτής είναι:

- ***Είσοδος:*** Η μονάδα LSTM παίρνει ως είσοδο το τρέχον διάνυσμα x_t και την έξοδο από την προηγούμενη χρονική στιγμή (μέσω των επαναλαμβανόμενων άκρων) που υποδηλώνεται με h_{t-1}. Στη συνέχεια οι δύο είσοδοι αθροίζονται και εφαρμόζεται $tanh$ συνάρτηση ενεργοποίησης, επιστρέφοντας στο τέλος ένα νέο διάνυσμα z_t.

$$z_t = \tanh\left(W^2 x_t + R^2 h_{t-1} + b^2\right)$$

- ***Πύλη Εισόδου:*** Η πύλη εισόδου παίρνει ως είσοδο τα x_t και h_{t-1}, υπολογίζει το άθροισμα τους και εφαρμόζει τη σιγμοειδή συνάρτηση ενεργοποίησης. Το αποτέλεσμα i_t πολλαπλασιάζεται με το z_t.

$$i_t = \sigma(W^i x_t + R^i h_{t-1} + b^i)$$

- ***Πύλη Forget:*** Η πύλη forget αποτελεί ένα μηχανισμό, μέσω του οποίου τα LSTM μαθαίνει να επαναφέρει τα περιεχόμενα της μνήμης όταν πλέον δεν είναι σχετικά. Αυτό μπορεί να συμβεί όταν το δίκτυο επεξεργάζεται μια νέα ακολουθία. Παίρνει σαν είσοδο τα x_t και h_{t-1} και εφαρμόζει τη σιγμοειδή συνάρτηση ενεργοποίησης. Το αποτέλεσμα f_t πολλαπλασιάζεται με την κατάσταση της προηγούμενης μονάδας στην προηγούμενη χρονική στιγμή s_{t-1} όπου επιτρέπει να ‘’ξεχαστούν’’ τα δεδομένα που δε χρειάζονται.

$$f_t = \sigma(W^f x_t + R^f h_{t-1} + b^f)$$

- ***Μονάδα Μνήμης:*** Η μονάδα μνήμης αποτελείται από το CEC, η οποία περιέχει το επαναλαμβανόμενο άκρο και το βάρος του. Η τρέχουσα κατάσταση της μονάδας αυτής s_t υπολογίζεται ‘’ξεχνώντας’’ άσχετες πληροφορίες (εάν υπάρχουν) από την προηγούμενη χρονική στιγμή και ‘’αποδέχοντας’’ σχετικές (εάν υπάρχουν) από την τρέχουσα είσοδο.
- ***Πύλη Εξόδου:*** Η πύλη εξόδου λαμβάνει ως είσοδο το άθροισμα x_t και h_{t-1} και εφαρμόζει την σιγμοειδή συνάρτηση ενεργοποίησης, με στόχο να ελέγξει τι πληροφορία θα εξαχθεί έξω από την LSTM μονάδα.

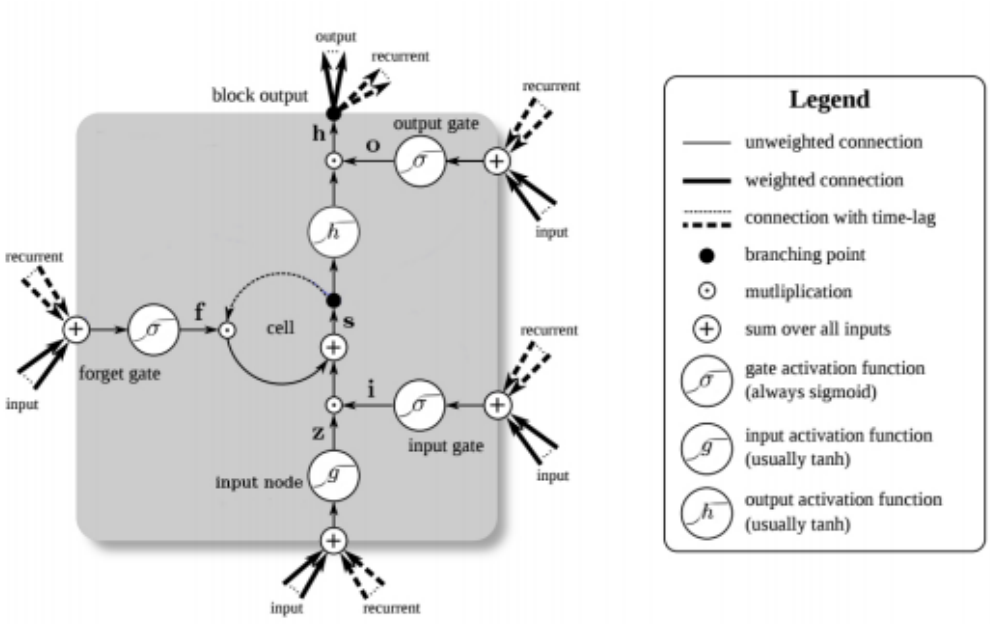
$$O_t = \sigma(W \odot x_t + R \odot h_{t-1} + b^o)$$

- ***Έξοδος:*** Η έξοδος της μονάδας LSTM h_t υπολογίζεται πολλαπλασιάζοντας την κατάσταση της μονάδας s_t με την συνάρτηση $tanh$, το γινόμενο αυτό στην συνέχεια πολλαπλασιάζεται με την πύλη εξόδου o_t.

$$s_t = z_t \circ i_t + s_{t-1} \circ f_t \quad (\kappa\alpha\tau\acute{\alpha}\sigma\tau\alpha\sigma\eta\ \mu o\nu\acute{\alpha}\delta\alpha\varsigma)$$

$$h_t = \tanh(s_t) \circ o_t \quad (\acute{\epsilon}\xi o\delta o\varsigma)$$

Στις παραπάνω εξισώσεις W^* είναι τα βάρη εισόδου ,R^* είναι τα επαναλαμβανόμενα βάρη και b^* είναι η σταθερά bias.



Εικόνα 3.6-1: Μια μονάδα LSTM με πύλες Forget. (Singh, 2017)

Με αυτόν τον τρόπο, γίνεται κατανοητό ότι οι ανθρώπινες δραστηριότητες δομούνται από ενέργειες οι οποίες δύναται να απέχουν από αυθαίρετες αποστάσεις, να εμφανίζονται περιοδικά ή με εναλλαγές στη σειρά εμφάνισης τους. Ο τρόπος λειτουργίας των LSTM ως προς τις χρονικές εξαρτήσεις έχουν οδηγήσει στην εφαρμογή αυτών σε εφαρμογές αναγνώρισης δραστηριότητας, ειδικά όταν η είσοδος των δεδομένων είναι σε μορφή βίντεο. Συνδυάζοντας σε αυτή την περίπτωση τα συνελικτικά δίκτυα με στόχο την εξαγωγή χαρακτηριστικών από τα βίντεο αλλά και τα LSTMs με την σειρά τους στη συνέχεια αναλαμβάνουν να εντοπίσουν τους χρονικούς συσχετισμούς (Μενύχτας, 2019).

3.6.2 Αναδρομική Μονάδα με Πύλες

Μια γενικευμένη μορφή του LSTM δικτύου αποτελούν οι λεγόμενες Αναδρομικές μονάδες με πύλες ή αλλιώς GRUs (Gated Recurrent Units), παρόλο που ένα LSTM δίκτυο ασχολείται με τα προβλήματα, τα οποία αφορούν τις κλίσεις που τείνουν στο μηδέν. Οι Αναδρομικές μονάδες με πύλες παρουσιάστηκαν το 2014 από τους (Chung, Gulcehre, Cho, & Bengio, 2014). Αναφορικά με την αρχιτεκτονική των Αναδρομικών μονάδων με πύλες είναι παρόμοια με αυτήν των LSTM, με μόνη διαφορά την ύπαρξη πυλών, οι οποίες ρυθμίζουν τη ροή των πληροφοριών εντός της μονάδας, χωρίς να έχουν ξεχωριστή μονάδα μνήμης. Οι μονάδες αυτές διαθέτουν δύο πύλες, οι οποίες ονομάζονται πύλες ενημέρωσης και επαναφοράς με σκοπό τον έλεγχο των πληροφοριών που εισέρχονται στη μονάδα. Παρακάτω δίνεται το αντίστοιχο σχεδιάγραμμα με την αρχιτεκτονική μιας GRU μονάδας, καθώς και οι αντίστοιχες εξισώσεις υπολογισμού (Sarika, 2018):

- ***Πύλη Ενημέρωσης:*** Αυτή η πύλη ελέγχει το πόση πληροφορία από την προηγούμενη κρυφή κατάσταση h θα μεταδοθεί στην τωρινή.

$$z_t = \sigma(W_z * [h_{t-1}, x_t])$$

- ***Πύλη Επαναφοράς:*** Αυτή η πύλη ελέγχει κατά πόσο είναι σχετική η πληροφορία της προηγούμενης κρυφής κατάστασης. Αν δεν είναι την αντικαθιστά με την τωρινή πληροφορία.

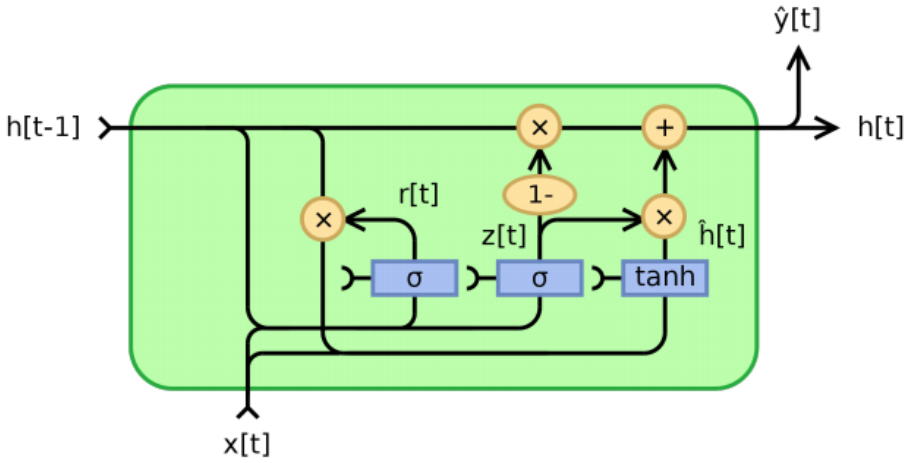
$$r_t = \sigma(W_r * [h_{t-1}, x_t])$$

- ***Νέα Μνήμη:***

$$\widetilde{h_t} = tanh(W * [r_t \odot h_{t-1}, x_t])$$

- ***Τελική Μνήμη:***

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h_t})$$



Εικόνα 3.6-2: Κλασική αρχιτεκτονική μιας GRU (Κερατζάκης, 2019)

Όπου W αναπαρίστανται τα βάρη σε κάθε περίπτωση, σ είναι η σιγμοειδής συνάρτηση και ο σύμβολο $\odot$ δηλώνει την πράξη του γινομένου Hadamard.

3.7 Βαθιά Μάθηση στην Αναγνώριση Ανθρώπινης Δραστηριότητας

Στην πλειοψηφία των προηγούμενων προσεγγίσεων έχουν εφαρμοστεί μέθοδοι μηχανικής μάθησης στην αναγνώριση ανθρώπινης δραστηριότητας (Oscar & Labrador, 2013). Οι περισσότερες από αυτές βασίζονται σε τεχνικές εξαγωγής χαρακτηριστικών, συμπεριλαμβάνοντας τον μετασχηματισμό συχνότητας - χρόνου. Παρόλο που τα χαρακτηριστικά εξάγονται με προσοχή και σωστό σχεδιασμό, δεν υπάρχει κάποια καθολική προσέγγιση για την αποτελεσματική εξαγωγή διακριτών χαρακτηριστικών για ανθρώπινες δραστηριότητες.

Τα τελευταία χρόνια, η εφαρμογή της βαθιάς μάθησης παρουσιάζει ραγδαία αύξηση σε εξαγωγές περίπλοκων δεδομένων σε πολυάριθμους τομείς, όπως είναι η όραση υπολογιστών, η επεξεργασία φυσικής γλώσσας και ομιλίας. Μετά από τις πρώτες προσεγγίσεις των (Hammerla, et al., 2015) (Lane & Georgiev, 2015) οι οποίοι εξέτασαν την αποτελεσματικότητα της βαθιάς μάθησης στην αναγνώριση ανθρώπινης δραστηριότητας, σχετικές μελέτες ξεκίνησαν στον τομέα αυτό. Έτσι σε

συνδυασμό με την αναπόφευκτη ανάπτυξη της βαθιάς μάθησης στην αναγνώριση της ανθρώπινης δραστηριότητας, οι προκλήσεις αυξάνονταν με αποτέλεσμα οι σύγχρονες προσεγγίσεις να προσπαθούν να τις αντιμετωπίσουν. Ωστόσο, η βαθιά μάθηση παρόλο των πολλών προκλήσεων που έχει, εξακολουθεί να απολαμβάνει τεράστιας αποδοχής από τους ερευνητές, λόγω της έντονης καινοτομίας που τη διέπει. Επομένως, είναι απαραίτητο να αποδειχθούν όλοι οι λόγοι πίσω από τη μεγάλη αυτή επιτυχία της βαθιάς μάθησης στην αναγνώριση ανθρώπινης δραστηριότητας.

- Το πιο ελκυστικό χαρακτηριστικό της βαθιάς μάθησης αποτελεί η λέξη ‘’βαθιά’’. Τα μοντέλα επιτρέπουν την εκμάθηση χαρακτηριστικών με την δυνατότητα κλιμάκωσης. Επίσης, χάρη στην προηγμένη ανάπτυξη των πόρων της πληροφορικής -όπως οι κάρτες γραφικών-, παρέχουν σε βαθιά μοντέλα μια ισχυρή ικανότητα εκμάθησης χαρακτηριστικών από περίπλοκα δεδομένα. Έτσι, η εξαιρετική αυτή μαθησιακή ικανότητα επιτρέπει στο σύστημα αναγνώρισης δραστηριότητας να αναλύσει σε βάθος πολυτροπικά αισθητήρια δεδομένα για την ακριβέστερη αναγνώριση.
- Οι διαφορετικές αρχιτεκτονικές και δομές νευρωνικών δικτύων πλέον είναι σε θέση να κωδικοποιούν χαρακτηριστικά από πολλές απόψεις. Χαρακτηριστικό παράδειγμα αποτελούν τα συνελικτικά δίκτυα, τα οποία είναι ικανά να αναγνωρίζουν πολλά διαφορετικά και πολυτροπικά χαρακτηριστικά, μέσω των αισθητήρων. Αντίστοιχα, τα RNN δύναται να εξάγουν χρονικές εξαρτήσεις των δραστηριοτήτων ενός ατόμου και σταδιακά να εκπαιδεύονται από αυτές. Επίσης, μπορεί να χρησιμοποιηθεί συνδυασμός των δίκτυών αυτών.
- Τέλος, τα βαθιά νευρωνικά δίκτυα είναι αποσπώμενα και μπορούν να συνδυαστούν με ευελιξία σε ενοποιημένα δίκτυα με στόχο την συνολική βελτιστοποίηση. Με αυτόν τον τρόπο μπορούν να χρησιμοποιηθούν σε διάφορες τεχνικές βαθιάς μάθησης, όπως είναι η βαθιά μάθηση μεταφοράς (deep transfer learning), η βαθιά ενεργή μάθηση (deep active learning), ο μηχανισμός βαθιάς προσοχής (deep attention mechanism), καθώς και σε άλλες αποτελεσματικές λύσεις (Kaixuan, et al., 2018).

Κεφάλαιο 4 – Πειραματική Αξιολόγηση

4.1 Εισαγωγή

Στα πλαίσια της παρούσας διπλωματικής υλοποιήθηκαν δύο προσεγγίσεις βαθιάς μάθησης στην αναγνώριση ανθρώπινης δραστηριότητας (HAR). Η πρώτη προσέγγιση αφορά την αναγνώριση ανθρώπινης δραστηριότητας από αισθητήρες smartphone, ενώ η δεύτερη προσέγγιση αφορά το σύστημα αναγνώρισης ανθρώπινης δραστηριότητας που έχει εκπαιδευτεί πάνω σε RGB-D εικόνες.

Στην πρώτη προσέγγιση έχουν χρησιμοποιηθεί τρία διαφορετικά είδη νευρωνικών δικτύων. Το πρώτο είδος είναι ένα τυπικό αναδρομικό δίκτυο μακράς βραχυπρόθεσμης μνήμης (LSTM), το δεύτερο είδος αποτελεί ένα τυπικό αναδρομικό δίκτυο (RNN) και το τρίτο είδος είναι μια αναδρομική μονάδα με πύλες (GRU). Όλα τα δίκτυα χρησιμοποιούνται για συγκρίσεις απόδοσης μεταξύ τους. Επιπλέον, όλα τα δίκτυα έχουν γίνει σε Python με τη βοήθεια της βιβλιοθήκης ανοιχτού κώδικα TensorFlow, η οποία χρησιμοποιείται ευρέως σε εφαρμογές βαθιάς μάθησης. Το σύνολο δεδομένων το οποίο επιλέχτηκε για το πρόβλημα είναι το ‘Human Activity Recognition Using Smartphones Data Set’, το οποίο είναι διαθέσιμο από το UCL Machine Learning Repository. Από αυτό επιλέχθηκε ένα υποσύνολο των δεδομένων, το οποίο υποβλήθηκε σε προ-επεξεργασία για την εκπαίδευση των δικτύων και ένα άλλο υποσύνολο για τη δοκιμή. Τελικώς, η αξιολόγηση των δικτύων έγινε πάνω σε έξι κατηγορίες κλάσεων, οι οποίες αφορούσαν ανθρώπινες δραστηριότητες.

Στη δεύτερη προσέγγιση έχουν χρησιμοποιηθεί τρείς συνδυασμοί νευρωνικών δικτύων με χρήση δύο διαφορετικών τεχνικών αντίστοιχα σε κάθε περίπτωση. Τα δίκτυα τα οποία χρησιμοποιούνται παρουσιάζουν μια ιδιομορφία σε σχέση με τις προηγούμενες προσεγγίσεις, καθώς το σύνολο δεδομένων το απαιτεί. Πιο συγκεκριμένα, ο πρώτος συνδυασμός νευρωνικών δικτύων είναι ένα 2D συνελικτικό δίκτυο (Conv2D) μαζί με ένα τυπικό δίκτυο βραχυπρόθεσμης μνήμης (LSTM), ο δεύτερος συνδυασμός πρόκειται πάλι για ένα 2D συνελικτικό δίκτυο με ένα τυπικό αναδρομικό δίκτυο (RNN) και ο τρίτος συνδυασμός αφορά ένα συνελικτικό δίκτυο με μια αναδρομική μονάδα με πύλες (GRU). Αντίστοιχα οι τεχνικές, οι οποίες χρησιμοποιούνται αφορούν την επεξεργασία του συνόλου δεδομένων πριν (Early Fusion) την εκπαίδευση ή το συνδυασμό των αποτελεσμάτων μετά (Late Fusion). Ομοίως, σε αυτή την προσέγγιση όλα τα δίκτυα χρησιμοποιούνται για συγκρίσεις μεταξύ τους και έχουν γίνει σε Python με τη βοήθεια της

βιβλιοθήκης TensorFlow. Όσον αφορά το σύνολο δεδομένων στο οποίο εκπαιδεύονται τα δίκτυα είναι το ‘CAD-120’, το οποίο είναι ελεύθερα διαθέσιμο από ερευνητές του πανεπιστημίου Cornell. Στο εν λόγω σύνολο δεδομένων, επιλέχθηκε ένα υποσύνολο των δεδομένων, το οποίο υποβλήθηκε σε προ-επεξεργασία για την εκπαίδευση των δικτύων και ένα άλλο υποσύνολο για τη δοκιμή. Καταληκτικά, η αξιολόγηση των δικτύων έγινε πάνω σε δέκα κατηγορίες κλάσεων, οι οποίες αφορούσαν ανθρώπινες δραστηριότητες.

4.2 Προσέγγιση σε Δεδομένα από Smartphone Αισθητήρες

Τα smartphone πλέον αποτελούν τα πιο χρήσιμα εργαλεία της σύγχρονης καθημερινότητας, καθώς με την ραγδαία εξέλιξη της τεχνολογίας αποκτούν ολοένα και περισσότερες ικανότητες αναφορικά με την εξυπηρέτηση των αναγκών των χρηστών. Προκειμένου τα smartphone να γίνουν πιο λειτουργικά και ισχυρά, οι σχεδιαστές και οι προγραμματιστές προσθέτουν συνεχώς νέα δομικά στοιχεία στο υλικό κατασκευής τους. Επομένως, γίνεται κατανοητό πως αναφορικά με τη λειτουργικότητα τους, καθώς και την καλύτερη κατανόηση του περιβάλλοντος γύρω τους παίζουν πολύ σημαντικό ρόλο οι αισθητήρες. Τα περισσότερα smartphone διαθέτουν διαφορετικούς αισθητήρες και αυτό καθιστά δυνατή την συλλογή μεγάλου μεγέθους πληροφοριών σχετικά με την καθημερινή ζωή και τις δραστηριότητες του χρήστη.

Ένας από αυτούς τους αισθητήρες είναι το επιταχυνσιόμετρο. Πιο συγκεκριμένα περιλαμβάνεται σχεδόν σε όλους τους κατασκευαστές smartphone. Όπως υποδηλώνει και το όνομα του, το επιταχυνσιόμετρο μετρά την αλλαγή στην ταχύτητα, αξίζει να τονιστεί ότι δε μετρά την ταχύτητα καθαυτή. Τα δεδομένα τα οποία καταγράφονται από ένα επιταχυνσιόμετρο μπορούν να υποστούν επεξεργασία, προκειμένου να ανιχνευθούν ξαφνικές αλλαγές στην κίνηση. Αντίστοιχα, ένας άλλος αισθητήρας, ο οποίος χρησιμοποιείται κατά κύριο λόγο, είναι το γυροσκόπιο, το οποίο μετρά τον προσανατολισμό και τη γωνία θέσης, βάσει τη βαρύτητα. Τα σήματα τα οποία καταγράφονται από το γυροσκόπιο δύνανται να υποστούν επεξεργασία και να ανιχνεύσουν τη θέση και την ευθυγράμμιση της συσκευής. Επομένως, γίνεται αντιληπτό ότι λόγω αυτής της μεγάλης διαφοράς στα δεδομένα που προσφέρουν οι δύο προαναφερθέντες αισθητήρες, μπορούν να εξαχθούν χαρακτηριστικά, τα οποία μπορούν να οδηγήσουν στην αναγνώριση δραστηριότητας ενός ανθρώπου (Bulbul, Cetin, & Dogru, 2018).

4.2.1 Σχετική Δουλειά

Ένα από τα εμπόδια τα οποία επικαλείται να αντιμετωπίσει η αναγνώριση ανθρώπινης δραστηριότητας αποτελεί η επεξεργασία των σωστών πληροφοριών, λόγω του μεγάλου όγκου που υπάρχει διαθέσιμος. Η εποπτευόμενη μάθηση χρησιμοποιεί πληροφορίες και δεδομένα, τα οποία μπορούν να εγείρουν ανησυχίες σχετικά με το απόρρητο λόγω της καθημερινής ανάγκης παρακολούθησης των ατόμων και των δραστηριοτήτων που πραγματοποιούν (Hossain, Khan,, & Roy, 2016). Ωστόσο για να επιλυθούν τέτοια ζητήματα, οι ερευνητές βρήκαν έναν αισθητήρα, ο οποίος μπορεί να δουλεύει σαν λειτουργία στο παρασκήνιο ενός κινητού σε χαμηλό κόστος με υψηλή ακρίβεια (Martín, Bernardos, Iglesias, & Casar, 2013). Επιπροσθέτως, σήμερα οι ερευνητές αναζητούν νέες μεθόδους και τεχνικές, οι οποίες μπορούν να βοηθήσουν στη βελτίωση της ακρίβειας των αισθητήρων των smartphone. Αναλυτικότερα, ανακάλυψαν μια νέα μέθοδο, η οποία χρησιμοποιεί ένα ζεύγος επιταχυνσιόμετρου ενός smartphone μαζί με ένα ειδικό αισθητήρα στήθους, για να αναγνωρίσει την ανθρώπινη δραστηριότητα (Guiry, Ven, Nelson, Warmerdam, & Riper, 2014). Στη δημοσίευση τους οι (Lu, et al., 2017) παρουσίασαν μια νέα τεχνική αναγνώρισης ανθρώπινης δραστηριότητας, η οποία βασίζεται στη θεωρία συμπιεσμένης αίσθησης με αποτέλεσμα να πετυχαίνουν ποσοστά ακρίβειας που φτάνουν το 86%. Προκειμένου να βελτιωθεί ακόμα πιο πολύ το πεδίο της αναγνώρισης ανθρώπινης δραστηριότητας, οι ερευνητές μελέτησαν πολλές περιπτώσεις και σενάρια, των οποίων η ακρίβεια των αποτελεσμάτων θεωρήθηκε αμφιλεγόμενη. Για αυτό το λόγο, η ποιότητα των μετρήσεων και η συνεχή εξέλιξη στη χρήση μετρικών και εργαλείων για την καλύτερη αποτελεσματικότητα υπήρξε συχνά διαλεκτικό ζήτημα μεταξύ τους. Αυτό είχε ως αποτέλεσμα να δημιουργηθεί μια νέα μέθοδος, η οποία χρησιμοποιείται από τις παραμέτρους βελτιστοποίησης του μοντέλου με στόχο τη βελτίωση της ποιότητας της αναγνώρισης. Ωστόσο, η αναγνώριση ανθρώπινης δραστηριότητας δεν αφορά αποκλειστικά το περιβάλλον της τεχνολογίας πληροφοριών. Χρησιμοποιείται, επίσης, από τους ερευνητές της ιατρικής, για να βελτιώσει τις υπηρεσίες υγείας (Woznowski, Kaleshi, Oikonomou, & Craddock, 2016). Επιπλέον, οι (Ronao & Cho, 2016) υποστήριξαν πως το βαθύ δίκτυο μάθησης μπορεί να βοηθήσει στην αναγνώριση της ανθρώπινης δραστηριότητας σε smartphone. Αυτός αποτέλεσε ο κύριος λόγος, ο οποίος χρησιμοποιήσαν αυτή την τεχνική για να εξάγουν τα περίπλοκα χαρακτηριστικά από κάθε δραστηριότητα με απώτερο σκοπό να αυξήσουν την τελική αποτελεσματικότητα του μοντέλου. Η έρευνα των (Reiss, Hendeby, & Stricker, 2015) αποτελεί μια καλή βιβλιογραφική αναφορά στην εξήγηση σε μια από τις μεγαλύτερες προκλήσεις στην αναγνώριση της ανθρώπινης δραστηριότητας που σχετίζονται με σύνθετους ταξινομητές κατηγοριοποίησης. Από την άλλη μεριά οι (Martín, Bernardos, Iglesias, & Casar, 2013) επικεντρώθηκαν στην αναγνώριση ανθρώπινης δραστηριότητας μέσω smartphone, χωρίς να επηρεάζουν τον τρόπο ζωής του χρήστη.

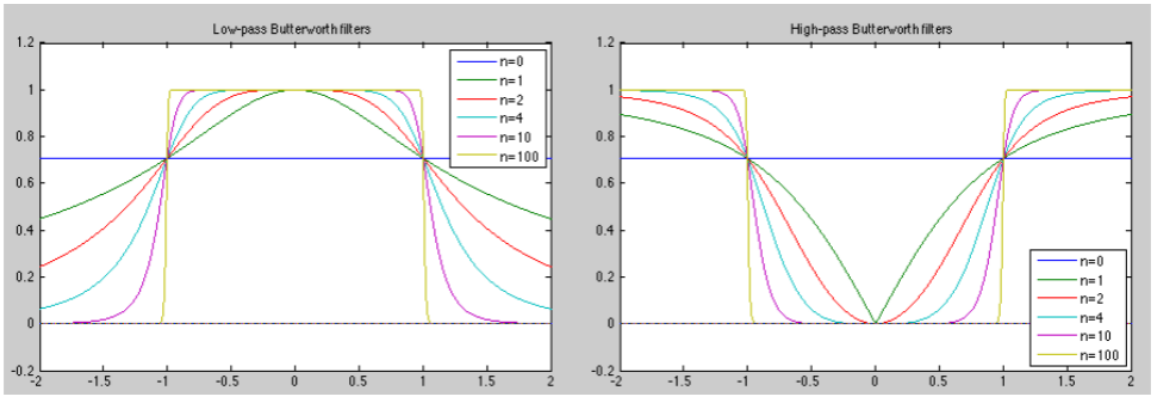
4.2.2 Σύνολο Δεδομένων Human Activity Recognition Using Smartphones

Ένα από τα πιο γνωστά σύνολα δεδομένων, το οποίο είναι ελεύθερο για χρήση αποτελεί το λεγόμενο 'Human Activity Recognition Using Smartphones' το οποίο έγινε διαθέσιμο μέσα από το UCL Machine Learning Repository το 2013. Ειδικότερα, δημιουργήθηκε και διατέθηκε από τους (Anguita, Ghio, Oneto, Parra Perez, & Reyes Ortiz, 2013) στο πανεπιστήμιο της Γένοβα. Τα δεδομένα συλλέχθηκαν από 30 άτομα ηλικίας μεταξύ 19 και 48 ετών, τα οποία εκτελούσαν μία από τις έξι τυπικές δραστηριότητες, ενώ φορούσαν ένα smartphone στη μέση τους, με στόχο την καταγραφή των δεδομένων κίνησης. Οι έξι δραστηριότητες, οι οποίες καταγράφηκαν είναι:

1. Περπάτημα
2. Ανέβασμα Σκάλας
3. Κατέβασμα Σκάλας
4. Καθιστή στάση
5. Όρθια στάση
6. Ξαπλωμένη στάση

Αναλυτικότερα, τα δεδομένα κίνησης καταγράφηκαν από το επιταχυνσιόμετρο (γραμμική επιτάχυνση στους άξονες x, y και z) και από το γυροσκόπιο (γωνιακή ταχύτητα) ενός Samsung Galaxy S II smartphone.

Όλες οι παρατηρήσεις καταγράφηκαν στα 50Hz, δηλαδή πενήντα σημεία δεδομένων ανά δευτερόλεπτο. Στη συνέχεια ο θόρυβος φιλτραρίστηκε χρησιμοποιώντας ένα median και ένα 20 Hz Butterworth φίλτρο με στόχο ακριβέστερα αποτελέσματα. Έπειτα, εφαρμόστηκε ένα δεύτερο 3 Hz Butterworth φιλτράρισμα για να εξαλειφθεί η επίδραση της βαρύτητας στα σήματα του επιταχυνσιόμετρου. Ύστερα όλες οι τιμές



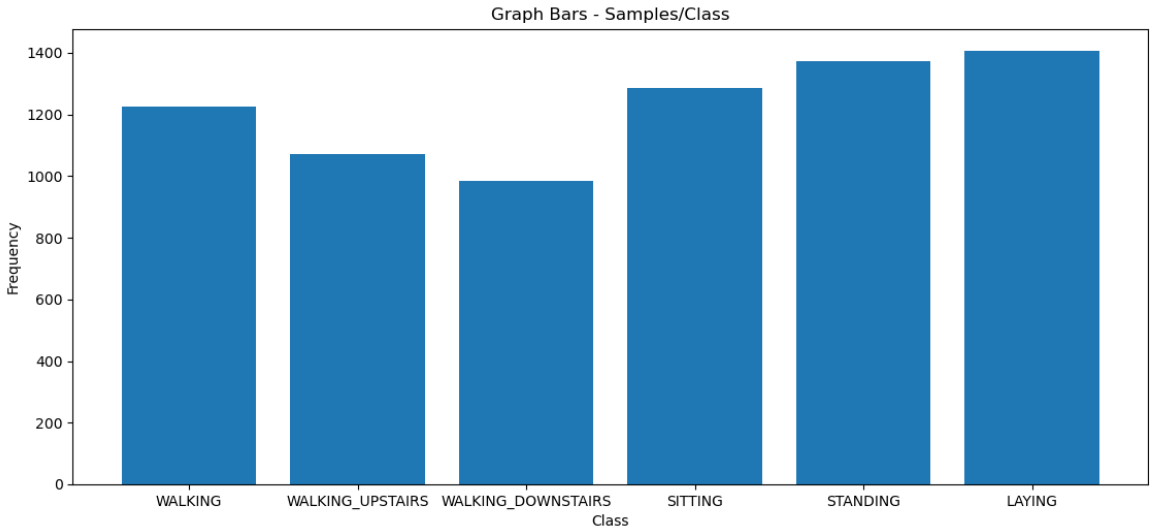
Εικόνα 4.2-1: Αναπαράσταση των υψηλών και χαμηλών συχνοτήτων φίλτρων (Bulbul, Cetin, & Dogru, 2018).

ομαλοποιήθηκαν σε διάστημα (-1,1). Ακόμα τα μεγέθη των ευκλειδείων τιμών των τριών διαστάσεων υπολογίζονται, προκειμένου να συγχωνεύσουν το τρισδιάστατο σήμα σε ένα σύνολο δεδομένων. Τελικά κάθε άτομο εκτελούσε την ακολουθία δραστηριοτήτων δύο φορές. Μία φορά με την συσκευή στο αριστερό χέρι και μία στο δεξί. Επιπρόσθετα, η διαδικασία καταγραφής πραγματοποιήθηκε εντός κλειστού ελεγχόμενου χώρου εργαστηρίου προς αποφυγή περεταίρω θορύβων.

4.2.3 Δεδομένα Πειράματος

Προτού παρουσιαστούν οι αρχιτεκτονικές των δικτύων που χρησιμοποιήθηκαν, παρακάτω θα παρουσιαστούν και θα αναλυθούν τα δεδομένα με τα οποία επεξεργάζονται και εκπαιδεύονται τα δίκτυα.

Όπως προαναφέρθηκε στην προηγούμενη ενότητα τα αρχεία προέρχονται από το UCL Machine Learning Repository και αποτελούνται από δύο κύριους φακέλους (εκπαίδευσης και δοκιμής), όπου κάθε φάκελος απαρτίζεται από εννέα διαφορετικά αρχεία .txt τα οποία περιέχουν τις μετρήσεις του γυροσκοπίου και του επιταχυνσιόμετρου στους x, y, z άξονες με τις αντίστοιχες ετικέτες τους. Ομοίως, υπάρχουν και άλλα εννέα αρχεία με διαφορετικές μετρήσεις στον φάκελο δοκιμής για την αξιολόγηση των δικτύων. Τα συνολικά δεδομένα εκπαίδευσης είναι 7352 με τις αντίστοιχες ετικέτες τους. Στο παρακάτω σχήμα φαίνεται



Εικόνα 4.2-2: Γραφική απεικόνιση των πειραματικών δεδομένων ως προς τις ετικέτες τους.

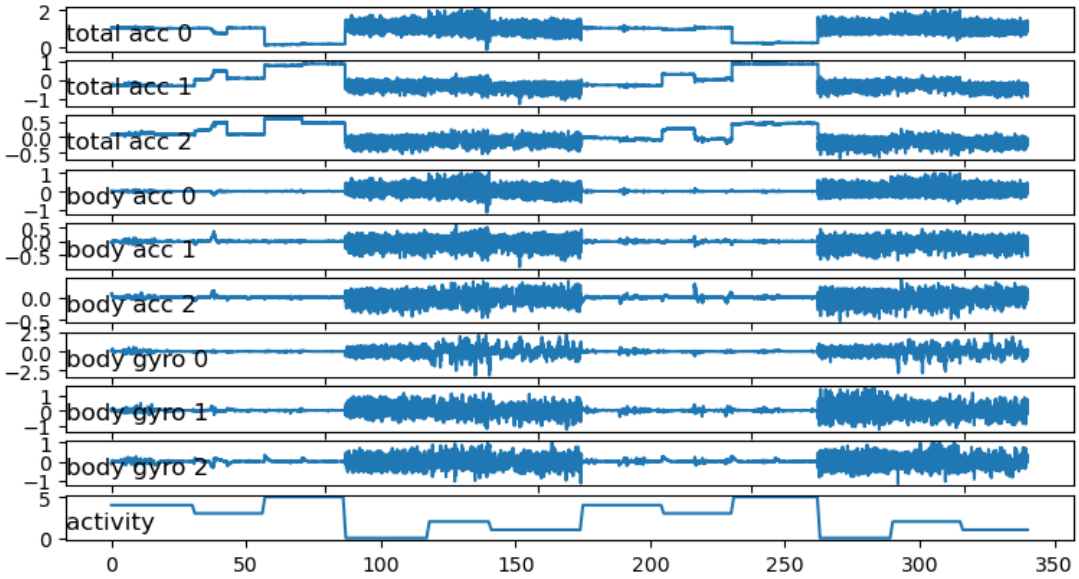
η κατανομή των δειγμάτων ως προς τις ετικέτες τους.

Επομένως, γίνεται κατανοητό πως οι κλάσεις δεν είναι ισοδύναμα μοιρασμένες, αντιθέτως μοιράζονται σε τέτοιο βαθμό που δε χρειάζεται περαιτέρω προ-επεξεργασία. Η κλάση με το μεγαλύτερο ποσοστό (∼19%) είναι η ετικέτα με την ‘’Ξαπλωμένη Στάση’’, ακολουθούν οι ετικέτες με την ‘’Καθιστή και Όρθια Στάση’’ με ποσοστό περίπου ∼18% και τέλος οι υπόλοιπες τρείς. Η ετικέτα με το λιγότερο ποσοστό είναι κλάση ‘’Κατέβασμα Σκάλας’’ με ποσοστό περίπου ∼13%.

	Συχνότητα Εμφάνισης (~%)
Walking	16.6
Waling Upstairs	14.5
Walking Downstairs	13.4
Sitting	17.5
Standing	18.6
Laying	19.1

Πίνακας 4.2-4.2-1: Πίνακας Συχνότητας εμφάνισης δειγμάτων ανά κατηγορία..

Στην συνέχεια, στο παρακάτω διάγραμμα αναπαρίστανται οι εννέα τιμές των αισθητήρων, οι οποίοι καταγράφηκαν κατά τη διάρκεια εκτέλεσης και των έξι δραστηριοτήτων ενός ατόμου. Το διάγραμμα αυτό αποτελείται από δέκα γραφικές παραστάσεις, όπου οι εννέα αναπαριστούν την τιμή του κάθε αισθητήρα την χρονική στιγμή t. Αναφορικά με την τελευταία γραφική παράσταση, αναπαριστά την συνολική δραστηριότητα του ατόμου.



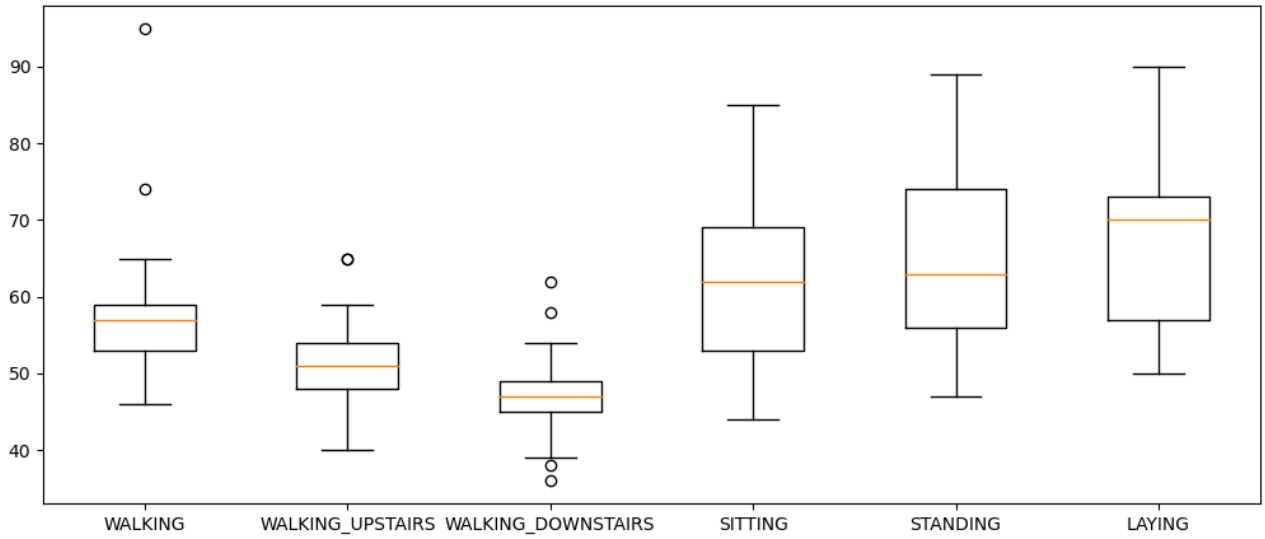
Εικόνα 4.2-3: Γραφική απεικόνιση των τιμών των αισθητήρων κατά την διάρκεια μιας δραστηριότητας του ατόμου.

Στο παραπάνω διάγραμμα παρατηρούνται περίοδοι μεγάλης κίνησης, οι οποίοι αντιστοιχούν στις τρείς δραστηριότητες που έχουν να κάνουν με κίνηση (Περπάτημα, Ανέβασμα Σκάλας, Κατέβασμα Σκάλας). Εντούτοις, υπάρχουν και περίοδοι με λιγότερη δραστηριότητα (δηλαδή σχετικά ευθεία γραμμή) που αντιστοιχούν στις άλλες τρεις δραστηριότητες, οι οποίες έχουν να κάνουν με ακινησία (Όρθια Στάση, Καθιστή Στάση, Ξαπλωμένη Στάση). Αξίζει να σημειωθεί σε αυτό το σημείο ότι το κάθε άτομο έχει εκτελέσει τις δραστηριότητες τουλάχιστον από δύο φορές. Αυτό υποδηλώνει ότι για κάθε άτομο, δεν πρέπει να γίνονται υποθέσεις σχετικά με την σειρά των δραστηριοτήτων που μπορεί να έχουν εκτελεσθεί.

Μπορεί να υπάρξει, επίσης, σε κάποια δεδομένα ατόμων σχετικά μεγάλη κινητικότητα (υψηλές τιμές στις μετρήσεις) για ορισμένες ‘’στάσιμες’’ δραστηριότητες, όπως η ‘’Ξαπλωμένη Στάση’’, είναι πιθανό να είναι ακραίες τιμές ή να σχετίζονται με τη μετάβαση σε άλλη δραστηριότητα. Σε τέτοιες περιπτώσεις, συνήθως η εξομάλυνση ή η κατάργηση τέτοιων τιμών μπορεί να αντιμετωπίσει το πρόβλημα.

Ένας τελευταίος τομέας, ο οποίος πρέπει να ληφθεί υπόψιν αποτελεί το χρονικό διάστημα, το οποίο ένα άτομο ξοδεύει σε κάθε δραστηριότητα. Αυτό είναι άρρηκτα συνδεδεμένο με την ισορροπία μεταξύ των κλάσεων. Εάν οι κλάσεις σε ένα σύνολο δεδομένων είναι σχετικά ισορροπημένες, τότε αναμένεται να υπάρχει ισορροπία και στη διάρκεια που αφιερώνει ένα άτομο μεταξύ των δραστηριοτήτων. Ο τρόπος με τον οποίο εξετάζεται αυτή ισορροπία είναι απλή. Υπολογίζεται το πόσο καιρό (σε δείγματα ή ακολουθίες) κάθε άτομο ξοδεύει σε κάθε δραστηριότητα (κλάση) και εξετάζεται η κατανομή των χρονικών διαστημάτων για κάθε δραστηριότητα. Ο πιο γνωστός τρόπος να επιτευχθεί αυτό παρουσιάζεται παρακάτω

στα διαγράμματα boxplot στα οποία συνοψίζεται το εύρος των δεδομένων, η μέση τιμή αυτών 50% (το πλαίσιο), το median (μια γραμμή) και τα ακραία σημεία (οι τελείες).



Εικόνα 4.2-4: Γραφική αναπαράσταση boxplot για το σύνολο δεδομένων

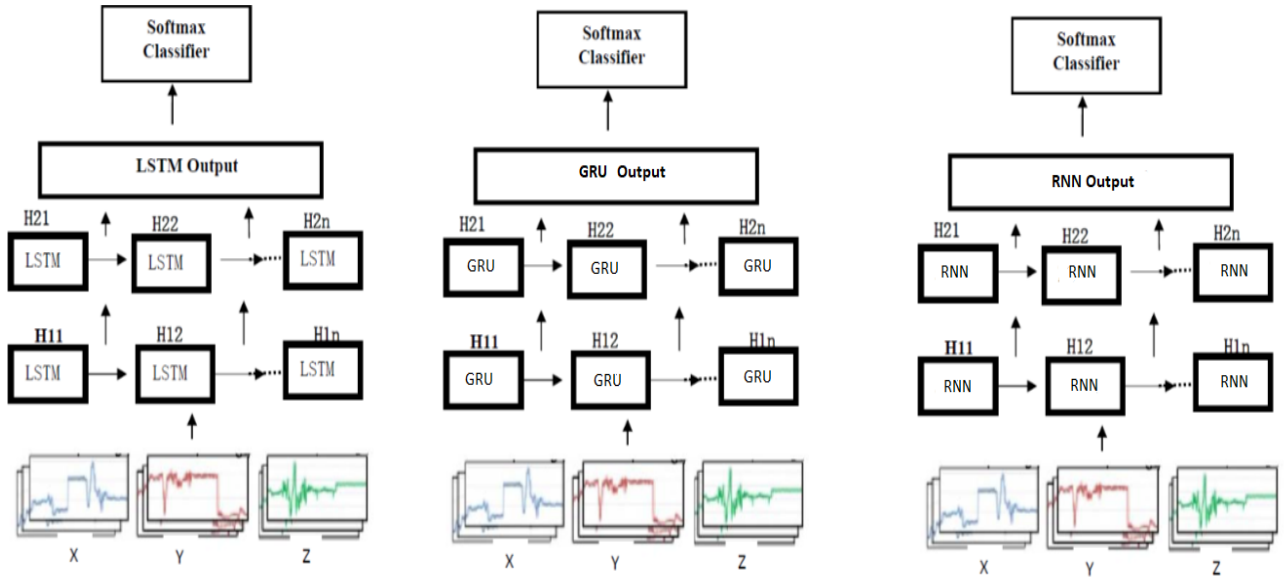
Κάθε boxplot συνοψίζει πόσο καιρό (σε σειρές η σε αριθμό παραθύρων) τα άτομα στο σύνολο δεδομένων δαπάνησαν για την κάθε δραστηριότητα. Παραπάνω φαίνεται ότι τα άτομα ξόδεψαν περισσότερο χρόνο σε στάσιμες δραστηριότητες (Καθιστική Στάση', 'Όρθια Στάση' , 'Ξαπλωμένη στάση') και λιγότερο χρόνο σε δραστηριότητες κίνησης ('Περπάτημα' ,'Ανέβασμα Σκαλών' ,'Κατέβασμα Σκαλών'). Η εξάπλωση μεταξύ των δραστηριοτήτων δεν είναι μεγάλη, γεγονός που υποδηλώνει ότι δεν υπάρχει ανάγκη περικοπής και προ-επεξεργασίας των δραστηριοτήτων κίνησης.

4.2.4 Αρχιτεκτονική Δικτύων

Η αναγνώριση ανθρώπινης δραστηριότητας όπως έχει αναφερθεί και προηγουμένως είναι ένα κλασικό πρόβλημα ακολουθιακών εξαρτήσεων , έτσι τα πιο κατάλληλα δίκτυα που μπορούν να το αντιμετωπίσουν είναι τα RNN. Όπως περιγράφηκε και στην εισαγωγή, στα πλαίσια της διπλωματικής, υλοποιούνται για την προσέγγιση με αισθητήρες από Smartphone, τρία διαφορετικά δίκτυα ακολουθιακών εξαρτήσεων ένα LSTM , ένα RNN και ένα GRU. Στα παρακάτω σχήματα παρουσιάζονται τα τρία παρόμοια μοντέλα. Δεδομένης της τρισδιάστατης πληροφορίας που προσφέρει το επιταχυνσιόμετρο και το

γυροσκόπιο , χρησιμοποιείται ένα ‘’συρόμενο παράθυρο’’ με μήκος τιμών N για την εξαγωγή δεδομένων εισόδου στα μοντέλα. Προκειμένου να υπάρχει μια πιο πλούσια αναπαράσταση των δεδομένων, το μοντέλο έχει δύο επίπεδα LSTM,RNN και GRU αντίστοιχα.

Στο παρακάτω σχήμα X είναι τα δεδομένα του επιταχυνσιόμετρου και του γυροσκόπιου στην κατεύθυνση X. Ομοίως Y για την κατεύθυνση Y και Z για την κατεύθυνση Z. Στην διπλωματική αυτή συνδυάζονται και οι τρείς είσοδοι για κάθε αισθητήρα σ ’ένα τρισδιάστατο διάνυσμα (9 χαρακτηριστικά συνολικά) και τροφοδοτούνται στο μοντέλο ακολουθιακά βάση την τιμή του συρόμενου παραθύρου N=128. Έτσι τα δεδομένα εισαγωγής για τα μοντέλα είναι μια χρονική ακολουθία $NX9$ πίνακα. Στην συνέχεια η εξαγωγή των χαρακτηριστικών και των ακολουθιακών εξαρτήσεων με χρονικά διαστήματα N γίνεται από τις μονάδες LSTM,GRU,RNN αντίστοιχα. Τέλος η παραγόμενη έξοδος των μονάδων αυτών τροφοδοτείτε σε έναν πολύ-ταξινομητή ώστε να κατηγοριοποιήσει τα δεδομένα στις αντίστοιχες κλάσεις.



Εικόνα 4.2-5: Αρχιτεκτονική των μοντέλων RNN,LSTM και GRU που χρησιμοποιήθηκαν στην πειραματική διαδικασία.

4.2.5 Αποτελέσματα Δικτύων

Σε αυτή την ενότητα θα παρουσιαστούν τα αποτελέσματα των προαναφερθέντων μοντέλων για το σύνολο δεδομένων που αφορά την προσέγγιση με αισθητήρες από Smartphone. Αξίζει να σημειωθεί ότι η αξιολόγηση των μοντέλων έγινε υπολογίζοντας τέσσερις μετρικές. Η πρώτη μετρική αφορά τη συνολική ακρίβεια ή αλλιώς overall accuracy και ορίζεται ως:

$$Accuracy = \frac{Αριθμός\ σωστών\ προβλέψεων}{Συνολικός\ αριθμός\ προβλέψεων} \text{ ή } Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Η δεύτερη μετρική, η οποία χρησιμοποιείται είναι η Ανάκληση (Recall) όπου σχετίζεται με το ερώτημα ‘’Δοθέντος ενός δείγματος, το οποίο προέρχεται από την κλάση i , πόσο πιθανό είναι να ταξινομηθεί σωστά στον ταξινομητή;’’. Με άλλα λόγια πρόκειται για το ποσοστό διανυσμάτων, τα οποία διανύσματα προέρχονται από την κλάση i και ταξινομούνται στην κλάση αυτή:

$$Recall = \frac{TP}{TP + FN} \text{ ή } R_i = \frac{A(i,i)}{\sum_{j=1}^{N} A(i,j)}$$

Όσον αφορά την τρίτη μετρική είναι παρόμοιας λογικής με την Ανάκληση, μόνο που ικανοποιεί το ερώτημα ‘’Δοθέντος ενός δείγματος, το οποίο ταξινομήθηκε στην κλάση i , πόσο πιθανό είναι η ταξινόμηση αυτή να είναι σωστή; ‘’Ονομάζεται ακρίβεια ή αλλιώς Precision και αναπαριστά το ποσοστό των διανυσμάτων που ταξινομούνται στην κλάση i και πράγματι ανήκουν στην κλάση αυτή:

$$Precision = \frac{TP}{TP + FP} \text{ ή } P_i = \frac{A(i,i)}{\sum_{j=1}^{N} A(j,i)}$$

Τελευταία μετρική, η οποία χρησιμοποιείται είναι η F1 η οποία λαμβάνει υπόψη τόσο την ανάκληση όσο και την ακρίβεια και υπολογίζει την τελική ακρίβεια του δικτύου για κάθε κλάση, δίνοντας ίση βαρύτητα και στις δύο. Ορίζεται ως:

$$F_1 = \frac{2}{N} \sum_{j}^{N} \frac{precision_j * recall_j}{precision_j + recall_j}$$

Στους παραπάνω τύπους όπου j η κλάση δραστηριότητας, N το σύνολο των κλάσεων, TP = Αληθώς Θετικά, FN = Ψευδώς Αρνητικά, FP = Ψευδώς Θετικά και TN = Αληθώς Αρνητικά.

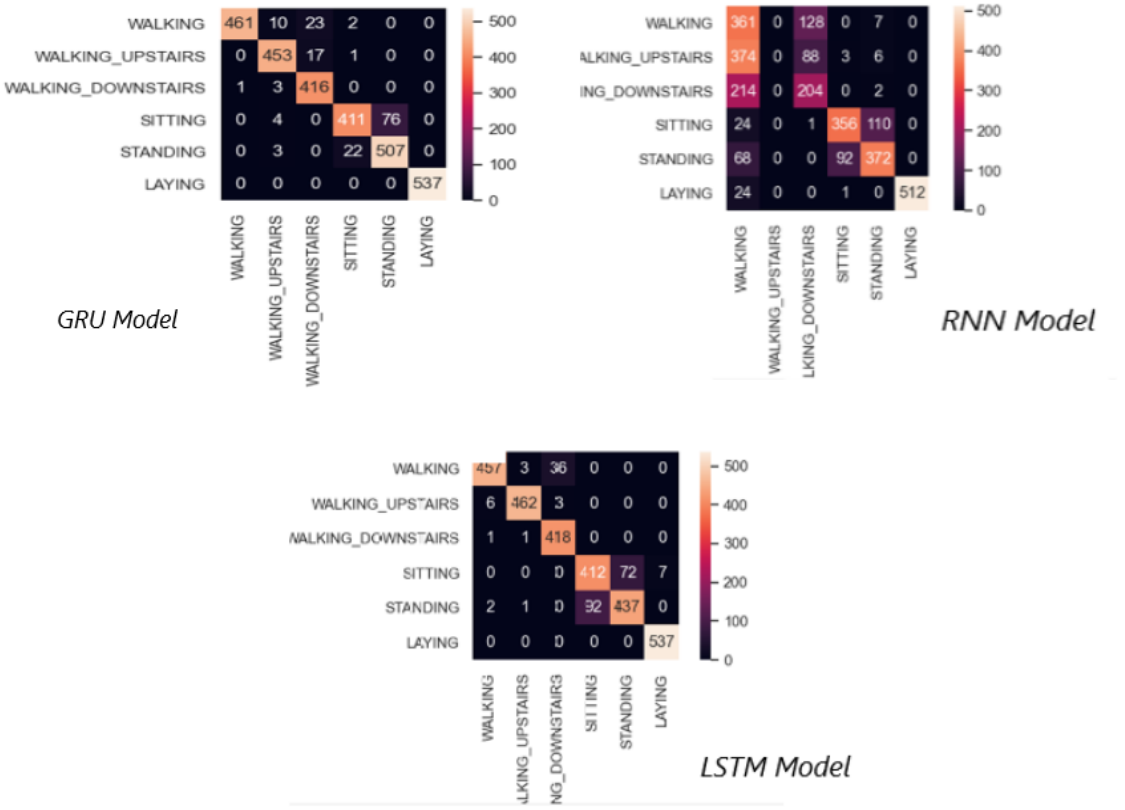
Στον πίνακα 4.2.1 που ακολουθεί παρουσιάζεται για κάθε μοντέλο τα αποτελέσματα, τα οποία προέκυψαν δοκιμάζοντας τα σύνολα δεδομένων αξιολόγησης, να σημειωθεί ότι τα δίκτυα εκπαιδεύτηκαν με 7352 δείγματα και 9 χαρακτηριστικά και αξιολογήθηκαν με 2947 δείγματα.

	RNN	LSTM	GRU
Accuracy	0.61	0.92	0.95
Recall	0.60	0.93	0.95
Precision	0.56	0.92	0.95
F_1	0.57	0.92	0.94

Πίνακας 4.2-4.2-2: Επιδόσεις των RNN,LSTM και GRU στο σετ αξιολόγησης.

Όπως γίνεται αντιληπτό από τα αποτελέσματα, το δίκτυο GRU έχει τη μεγαλύτερη ακρίβεια. Αυτή η πληροφορία αναμενόμενο μιας και οι μονάδες GRU είναι μια γενικευμένη μορφή των LSTM η οποία χρησιμοποιείται, για να αποτρέψει τα προβλήματα εξαφάνισης. Αντίθετα, παρατηρείται ότι τα RNN είναι αυτά με τη μικρότερη απόδοση, αυτό συμβαίνει καθώς τα LSTM και GRU μοντέλα είναι εξέλιξη των απλών RNN δικτύων και μπορούν να αντιμετωπίσουν προβλήματα εξαφάνισης ή έκρηξης.

Προκειμένου να εξαχθούν πιο ολοκληρωμένα συμπεράσματα για την λειτουργία των μοντέλων, είναι αναγκαίο να φανούν τα ακριβή σημεία στα οποία υστερούν και ποιες κλάσεις αναγνωρίζονται με μεγαλύτερη ευκολία και ποιες όχι. Σε αυτό το ζήτημα βοηθάει η εικόνα 4.2.6 βάσει της οποίας φαίνεται ο πίνακα σύγχυσης (confusion matrix) για καθένα από τα τρία μοντέλα. Οι στήλες του πίνακα αντιστοιχούν στις προβλέψεις του μοντέλου, ενώ οι γραμμές του αντιστοιχούν στις πραγματικές κλάσεις.



Εικόνα 4.2-6: Πίνακες σύγχυσης με τα αποτελέσματα κατηγοριοποίησης για κάθε εκτέλεση δικτύου πάνω στα σετ αξιολόγησης.

Όπως παρατηρείται, ειδικά στον πίνακα του RNN, εντοπίζονται ορισμένα λάθη όπου υπάρχει σύγχυση μεταξύ παρόμοιων ενεργειών (στατικές-ενεργητικές). Τα λάθη αυτά οφείλονται σε πολλούς παράγοντες, όπως είναι η επιλογή του σωστού χρονικού παραθύρου, καθώς η επιλογή λανθασμένου, μπορεί να εμποδίσει το μοντέλο να εντοπίσει σωστά τους χρονικούς συσχετισμούς. Επιπροσθέτως, ένα άλλο λάθος το οποίο ενδέχεται να προκύψει, μπορεί να προέρχεται μεταξύ των κλάσεων και της χρονικής τους οριοθέτησης. Πιο συγκεκριμένα, είναι δυνατό να μην υπάρχει σαφής χρονική διαφορά μεταξύ δύο δραστηριοτήτων κατά την εκτέλεση τους από ένα άτομο.

4.3 Προσέγγιση σε Πολυτροπικά Δεδομένα Αισθητήρων

Οι ανθρώπινες δραστηριότητες είναι εγγενώς πολυτροπικές. Αυτό το γεγονός έχει ως απόρροια, η αναγνώριση τους να αποτελεί ένα πρόβλημα πολλαπλών επιπέδων και παραγόντων, καθώς περιλαμβάνει οπτικοακουστικά στοιχεία, τα οποία θέτουν πολλές προκλήσεις σε επίπεδο χαρακτηριστικών και στο συνδυασμό αυτών. Οι τεχνολογίες οι οποίες χρησιμοποιούνται για αναγνώριση τέτοιων δραστηριοτήτων ποικίλουν και πολλές φορές βασίζονται σε διαφορετικές προσεγγίσεις. Χαρακτηριστικό παράδειγμα αποτελεί η χρήση αισθητήρων περιβάλλοντος και ακουστικών επιτρέπει να ανιχνευθεί η δραστηριότητα, κατά την οποία αλληλεπιδρά ο χρήστης με το περιβάλλον και με τα αντικείμενα που βρίσκονται σε αυτό. Όμως, η χρήση οπτικών μέσων και φορητών αισθητήρων είναι η πιο διαδεδομένη σε τέτοιου είδους προβλήματα. Οι αισθητήρες RGB-D (Red-Green-Blue Depth), δηλαδή κόκκινη-πράσινη-μπλε και αισθητήρες βάθους, μπορούν να θεωρηθούν ως βελτιωμένες συσκευές, οι οποίες βασίζονται στην όραση, καθώς μπορούν να παρέχουν επιπλέον δεδομένα βάθους, τα οποία με τη σειρά τους στη συνέχεια μπορούν να διευκολύνουν στην ανίχνευση ανθρώπινων δραστηριοτήτων. Αναλυτικότερα, γνωρίζοντας πληροφορίες από τους αισθητήρες βάθους είναι πιο εύκολο να εξαχθούν πληροφορίες, οι οποίες αφορούν την ανθρώπινη σιλουέτα ανεξάρτητα από σκιές, αντανακλάσεις φωτός και ομοιότητα χρωμάτων. Ακόμα, μπορούν να εξαχθούν σκελετικά στοιχεία, σχετικά με την στάση του ανθρώπου και κατά επέκταση να αξιοποιηθούν και αυτά στην αναγνώριση της δραστηριότητας.

Ο σκοπός της προσέγγισης αυτής, εκμεταλλευόμενοι τους αισθητήρες RGB-D, είναι να εφαρμοσθούν τεχνικές βαθιάς μάθησης σε σύνολο δεδομένων, το οποίο προέρχεται από αυτούς. Πιο συγκεκριμένα, θα χρησιμοποιηθούν δύο ειδών τεχνικές σύντηξης των δεδομένων και τρεις διαφορετικοί συνδυασμοί αναδρομικών δικτύων με συνελικτικά δίκτυα, με στόχο πάντα την εξαγωγή χαρακτηριστικών, τα οποία κατά επέκταση θα βοηθήσουν στην αναγνώριση δραστηριοτήτων.

4.3.1 Σχετική Δουλειά

Τα συστήματα αναγνώρισης ανθρώπινης δραστηριότητας, τα οποία βασίζονται σε βίντεο και ειδικά σε αισθητήρες RGB-D, όπως αναφέρθηκε και προηγουμένως, επιτρέπουν την εξαγωγή τέτοιων χαρακτηριστικών, τα οποία οδηγούν στην αναγνώριση της κίνησης του σώματος, χωρίς να είναι ενοχλητικές ως προς τον χρήστη. Πιο αναλυτικά, οι αισθητήρες αυτοί δεν χρειάζονται ούτε ιδιαίτερη εγκατάσταση, όπως συμβαίνει στους περιβαλλοντικούς αισθητήρες, ούτε εγείρουν προβλήματα που σχετίζονται με τις επιπτώσεις από την ακτινοβολία, όπως θα μπορούσε να είχε ένα κινητό ή ένα ραντάρ. Από την άλλη πλευρά, τέτοιοι αισθητήρες μπορεί να θεωρηθούν μη αποδεκτοί λόγω απορρήτου. Όταν, βέβαια, χρησιμοποιούνται μόνο οι αισθητήρες βάθους διατηρείται το απόρρητο, καθώς δεν συλλέγονται απλές εικόνες, ωστόσο η εξαγωγή δεν μπορεί να είναι αποδοτική μόνο με την χρήση αυτών. Έτσι, γίνεται κατανοητό πώς παρόλο που η πολυτροπική αναγνώριση ανθρώπινης δραστηριότητας είναι ακόμα ανοικτή στη βιβλιογραφία, προσεγγίσεις που βασίζονται σε βίντεο και σε RGB-D αισθητήρες είναι περιορισμένες. Παρόλα αυτά, έχουν δημοσιευτεί στο παρελθόν αρκετές διαφορετικές δημοσιεύσεις, αναφορικά με την αναγνώριση ανθρώπινης δραστηριότητας που βασίζεται σε RGB-D αισθητήρες, κάθε μια από τις οποίες προτείνει τη δική της προσέγγιση για το πρόβλημα. Οι (Aggarwal & Xia, 2014) στη δημοσίευση τους προτείνουν μεθόδους που βασίζονται σε 3D δεδομένα, τα οποία μπορούν να ληφθούν από τρείς διαφορετικές τεχνολογίες: από συστήματα βασισμένα σε δείκτες, από stereo εικόνες ή από αισθητήρες εύρους. Ακόμη, οι (Li, Zhang, & Liu, 2010) εκμεταλλευόμενοι την απλότητα στην εξαγωγή πληροφορίας από τους αισθητήρες βάθους δημοσίευσαν μια προσέγγιση, η οποία υπολογίζει τα 3D σημεία της σιλουέτας του ανθρώπου, βασιζόμενη στα περιγράμματα των επίπεδων προβολών του τρισδιάστατου χάρτη βάθους. Έτσι δημιούργησαν ένα γράφημα, στο οποίο ο κάθε κόμβος απεικονίζει μια στάση. Χαρακτηριστικά από 2D σημεία έχουν ληφθεί υπόψη από τους (Chaaraoui, Climent-Perez, & Florez-Revuelta, 2013), όπου μια δραστηριότητα διαμορφώνεται ως ακολουθία πολλών θέσεων και εξάγεται μέσω ενός αλγορίθμου ομαδοποίησης, από ένα σύνολο εκπαίδευσης. Από την άλλη πλευρά οι (Voulodimos, Doulamis, Doulamis, & Lalos, 2016) επικεντρώθηκαν στη δημιουργία μιας μεθόδου για την αναγνώριση δραστηριότητας μέσα από ροές βίντεο (Voulodimos et al., 2011). Ειδικότερα, προτείνουν έναν αλγόριθμο παρακολούθησης, ο οποίος χρησιμοποιεί ροές αντικειμένων ως μοντέλο κίνησης (Lalos et al., 2014) και εκτιμά την μετατόπιση και κατεύθυνση αυτών σε ροές εικόνων. Άλλες προσεγγίσεις οι οποίες αξιοποιούν τα δεδομένα βάθους θεωρούνται επίσης η εξαγωγή τοπικών και ολιστικών περιγραφών. Μία από αυτές είναι των (Vieira, Nascimento, Oliveira, Liu, & Campos, 2012) όπου μια χώρο-χρονική υποδιαίρεση του χώρου χωρίζεται σε πολλά τμήματα, όπου τα μοτίβα πληρότητας εξάγονται από ένα 4D πλέγμα. Ακόμα, ολιστικοί περιγραφείς, ή αλλιώς HON4D και τα HOPC έχουν αξιοποιηθεί αντίστοιχα από τους (Oreifej & Liu, 2013), (Rahmani,

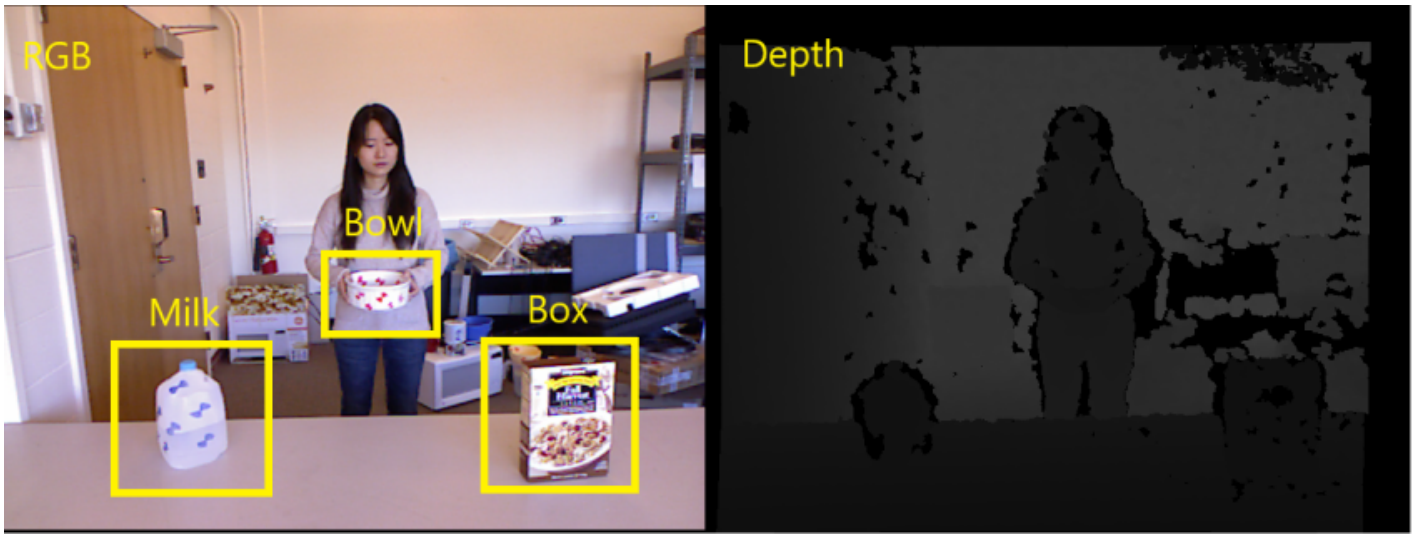
Mahmood, Huynh, & Mian, 2014). Το HON4D βασίζεται στον προσανατολισμό των κανονικών επιφανειών στο 4D χώρο, ενώ το HOPC μπορεί να αντιπροσωπεύει τα γεωμετρικά χαρακτηριστικά μιας ακολουθίας τρισδιάστατων σημείων.

4.3.2 Σύνολο Δεδομένων CAD-120

Το Cornell Activity Dataset– 120 αποτελεί ένα από τα πιο γνωστά και ολοκληρωμένα σύνολα δεδομένων στο πεδίο της ανθρώπινης αναγνώρισης δραστηριότητας με βίντεο. Όπως προκύπτει από το όνομα του έχει συλλεχθεί από τους (Sung, Ponce, Selman, & Saxena, 2011) ερευνητές του πανεπιστημίου Cornell και εστιάζει σε υψηλού επιπέδου ανθρώπινες δραστηριότητες με αλληλεπιδράσεις αντικειμένων. Αποτελείται από δέκα δραστηριότητες, οι οποίες εκτελέστηκαν από τέσσερα διαφορετικά άτομα με κάθε δραστηριότητα να πραγματοποιείται τρεις φορές με διαφορετικά αντικείμενα κάθε φορά. Οι δέκα δραστηριότητες είναι:

1. Οργάνωση Αντικειμένων (Arranging Obj.)
2. Καθαρισμός Αντικειμένων (Cleaning Obj.)
3. Κατανάλωση Φαγητού (Having Meal)
4. Παρασκευή Δημητριακών (Making Cereal)
5. Τοποθέτηση Φαγητού (σε φούρνο μικροκυμάτων) (Microwaving Food)
6. Παραλαβή Αντικειμένων (Picking Obj.)
7. Στοίβαγμα Αντικειμένων (Stacking Obj.)
8. Παραλαβή Φαγητού (από φούρνο μικροκυμάτων) (Taking Food)
9. Λήψη Φαρμάκου (Taking Med)
10. Διαχωρισμός Αντικειμένων (Unstacking Obj.)

Να σημειωθεί σε αυτό το σημείο ότι οι δραστηριότητες καταγράφηκαν με τη βοήθεια Kinect αισθητήρα. Επιπλέον, κάθε δραστηριότητα αποτελείται από μια ακολουθία υπό-δραστηριοτήτων. Διαφορετικά άτομα πραγματοποίησαν τις υπό-δραστηριότητες για διαφορετικό χρονικό διάστημα και με διαφορετική σειρά και τρόπο εκτέλεσης. Αυτό έχει ως αποτέλεσμα κάθε άτομο να μπορεί να εκτελεί την ίδια δραστηριότητα με διαφορετικά αντικείμενα. Η μορφή του συνόλου δεδομένων είναι σε τελικό στάδιο, δηλαδή έχει επεξεργαστεί και έχει μετατραπεί από βίντεο σε ακολουθίες RGB-D εικόνων.



Εικόνα 4.3-1: Αναπαράσταση ενός RGB-D δείγματος από το σύνολο δεδομένων CAD-120

Παραπάνω φαίνεται ένα frame για την κατηγορία 'Παρασκευή Δημητριακών', στην αριστερή πλευρά είναι η RGB εικόνα και στη δεξιά η εικόνα βάθους. Επίσης με κίτρινο πλαίσιο απεικονίζονται τα αντικείμενα με τα οποία αλληλεπιδρά το άτομο. Τέλος, αξίζει να σημειωθεί ότι και τα RGB, καθώς και τα frames βάθους δίνονται σε μέγεθος 640 x 480 με τρία κανάλια χρωμάτων και ένα κανάλι αντίστοιχα.

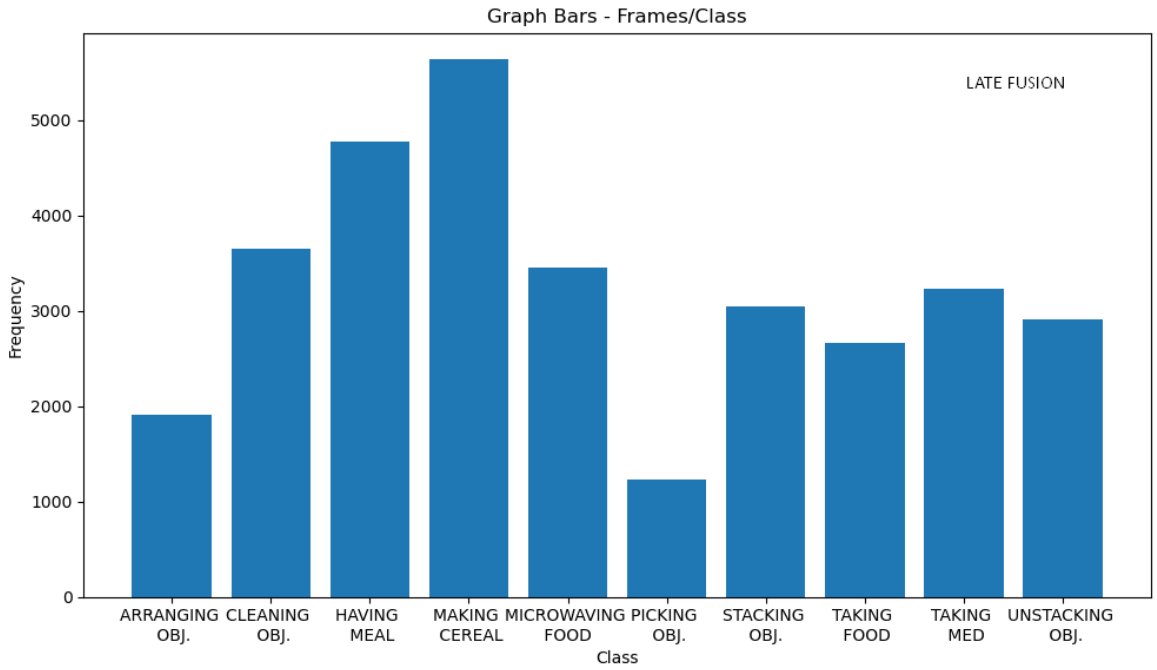
4.3.3 Δεδομένα Πειράματος

Προτού παρουσιαστούν οι αρχιτεκτονικές των δικτύων και οι διαφορετικές τεχνικές, οι οποίες χρησιμοποιήθηκαν, παρακάτω θα παρουσιαστούν και θα αναλυθούν τα δεδομένα αυτά με τα οποία επεξεργάζονται και εκπαιδεύονται τα δίκτυα.

Αρχικά, τα αρχεία προέρχονται από το CAD-120 σύνολο δεδομένων του πανεπιστημίου Cornell, αποτελούνται από δέκα κύριους φακέλους, οι οποίοι υποδεικνύουν την κάθε δραστηριότητα του κάθε ατόμου, όπου κάθε φάκελος αποτελείται από τρεις υπό-φακέλους που ο καθένας αποτελείται από τα RGB και Depth frames για την εκάστοτε εκτέλεση της συγκεκριμένης δραστηριότητας. Σε αυτό το σημείο αξίζει να σημειωθεί ότι τα RGB frames πρόκειται για έγχρωμες εικόνες (Red, Green, Blue) με απεικόνιση τριών διαστάσεων R_i, G_i, B_i και κατά συνέπεια τριών καναλιών, άρα το τελικό σχήμα του πίνακα αποθήκευσης τους να είναι της μορφής (Height,Width, Number of Channels) για κάθε frame. Τα συνολικά δεδομένα για

κάθε μέθοδο υλοποίησης είναι διαφορετικά. Αυτό οφείλεται στους διαφορετικούς πόρους, οι οποίοι απαιτούνται για την εκπαίδευση των δικτύων. Λόγο των διαφορετικών τεχνικών σύμμειξης των δεδομένων.

Για την πρώτη τεχνική, η οποία είναι η λεγόμενη ‘late fusion’, θα γίνει εκτενή αναφορά στις παρακάτω ενότητες για το πως λειτουργεί, χρησιμοποιούνται συνολικά 65008 frames από έγχρωμες και βάθους εικόνες, εκ των οποίων το 80% χρησιμοποιήθηκε για την εκπαίδευση των δικτύων και 20% για την αξιολόγηση τους. Στο παρακάτω σχήμα φαίνεται η κατανομή των δειγμάτων ως προς τις αντίστοιχες κλάσεις.



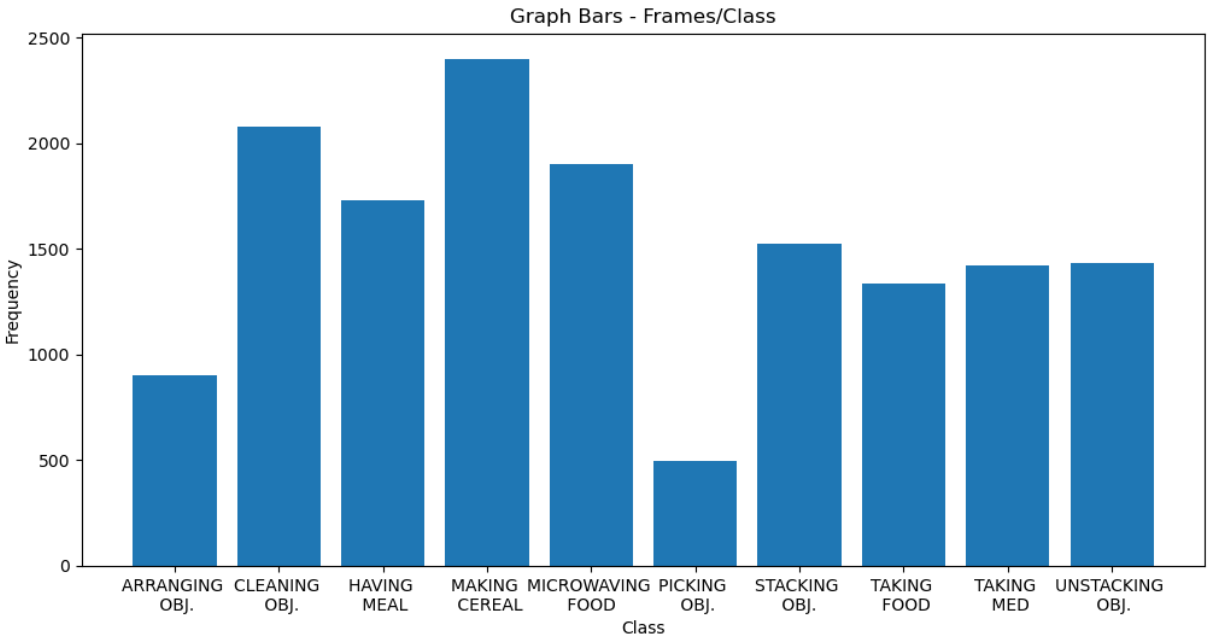
Εικόνα 4.3-2: Γραφική απεικόνιση των πειραματικών δεδομένων ως προς τις ετικέτες τους για την late fusion τεχνική.

Επομένως, γίνεται κατανοητό ότι οι κλάσεις δεν είναι ισοδύναμα μοιρασμένες, γεγονός το οποίο οφείλεται στο δοθέν σύνολο δεδομένων και στο πως έχει γίνει η συλλογή και καταγραφή των εικόνων από τα οπτικά μέσα. Αυτό, όμως, έχει ως συνέπεια οποιαδήποτε προσπάθεια κανονικοποίησης του συνόλου δεδομένων να μην επιφέρει κάποιο καλύτερο αποτέλεσμα. Όσον αφορά την συχνότητα εμφάνισης των frames, ανά ετικέτα προκύπτει ότι η κλάση με το μεγαλύτερο ποσοστό (~17.3%), όπως φαίνεται και στον παρακάτω πίνακα, αποτελεί η ετικέτα με την ‘’Παρασκευή Δημητριακών’’, ακολουθεί η ετικέτα ‘’Κατανάλωση Φαγητού’’ με ποσοστό περίπου ~17%, ενώ η ετικέτα με το λιγότερο ποσοστό είναι η ‘’Παραλαβή Αντικειμένου’’ με ποσοστό περίπου ~3.8%.

	Συχνότητα Εμφάνισης (~%)
Arranging Obj.	5.8
Cleaning Obj.	11.2
Having Meal	14.7
Making Cereal	17.3
Micro Food	10.6
Picking Obj.	3.8
Stacking Obj	9.4
Taking Food	8.2
Taking Med	9.9
Unstacking Obj	8.9

Πίνακας 4.3-1: Πίνακας Συχνότητας εμφάνισης δειγμάτων ανά κατηγορία για την late fusion μέθοδο.

Για τη δεύτερη τεχνική ή αλλιώς early fusion, θα γίνει εκτενή αναφορά στις παρακάτω ενότητες για το πως λειτουργεί, το σύνολο δεδομένων, το οποίο χρησιμοποιήθηκε είναι πιο μικρό, λόγω της μεγάλης υπολογιστικής ισχύς που χρειαζόταν η επεξεργασία του. Πιο συγκεκριμένα, αποτελείται συνολικά από 15277 RGB-D frames, εκ των οποίων όπως και στην προηγούμενη τεχνική το 80% χρησιμοποιήθηκε για εκπαίδευση του δικτύου, ενώ το 20% για αξιολόγηση. Παρακάτω φαίνεται η αντίστοιχη κατανομή δειγμάτων και ετικετών για τη μέθοδο αυτή.



Εικόνα 4.3-3: Γραφική απεικόνιση των πειραματικών δεδομένων ως προς τις ετικέτες τους για την early fusion τεχνική.

Όπως γίνεται αντιληπτό, τα δείγματα σε αυτή τη μέθοδο είναι αρκετά λιγότερα και με ακόμα περισσότερες ανισορροπίες, ως προς τις ποσότητες στις κλάσεις. Σε αυτή τη μέθοδο έγινε προσπάθεια κανονικοποίησης των κλάσεων σε σχετικά παρόμοια επίπεδα, όσο είναι επιτρεπτό, αφού χρειάστηκε να αποκοπούν πολλά δεδομένα από το κύριο σύνολο δεδομένων. Έτσι, παρατηρείται με μεγαλύτερη συχνότητα εμφάνισης η κλάση ‘’Παρασκευή Δημητριακών’’ με ποσοστό ~15.7% και ακολουθεί η ετικέτα ‘’Καθαρισμός Αντικειμένων’’ με ποσοστό ~13.2%, ενώ η ετικέτα με το μικρότερο ποσοστό είναι η ‘’Παραλαβή Αντικειμένου” με ποσοστό περίπου ~3.2%.

	Συχνότητα Εμφάνισης (~%)
Arranging Obj.	5.9
Cleaning Obj.	13.6
Having Meal	11.3

Making Cereal	15.7
Micro Food	12.5
Picking Obj.	3.2
Stacking Obj	10
Taking Food	8.7
Taking Med	9.3
Unstacking Obj	9.4

Πίνακας 4.3-2: Πίνακας Συχνότητας εμφάνισης δειγμάτων ανά κατηγορία για την early fusion Μέθοδο.

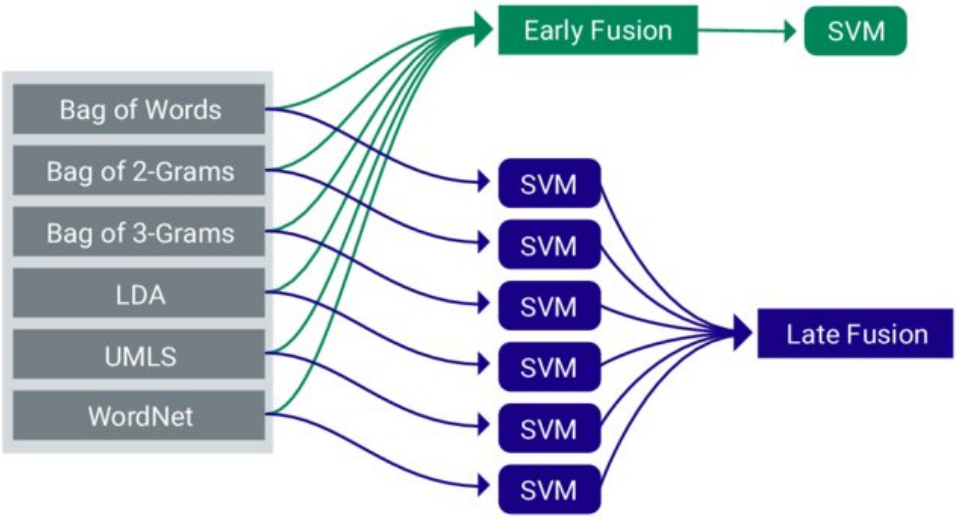
Σε αυτό το σημείο αξίζει να σημειωθεί ότι τα frames και για τις δύο μεθόδους, ενώ αρχικά ήταν μεγέθους 640 x 480, εν τέλει υπέστησαν επεξεργασία και μετασχηματίσθηκαν σε μεγέθους 200 x 200, λόγω περιορισμένης υπολογιστικής ισχύς, η οποία κατείχε το σύστημα εκπαίδευσης.

4.3.4 Μέθοδοι Σύμμειξης (Fusion)

Στην εν λόγω προσέγγιση χρησιμοποιήθηκαν δύο διαφορετικές μέθοδοι επεξεργασίας των δεδομένων. Η πρώτη είναι η λεγόμενη Early Fusion, βάσει της οποίας τα διανύσματα χαρακτηριστικών από διαφορετικές πηγές δεδομένων συνενώνονται πριν χρησιμοποιηθούν για ταξινόμηση από το δίκτυο. Με αυτόν τον τρόπο, δημιουργούνται νέα συνενωμένα δεδομένα αποτελούμενα από πολλά χαρακτηριστικά με αποτέλεσμα η εκπαίδευση, ο χρόνος κατηγοριοποίησης, αλλά και η υπολογιστική ισχύ να αυξηθεί. Ωστόσο, πολλές φορές σε συνδυασμό με τις κατάλληλες μεθόδους μάθησης μπορεί να οδηγήσει σε πολύ καλύτερη απόδοση στο τέλος.

Η δεύτερη μέθοδος αποτελεί η Late Fusion, η οποία υποδηλώνει την συνένωση των αποτελεσμάτων των εμπλεκόμενων ταξινομητών μετά τη διαδικασία της κατηγοριοποίησης. Αυτή η διαδικασία, παράγει το τελικό αποτέλεσμα σύμφωνα με τις βαθμολογίες των εμπλεκόμενων δικτύων η οποία είναι βασισμένη σε κανόνες απόφασης. Κάποιοι από αυτούς τους κανόνες είναι: η majority vote

όπου ύστερα από διαδικασία ψηφοφορίας προτείνετε η κλάση με την πλειοψηφία , η maximum όπου η κλάση με το μεγαλύτερο score μεταξύ των δικτύων προτείνετε και η average όπου η κλάση με το μεγαλύτερο μέσο όρο score προκρίνεται.



Εικόνα 4.3-4: Αναπαράσταση των Fusion μεθόδων, εικόνα είναι από: (Ebersmach, Herms, & Eibl, 2017)

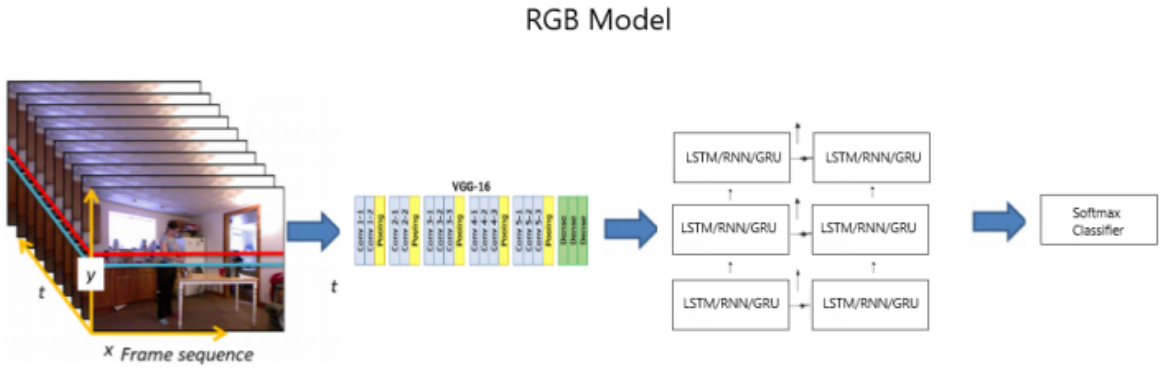
Όπως φαίνεται και στην παραπάνω εικόνα στην περίπτωση του Early Fusion τα αρχικά δεδομένα συνδυάζονται σε ένα σύνολο και στην συνέχεια τροφοδοτούνται στον ταξινομητή. Αντίθετα, στην Late Fusion μέθοδο τα αρχικά δεδομένα τροφοδοτούνται σε πολλούς ταξινομητές,03C8 για να πραγματοποιηθεί η κατηγοριοποίηση τους και στο τέλος τα αποτελέσματα αυτά συνδυάζονται μέσω της μεθόδου.

4.3.5 Αρχιτεκτονική Δικτύων με Late Fusion Μέθοδο

Η αναγνώριση ανθρώπινης δραστηριότητας όπως έχει αναφερθεί στην προηγούμενη προσέγγιση αποτελεί ένα κλασικό πρόβλημα ακολουθιακών εξαρτήσεων, ως εκ τούτου τα δίκτυα RNN, GRU & LSTM να είναι τα καταλληλότερα. Ωστόσο, αντίθετα με την προηγούμενη προσέγγιση, το σύνολο δεδομένων σε αυτή την περίπτωση είναι λίγο διαφοροποιημένο και κατά επέκταση και η αρχιτεκτονική των δικτύων. Καταρχάς το δοθέν σύνολο δεδομένων περιέχει έγχρωμες και βάθους εικόνες, με αποτέλεσμα αυτό από μόνο του να αποτελεί καθοριστικό παράγοντα διαφοροποίησης της αρχιτεκτονικής. Επιπροσθέτως, από την στιγμή που υπάρχουν δύο διαφορετικά είδη εικόνων (RGB και Depth) και χρησιμοποιείται η τεχνική Late Fusion, θα πρέπει να υπάρχει και διαφορετική αρχιτεκτονική των δικτύων για κάθε είδος εικόνας. Πιο συγκεκριμένα, βάσει της ανάλυσης του τρίτου κεφάλαιο, ο καλύτερος τρόπος εξαγωγής χαρακτηριστικών

από εικόνες, αποτελεί η χρήση των συνελικτικών δικτύων, αλλά λόγω της ακολουθιακής εξάρτησης, της οποίας παρουσιάζουν τα δεδομένα, ο συνδυασμός τους με αναδρομικά δίκτυα θεωρείται απαραίτητος.

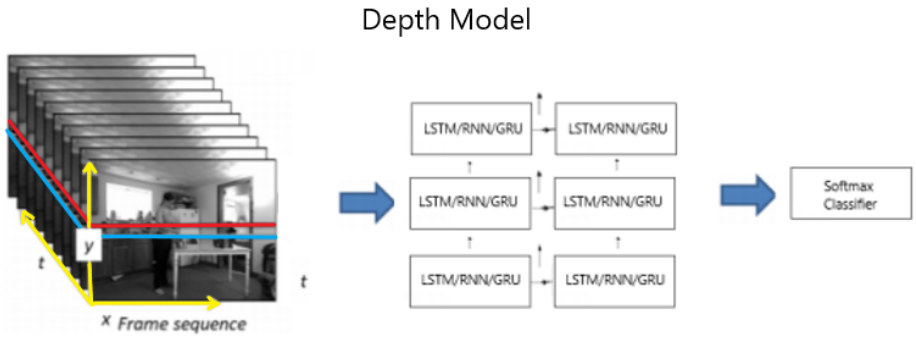
Όσον αφορά την εκπαίδευση των έγχρωμων εικόνων, στην αρχή για την εξαγωγή των χαρακτηριστικών τους, επιλέχτηκε να χρησιμοποιηθεί ένα προ-εκπαιδευμένο συνελικτικό δίκτυο εν ονόματι VGG16. Το VGG16 προτάθηκε από τους (Simonyan & Zisserman, 2014) στο πανεπιστήμιο της Οξφόρδης, ως ένα από τα κορυφαία δίκτυα στην κατηγοριοποίηση εικόνων, μάλιστα έχει πετύχει ποσοστό 92.7% στο ImageNet σύνολο δεδομένων, το οποίο περιέχει 14 εκατομμύρια εικόνες με 1000 κλάσεις. Επομένως, γίνεται κατανοητό πως η χρήση του δικτύου επιφέρει μεγάλη αποτελεσματικότητα στο συνολικό δίκτυο, καθώς μειώνει κατά μεγάλο βαθμό την πολυπλοκότητα των δεδομένων. Έπειτα, τα χαρακτηριστικά, τα οποία έχουν εξαχθεί από το VGG16 εισάγονται στα αναδρομικά δίκτυα LSTM, RNN & GRU αντίστοιχα για να μπορέσουν με τη βοήθεια των ακολουθιακών εξαρτήσεων να κατηγοριοποιήσουν σε τελικό στάδιο τις εικόνες στις σωστές κλάσεις. Παρακάτω απεικονίζεται ένα ενδεικτικό σχεδιάγραμμά της αρχιτεκτονικής για την εκπαίδευση των έγχρωμων εικόνων.



Εικόνα 4.3-5: Αναπαράσταση αρχιτεκτονικής μοντέλου εκπαίδευσης για RGB εικόνες.

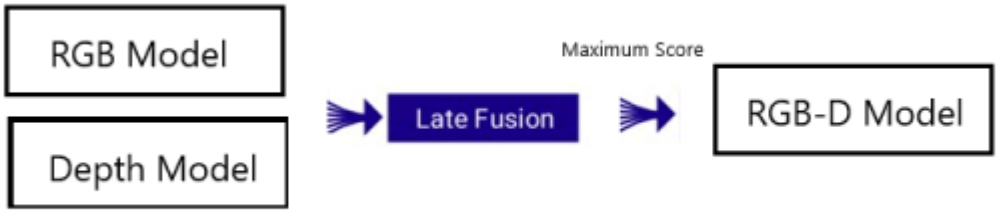
Σχετικά με την εκπαίδευση των εικόνων βάθους, ακολουθείται μια διαφορετική και πιο απλή αρχιτεκτονική, αφού δεν υπάρχει κανένα προ-εκπαιδευμένο δίκτυο εικόνων βάθους, έτσι επιλέχτηκε να γίνει χρήση κατευθείαν των αναδρομικών δικτύων. Πιο συγκεκριμένα, η εξαγωγή των χαρακτηριστικών καθώς και η ακολουθιακή εξάρτηση πραγματοποιείται κατευθείαν από τα LSTM, RNN & GRU δίκτυα, αφού τα δεδομένα λόγω μικρότερης πληροφορίας δεν είναι τόσο περίπλοκα, ώστε να απαιτούν προ

επεξεργασία από συνελικτικά δίκτυα. Παρακάτω φαίνεται ένα ενδεικτικό σχεδιάγραμμά της αρχιτεκτονικής για την εκπαίδευση των εικόνων βάθους.



Εικόνα 4.3-6: Αναπαράσταση αρχιτεκτονικής μοντέλου εκπαίδευσης για Depth εικόνες.

Καταλήγοντας στο τελευταίο μέρος της αρχιτεκτονικής των δικτύων, αξίζει να αναφερθεί ότι έχει άμεση σχέση με την Late Fusion τεχνική. Όπως αναφέρθηκε και στην προηγούμενη ενότητα, η Late Fusion τεχνική συνδυάζει τα αποτελέσματα των επιμέρους δικτύων και βάση ενός κανόνα παράγει το τελικό αποτέλεσμα. Στην συγκεκριμένη περίπτωση εφαρμόζεται η τεχνική στα αποτελέσματα των RGB και Depth δικτύων, με κανόνα απόφασης το λεγόμενο maximum score (προτείνεται η κλάση μεταξύ των δύο δικτύων, αυτή που έχει το μεγαλύτερο score) με στόχο να παραχθεί το τελικό μοντέλο, το οποίο θα τεθεί προς αξιολόγηση. Στην παρακάτω εικόνα αναπαρίσταται το σχεδιάγραμμα με τα αποτελέσματα των RGB και Depth δικτύων να συνδυάζεται με την Late Fusion τεχνική και στο τέλος να παράγεται το επιθυμητό μοντέλο.



Εικόνα 4.3-7: Αναπαράσταση παραγόμενου μοντέλου Late Fusion.

4.3.5.1 Αποτελέσματα Μεθόδου

Σε αυτή την ενότητα θα παρουσιαστούν τα αποτελέσματα των προαναφερθέντων μοντέλων για το σύνολο δεδομένων που αφορά την προσέγγιση σε πολυτροπικά δεδομένα, για την τεχνική Late Fusion. Αξίζει να σημειωθεί ότι η αξιολόγηση των μοντέλων έγινε υπολογίζοντας τέσσερις μετρικές, όπως και στην προσέγγιση με το σύνολο δεδομένων Smartphone.

Ειδικότερα, παρακάτω παρατίθενται δύο πίνακες. Ο πίνακας 4.3-3 αφορά τα αποτελέσματα των μοντέλων πριν γίνει το late fusion για την εκπαίδευση των έγχρωμων frames, ενώ ο άλλος πίνακας αφορά την εκπαίδευση των εικόνων βάθους. Να σημειωθεί ότι τα δίκτυα εκπαιδεύτηκαν με 52000 frames και αξιολογήθηκαν με 6501 frames διαφόρων κλάσεων, σε δέκα κλάσεις δραστηριότητας στο σύνολο.

Pre - Fusion Results (RGB)	**RNN**	**LSTM**	**GRU**
Accuracy	0.95	0.91	0.83
Recall	0.96	0.91	0.83
Precision	0.96	0.91	0.83
F_1	0.96	0.91	0.83

Πίνακας 4.3-3: Αναπαράσταση των μετρικών απόδοσης των δικτύων RGB πριν την εφαρμογή της μεθόδου Late Fusion.

Pre - Fusion Results (Depth)	RNN	LSTM	GRU
Accuracy	0.57	0.93	0.94
Recall	0.58	0.93	0.94
Precision	0.58	0.93	0.94
F_1	0.58	0.93	0.94

Πίνακας 4.3-4: Αναπαράσταση των αποδόσεων των δικτύων Depth πριν την εφαρμογή της μεθόδου Late Fusion.

Γίνεται κατανοητό πως οι τιμές των μετρικών μεταξύ των δύο εκπαιδεύσεων δεν ταυτίζονται και πολλές φορές κάποια υπερέχει της άλλης μετρικής. Αυτό συμβαίνει διότι το σύνολο δεδομένων είναι διαφορετικό σε κάθε περίπτωση, όπως άλλωστε και η αρχιτεκτονική του κάθε δικτύου. Στη συνέχεια παρατίθεται ο συνολικός πίνακας, ο οποίος προκύπτει ύστερα από την εφαρμογή της μεθόδου late fusion στα μοντέλα που προκύπτουν από τις δύο διαφορετικές εκπαιδεύσεις με την χρήση του κανόνα απόφασης maximum.

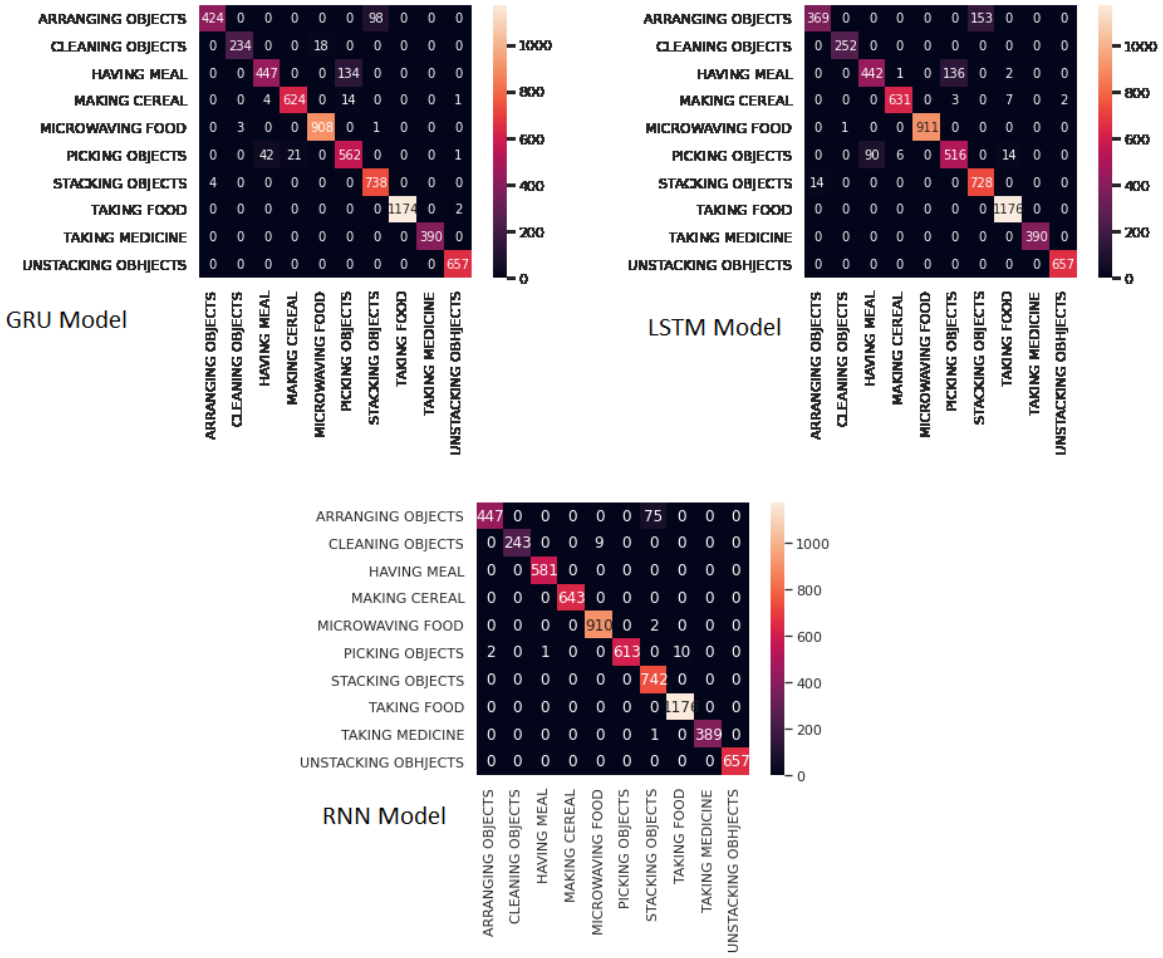
Overall Results (RGB-D)	RNN	LSTM	GRU
Accuracy	0.98	0.93	0.95
Recall	0.98	0.93	0.95
Precision	0.99	0.93	0.95
F_1	0.98	0.93	0.94

Πίνακας 4.3-5: Αναπαράσταση των συνολικών αποδόσεων των δικτύων μετά την εφαρμογή της μεθόδου Late Fusion.

Όπως φαίνεται παραπάνω η συνένωση των δύο εκπαιδεύσεων παράγει ένα αποτέλεσμα πολύ καλύτερο, σε σύγκριση με αυτό το οποίο παράγει η κάθε εκπαίδευση από μόνη της. Για παράδειγμα, κάποιες φορές παρατηρείται, ότι η εκπαίδευση των έγχρωμων εικόνων είναι καλύτερη, είναι πιθανό αυτό να οφείλεται στην χρήση του προ-εκπαιδευμένου μοντέλου που χρησιμοποιείται. Όμως, με την εφαρμογή της μεθόδου ισορροπείται και η απόδοση της εκπαίδευσης των εικόνων βάθους, ακόμα και αν είναι μικρότερη, με στόχο στο τέλος να υπάρχει το πιο αποδοτικό αποτέλεσμα. Συνολικά τα αποτελέσματα είναι ενθαρρυντικά ακόμα

και αν τα μοντέλα πολλές φορές ξεπερνούν σε απαιτήσεις το σύνολο δεδομένων. Επιπλέον παρατηρείται ότι τα RNN έχουν πολύ καλύτερη απόδοση από τα άλλα δύο δίκτυα, στη συνέχεια ακολουθούν τα GRU και τέλος τα LSTM.

Προκειμένου να βγουν πιο ολοκληρωμένα συμπεράσματα για την λειτουργία των μοντέλων, πρέπει να φανεί το ακριβές σημείο που υστερούν και ποιες κλάσεις αναγνωρίζονται με μεγαλύτερη ευκολία και ποιες όχι. Σε αυτό το ζήτημα, βοηθητική είναι η εικόνα 4.3-8, στην οποία απεικονίζεται ο πίνακας σύγχυσης (confusion matrix) για κάθε από τα τρία μοντέλα μετά από την εφαρμογή της μεθόδου late fusion.



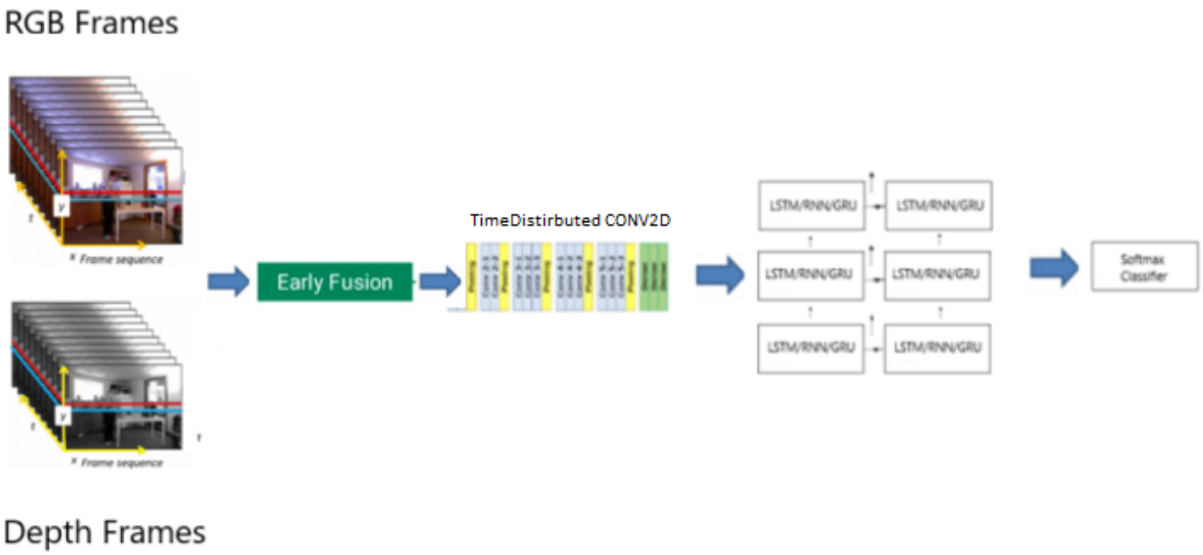
Εικόνα 4.3-8: Πίνακες σύγχυσης με τα αποτελέσματα κατηγοριοποίησης για κάθε εκτέλεση δικτύου πάνω στα σετ αξιολόγησης για την Late Fusion μέθοδο.

Παρατηρείται ότι οι κλάσεις με τις μεγαλύτερες απώλειες αποτελούν η "Κατανάλωση Φαγητού'' και η ''Παραλαβή αντικειμένων''. Ειδικότερα για την δεύτερη κλάση είναι κάτι αναμενόμενο, αφού το σύνολο δεδομένων από μόνο του δεν παρέχει πληθώρα δειγμάτων. Γίνεται κατανοητό, επίσης, πως και τα τρία

δίκτυα παρότι οι δραστηριότητες πολλές φορές πραγματοποιούνταν από διαφορετικούς ανθρώπους με διαφορετικά χαρακτηριστικά αλλά και από διαφορετικές γωνίες καταγραφής αποδίδουν σε έναν πολύ καλό βαθμό, χωρίς μεγάλες απώλειες.

4.3.6 Αρχιτεκτονική Δικτύων με Early Fusion Μέθοδο

Όπως και στην προηγούμενη αρχιτεκτονική με την μέθοδο Late Fusion, ο συνδυασμός αναδρομικών και συνελικτικών δικτύων είναι απαραίτητος. Ωστόσο, η ειδοποιός διαφορά, σε αυτή την περίπτωση αποτελεί η συνένωση των RGB και Depth εικόνων πριν την εκτέλεση των δικτύων. Πιο συγκεκριμένα, οι έγχρωμες εικόνες, όπως αναφέρθηκε και σε προηγούμενα κεφάλαια, αποτελούνται από τρία κανάλια χρωμάτων (R_i, G_i, B_i) ενώ οι εικόνες βάθους από ένα κανάλι. Αυτό που επιτυγχάνεται με τη μέθοδο early fusion είναι η προσθήκη ενός τέταρτου καναλιού (το κανάλι βάθους) στα ήδη υπάρχοντα τρία κανάλια των έγχρωμων εικόνων, με αποτέλεσμα την συνένωση τους. Στη συνέχεια, γίνεται η εξαγωγή των χαρακτηριστικών των τεσσάρων καναλιών εικόνων από συνελικτικό δίκτυο. Στη συγκεκριμένη μέθοδο δεν μπορεί να χρησιμοποιηθεί ένα προ-εκπαιδευμένο μοντέλο, όπως ήταν το VGG16 στην περίπτωση του Late Fusion, καθώς δεν είναι εκπαιδευμένο σε εικόνες με τέσσερα κανάλια. Επομένως, μια ιδιόμορφη χρήση των συνελικτικών δικτύων με ακολουθιακές εξαρτήσεις είναι απαραίτητη για την σωστή λειτουργία συνολικά των δικτύων.



Εικόνα 4.3-9: Αναπαράσταση αρχιτεκτονικής μοντέλου εκπαίδευσης για την Early Fusion μέθοδο

Στη συνέχεια με την γνωστή διαδικασία τα χαρακτηριστικά, τα οποία έχουν εξαχθεί από το συνελικτικό δίκτυο εισάγονται στα αναδρομικά δίκτυα LSTM, RNN & GRU αντίστοιχα, για να μπορέσουν με τη βοήθεια των ακολουθιακών εξαρτήσεων να κατηγοριοποιήσουν τις εικόνες στις σωστές κλάσεις. Παραπάνω αναπαρίσταται το διάγραμμα της αρχιτεκτονικής, το οποίο χρησιμοποιήθηκε για την Early Fusion μέθοδο. Τέλος, αξίζει να σημειωθεί ότι η διαδικασία της προ-επεξεργασίας των δεδομένων με την τεχνική early fusion αλλά και η εκπαίδευση των δικτύων αυτή καθαυτή, χρειάζεται αρκετό χρόνο και υπολογιστική ισχύ για να πραγματοποιηθεί.

4.3.6.1 Αποτελέσματα Μεθόδου

Σε αυτήν την ενότητα θα παρουσιαστούν τα αποτελέσματα των προαναφερθέντων μοντέλων για το σύνολο δεδομένων που αφορά την προσέγγιση σε πολυτροπικά δεδομένα, για την τεχνική Early Fusion. Αξίζει να σημειωθεί ότι η αξιολόγηση των μοντέλων έγινε υπολογίζοντας τέσσερις μετρικές, όπως με την προηγούμενη τεχνική Late Fusion. Αντίθετα, όμως, με την προηγούμενη μέθοδο η αξιολόγηση και η εκπαίδευση των δικτύων που χρησιμοποιήθηκαν περιορίζεται στην χρήση μόνο δύο LSTM και GRU, διότι η απόδοση των RNN ήταν αρκετά χαμηλή (<50%). Όποτε κρίθηκε αναγκαίο η παράλειψη τους.

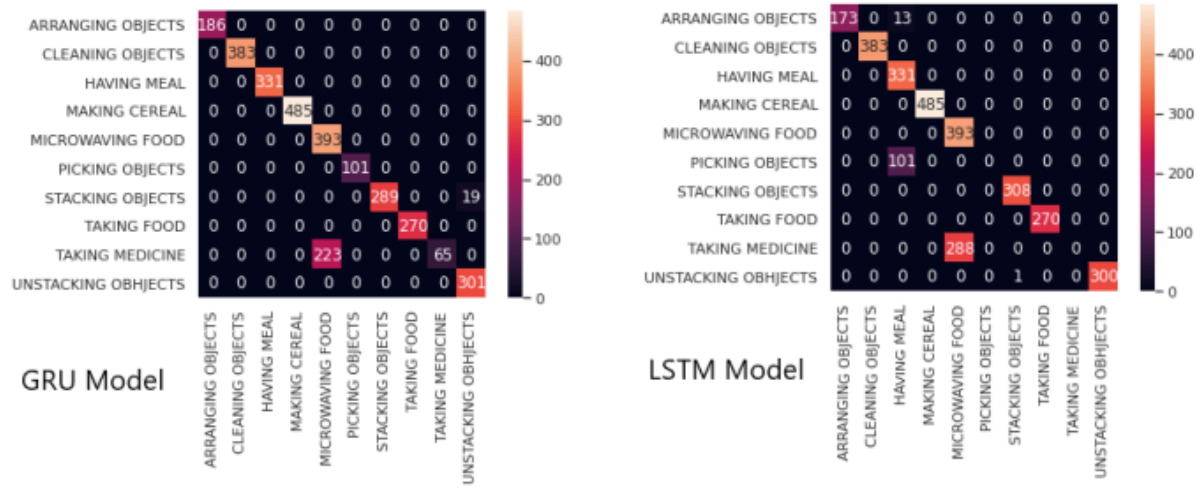
Αναλυτικότερα, παρακάτω παρατίθενται ο πίνακας 4.3-6 ο οποίος αφορά τα συνολικά αποτελέσματα των δικτύων πάνω στις τέσσερις μετρικές. Να σημειωθεί ότι τα δίκτυα εκπαιδεύτηκαν με ~13000 frames, και αξιολογήθηκαν με ~2000 frames διαφόρων κλάσεων, σε δέκα κλάσεις δραστηριότητας στο σύνολο. Η μεγάλη διαφορά στο μέγεθος των frames με την προηγούμενη μέθοδο οφείλεται στο ότι η τεχνική Early Fusion απαιτεί πολύ περισσότερη υπολογιστική ισχύ με αποτέλεσμα το σύστημα πάνω από ένα σύνολο εικόνων να σταματάει να λειτουργεί.

Overall Results (RGB-D)	**LSTM**	**GRU**
Accuracy	0.87	0.95
Recall	0.73	0.95
Precision	0.79	0.95
F_1	0.75	0.94

Πίνακας 4.3-6: : Αναπαράσταση των συνολικών αποδόσεων των δικτύων μετά την εφαρμογή της μεθόδου Early Fusion.

Γίνεται κατανοητό πως τα αποτελέσματα των δύο δικτύων είναι αρκετά ενθαρρυντικά, αν ληφθεί υπόψιν ο περιορισμός στο σύνολο δεδομένων. Επιπλέον, η σύγκριση των μετρικών αυτών με της προηγούμενης τεχνικής είναι αρκετά όμοια, ιδιαίτερα στην εκπαίδευση και αξιολόγηση των GRU δικτύων. Τέλος, παρατηρείται ότι τα GRU για άλλη μια φορά έχουν καλύτερη απόδοση από τα LSTM, γεγονός το οποίο οφείλεται κυρίως στην αρχιτεκτονική και στον καλύτερο τρόπο που ρυθμίζουν τη ροή των πληροφοριών εντός της μονάδας.

Στη συνέχεια, προκειμένου να υπάρξει μια πιο ολοκληρωμένη άποψη, καθώς και σύγκριση μεταξύ των τεχνικών και των δραστηριοτήτων που αναγνωρίζει το κάθε δίκτυο, παρατίθενται οι πίνακες σύγχυσης των δύο δικτύων για τη μέθοδο Early Fusion.



Εικόνα 4.3-10: Πίνακες σύγχυσης με τα αποτελέσματα κατηγοριοποίησης για κάθε εκτέλεση δικτύου πάνω στα σετ αξιολόγησης για την Early Fusion μέθοδο.

Όπως ήταν αναμενόμενο τα αποτελέσματα των πινάκων είναι αρκετά διαφοροποιημένα ως προς αυτά της προηγούμενης τεχνικής, λόγω του περιορισμένου αριθμού δεδομένων, τα οποία διατίθενται για εκπαίδευση. Παρόλα αυτά, παρατηρείται καλή απόδοση των δικτύων στις περισσότερες κλάσεις. Αξίζει να προστεθεί ότι οι περισσότερες απώλειες για το δίκτυο GRU παρατηρούνται στην κλάση ’’Λήψη Φαρμάκου’’, ενώ για το δίκτυο LSTM παρατηρούνται στην κλάση ‘’Παραλαβή Αντικειμένου’’ αφού αποτελεί την κλάση με τα λιγότερα διαθέσιμα δεδομένα. Αντίθετα υπάρχουν κλάσεις και στα δύο δίκτυα, τα οποία έχουν αρκετά μεγάλη ακρίβεια κοντά στο 1.0 (αν όχι 1.0).

Κεφάλαιο 5 – Συμπεράσματα - Επίλογος

Η αναγνώριση της ανθρώπινης δραστηριότητας αποτελεί ένα από τα πιο καυτά θέματα στον τομέα της μηχανικής μάθησης και κατά επέκταση της τεχνητής νοημοσύνης στο σύνολο, καθώς σχετίζεται με πάρα πολλές εφαρμογές οι οποίες αφορούν την αλληλεπίδραση ανθρώπου – υπολογιστή. Πέρα τον συνήθη εμποδίων που αντιμετωπίζονται σε τέτοια προβλήματα, η ανθρώπινη αναγνώριση δραστηριότητας εισάγει, επίσης, τον παράγοντα άνθρωπο, γεγονός το οποίο καθιστά το πρόβλημα ακόμα πιο σύνθετο, καθώς κάθε άτομο που συμμετέχει σε μια τέτοια διαδικασία διαθέτει διαφορετικά χαρακτηριστικά, είτε σωματικά, είτε αναφορικά με τον τρόπο τον οποίο εκτελεί μια δραστηριότητα.

Στην παρούσα διπλωματική, παρουσιάστηκαν τα αποτελέσματα από δύο προσεγγίσεις ύστερα από την εφαρμογή τεχνικών βαθιάς μάθησης στα εκάστοτε σύνολα δεδομένων. Πιο συγκεκριμένα, και στις δύο προσεγγίσεις αναπτύχθηκαν τριών ειδών αναδρομικά δίκτυα (LSTM, RNN & GRU) και σε κάποιες περιπτώσεις συνδυάστηκαν με συνελικτικά. Τα εν λόγω είδη μοντέλων έχουν αποδειχθεί ισχυρά σε μια πληθώρα διαφορετικών προβλημάτων, καθώς και τα πιο κατάλληλα για δεδομένα που έχουν ακολουθιακές εξαρτήσεις.

Υλοποιώντας, αρχικά, τα δίκτυα για την προσέγγιση του συνόλου δεδομένων Smartphone διαπιστώνεται ότι τα GRU δίκτυα έχουν την καλύτερη απόδοση στο συγκεκριμένο σύνολο δεδομένων με ποσοστό ~95%, σε αντίθεση με τα RNN, τα οποία έχουν λιγότερη καλή απόδοση. Επιπροσθέτως, έγινε αντιληπτό η πολύ παραμετρικότητα στο πρόβλημα της ανθρώπινης αναγνώρισης δραστηριότητας, καθώς πολλές δραστηριότητες οι οποίες μεταξύ τους είναι πιθανό να έμοιαζαν, το δίκτυο δεν ήταν εύκολο να τις ξεχωρίσει, με αποτέλεσμα πολλές φορές να τις κατηγοριοποιεί σε λάθος κλάση. Κρίνεται αναγκαίο να τεθεί ένα παράδειγμα, πολλές φορές έγινε λάθος κατηγοριοποίηση μεταξύ των κλάσεων ‘’Κατέβασμα Σκάλας’’ και ‘’Ανέβασμα Σκάλας’’, τα οποία φαινομενικά στο ανθρώπινο μάτι αποτελούν δύο διαφορετικές δραστηριότητες, ωστόσο για το σύστημα δεν είναι τόσο διαφορετικές δραστηριότητες. Ένα ακόμα σημαντικό χαρακτηριστικό, το οποίο έρχεται να επιβεβαιώσει την πολύ παραμετρικότητα του προβλήματος, αποτέλεσε η λανθασμένη οριοθέτηση των δραστηριοτήτων. Πιο συγκεκριμένα, υπήρξαν πολλές περιπτώσεις κατά τις οποίες οι χρονικές οριοθετήσεις μεταξύ των δραστηριοτήτων μπορεί να διέφεραν σε μικρό βαθμό, με αποτέλεσμα το εκάστοτε δίκτυο να κατηγοριοποιεί λανθασμένα το κάθε frame και κατά επέκταση συνολικά όλη την δραστηριότητα. Γενικότερα, οι αποδόσεις και οι τιμές στις μετρικές των τριών διαφορετικών δικτύων ήταν αρκετά ικανοποιητικές λαμβάνοντας υπόψιν την ομοιομορφία κάποιων κλάσεων, καθώς και τις λανθασμένες οριοθετήσεις, οι οποίες υπήρχαν εξ’ αρχής στο σύνολο δεδομένων.

Αναφορικά με την προσέγγιση του συνόλου δεδομένων, η οποία βασίζεται σε πολυτροπικά δεδομένα, στο σύνολο δεδομένων εν ονόματι CAD-120. Σε αρχικό στάδιο, γίνεται κατανοητό πως διαφέρει αρκετά από την προηγούμενη προσέγγιση ως προς την υλοποίηση της. Στη συγκεκριμένη προσέγγιση χρησιμοποιήθηκαν δύο πολύ γνωστοί μέθοδοι σύμμειξης των πολυτροπικών δεδομένων. Η πρώτη μέθοδος, γνωστή ως Late Fusion, οδήγησε στη διαπίστωση ότι τα RNN δίκτυα έχουν την καλύτερη απόδοση με ποσοστά, τα οποία αγγίζουν το 98%, ενώ τα GRU και LSTM ακολουθούν με εξίσου καλές τιμές. Επιπλέον έγινε αντιληπτή, η χρησιμότητα της συνένωσης των αποτελεσμάτων, μιας και τα αποτελέσματα ξεχωριστά για κάθε εκπαίδευση ήταν πολύ χαμηλότερα από το τελικά. Ακόμη, αξίζει να σημειωθεί, ότι τα δίκτυα, ενώ εκπαιδεύτηκαν σε ένα μεγάλο σύνολο δεδομένων, με διαφορετικά άτομα να εκτελούν την ίδια δραστηριότητα πολλές φορές, καθώς και με διαφορετικές γωνίες ή λήψης, τα αποτελέσματα και οι τιμές των μετρικών απόδοσης είναι αρκετά ενθαρρυντικές.

Όσον αφορά τη δεύτερη μέθοδο, γνωστή ως Early Fusion, από την πρώτη στιγμή έγινε αντιληπτό πόσο απαιτητική και ιδιόμορφη είναι για να πραγματοποιηθεί. Αναλυτικότερα, η διαδικασία, η οποία απαιτείται προκειμένου να εφαρμοσθεί αυτή η μέθοδος τόσο στο σύνολο δεδομένων, καθώς και στην εκπαίδευση του δικτύου αργότερα, απαιτεί τεράστια υπολογιστική μνήμη. Ενδεικτικό των απαιτήσεων αυτών, αποτελεί το γεγονός ότι για να εκπαιδευτεί ένα δίκτυο με το συνολικό σύνολο δεδομένων, που πραγματοποιήθηκε στην προηγούμενη προσέγγιση, χρειάζεται τουλάχιστον 32GB μνήμη RAM. Παρόλα αυτά, ύστερα από περιορισμούς του συνόλου δεδομένων, τέτοιους ώστε να βρεθεί η χρυσή τομή μεταξύ μη κατάρρευσης του συστήματος αλλά και σωστής εκπαίδευσης, παρατηρήθηκε ότι το δίκτυο αυτό με την καλύτερη απόδοση ήταν το GRU με ποσοστά της τάξεως του 95%. Επιπλέον, αναφορικά με τις δύο μεθόδους σύμμειξης, παρατηρήθηκε πως πολύ σημαντικό ρόλο στην απόδοση των δικτύων παίζει η ποικιλία και η πληθώρα των εικόνων. Χαρακτηριστικό παράδειγμα αποτελεί η κλάση ‘’Επιλογή Αντικειμένου’’ όπου κοιτάζοντας τις ανακατανομές των δειγμάτων με τις κλάσεις, ήταν η κλάση με τα λιγότερα δείγματα. Αυτό το γεγονός είχε εμφανή αποτελέσματα στους πίνακες σύγχυσης, καθώς η κατηγοριοποίηση σε αυτή την κλάση ήταν αρκετά δύσκολη, λόγω του περιορισμού των δεδομένων.

Το πεδίο της αναγνώρισης ανθρώπινης δραστηριότητας έχει εξαιρετικό ενδιαφέρον και οι εφαρμογές του έχουν πολλές προοπτικές για βελτίωση και περαιτέρω ανάπτυξη των δυνατοτήτων τους. Η συνεχή εξέλιξη των ενσωματωμένων αισθητήρων στα Smartphones, καθώς και η τεχνολογική εξέλιξη των οπτικών μέσων πλέον παίζουν πολύ σημαντικό ρόλο. Αντίστοιχα η έρευνα και ανάπτυξη νέων τεχνικών στη βαθιά μάθηση και γενικότερα στην τεχνητή νοημοσύνη αποτελεί έναν εξίσου σημαντικό παράγοντα δημιουργίας βελτιωμένων τέτοιων συστημάτων. Υπάρχουν, βέβαια, ακόμα πολλές πτυχές βελτίωσης όλων αυτών των συστημάτων, αλλά τα ήδη υπάρχοντα αποτελέσματα αποτελούν ο προθάλαμος για το τι έπεται.

Βιβλιογραφία

Adil, M. K. (2011). *Human Activity Recognition Using A Single Tri-axial Accelometer.* Seoul .

Aggarwal, J. K., & Xia, L. (2014). *Human activity recognition from 3D data: A review.* Pattern Recognition Letters.

Álvarez, M. Á., Soria Morillo, L. M., & Álvarez García, J. A. (2017). Mobile activity recognition and fall detection. *Pervasive and Mobile*, (σσ. 3-13).

Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., & Reyes Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. i6doc.

Annalisa, F., Magnani, A., & Maio, D. (2020). on, A multimodal approach for human activity recognition based. *Pattern Recogniton Letters*, σσ. 293-299.

Bakalos, N. et al. (2019). Protecting Water Infrastructure From Cyber and Physical Threats: Using Multimodal Data Fusion and Adaptive Deep Learning to Monitor Critical Systems. *IEEE Signal Processing Magazine*, *36* (2), (σσ. 36-48), doi: 10.1109/MSP.2018.2885359.

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. *Pervasive Computing* (σσ. 158–175). Heidelberg: Springer Berlin.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE transactions on neural networks* (σσ. 157–166). IEEE.

Bulbul, E., Cetin, A., & Dogru, I. (2018). Human Activity Recognition Using Smartphones. Turkey: IEEE.

Chaaraoui, A., Climent-Perez, P., & Florez-Revuelta, F. (2013). *Silhouete-based human action recognition using sequences of key poses.* Pattern Recognition Letters.

Charmi, J., Jatna, B., & Nishant, D. (2019). Human Activity Recognition : A Survey. *2nd International Workshop on Recent advances on IOT: Technology and Application Approaches* (σ. 6). Halifax , Canada: ScienceDirect.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling.*

Cippitelli, E., Gambi, E., & S., S. (2017). Human Action Recognition with RGB-D Sensors. Στο M. ,.-G. Carlos, *Motion Tracking and Gesture Recognition.* doi:10.5772/68121

Ebersmach, M., Herms, R., & Eibl, M. (2017). Fusion Methods for ICD10 Code Classification of Death Certificates in Multilingual Corpora. *Conference and Labs of the Evaluation Forum (CLEF) 2017.* Chemnitz.

Frank, R. J., Davey, N., & Hunt, S. P. (2001). Time Series Prediction and Neural. *Intelligent & Robotic Systems,*, 91–103.

Gers, F. A., & Cummins, F. (2000). Learning to forget: Continual Prediction with LSTM. *Neural computation*, (σσ. 2451–2471).

Guiry, J. J., Ven, P. v., Nelson, J., Warmerdam, L., & Riper, H. (2014). Activity recognition with smartphone support. *Medical Engineering and Physics*, (σσ. 670–675).

Guyon, I. (1997). *A Scaling Law for the Validation-Set Training-Set Size.* AT & T Bell Laboratories.

Hammerla, N. Y., Fisher, J., Andras, P., Rochester, L., Walker, R., & Plötz, T. (2015). Disease state assessment in naturalistic environments using deep learning. *Twenty-Ninth AAAI.*

Haykin, S. (1999). Neural networks : a comprehensive foundation. Second edition. Delhi: Pearson Education.

Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition: A survey. *Image Vision Comput.*

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural computation*, (σσ. 1735–1780).

Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). *Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies.* Ανάκτηση από http://www.bioinf.jku.at/publications/older/ch7.pdf

Hossain, H. S., K. M., & Roy, N. (2016). Active learning enabled activity recognition. *Pervasive and Mobile Computing.*

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of. *Journal of Physiology*, 215-243.

Jin, Q., Zhangjing, W., Xiancheng, L., & Chunming, L. (2018). Learning Complex Spatio-Temporal Configurations of Body Joints for Online Activity Recognition. *EEE Transactions on Human-Machine Systems* (σσ. 637 - 647). IEEE. doi:10.1109/THMS.2018.2850301

Kaixuan, C., Dalin, Z., Lina, Y., Bin, G., Zhiwn, Y., & Yunhao, L. (2018, August). Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges and Opportunities. *J. ACM*.

Kong, Y., & Fu, Y. (2018). Human Action Recognition and Prediction: A Survey. *IEEE.*

Lalos, C., Voulodimos, A., Doulamis, A. et al. (2014). Efficient tracking using a robust motion estimation technique. *Multimedia Tools and Applications*, 69, (σσ. 277–292). https://doi.org/10.1007/s11042-012-0994-3

Lane, N. D., & Georgiev, P. (2015). Can deep learning revolutionize mobile sensing? *16th International Workshop on Mobile Computing Systems and Applications*, (σσ. 117–122).

Laptev, I. (2004). On Space-Time Interest Points. *International Journal of Computer Vision*, σσ. 107–123. doi:https://doi.org/10.1007/s11263-005-1838-7

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *IEEE 86*, (σσ. 499–514).

Li, W., Zhang, Z., & Liu, Z. (2010). Action Recognition Based on a Bag of 3d Points. *Computer Vision and Pattern Recognition Workshops (CVPRW)* (σσ. 9-14). San Francisco: IEEE.

Lu, Y., Wei, Y., Liu, L., Zhong, J., Sun, L., & Liu, Y. (2017). Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimed. Tools Appl.*, (σσ. 10701–10719)

Martín, H., Bernardos, A. M., Iglesias, J., & Casar, J. R. (2013). Activity logging using lightweight classification techniques in mobile devices. *Personal and Ubiquitous Computing*, (σσ. 675–695).

Mitchell, T. M. (1997). *Machine Learning.* McGraw Hill.

Noury, N., Dittmar, A., Corroy, C., Baghai, R., Weber, D. B., Klefstat, F., & Blinovska, A. (2004). Vtamna smart clothe for ambulatory remote monitoring of physiological parameters and activity. *26th Annual IEEE International Conference.*

Oreifej, O., & Liu, Z. (2013). HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences. *Computer Vision and Pattern Recognition*, (σσ. 716-723). Portland.

Oscar, D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials 15*, (σσ. 1192–1209).

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the Difficulty of Training Recurrent Neural Networks. *ICML*, (σσ. 1310–1318).

Rahmani, H., Mahmood, A., Huynh, D., & Mian, A. (2014). Histogram of Oriented Principal Components of 3D Pointclouds for Action Recognition. *Computer Vision - ECCV 2014*, σσ. 742-757.

Reiss, A., Hendeby, G., & Stricker, D. (2015). A novel confidence-based multiclass boosting algorithm for mobile physical activity monitoring. *Personal and Ubiquitous Computing*, (σσ. 105–121).

Ronao, C. A., & Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications*, (σσ. 235–244).

Santiago, J., Cotto, E., Jaimes, L. G., & I., V.-L. (2017). Fall detection system for the elderly. *IEEE 7th Annual Computing and.*

Sarika, P. K. (2018). Comparing LSTM and GRU for Multiclass Sentiment Analysis of Movie Reviews. Karlskrona.

Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition.*

Singh, A. (2017). Anomaly Detection for Temporal Data using LSTM. EINDHOVEN

.Statista. (2021). *Statista*. Ανάκτηση από Statista: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Sung, J., Ponce, C., Selman, B., & Saxena, A. (2011). Human Activity Detection from RGBD Images Recognition (PAIR). *Activity and Intent.* AAAI.

Tapia, E. M., Intille, S. S., & Larson, K. (2004). Activity Recognition in the home using simple and ubiquitous sensors. *Pervasive Computing* (σσ. 158-175). Heidelberg: Springer Berlin.

Van Kasteren, T., Noylas, A., Englebienne, G., & Krose, B. (2008). Accurate Activity Recognition in a home setting". *Proceedings of 10th international conference on Ubiquitous computing* (σσ. 1-9). New York: ACM.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory.* Springer , New York.

Vieira, A., Nascimento, E., Oliveira, G., Liu, Z., & Campos, M. (2012). STOP: Space-Time Occupancy Patterns for 3D Action Recognition from Depth Map Sequences. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications.* , σσ. 255-259.

Vojt, B. J. (2016). Deep neural networks and their implementation. Πράγα.

Voulodimos, A., Doulamis, N., Doulamis, A., & Lalos, C. (2016). Human Tracking Driven Activity Recognition in Video Streams. *Imaging Systems and Techniques (IST) 2016* (σσ. 554-559). Chania: IEEE .

Voulodimos, A., et al. (2011). Online classification of visual tasks for industrial workflow monitoring. *Neural Networks*, 24(8), 2011, σσ. 852-860, https://doi.org/10.1016/j.neunet.2011.06.001.

Werbos, P. J. (1990). Backpropagation Through Time: What It Does and How to Do. *IEEE*, (σσ. 1550–1560).

Woditsch, S. (2017). Classification of multi-spectral and multi-temporal. Munster.

Woznowski, P., Kaleshi, D., Oikonomou, G., & Craddock, I. (2016). Classification and suitability of sensing technologies for activity recognition. *Computer Communications*, (σσ. 89–90 ; 34-50).

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., . . . Steinberg, D. (2007). *Top 10 Algorithms in Data Mining, Knowledge and Information Systems.* Springer.

Zhang, J., Li, W., Ogunbona, P. O., Wanga, P., & Tanga, C. (2016). *RGB-D-based Action Recognition Datasets: A Survey.* doi:10.1016/j.patcog.2016.05.019

Αγαπητός, Α. (2018). Αξιολόγηση Μοντέλων Αναγνώρισης Δραστηριότητας μέσω κινητών συσκευών. Θεσσαλονική.

Γουδέλη, Γ. (2018). *Έυρωστη αναγνώριση ανθρώπινης δραστηριότητας σε ρεαλιστικά περιβάλλοντα.* Αθήνα.

Κερατζάκης, Ι. Ε. (2019). Ανάλυση Συναισθήματος από Κείμενα με Χρήση Τεχνικών Μηχανικής Μάθησης. Αθήνα.

Μενύχτας, Β. (2019). *Αναγνώριση Ανθρώπινης Δραστηριότητας με χρήση μεθόδων Βαθιάς Μάθησης.* Πάτρα.