



# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

## Δίκτυα Αραιής Ενεργοποίησης: Μια νέα μέθοδος αποσύνθεσης και συμπίεσης δεδομένων

Διδακτορική Διατριβή

ΜΠΙΖΟΠΟΥΛΟΥ ΠΑΣΧΑΛΗ

Επιβλέπων : Δημήτριος Κουτσούρης  
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΪΑΤΡΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

Αθήνα, Οκτώβριος 2019





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Συστημάτων Μετάδοσης Πληροφορίας  
και Τεχνολογίας Υλικών  
Εργαστήριο Βιοϊατρικής Τεχνολογίας

## Δίκτυα Αραιής Ενεργοποίησης: Μια νέα μέθοδος αποσύνθεσης και συμπύεσης δεδομένων

Διδακτορική Διατριβή  
ΜΠΙΖΟΠΟΥΛΟΥ ΠΑΣΧΑΛΗ

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 30η Οκτωβρίου 2019.

.....  
Δημήτριος Κουτσούρης  
Καθηγητής Ε.Μ.Π.

.....  
Δημήτριος Φωτιάδης  
Καθηγητής Παν. Ιωαννίνων

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

.....  
Εμμανουήλ Βαβουρανάκης  
Καθηγητής Ε.Κ.Π.Α.

.....  
Γεώργιος Ματσόπουλος  
Καθηγητής Ε.Μ.Π.

.....  
Θεόδωρος Παπαϊωάννου  
Αν.Καθηγητής Ε.Κ.Π.Α.

.....  
Ανδριάννα Πρέντζα  
Καθηγήτρια Παν. Πειραιώς



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Συστημάτων Μετάδοσης Πληροφορίας  
και Τεχνολογίας Υλικών  
Εργαστήριο Βιοϊατρικής Τεχνολογίας

.....

Μπιζόπουλος Πασχάλης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Α.Π.Θ.

Copyright ©—All rights reserved Μπιζόπουλος Πασχάλης, 2019.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή κ. Κουτσούρη για την δυνατότητα που με έδωσε να ξεκινήσω διδακτορικές σπουδές, την ελευθερία στην ερευνητική διαδικασία αλλά και για τις πολύτιμες επιστημονικές συμβουλές του κατά τη διάρκεια του διδακτορικού. Θέλω επίσης να ευχαριστήσω τον κ. Φωτιάδη για την πολύτιμη βοήθειά του και τις κατατοπιστικές ερευνητικές συμβουλές, ιδίως στα πρώτα χρόνια του διδακτορικού. Θα ήθελα να ευχαριστήσω τα μέλη της επταμελούς επιτροπής και όλα τα μέλη του Εργαστηρίου Βιοϊατρικής Τεχνολογίας για την συμπαράσταση και βοήθεια που προσέφεραν. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου στην οποία οφείλω τα πάντα.



# Περίληψη

Η πρόσφατη βιβλιογραφία σχετικά με τη μη-επιβλεπόμενη μάθηση επικεντρώθηκε στο σχεδιασμό δομών με στόχο την μάθηση χαρακτηριστικών. Αυτό όμως γινόταν χωρίς να ληφθεί υπόψη το μήκος περιγραφής των αναπαραστάσεων, το οποίο είναι ένα άμεσο και αμερόληπτο μέτρο της πολυπλοκότητας του μοντέλου. Στο πλαίσιο της διδακτορικής διατριβής προτείνουμε ένα μέτρο  $\varphi$  το οποίο αξιολογεί μη-επιβλεπόμενα μοντέλα με βάση την ακρίβεια ανακατασκευής και το βαθμό συμπίεσης των εσωτερικών αναπαραστάσεων. Έπειτα παρουσιάζουμε και ορίζουμε δύο συναρτήσεις ενεργοποίησης (Ταυτότητα, ReLU) ως βάσεις αναφοράς και τρεις αραιές συναρτήσεις ενεργοποίησης (Απόλυτα κ-μέγιστα, Δείκτες συγκέντρωσης ακρότατων, Ακρότατα) ως υποψήφιες δομές για την ελαχιστοποίηση του προηγουμένως ορισμένου μέτρου  $\varphi$ . Τέλος προτείνουμε μια νέα αρχιτεκτονική Νευρωνικών Δικτύων, τα **Δίκτυα Αραιής Ενεργοποίησης** (SANs), τα οποία αποτελούνται από πυρήνες με κοινά βάρη που κατά την κωδικοποίηση συνελλίσονται με την είσοδο και στη συνέχεια διέρχονται μέσω μιας συνάρτησης αραιής ενεργοποίησης. Κατά τη διάρκεια της αποκωδικοποίησης, τα ίδια βάρη συνελλίσονται με τον χάρτη αραιής ενεργοποίησης και έπειτα οι μερικές ανακατασκευές από κάθε βάρος αθροίζονται για να ανακατασκευάσουν την είσοδο. Συγκρίνουμε τα SANs χρησιμοποιώντας τις προηγουμένως ορισμένες συναρτήσεις ενεργοποίησης σε ένα σύνολο από βάσεις δεδομένων (15 βάσεις δεδομένων από την Physionet και EEG από την UCI για ταξινόμηση επιληπτικών κρίσεων) και δείχνουμε ότι τα SANs που επιλέγονται με χρήση του  $\varphi$  έχουν αναπαραστάσεις με μικρό μήκος περιγραφής και περιέχουν ερμηνεύσιμους πυρήνες.

## Λέξεις Κλειδιά

νευρωνικά δίκτυα, αυτοκωδικοποιητές, αραιότητα, συμπίεση





# Abstract

Recent literature on unsupervised learning focused on designing structural priors with the aim of learning meaningful features, but without considering the description length of the representations. In this thesis, first we introduce the  $\varphi$  metric that evaluates unsupervised models based on their reconstruction accuracy and the degree of compression of their internal representations. We then present and define two activation functions (Identity, ReLU) as base of reference and three sparse activation functions (top-k absolutes, Extrema-Pool indices, Extrema) as candidate structures that minimize the previously defined metric  $\varphi$ . We lastly present Sparsely Activated Networks (SANs) that consist of kernels with shared weights that, during encoding, are convolved with the input and then passed through a sparse activation function. During decoding, the same weights are convolved with the sparse activation map and subsequently the partial reconstructions from each weight are summed to reconstruct the input. We compare SANs using the five previously defined activation functions on a variety of datasets (Physionet, UCI-epilepsy, MNIST, FMNIST) and show that models that are selected using  $\varphi$  have small description representation length and consist of interpretable kernels.

## Keywords

neural networks, autoencoders, sparsity, compression



# Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
Abstract	5
Περιεχόμενα	9
Λίστα εικόνων	11
Λίστα πινάκων	14
<b>1 Εισαγωγή</b>	<b>15</b>
1.1 Οργάνωση τόμου . . . . .	16
<b>2 Νευρωνικά δίκτυα</b>	<b>17</b>
2.1 Εισαγωγή . . . . .	17
2.2 Επισκόπηση θεωρίας . . . . .	18
2.3 Επισκόπηση αρχιτεκτονικών . . . . .	19
Βιβλιογραφία . . . . .	23
<b>3 Βαθιά μάθηση με δομημένα δεδομένα και σήματα</b>	<b>25</b>
3.1 Εισαγωγή . . . . .	25
3.2 Βαθιά μάθηση με δομημένα δεδομένα . . . . .	26
3.3 Βαθιά μάθηση με σήματα . . . . .	29
3.3.1 Ηλεκτροκαρδιογράφημα . . . . .	30
3.3.2 Φωνοκαρδιογράφημα με χρήση της Physionet 2016 . . . . .	38
3.3.3 Άλλα σήματα . . . . .	40
Βιβλιογραφία . . . . .	43

<b>4 Βαθιά μάθηση με απεικονίσεις</b>	<b>53</b>
4.1 Εισαγωγή . . . . .	53
4.2 Τομογραφία μαγνητικού συντονισμού . . . . .	53
4.2.1 Κατάτμηση LV . . . . .	53
4.2.2 Κατάτμηση LV/RV . . . . .	58
4.2.3 Κατάτμηση της καρδιάς . . . . .	61
4.2.4 Άλλες εφαρμογές . . . . .	62
4.2.5 Συμπεράσματα της χρήσης βαθιάς μάθησης με MRI . . . . .	65
4.3 Fundus . . . . .	65
4.3.1 Κατάτμηση αγγείων . . . . .	65
4.3.2 Ανίχνευση μικροανευρυσμάτων και αιμορραγίας . . . . .	68
4.3.3 Άλλες εφαρμογές . . . . .	69
4.3.4 Συμπεράσματα της χρήσης βαθιάς μάθησης με Fundus . . . . .	71
4.4 Ηλεκτρονική τομογραφία . . . . .	71
4.5 Ηχοκαρδιογράφημα . . . . .	75
4.6 Τομογραφία οπτικής συνοχής . . . . .	77
4.7 Άλλες απεικονιστικές τεχνικές . . . . .	78
4.8 Συμπεράσματα της χρήσης βαθιάς μάθησης με CT, ηχοκαρδιογράφημα, OCT και άλλων απεικονιστικών τεχνικών . . . . .	81
4.9 Συζήτηση και μελλοντικές κατευθύνσεις . . . . .	81
Βιβλιογραφία . . . . .	87
<b>5 Επίπεδα Signal2Image σε βαθιά νευρωνικά δίκτυα για ταξ- ινόμηση EEG</b>	<b>107</b>
5.1 Εισαγωγή . . . . .	107
5.2 Δεδομένα . . . . .	109
5.3 Μέθοδοι . . . . .	109
5.3.1 Ορισμοί . . . . .	109
5.3.2 Signal2Image . . . . .	111
5.4 Ρύθμιση πειραμάτων . . . . .	112
5.5 Αποτελέσματα . . . . .	113
5.6 Συμπεράσματα . . . . .	113
Βιβλιογραφία . . . . .	116
<b>6 Δίκτυα αραιής ενεργοποίησης</b>	<b>119</b>
6.1 Εισαγωγή . . . . .	119
6.2 Μέτρο $\varphi$ . . . . .	122
6.3 Δίκτυα αραιής ενεργοποίησης . . . . .	124
6.3.1 Συναρτήσεις αραιής ενεργοποίησης . . . . .	124

---

6.3.2	Αρχιτεκτονική και εκπαίδευση των SANs . . . . .	127
6.4	Πειράματα . . . . .	130
6.4.1	Σύγκριση του μέτρου $\varphi$ για συναρτήσεις αραιής ενεργοποίησης και διάφορα μεγέθη πυρήνα στην Physionet . . . . .	131
6.4.2	Αξιολόγηση της ανακατασκευής των SANs με χρήση επιβλεπόμενου CNN για αναγνώριση επιληψίας στην UCI . . . . .	132
6.4.3	Αξιολόγηση της ανακατασκευής των SANs με χρήση επιβλεπόμενου FNN στην MNIST και FMNIST . . . . .	137
6.5	Συζήτηση . . . . .	139
	Βιβλιογραφία . . . . .	142
<b>7</b>	<b>Επίλογος</b>	<b>147</b>
	<b>Ακρωνύμια</b>	<b>149</b>
	<b>Κατάλογος Δημοσιεύσεων</b>	<b>151</b>



# Λίστα εικόνων

2.1	Αντιλήπτορ . . . . .	19
2.2	Συνελικτικό νευρωνικό δίκτυο . . . . .	21
5.1	Αρχιτεκτονική Signal2Image . . . . .	110
5.2	Σήματα και έξοδοι των Signal2Image για κάθε κλάση . . . . .	112
6.1	Οπτικοποίηση χαρτών ενεργοποίησης πέντε συναρτήσεων αραιής ενεργοποίησης (Identity, ReLU, top-k absolutes, Extrema-Pool indices και Extrema) για 1D και 2D είσοδο. . . . .	121
6.2	Διαγράμματα της προς-τα-εμπρός διάδοσης ενός 1D και 2D SAN με δύο πυρήνες για παραδείγματα από την βάση δεδομένων επιληψίας UCI και MNIST αντίστοιχα. . . . .	128
6.3	Αντίστροφος λόγος συμπίεσης ( $CR^{-1}$ ) έναντι κανονικοποιημένης απώλειας ανακατασκευής ( $\tilde{\mathcal{L}}$ ) για 15 βάσεις δεδομένων της Physionet και για διάφορα μεγέθη πυρήνα. . . . .	133
6.4	Συνολικά αποτελέσματα της αξιολόγησης των βάσεων δεδομένων της Physionet με τη χρήση του μέτρου $\varphi$ . Το διάγραμμα πυκνότητας δημιουργήθηκε χρησιμοποιώντας εκτίμηση πυκνότητας γκαουσιανών πυρήνων και τα διαστήματα εμπιστοσύνης απεικονίζουν μια τυπική απόκλιση. . . . .	134
6.5	Οπτικοποίηση των πυρήνων για κάθε αραιή συνάρτηση ενεργοποίησης (γραμμή) και για κάθε βάση δεδομένων της Physionet (στήλη). . . . .	135





# Λίστα πινάκων

3.1	Δημόσιες καρδιολογικές βάσεις δεδομένων, δομημένων δεδομένων και σημάτων . . . . .	27
3.2	Εφαρμογές βαθιάς μάθησης με χρήση δομημένων δεδομένων . . . . .	28
3.3	Εφαρμογές βαθιάς μάθησης για ανίχνευση αρρυθμιών με την MITDB . . . . .	31
3.4	Εφαρμογές βαθιάς μάθησης με χρήση ECG για ανίχνευση αρρυθμιών και AF . . . . .	34
3.5	Εφαρμογές βαθιάς μάθησης με χρήση ECG σε άλλες εφαρμογές . . . . .	36
3.6	Εφαρμογές βαθιάς μάθησης με χρήση PCG και άλλων σημάτων . . . . .	42
4.1	Δημόσιες καρδιολογικές βάσεις δεδομένων, MRI . . . . .	54
4.2	Δημόσιες καρδιολογικές βάσεις δεδομένων, Fundus και άλλες . . . . .	55
4.3	Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάτμηση LV . . . . .	56
4.4	Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάτμηση LV/RV . . . . .	60
4.5	Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάτμηση της καρδιάς και άλλα . . . . .	63
4.6	Εφαρμογές βαθιάς μάθησης με χρήση Fundus, για κατάτμηση αγγείων . . . . .	67
4.7	Εφαρμογές βαθιάς μάθησης με χρήση Fundus, εκτός της κατάτμησης αγγείων . . . . .	70
4.8	Εφαρμογές βαθιάς μάθησης με χρήση CT . . . . .	72
4.9	Εφαρμογές βαθιάς μάθησης με χρήση Ηχοκαρδιογραφήματος . . . . .	76
4.10	Εφαρμογές βαθιάς μάθησης με χρήση OCT και άλλων απεικονιστικών τεχνικών . . . . .	80
4.11	Βιβλιογραφικές αναφορές βαθιάς μάθησης με εφαρμογή στην καρδιολογία . . . . .	82
5.1	Ακρίβειες στα δεδομένα δοκιμής (%) για τα συνδυασμένα μοντέλα . . . . .	114
6.1	Μέγεθος πυρήνα $m$ με το καλύτερο $\varphi$ για κάθε συνάρτηση αραιής ενεργοποίησης για κάθε βάση δεδομένων της Physionet . . . . .	136
6.2	SANs με επιβλεπόμενο στοιβαγμένο CNN για αναγνώριση επιληψίας στην UCI . . . . .	138
6.3	SANs με επιβλεπόμενο στοιβαγμένο FNN στην MNIST . . . . .	140



# 1. Εισαγωγή

Οι γιατροί κάνουν διαγνώσεις με βάση το ιατρικό ιστορικό, βιοδείκτες και τις εξετάσεις μεμονωμένων ασθενών, τα οποία ερμηνεύουν σύμφωνα με την προσωπική τους κλινική εμπειρία. Στη συνέχεια, αντιστοιχούν τον κάθε ασθενή στην παραδοσιακή ταξινόμηση των ιατρικών νοσημάτων/παθήσεων σύμφωνα με μια υποκειμενική ερμηνεία της ιατρικής βιβλιογραφίας. Αυτή η διαδικασία είναι όλο και περισσότερο επιρρεπής σε σφάλματα. Επιπλέον, οι τεχνολογίες απεικονίσεων αυξάνουν συνεχώς την ικανότητά τους να παράγουν μεγάλες ποσότητες δεδομένων, τα οποία οι ειδικοί αδυνατούν να καταλάβουν και να χρησιμοποιήσουν αποτελεσματικά, καθιστώντας το έργο τους δυσκολότερο. Επομένως, απαιτείται η αυτοματοποίηση των ιατρικών διαδικασιών για την αύξηση της ποιότητας της υγείας των ασθενών και τη μείωση του κόστους των συστημάτων υγειονομικής περίθαλψης.

Η ανάγκη για αυτοματοποίηση των ιατρικών διαδικασιών κυμαίνεται από τη διάγνωση έως τη θεραπεία και ιδίως σε περιπτώσεις/περιοχές όπου υπάρχει έλλειψη υγειονομικής περίθαλψης. Προηγούμενες προσπάθειες αυτοματοποίησης περιλαμβάνουν συστήματα εμπειρογνομόνων βασισμένα σε κανόνες, τα οποία έχουν σχεδιαστεί να μιμούνται τη διαδικασία που ακολουθούν οι ιατρικοί εμπειρογνώμονες όταν επιλύουν ιατρικά προβλήματα. Αυτά τα συστήματα έχουν αποδειχθεί αναποτελεσματικά επειδή απαιτούν χειροκίνητη δημιουργία χαρακτηριστικών και γνώση ειδικών πάνω στο πρόβλημα για να επιτευχθεί επαρκής ακρίβεια και επίσης είναι δύσκολο να εμφανίσουν γραμμική βελτίωση με την παρουσία νέων δεδομένων.

Η βαθιά μάθηση έχει αναδειχθεί ως μια πιο ακριβής και αποτελεσματική τεχνολογία σε ένα ευρύ φάσμα ιατρικών προβλημάτων όπως η διάγνωση, η πρόβλεψη και η παρέμβαση. Είναι μια μέθοδος μάθησης αναπαραστάσεων που αποτελείται από επίπεδα που μετασχηματίζουν τα δεδομένα με μη-γραμμικό τρόπο, αποκαλύπτοντας έτσι ιεραρχικές σχέσεις και δομές. Παρόλο που η βαθιά μάθηση έχει εφαρμοστεί με επιτυχία σε πολλούς τομείς, αυτό έρχεται με κόστος της ερμηνευσιμότητας των αναπαραστάσεων, καθώς αυτές αποτελούνται από εκατομμύρια παραμέτρους.

Σκοπός της διδακτορικής διατριβής είναι η δημιουργία μιας νέας αρχιτεκτονικής νευρωνικών δικτύων, τα **Δίκτυα Αραιής Ενεργοποίησης** (Sparsely Activated Networks, SANs), τα οποία αποσυνθέτουν τις εισόδους τους σε ένα σύνολο από αραιά

επαναλαμβανόμενα πρότυπα διαφορετικού πλάτους και σε συνδυασμό με ένα προτεινόμενο μέτρο  $\varphi$  μαθαίνουν αναπαραστάσεις με ελάχιστα μήκη περιγραφής, αυξάνοντας έτσι την ερμηνευσιμότητα του μοντέλου.

## 1.1. Οργάνωση τόμου

Στο Κεφάλαιο 2, παρουσιάζουμε τις θεμελιώδεις έννοιες των νευρωνικών δικτύων και της βαθιάς μάθησης καθώς επίσης και τις γενικές ιδιότητες των πιο συχνά χρησιμοποιούμενων αρχιτεκτονικών.

Στα Κεφάλαια 3 και 4 παρουσιάζουμε μια βιβλιογραφική επισκόπηση των αρχιτεκτονικών βαθιάς μάθησης σε δομημένα δεδομένα, σήματα και απεικονίσεις που έχουν χρησιμοποιηθεί στην ιατρική και ιδιαίτερα στην καρδιολογία. Συζητάμε τα πλεονεκτήματα και τους περιορισμούς των εφαρμογών της βαθιάς μάθησης στην ιατρική, ενώ προτείνουμε ορισμένες κατευθύνσεις ως τις πιο βιώσιμες για κλινική χρήση. Πιο συγκεκριμένα στις ενότητες 3.2 και 3.3 παρουσιάζουμε τις εφαρμογές βαθιάς μάθησης με χρήση δομημένων δεδομένων και μορφές σημάτων, ενώ στην ενότητα 4.9 παρουσιάζουμε πλεονεκτήματα και περιορισμούς των εφαρμογών βαθιάς μάθησης στην καρδιολογία και προτείνουμε κάποιες κατευθύνσεις για την υλοποίηση μοντέλων βαθιάς μάθησης που μπορούν να εφαρμοστούν κλινικά.

Στο Κεφάλαιο 5 προτείνουμε τα Signal2Image (S2Is) ως εκπαιδευσιμα επίπεδα προθέματος νευρωνικών δικτύων, τα οποία μετατρέπουν σήματα, όπως το ηλεκτροεγκεφαλογράφημα (Electroencephalogram, EEG), σε αναπαραστάσεις εικόνων, καθιστώντας τα κατάλληλα για την εκπαίδευση βαθιών νευρωνικών δικτύων βασισμένα σε εικόνες, τα οποία ορίζονται ως ‘μοντέλα βάσης’. Συγκρίνουμε την ακρίβεια και τις επιδόσεις τεσσάρων S2Is (‘σήμα ως εικόνα’, φασματογράφημα, CNN ενός και δύο επιπέδων) σε συνδυασμό με ένα σύνολο ‘μοντέλων βάσης’ μαζί με μοντέλα διαφορετικού βάθους και 1D παραλλαγές των τελευταίων. Παρέχουμε επίσης εμπειρικές αποδείξεις ότι το CNN S2I ενός επιπέδου αποδίδει καλύτερα σε 11 από τα 15 μοντέλα που δοκιμάστηκαν σε σύγκριση με τα μη-εκπαιδευσιμα S2Is για την ταξινόμηση σημάτων EEG και οπτικοποιούμε τις εξόδους κάποιων από τα S2Is.

Στο Κεφάλαιο 6, ενότητα 6.2 ορίζουμε το μέτρο  $\varphi$ , έπειτα στην ενότητα 6.3 ορίζουμε τις πέντε συναρτήσεις ενεργοποίησης που θα συγκριθούν καθώς επίσης και την αρχιτεκτονική και διαδικασία εκπαίδευσης των SAN, στην ενότητα 6.4 δοκιμάζουμε τα SANs στις βάσεις δεδομένων Physionet, UCI-epilepsy και MNIST και οπτικοποιούμε τις ενδιάμεσες αναπαραστάσεις και αποτελέσματα. Στην ενότητα 6.5 συζητάμε τα ευρήματα των πειραμάτων και τους περιορισμούς των SAN.

Τέλος στο Κεφάλαιο 7 παρουσιάζουμε τα τελικά συμπεράσματα και προτείνουμε πιθανές μελλοντικές κατευθύνσεις.

## 2. Νευρωνικά δίκτυα

### 2.1. Εισαγωγή

Η μηχανική μάθηση είναι ένα σύνολο μεθόδων Τεχνητής Νοημοσύνης (Artificial Intelligence, AI) που επιτρέπει στους υπολογιστές να μαθαίνουν μια διαδικασία χρησιμοποιώντας δεδομένα, αντί να προγραμματίζονται ρητά. Έχει αναδειχθεί ως ένας αποτελεσματικός τρόπος χρήσης και συνδυασμού βιολογικών δεικτών, απεικόνισης, συσσωρευμένης κλινικής έρευνας από τη βιβλιογραφία και τους Ηλεκτρονικούς Φακέλους Υγείας (Electronic Health Record, EHR) για την αύξηση της ακρίβειας λύσεων ενός ευρύ φάσματος ιατρικών προβλημάτων. Οι ιατρικές διαδικασίες που χρησιμοποιούν την μηχανική μάθηση εξελίσσονται από τέχνη σε επιστήμη με γνώμονα τα δεδομένα, προσφέροντας διόραση από τα πληθυσμιακά δεδομένα στην προσωποποιημένη ιατρική.

Η βαθιά μάθηση, και η εφαρμογή της στα νευρωνικά δίκτυα τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks, DNN), είναι ένα σύνολο μεθόδων μηχανικής μάθησης που αποτελούνται από πολλαπλά στοιβαγμένα επίπεδα και χρησιμοποιούν δεδομένα για να μάθουν ιεραρχικά επίπεδα. Η βαθιά μάθηση προέκυψε λόγω της αύξησης της υπολογιστικής ισχύος των μονάδων επεξεργασίας γραφικών και της διαθεσιμότητας δεδομένων μεγάλου όγκου και έχει αποδειχθεί ότι είναι μια ισχυρή λύση για προβλήματα όπως ταξινόμηση εικόνων [1], κατάτμηση εικόνας [2], επεξεργασία φυσικής γλώσσας [3], αναγνώριση ομιλίας [4] και γονιδιωματική [5].

Πλεονεκτήματα των DNN έναντι των παραδοσιακών τεχνικών μηχανικής μάθησης περιλαμβάνουν ότι απαιτούν λιγότερη εξειδικευμένη γνώση για το πρόβλημα που προσπαθούν να λύσουν και επίσης είναι ευκολότερο να αυξήσουν την ακρίβεια τους είτε με την αύξηση του συνόλου δεδομένων εκπαίδευσης ή την αύξηση της χωρητικότητας του δικτύου. Τα ρητά μοντέλα μηχανικής μάθησης, όπως τα δέντρα αποφάσεων και οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVM), είναι 'ανεπαρκή'; το οποίο σημαίνει ότι απαιτούν μεγάλο αριθμό υπολογισμών κατά τη διάρκεια της εκπαίδευσης/συμπερασμού, μεγάλο αριθμό παρατηρήσεων για την επίτευξη γενίκευσης και σημαντική ανθρώπινη εργασία για τον προσδιορισμό της πρότερης γνώσης του μοντέλου [6].

## 2.2. Επισκόπηση θεωρίας

Τα νευρωνικά δίκτυα είναι ένα σύνολο τεχνικών μηχανικής μάθησης εμπνευσμένων από τον εγκέφαλο αλλά χωρίς πρωταρχικό στόχο να τον προσομοιώνουν. Πρόκειται για μεθόδους προσέγγισης συναρτήσεων όπου η είσοδος  $\mathbf{x}$  μπορεί να είναι κείμενο, εικόνα, ήχος, σήμα, τρισδιάστατος όγκος, βίντεο (ή συνδυασμός αυτών) και η έξοδος  $\mathbf{y}$  είναι από το ίδιο σετ με το  $\mathbf{x}$  αλλά αυξημένου ενημερωτικού περιεχομένου. Με μαθηματικούς όρους, ο στόχος ενός νευρωνικού δικτύου είναι να βρει το σύνολο παραμέτρων  $\theta$  (το οποίο αποτελείται από βάρη  $\mathbf{w}$  και πολώσεις  $\mathbf{b}$ ):

$$f(\mathbf{x}; \theta) = \hat{\mathbf{y}} \quad (2.1)$$

όπου  $f$  είναι μια προκαθορισμένη συνάρτηση και  $\hat{\mathbf{y}}$  είναι η πρόβλεψη. Ο περιορισμός για το  $\theta$  είναι να έχουμε ένα όσο το δυνατόν χαμηλότερη τιμή για μια συνάρτηση κόστους  $J(\theta)$  μεταξύ του  $\mathbf{y}$  και του  $\hat{\mathbf{y}}$ .

Η βασική μονάδα των νευρωνικών δικτύων είναι ο αντιλήπτορ που απεικονίζεται στην Εικ.2.1, ο οποίος δημοσιεύθηκε για πρώτη φορά από τον Rosenblatt [7] το 1958. Αποτελείται από ένα σύνολο εισόδων που συμβολίζεται με το διάνυσμα  $\mathbf{x} = [x_1, \dots, x_j, \dots, x_n]$ , ένα σύνολο βαρών για την κάθε είσοδο, που συμβολίζεται με το διάνυσμα  $\mathbf{w} = [w_1, \dots, w_j, \dots, w_n]$  και η πόλωση  $b$ . Η απόφαση του κόμβου να πυροδοτήσει ένα σήμα  $\alpha$  στον επόμενο νευρώνα ή στην έξοδο καθορίζεται από τη συνάρτηση ενεργοποίησης  $\phi$ , το σταθμισμένο άθροισμα  $\mathbf{w}$  και τη πόλωση  $b$ , με τον ακόλουθο τρόπο:

$$\alpha = \phi\left(\sum_j w_j x_j + b\right) \quad (2.2)$$

Οι τιμές του διανύσματος βαρών  $\mathbf{w}$  και των πολώσεων  $\mathbf{b}$  στα DNN ρυθμίζονται επαναληπτικά χρησιμοποιώντας αλγόριθμους βελτιστοποίησης που βασίζονται στην κλίση καθόδου και την αντίστροφη διάδοση σφάλματος (backpropagation) [8] ο οποίος υπολογίζει την κλίση της συνάρτησης κόστους σε σχέση με τις παραμέτρους  $\nabla_{\theta} J(\theta)$  [9]. Η αξιολόγηση της γενίκευσης ενός νευρωνικού δικτύου απαιτεί τη διάσπαση του αρχικού συνόλου δεδομένων  $D = (\mathbf{x}, \mathbf{y})$  σε τρία μη-αλληλεπικαλυπτόμενα σύνολα δεδομένων:

$$D_{train} = (\mathbf{x}_{train}, \mathbf{y}_{train}) \quad (2.3)$$

$$D_{validation} = (\mathbf{x}_{validation}, \mathbf{y}_{validation}) \quad (2.4)$$

$$D_{test} = (\mathbf{x}_{test}, \mathbf{y}_{test}) \quad (2.5)$$

Το  $D_{train}$  χρησιμοποιείται για την μάθηση των  $\mathbf{w}$  και  $\mathbf{b}$ , των οποίων οι τιμές κάνουν το δίκτυο να ελαχιστοποιεί την επιλεγμένη συνάρτηση κόστους, ενώ το  $D_{validation}$

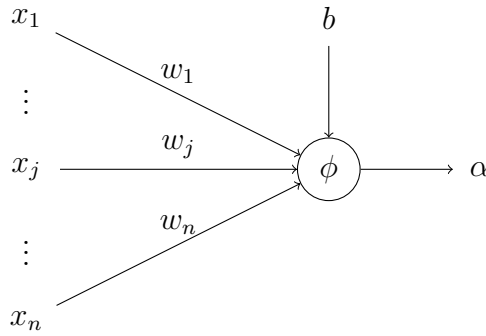


Figure 2.1: Αντιλήπτορ. Περιέχει ένα σύνολο συνδέσεων  $x_1, \dots, x_n$  ως εισόδους, τα βάρη  $w_1, \dots, w_n$ , τη πόλωση  $b$ , τη συνάρτηση ενεργοποίησης  $\phi$  και την έξοδο  $\alpha$ .

χρησιμοποιείται για την επιλογή των υπερπαραμέτρων του δικτύου. Το  $D_{test}$  χρησιμοποιείται για την αξιολόγηση της γενίκευσης του δικτύου και θα πρέπει ιδανικά να προέρχεται από διαφορετικές μηχανές/ασθενείς/οργανισμούς ανάλογα με το ερευνητικό ερώτημα.

Η απόδοση των DNN έχει βελτιωθεί σημαντικά με τη χρήση της ανορθωμένης γραμμικής μονάδας (Rectified Linear Unit, ReLU) ως συνάρτηση ενεργοποίησης σε σύγκριση με την σιγμοειδής και την υπερβολική εφαπτομένη [10]. Η συνάρτηση ενεργοποίησης του τελευταίου επιπέδου επιλέγεται με βάση τη φύση του ερευνητικού ερωτήματος που πρέπει να απαντηθεί (π.χ. softmax για ταξινόμηση, σιγμοειδής για παλινδρόμηση).

Οι συναρτήσεις κόστους που χρησιμοποιούνται στα νευρωνικά δίκτυα εξαρτώνται επίσης από το πρόβλημα που πρέπει να επιλυθεί. Η διασταυρωμένη εντροπία (cross-entropy) ποσοτικοποιεί τη διαφορά μεταξύ της πραγματικής και της προβλεπόμενης κατανομής πιθανοτήτων και επιλέγεται συνήθως για προβλήματα ανίχνευσης και ταξινόμησης. Το εμβαδόν της περιοχής κάτω από την καμπύλη λειτουργίας δέκτη (Area Under Curve, AUC) αντιπροσωπεύει την πιθανότητα ένα τυχαίο ζεύγος κανονικών και μη κανονικών εικονοστοιχείων/σημάτων/εικόνων να ταξινομηθεί σωστά [11] και χρησιμοποιείται σε προβλήματα δυαδικής τμηματοποίησης. Ο συντελεστής Dice [12] είναι ένα μέτρο ομοιότητας που χρησιμοποιείται στα προβλήματα τμηματοποίησης και οι τιμές του κυμαίνονται μεταξύ μηδέν (συνολική αναντιστοιχία) και μονάδας (τέλεια αντιστοιχία).

## 2.3. Επισκόπηση αρχιτεκτονικών

Τα Πλήρως Συνδεδεμένα Δίκτυα (Fully Connected Networks, FNN) είναι δίκτυα που αποτελούνται από πολλαπλούς αντιλήπτορες στοιβαγμένους σε πλάτος και βάθος,

που σημαίνει ότι κάθε αντιλήπτωρ σε κάθε επίπεδο συνδέεται με κάθε αντιλήπτωρ στα επίπεδα αμέσως πριν και μετά. Αν και έχει αποδειχθεί [13] ότι τα FNN ενός επιπέδου με επαρκή αριθμό κρυφών μονάδων είναι καθολικοί προσεγγιστές συναρτήσεων, δεν είναι υπολογιστικά αποδοτικά για την προσέγγιση σύνθετων συναρτήσεων. Τα Δίκτυα Βαθιάς Πίστης (Deep Belief Networks, DBN) [14] είναι στοιβαγμένες Περιορισμένες Μηχανές Boltzmann (Restricted Boltzmann Machines, RBMs) όπου κάθε επίπεδο κωδικοποιεί τις στατιστικές εξαρτήσεις μεταξύ των μονάδων στο προηγούμενο επίπεδο; εκπαιδεύονται για να μεγιστοποιήσουν την πιθανοφάνεια των δεδομένων εκπαίδευσης.

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN), όπως φαίνεται στο Σχήμα 2.2, αποτελούνται από ένα συνελικτικό μέρος στο οποίο γίνεται ιεραρχική εξαγωγή χαρακτηριστικών (χαρακτηριστικά χαμηλού επιπέδου όπως ακμές και γωνίες και χαρακτηριστικά υψηλού επιπέδου, όπως τμήματα αντικειμένων) και ένα πλήρως συνδεδεμένο μέρος για ταξινόμηση ή παλινδρόμηση, ανάλογα με τη φύση της εξόδου  $y$ . Τα συνελικτικά επίπεδα είναι πολύ καλοί βελτιστοποιητές χαρακτηριστικών που χρησιμοποιούν τις τοπικές σχέσεις των δεδομένων, ενώ τα πλήρως συνδεδεμένα επίπεδα είναι καλοί ταξινομητές, επομένως χρησιμοποιούνται ως τα τελευταία επίπεδα ενός CNN. Επιπλέον, τα συνελικτικά επίπεδα δημιουργούν χάρτες ενεργοποιήσεων χρησιμοποιώντας κοινόχρηστα βάρη που έχουν σταθερό αριθμό παραμέτρων σε αντίθεση με τα πλήρως συνδεδεμένα επίπεδα, καθιστώντας τα έτσι πολύ ταχύτερα. Το VGGnet [15] είναι μια απλή CNN αρχιτεκτονική που χρησιμοποιεί μικρά συνελικτικά φίλτρα ( $3 \times 3$ ) και η απόδοση του αυξάνεται απλά αυξάνοντας το βάθος του δικτύου. Το GoogleNet [16] είναι μια άλλη CNN αρχιτεκτονική η οποία χρησιμοποιεί την μονάδα inception. Η μονάδα inception χρησιμοποιεί παράλληλα πολλαπλά επίπεδα περιελίξεων από τα οποία το αποτέλεσμα συσσωρεύεται σειριακά, επιτρέποντας έτσι στο δίκτυο να μαθαίνει χαρακτηριστικά πολλαπλών επιπέδων. Το ResNet [17] είναι μια CNN αρχιτεκτονική η οποία διαμορφώνει τα επίπεδα ως συναρτήσεις μάθησης υπολειμμάτων με βάση τις εισόδους των επιπέδων, επιτρέποντας την δημιουργία πολύ πιο βαθιών δικτύων, σε σχέση με τα προηγούμενως αναφερθέντα.

Οι Αυτο-κωδικοποιητές (Autoencoders, AE) είναι νευρωνικά δίκτυα που εκπαιδεύονται με στόχο να αντιγράψουν την είσοδο στην έξοδο με τέτοιο τρόπο ώστε να κωδικοποιούν χρήσιμες ιδιότητες των δεδομένων. Συνήθως αποτελείται από ένα τμήμα κωδικοποίησης που υποδειγματολειτουργεί την είσοδο μέχρι αυτό να γίνει γραμμικό χαρακτηριστικό και ένα τμήμα αποκωδικοποίησης που υπερδειγματολειτουργεί προς τις αρχικές διαστάσεις. Μια κοινή αρχιτεκτονική AE είναι ο Στοιβαγμένος Αυτο-κωδικοποιητής Αποθρομβοποίησης (Stacked Denoised Autoencoder, SDAE) που έχει ως στόχο την ανακατασκευή της εισόδου από μια τεχνητά αλλοιωμένη έκδοση της [18], η οποία εμποδίζει το μοντέλο να μάθει προφανείς λύσεις. Μια άλλη αρχιτεκτονική τύπου AE είναι το u-net [2], το οποίο παρουσιάζει ιδιαίτερο ενδιαφέρον για τη



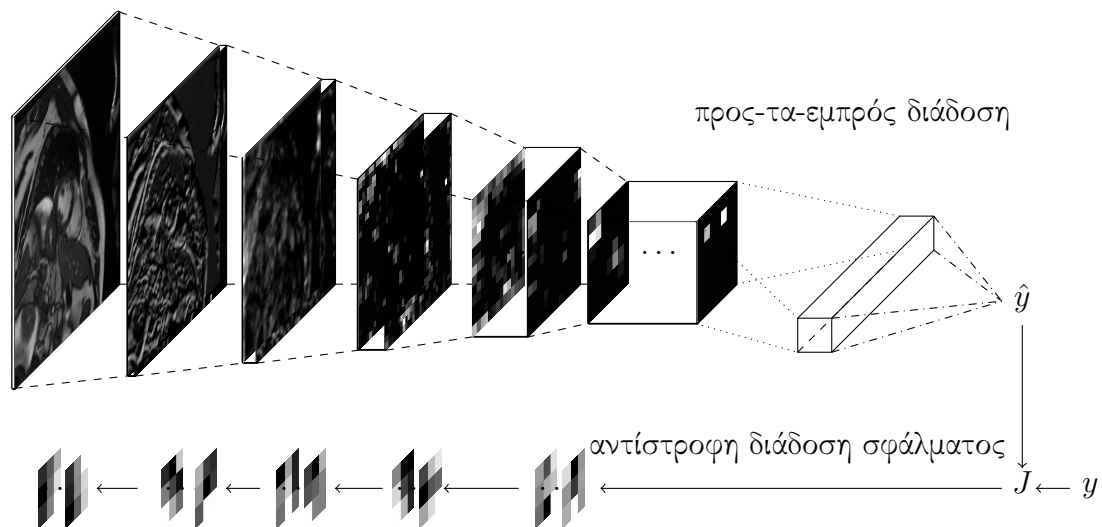


Figure 2.2: Ένα συνελικτικό νευρωνικό δίκτυο που υπολογίζει την εμβαδόν της αριστερής καρδιακής κοιλίας ( $\hat{y}$ ) από μια εικόνα MRI ( $x$ ). Η πυραμιδοειδής δομή στο πάνω μέρος της εικόνας υποδηλώνει τη ροή των υπολογισμών κατά τη διάρκεια της προς-τα-εμπρός διάδοσης, ξεκινώντας από την εικόνα εισόδου μέσω του συνόλου των χαρτών ενεργοποιήσεων που απεικονίζονται ως τρισδιάστατα ορθογώνια στην έξοδο  $\hat{y}$ . Το ύψος και το πλάτος των ορθογωνίων είναι ανάλογο του ύψους και του πλάτους των χαρτών ενεργοποιήσεων, ενώ το βάθος είναι ανάλογο του αριθμού των χαρτών ενεργοποιήσεων. Τα βέλη στο κάτω μέρος υποδηλώνουν τη ροή της αντίστροφης διάδοσης του σφάλματος (backpropagation) ξεκινώντας από τον υπολογισμό του σφάλματος χρησιμοποιώντας τη συνάρτηση κόστους  $J$ , την πραγματική έξοδο  $y$  και την πρόβλεψη  $\hat{y}$ . Αυτή η απώλεια διαδίδεται προς τα πίσω μέσω των φίλτρων του δικτύου, τα οποία προσαρμόζουν τις τιμές τους. Οι διακεκομμένες γραμμές με παύλες υποδηλώνουν ένα 2D συνελικτικό επίπεδο με ReLU και Extrema-Pooling (το οποίο μειώνει το ύψος και το πλάτος των χαρτών ενεργοποιήσεων), η διακεκομμένη γραμμή με τελείες δηλώνει το πλήρως συνδεδεμένο επίπεδο και τέλος οι διακεκομμένες γραμμές με παύλες-τελείες υποδηλώνουν σιγμοειδές επίπεδο. Για ευκολία απεικόνισης εμφανίζονται μόνο μερικοί από τους χάρτες ενεργοποιήσεων και τα φίλτρα και δεν εμφανίζονται σε κλίμακα.

βιοϊατρική κοινότητα καθώς αρχικά εφαρμόστηκε για κατάτμηση βιοϊατρικών εικόνων. Το u-net εισήγαγε τις συνδέσεις παράλειψης (skip connections) που συνδέουν τα επίπεδα του κωδικοποιητή με τα αντίστοιχα επίπεδα του αποκωδικοποιητή.

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Network, RNN) είναι δίκτυα που αποτελούνται από βρόχους ανατροφοδότησης και σε αντίθεση με τις προηγούμενες καθορισμένες αρχιτεκτονικές μπορούν να χρησιμοποιήσουν την εσωτερική τους κατάσταση για να επεξεργαστούν την είσοδο. Τα απλά RNN έχουν το πρόβλημα της εξαφάνισης των κλίσεων (vanishing gradients) και για το λόγο αυτό προτάθηκε

η Μακρο-Βραχυ Πρόθεσμη Μονάδα Μνήμης (Long-Short Term Memory, LSTM) ως λύση για την αποθήκευση πληροφοριών για παρατεταμένο χρόνο. Η Κεκλεισμένη Επαναληπτική Μονάδα (Gated Recurrent Unit, GRU) προτάθηκε αργότερα ως μια απλούστερη εναλλακτική λύση έναντι του LSTM.

## Βιβλιογραφία

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [3] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.
- [4] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on*, pages 6645–6649. IEEE, 2013.
- [5] Babak Alipanahi, Andrew DeLong, Matthew T Weirauch, and Brendan J Frey. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. *Nature biotechnology*, 33(8):831, 2015.
- [6] Yoshua Bengio, Yann LeCun, et al. Scaling learning algorithms towards ai. *Large-scale kernel machines*, 34(5):1–41, 2007.
- [7] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [8] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986.
- [9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*, volume 1. MIT press Cambridge, 2016.
- [10] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [11] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

- 
- [12] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [13] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [14] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [18] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

## 3. Βαθιά μάθηση με δομημένα δεδομένα και σήματα

### 3.1. Εισαγωγή

Οι καρδιαγγειακές παθήσεις (Cardiovascular Diseases, CVDs) είναι η κύρια αιτία θανάτου παγκοσμίως αντιπροσωπεύοντας το 30% των θανάτων το 2014 στις Ηνωμένες Πολιτείες [1], το 45% των θανάτων στην Ευρώπη και εκτιμάται ότι κοστίζουν 210 δισ. Ευρώ ετησίως μόνο για την Ευρωπαϊκή Ένωση [2].

Στα επόμενα δύο κεφάλαια θα παρουσιάσουμε μια συστηματική βιβλιογραφική ανασκόπηση, που έγινε στο πλαίσιο της παρούσας διδακτορικής διατριβής, των εφαρμογών της βαθιάς μάθησης σε δομημένα δεδομένα, σήματα και απεικόνιση από την καρδιολογία, που σχετίζονται με καρδιακές δομές και αγγεία. Η φράση αναζήτησης της βιβλιογραφίας είναι η συνδυασμός καθενός από τους όρους καρδιολογίας και βαθιάς μάθησης που αναφέρονται στα Ακρωνύμια, χρησιμοποιώντας το Google Scholar<sup>1</sup>, το Pubmed<sup>2</sup> και το Scopus<sup>3</sup>. Έπειτα τα αποτελέσματα επιλέγονται έτσι ώστε να ταιριάζουν με τα κριτήρια επιλογής της ανασκόπησης τα οποία συνοψίζονται σε δύο κύριους άξονες: την αρχιτεκτονική του νευρωνικού δικτύου και ο τύπος δεδομένων που χρησιμοποιήθηκε για εκπαίδευση/επικύρωση/δοκιμή. Οι αξιολογήσεις αναφέρονται για περιοχές που χρησιμοποίησαν ένα συνεκτικό σύνολο μετρήσεων με την ίδια μη-τροποποιημένη βάση δεδομένων και το ίδιο ερευνητικό ερώτημα. Δημοσιεύσεις που δεν παρέχουν πληροφορίες σχετικά με την αρχιτεκτονική του νευρωνικού δικτύου ή δημοσιεύσεις που απλά αναπαράγουν μεθόδους προηγούμενων δημοσιεύσεων ή προκαταρκτικές δημοσιεύσεις αποκλείονται από αυτήν την ανασκόπηση. Όταν αναφέρεται η λέξη *πολλαπλά* στη στήλη 'Αποτελέσματα', αυτά αναφέρονται στο κυρίως κείμενο όπου είναι κατάλληλο και ειδικά για τις αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων, η χρήση του όρου *επίπεδο* συνεπάγεται 'συνελικτικό επίπεδο' για χάρη συντομίας. Οι πίνακες 3.1, 4.1 και 4.2 παρέχουν μια επισκόπηση των δημόσια

<sup>1</sup><https://scholar.google.com>

<sup>2</sup><https://ncbi.nlm.nih.gov/pubmed/>

<sup>3</sup><https://www.scopus.com/search/form.uri?=display=basic>

διαθέσιμων βάσεων δεδομένων καρδιολογίας που έχουν χρησιμοποιηθεί για εκπαίδευση μοντέλων βαθιάς μάθησης.

### 3.2. Βαθιά μάθηση με δομημένα δεδομένα

Τα δομημένα δεδομένα περιλαμβάνουν κυρίως EHRs και συνήθως διατηρούνται σε σχεσιακές βάσεις δεδομένων. Μια σύνοψη των εφαρμογών βαθιάς μάθησης που χρησιμοποιούν δομημένα δεδομένα παρουσιάζεται στον Πίνακα 3.2.

Τα RNNs έχουν χρησιμοποιηθεί για τη διάγνωση καρδιαγγειακών νοσημάτων με χρήση δομημένων δεδομένων. Στο [24] οι συγγραφείς προβλέπουν την πίεση του αίματος (Blood Pressure, BP) κατά τη διάρκεια χειρουργικής επέμβασης και μετά, χρησιμοποιώντας LSTM. Πραγματοποίησαν πειράματα σε ένα σύνολο 12036 χειρουργικών επεμβάσεων που περιέχουν πληροφορίες για ενδοεγχειρητικά σήματα (θερμοκρασία σώματος, αναπνευστικός ρυθμός, καρδιακός ρυθμός, διαστολική BP (Diastolic Blood Pressure, DBP), συστολική BP (Systolic Blood Pressure, SBP), κλάσμα εισπνεόμενου  $O_2$ ), επιτυγχάνοντας καλύτερα αποτελέσματα από τα KNN και SVM. Οι Choi et al. [22] εκπαίδευσαν ένα GRU με διαχρονικά δεδομένα EHR, ανιχνεύοντας σχέσεις μεταξύ χρονικών συμβάντων (διάγνωση ασθενειών, εντολές φαρμάκων κ.λπ.), χρησιμοποιώντας ένα παράθυρο παρατήρησης. Κάνουν διάγνωση καρδιακής ανεπάρκειας (Heart Failure, HF) με AUC 0.777 για παράθυρο 12 μηνών και 0.883 για παράθυρο 18 μηνών, καλύτερα από τα MLP, SVM και KNN. Οι Purushotham et al. [25] σύγκριναν τον super-learner (σύνολο ρηχών αλγορίθμων μηχανικής μάθησης) [29] με FNN, RNN και ένα πολυτροπικό μοντέλο βαθιάς μάθησης που προτάθηκε από τους συγγραφείς στη βάση δεδομένων MIMIC. Το προτεινόμενο πλαίσιο χρησιμοποιεί FNN και GRU για το χειρισμό των μη-χρονικών και χρονικών χαρακτηριστικών αντίστοιχα, μαθαίνοντας έτσι τις κοινές λανθάνουσες αναπαραστάσεις τους για πρόβλεψη. Τα αποτελέσματα δείχνουν ότι οι μέθοδοι βαθιάς μάθησης υπερβαίνουν κατά πολύ τον super-learner στην πλειονότητα των προβλημάτων πρόβλεψης της MIMIC (πρόβλεψη θνησιμότητας εντός νοσοκομείου με AUC 0.873, πρόβλεψη βραχυπρόθεσμης θνησιμότητας με AUC 0.871, πρόβλεψη μακροπρόθεσμης θνησιμότητας με AUC 0.87 και πρόβλεψη κωδικού ICD-9 με AUC 0.777). Οι Kim et al. [26] δημιούργησαν δύο μοντέλα πρόγνωσης ιατρικού ιστορικού χρησιμοποιώντας δίκτυα προσοχής και τα αξιολόγησαν σε 50000 ασθενείς με υπέρταση. Έδειξαν ότι η χρήση ενός GRU διπλής κατεύθυνσης παρέχει καλύτερη διακριτική ικανότητα από ένα αντίστοιχο συνελικτικό δίκτυο, το οποίο όμως έχει μικρότερο χρόνο εκπαίδευσης με ανταγωνιστική ακρίβεια.

Τα AE χρησιμοποιήθηκαν για τη διάγνωση καρδιαγγειακών νοσημάτων με δομημένα δεδομένα. Οι Hsiao et al. [23] εκπαίδευσαν ένα AE και ένα επίπεδο softmax για την ανάλυση κινδύνου τεσσάρων κατηγοριών καρδιαγγειακών παθήσεων. Η εί-

Table 3.1: Δημιόσιες καρδιολογικές βάσεις δεδομένων, δομημένων δεδομένων και σημάτων

Βάση δεδομένων	Ακρωνύμιο	Ασθενείς	Πρόβλημα
<b>Δομημένες βάσεις δεδομένων</b>			
Medical Information Mart for Intensive Care III [3]	MIMIC	38597	53423 νοσοκομειακές εισαγωγές για πρόβλεψη ICD-9
KNHANES-VI [4]	KNH	8108	δημογραφικά, τεστ αίματος και τρόπος ζωής
<b>Βάσεις δεδομένων σημάτων (όλα ECG εκτός του [5])</b>			
IEEE-TBME PPG Respiratory Rate Benchmark Dataset [5]	PPGDB	42	επίμηση ρυθμού αναπνοής με χρήση PPG
Creighton University Ventricular Tachyarrhythmia [6]	CREI	35	εντοπισμός κοιλιακής ταχυαρρυθμίας
MIT-BIH Atrial Fibrillation Database [7]	AFDB	25	πρόβλεψη HF
BIH Deaconess Medical Center CHF Database [8]	CHFDB	15	ταξινόμηση CHF
St.Petersburg Institute of Cardiological Technics [9]	INDB	32	εντοπισμός QRS και ταξινόμηση χτύπων ECG
Long-Term Atrial Fibrillation Database [10]	LTAfDB	84	εντοπισμός QRS και ταξινόμηση χτύπων ECG
Long-Term ST Database [11]	LTSTDB	80	ανίχνευση χτύπων ST και ταξινόμηση
MIT-BIH Arrhythmia Database [12]	MITDB	47	ανίχνευση αρρυθμιών
MIT-BIH Noise Stress Test Database [13]	NSTDB	12	χρήση για δοκιμή αντοχής θορύβου μοντέλων
MIT-BIH Normal Sinus Rhythm Database [9]	NSRDB	18	ανίχνευση αρρυθμιών
MIT-BIH Normal Sinus Rhythm RR Interval Database [14]	NSR2DB	54	ταξινόμηση χτύπων ECG
Fantasia Database [15]	FAN	40	ταξινόμηση χτύπων ECG
AF Classification short single lead ECG Physionet 2017 [12]	PHY17	— <sup>a</sup>	ταξινόμηση χτύπων ECG (12186 σήματα μονού καναλιού)
Physionet 2016 Challenge [16]	PHY16	1297	ταξινόμηση ήχων καρδιάς με χρήση PCG (3126 σήματα)
Physikalisch-Technische Bundesanstalt ECG Database [17]	PTBDB	268	διάγνωση καρδιακών ασθενειών
QT Database [18]	QTDB	105	ανίχνευση χτύπων QT και ταξινόμηση
MIT-BIH Supraventricular Arrhythmia Database [19]	SVDB	78	ανίχνευση υπερκοιλιακών αρρυθμιών
Non-Invasive Fetal ECG Physionet Challenge Dataset [20]	PHY13	447	μέτρηση εμβρυϊκού HR, διαστήματος RR και QT
DeepQ Arrhythmia Database [21]	DeepQ	299	ταξινόμηση χτύπων ECG (897 σήματα)

<sup>a</sup>Ο αριθμός των ασθενών δεν αναφέρεται.

Table 3.2: Εφαρμογές βαθιάς μάθησης με χρήση δομημένων δεδομένων

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Αποτέλεσμα <sup>b</sup>
Gopalswamy 2017 [24]	LSTM	πρόβλεψη BP και διάφορα περιθώρια με χρήση πολλαπλών βιοδεικτών (μη-δημόσια)	73.1%
Choi 2016 [22]	GRU	διάγνωση του HF με χρήση GRU και παραθύρου παρατήρησης (μη-δημόσια)	0.883 <sup>b</sup>
Purushotham 2018 [25]	FNN, GRU	προβλεψια πρόβλεψης της MIMIC με χρήση ενός FNN και ενός μοντέλου GRU (MIMIC)	πολλαπλά
Kim 2017 [26]	GRU, CNN	πρόβλεψη υψηλού ρίσκου αγγειακής ασθένειας με ένα Bi-GRU και ενός ID CNN (μη-δημόσια)	πολλαπλά
Hsiao 2016 [23]	AE	ανάλυση ρίσκου CVD με AE και softmax σε ασθeneίες, μετεωρολογικά δεδομένα (μη-δημόσια)	-- <sup>b</sup>
Kim 2017 [27]	DBN	πρόβλεψη καρδιαγγειακού ρίσκου με χρήση DBN (KINH)	83.9%
Huang 2018 [28]	SDAE	πρόβλεψη ACS με SDAE με δύο περιρισμούς συστηματοποίησης και softmax (μη-δημόσια)	73.0%

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>Υπάρχει μεγάλη μεταβλητότητα στην αναφορά αποτελεσμάτων. Όλα τα αποτελέσματα είναι ακρίβειες εκτός από το [22] που αναφέρει AUC και το [23] που είναι στατιστική μέγεθος.



σοδος περιελάμβανε δημογραφικά στοιχεία, κωδικούς ICD-9 από αρχεία εξωτερικών ασθενών, ρύπους συγκέντρωσης και μετεωρολογικές παραμέτρους από περιβαλλοντικά αρχεία. Οι Huang et al. [28] εκπαίδευσαν ένα SDAE χρησιμοποιώντας ένα σύνολο δεδομένων EHR 3464 ασθενών για να προβλέψουν Οξύ Στεφανιαίο Σύνδρομο (Acute Coronary Syndrome, ACS). Το SDAE έχει δύο περιορισμούς συστηματοποίησης (regularization) που κάνουν τις ανακατασκευασμένες αναπαραστάσεις χαρακτηριστικών να περιέχουν περισσότερες πληροφορίες σχετικές με το επίπεδο κινδύνου, καταγράφοντας έτσι τα χαρακτηριστικά των ασθενών σε παρόμοια επίπεδα κινδύνου και διατηρώντας τις διακριτικές πληροφορίες σε διαφορετικά επίπεδα κινδύνου. Στη συνέχεια, τοποθέτησαν ένα επίπεδο softmax, το οποίο προσαρμόζεται στο πρόβλημα της κλινικής πρόβλεψης κινδύνου.

Τα DBN έχουν επίσης χρησιμοποιηθεί σε συνδυασμό με δομημένα δεδομένα εκτός από τα RNNs και AEs. Στο [27] οι συγγραφείς πρώτα εφήρμοσαν μια στατιστική μελέτη ενός συνόλου δεδομένων 4244 εγγραφών για την εύρεση μεταβλητών που σχετίζονται με καρδιαγγειακές παθήσεις (ηλικία, φύλο, χοληστερόλη, λιποπρωτεΐνη υψηλής πυκνότητας, SBP, DBP, κάπνισμα, διαβήτης). Στη συνέχεια, ανέπτυξαν ένα μοντέλο DBN για την πρόβλεψη καρδιαγγειακών παθήσεων (υπέρταση, υπερλιπιδαιμία, μυοκαρδιακό έμφραγμα (Myocardial Infarction, MI), στηθάγχη). Σύγκριναν το μοντέλο τους με Naive Bayes, λογιστική παλινδρόμηση, SVM, RF και ένα βασικό DBN πετυχαίνοντας καλύτερα αποτελέσματα.

Σύμφωνα με τη βιβλιογραφία, τα RNN χρησιμοποιούνται ευρέως σε δομημένα δεδομένα καρδιολογίας επειδή είναι σε θέση να βρουν χρονικά χαρακτηριστικά καλύτερα από άλλες μεθόδους βαθιάς/μηχανικής μάθησης. Από την άλλη πλευρά, οι εφαρμογές σε αυτόν τον τομέα είναι σχετικά λίγες και αυτό συμβαίνει κυρίως επειδή υπάρχει μικρός αριθμός δημόσιων βάσεων δεδομένων, κάτι το οποίο εμποδίζει την περαιτέρω αξιολόγηση και σύγκριση διαφορετικών αρχιτεκτονικών. Επιπλέον, οι δομημένες βάσεις δεδομένων λόγω σχεδιασμού τους περιέχουν λιγότερες πληροφορίες για έναν μεμονωμένο ασθενή και επικεντρώνονται περισσότερο σε ομάδες ασθενών, καθιστώντας αυτές πιο κατάλληλες για επιδημιολογικές μελέτες παρά για τον τομέα της καρδιολογίας.

### 3.3. Βαθιά μάθηση με σήματα

Τα σήματα περιλαμβάνουν χρονοσειρές όπως ηλεκτροκαρδιογράφημα (Electrocardiogram, ECG), φωνοκαρδιογράφημα (Phonocardiogram, PCG), παλμομετρικά δεδομένα και δεδομένα από φορητά εξαρτήματα. Ένας λόγος που η παραδοσιακή μηχανική μάθηση έχει δουλέψει αρκετά καλά σε αυτόν τον τομέα τα προηγούμενα χρόνια οφείλεται στη χρήση χειροποίητων και προσεκτικά σχεδιασμένων χαρακτηριστικών από εμπειρογνώμονες, όπως τα στατιστικά μέτρα από το ECG beats και το διάστημα RR [30].

Η βαθιά μάθηση μπορεί να βελτιώσει τα μοντέλα όταν οι επισημάνσεις των ειδικών είναι χαμηλής ποιότητας ή όταν είναι δύσκολο να δημιουργηθεί ένα μοντέλο με χρήση χειροποίητων χαρακτηριστικών. Μια σύνοψη των εφαρμογών βαθιάς μάθησης που χρησιμοποιούν σήματα εμφανίζεται στους πίνακες 3.3, 3.4, 3.5 και 3.6.

### 3.3.1. Ηλεκτροκαρδιογράφημα

Το ECG είναι η μέθοδος μέτρησης των ηλεκτρικών δυναμικών της καρδιάς για τη διάγνωση καρδιακών προβλημάτων [39]. Είναι μη-επεμβατική, εύκολη στην απόκτηση και παρέχει χρήσιμες πληροφορίες για τη διάγνωση ασθενειών. Έχει χρησιμοποιηθεί κυρίως για την ανίχνευση αρρυθμιών χρησιμοποιώντας τον μεγάλο αριθμό δημόσιων βάσεων δεδομένων ECG όπως φαίνεται στον Πίνακα 3.1.

#### Ανίχνευση αρρυθμιών με την MITDB

Τα CNN έχουν χρησιμοποιηθεί για την ανίχνευση αρρυθμιών με την MITDB. Οι Zubair et al. [34] ανίχνευσαν τις κορυφές R χρησιμοποιώντας ένα μη-γραμμικό μετασχηματισμό και τμηματοποιούν χτύπους γύρω από αυτό. Στη συνέχεια, χρησιμοποίησαν τα τμήματα για να εκπαιδεύσουν ένα 1D CNN τριών επιπέδων με μεταβλητό ρυθμό μάθησης, βασισμένο στο μέσο τετραγωνικό σφάλμα πετυχαίνοντας καλύτερα αποτελέσματα από τις προηγούμενες καλύτερες μεθόδους. Οι Li et al. [35] χρησιμοποίησαν Μετασχηματισμό Κυματιδίων (Wavelet Transform, WT) για την απομάκρυνση του θορύβου υψηλής συχνότητας και της μετατόπισης της γραμμής βάσης και διορθωτικού spline για την ανίχνευση των κορυφών R. Έπειτα, δημιούργησαν τμήματα γύρω από την κορυφή R τα οποία τροφοδότησαν σε ένα 1D CNN δυο επιπέδων. Στο άρθρο τους οι Kiranyaz et al. [31] εκπαιδεύσαν CNNs που μπορούν να χρησιμοποιηθούν για την ταξινόμηση δεδομένων ECG μακράς διάρκειας και για την παρακολούθηση ECG σε πραγματικό χρόνο, ως μέρος συστήματος έκτακτης προειδοποίησης σε φορητή συσκευή. Το CNN αποτελούνταν από τρία προσαρμόσιμα 1D συνελικτικά επίπεδα. Επιτεύχθηκε 99% και 97.6% για την ταξινόμηση κοιλιακών και υπερκοιλιακών εκτοπικών παλμών αντίστοιχα. Στο [32] οι συγγραφείς χρησιμοποίησαν αφαίρεση μέσου όρου για απομάκρυνση της βάσης αναφοράς, κινητό φίλτρο μέσου όρου για απομάκρυνση των υψηλών συχνοτήτων, διαφορικό φίλτρο για απομάκρυνση της γραμμής βάσης και φίλτρο χτένας (comb) για την απομάκρυνση του θορύβου γραμμής ισχύος. Ανίχνευσαν τα QRS με τον αλγόριθμο Pan-Tompkins [40], εξήγαγαν τμήματα χρησιμοποιώντας δείγματα μετά την κορυφή R τα οποία μετέτρεψαν σε  $256 \times 256 \times 3$  δυαδικές εικόνες. Οι εικόνες έπειτα τροφοδοτήθηκαν σε ένα εξαγωγέα χαρακτηριστικών AlexNet προεκπαιδευμένο στην βάση δεδομένων ImageNet και έπειτα στην Ανάλυση Κύριων Συνιστωσών (Principal

Table 3.3: Εφαρμογές βαθιάς μάθησης για ανίχνευση αρρυθμιών με την MITDB

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Ακρίβεια <sup>b</sup>
Zubair 2016 [34]	CNN	μη-γραμμικός μετασχηματισμός για ανίχνευση R-peak και 1D CNN με μεταβλητό ρυθμό μάθησης	92.7%
Li 2017 [35]	CNN	WT για αποθρομβοποίηση και ανίχνευση των R-peak και 1D CNN δύο επιπέδων	97.5%
Kiranyaz 2016 [31]	CNN	CNN ειδικό για κάθε ασθενή με χρήση προσαρμοσμένων 1D συνελκτικών επιπέδων	99%, 97.6% <sup>b</sup>
Isin 2017 [32]	CNN	φίλτρα αποθρομβοποίησης, Pan-Tomkins, AlexNet για χαρακτηριστικά και PCA για ταξινόμηση	92.0% <sup>b</sup>
Luo 2017 [36]	SDAE	φίλτρα αποθρομβοποίησης, ανίχνευση R-peak με χρήση παραγώγων, WT, SDAE και softmax	97.5%
Jiang 2017 [37]	SDAE	φίλτρα αποθρομβοποίησης, Pan-Tomkins, SDAE και FNN	97.99%
Yang 2017 [38]	SSAE	κανονικοποίηση του ECG, SSAE	99.45%
Wu 2016 [33]	DBN	φίλτρα αποθρομβοποίησης, ecgruwave, δύο τύποι RBMs	99.5% <sup>b</sup>

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.<sup>b</sup>Υπάρχει μεγάλη μεταβλητότητα στην αναφορά αποτελεσμάτων. Τα αποτελέσματα του [31] είναι για κοιλιακούς/υπερκοιλιακούς εκτοπικούς χτύπους, το [32] είναι για τρεις τύπους αρρυθμίας, το [33] είναι για πέντε τύπους αρρυθμίας.

Component Analysis, PCA). Πέτυχαν υψηλή ακρίβεια στην ταξινόμηση τριών τύπων αρρυθμιών της MITDB.

Τα AE έχουν επίσης χρησιμοποιηθεί για ανίχνευση αρρυθμιών με MITDB. Στο άρθρο τους οι Luo et al. [36] χρησιμοποίησαν αξιολόγηση ποιότητας για την αφαίρεση των καρδιακών παλμών χαμηλής ποιότητας, δύο φίλτρα μέσης τιμής για την αφαίρεση του θορύβου της γραμμής ισχύος, του θορύβου υψηλής συχνότητας και της μετατόπισης της γραμμής βάσης. Στη συνέχεια, χρησιμοποίησαν έναν διαφορικό αλγόριθμο για να ανιχνεύσουν τις κορυφές R και τα χρονικά παράθυρα για την κατάτμηση κάθε καρδιακού παλμού. Χρησιμοποίησαν επίσης WT για τον υπολογισμό του φάσματος κάθε καρδιακού παλμού και ενός SDAE για την εξαγωγή χαρακτηριστικών από το φάσμα. Έπειτα, δημιούργησαν έναν ταξινομητή τεσσάρων αρρυθμιών από τον κωδικοποιητή του SDAE και ένα softmax, επιτυγχάνοντας ακρίβεια 97.5%. Στο [37] οι συγγραφείς αποθρομβοποίησαν τα σήματα με ένα χαμηλοπερατό, ένα ζωνοπερατό και ένα φίλτρο διάμεσου. Εντόπισαν κορυφές R χρησιμοποιώντας τον αλγόριθμο Pan-Tomkins και κατάτμησαν/ανακατασκεύασαν τους καρδιακούς παλμούς. Χαρακτηριστικά εξήχθησαν από το σήμα της καρδιάς χρησιμοποιώντας ένα SDAE και χρησιμοποιήθηκε ένα FNN για την ταξινόμηση καρδιακών παλμών από 16 τύπους αρρυθμίας. Παρατηρήθηκε συγκρίσιμη απόδοση σε σχέση με προηγούμενες μεθόδους βασισμένες στην χειροκίνητη εξαγωγή χαρακτηριστικών. Οι Yang et al. [38] κανονικοποίησαν το ECG και στη συνέχεια το τροφοδότησαν σε ένα Στοιβαγμένο Αραιό AE (Stacked Sparse Autoencoder, SSAE). Ταξινομούν έξι τύπους αρρυθμιών επιτυγχάνοντας ακρίβεια 99.5% ενώ ταυτόχρονα αποδεικνύουν την ανθεκτικότητα του έναντι στο θόρυβο με χρήση τεχνητά προστιθέμενου θορύβου.

Τα DBN έχουν επίσης χρησιμοποιηθεί για αυτό το πρόβλημα εκτός από CNN και AE. Οι Wu et al. [33] χρησιμοποίησαν φίλτρα διάμεσου για την απομάκρυνση της γραμμής βάσης, ένα χαμηλοπερατό φίλτρο για την απομάκρυνση του θορύβου γραμμής ισχύος και του θορύβου υψηλής συχνότητας. Εντόπισαν κορυφές R χρησιμοποιώντας το λογισμικό ecgriwave από την Physionet και κατάτμησαν τους ECG χτύπους. Δύο τύποι RBMs, εκπαιδεύτηκαν για την εξαγωγή χαρακτηριστικών από το ECG για ανίχνευση αρρυθμιών. Επιτεύχθηκε ακρίβεια 99.5% σε πέντε κατηγορίες της MITDB.

### Ανίχνευση αρρυθμιών με άλλες βάσεις δεδομένων

Τα CNN χρησιμοποιήθηκαν για ανίχνευση αρρυθμιών χρησιμοποιώντας άλλες βάσεις δεδομένων εκτός της MITDB. Στο [41] οι συγγραφείς δημιούργησαν ένα CNN δύο επιπέδων χρησιμοποιώντας την DeepQ και την MITDB για να ταξινομήσουν τέσσερις τύπους αρρυθμιών. Τα σήματα υπόκεινται σε έντονη προεπεξεργασία με φίλτρα απομάκρυνσης (μεσαία, υψηλή, και χαμηλή διέλευση, αφαίρεση ακραίων τιμών) και κατατμήνονται σε 0.6 δευτερόλεπτα γύρω από την κορυφή R. Στη συνέχεια, τρο-

φοδοτούνται στο CNN μαζί με το διάστημα RR για εκπαίδευση. Οι συγγραφείς χρησιμοποιούν επίσης μια ενεργή μέθοδο μάθησης για την επίτευξη εξατομικευμένων αποτελεσμάτων και βελτιωμένης ακρίβειας, επιτυγχάνοντας υψηλή ευαισθησία και θετική προβλεψιμότητα και στα δύο σύνολα δεδομένων. Οι Hannun et al. [42] δημιούργησαν ένα σύνολο δεδομένων με ECG από φορητές συσκευές το οποίο περιέχει τον μεγαλύτερο αριθμό ασθενών (30000) σε σχέση με προηγούμενες βάσεις δεδομένων και το χρησιμοποίησαν για να εκπαιδεύσουν ένα Residual-CNN με 34 επίπεδα. Το μοντέλο τους ανιχνεύει ένα ευρύ φάσμα αρρυθμιών συνολικά 14 κατηγορίες, ξεπερνώντας τον μέσο καρδιολόγο σε ακρίβεια. Στο άρθρο τους, οι Acharya et al. [43] εκπαιδεύσαν ένα CNN τεσσάρων επιπέδων στην AFDB, MITDB και CREI, για την ταξινόμηση μεταξύ φυσιολογικού, AF, κολπικού περυσισμού και κοιλιακής μαρμαρυγής. Χωρίς την ανίχνευση του QRS, πέτυχαν συγκρίσιμες επιδόσεις με προηγούμενες μεθόδους που βασίστηκαν στην ανίχνευση κορυφών R και στην χειροκίνητη εξαγωγή χαρακτηριστικών. Οι ίδιοι συγγραφείς επίσης εκπαιδεύσαν την προηγούμενη CNN αρχιτεκτονική για τον εντοπισμό των απινιδωτικών και μη-απινιδωτικών κοιλιακών αρρυθμιών [44], εντοπισμό CAD ασθενών με τις FAN και INDB [45], ταξινόμηση CHF με τις CHFDB, NSTDB, FAN [46] και την δοκιμασία της αντοχής τους στο θόρυβο με αποθορυβοποίηση WT [47].

Μια εφαρμογή των RNN σε αυτή την περιοχή είναι από τους Schwab et al. [50] που δημιούργησαν ένα ensemble από RNN που διακρίνει μεταξύ φυσιολογικών φλεβοκομβικών ρυθμών, AF, άλλων τύπων αρρυθμίας και θορυβώδους σήματος. Εισηγάγαν μια μορφοποίηση του προβλήματος κατά την οποία τμηματοποιούν το ECG σε καρδιακούς παλμούς για να μειώσουν τον αριθμό των χρονικών βημάτων ανά ακολουθία. Επέκτειναν επίσης τα RNNs με έναν μηχανισμό προσοχής που τους επιτρέπει να εκτιμήσουν σε ποιους καρδιακούς παλμούς επικεντρώνεται το RNN για να λάβει τις αποφάσεις του και να επιτύχει συγκρίσιμη αποτελεσματικότητα με άλλες μεθόδους χρησιμοποιώντας λιγότερες παραμέτρους.

### Ανίχνευση AF

Τα CNN χρησιμοποιήθηκαν για την ανίχνευση AF. Οι Yao et al. [51] εξήγαγαν την ακολουθία του στιγμιαίου καρδιακού ρυθμού, η οποία τροφοδοτείται σε ένα CNN πολλαπλής κλίμακας που εξάγει το αποτέλεσμα ανίχνευσης AF, επιτυγχάνοντας καλύτερα αποτελέσματα από τις προηγούμενες μεθόδους όσον αφορά την ακρίβεια. Οι Xia et al. [52] σύγκριναν δύο CNN, με τρία και δύο επίπεδα, τα οποία τροφοδοτήθηκαν με φάσματα σημάτων από την AFDB χρησιμοποιώντας βραχυπρόθεσμο μετασχηματισμό Fourier (Short-Time Fourier Transform, STFT) και στατικό WT αντίστοιχα. Τα πειράματά τους κατέληξαν στο συμπέρασμα ότι η χρήση του στατικού WT επιτυγχάνει μια ελαφρώς καλύτερη ακρίβεια για αυτό το πρόβλημα.

Table 3.4: Εφαρμογές βαθιάς μάθησης με χρήση ECG για ανίχνευση αρρυθμιών και AF

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Ακρίβεια <sup>b</sup>
<b>Ανίχνευση αρρυθμιών</b>			
Wu 2018 [41]	CNN	ενεργή μάθηση και CNN δύο επιπέδων με ECG και διαστήματα RR (MITDB, DeepQ)	πολλαπλά
Hannun 2019 [42]	CNN	CNN 34-επιπέδων (μη-δημόσια)	80% <sup>b</sup>
Acharya 2017 [43]	CNN	CNN τεσσάρων επιπέδων (AFDB, MITDB, Creighton)	92.5%
Schwab 2017 [50]	RNN	ensemble από RNNs με μηχανισμό προσοχής (PHY17)	79%

#### Ανίχνευση AF

Yao 2017 [51]	CNN	πολυ-κλιμακωτό CNN (AFDB, LIAFDB, μη-δημόσια)	98.18%
Xia 2018 [52]	CNN	CNN με φασματογράφημα από STFT ή στατικού WT (AFDB)	98.29%
Andersen 2018 [53]	CNN, LSTM	διαστήματα RR με CNN-LSTM (MITDB, AFDB, NSRDB)	87.40%
Xiong 2015 [48]	AE	κατωφλίωση προσομοιωμένης κλιμακίας WT και AE αποθρομβοποίησης (MITDB, NSTDB)	~18, <sup>7b</sup>
Taji 2017 [49]	DBN	μέλωση false alarm κατά τη διάκριση ανίχνευσης AF σε θορυβώδη ECG σήματα (AFDB, NSTDB)	87% <sup>b</sup>

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>Υπόκειται μεγάλη μεταβλητότητα στην αναφορά αποτελεσμάτων. Τα αποτελέσματα του [42] είναι precision, το [48] αναφέρει SNR και πολλαπλά αποτελέσματα ανάλογα με τον προστιθέμενο θορυβό, το αποτέλεσμα του [49] αφορά μελέτη αντοχής θορύβου.

Εκτός από τα CNN έχουν χρησιμοποιηθεί και άλλες αρχιτεκτονικές για την ανίχνευση AF. Οι Andersen et al. [53] μετέτρεψαν τα σήματα ECG από την AFDB σε διαστήματα RR για να τα ταξινομήσουν για ανίχνευση AF. Στη συνέχεια, κατάτμησαν τα διαστήματα RR σε 30 δείγματα το καθένα και τα τροφοδότησαν σε ένα δίκτυο με δύο επίπεδα ακολουθούμενα από ένα επίπεδο συγκέντρωσης (pooling) και ένα επίπεδο LSTM με 100 νευρώνες. Η μέθοδος επικυρώθηκε στην MITDB και την NSRDB επιτυγχάνοντας ακρίβεια που δείχνει ότι μπορεί να γενικεύσει. Στο [48] οι συγγραφείς πρόσθεσαν σήματα θορύβου από την NSTDB στην MITDB και στη συνέχεια χρησιμοποίησαν καταωφλίωση προσαρμοσμένης κλίμακας WT, για να απομακρύνουν το μεγαλύτερο μέρος του θορύβου και ένα SDAE για να αφαιρεθεί ο υπολειπόμενος θόρυβος. Τα πειράματά τους έδειξαν ότι η αύξηση του αριθμού των δεδομένων εκπαίδευσης σε 1000, αυξάνει δραματικά το λόγο σήματος προς θόρυβο μετά την αποθορυβοποίηση. Οι Taji et al. [49] εκπαίδευσαν ένα DBN για να ταξινομήσουν αποδεκτά από μη-αποδεκτά τμήματα ECG, έτσι ώστε να μειώσουν το ποσοστό ψευδούς συναγερού που προκαλείται από κακής ποιότητας ECG κατά την ανίχνευση AF. Οκτώ διαφορετικά επίπεδα ποιότητας ECG παρέχονται, προσθέτοντας στο ECG θόρυβο κίνησης από την NSTDB. Με SNR  $-20dB$  στο ECG, η μέθοδος τους πέτυχε αύξηση 22% της ακρίβειας σε σύγκριση με ένα μοντέλο βάσης αναφοράς.

### Άλλες εφαρμογές με δημόσιες βάσεις δεδομένων

Ταξινόμηση χτύπων του ECG πραγματοποιήθηκε επίσης από μια σειρά μελετών που χρησιμοποίησαν δημόσιες βάσεις δεδομένων. Στο [54] οι συγγραφείς μετεκπαίδευσαν ένα Inception-v3 προεκπαιδευμένο στην ImageNet, χρησιμοποιώντας σήματα από την LTSTDB για την ταξινόμηση συμβάντων ST. Τα δείγματα εκπαίδευσης ήταν πάνω από 500000 τμήματα ST και μη-ST ECG σημάτων με διάρκεια δέκα δευτερολέπτων που έπειτα μετατράπηκαν σε εικόνες. Επιτυγχάνουν συγκρίσιμες επιδόσεις με προηγούμενες πολύπλοκες μεθόδους βασισμένες σε χειροκίνητο ορισμό κανόνων. Οι Rahhal et al. [55] εκπαίδευσαν ένα SDAEs με περιορισμό αραιότητας και ένα softmax για την ταξινόμηση των ECG χτύπων. Σε κάθε επανάληψη, ο εμπειρογνώμονας επισημάνει τους πιο αβέβαιους χτύπους ECG στο σετ δοκιμών, τα οποία στη συνέχεια χρησιμοποιούνται για εκπαίδευση, ενώ η έξοδος του δικτύου εκχωρεί τα μέτρα εμπιστοσύνης σε κάθε χτύπο. Πειράματα που εκτελούνται στις MITDB, INDB, SVDB δείχνουν την αξιοπιστία και την υπολογιστική αποτελεσματικότητα της μεθόδου. Στο [56] οι συγγραφείς εκπαίδευσαν τρεις ξεχωριστές αρχιτεκτονικές για να προσδιορίσουν τα κύματα P-QRS-T στο ECG με την QTDB. Σύγκριναν ένα FNN δύο επιπέδων, ένα CNN δύο επιπέδων και ένα CNN δύο επιπέδων με dropout, με το δεύτερο να πετυχαίνει τα καλύτερα αποτελέσματα.

Το ECG έχει επίσης χρησιμοποιηθεί για τον εντοπισμό και την ταξινόμηση του

Table 3.5: Εφαρμογές βαθιάς μάθησης με χρήση ECG σε άλλες εφαρμογές

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Ακρίβεια <sup>b</sup>
<b>Άλλες εφαρμογές</b>			
Xiao 2018 [54]	CNN	τάξινοηση γεγονότων ST από ECG με μεταφοράς μάθησης στο Inception-v3 (LTSTDB)	0.867 <sup>b</sup>
Rahhal 2016 [55]	SDAE	SDAE με περιορισμό φαιρότητας και softmax (MITDB, INDB, SVDB)	>99% <sup>b</sup>
Abrishami 2018 [56]	Multiple	σύνθετων ένα FNN, CNN και CNN με dropout για εντοπισμό κυμάτων ECG (QTDB)	96.2%
Wu 2016 [57]	SAE	εντοπισμός και τάξινοηση MI με SAE και πολυ-χλιμακωτό διακριτό WT (PTBDB)	~99% <sup>b</sup>
Reasat 2017 [59]	Inception	ανίχνευση MI με Inception μπλόκ για κάθε κανάλι ECG (PTBDB)	84.54%
Zhong 2018 [60]	CNN	CNN τριών επιπέδων για την τάξινοηση εμβρυϊκών ECG (PHY13)	77.85%

**Άλλες εφαρμογές (μη-δημόσιες βάσεις)**

Ripoll 2016 [61]	RBM	ανίχνευση μη-καρδιακών ECG με προεκπαιδευμένα RBMs	85.52%
Jin 2017 [62]	CNN	ανίχνευση μη-καρδιακών ECG με lead-CNN και κανόνα συμπερασμού	86.22%
Liu 2018 [63]	Multiple	συνέχρισαν το Inception και ένα 1D CNN για πρόωση κοιλιακή συστολή με ECG	88.5%
Hwang 2018 [58]	CNN, RNN	ανίχνευση στρες με ένα CNN ενός επιπέδου με dropout και δύο RNNs	87.39% <sup>b</sup>

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>Υπάρχει μεγάλη μεταβλητότητα στην αναφορά αποτελεσμάτων. Τα αποτελέσματα του [54] αναφέρει AUC, το [55] αναφέρει πολλαπλές ακρίβειες για υπερκοιλιακούς/κοιλιακούς εκτοπισμούς χτύπους, το [57] αναφέρει ευαισθησία και εξειδίκευση (specificity), το [58] αναφέρει αποτελέσματα για δύο περιπτώσεις.



MI. Στο άρθρο τους οι Wu et al. [57] εντόπισαν και ταξινομήσαν MI στην PTBDB. Χρησιμοποίησαν διακριτό WT πολλαπλών κλιμάκων, για να διευκολύνουν την εξαγωγή χαρακτηριστικών για MI σε συγκεκριμένες αναλυτικότητες συχνοτήτων και ένα επίπεδο παλινδρόμησης softmax για να δημιουργήσουν έναν ταξινομητή πολλαπλών κατηγοριών. Τα πειράματα επικύρωσης δείχνουν ότι η μέθοδος τους απέδωσε καλύτερα από τις προηγούμενες μεθόδους, όσον αφορά την ευαισθησία και την εξειδίκευση. Η PTBDB χρησιμοποιήθηκε επίσης από τους Reasat et al. [59] για να εκπαιδεύσουν ένα μοντέλο CNN βασισμένο στο inception. Κάθε ηλεκτρόδιο του ECG τροφοδοτείται σε ένα inception μπλοκ, ακολουθούμενο από επίπεδα συγκόλλησης (concatenation), παγκόσμιας μέσης συγκέντρωσης (global average pooling) και ένα softmax. Οι συγγραφείς συνέκριναν τη μέθοδο τους με μια προηγούμενη σύγχρονη μέθοδο που χρησιμοποιεί SWT, πετυχαίνοντας καλύτερα αποτελέσματα.

Συμπλέγματα εμβρυϊκού QRS ταυτοποιήθηκαν με ένα CNN τριών στρωμάτων με dropout από τους Zhong et al. [60] με την PHY13. Αρχικά, τα σήματα κακής ποιότητας απορρίπτονται χρησιμοποιώντας την εντροπία του δείγματος και στη συνέχεια κανονικοποιημένα τμήματα διάρκειας 100ms τροφοδοτούνται στο CNN για εκπαίδευση. Οι συγγραφείς συνέκριναν τη μέθοδο τους με τα KNN, Naive Bayes και SVM επιτυγχάνοντας καλύτερα αποτελέσματα.

### Άλλες εφαρμογές με μη-δημόσιες βάσεις δεδομένων

Η ανίχνευση μη-φυσιολογικών ECG μελετήθηκε από μια σειρά δημοσιεύσεων. Οι Ripoll et al. [61] χρησιμοποίησαν προεκπαιδευμένα μοντέλα με ECG από 1390 ασθενείς, για να αξιολογήσουν εάν ένας ασθενής στο ασθενοφόρο ή στα επείγοντα πρέπει να παραπεμφθεί σε μια καρδιολογική υπηρεσία. Σύγκριναν το μοντέλο τους με τα KNN, SVM, μηχανές ακραίας μάθησης (Extreme Learning Machines, ELMs) και με ένα σύστημα εμπειρογνομόνων πετυχαίνοντας καλύτερα αποτελέσματα στην ακρίβεια και την εξειδίκευση. Στο [62] οι συγγραφείς εκπαιδεύσαν ένα μοντέλο το οποίο ταξινομεί φυσιολογικούς και μη-φυσιολογικούς ασθενείς με 193690 αρχεία ECG 10 έως 20 δευτερολέπτων. Το μοντέλο τους αποτελείται από δύο παράλληλα μέρη; την στατιστική μάθηση και ένα κανόνα συμπεράσματος. Κατά τη διάρκεια της μάθησης, τα ECG υπόκεινται σε προεπεξεργασία με χρήση χαμηλοπερατών και ζωνοπερατών φίλτρων, στη συνέχεια τροφοδοτούνται σε δύο παράλληλα lead-CNNs και τέλος χρησιμοποιείται Bayesian σύντηξη για τον συνδυασμό των εξόδων πιθανότητας. Κατά τη διάρκεια του συμπερασμού, ανιχνεύονται οι θέσεις κορυφής R στο αρχείο ECG και χρησιμοποιούνται τέσσερις κανόνες ασθένειας για την ανάλυση. Τέλος, χρησιμοποιούν το μέσο όρο bias για τον προσδιορισμό του αποτελέσματος.

Άλλες εφαρμογές περιλαμβάνουν την ταξινόμηση της πρόωρης κοιλιακής συστολής και την ανίχνευση του στρες. Οι Liu et al. [63] χρησιμοποίησαν ένα σύνολο δε-

δομένων μονού καναλιού με 2400 φυσιολογικά και πρόωρης κοιλιακής συστολής ECG από το Νοσοκομείο Παιδών της Σαγκάης για εκπαίδευση. Δύο διαφορετικά μοντέλα εκπαιδεύτηκαν χρησιμοποιώντας τις εικόνες των κυματομορφών. Το πρώτο ήταν ένα CNN δύο επιπέδων με dropout και το δεύτερο ένα Inception-v3 εκπαιδευμένο στην Imagenet. Άλλα τρία μοντέλα εκπαιδεύτηκαν χρησιμοποιώντας τα 1D σήματα. Το πρώτο μοντέλο ήταν ένα FNN με dropout, το δεύτερο ένα 1D CNN τριών επιπέδων και το τρίτο ένα 2D CNN το ίδιο με το πρώτο αλλά εκπαιδευμένο με μια στοιβαγμένη έκδοση του σήματος (με επαύξηση δεδομένων). Τα πειράματα τους έδειξαν ότι το 1D CNN με τα τρία επίπεδα είχαν καλύτερα και πιο σταθερά αποτελέσματα. Στο [58] οι συγγραφείς εκπαιδύσαν ένα δίκτυο με ένα συνελικτικό επίπεδο με dropout, ακολουθούμενο από δύο RNNs για τον εντοπισμό του άγχους χρησιμοποιώντας βραχυπρόθεσμα δεδομένα ECG. Έδειξαν ότι το δίκτυο τους πέτυχε τα καλύτερα αποτελέσματα σε σύγκριση με παραδοσιακές μεθόδους μάθησης μηχανών και DNN βάσης αναφοράς.

### Συνολική άποψη για την χρήση της βαθιάς μάθησης στο ECG

Πολλές μέθοδοι βαθιάς μάθησης έχουν χρησιμοποιήσει ECG για την εκπαίδευση μοντέλων χρησιμοποιώντας τον μεγάλο αριθμό διαθέσιμων βάσεων δεδομένων. Είναι προφανές από τη βιβλιογραφία ότι οι περισσότερες μέθοδοι βαθιάς μάθησης (κυρίως CNN και SDAE) σε αυτή την περιοχή αποτελούνται από τρία μέρη: φιλτράρισμα για αποθρομβοποίηση, ανίχνευση κορυφών R για κατάτμηση χτύπων και ένα νευρωνικό δίκτυο για την εξαγωγή χαρακτηριστικών. Ένα άλλο δημοφιλές σύνολο μεθόδων είναι η μετατροπή των ECG σε εικόνες, για την αξιοποίηση ενός ευρύ φάσματος των αρχιτεκτονικών και των προεκπαιδευμένων μοντέλων που έχουν ήδη κατασκευαστεί για τις μορφές απεικόνισης. Αυτό έγινε χρησιμοποιώντας τεχνικές ανάλυσης φάσματος [36, 52] και μετατροπών σε δυαδική εικόνα [54, 63, 32].

#### 3.3.2. Φωνοκαρδιογράφημα με χρήση της Physionet 2016

Ο διαγωνισμός της Physionet/Computing στην Καρδιολογία (Cinc) 2016 (PHY16) αφορούσε την ταξινόμηση των φυσιολογικών/παθολογικών καρδιακών ηχογραφήσεων. Το δεδομένα εκπαίδευσης αποτελούνται από πέντε βάσεις δεδομένων (A έως E) που περιέχουν 3126 φωνοκαρδιογράφημα (PCGs), τα οποία διαρκούν από 5 δευτερόλεπτα έως 120 δευτερόλεπτα.

Οι περισσότερες από τις μεθόδους μετατρέπουν τα PCG σε εικόνες χρησιμοποιώντας τεχνικές φασματογραφίας. Οι Rubin et al. [64] χρησιμοποίησαν ένα κρυφό ημι-Markov μοντέλο για την κατάτμηση της έναρξης κάθε παλμού της καρδιάς, το οποίο στη συνέχεια μετατράπηκε σε φασματογράφημα με τη χρήση μετρητών συχνότητας Mel-Frequency Cepstral (MFCCs). Κάθε φασματογράφημα ταξινομήθηκε σε φυ-

σιολογικό ή μη-φυσιολογικό χρησιμοποιώντας ένα CNN δύο επιπέδων με τροποποιημένη συνάρτηση απώλειας που μεγιστοποιεί την ευαισθησία και την εξειδίκευση, μαζί με μια παράμετρο συστηματοποίησης. Η τελική ταξινόμηση του σήματος ήταν η μέση πιθανότητα όλων των πιθανοτήτων των τμημάτων. Πέτυχαν συνολική ακρίβεια 83.99% τοποθετώντας την μέθοδο όγδοη κατά τη διάρκεια του διαγωνισμού στο PHY16. Οι Kucharski et al. [65] χρησιμοποίησαν ένα φασματογράφημα οκτώ δευτερολέπτων για τα τμήματα, πριν τροφοδοτηθούν σε ένα CNN πέντε επιπέδων με dropout. Η μέθοδος τους πέτυχε ευαισθησία 99.1% και ειδικότητα 91.6% οι οποίες είναι συγκρίσιμες με αποτελέσματα προηγούμενων μεθόδων. Οι Dominguez et al. [66] ταξινόμησαν τα σήματα και τα προεπεξεργάστηκαν χρησιμοποιώντας τον νευρομορφικό ακουστικό αισθητήρα [67] για να αποσυνθέσουν τις πληροφορίες ήχου σε ζώνες συχνότητας. Στη συνέχεια, υπολογίζουν τα φασματογραφήματα που τροφοδοτούνται σε μια τροποποιημένη έκδοση του δικτύου AlexNet. Το μοντέλο τους πέτυχε ακρίβεια 94.16%, η οποία είναι μια σημαντική βελτίωση σε σύγκριση με το μοντέλο που κέρδισε το PHY16. Στο [68] οι συγγραφείς χρησιμοποίησαν το Adaboost το οποίο τροφοδοτήθηκε με χαρακτηριστικά φασματογραφίας από PCG και ένα CNN το οποίο εκπαιδεύτηκε χρησιμοποιώντας καρδιακούς κύκλους αποσυνθεμένους σε τέσσερις ζώνες συχνότητας. Τέλος, οι έξοδοι του Adaboost και του CNN συνδυάστηκαν για να παράξουν το τελικό αποτέλεσμα ταξινόμησης χρησιμοποιώντας έναν απλό κανόνα απόφασης. Η συνολική ακρίβεια ήταν 89%, τοποθετώντας τη μέθοδο πρώτη στο διαγωνισμό του PHY16.

Τα μοντέλα που δεν μετατρέπουν τα PCG σε φασματογραφήματα φαίνεται να έχουν μικρότερη απόδοση. Οι Ryu et al. [69] εφάρμοσαν ένα φίλτρο με παράθυρο Hamming Window-sinc για αποθορυβοποίηση, έπειτα κλιμάκωσαν το σήμα και χρησιμοποίησαν ένα σταθερό παράθυρο για κατάτμηση. Εκπαίδευσαν ένα 1D CNN τεσσάρων επιπέδων χρησιμοποιώντας τα τμήματα ενώ η τελική ταξινόμηση ήταν ο μέσος όρος όλων των πιθανοτήτων των τμημάτων. Επιτεύχθηκε συνολική ακρίβεια 79.5% στην επίσημη φάση του PHY16.

Τα φωνοκαρδιογραφήματα έχουν επίσης χρησιμοποιηθεί για προβλήματα όπως η αναγνώριση ήχου καρδιάς S1 και S2 από τον Chen et al. [70]. Μετασχημάτισαν τα ηχητικά σήματα της καρδιάς σε μια ακολουθία MFCCs και έπειτα εφάρμοσαν K-μέσους για να συσσωρεύσουν τα χαρακτηριστικά MFCC σε δύο ομάδες για να βελτιώσουν την εκπροσώπηση και τη διακριτική τους ικανότητα. Τα χαρακτηριστικά τροφοδοτούνται στη συνέχεια σε ένα DBN για την εκτέλεση ταξινόμησης S1 και S2. Οι συγγραφείς συνέχισαν τη μέθοδο τους με τα μοντέλα KNN, Gaussian mixture, λογιστική παλινδρόμηση και SVM επιτυγχάνοντας καλύτερα αποτελέσματα.

Σύμφωνα με τη βιβλιογραφία, τα CNN αποτελούν την πλειονότητα των αρχιτεκτονικών νευρωνικών δικτύων που χρησιμοποιήθηκαν για την επίλυση προβλημάτων με PCG. Επιπλέον, όπως και στο ECG, πολλές μέθοδοι βαθιάς μάθησης μετασχηματί-

ζουν τα σήματα σε εικόνες χρησιμοποιώντας τεχνικές φασματογραφίας [68, 64, 65, 66, 71, 72].

### 3.3.3. Άλλα σήματα

#### Παλμομετρικά δεδομένα

Τα παλμομετρικά δεδομένα χρησιμοποιούνται για την εκτίμηση της SBP και της DBP που είναι οι αιμοδυναμικές πιέσεις που ασκούνται στο αρτηριακό σύστημα κατά τη διάρκεια της συστολής και της διαστολής αντίστοιχα [73].

Τα DBN έχουν χρησιμοποιηθεί για την εκτίμηση των SBP και DBP. Στο άρθρο τους οι Lee et al. [74] χρησιμοποίησαν bootstrap-aggregation για να δημιουργήσουν ensemble παραμέτρους και χρησιμοποίησαν τον Adaboost για την εκτίμηση των SBP και DBP. Στη συνέχεια, χρησιμοποίησαν bootstrap και Monte-Carlo για να προσδιορίσουν τα διαστήματα εμπιστοσύνης βασισμένα στο BP, τα οποία εκτιμήθηκαν χρησιμοποιώντας τον ensemble εκτιμητή παλινδρόμησης του DBN. Αυτή η τροποποίηση βελτίωσε σημαντικά την εκτίμηση του BP σε σχέση με το βασικό μοντέλο DBN. Παρόμοια κατεύθυνση έχει ακολουθηθεί για το ίδιο πρόβλημα από τους ίδιους συγγραφείς στα [75, 76, 77].

Παλμομετρικά δεδομένα έχουν επίσης χρησιμοποιηθεί από τους Pan et al. [71] για την εκτίμηση της μεταβολής των ήχων Korotkoff. Οι χτύποι χρησιμοποιήθηκαν για να δημιουργήσουν παράθυρα κεντραρισμένα στις κορυφές των παλμών, οι οποίες μετά εξήχθησαν. Ανάλυση φάσματος λήφθηκε από κάθε χτύπο και όλοι οι χτύποι μεταξύ των χειροκίνητα επισημασμένων SBPs και DBPs επισημάνθηκαν ως Korotkoff. Στη συνέχεια χρησιμοποιήθηκε ένα CNN τριών επιπέδων, για την ανάλυση της συνέπειας στα ηχητικά μοτίβα που συσχετίστηκαν με τους ήχους Korotkoff. Σύμφωνα με τους συγγραφείς, αυτή ήταν η πρώτη μελέτη που διεξήχθη για τέτοιου είδους πρόβλημα, αποδεικνύοντας ότι είναι δύσκολο να προσδιοριστούν οι ήχοι Korotkoff στη συστολή και διαστολή.

#### Δεδομένα από φορητές συσκευές

Οι φορητές συσκευές, οι οποίες επιβάλλουν περιορισμούς στο μέγεθος, την ισχύ και την κατανάλωση μνήμης για τα μοντέλα, έχουν επίσης χρησιμοποιηθεί για τη συλλογή δεδομένων καρδιολογίας για εκπαίδευση μοντέλων βαθιάς μάθησης ανίχνευσης AF.

Οι Shashikumar et al. [72] πήραν ECG, φωτοπληθυσμογραφία (PPG) και δεδομένα από επιταχυνσιόμετρο από 98 άτομα χρησιμοποιώντας μια φορητή συσκευή καρπού και παρήξαν την ανάλυση φάσματος χρησιμοποιώντας συνεχή WT. Εκπαίδευσαν ένα CNN πέντε επιπέδων σε μια σειρά βραχέων παραθύρων με θόρυβο κίνησης και η έξοδός τους συνδυάστηκε με χαρακτηριστικά που υπολογίστηκαν βάσει της μεταβλητότητας του

χτύπου-προς-χτύπο και του δείκτη ποιότητας του σήματος. Η μέθοδος πέτυχε ακρίβεια 91.8% στην ανίχνευση AF και σε συνδυασμό με την υπολογιστική αποτελεσματικότητά της είναι πολλά υποσχόμενη για κλινική εφαρμογή σύμφωνα με τους συγγραφείς. Οι Gotlibovych et al. [78] εκπαίδευσαν ένα δίκτυο CNN ενός επιπέδου ακολουθούμενο από ένα LSTM, χρησιμοποιώντας 180 ώρες PPG από δεδομένα φορητών συσκευών για την ανίχνευση AF. Η χρήση του επιπέδου LSTM επιτρέπει στο δίκτυο να μαθαίνει συσχετίσεις μεταβλητού μήκους σε αντίθεση με το σταθερό μήκος του συνελικτικού επιπέδου. Οι Poh et al. [79] δημιούργησαν μια μεγάλη βάση δεδομένων PPG (πάνω από 180000 σήματα από 3373 ασθενείς), συμπεριλαμβανομένων των δεδομένων από την MIMIC για την ταξινόμηση τεσσάρων ρυθμών. Ένα πυκνά συνδεδεμένο (densely) CNN με έξι επίπεδα χρησιμοποιήθηκε για ταξινόμηση, το οποίο τροφοδοτήθηκε με τμήματα 17 δευτερολέπτων που έχουν αποθορυβοποιηθεί χρησιμοποιώντας ένα ζωνοπερατό φίλτρο. Τα αποτελέσματα ελήφθησαν χρησιμοποιώντας ένα ανεξάρτητο σύνολο δεδομένων 3039 PPG πετυχαίνοντας καλύτερα αποτελέσματα από τις προηγούμενες μεθόδους που βασίστηκαν σε χειροποίητα χαρακτηριστικά.

Εκτός από την ανίχνευση AF, δεδομένα από φορητές συσκευές χρησιμοποιήθηκαν και για την αναζήτηση καλύτερων προγνωστικών καρδιοαγγειακών παθήσεων. Στο [80] οι συγγραφείς εκπαίδευσαν ένα ημι-επιβλεπόμενο, διπλής κατεύθυνσης LSTM σε δεδομένα από 14011 χρήστες της εφαρμογής Cardiogram για την ανίχνευση του διαβήτη, της υψηλής χοληστερόλης, της υψηλής BP και της άπνοιας. Τα αποτελέσματά τους δείχνουν ότι η ανταπόκριση της καρδιάς στη σωματική δραστηριότητα είναι ένας σημαντικός βιοδείκτης για την πρόβλεψη της εμφάνισης μιας νόσου και μπορεί να εντοπιστεί χρησιμοποιώντας βαθιά μάθηση.

Table 3.6: Εφαρμογές βαθιάς μάθησης με χρήση PCG και άλλων σημάτων

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Ακρίβεια <sup>b</sup>
<b>Διαγωνισμός PCG/Physionet 2016</b>			
Rubin 2017 [64]	CNN	λογιστική πιθανοδρόμηση με ένα κρυφό ημι-Markov μοντέλο, MFCCs και ένα CNN δύο επιπέδων	83.99%
Kucharski 2017 [65]	CNN	ανάυση φάσματος και ένα CNN πέντε επιπέδων CNN με dropout	91.6% <sup>b</sup>
Dominguez 2018 [66]	CNN	ανάυση φάσματος και τροποποιημένο AlexNet	94.16%
Potes 2016 [68]	CNN	ensemble από Adaboost και CNN, η έξοδος συνδυάζεται με ένα κανόνα απόφασης	86.02%
Ryu 2016 [69]	CNN	φίλτρα αποθορυβοποίησης και ένα CNN τεσσάρων επιπέδων	79.5%
Chen 2017 [70]	DBN	αναγνώριση S1 και S2 ήχων καρδιάς με χρήση MFCCs, K-means και DBN (μη-δημόσια)	91%

**Άλλα σήματα**

Lee 2017 [74]	DBN	εκτίμηση BP με χρήση bootstrap-aggregation, Monte-Carlo και DBN με παλμομετρικά δεδομένα	πολλαπλά
Pan 2017 [71]	CNN	εύρεση ήχων Korotkoff με χρήση ενός CNN τριών επιπέδων με παλμομετρικά δεδομένα	πολλαπλά <sup>b</sup>
Shashikumar 2017 [72]	CNN	ανάγνωση AF με χρήση ECG, φωτοαγλυθισμογραφία, και επιταχυνσιόμετρο με WT και ένα CNN	91.8%
Gotlibovych 2018 [78]	CNN, LSTM	ανάγνωση AF με χρήση PPG από φορητή συσκευή και ένα LSTM-CNN	>99% <sup>b</sup>
Poh 2018 [79]	CNN	ανάγνωση τεσσάρων ρυθμών με PPG και ένα densely CNN (MIMIC, VORTAL, PPGDB)	87.5% <sup>b</sup>
Ballinger 2018 [80]	LSTM	πρόβλεψη διαβήτη, υψηλής χαλινοστερόλης, BP και άπνοιας από αισθητήρα	0.845 <sup>b</sup>

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν. Στον υποτίτλοκα 'Διαγωνισμός PCG/Physionet 2016' όλες οι δημοσιεύσεις χρησιμοποιούν την PHX εκτός του [70], και στον υποτίτλοκα 'Άλλα Σήματα' όλες οι δημοσιεύσεις χρησιμοποιούν μη-δημόσιες βάσεις δεδομένων εκτός του [79].

<sup>b</sup>Υπόκειται μεγάλη μεταβλητότητα στην αναφορά αποτελεσμάτων. Το [65] αναφέρει specificity, το [71] αναφέρει αποτελέσματα για το SBP και το DBP, το [78] αναφέρει ευαισθησία, ειδικότητα, το [79] αναφέρει την τιμή της θετικής προβλεψιμότητας, το [80] αναφέρει AUC για το diabetest, αποτελέσματα επίσης αναφέρονται για υψηλή χαλινοστερόλη, άπνοια και υψηλό BP.

## Βιβλιογραφία

- [1] Emelia J Benjamin, Michael J Blaha, Stephanie E Chiuve, Mary Cushman, Sandeep R Das, Rajat Deo, Sarah D de Ferranti, James Floyd, Myriam Fornage, Cathleen Gillespie, et al. Heart disease and stroke statistics—2017 update: a report from the american heart association. *Circulation*, 135(10):e146–e603, 2017.
- [2] Elizabeth Wilkins, L Wilson, Kremlin Wickramasinghe, Prachi Bhatnagar, Jose Leal, Ramon Luengo-Fernandez, R Burns, Mike Rayner, and Nick Townsend. European cardiovascular disease statistics 2017. 2017.
- [3] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.
- [4] Sanghui Kweon, Yuna Kim, Myoung-jin Jang, Yoonjung Kim, Kirang Kim, Sunhye Choi, Chaemin Chun, Young-Ho Khang, and Kyungwon Oh. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). *International journal of epidemiology*, 43(1):69–77, 2014.
- [5] Walter Karlen, Srinivas Raman, J Mark Ansermino, and Guy A Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–1953, 2013.
- [6] FM Nolle, FK Badura, JM Catlett, RW Bowser, and MH Sketch. Crei-gard, a new concept in computerized arrhythmia monitoring systems. *Computers in Cardiology*, 13:515–518, 1986.
- [7] George Moody. A new method for detecting atrial fibrillation using RR intervals. *Computers in Cardiology*, pages 227–230, 1983.
- [8] Donald S Baim, Wilson S Colucci, E Scott Monrad, Harton S Smith, Richard F Wright, Alyce Lanoue, Diane F Gauthier, Bernard J Ransil, William Grossman, and Eugene Braunwald. Survival of patients with severe congestive heart failure treated with oral milrinone. *Journal of the American College of Cardiology*, 7(3):661–670, 1986.
- [9] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220, 2000.

- [10] Simona Petrutiu, Alan V Sahakian, and Steven Swiryn. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace*, 9(7):466–470, 2007.
- [11] Franc Jager, Alessandro Taddei, George B Moody, Michele Emdin, G Antolič, Roman Dorn, Ales Smrdel, Carlo Marchesi, and Roger G Mark. Long-term st database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. *Medical and Biological Engineering and Computing*, 41(2):172–182, 2003.
- [12] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.
- [13] George B Moody, W Muldrow, and Roger G Mark. A noise stress test for arrhythmia detectors. *Computers in cardiology*, 11(3):381–384, 1984.
- [14] Rochelle L Goldsmith, J Thomas Bigger, Richard C Steinman, and Joseph L Fleiss. Comparison of 24-hour parasympathetic activity in endurance-trained and untrained young men. *Journal of the American College of Cardiology*, 20(3):552–558, 1992.
- [15] Nikhil Iyengar, CK Peng, Raymond Morin, Ary L Goldberger, and Lewis A Lipsitz. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 271(4):R1078–R1084, 1996.
- [16] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair EW Johnson, et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37(12):2181, 2016.
- [17] R Bousseljot, D Kreiseler, and A Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995.
- [18] Pablo Laguna, Roger G Mark, A Goldberg, and George B Moody. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg. In *Computers in cardiology 1997*, pages 673–676. IEEE, 1997.
- [19] Scott David Greenwald, Ramesh S Patil, and Roger G Mark. Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms



- using contextual information. In *Computers in Cardiology 1990, Proceedings.*, pages 461–464. IEEE, 1990.
- [20] Ikaro Silva, Joachim Behar, Reza Sameni, Tingting Zhu, Julien Oster, Gari D Clifford, and George B Moody. Noninvasive fetal ecg: the physionet/computing in cardiology challenge 2013. In *Computing in Cardiology Conference (CinC), 2013*, pages 149–152. IEEE, 2013.
- [21] Meng-Hsi Wu and Edward Y Chang. Deepq arrhythmia database: A large-scale dataset for arrhythmia detector evaluation. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, pages 77–80. ACM, 2017.
- [22] Edward Choi, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, 24(2):361–370, 2016.
- [23] Han CW Hsiao, Sean HF Chen, and Jeffrey JP Tsai. Deep learning for risk analysis of specific cardiovascular diseases using environmental data and outpatient records. In *Bioinformatics and Bioengineering (BIBE), 2016 IEEE 16th International Conference on*, pages 369–372. IEEE, 2016.
- [24] Shruthi Gopalswamy, Patrick J Tighe, and Parisa Rashidi. Deep recurrent neural networks for predicting intraoperative and postoperative outcomes and trends. In *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*, pages 361–364. IEEE, 2017.
- [25] Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmarking deep learning models on large healthcare datasets. *Journal of biomedical informatics*, 83:112–134, 2018.
- [26] You Jin Kim, Yun-Geun Lee, Jeong Whun Kim, Jin Joo Park, Borim Ryu, and Jung-Woo Ha. Highrisk prediction from electronic medical records via deep attention networks. *arXiv preprint arXiv:1712.00010*, 2017.
- [27] Jaekwon Kim, Ungu Kang, and Youngho Lee. Statistics and deep belief network-based cardiovascular risk prediction. *Healthcare informatics research*, 23(3):169–175, 2017.
- [28] Zhengxing Huang, Wei Dong, Huilong Duan, and Jiquan Liu. A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records. *IEEE Transactions on Biomedical Engineering*, 65(5):956–968, 2018.

- [29] Eric C Polley and Mark J Van Der Laan. Super learner in prediction. 2010.
- [30] Shameer Faziludeen and PV Sabiq. Ecg beat classification using wavelets and svm. In *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, pages 815–818. IEEE, 2013.
- [31] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2016.
- [32] Ali Isin and Selen Ozdalili. Cardiac arrhythmia detection using deep learning. *Procedia Computer Science*, 120:268–275, 2017.
- [33] Zhiyong Wu, Xiangqian Ding, Guangrui Zhang, Xiaowei Xu, Xiaodong Wang, Ye Tao, and Chuanxiang Ju. A novel features learning method for ecg arrhythmias using deep belief networks. In *Digital Home (ICDH), 2016 6th International Conference on*, pages 192–196. IEEE, 2016.
- [34] Muhammad Zubair, Jinsul Kim, and Changwoo Yoon. An automated ecg beat classification system using convolutional neural networks. In *IT Convergence and Security (ICITCS), 2016 6th International Conference on*, pages 1–5. IEEE, 2016.
- [35] Dan Li, Jianxin Zhang, Qiang Zhang, and Xiaopeng Wei. Classification of ecg signals based on 1d convolution neural network. In *e-Health Networking, Applications and Services (Healthcom), 2017 IEEE 19th International Conference on*, pages 1–6. IEEE, 2017.
- [36] Kan Luo, Jianqing Li, Zhigang Wang, and Alfred Cuschieri. Patient-specific deep architectural model for ecg classification. *Journal of healthcare engineering*, 2017, 2017.
- [37] Chaoyan Jiang, Shuang Song, and Max Q-H Meng. Heartbeat classification system based on modified stacked denoising autoencoders and neural networks. In *Information and Automation (ICIA), 2017 IEEE International Conference on*, pages 511–516. IEEE, 2017.
- [38] Jianli Yang, Yang Bai, Feng Lin, Ming Liu, Zengguang Hou, and Xiuling Liu. A novel electrocardiogram arrhythmia classification method based on stacked sparse auto-encoders and softmax regression. *International Journal of Machine Learning and Cybernetics*, pages 1–8, 2017.

- 
- [39] Almir Badnjević, Mario Cifrek, Ratko Magjarević, and Zijad Džemić. *Inspection of Medical Devices: For Regulatory Purposes*. Springer, 2017.
- [40] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.
- [41] Meng-Hsi Wu, Emily J Chang, and Tzu-Hsuan Chu. Personalizing a generic ecg heartbeat classification for arrhythmia detection: A deep learning approach. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 92–99. IEEE, 2018.
- [42] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65, 2019.
- [43] U Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Information sciences*, 405:81–90, 2017.
- [44] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, U Raghavendra, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Yuki Hagiwara. Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. *Future Generation Computer Systems*, 79:952–959, 2018.
- [45] U Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Muhammad Adam, Jen Hong Tan, and Chua Kuang Chua. Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network. *Knowledge-Based Systems*, 132:62–71, 2017.
- [46] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, and Ru San Tan. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ecg signals. *Applied Intelligence*, pages 1–12, 2018.
- [47] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals. *Information Sciences*, 415:190–198, 2017.

- [48] Peng Xiong, Hongrui Wang, Ming Liu, and Xiuling Liu. Denoising autoencoder for electrocardiogram signal enhancement. *Journal of Medical Imaging and Health Informatics*, 5(8):1804–1810, 2015.
- [49] Bahareh Taji, Adrian DC Chan, and Shervin Shirmohammadi. False alarm reduction in atrial fibrillation detection using deep belief networks. *IEEE Transactions on Instrumentation and Measurement*, 2017.
- [50] Patrick Schwab, Gaetano C Scebba, Jia Zhang, Marco Delai, and Walter Karlen. Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks. *Computing*, 44:1, 2017.
- [51] Zhenjie Yao, Zhiyong Zhu, and Yixin Chen. Atrial fibrillation detection by multi-scale convolutional neural networks. In *Information Fusion (Fusion), 2017 20th International Conference on*, pages 1–6. IEEE, 2017.
- [52] Yong Xia, Naren Wulan, Kuanquan Wang, and Henggui Zhang. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in biology and medicine*, 93:84–92, 2018.
- [53] Rasmus S Andersen, Abdolrahman Peimankar, and Sadasivan Puthusserypady. A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications*, 2018.
- [54] Ran Xiao, Yuan Xu, Michele M Pelter, Richard Fidler, Fabio Badilini, David W Mortara, and Xiao Hu. Monitoring significant st changes through deep learning. *Journal of electrocardiology*, 2018.
- [55] Mohamad M Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and Ronald R Yager. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, 345:340–354, 2016.
- [56] Hedayat Abrishami, Matthew Campbell, Chia Han, Richard Czosek, and Xuefu Zhou. P-qrs-t localization in ecg using deep learning. In *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on*, pages 210–213. IEEE, 2018.
- [57] JF Wu, YL Bao, Shing-Chow Chan, HC Wu, L Zhang, and Xi-Guang Wei. Myocardial infarction detection and classification - a new multi-scale deep feature learning approach. In *Digital Signal Processing (DSP), 2016 IEEE International Conference on*, pages 309–313. IEEE, 2016.

- 
- [58] Bosun Hwang, Jiwoo You, Thomas Vaessen, Inez Myin-Germeys, Cheolsoo Park, and Byoung-Tak Zhang. Deep ecgnet: An optimal deep learning framework for monitoring mental stress using ultra short-term ecg signals. *TELEMEDICINE and e-HEALTH*, 2018.
- [59] Tahsin Reasat and Celia Shahnaz. Detection of inferior myocardial infarction using shallow convolutional neural networks. In *Humanitarian Technology Conference (R10-HTC), 2017 IEEE Region 10*, pages 718–721. IEEE, 2017.
- [60] Wei Zhong, Lijuan Liao, Xuemei Guo, and Guoli Wang. A deep learning approach for fetal qrs complex detection. *Physiological measurement*, 39(4):045004, 2018.
- [61] Vicent J Ribas Ripoll, Anna Wojdel, Enrique Romero, Pablo Ramos, and Josep Brugada. Ecg assessment based on neural networks with pretraining. *Applied Soft Computing*, 49:399–406, 2016.
- [62] Linpeng Jin and Jun Dong. Classification of normal and abnormal ecg records using lead convolutional neural network and rule inference. *Science China Information Sciences*, 60(7):078103, 2017.
- [63] Yixiu Liu, Yujuan Huang, Jianyi Wang, Li Liu, and Jiajia Luo. Detecting premature ventricular contraction in children with deep learning. *Journal of Shanghai Jiaotong University (Science)*, 23(1):66–73, 2018.
- [64] Jonathan Rubin, Rui Abreu, Anurag Ganguli, Saigopal Nelaturi, Ion Matei, and Kumar Sricharan. Recognizing abnormal heart sounds using deep learning. *arXiv preprint arXiv:1707.04642*, 2017.
- [65] Dariusz Kucharski, Dominik Grochala, Marcin Kajor, and Eliaz Kańtoch. A deep learning approach for valve defect recognition in heart acoustic signal. In *International Conference on Information Systems Architecture and Technology*, pages 3–14. Springer, 2017.
- [66] Juan P Dominguez-Morales, Angel F Jimenez-Fernandez, Manuel J Dominguez-Morales, and Gabriel Jimenez-Moreno. Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. *IEEE transactions on biomedical circuits and systems*, 12(1):24–34, 2018.
- [67] Angel Jiménez-Fernández, Elena Cerezuela-Escudero, Lourdes Miró-Amarante, Manuel Jesus Domínguez-Morales, Francisco Gomez-Rodriguez, Alejandro

- Linares-Barranco, and Gabriel Jiménez-Moreno. A binaural neuromorphic auditory sensor for fpga: A spike signal processing approach. *IEEE Trans. Neural Netw. Learning Syst.*, 28(4):804–818, 2017.
- [68] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *Computing in Cardiology Conference (CinC), 2016*, pages 621–624. IEEE, 2016.
- [69] Heechang Ryu, Jinkyoo Park, and Hayong Shin. Classification of heart sound recordings using convolution neural network. In *Computing in Cardiology Conference (CinC), 2016*, pages 1153–1156. IEEE, 2016.
- [70] Tien-En Chen, Shih-I Yang, Li-Ting Ho, Kun-Hsi Tsai, Yu-Hsuan Chen, Yun-Fan Chang, Ying-Hui Lai, Syu-Siang Wang, Yu Tsao, and Chau-Chung Wu. S1 and s2 heart sound recognition using deep neural networks. *IEEE Transactions on Biomedical Engineering*, 64(2):372–380, 2017.
- [71] Fan Pan, Peiyu He, Chengyu Liu, Taiyong Li, Alan Murray, and Dingchang Zheng. Variation of the korotkoff stethoscope sounds during blood pressure measurement: Analysis using a convolutional neural network. *IEEE journal of biomedical and health informatics*, 21(6):1593–1598, 2017.
- [72] Supreeth Prajwal Shashikumar, Amit J Shah, Qiao Li, Gari D Clifford, and Shamim Nemati. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*, pages 141–144. IEEE, 2017.
- [73] George S Everly Jr and Jeffrey M Lating. *A clinical guide to the treatment of the human stress response*. Springer Science & Business Media, 2012.
- [74] Soojeong Lee and Joon-Hyuk Chang. Deep learning ensemble with asymptotic techniques for oscillometric blood pressure estimation. *Computer methods and programs in biomedicine*, 151:1–13, 2017.
- [75] Soojeong Lee and Joon Hyuk Chang. Oscillometric blood pressure estimation based on deep learning. *IEEE Transactions on Industrial Informatics*, 13(2):461–472, 2017.
- [76] Soojeong Lee and Joon-Hyuk Chang. Deep belief networks ensemble for blood pressure estimation. *IEEE Access*, 5:9962–9972, 2017.

- 
- [77] Soojeong Lee and Joon Hyuk Chang. Deep boltzmann regression with mimic features for oscillometric blood pressure estimation. *IEEE Sensors Journal*, 17(18):5982–5993, 2017.
- [78] Igor Gotlibovych, Stuart Crawford, Dileep Goyal, Jiaqi Liu, Yaniv Kerem, David Benaron, Defne Yilmaz, Gregory Marcus, and Yihan Li. End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables. *arXiv preprint arXiv:1807.10707*, 2018.
- [79] Ming-Zher Poh, Yukkee Cheung Poh, Pak-Hei Chan, Chun-Ka Wong, Louise Pun, Wangie Wan-Chiu Leung, Yu-Fai Wong, Michelle Man-Ying Wong, Daniel Wai-Sing Chu, and Chung-Wah Siu. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart*, pages heartjnl–2018, 2018.
- [80] Brandon Ballinger, Johnson Hsieh, Avesh Singh, Nimit Sohoni, Jack Wang, Geoffrey H Tison, Gregory M Marcus, Jose M Sanchez, Carol Maguire, Jeffrey E Olgin, et al. Deepheart: Semi-supervised sequence learning for cardiovascular risk prediction. *arXiv preprint arXiv:1802.02511*, 2018.





## 4. Βαθιά μάθηση με απεικονίσεις

### 4.1. Εισαγωγή

Οι μέθοδοι απεικόνισης που έχουν βρει χρήση στην καρδιολογία περιλαμβάνουν τη τομογραφία μαγνητικού συντονισμού (Magnetic Resonance Imaging, MRI), την Fundus, την ηλεκτρονική τομογραφία (Computerized Tomography, CT), το ηχοκαρδιογράφημα, την τομογραφία οπτικής συνοχής (Optical Coherence Tomography, OCT), το ενδοαγγειακό υπερηχογράφημα (Intravascular Ultrasound, IVUS) και άλλες. Η βαθιά μάθηση υπήρξε ως επί το πλείστον επιτυχής σε αυτόν τον τομέα, κυρίως λόγω αρχιτεκτονικών που χρησιμοποιούν μεγάλο αριθμό συνελικτικών επιπέδων. Μια σύνοψη των εφαρμογών βαθιάς μάθησης που χρησιμοποιούν απεικονίσεις παρουσιάζονται στους Πίνακες 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 και 4.10.

### 4.2. Τομογραφία μαγνητικού συντονισμού

Η τομογραφία μαγνητικού συντονισμού (MRI) βασίζεται στην αλληλεπίδραση μεταξύ ενός συστήματος ατομικών πυρήνων και ενός εξωτερικού μαγνητικού πεδίου παρέχοντας μια εικόνα του εσωτερικού ενός φυσικού αντικειμένου [25]. Οι κύριες χρήσεις της MRI περιλαμβάνουν την κατάτμηση της αριστερής κοιλίας (Left Ventricle, LV), της δεξιάς κοιλίας (Right Ventricle, RV) και ολόκληρης της καρδιάς.

#### 4.2.1. Κατάτμηση LV

Τα CNNs χρησιμοποιήθηκαν για κατάτμηση LV με MRI. Οι Tan et al. [26] χρησιμοποίησαν ένα CNN για να εντοπίσουν το ενδοκάρδιο του LV και ένα επιπλέον CNN για να προσδιορίσουν την ακτίνα του ενδοκαρδίου, χρησιμοποιώντας τις STA11 και SUN09 για εκπαίδευση και αξιολόγηση αντίστοιχα. Χωρίς φιλτράρισμα των εικόνων που απεικονίζουν τα apical ή την χρήση παραμορφώσιμων μοντέλων επιτυχάνουν συγκρίσιμες επιδόσεις με προηγούμενες μεθόδους. Στο [27] οι συγγραφείς εκπαίδευσαν ένα CNN πέντε επιπέδων χρησιμοποιώντας MRI από την SUN09. Εκπαίδευσαν το μοντέλο τους χρησιμοποιώντας SGD και RMSprop με το πρώτο να φτάνει Dice 92%.

Table 4.1: Δημόσιες καρδιολογικές βάσεις δεδομένων, MRI

Βάση Δεδομένων	Ακρωνύμιο	Ασθενείς	Πρόβλημα
MICCAI 2009 Sunnybrook [1]	SUN09	45	κατάπτωση LV
MICCAI 2011 Left Ventricle Segmentation STACOM [2]	STA11	200	κατάπτωση LV
MICCAI 2012 Right Ventricle Segmentation Challenge [3]	RV12	48	κατάπτωση RV
MICCAI 2013 SATA [4]	SAT13	— <sup>a</sup>	κατάπτωση LV
MICCAI 2016 HVSMR [5]	HVS16	20	κατάπτωση καρδιάς
MICCAI 2017 ACDC [6]	AC17	150	κατάπτωση LV/RV
York University Database [7]	YUDB	33	κατάπτωση LV
Data Science Bowl Cardiac Challenge Data [8]	DS16	1140	εκτίμηση όγκου LV μετά την συστολή και διαστολή

<sup>a</sup> Ο αριθμός των ασθενών δεν αναφέρεται.

Table 4.2: Δημόσιες καρδιολογικές βάσεις δεδομένων, Fundus και άλλες

Βάση Δεδομένων	Ακρωνύμιο	Ασθενείς	Πρόβλημα
<b>Βάσεις δεδομένων αμφιβληστροειδούς (όλα Fundus εκτός του [9])</b>			
Digital Retinal Images for Vessel Extraction [10]	DRIVE	40	κατάκτηση αγγείων
Structured Analysis of the Retina [11]	STARE	20	κατάκτηση αγγείων
Child Heart and Health Study in England Database [12]	CHDB	14	κατάκτηση αγγείων
High Resolution Fundus [13]	HRF	45	κατάκτηση αγγείων
Kaggle Retinopathy Detection Challenge 2015 [14]	KRI15	-- <sup>b</sup>	ταξινόμηση διαβητικής αμφιβληστροειδοπάθειας
TeleOptica [15]	e-optha	381	MA και ανίχνευση αιμορραγίας
Messidor [16]	Messidor	1200	διάγνωση διαβητικής αμφιβληστροειδοπάθειας
Messidor2 [16]	Messidor2	874	διάγνωση διαβητικής αμφιβληστροειδοπάθειας
DiaRetdb1 [17]	DIA	89	MA και ανίχνευση αιμορραγίας
Retinopathy Online Challenge [18]	ROC	100	ανίχνευση MA
IOSTAR [9]	IOSTAR	30	κατάκτηση αγγείων με χρήση SLO
RC-SLO [9]	RC-SLO	40	κατάκτηση αγγείων με χρήση SLO
<b>Άλλες βάσεις δεδομένων απεικόνισης</b>			
MICCAI 2011 Lumen+External Elastic Laminae [19]	IV11	32	κατάκτηση περιγράμματος lumen, external IVUS
UK Biobank [20]	UKBDB	-- <sup>a</sup>	πολλαπλές βάσεις δεδομένων απεικόνισης
Coronary Artery Stenoses Detection and Quantification [21]	CASDQ	48	αγγειογραφία CT για στένωση κορωνιαίων αρτηριών
<b>Πολυτροπικές βάσεις δεδομένων</b>			
VORTAL [22]	VORTAL	45	εκτίμηση ρυθμού αναπνοής με χρήση ECG και PCG
Left Atrium Segmentation Challenge STACOM 2013 [23]	STA13	30	κατάκτηση αριστερού καρδιακού κόλπου με MRI, CT
MICCAI MMWHS 2017 [24]	MM17	60	120 εικόνες για κατάκτηση της καρδιάς με MRI, CT

<sup>a</sup>Ο αριθμός των ασθενών δεν αναφέρεται.

Table 4.3: Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάρτιση LV

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Dice <sup>b</sup>
Tan 2016 [26]	CNN	CNN για εύρεση τοποθεσίας και CNN για ευθυγράμμιση ενδοκαρδιακών συνόρων (SUN09, STA11)	88%
Romaguera 2017 [27]	CNN	CNN πέντε επιπέδων με SGD και RMSprop (SUN09)	92%
Poudel 2016 [28]	u-net, RNN	συνδυασμός u-net και RNN (SUN09, μη-δημόσια)	Πολλαπλά
Rupprecht 2016 [29]	CNN	συνδυασμός ενός CNN τεσσάρων επιπέδων με Sobolev (STA11, non-medical)	85%
Ngo 2014 [30]	DBN	συνδυασμός DBN με level set (SUN09)	88%
Aveni 2016 [31]	CNN, AE	CNN για εντοπισμό καρδιακού θάλαμου, AEs για συμπιεστικό σχήματος (SUN09)	96.69%
Yang 2016 [32]	CNN	δίκτυο εξαγωγής χαρακτηριστικών και ένα μη-τοπικό δίκτυο σύντηξης επιστημάτων (SAT13)	81.6%
Lu 2016 [33]	CNN	άτλας αντιστοίχισης του LV και ένα CNN τριών επιπέδων (DS16)	4.98% <sup>c*</sup>
Yang 2017 [34]	CNN, u-net	εντοπισμός με CNN παλινδρόμησης και κατάρτιση με u-net (YUDB, SUN09)	91%, 93% <sup>§</sup>
Tan 2017 [35]	CNN	CNN παλινδρόμησης (STA11, DS16)	Πολλαπλά
Curiale 2017 [36]	u-net	residual u-net (SUN09)	90%
Liao 2017 [37]	CNN	τοπικό δυαδικό μοτίβο για εντοπισμό και FCN για κατάρτιση (DS16)	4.69% <sup>c*</sup>
Emad 2015 [38]	CNN	εντοπισμός LV με χρήση CNN και πυραμίδες χλιμάκων (YUDB)	98.66% <sup>-</sup>

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>το (§) υποδηλώνει 'για κάθε βάση', το (\*) υποδηλώνει μέσω τετραγωνικό σφάλμα για EF το (+) υποδηλώνει 'για ενδοκαρδιακά και επικαρδιακά', το (-) υποδηλώνει ακρίβεια, και το (#) υποδηλώνει 'για CT και MRI'

Χρησιμοποιήθηκαν επίσης συνδυασμοί CNN με RNNs. Στο [28] οι συγγραφείς δημιούργησαν ένα recurrent u-net που μαθαίνει αναπαραστάσεις από μια στοίβα από 2D και έχει την ικανότητα να αξιοποιεί τις χωρικές εξαρτήσεις μεταξύ των τμημάτων μέσω εσωτερικών μονάδων μνήμης. Συνδυάζει ανίχνευση ανατομίας και κατάτμηση σε μια ενιαία αρχιτεκτονική από-άκρο-σε-άκρο, επιτυγχάνοντας συγκρίσιμα αποτελέσματα με άλλες μεθόδους, ξεπερνώντας τις βάσεις αναφοράς για το DBN, τα recurrent DBN και FCN όσον αφορά το Dice.

Άλλες δημοσιεύσεις συνδυάζουν μεθόδους βαθιάς μάθησης με το level-set για την κατάτμηση της LV. Οι Rupprecht et al. [29] εκπαίδευσαν ένα CNN τεσσάρων επιπέδων, που προβλέπει ένα διάνυσμα που υποδεικνύει το σημείο στο εξελισσόμενο περιγράμμα προς το πλησιέστερο σημείο στο όριο του αντικειμένου ενδιαφέροντος. Αυτές οι προβλέψεις σχημάτισαν ένα διανυσματικό πεδίο το οποίο στη συνέχεια χρησιμοποιήθηκε για την εξέλιξη του περιγράμματος, χρησιμοποιώντας το πλαίσιο ενεργού περιγράμματος Sobolev. Οι Anh et al. [30] δημιούργησαν μια μέθοδο μη-άκαμπτης κατάτμησης βασισμένη στη level-set μέθοδο ρυθμισμένη ανάλογα με την απόσταση, που αρχικοποιήθηκε από τα αποτελέσματα μιας δομημένης εξαγωγής από ένα DBN. Οι Avendi et al. [31] χρησιμοποίησαν ένα CNN για να ανιχνεύσουν τον LV θάλαμο και στη συνέχεια χρησιμοποίησαν στοιβαγμένα AE για να συμπεράνουν το σχήμα της LV. Στη συνέχεια το αποτέλεσμα ενσωματώθηκε σε παραμορφώσιμα μοντέλα για να βελτιωθεί η ακρίβεια και η ευρωστία της κατάτμησης.

Μέθοδοι που βασίζονται σε άτλαντες έχουν επίσης χρησιμοποιηθεί για την επίλυση αυτού του προβλήματος. Οι Yang et al. [32] δημιούργησαν ένα δίκτυο βαθιάς σύντηξης συνδυάζοντας ένα δίκτυο εξαγωγής χαρακτηριστικών και ένα μη-τοπικό δίκτυο σύντηξης επισημάνσεων βασισμένο σε patch. Τα χαρακτηριστικά που δημιουργούνται κατά τη διάρκεια της μάθησης χρησιμοποιούνται περαιτέρω για τον ορισμό ενός μέτρου ομοιότητας για την επιλογή άτλα MRI. Σύγκριναν τη μέθοδος τους με την ψηφοφορία με πλειοψηφία, τη σύντηξη ετικετών με βάση τα patch, την αντιστοίχιση patch πολλαπλών ατλάντων και το SVM με επαυξημένα χαρακτηριστικά επιτυγχάνοντας καλύτερα αποτελέσματα όσον αφορά την ακρίβεια. Οι Luo et al. [33] υιοθέτησαν μια μέθοδο χαρτογράφησης άτλα LV για να επιτευχθεί ακριβής εντοπισμός με χρήση δεδομένων MRI από το DS16. Στη συνέχεια, ένα CNN τριών επιπέδων εκπαιδεύτηκε για την πρόβλεψη του όγκου LV, πετυχαίνοντας συγκρίσιμα αποτελέσματα με τους νικητές του διαγωνισμού ACDC 2017 με βάση αναφοράς την μέση τετραγωνική ρίζα της τελικής διαστολής και των τελικών συστολικών όγκων.

Μέθοδοι παλινδρόμησης έχουν χρησιμοποιηθεί για τον εντοπισμό της LV πριν την κατάτμηση της. Οι Yang et al. [34] πρώτα εντόπισαν το LV σε ολόκληρη την εικόνα χρησιμοποιώντας CNN παλινδρόμησης και στη συνέχεια το ταξινόμησαν μέσα στην περιοχή ενδιαφέροντος (Region of Interest, ROI), χρησιμοποιώντας μια αρχιτεκτονική

που βασίζεται στο u-net. Το μοντέλο τους επιτυγχάνει υψηλή ακρίβεια με καλές υπολογιστικές επιδόσεις κατά τη διάρκεια των συμπερασμών.

Επίσης χρησιμοποιήθηκαν διάφορες άλλες μέθοδοι για την κατάτμηση του LV. Οι Tan et al. [35] παραμετροποίησαν όλες τις εικόνες του προβλήματος της τμηματοποίησης του LV με βάση τις ακτινικές αποστάσεις μεταξύ του κεντρικού σημείου LV και των ενδοκαρδιακών και επικαρδιακών περιγραμμάτων στις πολικές συντεταγμένες. Στη συνέχεια, εκπαίδευσαν ένα CNN παλινδρόμησης στην STA11 για να συμπεραίνουν αυτές τις παραμέτρους και δοκίμασαν τη γενικευσιμότητα της μεθόδου στην DS16 παρουσιάζοντας καλά αποτελέσματα. Στο [36] οι συγγραφείς χρησιμοποίησαν την απόσταση Jaccard ως τη αντικειμενική συνάρτηση βελτιστοποίησης, ενσωματώνοντας μια residual στρατηγική μάθησης και εισάγοντας ένα επίπεδο κανονικοποίησης παρτίδας (batch normalization) για την εκπαίδευση ενός u-net. Αποδείχθηκε ότι αυτή η διαμόρφωση είχε καλύτερα αποτελέσματα από άλλα απλά u-nets σε σχέση με το δείκτη Dice. Στο άρθρο τους οι Liao et al. [37] ανίχνευσαν το ROI που περιείχε LV και έπειτα χρησιμοποίησαν FCN για να ταξινομήσουν τα LV μέσα στο ROI. Τα αποτελέσματα της 2D τμηματοποίησης ενσωματώθηκαν μεταξύ διαφορετικών εικόνων για την εκτίμηση του όγκου. Το μοντέλο εκπαιδεύτηκε εναλλάξ στην κατάτμηση του LV και στην εκτίμηση του όγκου, τοποθετώντας το τέταρτο στον διαγωνισμό του DS16. Οι Emad et al. [38] εντόπισαν το LV χρησιμοποιώντας ένα CNN και μια ανάλυση πυραμίδας κλιμάκων, για να λάβουν υπόψη διαφορετικά μεγέθη της καρδιάς με την YUDB. Πέτυχαν καλά αποτελέσματα, αλλά με σημαντικό κόστος υπολογισμού (10 δευτερόλεπτα ανά εικόνα κατά τη διάρκεια συμπερασμών).

#### 4.2.2. Κατάτμηση LV/RV

Μια βάση δεδομένων που χρησιμοποιήθηκε για την κατάτμηση LV/RV ήταν η MICCAI 2017 ACDC Challenge (AC17), που περιέχει εικόνες MRI από 150 ασθενείς χωρισμένες σε πέντε ομάδες (φυσιολογικοί, προηγούμενο MI, διαστολή της καρδιομυοπάθειας, υπερτροφική καρδιομυοπάθεια, μη-φυσιολογική RV). Οι Zotti et al. [39] χρησιμοποίησαν ένα μοντέλο που περιλαμβάνει μια καρδιακή μονάδα παλινδρόμησης κέντρου μάζας, που επιτρέπει την καταγραφή του σχήματος εκ των προτέρων και μια συνάρτηση απώλειας προσαρμοσμένη στην καρδιακή ανατομία. Τα χαρακτηριστικά δημιουργούνται με μια αρχιτεκτονική μετατροπής πολλαπλών αναλυτικών επιπέδων conv-deconv 'πλέγματος', η οποία αποτελεί επέκταση του u-net. Αυτό το μοντέλο σε σύγκριση με το απλό conv-deconv και το u-net, εμφανίζει καλύτερα αποτελέσματα κατά μέσο όρο 5% όσον αφορά τον δείκτη Dice. Οι Patravali et al. [40] εκπαίδευσαν ένα μοντέλο βασισμένο στο u-net, χρησιμοποιώντας το Dice σε συνδυασμό με το cross-entropy ως μέτρηση για την τμηματοποίηση των LV/RV και του μυοκαρδίου. Το μοντέλο σχεδιάστηκε για να δέχεται μια στοίβα εικόνων ως κανάλια εισόδου, ενώ

η έξοδος προβλέπει τη μεσαία εικόνα. Με βάση τα πειράματα που διεξήγαγαν, βγήκε το συμπέρασμα ότι τρεις εικόνες ήταν βέλτιστες ως είσοδος για το μοντέλο αντί για μια ή πέντε. Οι Isensee et al. [41] χρησιμοποίησαν ένα σύνολο 2D και 3D u-net για κατάτμηση των LV/RV και του μυοκαρδίου του LV σε κάθε επανάληψη του καρδιακού κύκλου. Πληροφορίες εξήχθησαν από την κατατμημένη χρονοσειρά με τη μορφή χαρακτηριστικών, που αντικατοπτρίζουν τις διαγνωστικές κλινικές διαδικασίες του σκοπού της ταξινόμησης. Με βάση αυτά τα χαρακτηριστικά, εκπαίδευσαν ένα ensemble perceptrons πολλαπλών επιπέδων και έναν ταξινομητή RF για την πρόβλεψη της παθολογικής κατηγορίας. Το μοντέλο τους κατέλαβε την πρώτη θέση στον διαγωνισμό του ACDC.

Διάφορες άλλες βάσεις δεδομένων έχουν επίσης χρησιμοποιηθεί για την επίλυση της κατάτμησης LV/RV με CNNs. Στο [43] οι συγγραφείς δημιούργησαν μια μέθοδο ημι-επιβλεπόμενης μάθησης, στην οποία το δίκτυο κατάτμησης για το LV/RV και το μυοκάρδιο εκπαιδεύτηκε από τα επισημασμένα και μη-επισημασμένα δεδομένα. Η αρχιτεκτονική του δικτύου βασίστηκε στο VGGnet16, παρόμοια με την αρχιτεκτονική του DeepLab [49], ενώ η τελική κατάτμηση βελτιώθηκε με τη χρήση ενός υπό όρους τυχαίου πεδίου (Conditional Random Field, CRF). Οι συγγραφείς καταδεικνύουν ότι η εισαγωγή μη-επισημασμένων δεδομένων, βελτιώνει την απόδοση τμηματοποίησης όταν τα δεδομένα εκπαίδευσης είναι λίγα. Στο [48] οι συγγραφείς υιοθετούν ένα 3D πολλαπλών κλιμάκων CNN για τον εντοπισμό των pixel που ανήκουν στον RV. Το δίκτυο έχει δύο συνελικτικές διαδρομές με τις εισόδους του κεντροθετημένες στην ίδια θέση εικόνας, ενώ το δεύτερο τμήμα εξάγεται από μια εκδοχή της εικόνας που έχει υποβληθεί σε δειγματοληψία. Τα αποτελέσματα που προέκυψαν ήταν καλύτερα από προηγούμενες μεθόδους, παρόλο που τα τελευταία βασίστηκαν σε χειροποίητα χαρακτηριστικά και εκπαιδεύτηκαν σε λιγότερο μεταβλητές βάσεις δεδομένων.

Τα FCNs έχουν επίσης χρησιμοποιηθεί για κατάτμηση LV/RV. Στο άρθρο τους, οι Tran et al. [42] εκπαίδευσαν ένα μοντέλο FCN τεσσάρων επιπέδων για την κατάτμηση των LV/RV με τις SUN09, STA11. Σύγκριναν προηγούμενες μεθόδους μαζί με δύο αρχικοποιήσεις του μοντέλου τους: μία fine-tuned έκδοση του μοντέλου τους χρησιμοποιώντας το STA11 και μια αρχικοποιημένη με Xavier, με το τελευταίο να έχει τις καλύτερες επιδόσεις σε σχεδόν όλα τα προβλήματα.

Τα FCN με skip-connections και u-net έχουν επίσης χρησιμοποιηθεί για την επίλυση αυτού του προβλήματος. Οι Lieman et al. [44] δημιούργησαν μια αρχιτεκτονική FCN με skip-connections με το όνομα FastVentricle βασισμένη στο ENet [45], η οποία είναι ταχύτερη και λειτουργεί με λιγότερη μνήμη από τις προηγούμενες αρχιτεκτονικές κοιλιακής κατάτμησης επιτυγχάνοντας υψηλή κλινική ακρίβεια. Στο [46] οι συγγραφείς εισάγουν το *u-net* το οποίο είναι μια παραλλαγή u-net για την κατάτμηση του ενδοκαρδίου και του επικαρδίου LV/RV με τη χρήση των DS16, SUN09 και RV12. Αυτή η μέθοδος απέδωσε καλύτερα από τον ειδικό καρδιολόγο σε

Table 4.4: Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάτμηση LV/RV

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Dice <sup>b</sup>
Zotti 2017 [39]	u-net	παράλλαξη του u-net με πολυ-χάμιξακωτό conv-deconv αρχιτεκτονική πλέγματος (AC17)	90%
Patravali 2017 [40]	u-net	2D/3D u-net (AC17)	πολλαπλά
Isensee 2017 [41]	u-net	ensemble u-net, με συστηματοποίηση πολυ-επιπέδων perception και τάζινομητή RF (AC17)	πολλαπλά
Tran 2016 [42]	CNN	FCN τεσσάρων επιπέδων (SUN09, ST11)	92%, 96%+
Bai 2017 [43]	CNN	VGGnet16 και DeepLab με χρήση CRF και πιο λεπτομερή αποτελέσματα (UKBDB)	90.3%
Lieman 2017 [44]	u-net	επέκταση του ENet [45] με skip-connections (μη-δημόσια)	πολλαπλά
Witther 2017 [46]	u-net	u-net παράλλαξη του u-net (DS16, SUN09, RV12, μη-δημόσια)	πολλαπλά
Du 2018 [47]	DBN	χαρακτηριστικά DAISY και DBN παλινδρόμησης με χρήση 2900 εικόνων (μη-δημόσια)	91.6%, 94.1%+
Giannakidis 2016 [48]	CNN	κατάτμηση RV με χρήση 3D πολυ-χάμιξακωτών CNN με δύο διαδοχικές (μη-δημόσια)	82.81%

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>το (§) υποδηλώνει 'για κάθε βάση', το (\*) υποδηλώνει μέσο τετραγωνικό σφάλμα για EF, το (+) υποδηλώνει 'για ενδοκαρδιακά και επικαρδιακά', το (-) υποδηλώνει ακρίβεια, και το (#) υποδηλώνει 'για CT και MRI'



αυτή τη μελέτη, ειδικά για την κατάτμηση του RV.

Ορισμένες άλλες μέθοδοι βασίστηκαν σε μοντέλα παλινδρόμησης. Στο άρθρο τους οι Du et al. [47] δημιούργησαν ένα πλαίσιο κατάτμησης παλινδρόμησης για να οριοθετήσουν το LV/RV. Πρώτον, εξάγονται χαρακτηριστικά DAISY και στη συνέχεια χρησιμοποιήθηκε μια μέθοδος αναπαράστασης βάσει σημείων, για την απεικόνιση των ορίων. Τέλος, τα χαρακτηριστικά DAISY χρησιμοποιήθηκαν ως είσοδος και τα σημεία ορίων ως ετικέτες για την εκπαίδευση του μοντέλου παλινδρόμησης με βάση το DBN. Η απόδοση του μοντέλου αξιολογείται χρησιμοποιώντας διαφορετικά χαρακτηριστικά από το DAISY (GIST, ιστόγραμμα πυραμίδων προσανατολισμένων διαβαθμίσεων) και επίσης συγκρίνεται με την παλινδρόμηση διανυσμάτων υποστήριξης (Support Vector Regression, SVR) και άλλες παραδοσιακές μεθόδους (γραφήματα, ενεργά περιγράμματα, level set), επιτυγχάνοντας καλύτερα αποτελέσματα.

### 4.2.3. Κατάτμηση της καρδιάς

Το MICCAI 2016 HVSMR (HVS16) χρησιμοποιήθηκε για την κατάτμηση της καρδιάς η οποία περιέχει εικόνες MRI από 20 ασθενείς. Οι Wolterink et al. [50] εκπαίδευσαν ένα CNN δέκα επιπέδων με αυξανόμενα επίπεδα διαστολής για την κατάτμηση του μυοκαρδίου και του αίματος στις αξονικές, σαγματοειδείς και στεφανιαίες εικόνες. Επίσης, χρησιμοποιούν βαθιά επίβλεψη [51] για να επιλύσουν το πρόβλημα του vanishing gradients και να βελτιώσουν την αποτελεσματικότητα της εκπαίδευσης του δικτύου τους χρησιμοποιώντας ένα μικρό σύνολο δεδομένων. Τα πειράματα που πραγματοποίησαν με και χωρίς διαστολές (dilations) σε αυτήν την αρχιτεκτονική έδειξαν τη χρησιμότητα της. Στο άρθρο τους οι Li et al. [52] ξεκίνησαν με ένα 3D FCN με επισήμανση των voxel και στη συνέχεια εισήγαγαν διαστελλόμενα συνελικτικά επίπεδα στο βασικό μοντέλο για να επεκτείνουν το δεκτικό πεδίο. Έπειτα χρησιμοποιούν μονοπάτια βαθιά επίβλεψης, για την επιταχύνουν την εκπαίδευση και την αξιοποίηση πληροφοριών πολλαπλών κλιμάκων. Σύμφωνα με τους συγγραφείς το μοντέλο παρουσιάζει καλή ακρίβεια κατάτμησης, σε συνδυασμό με χαμηλό υπολογιστικό κόστος. Οι Yu et al. [53] δημιούργησαν ένα φράκταλ δίκτυο 3D FCN για κατάτμηση της καρδιάς και των μεγάλων αγγείων. Εφαρμόζοντας αναδρομικά έναν κανόνα απλής επέκτασης, κατασκευάζουν το φράκταλ δίκτυο συνδυάζοντας ιεραρχικές ενδείξεις για ακριβή κατάτμηση. Επιτυγχάνουν επίσης καλά αποτελέσματα με χαμηλό υπολογιστικό κόστος (12 δευτερόλεπτα ανά όγκο).

Μια άλλη βάση δεδομένων που χρησιμοποιήθηκε για την κατάτμηση της καρδιάς ήταν η MM17 η οποία περιέχει 120 πολυτροπικές εικόνες από καρδιακή MRI/CT. Η μέθοδος των Payer et al. [54] βασίζεται σε δύο FCN για τον εντοπισμό και την κατάτμηση της καρδιάς. Αρχικά, το CNN εντοπισμού βρίσκει το κέντρο του πλαισίου οριοθέτησης γύρω από όλες τις δομές της καρδιάς, έτσι ώστε το CNN κατάτμησης να

μπορεί να επικεντρωθεί σε αυτήν την περιοχή. Έπειτα, το CNN κατάτμησης μετατρέπει τις προβλέψεις ενδιάμεσων ετικετών σε θέσεις άλλων ετικετών. Επομένως το δίκτυο μαθαίνει από τις σχετικές θέσεις μεταξύ των επισημάνσεων και επικεντρώνεται στις ανατομικά εφικτές διαμορφώσεις. Το μοντέλο συγκρίθηκε με το u-net επιτυγχάνοντας καλύτερα αποτελέσματα, ειδικά στο σύνολο δεδομένων του MRI. Οι Mortazi et al. [55] εκπαίδευσαν ένα πολυεπίπεδο CNN με μια προσαρμοστική στρατηγική σύντηξης, για την κατάτμηση επτά περιοχών της καρδιάς. Σχεδίασαν τρία CNN (ένα για κάθε κάθετο επίπεδο) με την ίδια αρχιτεκτονική και τα εκπαίδευσαν για επισήμανση των voxel. Από τα πειράματά τους καταλήγουν στο συμπέρασμα ότι το μοντέλο τους οριοθετεί τις καρδιακές δομές με υψηλή ακρίβεια και αποτελεσματικά. Στο [56] οι συγγραφείς χρησιμοποίησαν ένα FCN, το οποίο συνδύασαν με 3D τελεστές, μεταφορά μάθησης και έναν μηχανισμό βαθιάς επίβλεψης για την απόσταξη 3D συμφραζομένων πληροφοριών και την επίλυση πιθανών δυσκολιών στην εκπαίδευση. Χρησιμοποιήθηκε υβριδική απώλεια που καθοδηγεί τη διαδικασία εκπαίδευσης για την εξισορρόπηση των κατηγοριών και διατηρεί τις λεπτομέρειες των ορίων. Σύμφωνα με τα πειράματά τους, η χρήση της υβριδικής απώλειας επιτυγχάνει καλύτερα αποτελέσματα από το Dice μέτρο.

#### 4.2.4. Άλλες εφαρμογές

Μέθοδοι βαθιάς μάθησης έχουν επίσης χρησιμοποιηθεί και για ανίχνευση άλλων καρδιακών δομών με MRI. Οι Yang et al. [57] δημιούργησαν μια μέθοδο διάδοσης πολλαπλών ατλάντων, για να ενσωματώσουν την ανατομική δομή του μυοκαρδίου του αριστερού κόλπου και των πνευμονικών φλεβών. Αυτό ακολουθήθηκε από ένα μη-επιτηρούμενο εκπαιδευμένο SSAE με ένα softmax για την κατάτμηση της κολπικής ίνωσης, χρησιμοποιώντας 20 εικόνες από ασθενείς με AF. Στο άρθρο τους οι Zhang et al. [58] προσπάθησαν να ανιχνεύσουν εικόνες έλλειψης apical και basal. Ελέγχουν την παρουσία τυπικών basal και apical μοτίβων στις τελευταίες και πρώτες εικόνες της βάσης δεδομένων και εκπαιδεύουν δύο CNN για να κατασκευάσουν ένα σύνολο διακριτικών χαρακτηριστικών. Τα πειράματά τους έδειξαν ότι το μοντέλο με τέσσερα επίπεδα, έχει καλύτερη απόδοση από τις SAE και τις Βαθιές Μηχανές Boltzmann.

Άλλα ιατρικά προβλήματα με MRI μελετήθηκαν επίσης, όπως η ανίχνευση εικόνων ακραίας συστολής και διαστολής. Οι Kong et al. [59] δημιούργησαν ένα χρονικό δίκτυο παλινδρόμησης που είχε προεκπαιδευτεί στο ImageNet με την ενσωμάτωση ενός CNN με ένα RNN, για να προσδιορίσει τις εικόνες της τελικής διαστολής και της τελικής συστολής από τις ακολουθίες MRI. Το CNN κωδικοποιεί τις χωρικές πληροφορίες μίας καρδιακής αλληλουχίας ενώ το RNN αποκωδικοποιεί τις χρονικές. Επίσης σχεδίασαν μια συνάρτηση απώλειας για να περιορίσουν τη δομή των προβλεπόμενων επισημάνσεων. Το μοντέλο τους επιτυγχάνει καλύτερη μέση διαφορά εικόνων από τις προηγούμενες μεθόδους. Στο άρθρο τους οι Yang et al. [60] χρησιμοποίησαν ένα CNN

Table 4.5: Εφαρμογές βαθιάς μάθησης με χρήση MRI, για κατάρτιση της καρδιάς και άλλα

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	Dice <sup>b</sup>
<b>Τμηματοποίηση όλης της καρδιάς</b>			
Wolterink 2016 [50]	CNN	διατελλόμενα CNN με ορθογώνια patches (HVS16)	80%, 93%
Li 2016 [52]	CNN	βαθιά επιβλεπόμενο 3D FCN με διαστολές (HVS16)	69.5%
Yu 2017 [53]	CNN	βαθιά επιβλεπόμενο 3D FCN κατασκευασμένο με αυτο-επιβλεπόμενο φρακταλ τρόπο (HVS16)	πολλαπλά
Payer [54]	CNN	δύο ξεχωριστά FCN για εντοπισμό και κατάρτιση (MM17)	90.7%, 87%#
Mortazi 2017 [55]	CNN	πολυ-επίπεδο FCN (MM17)	90%, 85%#
Yang [56]	CNN	βαθιά επιβλεπόμενο 3D FCN εκπαιδευμένο με μεταφορά μάθησης (MM17)	84.3%, 77.8%#
<b>Άλλες εφαρμογές</b>			
Yang 2017 [57]	SSAE	κατάρτιση κολπικής ίνωσης με διάδοση πολλαπλών ατλάντων, SSAE, softmax (μη-δημόσια)	82%
Zhang 2016 [58]	CNN	αίχνηση έλλειψης apical και basal με δύο CNNs τεσσάρων επιπέδων (UKBDB)	πολλαπλά
Kong 2016 [59]	CNN, RNN	CNN για χωρική πληροφορία και RNN για χρονική πληροφορία	πολλαπλά
Yang 2017 [60]	CNN	CNN για την αίχνηση του τέλους της διαστολής και συστολής από το LV (STA11, μη-δημόσια)	76.5% <sup>-</sup>
Xu 2017 [61]	Multiple	αίχνηση MI με χρήση Fast R-CNN για εντοπισμό της καρδιάς, LSTM και SAE (μη-δημόσια)	94.3% <sup>-</sup>
Xue 2018 [62]	CNN, LSTM	CNN, δύο παράλληλα LSTMs και Bayesian μοντέλο για ποσοτικοποίηση του LV (μη-δημόσια)	πολλαπλά
Zhen 2016 [63]	RBM	πολυ-κλιμακωτό συνελκτικό RBM και RF για δι-κολπική εκτίμηση του όγκου (μη-δημόσια)	3.87%*
Biffi 2016 [64]	CNN	αίχνηση υπερτροφικής καρδιομυοπάθειας με χρήση μεταβολικών AE (AC17, μη-δημόσια)	90% <sup>-</sup>
Oktay 2016 [65]	CNN	αύξηση αναλυτικότητας εικόνας με residual CNN (μη-δημόσια)	πολλαπλά

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.<sup>b</sup>το (§) υποδηλώνει 'για κάθε βάση', το (\*) υποδηλώνει μέσο τετραγωνικό σφάλμα για EF το (+) υποδηλώνει 'για ενδοκαρδιακά και επικαρδιακά', το (-) υποδηλώνει αρέβεια, και το (#) υποδηλώνει 'για CT και MRI'

για να ανιχνεύσουν τις εικόνες της τελικής διαστολής και της τελικής συστολής από την LV, επιτυγχάνοντας ακρίβεια 76.5%.

Δημιουργήθηκαν επίσης μέθοδοι ποσοτικοποίησης διαφόρων καρδιαγγειακών χαρακτηριστικών. Στο [61] οι συγγραφείς εντοπίζουν τη θέση και το σχήμα του MI χρησιμοποιώντας ένα μοντέλο που αποτελείται από τρία επίπεδα; πρώτον, το επίπεδο εντοπισμού καρδιάς είναι ένα Fast R-CNN το οποίο απομονώνει τις ακολουθίες ROI συμπεριλαμβανομένου του LV; δεύτερον, τα στατιστικά επίπεδα κίνησης, τα οποία κατασκευάζουν μια αρχιτεκτονική χρονοσειράς για να καταγράψουν τα χαρακτηριστικά τοπικής κίνησης που παράγονται από το LSTM-RNN και τα χαρακτηριστικά κίνησης που παράγονται από βαθιές οπτικές ροές από την ακολουθία ROI; τρίτον, τα FNN διάκρισης, τα οποία χρησιμοποιούν το SAE για να μάθουν περαιτέρω τα χαρακτηριστικά από το προηγούμενο επίπεδο και τέλος έναν ταξινομητή softmax. Οι Xue et al. [62] εκπαίδευσαν ένα δίκτυο βαθιάς μάθησης πολλαπλών προβλημάτων, σε MRI από 145 άτομα με 20 εικόνες ο καθένας για πλήρη ποσοτικοποίηση του LV. Αποτελείται από ένα CNN τριών επιπέδων που εξάγει τις καρδιακές αναπαραστάσεις, και στη συνέχεια δύο παράλληλα LSTM-RNN για τη μοντελοποίηση της χρονικής δυναμικής των καρδιακών ακολουθιών. Τέλος, τοποθετείται ένα Bayesian πλαίσιο ικανό να μάθει τις σχέσεις μεταξύ των προβλημάτων και ένας ταξινομητής softmax. Εκτεταμένες συγκρίσεις με προηγούμενες μεθόδους δείχνουν την αποτελεσματικότητα αυτής της μεθόδου όσον αφορά το μέσο απόλυτο σφάλμα. Στο [63] οι συγγραφείς δημιούργησαν μια μέθοδο μη-επιβλεπόμενης μάθησης καρδιακής απεικόνισης χρησιμοποιώντας πολυεπίπεδο συνελικτικό RBM και μια άμεση εκτίμηση όγκου των δύο κοιλοτήτων χρησιμοποιώντας RF. Σύγκριναν το μοντέλο τους με ένα Bayesian μοντέλο, ένα μοντέλο βασισμένο σε χειροποίητα χαρακτηριστικά, τα level-set και την περικοπή γραφήματος, επιτυγχάνοντας καλύτερα αποτελέσματα από πλευράς συντελεστή συσχέτισης για όγκους LV/RV και εκτίμηση σφάλματος του EF.

Άλλες μέθοδοι χρησιμοποιήθηκαν επίσης για την ανίχνευση υπερτροφικής καρδιομυοκαρδιοπάθειας ή για την αύξηση της αναλυτικότητας των MRI. Οι Biffi et al. [64] εκπαίδευσαν ένα VAE για την ταυτοποίηση ασθενών με υπερτροφική μυοκαρδιοπάθεια χρησιμοποιώντας ένα ισορροπημένο σύνολο δεδομένων 1365 ασθενών και την AC17. Δείχνουν επίσης ότι το δίκτυο είναι σε θέση να απεικονίσει και να ποσοτικοποιήσει τα πρότυπα αναδιαμόρφωσης που είναι σχετικά με την παθολογία στον αρχικό χώρο εισόδου των εικόνων, αυξάνοντας έτσι την ερμηνευσιμότητα του μοντέλου. Στο [65] οι συγγραφείς δημιούργησαν μια μέθοδο επαύξησης της αναλυτικότητας της εικόνας βασισμένη σε residual CNN η οποία επιτρέπει τη χρήση δεδομένων εισόδου που αποκτήθηκαν από διαφορετικά επίπεδα προβολής για βελτιωμένη απόδοση. Σύγκριναν με άλλες μεθόδους παρεμβολής (γραμμική, spline, ταίριασμα patch πολλαπλών ατλάντων, ρηχό CNN, CNN), επιτυγχάνοντας καλύτερα

αποτελέσματα όσον αφορά το PSNR. Συγγραφείς από την ίδια ομάδα πρότειναν μια στρατηγική εκπαίδευσης [66] που ενσωματώνει ανατομική προηγούμενη γνώση σε CNNs μέσω ενός μοντέλου συστηματοποίησης, ενθαρρύνοντας την να ακολουθήσει την ανατομία μέσω μη-γραμμικών παραστάσεων του σχήματος.

#### 4.2.5. Συμπεράσματα της χρήσης βαθιάς μάθησης με MRI

Υπάρχει ένα ευρύ φάσμα αρχιτεκτονικών που έχουν εφαρμοστεί στα MRI. Οι περισσότερες είναι CNNs ή u-net οι οποίες είτε χρησιμοποιούνται αποκλειστικά είτε σε συνδυασμό με RNNs, AEs ή ensemble. Το πρόβλημα είναι ότι οι περισσότερες από αυτές δεν εκπαιδεύονται από-άχρο-σε-άχρο. Βασίζονται σε προεπεξεργασία, χρήση χειροποίητων χαρακτηριστικών, ενεργά περιγράμματα, level-set και σε άλλες μη-διαφοροποιήσιμες μεθόδους, χάνοντας έτσι μερικώς τη δυνατότητα κλιμάκωσης στην παρουσία νέων δεδομένων. Κύριος στόχος αυτού του τομέα πρέπει να είναι η δημιουργία μοντέλων από-άχρο-σε-άχρο, ακόμη και αν αυτό σημαίνει μικρότερη ακρίβεια βραχυπρόθεσμα; πιο αποδοτικές αρχιτεκτονικές θα μπορούσαν να καλύψουν το χάσμα στο μέλλον.

Ένα ενδιαφέρον εύρημα σχετικά με την κατάτμηση της καρδιάς έγινε στο [67] όπου οι συγγραφείς διερεύνησαν την καταλληλότητα των προηγούμενων 2D, 3D CNN αρχιτεκτονικών και των τροποποιήσεών τους. Διαπίστωσαν ότι η επεξεργασία ανά εικόνα χρησιμοποιώντας δίκτυα 2D ήταν καλύτερη λόγω του μεγάλου πάχους του τμήματος. Ωστόσο, η επιλογή της αρχιτεκτονικής δικτύου διαδραματίζει μικρό ρόλο.

### 4.3. Fundus

Η απεικόνιση Fundus είναι ένα κλινικό εργαλείο για την αξιολόγηση της αμφιβληστροειδοπάθειας σε ασθενείς στην οποία η ένταση αντιπροσωπεύει την ποσότητα του ανακλώμενου φωτός συγκεκριμένης ζώνης κυμάτων [68]. Μία από τις πιο ευρέως διαδεδομένες βάσεις δεδομένων στο Fundus είναι η DRIVE, η οποία περιέχει 40 εικόνες και τις αντίστοιχες επισημάνσεις της μάσκας των αγγείων.

#### 4.3.1. Κατάτμηση αγγείων

Τα CNN χρησιμοποιήθηκαν για την κατάτμηση αγγείων σε εικόνες Fundus. Στο [69] οι συγγραφείς αρχικά χρησιμοποίησαν ισορροπία ιστογράμματος και φιλτράρισμα Gauss για τη μείωση του θορύβου. Στη συνέχεια χρησιμοποιήθηκε ένα CNN τριών επιπέδων ως εξαγωγέας χαρακτηριστικών και ένα RF ως ταξινομητής. Σύμφωνα με τα πειράματά τους, η καλύτερη απόδοση επιτεύχθηκε από ένα ensemble νικητής-τα-παίρνει-όλα, σε σύγκριση με ένα μέσο, σταθμισμένο και διάμεσο ensemble.

Οι Zhou et al. [70] εφάρμοσαν προεπεξεργασία εικόνας για την εξάλειψη των ισχυρών άκρων γύρω από το οπτικό πεδίο και κανονικοποίησαν την φωτεινότητα και την αντίθεση μέσα σε αυτό. Στη συνέχεια, εκπαίδευσαν ένα CNN για να παράξουν χαρακτηριστικά για γραμμικά μοντέλα και εφάρμοσαν φίλτρα για την ενίσχυση των λεπτών αγγείων, μειώνοντας τη διαφορά έντασης μεταξύ λεπτών και ευρέων αγγείων. Στη συνέχεια ένας πυκνός CRF προσαρμόστηκε για να επιτευχθεί η τελική κατάτμηση του αγγείου, λαμβάνοντας τα διακριτικά χαρακτηριστικά για μοναδιαία δυναμικά και την εικόνα με τα ενισχυμένα λεπτά αγγεία. Μεταξύ των αποτελεσμάτων τους, στα οποία παρουσιάζουν μεγαλύτερη ακρίβεια από τις περισσότερες μεθόδους τελευταίας τεχνολογίας, παρέχουν επίσης στοιχεία υπέρ της χρήσης των πληροφοριών RGB του Fundus αντί για χρήση μόνο του πράσινου καναλιού. Οι Chen et al. [71] σχεδίασαν ένα σύνολο κανόνων για τη δημιουργία τεχνητών δειγμάτων εκπαίδευσης με πρότερη γνώση και χωρίς χειροκίνητη επισήμανση. Εκπαίδευσαν ένα FCN με ένα skip-connection που επιτρέπει την υψηλού επιπέδου πληροφορία να καθοδηγεί την εργασία σε χαμηλότερα επίπεδα. Αξιολογούν το μοντέλο τους στις DRIVE και STARE, επιτυγχάνοντας συγκρίσιμα αποτελέσματα με άλλες μεθόδους που χρησιμοποιούν πραγματική επισήμανση. Στο [72] οι συγγραφείς εκπαίδευσαν ένα ensemble 12 CNNs με τρία επίπεδα το καθένα στη DRIVE, όπου κατά τη διάρκεια του συμπερασμού οι απαντήσεις των CNNs υπολογίζονται κατά μέσο όρο για να σχηματίσουν την τελική κατάτμηση. Δείχνουν ότι το μοντέλο τους επιτυγχάνει υψηλότερη μέση ακρίβεια από τις προηγούμενες μεθόδους. Οι Fu et al. [73] εκπαιδεύουν ένα CNN στις DRIVE και STARE δημιουργώντας ένα χάρτη πιθανοτήτων και έπειτα χρησιμοποιούν έναν πλήρως συνδεδεμένο CRF για να συνδυάσουν τους χάρτες πιθανότητας και τις αλληλεπιδράσεις μεγάλης εμβέλειας μεταξύ των εικονοστοιχείων. Στο [74] οι συγγραφείς χρησιμοποίησαν ένα CNN για να μάθουν τα χαρακτηριστικά και μια αναζήτηση πλησιέστερων γειτόνων βασισμένη στο PCA που χρησιμοποιήθηκε για την εκτίμηση της τοπικής κατανομής δομών. Εκτός από την παρουσίαση καλών αποτελεσμάτων υποστηρίζουν ότι είναι σημαντικό για το CNN να ενσωματώσει πληροφορίες σχετικά με τη δομή δέντρων όσον αφορά την ακρίβεια.

Τα AE έχουν επίσης χρησιμοποιηθεί για την κατάτμηση των αγγείων. Οι Li et al. [75] εκπαίδευσαν ένα FNN και ένα AE αποθρομβοποίησης με τις DRIVE, STARE και CHDB. Υποστηρίζουν ότι τα χαρακτηριστικά του μοντέλου τους είναι πιο ανθεκτικά στο θόρυβο και τις διαφορετικές συνθήκες απεικόνισης, επειδή η διαδικασία μάθησης εκμεταλλεύεται τα χαρακτηριστικά των αγγείων σε όλες τις εικόνες εκπαίδευσης. Στο [76] οι συγγραφείς χρησιμοποίησαν μη-επιβλεπόμενα ιεραρχικά χαρακτηριστικά χρησιμοποιώντας ένα ensemble δύο επιπέδων SDAE. Το επίπεδο εκπαίδευσης εξασφαλίζει αποσύνδεση και το επίπεδο του ensemble εξασφαλίζει την αρχιτεκτονική αναθεώρηση. Δείχνουν επίσης ότι η εκπαίδευση του ensemble των AE

Table 4.6: Εφαρμογές βαθιάς μάθησης με χρήση Fundus, για κατάρτιση αγγείων

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	AUC <sup>b</sup>
Wang 2015 [69]	CNN, RF	CNN τριών επιπέδων συνδυασμένα με ensemble RF (DRIVE, STARE)	0.9475
Zhou 2017 [70]	CNN, CRF	CNN για εξαγωγή χαρακτηριστικών και CRF για τελικό αποτέλεσμα (DRIVE, STARE, CHDB)	0.7942
Chen 2017 [71]	CNN	τεχνητά δεδομένα, FCN (DRIVE, STARE)	0.9516
Maji 2016 [72]	CNN	12 CNNs ensemble με τρία επίπεδα (DRIVE)	0.9283
Fu 2016 [73]	CNN, CRF	CNN και CRF (DRIVE, STARE)	94.70%*
Wu 2016 [74]	CNN	τηματοποίηση αγγείου και ανίχνευση βρόχου με CNN και PCA (DRIVE)	0.9701
Li 2016 [75]	SDAE	FNN και SDAE (DRIVE, STARE, CHDB)	0.9738
Lahiri 2016 [76]	SDAE	ensemble δύο επιπέδων SDAE (DRIVE)	95.30%*
Oliveira 2017 [77]	u-net	επύξηση δεδομένων και u-net (DRIVE)	0.9768
Leopold 2017 [78]	CNN	CNN ως ένα πολυ-καναλικός ταξινομητής και φίλτρα Gabor (DRIVE)	94.78%*
Leopold 2017 [79]	AE	πλήρως residual AE με περιγραφόμενο ρεύμα βασισμένο στο u-net (DRIVE, STARE, CHDB)	0.8268
Mo 2017 [80]	CNN	επυξημένοι ταξινομητές και μεταφορά μάθησης (DRIVE, STARE, CHDB)	0.9782
Melinscak 2015 [81]	CNN	CNN τεσσάρων επιπέδων (DRIVE)	0.9749
Sengur 2017 [82]	CNN	CNN δύο επιπέδων με dropout (DRIVE)	0.9674
Meyer 2017 [83]	u-net	τηματοποίηση αγγείου με χρήση u-net στην SLO (IOSTAR, RC-SLO)	0.9771

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.<sup>b</sup>Το (\*) υποδηλώνει ακρίβεια.

ενισχύει την ποικιλομορφία στο λεξικό των χαρακτηριστικών που έχουν μαθευτεί για την κατάτμηση των αγγείων. Ο ταξινομητής Softmax χρησιμοποιήθηκε στη συνέχεια για το fine-tuning κάθε AE και διερευνήθηκαν στρατηγικές για συγχώνευση δύο επιπέδων μελών του ensemble.

Άλλες αρχιτεκτονικές χρησιμοποιήθηκαν επίσης για την κατάτμηση των αγγείων. Στο άρθρο τους οι Oliveira et al. [77] εκπαίδευσαν ένα u-net με την DRIVE, παρουσιάζοντας καλά αποτελέσματα και ενδείξεις των πλεονεκτημάτων της επαύξησης των δεδομένων εκπαίδευσης με χρήση ελαστικών μετασχηματισμών. Οι Leopold et al. [78] διερεύνησαν τη χρήση ενός CNN ως ταξινομητή πολλαπλών καναλιών και τη χρήση φίλτρων Gabor για να ενισχύσουν την ακρίβεια της μεθόδου που περιγράφεται στο [84]. Εφάρμοσαν το μέσο μιας σειράς φίλτρων Gabor με διάφορες συχνότητες και τιμές σίγμα στην έξοδο του δικτύου για να καθορίσουν εάν ένα εικονοστοιχείο αντιπροσωπεύει ένα αγγείο ή όχι. Εκτός από τη διαπίστωση ότι τα βέλτιστα φίλτρα διαφέρουν μεταξύ των καναλιών, οι συγγραφείς δηλώνουν επίσης την ‘ανάγκη’ να επιβάλλουν στα δίκτυα να ευθυγραμμίζονται με την ανθρώπινη αντίληψη, στο πλαίσιο της χειρωνακτικής επισήμανσης, ακόμη και αν απαιτεί πληροφορίες υποδειγματοληψίας, οι οποίες διαφορετικά θα μείωναν το υπολογιστικό κόστος. Οι ίδιοι συγγραφείς [79] δημιούργησαν το PixelBNN, το οποίο είναι ένα πλήρως residual AE. Είναι πάνω από οκτώ φορές χρονικά αποδοτικό από τις προηγούμενες μεθόδους κατά την διάρκεια του συμπερασμού με καλά αποτελέσματα, λαμβάνοντας υπόψη τη σημαντική μείωση των πληροφοριών από την αλλαγή μεγέθους των εικόνων κατά την προεπεξεργασία. Στο άρθρο τους οι Mo et al. [80] χρησιμοποίησαν βαθιά επίβλεψη με βοηθητικούς ταξινομητές στα ενδιάμεσα επίπεδα του δικτύου, για να βελτιώσουν τη διακριτική ικανότητα των χαρακτηριστικών στα χαμηλότερα επίπεδα του δικτύου και να καθοδηγήσουν το backpropagation να ξεπεράσει τα vanishing gradients. Επιπλέον, μεταφορά μάθησης χρησιμοποιήθηκε για να ξεπεραστεί το ζήτημα των ανεπαρκών δεδομένων εκπαίδευσης.

### 4.3.2. Ανίχνευση μικροανευρυσμάτων και αιμορραγίας

Οι Haloi et al. [85] εκπαίδευσαν ένα CNN τριών επιπέδων με dropout και maxout για ανίχνευση MA. Τα πειράματα στις ROC και DIA έδειξαν θετικά αποτελέσματα. Στο [86] οι συγγραφείς δημιούργησαν ένα μοντέλο που μαθαίνει έναν γενικό περιγραφέα της μορφολογίας των αγγείων χρησιμοποιώντας την εσωτερική αναπαράσταση ενός variational u-net. Έπειτα, εξέτασαν τις αγγειακές ενσωματώσεις (embeddings) σε ένα παρόμοιο πρόβλημα ανάκτησης εικόνων σύμφωνα με το αγγειακό σύστημα και σε ένα πρόβλημα ταξινόμησης της διαβητικής αμφιβληστροειδοπάθειας, στο οποίο δείχνουν πως τα embeddings των αγγείων μπορούν να βελτιώσουν την ταξινόμηση μιας μεθόδου η οποία βασίζεται στην ανίχνευση MA. Στο [87] οι συγγραφείς συνδυάζουν ενισχυμένα χαρακτηριστικά από ένα CNN με χειροποίητα χαρακτηριστικά. Αυτό το en-



semble διάνυσμα χαρακτηριστικών χρησιμοποιήθηκε στη συνέχεια για την αναγνώριση των υποψηφίων αιμορραγίας και των ΜΑ, χρησιμοποιώντας έναν ταξινομητή RF. Η ανάλυσή τους με χρήση t-SNE καταδεικνύει ότι τα χαρακτηριστικά από το CNN έχουν μεγάλη λεπτομέρεια όπως ένδειξη του προσανατολισμού της βλάβης ενώ τα χειροποίητα χαρακτηριστικά είναι σε θέση να διακρίνουν βλάβες χαμηλής αντίθεσης όπως αιμορραγίες. Στο [88] οι συγγραφείς εκπαιδύσαν ένα CNN πέντε επιπέδων, για ανίχνευση αιμορραγίας χρησιμοποιώντας 6679 εικόνες από τις βάσεις δεδομένων DS16 και Messidor. Εφάρμοσαν επιλεκτική δειγματοληψία σε ένα CNN, κάτι το οποίο αύξησε την ταχύτητα της εκπαίδευσης με δυναμική επιλογή λανθασμένα ταξινομημένων δειγμάτων κατά τη διάρκεια της εκπαίδευσης. Τα βάρη αποδίδονται στα δείγματα εκπαίδευσης και τα ενημερωμένα δείγματα περιλαμβάνονται στην επόμενη επανάληψη εκπαίδευσης.

### 4.3.3. Άλλες εφαρμογές

Το Fundus έχει χρησιμοποιηθεί και για ταξινόμηση αρτηριών/φλεβών. Στο άρθρο τους οι Girard et al. [89] εκπαιδύσαν ένα CNN τεσσάρων επιπέδων, που ταξινομεί τα εικονοστοιχεία των αγγείων σε αρτηρίες/φλέβες χρησιμοποιώντας επαύξηση των δεδομένων εκπαίδευσης με περιστροφικό μετασχηματισμό. Στη συνέχεια κατασκευάστηκε ένα γράφημα από το αγγειακό δίκτυο του αμφιβληστροειδούς, όπου οι κόμβοι ορίζονται ως οι κλάδοι των αγγείων και κάθε άκρη συνδέεται με ένα κόστος που εκτιμά εάν οι δύο κλάδοι θα πρέπει να έχουν την ίδια ετικέτα. Η ταξινόμηση του CNN διαδόθηκε μέσω του ελάχιστου δένδρου του γραφήματος. Τα πειράματα κατέδειξαν την αποτελεσματικότητα της μεθόδου, ιδίως στην παρουσία εμφραγμάτων. Οι Welikala et al. [90] εκπαιδύσαν και αξιολόγησαν ένα CNN τριών επιπέδων, χρησιμοποιώντας εικονοστοιχεία της κεντρικής γραμμής που προέρχονται από εικόνες αμφιβληστροειδούς. Από τα πειράματά τους διαπίστωσαν ότι η επαύξηση των δεδομένων εκπαίδευσης με χρήση περιστροφικών και κλιμακωτών μετασχηματισμών, δεν βοήθησε στην αύξηση της ακρίβειας αποδίδοντας το στην παρεμβολή μεταξύ των εντάσεων των εικονοστοιχείων, η οποία είναι προβληματική λόγω της ευαισθησίας του CNN στην κατανομή των εικονοστοιχείων.

Υπάρχουν επίσης και άλλες εφαρμογές του Fundus όπως η αναγνώριση διακλάδωσης/διέλευσης. Οι Pratt et al. [91] εκπαιδύσαν ένα ResNet 18 επιπέδων, για να εντοπίσουν μικρά patches που περιλαμβάνουν είτε διακλάδωση είτε διέλευση. Ένα άλλο ResNet18 εκπαιδεύτηκε σε patches που έχουν ταξινομηθεί ώστε να έχουν διακλαδώσεις και διελεύσεις έτσι ώστε να διακρίνουν σε ποια κατηγορία ανήκει. Παρόμοια επίλυση σε αυτό το πρόβλημα έχει γίνει από τους ίδιους συγγραφείς [93] χρησιμοποιώντας ένα CNN.

Ένα σημαντικό αποτέλεσμα στον τομέα της Καρδιολογίας χρησιμοποιώντας τη Fundus είναι από τους Poplin et al. [92] που χρησιμοποίησαν ένα Inception-v3 για να

Table 4.7: Εφαρμογές βαθιάς μάθησης με χρήση Fundus, εκτός της κατάκτησης αγγείων

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>	AUC <sup>b</sup>
<b>Μικροαπειρίσγια και αντίγνωση αιμορραγίας</b>			
Haloι 2015 [85]	CNN	ανίγνωση MA με χρήση CNN με dropout και maxout (ROC, Messidor, DIA)	0.98
Giamcardo 2017 [86]	u-net	ανίγνωση MA με χρήση εσωτερικών αναποστώσεων με εκπαιδευμένα u-net (DRIVE, Messidor)	πολλαπλά
Orlando 2018 [87]	CNN	ανίγνωση MA και αιμορραγιών με χρήση χαρρακτριστικών και CNN (DIA, e-ophtha, Messidor)	πολλαπλά
van Grinsven 2017 [88]	CNN	ανίγνωση αιμορραγιών με δειγματοληψία και χρήση CNN πέντε επιπέδων (KR15, Messidor)	πολλαπλά
<b>Άλλες εφαρμογές</b>			
Girard 2017 [89]	CNN	ταξινόμηση αφτηριών/φλέβών με χρήση CNN και διάδοση πιθανοφάνειας (DRIVE, Messidor)	πολλαπλά
Weikala 2017 [90]	CNN	ταξινόμηση αφτηριών/φλέβών με χρήση ενός CNN τριών επιπέδων (UKBDB)	82.26%*
Pratt 2017 [91]	ResNet	ταξινόμηση διασταυρώσεων με χρήση ενός ResNet18 (DRIVE, IOSTAR)	πολλαπλά
Poplin 2017 [92]	Inception	πρόβλεψη παραγόντων καρδιαγγειακών κινδύνου (UKBDB, μη-δημόσια)	πολλαπλά

<sup>a</sup>Σε παρένθεση οι βάσεις δεδομένων που χρησιμοποιήθηκαν.

<sup>b</sup>Το (\*) υποδηλώνει αχρίβεια.

προβλέφουν παράγοντες καρδιαγγειακού κινδύνου (ηλικία, φύλο, κατάσταση καπνίσματος, HbA1c, SBP) και μείζον καρδιακά επεισόδια. Τα μοντέλα τους χρησιμοποίησαν ξεχωριστές πτυχές της ανατομίας για τη δημιουργία κάθε πρόβλεψης, όπως ο οπτικός δίσκος ή τα αιμοφόρα αγγεία, όπως αποδείχθηκε χρησιμοποιώντας την τεχνική soft-attention. Τα περισσότερα αποτελέσματα ήταν σημαντικά καλύτερα από ότι θεωρούνταν προηγουμένως δυνατό με τη Fundus ( $>70\%$  AUC).

#### 4.3.4. Συμπεράσματα της χρήσης βαθιάς μάθησης με Fundus

Όσον αφορά τη χρήση των αρχιτεκτονικών, υπάρχει σαφής προτίμηση στα CNN ειδικά στον τομέα της κατάτμησης των αγγείων, ενώ μια ενδιαφέρουσα προσέγγιση από ορισμένες δημοσιεύσεις είναι η χρήση των CRF μετά την επεξεργασία για τη περαιτέρω βελτίωση της κατάτμησης των αγγείων. Το γεγονός ότι υπάρχουν πολλές βάσεις δεδομένων που είναι διαθέσιμες και ότι η βάση δεδομένων DRIVE χρησιμοποιείται κατά κύριο λόγο στην πλειοψηφία της βιβλιογραφίας, καθιστά τον τομέα αυτό ευκολότερο στη σύγκριση και επικύρωση νέων αρχιτεκτονικών. Επιπλέον, η μη-επεμβατική φύση του Fundus και η πρόσφατη χρήση του ως εργαλείου για την εκτίμηση καρδιαγγειακού κινδύνου, το καθιστά μια πολλά υποσχόμενη απεικονιστική αυξημένης χρησιμότητας στον τομέα της καρδιολογίας.

### 4.4. Ηλεκτρονική τομογραφία

Η ηλεκτρονική τομογραφία (CT) είναι μια μη-επεμβατική μέθοδος για την ανίχνευση της αποφρακτικής αρτηριακής νόσου. Ορισμένες από τις περιοχές στις οποίες εφαρμόστηκε η βαθιά μάθηση στην CT, περιλαμβάνουν την αξιολόγηση της βαθμολογίας ασβεστίου της στεφανιαίας αρτηρίας, τον εντοπισμό και τον κατάτμηση των καρδιακών περιοχών.

Η μέθοδος των Lessman et al. [95] για τη βαθμολόγηση του στεφανιαίου ασβεστίου χρησιμοποιεί τρία ανεξάρτητα εκπαιδευμένα CNNs για να εκτιμήσει ένα πλαίσιο οριοθέτησης γύρω από την καρδιά, στο οποίο τα συνδεδεμένα συστατικά (connected components) πάνω από ένα όριο μονάδας Hounsfield θεωρούνται υποψήφια για CACs. Η ταξινόμηση των voxels πραγματοποιήθηκε τροφοδοτώντας δισδιάστατα patches από τρία ορθογώνια επίπεδα ταυτόχρονα σε τρία CNNs, για να διαχωριστούν από άλλες περιοχές υψηλής έντασης. Οι ασθενείς ταξινομήθηκαν σε μία από τις πέντε πρότυπες κατηγορίες καρδιαγγειακού κινδύνου βάσει της βαθμολογίας Agatston. Συγγραφείς από την ίδια ομάδα δημιούργησαν μια μέθοδο [109] για την ανίχνευση ασβεστοποιήσεων σε θωρακικό CT χαμηλής δόσης χρησιμοποιώντας ένα CNN για ανατομική θέση και ένα

Table 4.8: Εφαρμογές βαθιάς μάθησης με χρήση CT

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>
Lessman 2016 [95]	CNN	ανίχνευση στεφανιαίου αββεστίου με χρήση τριών ανεξάρτητα εκπαιδευμένων CNNs
Shadmi 2018 [96]	DenseNet	συνέκριναν το DenseNet και το u-net για τον εντοπισμό στεφανιαίου αββεστίου
Caro 2018 [97]	CNN	3D CNN παλινδρόμησης για τον υπολογισμό της βαθμολογίας Agatston
Walterink 2016 [98]	CNN	ανίχνευση στεφανιαίου αββεστίου με χρήση τριών CNNs για τον εντοπισμό και δύο CNN για την ανίχνευση
Santini 2017 [99]	CNN	ανίχνευση στεφανιαίου αββεστίου με χρήση ενός CNN επτά επιπέδων σε patches εικόνας
Lopez 2017 [100]	CNN	χαρακτηρισμός όγκου θρόμβου με χρήση ενός 2D CNN και μετα-επέξεργασία
Hong 2016 [101]	DBN	ανίχνευση, τμηματοποίηση, ταξινόμηση αορτικού ανευρύσματος με χρήση DBN και patches εικόνας
Lin 2017 [94]	CNN	τμηματοποίηση αορστερού κόλπου καρδιάς με CNN δώδεκα επιπέδων και μοντέλα ενεργού σχήματος (STA13)
de Vos 2016 [102]	CNN	3D εντοπισμός των ανατομικών δομών με χρήση τριών CNNs, ένα για κάθε ορθογώνιο επίπεδο
Moradi 2016 [103]	CNN	ανίχνευση θέσης στο CT με προεκπαιδευμένο VGGnet, χειροποίητα χαρακτηριστικά και SVM
Zheng 2015 [104]	Multiple	ανίχνευση διααλάδωσης της καρδιάς με χρήση πολυ-επίπεδων perceptors και πιθανοτικά boosting-trees
Μοντογα 2018 [105]	ResNet	3D αναστρωσική αγγειογραφείας με χρήση ενός ResNet-30 επιπέδων
Zreik 2018 [106]	CNN, AE	ανίχνευση στεφανιαίας αρτηριακής στένωσης με CNN για LV τμηματοποίηση και AE, SVM για ταξινόμηση
Commandeur 2018 [107]	CNN	ποσοτικοποίηση του επικάρδιου και θωρακικού λιπώδους ιστού από χωρίς-αντίθεση CT
Gulsun 2016 [108]	CNN	εξαγωγή στεφανιαίων κεντρικών γραμμών με χρήση βέλτιστου μονοπατιού από πεδίο ραών και ένα CNN

<sup>a</sup> Αποτελέσματα από αυτές τις απεικονιστικές τεχνικές δεν δημοσιεύονται σε αυτήν την βιβλιογραφική αναφορά καθώς υπάρχει υψηλή μεταβλητότητα σε σχέση με το ερευνητικό ερώτημα προς προς απόκτηση και στην χρήση των μετρήσεων. Επιπλέον όλες οι δημοσιεύσεις χρησιμοποιούν μη-δημόσιες βάσεις δεδομένων εκτός της [94].

άλλο CNN για την ανίχνευση της ασβεστοποίησης. Στο [96] οι συγγραφείς συγκρίναν ένα u-net και DenseNet για τον υπολογισμό της βαθμολογίας Agatston χρησιμοποιώντας πάνω από 1000 εικόνες θωρακικού CT. Οι συγγραφείς επεξεργάστηκαν έντονα τις εικόνες χρησιμοποιώντας κατωφλίωση, ανάλυση connected components και μορφολογικές λειτουργίες για την ανίχνευση των πνευμόνων, της τραχείας και της τροπίδας. Τα πειράματά τους έδειξαν ότι το DenseNet είχε καλύτερη απόδοση όσον αφορά την ακρίβεια. Οι Cano et al. [97] εκπαίδευσαν ένα 3D CNN παλινδρόμησης που υπολόγισε την βαθμολογία Agatston χρησιμοποιώντας 5973 εικόνες CT χωρίς ECG περίφραξης επιτυγχάνοντας συσχέτιση Pearson 0.932. Στο [98] οι συγγραφείς δημιούργησαν μια μέθοδο για την ανίχνευση και τον ποσοτικό προσδιορισμό της CAC, χωρίς την εξαγωγή της στεφανιαίας αρτηρίας. Η ανίχνευση του πλαισίου οριοθέτησης γύρω από την καρδιά χρησιμοποιεί τρία CNNs, όπου το καθένα ανιχνεύει την καρδιά στο αξονικό, ισχαιμικό και στεφανιαίο επίπεδο. Ένα άλλο ζεύγος από CNNs χρησιμοποιήθηκε για την ανίχνευση CAC. Το πρώτο CNN αναγνωρίζει κύστες τύπου CAC, απορρίπτοντας έτσι την πλειονότητα των voxels που δεν είναι υποψήφια CAC όπως ο πνεύμονας και ο λιπώδης ιστός. Τα αναγνωρισμένα voxels τύπου CAC ταξινομούνται περαιτέρω από το δεύτερο CNN, το οποίο διακρίνει τα CAC. Αν και τα CNNs μοιράζονται την αρχιτεκτονική, δεδομένου ότι έχουν διαφορετικό έργο, δεν μοιράζονται βάρη. Επιτυγχάνουν μια συσχέτιση Pearson 0.95, συγκρίσιμη με προηγούμενες καλύτερες τεχνικές. Οι Santini et al. [99] εκπαίδευσαν ένα CNN επτά επιπέδων χρησιμοποιώντας patches, για την κατάτμηση και ταξινόμηση των στεφανιαίων περιοχών στις εικόνες CT. Εκπαίδευσαν, επικύρωσαν και δοκίμασαν το δίκτυό τους σε 45, 18 και 56 CT όγκους αντιστοίχως, επιτυγχάνοντας μια συσχέτιση Pearson 0.983.

Το CT έχει χρησιμοποιηθεί επίσης και για την κατάτμηση διαφόρων καρδιακών περιοχών. Οι Lopez et al. [100] εκπαίδευσαν ένα 2D CNN για την εκτίμηση του όγκου αορτής θρόμβου από προεγχειρητικές και μετεγχειρητικές κατατμήσεις, χρησιμοποιώντας περιστροφικές και κατοπτρικές επαυξήσεις δεδομένων. Μεταγενέστερη επεξεργασία περιλαμβάνει Gaussian φιλτράρισμα και ομαδοποίηση κ-μέσων. Στο άρθρο τους, οι Hong et al. [101] εκπαίδευσαν ένα DBN χρησιμοποιώντας patches εικόνας για την ανίχνευση, τμηματοποίηση και ταξινόμηση της σοβαρότητας του κοιλιακού αορτικού ανευρύσματος σε CT εικόνες. Οι Liu et al. [94] χρησιμοποίησαν ένα FCN με δώδεκα επίπεδα για την κατάτμηση του αριστερού κόλπου σε 3D όγκους CT και στη συνέχεια βελτίωσαν τα αποτελέσματα τμηματοποίησης του FCN με ένα μοντέλο ενεργού σχήματος επιτυγχάνοντας Dice 93%.

Το CT έχει επίσης χρησιμοποιηθεί για τον εντοπισμό των καρδιακών περιοχών. Στο [102] οι συγγραφείς δημιούργησαν μια μέθοδο για την ανίχνευση ανατομικών ROI (καρδιά, αορτική αψίδα και φθίνουσα αορτή) σε εικόνες από θωρακικό CT προκειμένου να εντοπιστούν σε 3D. Κάθε ROI εντοπίστηκε χρησιμοποιώντας έναν συνδυ-

ασμό τριών CNNs, κάθε ένα αναλύοντας ένα ορθογώνιο επίπεδο εικόνας. Ενώ ένα CNN προέβλεψε την παρουσία ενός συγκεκριμένου ROI στο δεδομένο επίπεδο, ο συνδυασμός των αποτελεσμάτων τους παρείχε ένα 3D πλαίσιο οριοθέτησης γύρω του. Στο άρθρο τους οι Moradi et al. [103] αντιμετωπίζουν το πρόβλημα της ανίχνευσης κατακόρυφης θέσης για μια δεδομένη καρδιακή εικόνα CT. Διαχωρίζουν την περιοχή του σώματος που απεικονίζεται στο θωρακικό CT σε εννέα σημασιολογικές κατηγορίες, οι οποίες κάθε μια αντιπροσωπεύουν μία περιοχή που είναι σχετική με τη μελέτη αντίστοιχων νόσων. Χρησιμοποιώντας ένα σύνολο χειροποίητων χαρακτηριστικών εικόνας μαζί με τα χαρακτηριστικά που προέρχονται από ένα προεκπαιδευμένο VGGnet με πέντε επίπεδα, χτίζουν ένα σχήμα ταξινόμησης για να αντιστοιχίσουν μια δεδομένη εικόνα CT στο σχετικό επίπεδο. Κάθε ομάδα χαρακτηριστικών χρησιμοποιήθηκε για την εκπαίδευση ενός ξεχωριστού ταξινομητή SVM και οι προβλεπόμενες ετικέτες στη συνέχεια συνδυάζονται σε ένα γραμμικό μοντέλο, το οποίο επίσης αντλήθηκε από τα δεδομένα εκπαίδευσης.

Η βαθιά μάθηση χρησιμοποιήθηκε σε συνδυασμό με το CT και για άλλες περιοχές εκτός από την καρδιά. Οι Zheng et al. [104] δημιούργησαν μια μέθοδο για 3D ανίχνευση σε ογκομετρικά δεδομένα, τα οποία αξιολογήθηκαν ποσοτικά για την ανίχνευση της διακλάδωσης της καρωτιδικής αρτηρίας σε CT. Χρησιμοποιήθηκε ένα δίκτυο κρυφών επιπέδων για τον αρχικό έλεγχο όλων των voxels για να αποκτηθεί ένας μικρός αριθμός υποψηφίων, ακολουθούμενος από μια ακριβέστερη ταξινόμηση με ένα βαθύ δίκτυο. Τα χαρακτηριστικά από το δίκτυο συνδυάζονται περαιτέρω με τα χαρακτηριστικά κυματιδίων Haar, για την αύξηση της ακρίβειας ανίχνευσης. Οι Montoya et al. [105] εκπαίδευσαν ένα ResNet με 30 επίπεδα, για να δημιουργήσουν 3D αγγειογραφήματα χρησιμοποιώντας τρεις τύπους ιστών (αγγειακό σύστημα, οστό και μαλακό ιστό). Δημιούργησαν τις επισημάνσεις χρησιμοποιώντας κατωφλίωση και connected components σε 3D, έχοντας ένα συνδυασμένο σύνολο 13790 εικόνων.

Το CT έχει επίσης χρησιμοποιηθεί για την επίλυση και άλλων προβλημάτων. Οι Zreik et al. [106] δημιούργησαν μια μέθοδο για την ταυτοποίηση ασθενών με στένωση στεφανιαίας αρτηρίας από το μυοκάρδιο του LV από CT. Χρησιμοποίησαν ένα CNN πολλαπλών κλιμάκων για να τμηματοποιήσουν το μυοκάρδιο του LV και έπειτα το κωδικοποίησαν χρησιμοποιώντας ένα μη-επιβλεπόμενο συνελικτικό AE. Η τελική ταξινόμηση έγινε με έναν ταξινομητή SVM με βάση τα εξαγόμενα και ομαδοποιημένα clustering. Παρόμοια δουλειά έχει γίνει από τους ίδιους συγγραφείς [110], οι οποίοι χρησιμοποίησαν τρία CNNs για να ανιχνεύσουν ένα πλαίσιο οριοθέτησης γύρω από την LV και εφάρμοσαν ταξινόμηση voxel LV μέσα στο κιβώτιο οριοθέτησης. Οι Commandeur et al. [107] χρησιμοποίησαν ένα συνδυασμό δύο βαθιών δικτύων για τον ποσοτικό προσδιορισμό του επικαρδιακού και θωρακικού λιπώδους ιστού, με CT από 250 ασθενείς με 55 εικόνες ανά ασθενή κατά μέσο όρο. Το πρώτο δίκτυο είναι ένα

CNN έξι επιπέδων, που ανιχνεύει την περιοχή που βρίσκεται μέσα στα όρια της καρδιάς και διαχωρίζει τις θωρακικές και επικαρδιακές-παρακαρδιακές μάσκες. Το δεύτερο δίκτυο είναι ένα CNN πέντε επιπέδων, που ανιχνεύει τη γραμμή του περικαρδίου από την αξονική τομογραφία σε κυλινδρικές συντεταγμένες. Στη συνέχεια ένα στατιστικό μοντέλο συστηματοποίησης μαζί με εφαρμογή κατωφλίωσης και φιλτράρισμα διαμέσου, παρέχουν τις τελικές κατατιμήσεις. Οι Gulsun et al. [108] δημιούργησαν μια μέθοδο για την εξαγωγή των κεντρικών γραμμών των αγγείων στην CT. Πρώτον, οι βέλτιστες διαδρομές ανιχνεύονται σε ένα υπολογισμένο πεδίο ροής και στη συνέχεια χρησιμοποιείται ένας ταξινομητής CNN για την αφαίρεση των εξωτερικών διαδρομών στις ανιχνευμένες κεντρικές γραμμές. Η μέθοδος ενισχύθηκε χρησιμοποιώντας μια ανίχνευση βασισμένη στο μοντέλο των συγκεκριμένων στεφανιαίων περιοχών και των κύριων κλάδων για να περιορίσουν τον χώρο αναζήτησης.

## 4.5. Ηχοκαρδιογράφημα

Το ηχοκαρδιογράφημα είναι μια μέθοδος απεικόνισης της περιοχής της καρδιάς χρησιμοποιώντας υπερηχητικά κύματα. Χρήσεις της βαθιάς μάθησης στο ηχοκαρδιογράφημα περιλαμβάνουν κυρίως την κατάτμηση του LV και την αξιολόγηση της ποιότητας της βαθμολογίας της εικόνας, μεταξύ άλλων.

Τα DBN έχουν χρησιμοποιηθεί για την κατάτμηση της LV στα ηχοκαρδιογραφήματα. Στο [111] οι συγγραφείς δημιούργησαν μια μέθοδο που αποσυνδέει τις άκαμπτες και μη-άκαμπτες ανιχνεύσεις, με ένα DBN που μοντελοποιεί το LV δείχνοντας ότι είναι πιο ισχυρό από τα level-sets και τα παραμορφώσιμα μοντέλα. Οι Nascimento et al. [112] χρησιμοποιούν μάθηση του πολυειδούς που χωρίζει τα δεδομένα σε patches όπου το κάθε ένα προτείνει μια κατάτμηση της LV. Η σύντηξη των patches πραγματοποιήθηκε από έναν πολλαπλό ταξινομητή DBN ο οποίος αποδίδει ένα βάρος σε κάθε patch. Με τον τρόπο αυτό, η μέθοδος δεν βασίζεται σε μια μόνο κατάτμηση και η διαδικασία εκπαίδευσης παράγει ισχυρά μοντέλα χωρίς την ανάγκη μεγάλων βάσεων δεδομένων εκπαίδευσης. Στο [113] οι συγγραφείς χρησιμοποίησαν ένα συστηματοποιημένο FCN και μεταφορά μάθησης. Συγκρίνουν τη μέθοδο τους με απλούστερες αρχιτεκτονικές FCN και μια προηγούμενη μέθοδο παρουσιάζοντας καλύτερα αποτελέσματα.

Το ηχοκαρδιογράφημα έχει επίσης χρησιμοποιηθεί για την ταξινόμηση των όψεων της απεικόνισης. Οι Madani et al. [114] εκπαίδευσαν CNN έξι επιπέδων για να ταξινομήσουν 15 προβολές (12 βίντεο και 3 ακίνητα) με διαθωρακικό υπερηχογράφημα, επιτυγχάνοντας καλύτερα αποτελέσματα από τους ειδικούς ηχοκαρδιογραφήματος. Στο [115] οι συγγραφείς δημιούργησαν ένα residual 3D CNN για ταξινόμηση κλάσματος εξώθησης από εικόνες διαθωρακικού ηχοκαρδιογραφήματος. Χρησιμοποίησαν 8715 εξετάσεις κάθε μια με 30 διαδοχικές εικόνες του apical θαλάμου για να εκπαιδευ-

Table 4.9: Εφαρμογές βαθιάς μάθησης με χρήση Ηχοκαρδιογραφήματος

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>
Carneiro 2012 [111]	DBN	τμηματοποίηση LV με αποσύνδεση στοιβαγών και μη-στοιβαγών ανηλεύσεων με χρήση DBN σε 480 εικόνες
Nascimento 2016 [112]	DBN	τμηματοποίηση LV με μάθηση επιφανειών και ένα DBN
Chen 2016 [113]	CNN	τμηματοποίηση LV με συστηματοποιημένο FCN πολλαπλών τομέων και μεταφορά μάθησης
Madani 2018 [114]	CNN	ταξινόμηση όψης διαθωρακικού ηχοκαρδιογραφήματος με CNN έξι επιπέδων
Silva 2018 [115]	CNN	ταξινόμηση κλάσματος εξώθησης με ένα residual 3D CNN και διαθωρακικά ηχοκαρδιογραφήματα
Gao 2017 [116]	CNN	ταξινόμηση όψης με σύντηξη δύο CNNs με επτά επίπεδα
Abdi 2017 [117]	CNN, LSTM	αξιολόγηση ποιότητας βαθμολογίας με συνεικτικώς και επαναληπτικά επίπεδα
Ghesu 2016 [118]	CNN	τμηματοποίηση καρτικής βαλβίδας με χρήση 2891 3D δεισοφωγραφικών ηχοκαρδιογραφικών εικόνων
Perrin 2017 [119]	CNN	ταξινόμηση εκ-γενετής καρδιακής ασθένειας με CNN
Moradi 2016 [120]	VGGnet, doc2vec	δημιουργία σημειωτικών περιγραφών για εικόνες

<sup>a</sup> Αποτέλεσμα από αυτές τις απεικονιστικές τεχνικές δεν δημοσιεύονται σε αυτήν την βιβλιογραφική αναφορά καθώς υπάρχει υψηλή μεταβλητότητα σε σχέση με το ερευνητικό ερώτημα προς απάντηση και στην χρήση των μέτρων. Επιπλέον όλες οι δημοσιεύσεις χρησιμοποιούν μη-δημόσιες βάσεις δεδομένων.



σουν και να δοκιμάσουν τη μέθοδο τους επιτυγχάνοντας ικανοποιητικά αποτελέσματα. Οι Gao et al. [116] ενσωμάτωσαν χωρική και χρονική πληροφορία από τις εικόνες του βίντεο της κινούμενης καρδιάς, με τη σύντηξη δύο CNNs με επτά επίπεδα το καθένα. Η μέτρηση της επιτάχυνσης σε κάθε σημείο υπολογίστηκε με τη χρήση μεθόδου πυκνής οπτικής ροής, για την απεικόνιση πληροφοριών χρονικής κίνησης. Στη συνέχεια η σύντηξη των CNN πραγματοποιήθηκε χρησιμοποιώντας γραμμικές ενσωματώσεις των διανυσμάτων των εξόδων τους. Συγκρίσεις έγιναν με προηγούμενες προσεγγίσεις βασισμένες σε χειροκίνητη δημιουργία χαρακτηριστικών, καταδεικνύοντας καλύτερα αποτελέσματα.

Τέλος, η αξιολόγηση της ποιότητας των εικόνων και άλλα προβλήματα επιλύθηκαν με τη χρήση του ηχοκαρδιογραφήματος. Στο [117] οι συγγραφείς δημιούργησαν μια μέθοδο για τη μείωση της μεταβλητότητας των δεδομένων κατά τη διάρκεια της επισήμανσης από τον χειριστή, υπολογίζοντας μια βαθμολογία ποιότητας που ανατροφοδοτείται σε πραγματικό χρόνο. Το μοντέλο συνίσταται από συνελικτικά επίπεδα για την εξαγωγή χαρακτηριστικών από την είσοδο και επαναλαμβανόμενα επίπεδα για χρήση των χρονικών πληροφοριών. Η μέθοδος των Ghesu et al. [118] ανίχνευσης αντικειμένων και κατάτμησης στο πλαίσιο της ογκομετρικής ανάλυσης εικόνων, γίνεται με την επίλυση της ανατομικής εκτίμησης της θέσης και της οριοθέτησης. Για το σκοπό αυτό εισάγουν βαθιά μάθηση της οριοθέτησης, η οποία παρέχει υψηλής απόδοσης χρόνο εκτέλεσης, μαθαίνοντας ταξινομητές σε περιοχές υψηλής πιθανότητας και σε χώρους σταδιακά αυξανόμενων διαστάσεων. Δεδομένου του εντοπισμού των αντικειμένων, προτείνουν ένα συνδυασμένο μοντέλο ενεργού σχήματος βαθιάς μάθησης για την εκτίμηση του μη-άκαμπτου ορίου αντικειμένου. Στο άρθρο τους οι Perrin et al. [119] εκπαίδευσαν και αξιολόγησαν το AlexNet με 59151 εικόνες, για να ταξινομήσουν μεταξύ πέντε παιδιατρικών πληθυσμών με συγγενή καρδιακή νόσο. Οι Moradi et al. [120] δημιούργησαν μια μέθοδο που βασίζεται σε VGGnet και doc2vec για να παράξουν σημασιολογικούς περιγραφείς εικόνων, οι οποίοι μπορούν να χρησιμοποιηθούν ως ασθενώς επισημασμένες περιπτώσεις ή να διορθωθούν από ιατρικούς εμπειρογνώμονες. Το μοντέλο τους ήταν σε θέση να αναγνωρίσει το 91% των ασθενειών και 77% των σοβαρών νόσων από εικόνες Doppler καρδιακών βαλβίδων.

## 4.6. Τομογραφία οπτικής συνοχής

Η Τομογραφία Οπτικής Συνοχής (Optical Coherence Tomography, OCT) είναι μια ενδοαγγειακή απεικόνιση που παρέχει εικόνες αρτηριών υψηλής ανάλυσης και ποσοτικές μετρήσεις της στεφανιαίας γεωμετρίας στο κλινικό περιβάλλον [121].

Στο άρθρο τους οι Roy et al. [122] χαρακτηρίζουν τον ιστό στο OCT, μαθαίνοντας την πολυκλιμακωτή κατανομή των δεδομένων με ένα AE. Ο κανόνας μάθησης του

δικτύου εισάγει μια παράμετρο κλίμακας που συνδέεται με το backpropagation. Σε σύγκριση με τρία προεκπαιδευμένα ΑΕ βάσης αναφοράς επιτυγχάνεται καλύτερη απόδοση όσον αφορά την ακρίβεια στην ανίχνευση πλακών/κανονικών εικονοστοιχείων. Οι Yong et al. [123] δημιούργησαν ένα CNN γραμμικής παλινδρόμησης με τέσσερα επίπεδα, για να χωρίσουν τον αυλό του αγγείου παραμετροποιημένο σε όρους ακτινικών αποστάσεων από το κέντρο του καθετήρα στις πολικές συντεταγμένες. Η υψηλή ακρίβεια αυτής της μεθόδου μαζί με την υπολογιστική αποτελεσματικότητά της (40.6ms/εικόνα), υποδεικνύουν τη δυνατότητα χρήσης σε πραγματικό κλινικό περιβάλλον. Στο [124] οι συγγραφείς σύγκριναν την διακριτική ικανότητα των βαθιών χαρακτηριστικών που εξάγονται από τα AlexNet, GoogleNet, VGGnet16 και VGGnet19 για την ταυτοποίηση του ινωαθερώματος. Η επαύξηση δεδομένων εφαρμόστηκε σε μια βάση δεδομένων OCT για κάθε σχήμα ταξινόμησης και επίσης εφαρμόστηκε γραμμικό SVM για την ταξινόμηση των εικόνων ως κανονικών ή ινωαθερωματικών. Τα αποτελέσματα υποδεικνύουν ότι το VGGnet19 είναι καλύτερο για τον εντοπισμό εικόνων που περιέχουν ινωαθέρωμα. Οι Abdolmanafi et al. [125] ταξινομούν τον ιστό στο OCT χρησιμοποιώντας ένα προεκπαιδευμένο AlexNet ως εξαγωγέα χαρακτηριστικών και συγκρίνουν τις προβλέψεις τριών ταξινομητών, CNN, RF και SVM, με το πρώτο να επιτυγχάνει τα καλύτερα αποτελέσματα.

## 4.7. Άλλες απεικονιστικές τεχνικές

Το ενδοαγγειακό υπερηχογράφημα (Intravascular Ultrasound, IVUS) χρησιμοποιεί μετασχηματιστές τοποθετημένους σε τροποποιημένους ενδοστεφανιαίους καθετήρες για την παροχή ακτινικής ανατομικής απεικόνισης ενδοστεφανιαίας ασβεστοποίησης και σχηματισμού πλάκας [126]. Οι Lekadir et al. [127] χρησιμοποίησαν ένα βασισμένο σε patches CNN τεσσάρων επιπέδων, για τον χαρακτηρισμό της σύνθεσης της πλάκας σε καρωτιδικές εικόνες υπερήχων. Τα πειράματα που έγιναν από τους συγγραφείς έδειξαν ότι το μοντέλο επιτυγχάνει καλύτερη ακρίβεια από SVM μονών και πολλαπλών κλιμάκων. Στο [128] οι συγγραφείς αυτοματοποίησαν ολόκληρη τη διαδικασία της ερμηνείας του πάχους του καρωτιδικού intima-media. Εκπαίδευσαν ένα CNN δύο επιπέδων με δύο εξόδους για την επιλογή της εικόνας και ένα CNN δύο επιπέδων με τρεις εξόδους για τον εντοπισμό του ROI και τις μετρήσεις πάχους intima-media. Το μοντέλο αυτό αποδίδει καλύτερα από προηγούμενη μέθοδο από τους ίδιους συγγραφείς, το οποίο δικαιολογεί την ικανότητα του CNN να μαθαίνει την εμφάνιση του QRS και του ROI, αντί να στηρίζεται σε εφαρμογή κατωφλίωσης με κατώτατο πλάτος και καμπυλότητα. Οι Tom et al. [129] δημιούργησαν μια μέθοδο βασισμένη στα Δίκτυα Γεννήτριας-Διάκρισης (Generative Adversarial Networks, GANs) για χρονικά αποδοτική προσομοίωση ρεαλιστικού IVUS. Η προσομοίωση στο πρώτο

στάδιο αφορούσε τον προσομοιωτή ψευδότροπου IVUS και την χαρτογράφηση του στίγματος της ψηφιακά καθορισμένης εικόνας. Στο δεύτερο στάδιο βελτιώθηκαν οι αντιστοιχίσεις για τη διατήρηση των συγκεκριμένων εντάσεων στίγματος με ιστούς χρησιμοποιώντας ένα GAN με τέσσερα residual επίπεδα, ενώ στο τελευταίο στάδιο το GAN παρήξε εικόνες υψηλής αναλυτικότητας με παθορεαλιστικά προφίλ στίγματος.

Άλλες εφαρμογές καρδιολογίας με χρήση μεθόδων βαθιάς μάθησης, περιλαμβάνουν μαστογραφίες, ακτίνες X, διαδερμική διαφραγματική αγγειοπλαστική, εικόνες βιοψίας και απεικόνιση διάχυσης του μυοκαρδίου. Στο άρθρο τους οι Wang et al. [130] εφάρμοσαν μια διαδικασία βασισμένη σε εικονοστοιχεία και patches για την ανίχνευση ασβεστοποιητικού αρτηριακού μαστού σε μαστογραφίες, χρησιμοποιώντας ένα CNN δέκα επιπέδων και μορφολογικές λειτουργίες για μετα-επεξεργασία. Οι συγγραφείς χρησιμοποίησαν 840 εικόνες και τα πειράματά τους οδήγησαν σε ένα μοντέλο που πέτυχε συντελεστή προσδιορισμού 96.2%. Οι Liu et al. [131] εκπαίδευσαν ένα CNN χρησιμοποιώντας 1768 εικόνες ακτίνων X με αντίστοιχες επισημάνσεις διαγνώσεων. Η μέση διαγνωστική ακρίβεια των μοντέλων έφτασε στο 0.89 όταν το βάθος του δικτύου ήταν οκτώ επίπεδα; μετά από αυτό η αύξηση της ακρίβειας ήταν περιορισμένη. Στο [132] οι συγγραφείς δημιούργησαν μια μέθοδο για την αποθρομβοποίηση εικόνων διαδερμικής αγγειοπλαστικής στεφανιαίας. Δοκίμασαν το μέσο τετραγωνικό σφάλμα και τη δομική ομοιότητα ως συναρτήσεις απωλειών σε δύο patch CNN με τέσσερα επίπεδα και τα σύγκριναν με διαφορετικούς τύπους και επίπεδα θορύβου. Οι Nirschl et al. [133] χρησιμοποίησαν εικόνες ενδομυοκαρδιακής βιοψίας από 209 ασθενείς για να εκπαιδεύσουν και να δοκιμάσουν ένα patch CNN έξι επιπέδων, για τον εντοπισμό της καρδιακής ανεπάρκειας. Επίσης εφάρμοσαν περιστροφική επαύξηση δεδομένων, ενώ υπολογίστηκε ο μέσος όρος των εξόδων του CNN σε κάθε patch για να ληφθεί η πιθανότητα για κάθε εικόνα. Αυτό το μοντέλο έδειξε καλύτερα αποτελέσματα από τα AlexNet, Inception και ResNet50. Στο [134] οι συγγραφείς εκπαίδευσαν ένα CNN τριών επιπέδων για την πρόβλεψη της αποφρακτικής αιμάτωσης μυοκαρδίου του CAD από 1638 ασθενείς και το συνέκριναν με το συνολικό έλλειμμα διάχυσης. Το μοντέλο υπολογίζει την πιθανότητα ανά αγγείο κατά τη διάρκεια της εκπαίδευσης, ενώ κατά τη διάρκεια της δοκιμής χρησιμοποιούνται οι μέγιστες πιθανότητες ανά αρτηρία ανά βαθμολογία ασθενούς. Τα αποτελέσματα δείχνουν ότι αυτή η μέθοδος υπερέχει του συνολικού ελλείμματος διάχυσης στο πρόβλημα της πρόβλεψης των αγγείων και των ασθενών.

Table 4.10: Εφαρμογές βαθιάς μάθησης με χρήση OCT και άλλων απεικονιστικών τεχνικών

Αναφορά	Μέθοδος	Εφαρμογή/Σημειώσεις <sup>a</sup>
<b>OCT</b>		
Roy 2016 [122]	AE	χρωακτηρισμός ιστού με χρήση ενός AE διατήρησης της κατανομής
Yong 2017 [123]	CNN	τήμηματοποίηση του lumen με χρήση ενός CNN γραμμικής-παλινδρόμησης με τέσσερα επίπεδα
Xu 2017 [124]	CNN	παραυσία του ενδοθρομβώματος με χρήση χαρακτηριστικών εξαρτημένων από προηγούμενες αρχιτεκτονικές και ένα SVM
Abdolmaleki 2017 [125]	CNN	τήμηματοποίηση του intima-media με χρήση ενός προεκπαιδευμένου AlexNet και σύγκριση διαφόρων ταξινόμητρών
<b>Άλλες απεικονιστικές τεχνικές</b>		
Lekadir 2017 [127]	CNN	χρωακτηρισμός καρωτιδικής πλάκας με χρήση CNN τεσσάρων επιπέδων με υπερήχους
Tajbakhsh 2017 [128]	CNN	εμφηγεία video του πάχους των καρωτιδικών intima-media με χρήση δύο CNNs δύο επίπεδα με υπερήχους
Tom 2017 [129]	GAN	δημιουργία IVUS εικόνων με χρήση δύο GANs (IV1)
Wang 2017 [130]	CNN	αφβεστοποίηση αθηρωκών με χρήση CNN δέκα επιπέδων σε μαστογραφίες
Lin 2017 [131]	CNN	ανίχνευση GAC με χρήση CNNs σε 1768 X-Rays
Pavoni 2017 [132]	CNN	αποθρομβοποίηση του διαδερμικού αυλού στεφανιαίων εικόνων αγγειοπλαστικής με ένα CNN τεσσάρων επιπέδων
Nirschl 2018 [133]	CNN	CNN έξι επιπέδων βασισμένο σε patch για ανίχνευση καρδιακής ανεπάρκειας σε ενδομυοκαρδιακές εικόνες βιοψίας
Betancur 2018 [134]	CNN	CNN τριών επιπέδων για προβλεψη CAD από απεικονίσεις του μυοκαρδίου

<sup>a</sup> Αποτελέσματα από αυτές τις απεικονιστικές τεχνικές δεν δημοσιεύονται σε αυτήν την βιβλιογραφική αναφορά καθώς υπάρχει υψηλή μεταβλητότητα σε σχέση με το ερευνητικό ερώτημα προς προς απόκτηση και στην χρήση των μετρήσεων. Επιπλέον όλες οι δημοσιεύσεις χρησιμοποιούν μη-δημόσιες βάσεις δεδομένων εκτός της [129].

## 4.8. Συμπεράσματα της χρήσης βαθιάς μάθησης με CT, ηχοκαρδιογράφημα, OCT και άλλων απεικονιστικών τεχνικών

Δημοσιεύσεις που χρησιμοποίησαν αυτούς τους τρόπους απεικόνισης παρουσίασαν υψηλή μεταβλητότητα από την πλευρά του ερευνητικού τους ερωτήματος και ήταν κατά πλειοψηφία ασυμβίβαστα σε σχέση με τη χρήση μέτρων αξιολόγησης για τα αποτελέσματα που ανέφεραν. Αυτές οι μέθοδοι απεικόνισης επίσης δεν έχουν διαθέσιμες δημόσιες βάσεις δεδομένων, περιορίζοντας έτσι τις ευκαιρίες για δημιουργία νέων αρχιτεκτονικών από ερευνητικές ομάδες που δεν έχουν κλινικούς συνεργάτες. Από την άλλη πλευρά, υπάρχει σχετικά υψηλή ομοιομορφία όσον αφορά τη χρήση αρχιτεκτονικών με το πιο διαδεδομένο το CNN, ειδικά τις προεκπαιδευμένες αρχιτεκτονικές του διαγωνισμού ImageNet (AlexNet, VGGnet, GoogleNet, ResNet).

## 4.9. Συζήτηση και μελλοντικές κατευθύνσεις

Είναι προφανές από τη βιβλιογραφία ότι οι μέθοδοι βαθιάς μάθησης θα αντικαταστήσουν τα συστήματα που βασίζονται σε χειροποίητους κανόνες και την παραδοσιακή μηχανική μάθηση. Στο [155] οι συγγραφείς υποστηρίζουν ότι η βαθιά μάθηση είναι καλύτερη στην απεικόνιση πολύπλοκων μοτίβων κρυμμένων σε ιατρικά δεδομένα υψηλής διαστάσεων. Οι Krittanawong et al. [150] υποστηρίζουν ότι η αυξανόμενη διαθεσιμότητα αυτοματοποιημένων εργαλείων AI πραγματικού χρόνου στα EHRs, θα μειώσει την ανάγκη για συστήματα βαθμολόγησης όπως το σκορ Framingham. Στο [139] οι συγγραφείς υποστηρίζουν ότι οι προγνωστικές αναλύσεις του AI και η εξατομικευμένη κλινική υποστήριξη για την αναγνώριση του ιατρικού κινδύνου, είναι ανώτερες από τις ανθρώπινες γνωσιακές ικανότητες. Επιπλέον, το AI μπορεί να διευκολύνει την επικοινωνία μεταξύ ιατρών και ασθενών, μειώνοντας τους χρόνους επεξεργασίας και αυξάνοντας έτσι την ποιότητα της περίθαλψης των ασθενών. Οι Loh et al. [149] υποστηρίζουν ότι οι τεχνολογίες βαθιάς μάθησης και κινητής τηλεφωνίας θα επιταχύνουν τον πολλαπλασιασμό των υπηρεσιών υγειονομικής περίθαλψης σε εκείνους που βρίσκονται σε φτωχές περιοχές, γεγονός που με τη σειρά του οδηγεί σε περαιτέρω μείωση των ποσοστών ασθενειών. Οι Mayer et al. [135] δηλώνουν ότι τα big data υπόσχονται να αλλάξουν την καρδιολογία μέσω της αύξησης των δεδομένων που συλλέγονται, αλλά ο αντίκτυπος τους υπερβαίνει τη βελτίωση των υπάρχουσών μεθόδων όπως η αλλαγή του τρόπου με τον οποίο δημιουργούνται νέες ιδέες.

Η βαθιά μάθηση απαιτεί δεδομένα εκπαίδευσης μεγάλου όγκου για την επίτευξη αποτελεσμάτων υψηλής ποιότητας [157]. Αυτό είναι ιδιαίτερα δύσκολο με τα ιατρικά

Table 4.11: Βιβλιογραφικές αναφορές βαθιάς μάθησης με εφαρμογή στην καρδιολογία

Αναφορά	Εφαρμογή/Σημειώσεις
Mayer 2015 [135]	Big data στην καρδιολογία αλλάζει τον τρόπο με τον οποίο δημιουργούνται καινούργιες ιδέες
Austin 2016 [136]	σφαιρική εικόνα των big data, τα παρονεκτάγματα, πιθανά μειονεκτήματα και μελλοντική συνεισφορά στην καρδιολογία
Greenspan 2016 [137]	ανίχνευση ιστών, τμηματοποίηση και μοντελοποίηση ενεργού σχήματος
Miotto 2017 [138]	απεικονιστικές, EHR, γονιδιωματική, δεδομένα από φορητές συσκευές και η ανάγκη για αύξηση της ερμηνευσιμότητας
Krittanawong 2017 [139]	μελύτες πάλω στις τεχνολογίες αναγνώρισης εικόνας που προβλέπουν καλύτερα από τους ειδικούς
Lijiang 2017 [140]	ταξινόμηση εικόνων, εντοπισμός αντικειμένου, τμηματοποίηση και εγγραφή (registration)
Qayyum 2017 [141]	μέθοδοι βασισμένοι στο CNN στην τμηματοποίηση εικόνας, ταξινόμηση, διάγνωση και ανάκτηση εικόνας
Hengling 2017 [142]	επίτωση που θα έχει η μηχανική μάθηση στο μέλλον της καρδιαγγειακής απεικόνισης
Blair 2017 [143]	πρόοδος στη νευροεπιστήμη με MRI στην νόσο των μικρών αγγείων
Slomka 2017 [144]	πυρηνική καρδιολογία, CT αγγειογραφία Ηχοκαρδιολογράφημα, MRI
Campeiro 2017 [145]	μαστογραφία, καρδιαγγειακή και μικροσκοπική απεικόνιση
Johnson 2018 [146]	AI στην καρδιολογία, έννοιες προβλεπτικών μοντέλων, κοινά αγγόριθμοι και χρήση της βαθιάς μάθησης
Jiang 2017 [147]	εφαρμογές του AI στην αναγνώριση της καρδιακής προσβολής, διάγνωση, θεραπεία, πρόβλεψη και εκτίμηση πρόγνωσης
Lee 2017 [148]	AI στην απεικονιστική καρδιακής προσβολής εστιασμένο στις τεχνικές αρχές και κλινικές εφαρμογές
Loh 2017 [149]	διάγνωση καρδιακής ασθένειας και διαχείριση μέσα στο πλαίσιο της υγειονομικής περίθαλψης
Krittanawong 2017 [150]	καρδιαγγειακή κλινική περίθαλψη και ο ρόλος της ακριβής καρδιαγγειακής ιατρικής
Gomez 2018 [151]	πρόσφατη πρόοδος στην αυτοματοποίηση και ποσοτική ανάλυση στην πυρηνική καρδιολογία
Shamner 2018 [152]	υποσχέσεις και περιορισμοί της υλοποίησης της μηχανικής μάθησης στην καρδιαγγειακή ιατρική
Shrestha 2018 [153]	εφαρμογές μηχανικής μάθησης στην πυρηνική καρδιολογία
Kikuchi 2018 [154]	εφαρμογές του AI στην πυρηνική καρδιολογία και το πρόβλημα των περιορισμένων αριθμών δεδομένων
Awan 2018 [155]	εφαρμογές μηχανικής μάθησης στην διάγνωση καρδιακής ανεπάρκειας, ταξινόμηση και πρόβλεψη επανεισοδολογίας
Faust 2018 [156]	εφαρμογές βαθιάς μάθησης στα φυσιολογικά δεδομένα συμπεραδακμιασμένου του ECG

δεδομένα, λόγω του ότι η διαδικασία επισήμανσης των ιατρικών δεδομένων είναι δαπανηρή; απαιτεί χειρωνακτική εργασία από ιατρικούς εμπειρογνώμονες. Επιπλέον, τα περισσότερα ιατρικά δεδομένα ανήκουν στις κανονικές περιπτώσεις αντί στις μη-κανονικές, καθιστώντας τα εξαιρετικά μη-ισορροπημένα. Άλλες προκλήσεις της εφαρμογής της βαθιάς μάθησης στην ιατρική που έχει εντοπίσει η βιβλιογραφία είναι τα ζητήματα τυποποίησης/διαθεσιμότητας/διαστάσεων/όγκου/ποιότητας, δυσκολία στην απόκτηση των αντίστοιχων επισημάνσεων και θόρυβος στις επισημάνσεις [140, 137, 138, 144]. Πιο συγκεκριμένα, στο [143] οι συγγραφείς σημειώνουν ότι οι εφαρμογές βαθιάς μάθησης στην ασθένεια μικρών αγγείων, έχουν αναπτυχθεί χρησιμοποιώντας μόνο λίγα αντιπροσωπευτικά σύνολα δεδομένων και πρέπει να αξιολογηθούν σε μεγάλα σύνολα δεδομένων πολλαπλών κέντρων. Οι Kikuchi et al. [154] αναφέρουν ότι σε σύγκριση με τα CT και MRI, οι απεικονιστικές της πυρηνικής καρδιολογίας έχουν περιορισμένο αριθμό εικόνων ανά ασθενή και απεικονίζεται μόνο συγκεκριμένος αριθμός οργάνων. Ο Liebeskind [158] αναφέρει ότι οι μέθοδοι μηχανικής μάθησης δοκιμάζονται σε επίλεκτα και ομοιογενή κλινικά δεδομένα, αλλά η γενικευσιμότητα θα προέκυπτε χρησιμοποιώντας ετερογενή και πολύπλοκα δεδομένα. Το ισχαιμικό αγγειακό εγκεφαλικό επεισόδιο αναφέρεται ως παράδειγμα μιας ετερογενούς και σύνθετης ασθένειας, όπου η απόφραξη της μεσαίας εγκεφαλικής αρτηρίας μπορεί να οδηγήσει σε αποκλίνοσες μορφές απεικόνισης. Στο [151] οι συγγραφείς καταλήγουν στο συμπέρασμα ότι απαιτούνται πρόσθετα δεδομένα που επικυρώνουν αυτές τις εφαρμογές σε μη-ελεγχόμενες κλινικές ρυθμίσεις πολλαπλών κέντρων, πριν από την κλινική εφαρμογή τους. Πρέπει επίσης να διερευνηθεί ο αντίκτυπος αυτών των εργαλείων στην λήψη αποφάσεων, χρήση πόρων, και κόστος. Επιπλέον, η παρούσα βιβλιογραφία έδειξε ότι υπάρχει μια άνιση κατανομή των διαθέσιμων στο κοινό βάσεων δεδομένων ανάμεσα στις διάφορες απεικονιστικές τεχνικές στην καρδιολογία (π.χ. δεν υπάρχει διαθέσιμη δημόσια βάση δεδομένων για τα OCT σε αντίθεση με τα MRI).

Η προηγούμενη βιβλιογραφία αναφέρει ότι τα προβλήματα που σχετίζονται με τα δεδομένα μπορούν να λυθούν με τεχνικές επαύξησης δεδομένων, την ανοικτή συνεργασία μεταξύ των ερευνητικών οργανισμών και την αύξηση της χρηματοδότησης. Οι Hengling et al. [142] υποστηρίζουν ότι θα χρειαστούν σημαντικές επενδύσεις για τη δημιουργία επισημασμένων βάσεων δεδομένων υψηλής ποιότητας, οι οποίες είναι απαραίτητες για την επιτυχία των επιβλεπόμενων μεθόδων βαθιάς μάθησης. Στο [136] οι συγγραφείς υποστηρίζουν ότι η επιτυχία αυτού του τομέα εξαρτάται από τις τεχνολογικές εξελίξεις στην πληροφορική και την αρχιτεκτονική των υπολογιστών, καθώς και τη συνεργασία και την ανοικτή ανταλλαγή δεδομένων μεταξύ των γιατρών και άλλων ενδιαφερομένων. Οι Lee et al. [148] καταλήγουν στο συμπέρασμα ότι απαιτείται διεθνής συνεργασία για την κατασκευή ενός υψηλής ποιότητας πολυτροπικού συνόλου δεδομένων για απεικόνιση εγκεφαλικών επεισοδίων. Μια άλλη λύση για την καλύτερη αξιοποίηση

των δεδομένων μεγάλου όγκου στην καρδιολογία είναι η εφαρμογή μη-επιβλεπόμενων μεθόδων μάθησης, οι οποίες δεν απαιτούν επισημάνσεις. Η παρούσα βιβλιογραφική αναφορά έδειξε ότι η μη-επιβλεπόμενη μάθηση δεν χρησιμοποιείται ευρέως, καθώς η πλειονότητα των μεθόδων σε όλες τις απεικονιστικές τεχνικές είναι επιβλεπόμενες.

Όσον αφορά το πρόβλημα της έλλειψης ερμηνευσιμότητας όπως υποδεικνύεται από τον Hinton [159], είναι γενικά ανέφικτο να ερμηνευτούν τα μη-γραμμικά χαρακτηριστικά των βαθιών δικτύων, επειδή το νόημά τους εξαρτάται από πολύπλοκες αλληλεπιδράσεις με μη-ερμηνεύσιμα χαρακτηριστικά από άλλα επίπεδα. Επιπλέον, αυτά τα μοντέλα είναι στοχαστικά, που σημαίνει ότι κάθε φορά που ένα δίκτυο εκπαιδεύεται με τα ίδια δεδομένα αλλά με διαφορετικά αρχικά βάρη, μαθαίνει διαφορετικά χαρακτηριστικά. Πιο συγκεκριμένα σε μια εκτενή ανασκόπηση [134] του αν επιλύεται το πρόβλημα της τμηματοποίησης LV/RV, οι συγγραφείς δηλώνουν ότι αν και η πτυχή ταξινόμησης του προβλήματος επιτυγχάνει σχεδόν τέλεια αποτελέσματα, η χρήση ενός 'διαγνωστικού μαύρου κουτιού' δεν μπορεί να ενσωματωθεί στην κλινική πρακτική. Οι Miotto et al. [138] αναφέρουν την ερμηνευσιμότητα ως μία από τις κύριες προκλήσεις που αντιμετωπίζει η κλινική εφαρμογή της βαθιάς μάθησης στην υγειονομική περίθαλψη. Στο [148] οι συγγραφείς σημειώνουν ότι η ιδιότητα του μαύρου κουτιού που έχουν μερικές ΑΙ μέθοδοι, όπως η βαθιά μάθηση, είναι αντίθετη με την έννοια της τεκμηριωμένης ιατρικής και εγείρει νομικά και ηθικά ζητήματα στη χρήση τους στην κλινική πρακτική. Αυτή η έλλειψη ερμηνευσιμότητας είναι ο κύριος λόγος για τον οποίο οι ιατρικοί εμπειρογνώμονες αντιστέκονται στη χρήση αυτών των μοντέλων και υπάρχουν επίσης νομικοί περιορισμοί όσον αφορά την ιατρική χρήση των μη-ερμηνεύσιμων εφαρμογών [144]. Από την άλλη πλευρά, κάθε μοντέλο μπορεί να τοποθετηθεί σε έναν άξονα [160] 'ανθρώπου-μηχανής', συμπεριλαμβανομένων στατιστικών που οι ιατρικοί εμπειρογνώμονες βασίζονται στην καθημερινή λήψη κλινικών αποφάσεων. Για παράδειγμα, οι ανθρώπινες αποφάσεις όπως η επιλογή των μεταβλητών που πρέπει να συμπεριληφθούν στο μοντέλο, η σχέση εξαρτημένων και ανεξάρτητων μεταβλητών και μεταβλητών μετασχηματισμών, μετακινούν τον αλγόριθμο προς τον ανθρώπινο άξονα αποφάσεων, καθιστώντας τον πιο ερμηνεύσιμο αλλά ταυτόχρονα περισσότερο πιο επιρρεπή σε σφάλματα.

Όσον αφορά την επίλυση του προβλήματος ερμηνευσιμότητας, όταν νέες μέθοδοι είναι απαραίτητες οι ερευνητές είναι προτιμητέο να δημιουργούν απλούστερες μεθόδους βαθιάς μάθησης (από-άκρο-σε-άκρο και μη-ensembles) για να αυξήσουν την πιθανότητα κλινικής εφαρμογής τους, ακόμα κι αν αυτό σημαίνει μειωμένη αναφερόμενη ακρίβεια. Υπάρχουν επίσης επιχειρήματα κατά της δημιουργίας νέων μεθόδων, επικεντρώνοντας στην επικύρωση των ήδη υφιστάμενων. Στο [161] οι συγγραφείς καταλήγουν στο συμπέρασμα ότι υπάρχει πληθώρα μοντέλων που προβλέπουν περιστατικά CVD στο γενικό πληθυσμό. Η χρησιμότητα των περισσότερων μοντέλων είναι ασαφής λόγω



σφαλμάτων στη μεθοδολογία και έλλειψης εξωτερικών μελετών επικύρωσης. Αντί να αναπτυχθούν νέα μοντέλα πρόβλεψης κινδύνου CVD, η μελλοντική έρευνα πρέπει να επικεντρωθεί στην επικύρωση και σύγκριση υφιστάμενων μοντέλων και να διερευνήσει εάν μπορούν να βελτιωθούν περαιτέρω.

Μια δημοφιλής μέθοδος που χρησιμοποιείται για ερμηνεύσιμα μοντέλα είναι τα δίκτυα προσοχής [162]. Τα δίκτυα προσοχής είχαν ως αρχική έμπνευση την ικανότητα εστίασης της ανθρώπινης όρασης σε ένα συγκεκριμένο σημείο με υψηλή ανάλυση και την αντίληψη του περιβάλλοντος με χαμηλή ανάλυση. Έχουν χρησιμοποιηθεί από αρκετές δημοσιεύσεις στην καρδιολογία στην πρόβλεψη του ιατρικού ιστορικού [163], ταξινόμηση χτύπων ECG [164] και την πρόβλεψη CVD με χρήση Fundus [92]. Ένα άλλο απλούστερο εργαλείο για την ερμηνεία είναι οι χάρτες αξιοπιστίας (saliency maps) [165] που χρησιμοποιούν την κλίση της εξόδου σε σχέση με την είσοδο, η οποία δείχνει διασθητικά τις περιοχές που συνεισφέρουν περισσότερο στην έξοδο.

Εκτός από την επίλυση των προβλημάτων των δεδομένων και της ερμηνείας, οι ερευνητές στην καρδιολογία θα μπορούσαν να χρησιμοποιήσουν τις ήδη καθιερωμένες αρχιτεκτονικές βαθιάς μάθησης που δεν έχουν εφαρμοστεί ευρέως στην καρδιολογία, όπως τα δίκτυα καψουλών (Capsule Networks, CapsNets). Τα CapsNet [166] είναι βαθιά νευρωνικά δίκτυα που χρειάζονται λιγότερα δεδομένα εκπαίδευσης από τα CNN και τα επίπεδά τους αποτυπώνουν τον 'προσανατολισμό' των χαρακτηριστικών, καθιστώντας έτσι τις εσωτερικές τους λειτουργίες πιο ερμηνεύσιμες και πιο κοντά στον ανθρώπινο τρόπο αντίληψης. Εντούτοις, ένα σημαντικό μειονέκτημα που έχουν, το οποίο τους περιορίζει από την επίτευξη ευρύτερης χρήσης, είναι το υψηλό υπολογιστικό κόστος σε σύγκριση με τα CNN λόγω του αλγορίθμου 'δρομολόγησης με συμφωνία'. Μεταξύ των πρόσφατων χρήσεών τους στην ιατρική περιλαμβάνονται η ταξινόμηση όγκων στον εγκέφαλο [167] και η ταξινόμηση του καρκίνου του μαστού [168]. Τα CapsNet δεν έχουν χρησιμοποιηθεί ακόμη σε καρδιολογικά δεδομένα.

Μια άλλη λιγότερο χρησιμοποιούμενη αρχιτεκτονική βαθιάς μάθησης στην καρδιολογία είναι τα GANs [169], τα οποία αποτελούνται από ένα δίκτυο γεννήτρια που δημιουργεί ψεύτικες εικόνες από θόρυβο και ένα δίκτυο διάκρισης που είναι υπεύθυνο για τη διαφοροποίηση μεταξύ πλαστών εικόνων από τη γεννήτρια και πραγματικών εικόνων. Και τα δύο δίκτυα προσπαθούν να βελτιστοποιήσουν μια απώλεια σε ένα παιχνίδι με μηδενικό άθροισμα, με αποτέλεσμα μια γεννήτρια που παράγει ρεαλιστικές εικόνες. Τα GAN έχουν χρησιμοποιηθεί μόνο για την προσομοίωση παθο-ρεαλιστικών εικόνων IVUS [129] και ο τομέας της καρδιολογίας έχει πολλά να κερδίσει από τη χρήση αυτού του είδους των μοντέλων, ειδικά λόγω της έλλειψης επισήμασμένων δεδομένων υψηλής ποιότητας.

Οι ερευνητές θα μπορούσαν επίσης να χρησιμοποιήσουν CRF, τα οποία είναι γραφικά μοντέλα που συλλαμβάνουν πληροφορίες από γειτονικές περιοχές και είναι

σε θέση να ενσωματώσουν στατιστικά στοιχεία υψηλότερης τάξης, τα οποία οι παραδοσιακές μέθοδοι βαθιάς μάθησης δεν είναι σε θέση να κάνουν. Εκπαιδευμένα CRFs από κοινού με CNNs έχουν χρησιμοποιηθεί σε εκτίμηση του βάθους στην ενδοσκόπηση [170] και την κατάτμηση του ήπατος στην CT [171]. Υπάρχουν επίσης εφαρμογές καρδιολογίας που χρησιμοποίησαν CRF με βαθιά μάθηση για βελτίωση της τμηματοποίησης στα Fundus [70] και σε LV/RV [43]. Η πολυτροπική βαθιά μάθηση [172] μπορεί επίσης να χρησιμοποιηθεί για τη βελτίωση των διαγνωστικών αποτελεσμάτων, π.χ. τη δυνατότητα συνδυασμού δεδομένων fMRI και ECG. Ειδικές βάσεις δεδομένων πρέπει να δημιουργηθούν για να αυξηθεί η έρευνα στον τομέα αυτό, καθώς σύμφωνα με την τρέχουσα ανασκόπηση υπάρχουν μόνο τρεις βάσεις δεδομένων καρδιολογίας με πολυτροπικά δεδομένα. Εκτός από τις προηγούμενες βάσεις δεδομένων, η MIMIC-III έχει επίσης χρησιμοποιηθεί για πολυτροπική βαθιά μάθηση από [173] για την πρόβλεψη ενδονοσοκομειακής, βραχυπρόθεσμης/μακροπρόθεσμης θνησιμότητας και πρόβλεψης του κώδικα ICD-9.

## Βιβλιογραφία

- [1] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright. Evaluation framework for algorithms segmenting short axis cardiac MRI. In *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009.
- [2] Carissa G Fonseca, Michael Backhaus, David A Bluemke, Randall D Britten, Jae Do Chung, Brett R Cowan, Ivo D Dinov, J Paul Finn, Peter J Hunter, Alan H Kadish, et al. The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*, 27(16):2288–2295, 2011.
- [3] Caroline Petitjean, Maria A Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jérôme Caudron, Su Ruan, Ismail Ben Ayed, M Jorge Cardoso, Hsiang-Chou Chen, et al. Right ventricle segmentation from cardiac mri: a collation study. *Medical image analysis*, 19(1):187–202, 2015.
- [4] Andrew Asman, Alireza Akhondi-Asl, Hongzhi Wang, Nicholas Tustison, Brian Avants, Simon K Warfield, and Bennett Landman. Miccai 2013 segmentation algorithms, theory and applications (sata) challenge results summary. In *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*, 2013.
- [5] Danielle F Pace, Adrian V Dalca, Tal Geva, Andrew J Powell, Mehdi H Moghari, and Polina Golland. Interactive whole-heart segmentation in congenital heart disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 80–88. Springer, 2015.
- [6] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 2018.
- [7] Alexander Andreopoulos and John K Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri. *Medical Image Analysis*, 12(3):335–357, 2008.
- [8] Data science bowl cardiac challenge data, 2016.
- [9] Jiong Zhang, Behdad Dashtbozorg, Erik Bekkers, Josien PW Pluim, Remco Duits, and Bart M ter Haar Romeny. Robust retinal vessel segmentation via

- locally adaptive derivative frames in orientation scores. *IEEE transactions on medical imaging*, 35(12):2631–2644, 2016.
- [10] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [11] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
- [12] Christopher G Owen, Alicja R Rudnicka, Robert Mullen, Sarah A Barman, Dorothy Monekosso, Peter H Whincup, Jeffrey Ng, and Carl Paterson. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. *Investigative ophthalmology & visual science*, 50(5):2004–2010, 2009.
- [13] Jan Odstrcilik, Radim Kolar, Attila Budai, Joachim Hornegger, Jiri Jan, Jiri Gazarek, Tomas Kubena, Pavel Cernosek, Ondrej Svoboda, and Elli Angelopoulou. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, 2013.
- [14] Ben Graham. Kaggle diabetic retinopathy detection competition report. *University of Warwick*, 2015.
- [15] Etienne Decencière, Guy Cazuguel, Xiwei Zhang, Guillaume Thibault, J-C Klein, Fernand Meyer, Beatriz Marcotegui, Gwénolé Quéllec, Mathieu Lamard, Ronan Danno, et al. Teleophta: Machine learning and image processing methods for teleophthalmology. *Irbm*, 34(2):196–203, 2013.
- [16] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, et al. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014.
- [17] Tomi Kauppi, Joni-Kristian Kämäräinen, Lasse Lensu, Valentina Kalesnykiene, Iris Sorri, Hannu Uusitalo, and Heikki Kälviäinen. Constructing benchmark databases and protocols for medical image analysis: Diabetic retinopathy. *Computational and mathematical methods in medicine*, 2013, 2013.

- [18] Meindert Niemeijer, Bram Van Ginneken, Michael J Cree, Atsushi Mizutani, Gwénoélé Quéllec, Clara I Sánchez, Bob Zhang, Roberto Hornero, Mathieu Lamard, Chisako Muramatsu, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE transactions on medical imaging*, 29(1):185–195, 2010.
- [19] Simone Balocco, Carlo Gatta, Francesco Ciompi, Andreas Wahle, Petia Radeva, Stephane Carlier, Gozde Unal, Elias Sanidas, Josepa Mauri, Xavier Carillo, et al. Standardized evaluation methodology and reference database for evaluating ivus image segmentation. *Computerized medical imaging and graphics*, 38(2):70–90, 2014.
- [20] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.
- [21] HA Kirişli, Michiel Schaap, CT Metz, AS Dharampal, Willem Bob Meijboom, Stella-Lida Papadopoulou, Admir Dedic, Koen Nieman, MA De Graaf, MFL Meijs, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Medical image analysis*, 17(8):859–876, 2013.
- [22] Peter H Charlton, Timothy Bonnici, Lionel Tarassenko, David A Clifton, Richard Beale, and Peter J Watkinson. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological measurement*, 37(4):610, 2016.
- [23] Catalina Tobon-Gomez, Arjan J Geers, Jochen Peters, Jürgen Weese, Karen Pinto, Rashed Karim, Mohammed Ammar, Abdelaziz Daoudi, Jan Margeta, Zulma Sandoval, et al. Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets. *IEEE transactions on medical imaging*, 34(7):1460–1473, 2015.
- [24] Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016.
- [25] Giovanni Sebastiani and Piero Barone. Mathematical principles of basic magnetic resonance imaging in medicine. *Signal Processing*, 25(2):227–250, 1991.

- [26] Li Kuo Tan, Yih Miin Liew, Einly Lim, and Robert A McLaughlin. Cardiac left ventricle segmentation using convolutional neural network regression. In *Biomedical Engineering and Sciences (IECBES), 2016 IEEE EMBS Conference on*, pages 490–493. IEEE, 2016.
- [27] Liset Vázquez Romaguera, Marly Guimarães Fernandes Costa, Francisco Perdigón Romero, and Cicero Ferreira Fernandes Costa Filho. Left ventricle segmentation in cardiac mri images using fully convolutional neural networks. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 101342Z. International Society for Optics and Photonics, 2017.
- [28] Rudra PK Poudel, Pablo Lamata, and Giovanni Montana. Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 83–94. Springer, 2016.
- [29] Christian Rupprecht, Elizabeth Huaroc, Maximilian Baust, and Nassir Navab. Deep active contours. *arXiv preprint arXiv:1607.05074*, 2016.
- [30] Tuan Anh Ngo and Gustavo Carneiro. Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3125, 2014.
- [31] MR Avendi, Arash Kheradvar, and Hamid Jafarkhani. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. *Medical image analysis*, 30:108–119, 2016.
- [32] Heran Yang, Jian Sun, Huibin Li, Lisheng Wang, and Zongben Xu. Deep fusion net for multi-atlas segmentation: Application to cardiac mr images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 521–528. Springer, 2016.
- [33] Gongning Luo, Suyu Dong, Kuanquan Wang, and Henggui Zhang. Cardiac left ventricular volumes prediction method based on atlas location and deep learning. In *Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on*, pages 1604–1610. IEEE, 2016.
- [34] Xulei Yang, Zeng Zeng, and Su Yi. Deep convolutional neural networks for automatic segmentation of left ventricle cavity from cardiac magnetic resonance images. *IET Computer Vision*, 11(8):643–649, 2017.

- [35] Li Kuo Tan, Yih Miin Liew, Einly Lim, and Robert A McLaughlin. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine mr sequences. *Medical image analysis*, 39:78–86, 2017.
- [36] Ariel H Curiale, Flavio D Colavecchia, Pablo Kaluza, Roberto A Isoardi, and German Mato. Automatic myocardial segmentation by using a deep learning network in cardiac mri. In *Computer Conference (CLEI), 2017 XLIII Latin American*, pages 1–6. IEEE, 2017.
- [37] Fangzhou Liao, Xi Chen, Xiaolin Hu, and Sen Song. Estimation of the volume of the left ventricle from mri images using deep neural networks. *IEEE Transactions on Cybernetics*, 2017.
- [38] Omar Emad, Inas A Yassine, and Ahmed S Fahmy. Automatic localization of the left ventricle in cardiac mri images using deep learning. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 683–686. IEEE, 2015.
- [39] Clément Zotti, Zhiming Luo, Olivier Humbert, Alain Lalande, and Pierre-Marc Jodoin. Gridnet with automatic shape prior registration for automatic mri cardiac segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 73–81. Springer, 2017.
- [40] Jay Patravali, Shubham Jain, and Sasank Chilamkurthy. 2d-3d fully convolutional neural networks for cardiac mr segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 130–139. Springer, 2017.
- [41] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 120–129. Springer, 2017.
- [42] Phi Vu Tran. A fully convolutional neural network for cardiac segmentation in short-axis mri. *arXiv preprint arXiv:1604.00494*, 2016.
- [43] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2017.

- [44] Jesse Lieman-Sifry, Matthieu Le, Felix Lau, Sean Sall, and Daniel Golden. Fastventricle: Cardiac segmentation with enet. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 127–138. Springer, 2017.
- [45] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.
- [46] Hinrich B Winther, Christian Hundt, Bertil Schmidt, Christoph Czerner, Johann Bauersachs, Frank Wacker, and Jens Vogel-Claussen. nu-net: Deep learning for generalized biventricular cardiac mass and function parameters. *arXiv preprint arXiv:1706.04397*, 2017.
- [47] Xiuquan Du, Weiwei Zhang, Heye Zhang, Jun Chen, Yanping Zhang, James Claude Warrington, Gary Brahm, and Shuo Li. Deep regression segmentation for cardiac bi-ventricle mr images. *IEEE Access*, 2018.
- [48] Archontis Giannakidis, Konstantinos Kamnitsas, Veronica Spadotto, Jennifer Keegan, Gillian Smith, Ben Glocker, Daniel Rueckert, Sabine Ernst, Michael A Gatzoulis, Dudley J Pennell, et al. Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3d convolutional neural network. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on*, pages 42–46. IEEE, 2016.
- [49] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.
- [50] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Dilated convolutional neural networks for cardiovascular mr segmentation in congenital heart disease. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 95–102. Springer, 2016.
- [51] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial Intelligence and Statistics*, pages 562–570, 2015.
- [52] Jinpeng Li, Rongzhao Zhang, Lin Shi, and Defeng Wang. Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3d fcn.



- In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 111–118. Springer, 2016.
- [53] Lequan Yu, Xin Yang, Jing Qin, and Pheng-Ann Heng. 3d fractalnet: dense volumetric segmentation for cardiovascular mri volumes. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 103–110. Springer, 2016.
- [54] Christian Payer, Darko Štern, Horst Bischof, and Martin Urschler. Multi-label whole heart segmentation using cnns and anatomical label configurations. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 190–198. Springer, 2017.
- [55] Aliasghar Mortazi, Jeremy Burt, and Ulas Bagci. Multi-planar deep segmentation networks for cardiac substructures from mri and ct. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 199–206. Springer, 2017.
- [56] Xin Yang, Cheng Bian, Lequan Yu, Dong Ni, and Pheng-Ann Heng. Hybrid loss guided convolutional networks for whole heart parsing. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 215–223. Springer, 2017.
- [57] Guang Yang, Xiahai Zhuang, Habib Khan, Shouvik Haldar, Eva Nyktari, Xu-jiong Ye, Greg Slabaugh, Tom Wong, Raad Mohiaddin, Jennifer Keegan, et al. Segmenting atrial fibrosis from late gadolinium-enhanced cardiac mri by deep-learned features with stacked sparse auto-encoders. In *Annual Conference on Medical Image Understanding and Analysis*, pages 195–206. Springer, 2017.
- [58] Le Zhang, Ali Gooya, Bo Dong, Rui Hua, Steffen E Petersen, Pau Medrano-Gracia, and Alejandro F Frangi. Automated quality assessment of cardiac mr images using convolutional neural networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 138–145. Springer, 2016.
- [59] Bin Kong, Yiqiang Zhan, Min Shin, Thomas Denny, and Shaoting Zhang. Recognizing end-diastole and end-systole frames via deep temporal regression network. In *International conference on medical image computing and computer-assisted intervention*, pages 264–272. Springer, 2016.
- [60] Fan Yang, Yan He, Mubashir Hussain, Hong Xie, and Pinggui Lei. Convolutional neural network for the detection of end-diastole and end-systole frames in free-breathing cardiac magnetic resonance imaging. *Computational and mathematical methods in medicine*, 2017, 2017.

- [61] Chenchu Xu, Lei Xu, Zhifan Gao, Shen Zhao, Heye Zhang, Yanping Zhang, Xiuquan Du, Shu Zhao, Dhanjoo Ghista, and Shuo Li. Direct detection of pixel-level myocardial infarction areas via a deep-learning algorithm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 240–249. Springer, 2017.
- [62] Wufeng Xue, Gary Brahm, Sachin Pandey, Stephanie Leung, and Shuo Li. Full left ventricle quantification via deep multitask relationships learning. *Medical image analysis*, 43:54–65, 2018.
- [63] Xiantong Zhen, Zhijie Wang, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li. Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. *Medical image analysis*, 30:120–129, 2016.
- [64] Carlo Biffi, Ozan Oktay, Giacomo Tarroni, Wenjia Bai, Antonio De Marvao, Georgia Doumou, Martin Rajchl, Reem Bedair, Sanjay Prasad, Stuart Cook, et al. Learning interpretable anatomical features through deep generative models: Application to cardiac remodeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 464–471. Springer, 2018.
- [65] Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O’Regan, and Daniel Rueckert. Multi-input cardiac image super-resolution using convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 246–254. Springer, 2016.
- [66] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A Cook, Antonio de Marvao, Timothy Dawes, Declan P O’Regan, et al. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging*, 37(2):384–395, 2018.
- [67] Ender Konukoglu. An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers*, volume 10663, page 111. Springer, 2018.
- [68] Michael D Abràmoff, Mona K Garvin, and Milan Sonka. Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, 3:169–208, 2010.

- [69] Shuangling Wang, Yilong Yin, Guibao Cao, Benzhen Wei, Yuanjie Zheng, and Gongping Yang. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, 149:708–717, 2015.
- [70] Lei Zhou, Qi Yu, Xun Xu, Yun Gu, and Jie Yang. Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement. *Computer methods and programs in biomedicine*, 148:13–25, 2017.
- [71] Yongliang Chen. A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation. *arXiv preprint arXiv:1704.07502*, 2017.
- [72] Debapriya Maji, Anirban Santara, Pabitra Mitra, and Debdoot Sheet. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. *arXiv preprint arXiv:1603.04833*, 2016.
- [73] Huazhu Fu, Yanwu Xu, Damon Wing Kee Wong, and Jiang Liu. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 698–701. IEEE, 2016.
- [74] Aaron Wu, Ziyue Xu, Mingchen Gao, Mario Buty, and Daniel J Mollura. Deep vessel tracking: A generalized probabilistic approach via deep learning. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1363–1367. IEEE, 2016.
- [75] Qiaoliang Li, Bowei Feng, LinPei Xie, Ping Liang, Huisheng Zhang, and Tianfu Wang. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE transactions on medical imaging*, 35(1):109–118, 2016.
- [76] Avisek Lahiri, Abhijit Guha Roy, Debdoot Sheet, and Prabir Kumar Biswas. Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 1340–1343. IEEE, 2016.
- [77] Americo Oliveira, Sérgio Pereira, and Carlos A Silva. Augmenting data when training a cnn for retinal vessel segmentation: How to warp? In *Bioengineering (ENBENG), 2017 IEEE 5th Portuguese Meeting on*, pages 1–4. IEEE, 2017.
- [78] Henry A Leopold, Jeff Orchard, John Zelek, and Vasudevan Lakshminarayanan. Use of gabor filters and deep networks in the segmentation

- of retinal vessel morphology. In *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XV*, volume 10068, page 100680R. International Society for Optics and Photonics, 2017.
- [79] Henry A Leopold, Jeff Orchard, John S Zelek, and Vasudevan Lakshminarayanan. Pixelbnn: Augmenting the pixelcnn with batch normalization and the presentation of a fast architecture for retinal vessel segmentation. *arXiv preprint arXiv:1712.06742*, 2017.
- [80] Juan Mo and Lei Zhang. Multi-level deep supervised networks for retinal vessel segmentation. *International journal of computer assisted radiology and surgery*, 12(12):2181–2193, 2017.
- [81] Martina Melinščak, Pavle Prentašić, and Sven Lončarić. Retinal vessel segmentation using deep neural networks. In *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*, 2015.
- [82] Abdulkadir Sengur, Yanhui Guo, Ümit Budak, and Lucas J Vespa. A retinal vessel detection approach using convolution neural network. In *Artificial Intelligence and Data Processing Symposium (IDAP), 2017 International*, pages 1–4. IEEE, 2017.
- [83] Maria Ines Meyer, Pedro Costa, Adrian Galdran, Ana Maria Mendonça, and Aurélio Campilho. A deep neural network for vessel segmentation of scanning laser ophthalmoscopy images. In *International Conference Image Analysis and Recognition*, pages 507–515. Springer, 2017.
- [84] Henry Leopold, Jeff Orchard, John Zelek, and Vasudevan Lakshminarayanan. Segmentation and feature extraction of retinal vascular morphology. In *Medical Imaging 2017: Image Processing*, volume 10133, page 101330V. International Society for Optics and Photonics, 2017.
- [85] Mrinal Haloi. Improved microaneurysm detection using deep neural networks. *arXiv preprint arXiv:1505.04424*, 2015.
- [86] Luca Giancardo, Kirk Roberts, and Zhongming Zhao. Representation learning for retinal vasculature embeddings. In *Fetal, Infant and Ophthalmic Medical Image Analysis*, pages 243–250. Springer, 2017.
- [87] José Ignacio Orlando, Elena Prokofyeva, Mariana del Fresno, and Matthew B Blaschko. An ensemble deep learning based approach for red lesion detection in fundus images. *Computer methods and programs in biomedicine*, 153:115–127, 2018.

- [88] Mark JJP van Grinsven, Bram van Ginneken, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE transactions on medical imaging*, 35(5):1273–1284, 2016.
- [89] Fantin Girard and Farida Cheriet. Artery/vein classification in fundus images using cnn and likelihood score propagation. In *Signal and Information Processing (GlobalSIP), 2017 IEEE Global Conference on*, pages 720–724. IEEE, 2017.
- [90] RA Welikala, PJ Foster, PH Whincup, AR Rudnicka, CG Owen, DP Strachan, SA Barman, et al. Automated arteriole and venule classification using deep learning for retinal images from the uk biobank cohort. *Computers in biology and medicine*, 90:23–32, 2017.
- [91] Harry Pratt, Bryan M Williams, Jae Yee Ku, Charles Vas, Emma McCann, Baidaa Al-Bander, Yitian Zhao, Frans Coenen, and Yalin Zheng. Automatic detection and distinction of retinal vessel bifurcations and crossings in colour fundus photography. *Journal of Imaging*, 4(1):4, 2017.
- [92] Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. *arXiv preprint arXiv:1708.09843*, 2017.
- [93] Harry Pratt, Bryan M Williams, Jae Ku, Frans Coenen, and Yalin Zheng. Automatic detection and identification of retinal vessel junctions in colour fundus photography. In *Annual Conference on Medical Image Understanding and Analysis*, pages 27–37. Springer, 2017.
- [94] Honghui Liu, Jianjiang Feng, Zishun Feng, Jiwen Lu, and Jie Zhou. Left atrium segmentation in ct volumes with fully convolutional networks. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 39–46. Springer, 2017.
- [95] Nikolas Lessmann, Ivana Išgum, Arnaud AA Setio, Bob D de Vos, Francesco Ciompi, Pim A de Jong, Matthjis Oudkerk, P Th M Willem, Max A Viergever, and Bram van Ginneken. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest ct. In *Medical Imaging 2016: Computer-Aided Diagnosis*, volume 9785, page 978511. International Society for Optics and Photonics, 2016.

- [96] Ran Shadmi, Victoria Mazo, Orna Bregman-Amitai, and Eldad Elnekave. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest ct. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 24–28. IEEE, 2018.
- [97] Carlos Cano-Espinosa, Germán González, George R Washko, Miguel Cazorla, and Raúl San José Estépar. Automated agatston score computation in non-ecg gated ct scans using deep learning. In *Proceedings of SPIE—the International Society for Optical Engineering*, volume 10574, 2018.
- [98] Jelmer M Wolterink, Tim Leiner, Bob D de Vos, Robbert W van Hamersvelt, Max A Viergever, and Ivana Išgum. Automatic coronary artery calcium scoring in cardiac ct angiography using paired convolutional neural networks. *Medical image analysis*, 34:123–136, 2016.
- [99] Gianmarco Santini, Daniele Della Latta, Nicola Martini, Gabriele Valvano, Andrea Gori, Andrea Ripoli, Carla L Susini, Luigi Landini, and Dante Chiappino. An automatic deep learning approach for coronary artery calcium segmentation. In *EMBECC & NBC 2017*, pages 374–377. Springer, 2017.
- [100] Karen López-Linares, Luis Kabongo, Nerea Lete, Gregory Maclair, Mario Ceresa, Ainhoa García-Familiar, Iván Macía, and Miguel Ángel González Ballester. Dcnn-based automatic segmentation and quantification of aortic thrombus volume: Influence of the training approach. In *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 29–38. Springer, 2017.
- [101] Ho Aik Hong and UU Sheikh. Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning. In *Signal Processing & Its Applications (CSPA), 2016 IEEE 12th International Colloquium on*, pages 242–246. IEEE, 2016.
- [102] Bob D de Vos, Jelmer M Wolterink, Pim A de Jong, Max A Viergever, and Ivana Išgum. 2d image classification for 3d anatomy localization: employing deep convolutional neural networks. In *Medical Imaging 2016: Image Processing*, volume 9784, page 97841Y. International Society for Optics and Photonics, 2016.
- [103] Mehdi Moradi, Yaniv Gur, Hongzhi Wang, Prasanth Prasanna, and Tanveer Syeda-Mahmood. A hybrid learning approach for semantic labeling of cardiac ct slices and recognition of body position. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1418–1421. IEEE, 2016.

- [104] Yefeng Zheng, David Liu, Bogdan Georgescu, Hien Nguyen, and Dorin Comaniciu. 3d deep learning for efficient and robust landmark detection in volumetric data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 565–572. Springer, 2015.
- [105] Juan C Montoya, Yinsheng Li, Charles Strother, and Guang-Hong Chen. Deep learning angiography (dla): three-dimensional c-arm cone beam ct angiography generated from deep learning method using a convolutional neural network. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105731N. International Society for Optics and Photonics, 2018.
- [106] Majd Zreik, Nikolas Lessmann, Robbert W van Hamersvelt, Jelmer M Wolterink, Michiel Voskuil, Max A Viergever, Tim Leiner, and Ivana Išgum. Deep learning analysis of the myocardium in coronary ct angiography for identification of patients with functionally significant coronary artery stenosis. *Medical image analysis*, 44:72–85, 2018.
- [107] Frederic Commandeur, Markus Goeller, Julian Betancur, Sebastien Cadet, Mhairi Doris, Xi Chen, Daniel S Berman, Piotr J Slomka, Balaji K Tamarrappoo, and Damini Dey. Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast ct. *IEEE Transactions on Medical Imaging*, 2018.
- [108] Mehmet A Gülsün, Gareth Funka-Lea, Puneet Sharma, Saikiran Rapaka, and Yefeng Zheng. Coronary centerline extraction via optimal flow paths and cnn path pruning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 317–325. Springer, 2016.
- [109] Nikolas Lessmann, Bram van Ginneken, Majd Zreik, Pim A de Jong, Bob D de Vos, Max A Viergever, and Ivana Išgum. Automatic calcium scoring in low-dose chest ct using deep neural networks with dilated convolutions. *IEEE Transactions on Medical Imaging*, 2017.
- [110] Majd Zreik, Tim Leiner, Bob D de Vos, Robbert W van Hamersvelt, Max A Viergever, and Ivana Išgum. Automatic segmentation of the left ventricle in cardiac ct angiography using convolutional neural networks. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 40–43. IEEE, 2016.
- [111] Gustavo Carneiro, Jacinto C Nascimento, and António Freitas. The segmentation of the left ventricle of the heart from ultrasound data using deep learning

- architectures and derivative-based search methods. *IEEE Transactions on Image Processing*, 21(3):968–982, 2012.
- [112] Jacinto C Nascimento and Gustavo Carneiro. Multi-atlas segmentation using manifold learning with deep belief networks. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 867–871. IEEE, 2016.
- [113] Hao Chen, Yefeng Zheng, Jin-Hyeong Park, Pheng-Ann Heng, and S Kevin Zhou. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 487–495. Springer, 2016.
- [114] Ali Madani, Ramy Arnaout, Mohammad Mofrad, and Rima Arnaout. Fast and accurate view classification of echocardiograms using deep learning. *npj Digital Medicine*, 1(1):6, 2018.
- [115] João Figueira Silva, Jorge Miguel Silva, António Guerra, Sérgio Matos, and Carlos Costa. Ejection fraction classification in transthoracic echocardiography using a deep learning approach. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 123–128. IEEE, 2018.
- [116] Xiaohong Gao, Wei Li, Martin Loomes, and Lianyi Wang. A fused deep learning architecture for viewpoint classification of echocardiography. *Information Fusion*, 36:103–113, 2017.
- [117] Amir H Abdi, Christina Luong, Teresa Tsang, John Jue, Ken Gin, Darwin Yeung, Dale Hawley, Robert Rohling, and Purang Abolmaesumi. Quality assessment of echocardiographic cine using recurrent neural networks: Feasibility on five standard view planes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 302–310. Springer, 2017.
- [118] Florin C Ghesu, Edward Krubasik, Bogdan Georgescu, Vivek Singh, Yefeng Zheng, Joachim Hornegger, and Dorin Comaniciu. Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE transactions on medical imaging*, 35(5):1217–1228, 2016.
- [119] Douglas P Perrin, Alejandra Bueno, Andrea Rodriguez, Gerald R Marx, and J Pedro. Application of convolutional artificial neural networks to echocardiograms for differentiating congenital heart diseases in a pediatric population.



- In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 1013431. International Society for Optics and Photonics, 2017.
- [120] Mehdi Moradi, Yufan Guo, Yaniv Gur, Mohammadreza Negahdar, and Tanveer Syeda-Mahmood. A cross-modality neural network transform for semi-automatic medical image annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–307. Springer, 2016.
- [121] Takashi Kubo, Takashi Akasaka, Junya Shite, Takahiko Suzuki, Shiro Uemura, Bo Yu, Ken Kozuma, Hironori Kitabata, Toshiro Shinke, Maoto Habara, et al. Oct compared with ivus in a coronary lesion assessment: the opus-class study. *JACC: Cardiovascular Imaging*, 6(10):1095–1104, 2013.
- [122] Abhijit Guha Roy, Sailesh Conjeti, Stephane G Carlier, Khalil Houissa, Andreas König, Pranab K Dutta, Andrew F Laine, Nassir Navab, Amin Kattouzian, and Debodoot Sheet. Multiscale distribution preserving autoencoders for plaque detection in intravascular optical coherence tomography. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1359–1362. IEEE, 2016.
- [123] Yan Ling Yong, Li Kuo Tan, Robert A McLaughlin, Kok Han Chee, and Yih Miin Liew. Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography. *Journal of biomedical optics*, 22(12):126005, 2017.
- [124] Mengdi Xu, Jun Cheng, Annan Li, Jimmy Addison Lee, Damon Wing Kee Wong, Akira Taruya, Atsushi Tanaka, Nicolas Foin, and Philip Wong. Fibroatheroma identification in intravascular optical coherence tomography images using deep features. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 1501–1504. IEEE, 2017.
- [125] Atefeh Abdolmanafi, Luc Duong, Nagib Dahdah, and Farida Cheriet. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomedical optics express*, 8(2):1203–1220, 2017.
- [126] Joseph E Parrillo and R Phillip Dellinger. *Critical Care Medicine E-Book: Principles of Diagnosis and Management in the Adult*. Elsevier Health Sciences, 2013.

- [127] Karim Lekadir, Alfiya Galimzianova, Àngels Betriu, Maria del Mar Vila, Laura Igual, Daniel L Rubin, Elvira Fernández, Petia Radeva, and Sandy Napel. A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound. *IEEE journal of biomedical and health informatics*, 21(1):48–55, 2017.
- [128] Nima Tajbakhsh, Jae Y Shin, R Todd Hurst, Christopher B Kendall, and Jianming Liang. Automatic interpretation of carotid intima–media thickness videos using convolutional neural networks. In *Deep Learning for Medical Image Analysis*, pages 105–131. Elsevier, 2017.
- [129] Francis Tom and Debdoot Sheet. Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 1174–1177. IEEE, 2018.
- [130] Juan Wang, Huanjun Ding, Fatemeh Azamian Bidgoli, Brian Zhou, Carlos Iribarren, Sabee Molloy, and Pierre Baldi. Detecting cardiovascular disease from mammograms with deep learning. *IEEE transactions on medical imaging*, 36(5):1172–1181, 2017.
- [131] Xiuming Liu, Shice Wang, Yufeng Deng, and Kuan Chen. Coronary artery calcification (cac) classification with deep convolutional neural networks. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 101340M. International Society for Optics and Photonics, 2017.
- [132] Marco Pavoni, Yongjun Chang, and Örjan Smedby. Image denoising with convolutional neural networks for percutaneous transluminal coronary angioplasty. In *European Congress on Computational Methods in Applied Sciences and Engineering*, pages 255–265. Springer, 2017.
- [133] Jeffrey J Nirschl, Andrew Janowczyk, Eliot G Peyster, Renee Frank, Kenneth B Margulies, Michael D Feldman, and Anant Madabhushi. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of h&e tissue. *PloS one*, 13(4):e0192726, 2018.
- [134] Julian Betancur, Frederic Commandeur, Mahsaw Motlagh, Tali Sharir, Andrew J Einstein, Sabahat Bokhari, Mathews B Fish, Terrence D Ruddy, Philipp Kaufmann, Albert J Sinusas, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion spect: a multicenter study. *JACC: Cardiovascular Imaging*, 2018.

- [135] Viktor Mayer-Schönberger. Big data for cardiology: novel discovery? *European heart journal*, 37(12):996–1001, 2015.
- [136] Christopher Austin and Fred Kusumoto. The application of big data in medicine: current implications and future directions. *Journal of Interventional Cardiac Electrophysiology*, 47(1):51–59, 2016.
- [137] Hayit Greenspan, Bram van Ginneken, and Ronald M Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.
- [138] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 2017.
- [139] C Krittanawong. The rise of artificial intelligence and the uncertain future for physicians, 2017.
- [140] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [141] Adnan Qayyum, Syed Muhammad Anwar, Muhammad Majid, Muhammad Awais, and Majdi Alnowami. Medical image analysis using convolutional neural networks: A review. *arXiv preprint arXiv:1709.02250*, 2017.
- [142] Mir Henglin, Gillian Stein, Pavel V Hushcha, Jasper Snoek, Alexander B Wiltschko, and Susan Cheng. Machine learning approaches in cardiovascular imaging. *Circulation: Cardiovascular Imaging*, 10(10):e005614, 2017.
- [143] Gordon W Blair, Maria Valdez Hernandez, Michael J Thrippleton, Fergus N Doubal, and Joanna M Wardlaw. Advanced neuroimaging of cerebral small vessel disease. *Current treatment options in cardiovascular medicine*, 19(7):56, 2017.
- [144] Piotr J Slomka, Damini Dey, Arkadiusz Sitek, Manish Motwani, Daniel S Berman, and Guido Germano. Cardiac imaging: working towards fully-automated machine analysis & interpretation. *Expert review of medical devices*, 14(3):197–212, 2017.

- [145] Gustavo Carneiro, Yefeng Zheng, Fuyong Xing, and Lin Yang. Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pages 11–32. Springer, 2017.
- [146] Kipp W Johnson, Jessica Torres Soto, Benjamin S Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley, and Joel T Dudley. Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, 71(23):2668–2679, 2018.
- [147] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, pages svn–2017, 2017.
- [148] Eun-Jae Lee, Yong-Hwan Kim, Namkug Kim, and Dong-Wha Kang. Deep into the brain: Artificial intelligence in stroke imaging. *Journal of stroke*, 19(3):277, 2017.
- [149] Brian CS Loh and Patrick HH Then. Deep learning for cardiac computer-aided diagnosis: benefits, issues & solutions. *mHealth*, 3, 2017.
- [150] Chayakrit Krittanawong, HongJu Zhang, Zhen Wang, Mehmet Aydar, and Takeshi Kitai. Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, 69(21):2657–2664, 2017.
- [151] Javier Gomez, Rami Doukky, Guido Germano, and Piotr Slomka. New trends in quantitative nuclear cardiology methods. *Current Cardiovascular Imaging Reports*, 11(1):1, 2018.
- [152] Khader Shameer, Kipp W Johnson, Benjamin S Glicksberg, Joel T Dudley, and Partho P Sengupta. Machine learning in cardiovascular medicine: are we there yet? *Heart*, pages heartjnl–2017, 2018.
- [153] Sirish Shrestha and Partho P Sengupta. Machine learning for nuclear cardiology: The way forward, 2018.
- [154] Akihiro Kikuchi and Takashi Kawakami. Future of artificial intelligence and nuclear cardiology. *Annals of Nuclear Cardiology*, 4(1):79–82, 2018.
- [155] Saqib Ejaz Awan, Ferdous Sohel, Frank Mario Sanfilippo, Mohammed Benamoun, and Girish Dwivedi. Machine learning in heart failure: ready for prime time. *Current opinion in cardiology*, 33(2):190–195, 2018.

- [156] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: a review. *Computer methods and programs in biomedicine*, 161:1–13, 2018.
- [157] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [158] David S Liebeskind. Artificial intelligence in stroke care: Deep learning or superficial insight? *EBioMedicine*, 2018.
- [159] Geoffrey Hinton. Deep learning—a technology with the potential to transform health care. *JAMA*, 2018.
- [160] Andrew L Beam and Isaac S Kohane. Big data and machine learning in health care. *Jama*, 319(13):1317–1318, 2018.
- [161] Johanna AAG Damen, Lotty Hooft, Ewoud Schuit, Thomas PA Debray, Gary S Collins, Ioanna Tzoulaki, Camille M Lassale, George CM Siontis, Virginia Chiochia, Corran Roberts, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *bmj*, 353:i2416, 2016.
- [162] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR*, 2015.
- [163] You Jin Kim, Yun-Geun Lee, Jeong Whun Kim, Jin Joo Park, Borim Ryu, and Jung-Woo Ha. Highrisk prediction from electronic medical records via deep attention networks. *arXiv preprint arXiv:1712.00010*, 2017.
- [164] Patrick Schwab, Gaetano C Scebba, Jia Zhang, Marco Delai, and Walter Karlen. Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks. *Computing*, 44:1, 2017.
- [165] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.
- [166] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3856–3866, 2017.

- [167] Parnian Afshar, Arash Mohammadi, and Konstantinos N Plataniotis. Brain tumor type classification via capsule networks. *arXiv preprint arXiv:1802.10200*, 2018.
- [168] Tomas Iesmantas and Robertas Alzbutas. Convolutional capsule network for classification of breast cancer histology images. In *International Conference Image Analysis and Recognition*, pages 853–860. Springer, 2018.
- [169] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [170] Faisal Mahmood and Nicholas J Durr. Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. *Medical Image Analysis*, 2018.
- [171] Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettliger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D’Anastasi, et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer, 2016.
- [172] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.
- [173] Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmarking deep learning models on large healthcare datasets. *Journal of biomedical informatics*, 83:112–134, 2018.

## 5. Επίπεδα Signal2Image σε βαθιά νευρωνικά δίκτυα για ταξινόμηση EEG

### 5.1. Εισαγωγή

Η βαθιά μάθηση έχει φέρει επανάσταση στην όραση υπολογιστών, χρησιμοποιώντας την αυξημένη διαθεσιμότητα βάσεων δεδομένων μεγάλου όγκου και τη δύναμη των παράλληλων υπολογιστικών μονάδων, όπως είναι οι μονάδες επεξεργασίας γραφικών. Η συντριπτική πλειοψηφία της έρευνας στην βαθιά μάθηση γίνεται με τη χρήση εικόνων ως δεδομένων εκπαίδευσης, ωστόσο ο βιοϊατρικός τομέας είναι πλούσιος σε σήματα φυσιολογίας που χρησιμοποιούνται για προβλήματα διάγνωσης και πρόβλεψης. Είναι ακόμα ανοιχτό ερευνητικό ερώτημα, η καλύτερη αξιοποίηση σημάτων για την εκπαίδευση βαθιών νευρωνικών δικτύων.

Οι περισσότερες μέθοδοι για την επίλυση βιοϊατρικών προβλημάτων μέχρι πρόσφατα, περιελάμβαναν χειροποίητη δημιουργία χαρακτηριστικών και προσπάθεια μίμησης ανθρώπινων εμπειρογνομώνων, τα οποία αποδεικνύονται όλο και περισσότερο ανεπαρκή και επιρρεπή σε σφάλματα. Η βαθιά μάθηση αναδεικνύεται ως μια ισχυρή λύση για ένα ευρύ φάσμα προβλημάτων στη βιοϊατρική, που επιτυγχάνει καλύτερα αποτελέσματα σε σύγκριση με την παραδοσιακή μηχανική μάθηση. Το κύριο πλεονέκτημα των μεθόδων που χρησιμοποιούν τη βαθιά μάθηση είναι ότι μαθαίνουν ιεραρχικά χαρακτηριστικά από δεδομένα εκπαίδευσης με αυτόματο τρόπο, κάτι το οποίο τα καθιστά πιο κλιμακωτά και γενικεύσιμα από παραδοσιακές μεθόδους. Αυτό επιτυγχάνεται με τη χρήση δικτύων πολλαπλών επιπέδων που αποτελούνται από εκατομμύρια παραμέτρους [1], εκπαιδευμένα με backpropagation [2] σε μεγάλο όγκο δεδομένων. Παρόλο που η βαθιά μάθηση χρησιμοποιείται κυρίως σε βιοϊατρικές εικόνες υπάρχει επίσης ένα ευρύ φάσμα φυσιολογικών σημάτων, όπως το EEG, τα οποία χρησιμοποιούνται για προβλήματα διάγνωσης και πρόβλεψης. Το EEG είναι μια μέτρηση του ηλεκτρικού πεδίου που παράγεται από τον εγκέφαλο και χρησιμοποιείται για την ταξινόμηση σταδίων του ύπνου [3], τις

διεπαφές υπολογιστή-εγκεφάλου [4], την συναισθηματική παρακολούθηση [5] και την ταξινόμηση της επιληψίας [6].

Οι Yannick et al. [7] εξέτασαν δημοσιεύσεις βαθιάς μάθησης με χρήση του EEG και εντόπισαν μια γενική αύξηση της ακρίβειας όταν χρησιμοποιείται βαθιά μάθηση και συγκεκριμένα όταν χρησιμοποιούνται CNN αντί παραδοσιακών μεθόδων μηχανικής μάθησης. Ωστόσο δεν αναφέρουν συγκεκριμένα χαρακτηριστικά των αρχιτεκτονικών CNN που ευθύνονται για την αύξηση της απόδοσης. Είναι ακόμα ανοιχτό ερευνητικό ερώτημα, το πως μπορεί να χρησιμοποιηθεί το EEG για την εκπαίδευση μοντέλων βαθιάς μάθησης.

Μια κοινή προσέγγιση που χρησιμοποιήσαν προηγούμενες μελέτες για την ταξινόμηση των σημάτων EEG ήταν η εξαγωγή χαρακτηριστικών από το πεδίο συχνοτήτων και χρόνου-συχνοτήτων, που χρησιμοποιούν τη θεωρία πίσω από τις συχνότητες των ζωνών EEG [8]: δέλτα (0.5–4 Hz), θήτα (4–8 Hz), άλφα (8–13 Hz), βήτα (13–20 Hz) και γάμμα (20–64 Hz). Οι Truong et al. [9] χρησιμοποίησαν STFT σε παράθυρο ολίσθησης 30 δευτερολέπτων, για να εκπαιδεύσουν ένα CNN τριών επιπέδων σε αναπαραστάσεις πεδίου χρόνου-συχνοτήτων, για πρόβλεψη επιληπτικών κρίσεων και αξιολόγησαν τη μέθοδο τους σε τρεις βάσεις δεδομένων EEG. Οι Khan et al. [10] μετασχημάτισαν τα EEG στο πεδίο χρόνου-συχνοτήτων χρησιμοποιώντας κυματίδια πολλαπλών κλιμάκων και έπειτα εκπαιδεύσαν ένα CNN έξι επιπέδων, για την πρόβλεψη της εστιακής έναρξης παρουσιάζοντας υποσχόμενα αποτελέσματα.

Η εξαγωγή χαρακτηριστικών από το πεδίο χρόνου-συχνοτήτων έχει επίσης χρησιμοποιηθεί και σε άλλες δημοσιεύσεις που σχετίζονται με το EEG, εκτός από την πρόγνωση επιληπτικών κρίσεων. Οι Zhang et al. [11] εκπαιδεύσαν ένα ensemble CNN που περιείχε δύο έως δέκα επίπεδα, χρησιμοποιώντας χαρακτηριστικά STFT που εξήχθησαν από τις EEG συχνότητες για ταξινόμηση ψυχικού φόρτου εργασίας. Οι Giri et al. [12] εξήγαγαν στατιστικά και μέτρα πληροφορίας από το πεδίο συχνοτήτων για την εκπαίδευση ενός 1D CNN με δύο επίπεδα, για τον εντοπισμό ισχαιμικού εγκεφαλικού επεισοδίου.

Σε αυτό το κεφάλαιο και στο πλαίσιο της διδακτορικής διατριβής ορίζουμε τα επίπεδα νευρωνικών δικτύων Signal2Image (S2I) ως κάθε δομικό στοιχείο που τοποθετείται μετά την είσοδο μη-επεξεργασμένου σήματος και πριν από ένα 'μοντέλο βάσης' που είναι συνήθως μια καθιερωμένη αρχιτεκτονική για προβλήματα εικόνων. Μία σημαντική ιδιότητα ενός S2I είναι αν αποτελείται από παραμέτρους που μπορούν να εκπαιδευτούν όπως συνελκτικά και γραμμικά επίπεδα ή είναι μη-εκπαιδευσιμα όπως οι παραδοσιακές μέθοδοι χρόνου-συχνοτήτων. Χρησιμοποιώντας αυτόν τον ορισμό μπορούμε επίσης να καταλήξουμε στο συμπέρασμα ότι οι περισσότερες προηγούμενες μέθοδοι για την ταξινόμηση EEG χρησιμοποιούν μη-εκπαιδευσιμα S2I και ότι καμία προηγούμενη μελέτη δεν έχει συγκρίνει εκπαιδευσιμα με μη-εκπαιδευσιμα S2I.



Συγκρίνουμε τα εκπαιδύσιμα και μη-εκπαιδύσιμα S2Is σε συνδυασμό με τις γνωστές αρχιτεκτονικές νευρωνικών δικτύων ‘μοντέλων βάσης’, μαζί με τις 1D και τις σε βάθος παραλλαγές των τελευταίων. Μια επισκόπηση υψηλού επιπέδου αυτών των συνδυασμένων μεθόδων παρουσιάζεται στην Εικ. 5.1. Αν και επιλέγουμε το σύνολο δεδομένων αναγνώρισης επιληπτικών ελλείψεων EEG από το Πανεπιστήμιο της Καλιφόρνιας, Irvine (University of California, UCI) [13] για ταξινόμηση EEG, οι συνέπειες αυτής της μελέτης θα μπορούσαν να γενικευτούν σε οποιοδήποτε είδος προβλήματος ταξινόμησης σημάτων. Εδώ επίσης αναφερόμαστε στο CNN ως ένα νευρωνικό δίκτυο που αποτελείται από εναλλασσόμενα συνελικτικά επίπεδα ακολουθούμενα από ένα ReLU και ένα επίπεδο μέγιστης συγκέντρωσης και ένα πλήρως συνδεδεμένο επίπεδο στο τέλος, ενώ ο όρος ‘επίπεδο’ υποδηλώνει τον αριθμό των συνελικτικών επιπέδων.

## 5.2. Δεδομένα

Η βάση δεδομένων αναγνώρισης επιληπτικών κρίσεων UCI EEG [13] αποτελείται από 500 σήματα το καθένα με 4097 δείγματα (23.5 δευτερόλεπτα). Αποτελείται από πέντε κατηγορίες με 100 σήματα για κάθε κλάση (σε παρένθεση τα συντομευμένα ονόματα των επισημάνσεων που χρησιμοποιούνται στα Σχήματα 5.1 και 5.2):

1. υγιής ασθενής ενώ έχει τα μάτια του ανοιχτά (Open),
2. υγιής ασθενής, έχοντας τα μάτια κλειστά (Closed),
3. ασθενής με όγκο με σήμα που λαμβάνεται από υγιή περιοχή (Healthy),
4. ασθενής με όγκο με σήμα που λαμβάνεται από την περιοχή του όγκου (Tumor),
5. ασθενής ενώ έχει επιληπτική δραστηριότητα (Epilepsy)

Για τους σκοπούς αυτού του κεφαλαίου χρησιμοποιούμε μια παραλλαγή της βάσης δεδομένων<sup>1</sup> στην οποία τα σήματα EEG χωρίζονται σε τμήματα με 178 δείγματα το καθένα, καταλήγοντας σε ένα ισορροπημένο σύνολο δεδομένων που αποτελείται από 11500 σήματα EEG.

## 5.3. Μέθοδοι

### 5.3.1. Ορισμοί

Ορίζουμε το σύνολο δεδομένων  $D = \{x_i, y_i\}_{i=1..N}$  όπου  $x_i \in \mathbb{Z}^n$  και  $y_i \in \{1, 2, 3, 4, 5\}$  υποδηλώνει το  $i^{th}$  σήμα εισόδου με διαστάσεις  $n = 178$  και

<sup>1</sup><https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition>

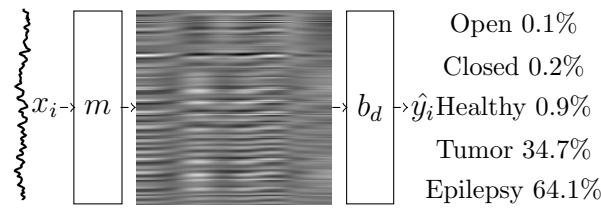


Figure 5.1: Αρχιτεκτονική Signal2Image.  $x_i$  είναι η είσοδος,  $m$  είναι το Signal2Image,  $b_d$  είναι η 1D ή 2D αρχιτεκτονική του ‘μοντέλου βάσης’ για  $d = 1, 2$  αντίστοιχα και  $\hat{y}_i$  είναι η προβλεπόμενη έξοδος. Τα ονόματα των επισημάνσεων απεικονίζονται δεξιά μαζί με τις προβλέψεις για το απεικονιζόμενο σήμα. Η εικόνα μεταξύ του  $m$  και του  $b_d$  απεικονίζει την έξοδο του CNN Signal2Image ενός επιπέδου, ενώ το ‘σήμα ως εικόνα’ και φασματογράφημα έχουν ενδιάμεσες εικόνες όπως απεικονίζονται στην δεύτερη και τρίτη σειρά του σχήματος 5.2. Τα βέλη απεικονίζουν την ροή της προς-τα-εμπρός διάδοσης. Για την 1D αρχιτεκτονική το  $m$  παραλείπεται και δεν δημιουργείται ενδιάμεση εικόνα.

την  $i^{th}$  επισήμανση με πέντε πιθανές κατηγορίες αντίστοιχα.  $N = 11500$  είναι ο αριθμός των παρατηρήσεων.

Επίσης, ορίζουμε το σύνολο των S2Is ως  $M$  και το μέλος αυτού του συνόλου ως  $m$  το οποίο περιλαμβάνει τα ακόλουθα:

- ‘σήμα ως εικόνα’ (μη-εκπαιδευσιμο)
- φασματογράφημα (μη-εκπαιδευσιμο)
- CNN ενός και δύο επιπέδων (εκπαιδευσιμο)

Καθορίζουμε στη συνέχεια το σύνολο των ‘μοντέλων βάσης’ ως  $B$  και το μέλος αυτού του συνόλου ως  $b_d$  όπου  $d = [1, 2]$  υποδηλώνει τη διαστασιμότητα των επιπέδων συνέλιξης (convolutional), μέγιστης συγκέντρωσης (max-pooling) και κανονικοποίησης παρτίδας (batch-norm). Το  $B$  περιλαμβάνει τα παρακάτω  $b_d$  μαζί με τις παραλλαγές σε βάθος και τις ισοδύναμες 1D αρχιτεκτονικές για  $d = 1$  (για μια πλήρη λίστα ανατρέξτε στις δύο πρώτες σειρές του Πίνακα. 5.1):

- LeNet [14]
- AlexNet [1]
- VGGnet [15]
- ResNet [16]

- DenseNet [17]

Τέλος ορίσαμε τους συνδυασμούς των  $m$  και  $b_d$  ως μέλη  $c$  του συνόλου συνδυασμένων μοντέλων  $C$ . Χρησιμοποιώντας τους προηγούμενους ορισμούς, ο στόχος αυτού του κεφαλαίου είναι η αξιολόγηση του συνόλου των μοντέλων  $C$ , όπου  $C$  είναι το συνδυασμένο σύνολο των  $M$  και  $B$  δλδ.  $C = M \times B$  σε σχέση με την χρονική απόδοση και την ακρίβεια εκπαιδευμένα στο  $D$ .

### 5.3.2. Signal2Image

Σε αυτή την ενότητα περιγράφονται τα εσωτερικά στοιχεία κάθε μονάδας S2I. Για το S2I ‘σήμα ως εικόνα’ κανονικοποιήσαμε το πλάτος  $x_i$  στο εύρος  $[1, 178]$ . Τα αποτελέσματα αντιστράφηκαν κατά μήκος του άξονα  $y$ , στρογγυλοποιημένα στον πλησιέστερο ακέραιο και στη συνέχεια χρησιμοποιήθηκαν ως  $y$ -δείκτες για τα εικονοστοιχεία με πλάτος 255 σε μια εικόνα  $178 \times 178$  αρχικοποιημένη με μηδενικά.

Για το S2I φασματογράφου, το οποίο χρησιμοποιείται για την απεικόνιση της μεταβολής της συχνότητας ενός μη-στάσιμου σήματος με την πάροδο του χρόνου [18], χρησιμοποιήθηκε παράθυρο Tukey με παράμετρο σχήματος 0.25, μήκος τμήματος 8 δειγμάτων, αλληλεπικάλυψη μεταξύ τμημάτων 4 δειγμάτων και ταχύ Μετασχηματισμό Fourier των 64 δειγμάτων για την μετατροπή του  $x_i$  στο πεδίο χρόνου-συχνότητας. Το προκύπτον φασματογράφημα, το οποίο αντιπροσωπεύει το μέγεθος της φασματικής πυκνότητας ισχύος ( $V^2/Hz$ ) του  $x_i$ , στη συνέχεια υπερδειγματολήφθηκε σε  $178 \times 178$  με χρήση διγραμμικής παρεμβολής μεταξύ των εικονοστοιχείων.

Για τα S2I CNN με ένα και δύο επίπεδα, το  $x_i$  μετατρέπεται σε εικόνα χρησιμοποιώντας εκπαιδευσιμες παραμέτρους, αντί χρήσης κάποιας στατικής διαδικασίας. Το S2I ενός επιπέδου αποτελείται από ένα 1D συνελικτικό επίπεδο (μεγέθη πυρήνα 3 με 8 κανάλια). Το S2I δύο επιπέδων αποτελείται από δύο 1D συνελικτικά επίπεδα (μεγέθη πυρήνα 3 με 8 και 16 κανάλια) με το πρώτο επίπεδο να ακολουθείται από μια συνάρτηση ενεργοποίησης ReLU και ένα επίπεδο 1D max-pooling (μέγεθος πυρήνα 2). Οι χάρτες ενεργοποιήσεων του τελευταίου συνελικτικού επιπέδου και για τα δύο S2I συμπύσσονται σειριακά κατά μήκος του  $y$ -άξονα και στη συνέχεια μετατρέπονται σε μέγεθος  $178 \times 178$  χρησιμοποιώντας τη διγραμμική παρεμβολή.

Περιορίζουμε την έξοδο για όλα τα  $m$  σε μια εικόνα μεγέθους  $178 \times 178$  για να επιτρέψουμε οπτική σύγκριση. Τρία πανομοιότυπα κανάλια στοιβάζονται επίσης για όλες τις εξόδους των  $m$  για να ικανοποιήσουν τις απαιτήσεις μεγέθους εισόδου των  $b_d$ . Οι αρχιτεκτονικές όλων των  $b_d$  παραμένουν οι ίδιες, εκτός του αριθμού των εξόδων του τελευταίου γραμμικού επιπέδου που έχει οριστεί σε πέντε για να αντιστοιχεί στον αριθμό των κατηγοριών του  $D$ . Ένα παράδειγμα των αντίστοιχων εξόδων κάποιων από τα  $m$  (το CNN ενός/δύο επιπέδων παρήγαγαν παρόμοιες απεικονίσεις) απεικονίζονται

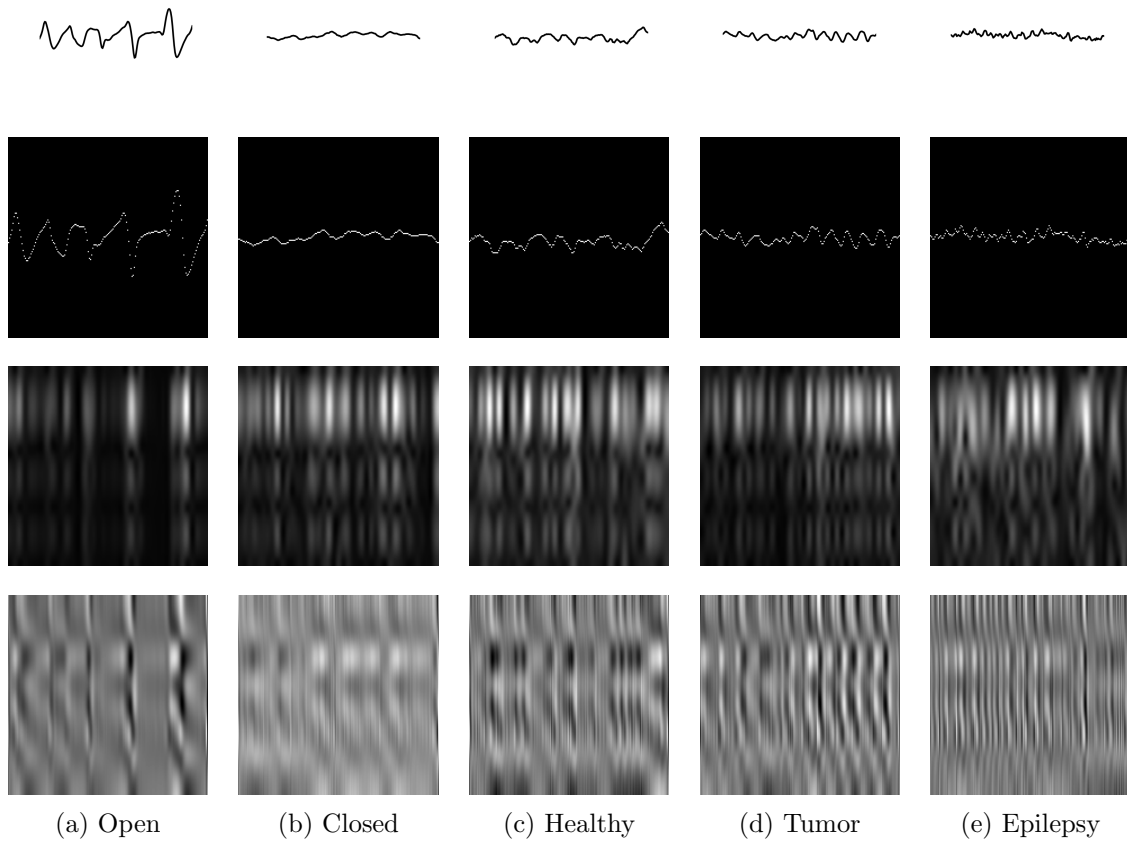


Figure 5.2: Σήματα και έξοδοι των Signal2Image για κάθε κλάση. Οι  $x$ ,  $y$ -άξονες της πρώτης σειράς είναι  $\mu\text{V}$  και χρονικά δείγματα αντίστοιχα. Οι  $x$ ,  $y$ -άξονες των υπόλοιπων εικόνων απεικονίζουν χωρική πληροφορία, επειδή δεν ενημερώνουμε το ‘μοντέλο βάσης’ με την έννοια του χρόνου κατά τον  $x$ -άξονα ή την έννοια της συχνότητας κατά τον  $y$ -άξονα. Υψηλότερη ένταση των εικονοστοιχείων υποδηλώνει υψηλότερο πλάτος.

στη δεύτερη, τρίτη και τέταρτη σειρά του σχήματος 5.2.

## 5.4. Ρύθμιση πειραμάτων

Τα συνελικτικά επίπεδα αρχικοποιήθηκαν με χρήση ομοιόμορφου θορύβου τύπου Kaiming [19]. Οι τιμές λαμβάνονται από την ομοιόμορφη κατανομή  $\mathcal{U}(-c, c)$ , όπου  $c$  είναι:

$$c = \sqrt{\frac{6}{(1 + a^2)k}} \quad (5.1)$$

, $a$  σε αυτήν τη μελέτη τίθεται μηδέν και  $k$  είναι το μέγεθος της εισόδου του επιπέδου. Τα γραμμικά επίπεδα του  $m$  αρχικοποιήθηκαν χρησιμοποιώντας  $\mathcal{U}(-\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}})$ . Τα συνε-

λικτικά και γραμμικά επίπεδα όλων των  $b_a$  αρχικοποιήθηκαν σύμφωνα με την αρχική τους εφαρμογή.

Χρησιμοποιήθηκε ο Adam [20] ως βελτιστοποιητής με ρυθμό μάθησης  $lr = 0.001$ , betas  $b_1 = 0.9$ ,  $b_2 = 0.999$ , epsilon  $\epsilon = 10^{-8}$  χωρίς απώλεια βαρών και τη διασταυρούμενη εντροπία ως συνάρτηση απώλειας. Το μέγεθος παρτίδας ήταν 20 και δεν χρησιμοποιήθηκε επιπλέον regularization εκτός από το προϋπάρχον όπως το dropout, το οποίο έχουν κάποια από τα ‘μοντέλα βάσης’ (AlexNet, VGGnet και DenseNet).

Από τα 11500 σήματα που χρησιμοποιήθηκαν 76% και 12% από τα δεδομένα (8740, 1380, 1380 σήματα) ως δεδομένα εκπαίδευσης, επικύρωσης και δοκιμής αντίστοιχα. Δεν πραγματοποιήθηκε αφαίρεση θορύβου ή άλλη προεπεξεργασία. Όλα τα δίκτυα εκπαιδεύτηκαν για 100 εποχές και η επιλογή των μοντέλων έγινε σύμφωνα με την καλύτερη ακρίβεια επικύρωσης από όλες τις εποχές. Χρησιμοποιήθηκε το PyTorch [21] για την υλοποίηση των αρχιτεκτονικών νευρωνικών δικτύων και η εκπαίδευση/προεπεξεργασία πραγματοποιήθηκε με τη χρήση της NVIDIA Titan X Pascal GPU 12GB RAM και 12 Core Intel i7-8700 CPU @ 3.20GHz σε λειτουργικό σύστημα Linux.

## 5.5. Αποτελέσματα

Όπως φαίνεται στον Πίνακα. 5.1 το CNN ενός επιπέδου DenseNet201 πέτυχε την καλύτερη ακρίβεια 85.3%, με διάρκεια εκπαίδευσης 70 δευτερόλεπτα/εποχή κατά μέσο όρο. Συνολικά, το CNN S2I ενός επιπέδου πέτυχε την καλύτερη ακρίβεια για έντεκα από τα δεκαπέντε ‘μοντέλα βάσης’. Το CNN S2I δύο επιπέδων παρουσίασε χειρότερα αποτελέσματα ακόμη και σε σύγκριση με τις 1D παραλλαγές, υποδεικνύοντας ότι η αύξηση του βάθους S2I δεν είναι ευεργετική. Το ‘σήμα ως εικόνα’ και το φασματογράφημα S2Is παρουσίασαν πολύ χειρότερα αποτελέσματα από τις 1D παραλλαγές και το CNN S2Is. Τα αποτελέσματα S2I του φασματογράφου είναι αντίθετα με την προσδοκία ότι η ερμηνεύσιμη αναπαράσταση του πεδίου χρόνου-συχνοτήτων θα βοηθούσε στην εύρεση καλών χαρακτηριστικών για ταξινόμηση. Υποθέτουμε ότι το φασματογράφημα S2I παρεμποδίστηκε από την έλλειψη μη-εκπαιδευσιμων παραμέτρων. Ένα άλλο αποτέλεσμα αυτών των πειραμάτων είναι ότι η αύξηση του βάθους των ‘μοντέλων βάσης’ δεν αύξησε την ακρίβεια, κάτι το οποίο είναι σύμφωνο με προηγούμενα αποτελέσματα [22].

## 5.6. Συμπεράσματα

Σε αυτό το κεφάλαιο δείξαμε εμπειρικά αποτελέσματα ότι οι παραλλαγές 1D ‘μοντέλου βάσης’ και τα εκπαιδευσιμα S2Is (ειδικά το CNN ενός επιπέδου) έχουν καλύτερη

Table 5.1: Ακρίβειες στα δεδομένα δοκιμής (%) για τα συνδυασμένα μοντέλα. Η δεύτερη σειρά απεικονίζει τον αριθμό των επιπέδων. Η έκτη γραμματοσειρά απεικονίζει τις καλύτερες ακρίβειες για κάθε μοντέλο βάσης.

Model \ Dim, S2I	LeNet		AlexNet		VGGnet					ResNet					DenseNet				
	2		5		11	13	16	19	18	34	50	101	152	121	161	169	201		
1D, none	72.6		78.8		76.9	<b>79.0</b>	79.5	<b>79.3</b>	81.5	82.5	81.4	78.8	81.4	81.8	<b>83.3</b>	82.1	82.0		
2D, signal as image	67.9		68.3		74.1	74.7	72.7	72.5	73.3	71.7	74.1	72.3	74.1	74.7	72.5	75.2	75.0		
2D, spectrogram	73.2		74.0		77.9	76.3	77.5	76.0	76.2	79.0	77.2	74.6	75.3	74.1	75.2	77.0	75.4		
2D, one layer CNN	<b>75.8</b>		<b>82.0</b>		<b>84.0</b>	77.9	80.7	78.4	<b>85.1</b>	<b>84.6</b>	<b>83.0</b>	<b>85.0</b>	<b>83.3</b>	<b>84.3</b>	80.7	<b>85.0</b>	<b>85.3</b>		
2D, two layer CNN	75.0		77.9		80.7	78.8	<b>81.1</b>	74.9	78.3	80.0	78.3	77.1	80.9	83.2	82.3	79.0	79.1		

απόδοση από τα μη-εκπαιδευσιμα S2I. Ωστόσο, πρέπει να καταβληθεί περισσότερη προσπάθεια για την πλήρη αντικατάσταση των μη-εκπαιδευσιμων S2I, όχι μόνο από την πλευρά επίτευξης υψηλότερων αποτελεσμάτων ακρίβειας, αλλά και για την αύξηση της ερμηνευσιμότητας του μοντέλου. Ένα άλλο σημείο αναφοράς είναι ότι τα συνδυασμένα μοντέλα εκπαιδύτηκαν από την αρχή, με βάση την υπόθεση ότι τα προρυθμισμένα χαρακτηριστικά χαμηλού επιπέδου των 'μοντέλων βάσης' μπορεί να μην είναι κατάλληλα για εικόνες που μοιάζουν με φασματογράφημα όπως εκείνες που δημιουργούνται από τα S2Is. Μια μελλοντική εργασία θα μπορούσε να περιλαμβάνει τη δοκιμή αυτής της υπόθεσης με την αρχικοποίηση ενός 'μοντέλου βάσης' με τη χρήση μεθόδων μεταφοράς μάθησης ή άλλων μεθόδων αρχικοποίησης. Επιπλέον, μπορούν να χρησιμοποιηθούν και άλλα εκπαιδευσιμα S2Is και 1D 'μοντέλου βάσης' για άλλα σήματα φυσιολογίας εκτός από το EEG, όπως το ECG, το ηλεκτρομυογράφημα και η γαλβανική απόκριση δέρματος.

## Βιβλιογραφία

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [2] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986.
- [3] Khald Aboalayon, Miad Faezipour, Wafaa Almuhammadi, and Saeid Mosleh-pour. Sleep stage classification using eeg signal analysis: a comprehensive survey and new investigation. *Entropy*, 18(9):272, 2016.
- [4] Abeer Al-Nafjan, Manar Hosny, Yousef Al-Ohali, and Areej Al-Wabil. Review and classification of emotion recognition based on eeg brain-computer interface system research: a systematic review. *Applied Sciences*, 7(12):1239, 2017.
- [5] Fabien Lotte, Laurent Bougrain, and Maureen Clerc. Electroencephalography (eeg)-based brain-computer interfaces. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–20, 1999.
- [6] U Rajendra Acharya, S Vinitha Sree, G Swapna, Roshan Joy Martis, and Jasjit S Suri. Automated eeg analysis of epilepsy: a review. *Knowledge-Based Systems*, 45:147–165, 2013.
- [7] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 2019.
- [8] Martin Längkvist, Lars Karlsson, and Amy Loutfi. Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems*, 2012:5, 2012.
- [9] Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Samuel Ippolito, and Omid Kavehei. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, 2018.
- [10] Haidar Khan, Lara Marcuse, Madeline Fields, Kalina Swann, and Bülent Yener. Focal onset seizure prediction using convolutional networks. *IEEE Transactions on Biomedical Engineering*, 65(9):2109–2118, 2018.



- 
- [11] Jianhua Zhang, Sunan Li, and Rubin Wang. Pattern recognition of momentary mental workload based on multi-channel electrophysiological data and ensemble convolutional neural networks. *Frontiers in neuroscience*, 11:310, 2017.
- [12] Endang Purnama Giri, Mohamad Ivan Fanany, Aniati Murni Arymurthy, and Sastra Kusuma Wijaya. Ischemic stroke identification based on eeg and eog using id convolutional neural network and batch normalization. In *Advanced Computer Science and Information Systems (ICACSIS), 2016 International Conference on*, pages 484–491. IEEE, 2016.
- [13] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, 2001.
- [14] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [18] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [21] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

- 
- [22] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggersperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

## 6. Δίκτυα αραιής ενεργοποίησης

### 6.1. Εισαγωγή

Τα DNN [1] χρησιμοποιούν πολλαπλά στοιβαγμένα επίπεδα με βάρη και συναρτήσεις ενεργοποίησης, που μετασχηματίζουν την είσοδο σε ενδιάμεσες αναπαραστάσεις κατά τη διάρκεια της προς-τα-εμπρός διάδοσης. Χρησιμοποιώντας backpropagation [2] η κλίση του κάθε βάρους σε σχέση με το σφάλμα της εξόδου υπολογίζεται και μεταβιβάζεται σε μια συνάρτηση βελτιστοποίησης όπως η Στοχαστική Κάθοδος Κλίσεων ή Adam [3] η οποία ενημερώνει τις τιμές των βαρών κάνοντας την έξοδο του δικτύου να συγκλίνει στην επιθυμητή έξοδο. Τα DNN χρησιμοποιώντας μεγάλο όγκο δεδομένων και ισχυρές μονάδες παράλληλης επεξεργασίας έχουν επιτύχει υψηλού επιπέδου αποτελέσματα σε προβλήματα όπως η αναγνώριση εικόνας [4] και ομιλίας [5]. Ωστόσο, αυτές οι καινοτομίες έχουν συμβεί σε βάρος της αύξησης του μήκους περιγραφής των αναπαραστάσεων, οι οποίες στα DNN είναι ανάλογες με τον αριθμό των:

1. βαρών του μοντέλου και
2. μη-μηδενικών ενεργοποιήσεων.

Η χρήση μεγάλου αριθμού βαρών ως επιλογή σχεδιασμού σε αρχιτεκτονικές όπως το Inception [6], VGGnet [7] και ResNet [8] (συνήθως με την αύξηση του βάθους), ακολουθήθηκε από έρευνα που επέδειξε τον πλεονασμό των βαρών των DNNs. Αποδείχθηκε ότι τα DNN προσαρμόζονται εύκολα σε τυχαίες επισημάνσεις δεδομένων [9] και ότι σε κάθε αρχικοποιημένο DNN υπάρχει ένα υποδίκτυο που μπορεί να επιλύσει το δεδομένο πρόβλημα, με την ίδια ακρίβεια με το αρχικά εκπαιδευμένο [10].

Επιπλέον τα DNN με μεγάλο αριθμό βαρών έχουν υψηλότερες απαιτήσεις αποθήκευσης και είναι πιο αργά κατά τη διάρκεια των συμπερασμών. Προηγούμενη έρευνα που έγινε πάνω σε αυτό το πρόβλημα, επικεντρώθηκε στο κλάδεμα βαρών από εκπαιδευμένα DNNs [11] και κλάδεμα βαρών κατά τη διάρκεια της εκπαίδευσης [12]. Το κλάδεμα ελαχιστοποιεί την χωρητικότητα του μοντέλου για χρήση σε περιβάλλοντα με χαμηλές υπολογιστικές δυνατότητες ή χαμηλές απαιτήσεις χρόνου συμπερασμών

και επίσης βοηθά στην μείωση της προσαρμογής των νευρώνων η οποία επίσης έχει αντιμετωπιστεί από το Dropout [13]. Ωστόσο οι στρατηγικές κλαδέματος λαμβάνουν υπόψη μόνο τον αριθμό των βαρών του μοντέλου.

Το άλλο στοιχείο που επηρεάζει το μήκος περιγραφής των αναπαραστάσεων των DNN, είναι ο αριθμός των μη-μηδενικών ενεργοποιήσεων στις ενδιάμεσες αναπαραστάσεις που σχετίζονται με την έννοια της αραιότητας. Στα νευρωνικά δίκτυα η αραιότητα μπορεί να εφαρμοστεί είτε στις συνδέσεις μεταξύ νευρώνων είτε στους χάρτες ενεργοποίησης [14]. Παρόλο που η αραιότητα στους χάρτες ενεργοποίησης συνήθως επιβάλλεται στη συνάρτηση απώλειας με την προσθήκη ενός όρου κανονικοποίησης  $L_{1,2}$  ή απόκλισης Kullback-Leibler [15], θα μπορούσαμε επίσης να επιτύχουμε αραιότητα στους χάρτες ενεργοποίησης με τη χρήση μιας κατάλληλης συνάρτησης ενεργοποίησης.

Αρχικά, χρησιμοποιήθηκαν φραγμένες συναρτήσεις όπως η σιγμοειδής και η υπερβολική εφαπτομένη (tanh), αλλά εκτός από την παραγωγή πυκνών χαρτών ενεργοποίησης παρουσιάζουν επίσης το πρόβλημα του vanishing gradients [16]. Το ReLU προτάθηκε αργότερα [17, 18] ως μια συνάρτηση ενεργοποίησης που λύνει το πρόβλημα της διαφυγής των κλίσεων και αυξάνει την αραιότητα των χαρτών ενεργοποίησης. Αν και το ReLU δημιουργεί μηδενικά (σε αντίθεση με τους προκατόχους του σιγμοειδής και tanh), ο χάρτης ενεργοποίησής του αποτελείται από αραιά χωρισμένες αλλά πυκνές περιοχές (Εικ. 6.1(b)) αντί για αραιές αιχμές. Το ίδιο ισχύει για άλλες γενικεύσεις του ReLU, όπως το Παραμετρικό ReLU [19] και το Maxout [20]. Πρόσφατα, στο  $k$ -Sparse Autoencoders [21] χρησιμοποιήθηκε μια συνάρτηση ενεργοποίησης που εφαρμόζει κατωφλίωση μέχρι να παραμείνουν οι  $k$  μεγαλύτερες ενεργοποιήσεις, ωστόσο αυτή η μη-γραμμικότητα καλύπτει μια περιορισμένη περιοχή του χάρτη ενεργοποίησης δημιουργώντας αραιά αποσυνδεδεμένες πυκνές περιοχές (Εικ. 6.1(c)), παρόμοιες με την περίπτωση ReLU.

Επιπλέον συναρτήσεις ενεργοποίησης που παράγουν συνεχείς χάρτες ενεργοποίησης (όπως το ReLU) είναι λιγότερο βιολογικά εύλογες, επειδή οι βιολογικοί νευρώνες σπάνια βρίσκονται στο μέγιστο επίπεδο κορεσμού τους [22] και επίσης χρησιμοποιούν αιχμές για να επικοινωνούν αντί για συνεχείς τιμές [23]. Προηγούμενη βιβλιογραφία έχει επίσης καταδείξει την αυξημένη βιολογική ευλογοφάνεια της αραιότητας στα τεχνητά νευρωνικά δίκτυα [24]. Η αραιότητα επιπέδου αιχμών στους χάρτες ενεργοποίησης έχει ερευνηθεί διεξοδικά στα βιολογικώς πιο αποδεκτά μοντέλα που επικοινωνούν με βάση το ρυθμό (rate-based) [25], αλλά δεν έχει διερευνηθεί διεξοδικά ως επιλογή σχεδιασμού για συναρτήσεις ενεργοποίησης σε συνδυασμό με συνελικτικά φίλτρα.

Ο αυξημένος αριθμός βαρών και των μη-μηδενικών ενεργοποιήσεων καθιστούν τα DNNs πολύπλοκα και έτσι είναι δυσκολότερο να χρησιμοποιηθούν σε προβλήματα που απαιτούν την αντιστοίχιση αιτιότητας της εξόδου με ένα συγκεκριμένο σύνολο

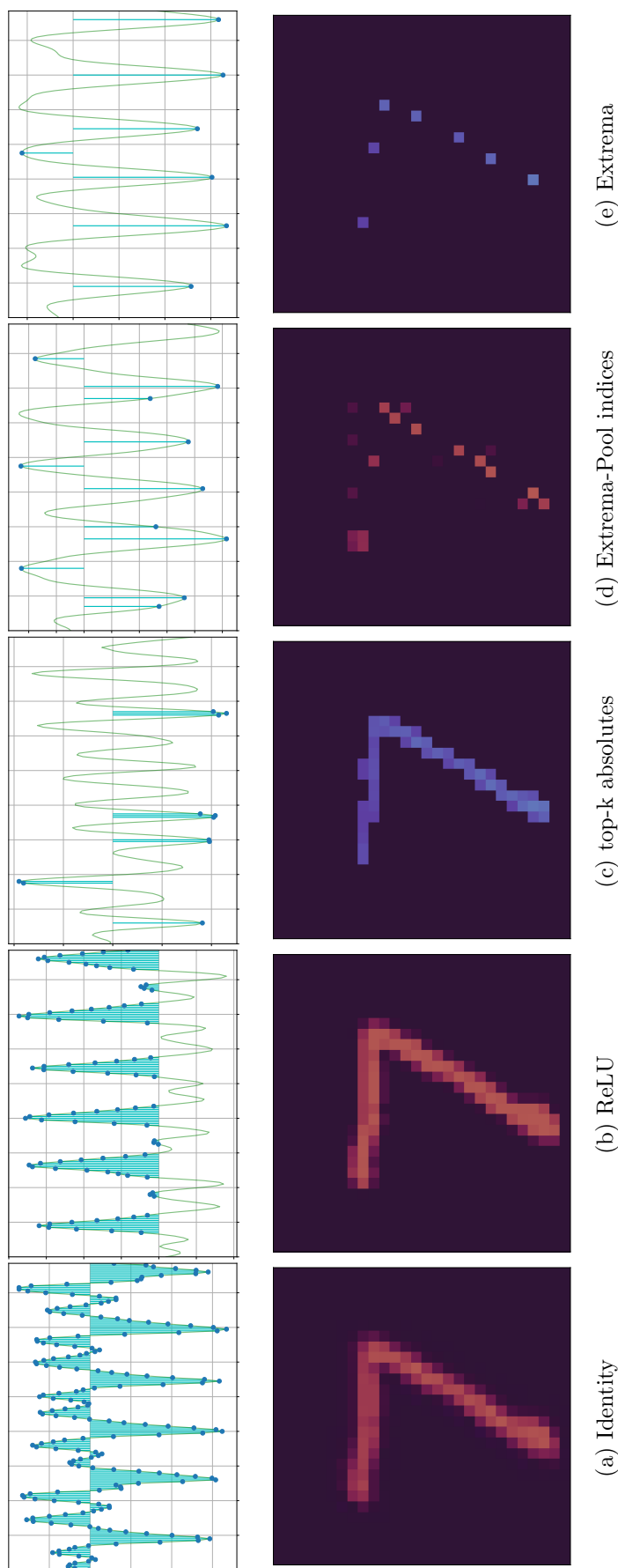


Figure 6.1: Οπτικοποίηση πέντε συναρτήσεων αραιής ενεργοποίησης (Identity, Max-Activations, Max-Pool indices και Extrema) για 1D και 2D είσοδο στην πρώτη και τη δεύτερη σειρά αντίστοιχα. Η 1D είσοδος στις συναρτήσεις ενεργοποίησης υποδηλώνεται με τη συνεχή διαφανή γραμμή, χρησιμοποιώντας ένα παράδειγμα από τη βάση δεδομένων επιληψίας του UCI. Η έξοδος κάθε συνάρτησης ενεργοποίησης απεικονίζεται με τις κυανές γραμμές με τους μπλε δείκτες. Το 2D παράδειγμα απεικονίζει μόνο την έξοδο των συναρτήσεων ενεργοποίησης, χρησιμοποιώντας ένα παράδειγμα από τη βάση δεδομένων MNIST.

νευρώνων. Η πλειοψηφία των τομέων που κάνουν χρήση της μηχανικής μάθησης συμπεριλαμβανομένων τομέων όπως η υγειονομική περίθαλψη [26] απαιτούν τα μοντέλα να είναι ερμηνεύσιμα και εξηγήσιμα πριν θεωρηθούν ως πιθανή λύση. Αν και αυτές οι ιδιότητες μπορούν να αυξηθούν χρησιμοποιώντας ανάλυση ευαισθησίας [27], μεθόδους αποσυνέλιξης [28], Layerwise-Relevance Propagation [29] και Local-Interpretable Model agnostic Explanations [30] θα ήταν προτιμότερο να έχουμε αυτο-ερμηνεύσιμα μοντέλα.

Επιπλέον λαμβάνοντας υπόψη ότι τα DNNs μαθαίνουν να αναπαριστούν δεδομένα χρησιμοποιώντας το συνδυαστικό σύνολο των βαρών και των μη-μηδενικών ενεργοποιήσεων κατά τη διάρκεια της προς-τα-εμπρός διάδοσης, προκύπτει ένα ενδιαφέρον ερώτημα:

*Ποιες είναι οι επιπτώσεις του συμβιβασμού μεγαλύτερου σφάλματος ανακατασκευής των αναπαραστάσεων και του λόγου συμπίεσης των αναπαραστάσεων σε σχέση με στα αρχικά δεδομένα;*

Προηγούμενη εργασία από τους Blier et al. [31] έδειξε την ικανότητα των DNN να συμπιέζουν χωρίς απώλειες τα δεδομένα εισόδου και τα βάρη, αλλά χωρίς να παίρνουν υπόψη τον αριθμό των μη-μηδενικών ενεργοποιήσεων. Σε αυτό το έργο, χαλαρώνουμε την απαίτηση της μη-απώλειας και επίσης θεωρούμε τα νευρωνικά δίκτυα καθαρά ως προσεγγιστές συναρτήσεων, αντί για πιθανοτικά μοντέλα. Οι συνεισφορές του παρόντος είναι οι ακόλουθες προτάσεις:

- Το μέτρο  $\varphi$  που αξιολογεί μη-επιβλεπόμενα μοντέλα με βάση το πόσο συμπιεσμένες είναι οι αναπαραστάσεις σε σχέση με τα αρχικά δεδομένα και κατά πόσο ακριβή είναι στις ανακατασκευές τους.
- Δίκτυα Αραιής Ενεργοποίησης (SANs) (Εικ. 6.2), στα οποία επιβάλλεται αραιότητα αιχμών στον χάρτη ενεργοποίησης μέσω μιας συνάρτησης αραιής ενεργοποίησης (Εικ. 6.1(d) και (e)).

## 6.2. Μέτρο $\varphi$

Έστω  $M$  ένα μοντέλο με  $q$  πυρήνες κάθε ένας από τους οποίους με  $m^{(i)}$  δείγματα και έστω  $\mathcal{L}$  η συνάρτηση απώλειας ανακατασκευής:

$$M : \mathbf{x} \mapsto \hat{\mathbf{x}} \quad (6.1)$$

, όπου  $\mathbf{x} \in \mathbb{R}^n$  είναι το διάνυσμα εισόδου και  $\hat{\mathbf{x}}$  είναι η ανακατασκευή του  $\mathbf{x}$ . Για τον ορισμό του μέτρου  $\varphi$  χρησιμοποιούμε ένα νευρωνικό δίκτυο το οποίο περιέχει

συνελικτικά φίλτρα, παρ'όλ'αυτά ο συγκεκριμένος ορισμός μπορεί να γενικευτεί και σε άλλες αρχιτεκτονικές. Το μέτρο  $\varphi$  αξιολογεί ένα μοντέλο με βάση δύο έννοιες: την 'φλυαρία' (verbosity) και την ακρίβειά του.

Η 'φλυαρία' στα νευρωνικά δίκτυα μπορεί να γίνει αντιληπτή ως αντιστρόφως ανάλογη με το λόγο συμπίεσης των αναπαραστάσεων. Πρώτον, υπολογίζουμε τον αριθμό των βαρών  $W$  ενός μοντέλου  $M$  ως εξής:

$$W = \sum_{i=1}^q m^{(i)} \quad (6.2)$$

Επίσης υπολογίζουμε τον αριθμό των μη-μηδενικών ενεργοποιήσεων  $A$  ενός μοντέλου  $M$  για είσοδο  $\mathbf{x}$  ως:

$$A_{\mathbf{x}} = \sum_{i=1}^q \|\boldsymbol{\alpha}^{(i)}\|_0 \quad (6.3)$$

, όπου  $\|\cdot\|_0$  δηλώνει την ψευδο-νόρμα  $\ell_0$  και  $\boldsymbol{\alpha}^{(i)}$  τον χάρτη ενεργοποίησης του  $i^{\text{th}}$  πυρήνα. Στη συνέχεια, χρησιμοποιώντας τις Εξισώσεις 6.2 και 6.3 ορίζουμε τη σχέση συμπίεσης  $CR$  του  $\mathbf{x}$  σε σχέση με  $M$  ως:

$$CR = \frac{n}{W + (\dim(\mathbf{x}) + 1)A_{\mathbf{x}}} \quad (6.4)$$

, όπου  $\dim$  ορίζει την διαστασιμότητα. Ο λόγος που πολλαπλασιάζουμε τη διαστασιμότητα του  $\mathbf{x}$  με τον αριθμό των ενεργοποιήσεων  $A_{\mathbf{x}}$ , είναι ότι πρέπει να εξετάσουμε τη χωρική θέση κάθε μη-μηδενικής ενεργοποίησης εκτός από το πλάτος του για την ανακατασκευή  $\mathbf{x}$ . Επιπλέον, χρησιμοποιώντας αυτόν τον ορισμό του  $CR$  δημιουργείται ένας επιθυμητός συμβιβασμός μεταξύ της χρήσης ενός μεγαλύτερου πυρήνα με λιγότερες εμφανίσεις και ενός μικρότερου πυρήνα με περισσότερες εμφανίσεις, βάσει του οποίου αποφασίζεται το μέγεθος του πυρήνα που ελαχιστοποιεί το  $CR$ .

Όσον αφορά την ακρίβεια ορίζουμε την κανονικοποιημένη απώλεια ανακατασκευής ως εξής:

$$\tilde{\mathcal{L}}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{\mathcal{L}(\hat{\mathbf{x}}, \mathbf{x})}{\mathcal{L}(0, \mathbf{x})} \quad (6.5)$$

Τέλος, χρησιμοποιώντας τις Εξισώσεις 6.4 και 6.5 ορίζουμε το μέτρο  $\varphi^1$  του  $\mathbf{x}$  σε σχέση με  $M$  ως εξής:

$$\varphi = \|(CR^{-1}, \tilde{\mathcal{L}}(\hat{\mathbf{x}}, \mathbf{x}))\|_2 \quad (6.6)$$

, όπου  $\|\cdot\|_2$  ορίζει την ευκλείδεια απόσταση.

<sup>1</sup>Η χρήση του συμβόλου  $\varphi$  προέρχεται από τον πρώτο χαρακτήρα του σύνθετου ελληνικού ουσιαστικού 'φλύθος' = φλύ + θος. Αποτελείται από το πρώτο μέρος της λέξης φλύ-αρος και το δεύτερο μέρος της λέξης λά-θος. Το φλύθος ορίζεται κυριολεκτικά ως: *Παροχή ανακριβών πληροφοριών χρησιμοποιώντας πολλές λέξεις. Η κατάσταση του να είναι κανείς λάθος και φλύαρος ταυτόχρονα.*

Όσον αφορά την επιλογή υπερπαραμέτρων, ορίζουμε επίσης το μέτρο  $\bar{\varphi}$  ενός συνόλου δεδομένων ή μια μίνι-παρτίδας σε σχέση με το  $M$  ως:

$$\bar{\varphi} = \frac{1}{l} \sum_{j=1}^l \varphi^{(j)} \quad (6.7)$$

, όπου  $l$  είναι ο αριθμός παρατηρήσεων της βάσης δεδομένων ή το μέγεθος της παρτίδας.

Το  $\bar{\varphi}$  είναι μη διαφοροποιήσιμο λόγω της παρουσίας της ψευδο-νόρμας  $\ell_0$  στην Εξ. 6.3. Ένας τρόπος για να ξεπεραστεί αυτό το πρόβλημα, είναι η χρήση του  $\mathcal{L}$  ως τη διαφοροποιήσιμη συνάρτηση βελτιστοποίησης κατά τη διάρκεια της εκπαίδευσης και η χρήση του  $\bar{\varphi}$  ως το μέτρο για την επιλογή μοντέλου κατά την επικύρωση, κατά την διάρκεια της οποίας παίρνονται οι αποφάσεις για τις επιλογές των τιμών των υπερπαραμέτρων (όπως το μέγεθος του πυρήνα).

## 6.3. Δίκτυα αραιής ενεργοποίησης

### 6.3.1. Συναρτήσεις αραιής ενεργοποίησης

Σε αυτή την υποενότητα ορίζουμε πέντε συναρτήσεις ενεργοποίησης  $\phi$  και την αντίστοιχη παράμετρο αραιότητας πυκνότητας  $d^{(i)}$  για τα οποία έχουμε:

$$\phi : s \mapsto a \quad (6.8)$$

Επιλέγουμε τιμές για το  $d^{(i)}$  για κάθε συνάρτηση αραιής ενεργοποίησης έτσι ώστε να έχουμε περίπου τον ίδιο αριθμό ενεργοποιήσεων για να έχουμε δίκαιη σύγκριση μεταξύ των συναρτήσεων αραιής ενεργοποίησης (εκτός από την Identity, η οποία δεν έχει παράμετρο αραιότητας).

#### Ταυτότητα

$\phi = \mathbb{1}$ . Η συνάρτηση ενεργοποίησης ταυτότητας (Identity) χρησιμεύει ως βάση αναφοράς και διατηρεί ακέραια την είσοδό του, όπως φαίνεται στην Εικ. 6.1(a). Για αυτήν την περίπτωση, δεν εφαρμόζεται παράμετρος αραιότητας  $d^{(i)}$ .

#### ReLU

$\phi = \text{ReLU}(s)$ . Η συνάρτηση ενεργοποίησης ReLU παράγει αραιά συνδεδεμένες αλλά πυκνές περιοχές, όπως φαίνεται στην Εικ. 6.1(b). Για αυτήν την περίπτωση, δεν εφαρμόζεται παράμετρος αραιότητας  $d^{(i)}$ .



**κ-μέγιστα απολύτων**

Η συνάρτηση κ-μεγίστων απολύτων (top-k absolutes) (που ορίζεται στον αλγόριθμο 1) διατηρεί τους δείκτες των  $k$  ενεργοποιήσεων με τη μεγαλύτερη απόλυτη τιμή και μηδενίζει τις υπόλοιπες, όπου  $1 \leq k < n \in \mathbb{N}$ . Ορίσαμε  $d^{(i)} = k$ , όπου  $k = \lfloor n/m \rfloor^{\dim(\mathbf{x})}$ . Η συνάρτηση ενεργοποίησης απόλυτων κ-μεγίστων είναι πιο αραιή από το ReLU αλλά κάποια ακρότατα ενεργοποιούνται πολλαπλά σε σχέση με άλλα ακρότατα που δεν ενεργοποιούνται καθόλου, όπως φαίνεται στην Εικ. 6.1(c).

**Algorithm 1** top-k absolutes**Input:**  $s, k$ **Output:**  $\alpha$ 

- 1:  $\alpha_i \leftarrow 0, i = 1 \dots \text{card}(s)$
- 2:  $p \leftarrow \text{topk}(|s|, k)$
- 3: **for**  $i = 0$  to  $\text{card}(s)$  **do**
- 4:    $\alpha_i(p_i) = s_i(p_i)$
- 5: **end for**
- 6: **return**  $\alpha$

**Δείκτες Συγκέντρωσης Ακρότατων**

Η συνάρτηση ενεργοποίησης των δεικτών συγκέντρωσης ακρότατων (Extrema-Pool indices) (που ορίζεται στον Αλγόριθμο 2) διατηρεί μόνο τον δείκτη της ενεργοποίησης με το μέγιστο απόλυτο πλάτος από κάθε περιοχή που περιγράφεται από ένα πλέγμα με την ίδια αναλυτικότητα όπως το μέγεθος πυρήνα  $m^{(i)}$  και μηδενίζει τις υπόλοιπες. Αποτελείται από ένα στρώμα μέγιστης συγκέντρωσης ακολουθούμενο από ένα στρώμα μέγιστης αποσυγκέντρωσης με τις ίδιες παραμέτρους, ενώ η παράμετρος αραιότητας  $d^{(i)}$  σε αυτή την περίπτωση ορίζεται ως  $d^{(i)} = m^{(i)} < n \in \mathbb{N}$ . Αυτή η συνάρτηση ενεργοποίησης δημιουργεί πιο αραιούς χάρτες ενεργοποίησης από τα απόλυτα κ-μέγιστα, παρ'όλ'αυτά σε περιπτώσεις που το πλέγμα της συγκέντρωσης είναι κοντά σε ακρότατο υπάρχει το ενδεχόμενο να ενεργοποιηθεί διπλά (όπως φαίνεται στην Εικ. 6.1(d)).

**Algorithm 2** Δείκτες συγκέντρωσης ακρότατων**Input:**  $s, m$ **Output:**  $\alpha$ 

- 1:  $p \leftarrow \text{maxpool}(|s|, m)$
- 2:  $\alpha \leftarrow \text{maxunpool}(s(p), p, m)$
- 3: **return**  $\alpha$

## Ακρότατα

Η συνάρτηση ενεργοποίησης ακρότατων (Extrema) (που ορίζεται στον Αλγόριθμο 3) ανιχνεύει τα υποψήφια ακρότατα χρησιμοποιώντας τη μηδενική διέλευση της πρώτης παραγώγου και στη συνέχεια τις ταξινομεί σε φθίνουσα σειρά και σταδιακά εξαλείφει εκείνες τα ακρότατα που έχουν μικρότερο πλάτος από ένα γειτονικό ακρότατο σύμφωνα με μια προκαθορισμένη απόσταση  $med$ . Η επιβολή ελάχιστης απόστασης ακρότατων στον αλγόριθμο ανίχνευσης ακρότατων καθιστά την  $\alpha$  πιο αραιή από τις προηγούμενες περιπτώσεις και λύνει το πρόβλημα της διπλής ενεργοποίησης ακρότατων που εμφανίζουν οι δείκτες συγκέντρωσης ακρότατων (όπως φαίνεται στην Εικ. 6.1(e)). Η παράμετρος αραιότητας σε αυτή την περίπτωση ορίζεται  $d^{(i)} = med$ , όπου  $1 \leq med < n \in \mathbb{N}$  είναι η ελάχιστη απόσταση ακρότατων. Ορίσαμε  $med = m^{(i)}$  για τη δίκαιη σύγκριση μεταξύ των συναρτήσεων αραιής ενεργοποίησης. Ειδικά για την συνάρτηση ενεργοποίησης ακρότατων εισάγουμε μια παράμετρο ‘συνοριακής ανοχής’ για να επιτρέψουμε μια πιο έγκαιρη ενεργοποίηση του νευρώνα.

---

### Algorithm 3 Ανίχνευση ακρότατων με ελάχιστη απόσταση $med$

---

**Input:**  $s, med$

**Output:**  $\alpha$

```

1:  $peaks \leftarrow \left( \frac{ds^+}{dt} \geq 0 \right) \wedge \left( \frac{ds^-}{dt} < 0 \right)$ 
2:  $valleys \leftarrow \left( \frac{ds^+}{dt} < 0 \right) \wedge \left( \frac{ds^-}{dt} \geq 0 \right)$ 
   # + and - denote one sample padding to the right and left respectively
3:  $z = peaks \vee valleys$ 
4:  $p_i \leftarrow z > 0$ 
5:  $p_{i_i} \leftarrow sort(z)$ 
6:  $p_{i_{sorted}} \leftarrow p_i(p_{i_i})$ 
7:  $q_i \leftarrow 0, i = 1 \dots card(s)$ 
8: for  $i = 0$  to  $card(s)$  do
9:   if  $\neg q_i$  then
10:     $p_{i_r} \leftarrow p_i \geq p_{i_i} - med$ 
11:     $p_{i_l} \leftarrow p_i \leq p_{i_i} + med$ 
12:     $p_{i_m} \leftarrow p_{i_r} \wedge p_{i_l}$ 
13:     $q \leftarrow q \vee p_{i_m}$ 
14:     $q_i \leftarrow 0$ 
15:   end if
16: end for
17:  $\alpha_{ind} \leftarrow p_{i_{sorted}}(\neg q)$ 
18:  $\alpha_i \leftarrow 0, i = 1 \dots card(s)$ 
19:  $\alpha(\alpha_i) \leftarrow s(\alpha_{ind})$ 
20: return  $\alpha$ 

```

---

### 6.3.2. Αρχιτεκτονική και εκπαίδευση των SANs

Έστω  $\mathbf{x} \in \mathbb{R}^n$  είναι ένα δεδομένο εισόδου, ωστόσο τα παρακάτω μπορεί να γενικευτούν σε εισόδους παρτίδων με διαφορετικά μήκη. Έστω  $\mathbf{w}^{(i)} \in \mathbb{R}^{m^{(i)}}$  ο πίνακας βάρους του  $i^{\text{th}}$  πυρήνα ο οποίος αρχικοποιείται χρησιμοποιώντας μια κανονική κατανομή με μέση τιμή  $\mu$  και τυπική απόκλιση  $\sigma$ :

$$\mathbf{w}^{(i)} \sim \mathcal{N}(\mu, \sigma) \quad (6.9)$$

, όπου  $0 \leq i < q \in \mathbb{N}$  είναι ο αριθμός των πυρήνων.

Αρχικά υπολογίζουμε τους πίνακες ομοιότητας<sup>2</sup> Το  $\mathbf{s}^{(i)}$  για καθένα από τους πίνακες βάρους  $\mathbf{w}^{(i)}$  είναι:

$$\mathbf{s}^{(i)} = \mathbf{x} * \mathbf{w}^{(i)} \quad (6.10)$$

, όπου  $*$  είναι η συνέλιξη<sup>3</sup> Για την διατήρηση του αρχικού μεγέθους του διανύσματος εισόδου συμπληρώνουμε με μηδενικά. Δεν χρειαζόμαστε όρο μεροληψίας επειδή θα εφαρμοζόταν συνολικά στο  $\mathbf{s}^{(i)}$  κάτι το οποίο είναι σχεδόν ισοδύναμο με την μείωση της γραμμής βάσης  $\mathbf{x}$ .

Έπειτα, περνάμε την  $\mathbf{s}^{(i)}$  και μια παράμετρο αραιότητας  $d^{(i)}$  στην συνάρτηση αραιής ενεργοποίησης  $\phi$  με αποτέλεσμα τον χάρτη ενεργοποίησης  $\boldsymbol{\alpha}^{(i)}$ :

$$\boldsymbol{\alpha}^{(i)} = \phi(\mathbf{s}^{(i)}, d^{(i)}) \quad (6.11)$$

, όπου  $\boldsymbol{\alpha}^{(i)}$  είναι ένας αραιός πίνακας, του οποίου τα μη-μηδενικά στοιχεία υποδηλώνουν τις χωρικές θέσεις των στιγμιότυπων του  $i^{\text{th}}$  πυρήνα. Η ακριβής μορφή του  $\phi$  και του  $d^{(i)}$  εξαρτώνται από την επιλογή της συνάρτησης αραιής ενεργοποίησης, οι οποίες παρουσιάζονται στην ενότητα 6.3.1.

Συνελλίσουμε κάθε  $\boldsymbol{\alpha}^{(i)}$  με το αντίστοιχο  $\mathbf{w}^{(i)}$ , έχοντας ως αποτέλεσμα ένα σύνολο ατομικών ανακατασκευών  $\mathbf{r}^{(i)}$  της εισόδου:

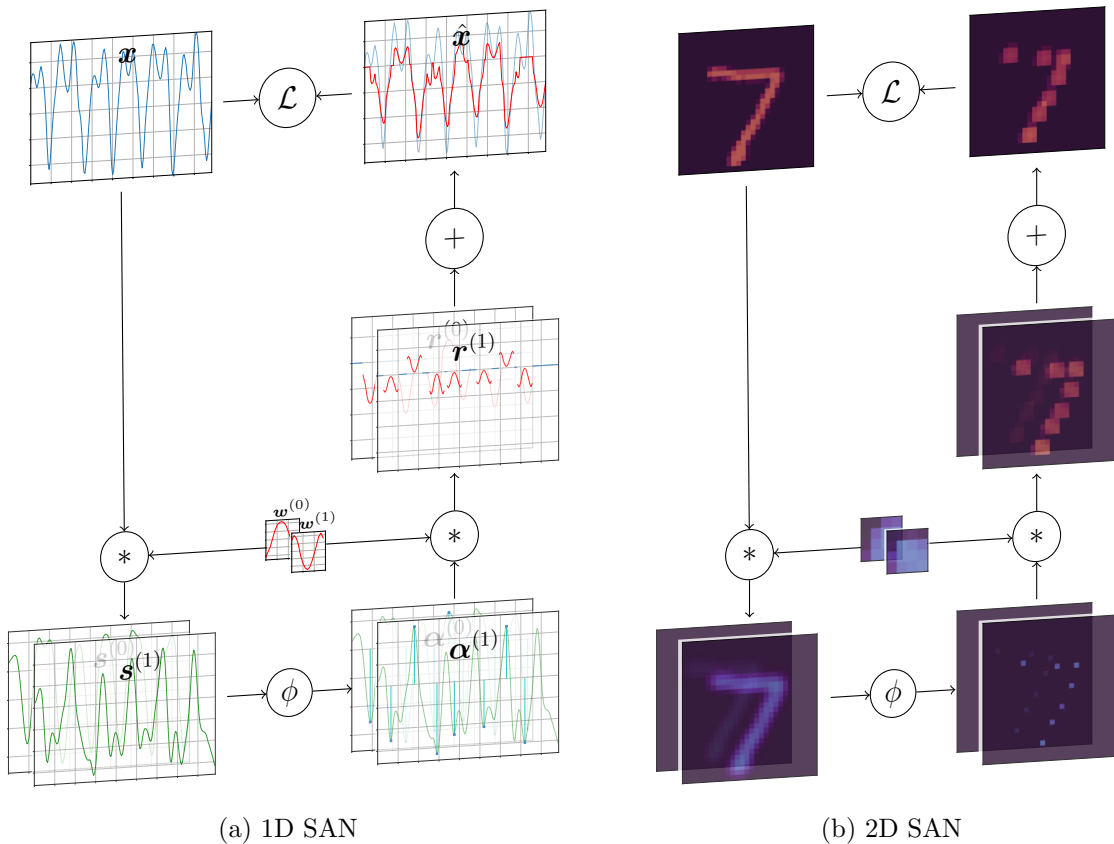
$$\mathbf{r}^{(i)} = \boldsymbol{\alpha}^{(i)} * \mathbf{w}^{(i)} \quad (6.12)$$

, που αποτελείται από αραιά επαναλαμβανόμενα πρότυπα  $\mathbf{w}^{(i)}$  με μεταβλητό πλάτος. Τέλος, μπορούμε να ανακατασκευάσουμε την είσοδο ως το άθροισμα των μεμονωμένων ανακατασκευών  $\mathbf{r}^{(i)}$  ως εξής:

$$\hat{\mathbf{x}} = \sum_{i=1}^q \mathbf{r}^{(i)} \quad (6.13)$$

<sup>2</sup>Η προηγούμενη βιβλιογραφία αναφέρεται σε αυτό ως ‘κρυφή μεταβλητή’, εδώ όμως χρησιμοποιούμε μια πιο άμεση ονομασία που ταιριάζει στο πλαίσιο του παρόντος κειμένου.

<sup>3</sup>Χρησιμοποιούμε την συνέλιξη αντί της αλληλοσυσχέτισης μόνο για λόγους συμβατότητας με την προηγούμενη βιβλιογραφία και τα υπολογιστικά πλαίσια. Η χρήση αλληλοσυσχέτισης θα παρήγαγε τα ίδια αποτελέσματα και επιπλέον δεν θα απαιτούσε την περιστροφή των πυρήνων κατά τη διάρκεια της οπτικοποίησης.



(a) 1D SAN

(b) 2D SAN

Figure 6.2: Διαγράμματα της προς-τα-εμπρός διάδοσης 1D και 2D SAN με δύο πυρήνες για παραδείγματα από την βάση δεδομένων επιληψίας UCI και MNIST αντίστοιχα. Οι εικόνες απεικονίζουν τις ενδιάμεσες αναπαραστάσεις; το  $\mathbf{x}$  δηλώνει το σήμα εισόδου (μπλε γραμμή), το  $\mathbf{w}^{(i)}$  τους πυρήνες (κόκκινη γραμμή), το  $\mathbf{s}^{(i)}$  τους πίνακες ομοιότητας (πράσινη γραμμή), το  $\alpha^{(i)}$  τους χάρτες ενεργοποίησης (κυανές γραμμές με μπλε δείκτες), το  $\mathbf{r}^{(i)}$  την μερική ανακατασκευή από κάθε  $\mathbf{w}^{(i)}$  και το  $\hat{\mathbf{x}}$  την ανακατασκευασμένη είσοδο (κόκκινη γραμμή). Για λόγους σύγκρισης, η διαφανής πράσινη γραμμή στο  $\alpha^{(i)}$  δηλώνει το αντίστοιχο  $\mathbf{s}^{(i)}$  και η διαφανή μπλε γραμμή στο  $\hat{\mathbf{x}}$  δηλώνει την είσοδο  $\mathbf{x}$ . Ο εκθέτης  $i = 0, 1$  αντιστοιχεί στον πρώτο και στον δεύτερο πυρήνα και στις ενδιάμεσες αναπαραστάσεις. Οι κύκλοι δηλώνουν συναρτήσεις. Το  $\mathcal{L}$  δηλώνει τη συνάρτηση απώλειας, το  $\phi$  τη συνάρτηση αραιής ενεργοποίησης, το  $*$  τη συνάρτηση συνέλιξης και το  $+$  τον τελεστή πρόσθεσης. Όλες οι συναρτήσεις εκτελούνται ξεχωριστά για κάθε  $\mathbf{w}^{(i)}$ , ωστόσο για οπτική διαύγεια απεικονίζουμε μόνο μία συνάρτηση για κάθε βήμα. Αποχρώσεις του κόκκινου και του μπλε απεικονίζουν θετικές και αρνητικές τιμές αντίστοιχα. Η Peak συνάρτηση ενεργοποίησης χρησιμοποιήθηκε και για τα δύο δίκτυα.

Το μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE) της εισόδου  $\mathbf{x}$  και η πρόβλεψη  $\hat{\mathbf{x}}$  υπολογίζεται ως εξής:

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{t=1}^n |\hat{\mathbf{x}}_t - \mathbf{x}_t| \quad (6.14)$$

, όπου ο δείκτης  $t$  δηλώνει το  $t^{\text{th}}$  δείγμα. Η επιλογή της MAE βασίζεται στην ανάγκη να αντιμετωπιστούν οι απομακρυσμένες τιμές των δεδομένων με το ίδιο βάρος με τις κανονικές τιμές. Ωστόσο τα SAN δεν περιορίζονται στη χρήση του MAE αλλά μπορούν να χρησιμοποιήσουν και άλλες συναρτήσεις απώλειας, όπως το μέσο τετραγωνικό σφάλμα.

Χρησιμοποιώντας backpropagation υπολογίζονται οι κλίσεις του σφάλματος απώλειας  $\mathcal{L}$  σε σχέση με το  $\mathbf{w}^{(i)}$ :

$$\nabla \mathcal{L} = \left( \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(1)}}, \dots, \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(q)}} \right) \quad (6.15)$$

Τέλος, το  $\mathbf{w}^{(i)}$  ενημερώνεται χρησιμοποιώντας τον ακόλουθο κανόνα μάθησης:

$$\Delta \mathbf{w}^{(i)} = -\lambda \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(i)}} \quad (6.16)$$

, όπου  $\lambda$  είναι ο ρυθμός μάθησης.

Μετά την εκπαίδευση, θεωρούμε τα  $\boldsymbol{\alpha}^{(i)}$  (τα οποία υπολογίζονται κατά τη διάρκεια της προς-τα-εμπρός διάδοσης από την Εξ. 6.11) και τα  $\mathbf{w}^{(i)}$  (τα οποία υπολογίζονται με τη χρήση του backpropagation από την Εξ. 6.16) ως τη συμπιεσμένη αναπαράσταση  $\mathbf{x}$ , η οποία μπορεί να ανακατασκευαστεί από τις Εξισώσεις 6.12 και 6.13:

$$\hat{\mathbf{x}} = \sum_{i=1}^q (\boldsymbol{\alpha}^{(i)} * \mathbf{w}^{(i)}) \quad (6.17)$$

Όσον αφορά το μέτρο  $\varphi$  και παίρνοντας υπόψη την Εξ. 6.17 στόχος μας είναι να υπολογίσουμε μια όσο το δυνατόν ακριβή αναπαράσταση του  $\mathbf{x}$  μέσω των  $\boldsymbol{\alpha}^{(i)}$  και  $\mathbf{w}^{(i)}$  με τον μικρότερο αριθμό μη-μηδενικών ενεργοποιήσεων και βαρών.

Η γενική διαδικασία εκπαίδευσης των SAN για πολλαπλές εποχές χρησιμοποιώντας παρτίδες (αντί για ένα παράδειγμα όπως παρουσιάστηκε προηγουμένως) παρουσιάζεται στον Αλγόριθμο 4.

---

**Algorithm 4** Εκπαίδευση δικτύων αραιής ενεργοποίησης

---

**Input:**  $\mathbf{x}$ **Output:**  $\alpha, \mathbf{w}$ *Hyperparameters:*  $m, \mu, \sigma, q, \phi, d, \lambda, epochs, batches$ 

```

1: for  $i = 1$  to  $q$  do
2:    $\mathbf{w}^{(i)} \sim \mathcal{N}(\mu, \sigma)$ 
3: end for
4: for  $e = 1$  to  $epochs$  do
5:   for  $b = 1$  to  $batches$  do
6:      $\mathbf{x}^{(b)} \sim \mathbf{x}$ 
7:     for  $i = 1$  to  $q$  do
8:        $\mathbf{s}^{(i)} \leftarrow \mathbf{x}^{(b)} * \mathbf{w}^{(i)}$ 
9:        $\alpha^{(i)} \leftarrow \phi(\mathbf{s}^{(i)}, d^{(i)})$ 
10:       $r^{(i)} \leftarrow \alpha^{(i)} * \mathbf{w}^{(i)}$ 
11:    end for
12:     $\hat{\mathbf{x}}^{(b)} \leftarrow \sum_{i=1}^q r^{(i)}$ 
13:     $\mathcal{L} \leftarrow \frac{1}{n} \sum_{t=1}^n |\hat{\mathbf{x}}_t^{(b)} - \mathbf{x}_t^{(b)}|$ 
14:     $\nabla \mathcal{L} = \left( \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(1)}}, \dots, \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(q)}} \right)$ 
15:     $\Delta \mathbf{w}^{(i)} = -\lambda \frac{\partial \mathcal{L}}{\partial \mathbf{w}^{(i)}}$ 
16:   end for
17: end for
18: return  $\alpha, \mathbf{w}$ 

```

---

## 6.4. Πειράματα

Για όλα τα πειράματα τα βάρη των πυρήνων SAN αρχικοποιούνται χρησιμοποιώντας την κανονική κατανομή  $\mathcal{N}(\mu, \sigma)$  με  $\mu = 0$  και  $\sigma = 0.1$ . Χρησιμοποιήσαμε τον Adam [3] ως βελτιστοποιητή με ρυθμό μάθησης  $\lambda = 0.01$ ,  $b_1 = 0.9$ ,  $b_2 = 0.999$ , epsilon  $\epsilon = 10^{-8}$  χωρίς σταδιακή απομείωση βαρών. Για την υλοποίηση και την εκπαίδευση των SANs χρησιμοποιήσαμε το Pytorch [32], με NVIDIA Titan X Pascal GPU 12GB RAM και 12 Core Intel i7-8700 CPU @ 3.20GHz σε λειτουργικό σύστημα Linux.

### 6.4.1. Σύγκριση του μέτρου $\varphi$ για συναρτήσεις αραιής ενεργοποίησης και διάφορα μεγέθη πυρήνα στην Physionet

Εδώ μελετάμε την επίδραση στο  $\bar{\varphi}$ , της επιλογής του μεγέθους πυρήνα  $m$  και των συναρτήσεων αραιής ενεργοποίησης που ορίστηκαν στην ενότητα 6.3.1.

#### Βάσεις δεδομένων

Χρησιμοποιούμε ένα σήμα από καθεμία από τις 15 βάσεις δεδομένων της Physionet που αναφέρονται στην πρώτη στήλη του Πίνακα 6.1. Κάθε σήμα αποτελείται από 12000 δείγματα τα οποία με τη σειρά τους κατανέμονται σε 12 σήματα με 1000 δείγματα το καθένα, για τη δημιουργία των σημάτων εκπαίδευσης (6), επικύρωσης (2 σήματα) και δοκιμής (4 σήματα). Η μόνη προεπεξεργασία που γίνεται είναι η αφαίρεση του μέσου όρου και διαίρεση με την τυπική απόκλιση στα σήματα των 1000 δειγμάτων.

#### Ρύθμιση πειράματος

Εκπαιδεύουμε τέσσερα SAN (ένα για κάθε συνάρτηση αραιής ενεργοποίησης) για καθεμία από τις 15 βάσεις δεδομένων της Physionet για 30 εποχές με μέγεθος παρτίδας 2 και μέγεθος πυρήνων που κυμαίνονται στην περιοχή  $[1, 250]$ . Κατά τη διάρκεια της επικύρωσης επιλέξαμε τα μοντέλα με μέγεθος πυρήνα που πέτυχαν το καλύτερο  $\bar{\varphi}$  από όλες τις εποχές. Κατά τη διάρκεια των δοκιμών τροφοδοτούμε τα δεδομένα δοκιμής στο επιλεγμένο μοντέλο και υπολογίζουμε τα  $CR^{-1}$ ,  $\tilde{\mathcal{L}}$  και  $\bar{\varphi}$  για αυτό το σύνολο υπερπαραμέτρων όπως φαίνεται στον Πίνακα 6.1. Για τη συνάρτηση ενεργοποίησης ακρότατων θέτουμε ‘συνοριακή ανοχή’ τριών δειγμάτων.

#### Αποτελέσματα

Η ύπαρξη των τριών ξεχωριστών ομάδων που απεικονίζονται στην Εικ. 6.3 και Εικ. 6.4(a) μεταξύ της Identity, του ReLU και των υπολοίπων είναι το αποτέλεσμα της επίδρασης της αραιότητας των συναρτήσεων ενεργοποίησης στις αναπαραστάσεις. Όσο αραιότερη η συνάρτηση ενεργοποίησης είναι τόσο περισσότερο συμπιέζεται η αναπαράσταση, μερικές φορές σε βάρος του σφάλματος ανακατασκευής. Ωστόσο, με οπτική επιθεώρηση της Εικ.6.5 μπορούμε να επιβεβαιώσουμε ότι τα βάρη των SANs με αραιότερους χάρτες ενεργοποίησης (Extrema-Pool indices και Extrema) αντιστοιχούν σε επαναλαμβανόμενα μοτίβα των βάσεων δεδομένων, επιτυγχάνοντας έτσι υψηλή ερμηνευσιμότητα. Αυτά τα αποτελέσματα δείχνουν ότι το σφάλμα ανακατασκευής από μόνο του δεν είναι επαρκές μέτρο για την αποσύνθεση δεδομένων σε ερμηνεύσιμα στοιχεία. Προσπαθώντας να επιτύχουμε αποκλειστικά μικρότερο σφάλμα ανακατασκευής (όπως

στην περίπτωση της Identity) έχουμε ως αποτέλεσμα θορυβώδεις πυρήνες, ενώ το συνδυασμένο μέτρο του σφάλματος ανακατασκευής και του λόγου συμπίεσης (μικρότερο  $\bar{\varphi}$ ) έχει ως αποτέλεσμα ερμηνεύσιμους πυρήνες. Συγκρίνοντας τις διαφορές στο  $\bar{\varphi}$  μεταξύ του Identity, του ReLU και των υπόλοιπων συναρτήσεων αραιής ενεργοποίησης στην Εικ. 6.4(c) παρατηρούμε ότι οι τελευταίες παράγουν μια περιοχή ελαχίστων στην οποία έχουμε ως αποτέλεσμα ερμηνεύσιμους πυρήνες.

### 6.4.2. Αξιολόγηση της ανακατασκευής των SANs με χρήση επιβλεπόμενου CNN για αναγνώριση επιληψίας στην UCI

Εδώ μελετάμε την ποιότητα των ανακατασκευών των SAN εκπαιδεύοντας ένα επιτηρούμενο 1D Συνελικτικό Νευρωνικό Δίκτυο (CNN) στις εξόδους του κάθε SAN. Επίσης, μελετάμε την επίδραση που έχει το  $m$  στο  $\bar{\varphi}$  και την ακρίβεια του ταξινομητή για τις πέντε συναρτήσεις αραιής ενεργοποίησης.

#### Βάση δεδομένων

Χρησιμοποιούμε τη βάση δεδομένων αναγνώρισης επιληψίας από την UCI που αποτελείται από 500 σήματα των 4097 δειγμάτων (23.5 δευτερόλεπτα) το καθένα. Η βάση δεδομένων απαριθμείται σε πέντε κατηγορίες με 100 σήματα για κάθε κατηγορία. Για τους σκοπούς αυτής της εργασίας χρησιμοποιούμε μια παραλλαγή της βάσης δεδομένων<sup>4</sup> στην οποία τα EEG σήματα χωρίζονται σε τμήματα των 178 δειγμάτων το καθένα, με αποτέλεσμα μια ισορροπημένη βάση δεδομένων που αποτελείται από 11500 σήματα EEG συνολικά.

#### Ρύθμιση πειράματος

Αρχικά, συγχωνεύουμε τις κατηγορίες όγκων (2 και 3) και των ματιών (4 και 5) με αποτέλεσμα μια τροποποιημένη βάση δεδομένων τριών κατηγοριών (όγκος, μάτια, επιληψία). Στη συνέχεια διαιρούμε τα 11500 σήματα σε 76%, 12% και 12% (8740, 1380, 1380) ως δεδομένα εκπαίδευσης, επικύρωσης και δοκιμής αντίστοιχα και κανονικοποιούμε στο εύρος  $[0, 1]$  χρησιμοποιώντας το ολικό μέγιστο και ελάχιστο. Για τα SAN χρησιμοποιήσαμε δύο πυρήνες  $q = 2$  με μεταβλητό μήκος εύρους  $[15, 22]$  και εκπαιδεύσαμε για 5 εποχές με μέγεθος παρτίδας 32. Μετά την εκπαίδευση, επιλέγουμε το μοντέλο που εμφάνισε το χαμηλότερο  $\bar{\varphi}$  από όλες τις εποχές.

Κατά τη διάρκεια της επιβλεπόμενης μάθησης, τα βάρη των πυρήνων παγώνουν και ένα CNN στοιβάζεται πάνω από τις ανακατασκευές των SAN. Ο εξαγωγέας χαρακ-

<sup>4</sup><https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition>



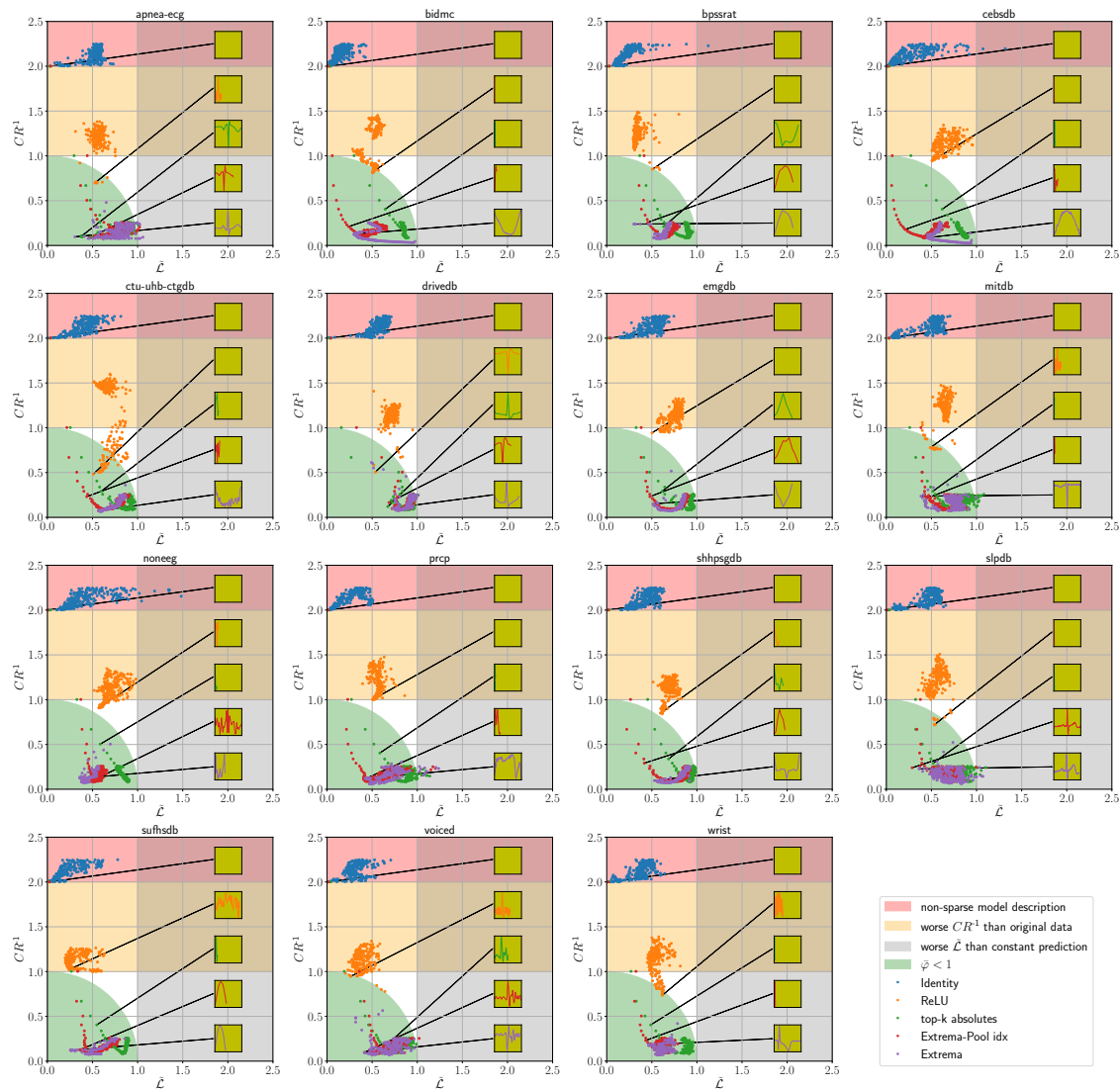


Figure 6.3: Αντίστροφος λόγος συμπίεσης ( $CR^{-1}$ ) έναντι κανονικοποιημένης απώλειας ανακατασκευής ( $\tilde{\mathcal{L}}$ ) για 15 βάσεις δεδομένων της Physionet και για διάφορα μεγέθη πυρήνα. Οι πέντε μικρές γραφικές παραστάσεις με το κίτρινο φόντο στα δεξιά της κάθε εικόνας, απεικονίζουν τον αντίστοιχο πυρήνα για το μέγεθος του πυρήνα που πέτυχε το καλύτερο  $\varphi$ .

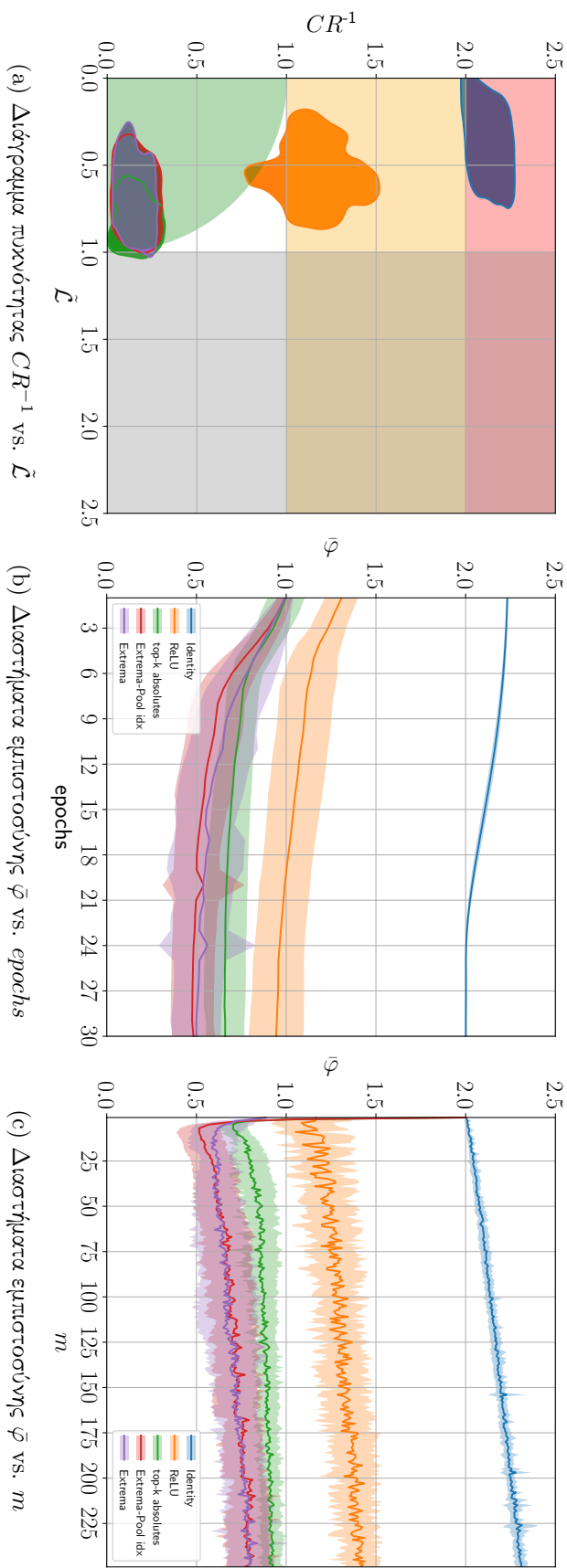


Figure 6.4: Συνοπτικά αποτελέσματα της αξιολόγησης των βάσεων δεδομένων της Physionet με τη χρήση του μέτρου  $\varphi$ . Το διάγραμμα πυκνότητας δημιουργήθηκε χρησιμοποιώντας εκτίμηση πυκνότητας γκαουσιανών πυρήνων και τα διαστήματα εμπιστοσύνης απεικονίζουν μια τυπική απόκλιση.

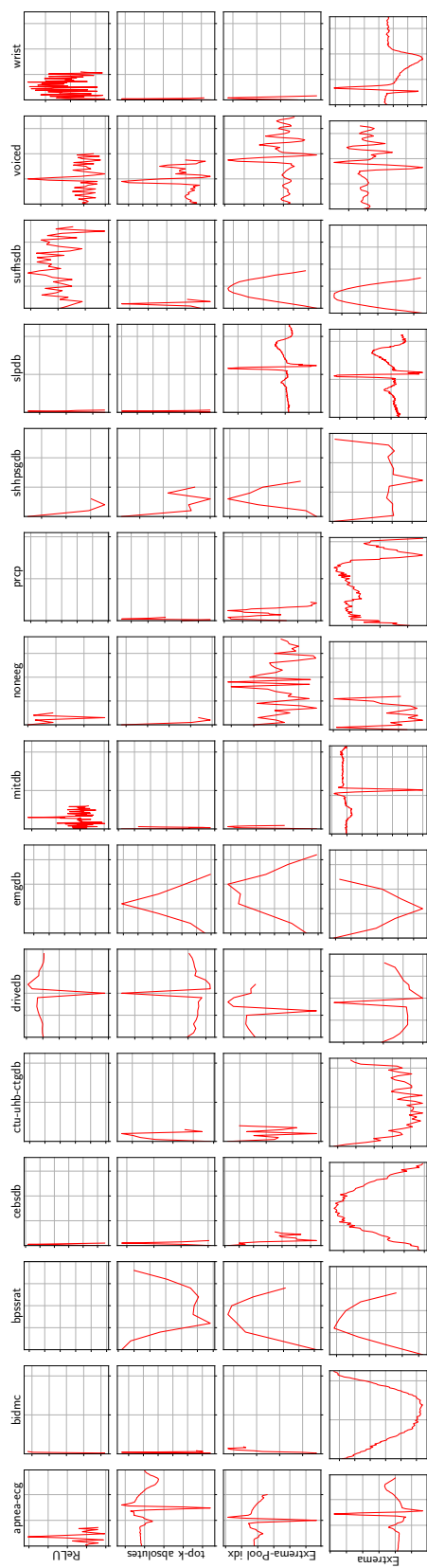


Figure 6.5: Οπτικοποίηση των πυρήνων για κάθε αραιή συνάρτηση ενεργοποίησης (γραμμή) και για κάθε βάση δεδομένων της Physionet (στήλη).

Table 6.1: Μέγεθος πυρήνα  $m$  με το καλύτερο  $\varphi$  για κάθε συνάρτηση αφαίξης ενεργοποίησης για κάθε βάση δεδομένων της Physionet

Datasets	Identity			ReLU			top-k absolutes			Extrema-Pool idx			Extrema							
	$m$	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\varphi$	$m$	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\varphi$	$m$	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\varphi$	$m$	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\varphi$				
apnea-ecg	1	2.00	0.03	2.00	19	0.70	0.53	0.87	74	0.10	0.37	0.39	51	0.09	0.47	0.48	72	0.10	0.31	0.32
bidmc	1	2.00	0.04	2.00	4	0.82	0.50	0.96	5	0.41	0.64	0.76	10	0.21	0.24	0.32	113	0.13	0.30	0.32
bpsrrat	1	2.00	0.02	2.00	1	0.85	0.51	0.99	10	0.21	0.63	0.67	8	0.26	0.45	0.52	8	0.24	0.30	0.38
cebsdb	1	2.00	0.01	2.00	3	0.95	0.51	1.07	5	0.41	0.62	0.74	12	0.18	0.21	0.28	71	0.09	0.45	0.46
ctu-uhb-ctgdb	1	2.00	0.01	2.00	1	0.48	0.51	0.71	7	0.29	0.60	0.66	9	0.23	0.44	0.49	45	0.07	0.57	0.57
drivedb	1	2.00	0.04	2.00	20	0.51	0.54	0.74	20	0.12	0.67	0.68	13	0.17	0.69	0.71	19	0.10	0.72	0.73
emgdb	1	2.00	0.04	2.00	1	0.94	0.50	1.07	7	0.29	0.62	0.68	9	0.23	0.48	0.53	7	0.15	0.51	0.53
mitdb	1	2.00	0.03	2.00	61	0.78	0.49	0.92	7	0.29	0.52	0.59	10	0.21	0.44	0.49	229	0.24	0.38	0.45
noneeg	1	2.00	0.01	2.00	6	0.91	0.57	1.08	4	0.50	0.59	0.77	37	0.09	0.49	0.50	15	0.12	0.36	0.38
prcp	1	2.00	0.03	2.00	1	1.00	0.51	1.12	5	0.41	0.59	0.71	23	0.11	0.41	0.42	105	0.12	0.42	0.44
shhpsgdb	1	2.00	0.02	2.00	4	0.85	0.60	1.05	6	0.34	0.69	0.77	7	0.29	0.42	0.51	15	0.10	0.53	0.54
slpdb	1	2.00	0.03	2.00	7	0.72	0.53	0.90	7	0.29	0.52	0.60	232	0.24	0.29	0.37	218	0.23	0.36	0.43
suffnsdb	1	2.00	0.03	2.00	38	1.02	0.24	1.05	5	0.41	0.55	0.68	18	0.13	0.36	0.39	17	0.12	0.26	0.28
voiced	1	2.00	0.01	2.00	41	0.95	0.26	0.98	36	0.09	0.56	0.57	70	0.10	0.41	0.43	67	0.10	0.41	0.43
wrist	1	2.00	0.04	2.00	56	0.74	0.62	0.96	5	0.41	0.49	0.63	9	0.23	0.43	0.49	173	0.18	0.46	0.50

τηριστικών του CNN αποτελείται από δύο συνελικτικά επίπεδα με φίλτρα 3 και 16 και μέγεθος πυρήνα 5, το καθένα από τα οποία ακολουθείται από ένα ReLU και Extrema-Pool με μέγεθος συγκέντρωσης 2. Ο ταξινομητής αποτελείται από τρία πλήρως συνδεδεμένα επίπεδα με μονάδες 656, 120 και 84. Τα πρώτα δύο από τα πλήρως συνδεδεμένα επίπεδα ακολουθούνται από ένα ReLU ενώ το τελευταίο περνάει από ένα log-softmax το οποίο παράγει τις προβλέψεις. Το CNN εκπαιδεύεται για 5 πρόσθετες εποχές με το ίδιο μέγεθος παρτίδας και τη διαδικασία επιλογής μοντέλου όπως και με τα SAN και την αρνητική λογαριθμική πιθανοφάνεια ως συνάρτηση απώλειας. Για τη συνάρτηση ενεργοποίησης ακρότατων θέτουμε ‘συνοριακή ανοχή’ δύο δειγμάτων.

### Αποτελέσματα

Όπως φαίνεται στον Πίνακα. 6.2, αν και χρησιμοποιούμε ένα σημαντικά μειωμένο μέγεθος αναπαράστασης, η ακρίβεια δεν πέφτει ανάλογα κάτι το οποίο δείχνει ότι τα SANs επιλέγουν τα πιο σημαντικά χαρακτηριστικά για να αναπαραστήσουν τα δεδομένα. Για παράδειγμα, για  $m = 15$  για την Peak συνάρτηση ενεργοποίησης, υπάρχει μια πτώση ακρίβειας 1.44% (το CNN βάση αναφοράς με τα αρχικά δεδομένα πέτυχε 82.14%) παρόλο που χρησιμοποιήθηκε μια μειωμένη αναπαράσταση με μόλις 34% μέγεθος σε σχέση με τα αρχικά δεδομένα.

### 6.4.3. Αξιολόγηση της ανακατασκευής των SANs με χρήση επιβλεπόμενου FNN στην MNIST και FMNIST

#### Βάση δεδομένων

Για την ίδια εργασία με την προηγούμενη, αλλά για 2D, χρησιμοποιούμε την MNIST [33] η οποία αποτελείται από μια βάση δεδομένων εκπαίδευσης 60000 χειρόγραφων ψηφίων στην κλίμακα του γκρι και μια βάση δεδομένων δοκιμής με 10000 εικόνες καθεμία με μέγεθος  $28 \times 28$ . Η ίδια διαδικασία ακολουθείται και για την FMNIST [34].

#### Ρύθμιση πειράματος

Τα μοντέλα αποτελούνται από δύο πυρήνες  $q = 2$  με μεταβλητό μήκος στο εύρος [1, 6]. Χρησιμοποιούμε 10000 εικόνες από τη βάση δεδομένων εκπαίδευσης για επικύρωση και εκπαιδεύουμε με τα υπόλοιπα 50000 για 5 εποχές και μέγεθος παρτίδας 64. Δεν εφαρμόζουμε προεπεξεργασία στις εικόνες.

Κατά τη διάρκεια της επιβλεπόμενης μάθησης, τα βάρη των πυρήνων παγώνουν και ένα μονοστρωματικό πλήρως συνδεδεμένο δίκτυο (FNN) στοιβάζεται πάνω από τις

Table 6.2: SANNs με επιβλεπόμενο στοιβαγμένο CNN για αναγνώριση επιδημίας στην UCI

$m$	Identity			ReLU			top-k absolutes			Extrema-Pool idx			Extrema							
	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%				
<b>15</b>	4.17	0.02	4.17	40.8	2.17	0.03	2.17	82.6	0.42	0.79	0.89	74.1	0.42	0.46	0.62	73.4	0.34	0.43	0.55	80.7
<b>16</b>	4.18	0.02	4.18	91.2	2.18	0.04	2.18	87.0	0.43	0.78	0.89	70.5	0.43	0.44	0.61	78.8	0.34	0.44	0.56	74.5
<b>17</b>	4.19	0.02	4.19	89.0	2.20	0.02	2.20	90.1	0.42	0.78	0.89	73.0	0.42	0.45	0.61	78.9	0.34	0.44	0.56	75.1
<b>18</b>	4.20	0.04	4.20	91.2	2.21	0.05	2.21	87.1	0.40	0.79	0.88	73.9	0.40	0.48	0.63	79.7	0.34	0.45	0.57	67.2
<b>19</b>	4.21	0.02	4.21	88.6	2.21	0.03	2.21	89.6	0.42	0.78	0.88	76.1	0.42	0.45	0.61	78.8	0.36	0.44	0.57	78.3
<b>20</b>	4.22	0.03	4.22	88.6	2.22	0.03	2.22	87.4	0.40	0.79	0.89	69.6	0.40	0.49	0.64	77.5	0.35	0.45	0.58	73.8
<b>21</b>	4.24	0.02	4.24	89.4	2.23	0.03	2.24	87.0	0.42	0.79	0.89	73.0	0.42	0.47	0.63	70.7	0.37	0.44	0.58	73.8
<b>22</b>	4.25	0.03	4.25	89.3	2.26	0.04	2.26	89.4	0.43	0.78	0.89	73.6	0.43	0.45	0.62	78.7	0.38	0.44	0.59	76.5

τελικές ανακατασκευές των SAN. Το FNN εκπαιδεύεται για 5 πρόσθετες εποχές με την ίδια διαδικασία επιλογής μοντέλου και μέγεθος παρτίδας όπως και με τα SANs και την αρνητική λογαριθμική πιθανοφάνεια ως συνάρτηση απώλειας. Για τη συνάρτηση ενεργοποίησης ακρότατων θέτουμε ‘συνοριακή ανοχή’ δύο δειγμάτων.

### Αποτελέσματα

Όπως φαίνεται στον Πίνακα 6.3, η ακρίβεια που επιτυγχάνεται με τις ανακατασκευές ορισμένων SANs είναι συγκρίσιμη με εκείνες ενός FNN που έχει εκπαιδευτεί στα αρχικά δεδομένα (92.17%), αν και έχουν συμπεστεί σε μεγάλο βαθμό. Είναι ενδιαφέρον να τονιστεί ότι σε μερικές περιπτώσεις οι ανακατασκευές των SANs, όπως των Extrema-Pool indices, πέτυχαν καλύτερη ακρίβεια από τα αρχικά δεδομένα. Αυτό υποδηλώνει τη συντριπτική παρουσία περιττής πληροφορίας που βρίσκεται στις αρχικές εικόνες των αρχικών δεδομένων και την δυνατότητα των SANs να εξάγουν τα πιο σημαντικά χαρακτηριστικά από τα δεδομένα.

## 6.5. Συζήτηση

Τα SANs σε συνδυασμό με το μέτρο  $\varphi$  συμπιέζουν την περιγραφή των δεδομένων σε  $\mathbf{w}^{(i)}$  και  $\mathbf{\alpha}^{(i)}$  κατά παρόμοιο τρόπο με ένα πλαίσιο γλώσσας ελάχιστης περιγραφής (Minimum Description Language). Τα πειράματα που έγιναν στην ενότητα 6.4 δείχνουν ότι η χρήση της ταυτότητας, του ReLU και (σε μικρότερο βαθμό) των Μέγιστων-Ενεργοποιήσεων παράγουν θορυβώδη χαρακτηριστικά, ενώ από την άλλη πλευρά οι δείκτες συγκέντρωσης ακρότατων και τα ακρότατα παράγουν σταθερά χαρακτηριστικά και μπορούν να προσαρμοστούν χρησιμοποιώντας παραμέτρους (μήκος πυρήνα, και *med*) των οποίων οι τιμές μπορούν να καθοριστούν με απλή επισκόπηση των δεδομένων.

Από την πλευρά της μάθησης αραιών λεξιλογίων (Sparse Dictionary Learning), οι πυρήνες των SANs θα μπορούσαν να θεωρηθούν ως άτομα ενός λεξικού που ειδικεύεται στην ερμηνευτική αντιστοίχιση μοτίβων (e.g. για είσοδο ECG οι πυρήνες των SAN είναι ECG beats) και ο χάρτης αραιής ενεργοποίησης ως η αναπαράσταση. Το γεγονός ότι τα SANs είναι ευρύ με λιγότερα και μεγαλύτερα μεγέθη πυρήνα αντί για βαθιά με μικρότερα και περισσότερα μεγέθη πυρήνα, τα καθιστούν πιο ερμηνεύσιμα από τα DNN και σε ορισμένες περιπτώσεις χωρίς να θυσιάζουν σημαντική ακρίβεια.

Ένα πλεονέκτημα των SANs σε σχέση με τους Αραιούς Αυτοκωδικοποιητές (Sparse Autoencoders) [35] είναι ότι ο περιορισμός της εγγύτητας των ενεργοποιήσεων μπορεί να εφαρμοστεί ατομικά για κάθε διάνυσμα εισόδου σε αντίθεση με τον υπολογισμό της προς-τα-εμπρός διάδοσης όλων των διανυσμάτων εισόδου.

Table 6.3: SANs με επιβλεπόμενο στοιβαγμένο FNN στην MNIST

$m$	Identity			ReLU			top-k absolutes			Extrema-Pool idx			Extrema							
	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%	$CR^{-1}$	$\tilde{\mathcal{L}}$	$\bar{\varphi}$	Acc%				
<b>1</b>	1.16	0.01	1.16	91.7	0.58	0.00	0.58	92.0	1.16	0.00	1.16	91.1	1.16	0.00	1.16	91.1	0.08	0.88	0.88	83.3
<b>2</b>	1.55	0.02	1.55	92.1	0.58	0.01	0.58	90.4	1.37	0.01	1.37	92.0	0.48	0.62	0.79	93.2	0.09	0.83	0.83	84.5
<b>3</b>	1.93	0.02	1.93	91.5	0.65	0.00	0.65	91.5	0.63	0.26	0.68	90.3	0.30	0.51	0.59	92.1	0.08	0.50	0.51	84.0
<b>4</b>	2.30	0.05	2.30	91.7	1.57	0.02	1.57	91.3	0.39	0.41	0.57	87.9	0.22	0.59	0.63	91.9	0.06	0.57	0.57	82.8
<b>5</b>	2.66	0.07	2.66	90.0	0.65	0.03	0.65	91.3	0.20	0.54	0.58	87.3	0.16	0.60	0.62	91.5	0.08	0.57	0.58	82.4
<b>6</b>	3.02	0.13	3.02	91.8	1.58	0.02	1.58	91.7	0.14	0.61	0.63	84.2	0.12	0.63	0.65	90.8	0.06	0.60	0.61	81.2



Επιπλέον, τα SANs δημιουργούν ακριβή μηδενικά αντί για περίπου-μηδενικά, κάτι το οποίο μειώνει την προσαρμογή μεταξύ των ενεργοποιήσεων των νευρώνων.

Το  $\varphi$  θα μπορούσε να θεωρηθεί ως μια εναλλακτική τυποποίηση του ζυραφιού του Occam [36], όπως η θεωρία του Solomonov για την επαγωγική εξαγωγή [37], αλλά με μια αιτιοκρατική ερμηνεία αντί για πιθανοτική. Το κόστος της περιγραφής των δεδομένων μπορεί να θεωρηθεί ότι είναι ανάλογο του αριθμού των βαρών και του αριθμού των μη-μηδενικών ενεργοποιήσεων, ενώ η ποιότητα της περιγραφής είναι ανάλογη με την απώλεια ανακατασκευής. Το μέτρο  $\varphi$  σχετίζεται επίσης με τη θεωρία του rate-distortion [38], όπου η μέγιστη παραμόρφωση ορίζεται σύμφωνα με την ανθρώπινη αντίληψη, η οποία όμως αναπόφευκτα εισάγει προκατάληψη. Υπάρχει επίσης σχέση με τον τομέα της συμπιεσμένης αντίχνευσης (Compressed Sensing) [39], στην οποία εκμεταλλευόμαστε την αραιότητα των δεδομένων, επιτρέποντάς μας να τα ανακατασκευάσουμε με λιγότερα δείγματα από αυτό που απαιτείται από το θεώρημα Nyquist-Shannon και τον τομέα της εξαγωγής σταθερών χαρακτηριστικών (Robust Feature Extraction) [40] όπου τα χαρακτηριστικά χρησιμοποιούνται για αντιπροσώπευση των αρχικών δεδομένων. Οι Olshausen et al. [41] παρουσίασαν μια συνάρτηση βελτιστοποίησης που εξετάζει υποκειμενικά μέτρα της αραιότητας των χαρτών ενεργοποίησης, ωστόσο σε αυτό το έργο χρησιμοποιούμε το άμεσο μέτρο του λόγου συμπίεσης. Προηγούμενες εργασίες όπως [42] χρησιμοποίησαν ένα σταθμισμένο συνδυασμό του αριθμού των νευρώνων, της διαφοράς μεταξύ των διαστημάτων root-mean-squared και ενός συντελεστή συσχέτισης για τη συνάρτηση βελτιστοποίησης ενός FNN ως μέτρο αλλά χωρίς να ληφθεί υπόψη ο αριθμός των μη-μηδενικών ενεργοποιήσεων.

Ένας περιορισμός των SAN είναι η χρήση μεταβλητών πυρήνων μόνο στο πλάτος, το οποίο δεν επαρκεί για πιο περίπλοκα δεδομένα και επίσης δεν αξιοποιείται πλήρως η συμπίεστικότητα των δεδομένων. Μια πιθανή λύση θα ήταν να χρησιμοποιηθεί ένας δειγματολήπτης πλέγματος [43] για τον πυρήνα, ο οποίος θα του επέτρεπε να μάθει πιο γενικές μεταβολές (όπως κλίμακα) από την απλή μεταβλητότητα πλάτους. Ωστόσο, οι πρόσθετες ιδιότητες του πυρήνα θα πρέπει να επιλέγονται παίρνοντας υπόψη τις επιπτώσεις στο μέτρο  $\varphi$ ; το μοντέλο θα πρέπει να συμπεστεί περισσότερο με μειωμένη απώλεια ανακατασκευής.

## Βιβλιογραφία

- [1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [2] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986.
- [3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [5] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.
- [6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [9] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.
- [10] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2019.
- [11] Alireza Aghasi, Afshin Abdi, Nam Nguyen, and Justin Romberg. Net-trim: Convex pruning of deep neural networks with performance guarantee. In *Advances in Neural Information Processing Systems*, pages 3177–3186, 2017.

- 
- [12] Ji Lin, Yongming Rao, Jiwen Lu, and Jie Zhou. Runtime neural pruning. In *Advances in Neural Information Processing Systems*, pages 2181–2191, 2017.
- [13] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [14] Simon B Laughlin and Terrence J Sejnowski. Communication in neuronal networks. *Science*, 301(5641):1870–1874, 2003.
- [15] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [16] Yoshua Bengio, Patrice Simard, Paolo Frasconi, et al. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
- [17] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [18] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [20] Ian Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, pages 1319–1327, 17–19 Jun 2013.
- [21] Alireza Makhzani and Brendan Frey. K-sparse autoencoders. *arXiv preprint arXiv:1312.5663*, 2013.
- [22] Paul Bush and Terrence Sejnowski. Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *Journal of computational neuroscience*, 3(2):91–110, 1996.
- [23] Yoshua Bengio, Dong-Hyun Lee, Jorg Bornschein, Thomas Mesnard, and Zhouhan Lin. Towards biologically plausible deep learning. *arXiv preprint arXiv:1502.04156*, 2015.

- [24] Martin Rehn and Friedrich T Sommer. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of computational neuroscience*, 22(2):135–146, 2007.
- [25] Thomas Heiberg, Birgit Kriener, Tom Tetzlaff, Gaute T Einevoll, and Hans E Plesser. Firing-rate models for neurons with a broad repertoire of spiking behaviors. *Journal of computational neuroscience*, 45(2):103–132, 2018.
- [26] Paschalis Bizopoulos and Dimitrios Koutsouris. Deep learning in cardiology. *IEEE reviews in biomedical engineering*, 12:168–193, 2019.
- [27] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.
- [28] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
- [29] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7):e0130140, 2015.
- [30] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.
- [31] Léonard Blier and Yann Ollivier. The description length of deep learning models. In *Advances in Neural Information Processing Systems*, pages 2216–2226, 2018.
- [32] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.
- [33] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [34] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

- 
- [35] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.
- [36] Andrei N Soklakov. Occam’s razor as a formal basis for a physical theory. *Foundations of Physics Letters*, 15(2):107–135, 2002.
- [37] Ray J Solomonoff. A formal theory of inductive inference. part i. *Information and control*, 7(1):1–22, 1964.
- [38] T Burger. Rate distortion theory, 1971.
- [39] David L Donoho et al. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [40] Yelin Kim, Honglak Lee, and Emily Mower Provost. Deep learning for robust feature generation in audiovisual emotion recognition. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 3687–3691. IEEE, 2013.
- [41] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996.
- [42] Bo Zhang, Jiasheng Zhao, Xiao Chen, and Jianhuang Wu. Ecg data compression using a neural network model based on multi-objective optimization. *PloS one*, 12(10):e0182500, 2017.
- [43] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.



## 7. Επίλογος

Με κάθε τεχνολογική πρόοδο η ιατρική, η οποία μέχρι τώρα είναι βαθιά εξαρτώμενη από τον ανθρώπινο παράγοντα, έρχεται πιο κοντά σε ένα αυτοματοποιημένο τομέα κατευθυνόμενο από την τεχνητή νοημοσύνη. Το AI όχι μόνο θα φτάσει στο σημείο που θα ανιχνεύει ασθένειες σε πραγματικό χρόνο, αλλά θα ερμηνεύει επίσης αμφιλεγόμενες καταστάσεις, φαινοτυπικές πολύπλοκες ασθένειες και θα παίρνει ιατρικές αποφάσεις. Ωστόσο, η πλήρης θεωρητική κατανόηση της βαθιάς μάθησης δεν είναι ακόμη διαθέσιμη και η κριτική κατανόηση των πλεονεκτημάτων και των περιορισμών της εσωτερικής λειτουργίας της είναι ζωτικής σημασίας για να κερδίσει τη θέση της στην καθημερινή κλινική χρήση. Η επιτυχής εφαρμογή του AI στον ιατρικό τομέα βασίζεται στην επίτευξη ερμηνεύσιμων μοντέλων και δημιουργία μεγάλων βάσεων δεδομένων.

Στο πλαίσιο της διδακτορικής διατριβής προτείναμε το μέτρο  $\varphi$  για να αξιολογήσουμε πόσο καλά τα μοντέλα μηχανικής μάθησης ανταλλάσσουν απώλεια ανακατασκευής με συμπίεση. Έπειτα προτείναμε μια νέα αρχιτεκτονική νευρωνικών δικτύων τα **SANs** τα οποία έχουν ελάχιστη δομή και με την χρήση των συναρτήσεων αφαιής ενεργοποίησης μαθαίνουν να συμπιέζουν δεδομένα χωρίς να χάνουν σημαντικές πληροφορίες. Χρησιμοποιώντας τις βάσεις δεδομένων Physionet και MNIST αποδείξαμε ότι τα SANs είναι σε θέση να δημιουργούν αναπαραστάσεις υψηλής ποιότητας με ερμηνεύσιμους πυρήνες.

Η ελάχιστη δομή των SANs καθιστά εύκολη τη χρήση τους για την εξαγωγή χαρακτηριστικών, την ομαδοποίηση και την πρόβλεψη χρονοσειρών. Άλλες μελλοντικές εργασίες σχετικά με τα SANs περιλαμβάνουν:

- Εφαρμογή αλγορίθμων σταδιακής μείωσης των ελάχιστων αποστάσεων των ακρότατων, για την αύξηση του βαθμού ελευθερίας των πυρήνων.
- Επιβολή ελάχιστης απόστασης ακρότατων κατά μήκος όλων των πινάκων ομοιότητας για πολλαπλούς πυρήνες, κάνοντας έτσι τους πυρήνες να ανταγωνίζονται για περιοχές.
- Εφαρμογή του dropout στις ενεργοποιήσεις για να διορθωθούν τα βάρη που έχουν υπερβεί, ειδικά όταν αρχικοποιούνται με υψηλές τιμές. Ωστόσο, η επίδραση

του dropout στα SAN θα ήταν γενικά αρνητική, καθώς τα SANs έχουν πολύ μικρότερο αριθμό βαρών από τα DNNs και επομένως δεν χρειάζονται ισχυρή κανονικοποίηση.

- Χρήση των SAN με δυναμικά δημιουργούμενους πυρήνες οι οποίοι θα μπορούσαν να μάθουν πολυτροπικά δεδομένα από μεταβλητές πηγές (e.g. από ECG σε αναπνευστικά) χωρίς να καταστρέψουν προηγούμενα βάρη.



# Ακρωνύμια

## Ακρωνύμιο Επεξήγηση

ACDC	Automated Cardiac Diagnosis Challenge
ACS	Acute Coronary Syndrome
AE	Autoencoder
AF	Atrial Fibrillation
AI	Artificial Intelligence
BIH	Beth Israel Hospital
BP	Blood Pressure
CAC	Coronary Artery Calcification
CAD	Coronary Artery Disease
CHF	Congestive Heart Failure
CNN	Convolutional Neural Network
CRF	Conditional Random Field
CT	Computerized Tomography
CVD	Cardiovascular Disease
DBN	Deep Belief Networks
DBP	Diastolic Blood Pressure
DNN	Deep Neural Networks
DWI	Diffusion Weighted Imaging
ECG	Electrocardiogram
EEG	Electroencephalogram
EHR	Electronic Health Record
FCN	Fully Convolutional Network
FECG	Fetal ECG
FNN	Fully Connected Networks
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
HF	Heart Failure
HT	Hemorrhagic Transformation
HVSMR	Heart & Vessel Segmentation from 3D MRI

ICD	International Classification of Diseases
IVUS	Intravascular Ultrasound
LSTM	Long-Short Term Memory
LV	Left Ventricle
MA	Microaneurysm
MFCC	Mel-Frequency Cepstral Coefficient
MICCAI	Medical Image Computing & Computer-Assisted Intervention
MI	Myocardial Infarction
MMWHS	Multi-Modality Whole Heart Segmentation Challenge
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
MRP	Magnetic Resonance Perfusion
OCT	Optical Coherence Tomography
PCA	Principal Component Analysis
PCG	Phonocardiogram
PPG	Pulsatile Photoplethysmography
RBM	Restricted Boltzmann Machine
RF	Random Forest
RNN	Recurrent Neural Network
ROI	Region of Interest
RV	Right Ventricle
SAE	Stacked Autoencoder
SATA	Segmentation Algorithms, Theory and Applications
SBP	Systolic Blood Pressure
SDAE	Stacked Denoised Autoencoder
SLO	Scanning Laser Ophthalmoscopy
SSAE	Stacked Sparse Autoencoder
STACOM	Statistical Atlases & Computational Modeling of the Heart
SVM	Support Vector Machine
VGG	Visual Geometry Group
WT	Wavelet Transform

# Κατάλογος Δημοσιεύσεων

## Άρθρα σε Επιστημονικά Περιοδικά

1. **Paschalis Bizopoulos**, Dimitris D.G Koutsouris, Sparsely Activated Networks, *Under Review/Considered for publication*, 2019.
2. **Paschalis Bizopoulos**, Dimitris D.G Koutsouris, Deep Learning in Cardiology, *IEEE Reviews in Biomedical Engineering*, 12:168–193, 2019.
3. Bourantas, Christos V and Papafaklis, Michail I and Lakkas, Lampros and Sakellarios, Antonis and Onuma, Yoshinobu and Zhang, Yao-Jun and Muramatsu, Takashi and Diletti, Roberto and **Bizopoulos, Paschalis** and Kalatzis, Fanis and Naka, Katerina and Fotiadis, Dimitrios and Wang, Jin and Garcia Garcia, Hector and Kimura, Takeshi and Michalis, Lampros and Serruys Patrick, Fusion of optical coherence tomographic and angiographic data for more accurate evaluation of the endothelial shear stress patterns and neointimal distribution after bioresorbable scaffold implantation: comparison with intravascular ultrasound-derived reconstructions, *The international journal of cardiovascular imaging*, 30, 3, 485–494, 2014, Springer Netherlands
4. Bourantas, Christos V and Papadopoulou, Stella-Lida and Serruys, Patrick W and Sakellarios, Antonis and Kitslaar, Pieter H and **Bizopoulos, Paschalis** and Girasis, Chrysafios and Zhang, Yao-Jun and de Vries, Ton and Boersma, Eric and Papafaklis, Michail and Naka, Katerina and Fotiadis, Dimitrios and Stone, Gregg and Reiber, Johan and Michalis, Lampros and Feyter, Pim and Garcia-Garcia, Hector, Noninvasive Prediction of Atherosclerotic Progression: The PROSPECT-MSCT Study, *JACC. Cardiovascular imaging*, 9, 8, 1009, 2016
5. Sakellarios, Antonis I and **Bizopoulos, Paschalis** and Papafaklis, Michail I and Athanasiou, Lambros and Exarchos, Themis and Bourantas, Christos V and Naka, Katerina K and Patterson, Andrew J and Young, Victoria EL and Gillard, Jonathan H and Parodi, Oberdan and Michalis, Lampros and

Fotiadis, Dimitrios, Natural history of carotid atherosclerosis in relation to the hemodynamic environment: a low-density lipoprotein transport modeling study with serial magnetic resonance imaging in humans, *Angiology*, 68, 2, 109–118, 2017, SAGE Publications Sage CA: Los Angeles, CA

## Ανακοινώσεις σε Συνέδρια

1. **Bizopoulos, Paschalis** and Koutsouris, Dimitrios D, Signal2Image Modules in Deep Neural Networks for EEG Classification, *Engineering in Medicine and Biology Society (EMBC), 2019 41th Annual International Conference of the IEEE, 2019, IEEE*
2. **Bizopoulos, Paschalis A** and Tsalikakis, Dimitrios G and Tzallas, Alexandros T and Koutsouris, Dimitrios D and Fotiadis, Dimitrios I, EEG epileptic seizure detection using k-means clustering and marginal spectrum based on ensemble empirical mode decomposition, *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on, 1–4, 2013, IEEE*
3. **Bizopoulos, Paschalis A** and Al-Ani, Tarik and Tsalikakis, Dimitrios G and Tzallas, Alexandros T and Koutsouris, Dimitrios D and Fotiadis, Dimitrios I, An automatic electroencephalography blinking artefact detection and removal method based on template matching and ensemble empirical mode decomposition, *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, 5853–5856, 2013, IEEE*
4. **Bizopoulos, Paschalis A** and Sakellarios, Antonis I and Koutsouris, Dimitrios D and Iliopoulou, Dimitra and Michalis, Lampros K and Fotiadis, Dimitrios I, Randomly generated realistic vessel geometry using spline interpolation and 2D Perlin noise, *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on, 157–160, 2014, IEEE*
5. **Bizopoulos, Paschalis A** and Sakellarios, Antonis I and Koutsouris, Dimitrios D and Kountouras, Jannis and Kostretzis, Lazaros and Karagergou, Stella and Michalis, Lampros K and Fotiadis, Dimitrios I, Prediction of atherosclerotic plaque evolution in carotids using features extracted from the arterial geometry, *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, 6556–6559, 2015, IEEE*
6. **Bizopoulos, Paschalis A** and Vavuranakis, Manolis and Papaioannou, Theodoros G and Vrachatis, Dimitrios A and Sakellarios, Antonis I and

- 
- Iliopoulou, Dimitra and Tousoulis, Dimitris and Koutsouris, Dimitrios D and Fotiadis, Dimitrios I, A Preliminary Study on In-Vivo 3-D Imaging of Bioprosthetic Aortic Valve Deformation, *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, 332–336, 2016, Springer Cham
7. **Bizopoulos, Paschalis** A and Sakellarios, Antonis and Michalis, Lampros K and Koutsouris, Dimitrios D and Fotiadis, Dimitrios I, 3-D Registration on Carotid Artery imaging data: MRI for different timesteps, *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 1159–1162, 2016, IEEE
  8. Sakellarios, Antonis I and **Bizopoulos, Paschalis** and Stefanou, Kostas and Athanasiou, Lambros S and Papafaklis, Michail I and Bourantas, Christos V and Naka, Katerina K and Michalis, Lampros K and Fotiadis, Dimitrios I, A proof-of-concept study for predicting the region of atherosclerotic plaque development based on plaque growth modeling in carotid arteries, *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 6552–6555, 2015, IEEE
  9. Salis, Christos I and Malissovass, Anastasios E and **Bizopoulos, Paschalis** A and Tzallas, Alexandros T and Angelidis, Pantelis A and Tsalikakis, Dimitrios G, Denoising simulated EEG signals: A comparative study of EMD, wavelet transform and kalman filter, *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, 1–4, 2013, IEEE

