



Ελληνικό Μεσογειακό Πανεπιστήμιο
Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών
Τμήμα Μουσικής Τεχνολογίας και Ακουστικής

**Επίλυση αντίστροφου προβλήματος εκτίμησης φυσικών
παραμέτρων δονούμενης μεμβράνης από τον παραγόμενο ήχο,
με χρήση νευρωνικών δικτύων βαθιάς μάθησης.**

Inverse physical parameter estimation of vibrating membranes
from sound using deep neural networks.

Διπλωματική Εργασία

Σταράκης Μιχάλης

Επιβλέποντες καθηγητές:

Αλεξανδράκη Χρυσούλα

Ζέρβας Παναγιώτης

Ρέθυμνο, Φεβρουάριος 2022

Ελληνικό Μεσογειακό Πανεπιστήμιο
Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών
Τμήμα Μουσικής Τεχνολογίας και Ακουστικής
74133, Ρέθυμνο, Κρήτη
Phone: +30 28310 21902
<https://www.hmu.gr/>
info@hmu.gr



Ελληνικό Μεσογειακό Πανεπιστήμιο

Σχολή Μουσικής και Οπτοακουστικών Τεχνολογιών

Τμήμα Μουσικής Τεχνολογίας και Ακουστικής

Επίλυση ανάστροφου προβλήματος εκτίμησης φυσικών παραμέτρων δονούμενης μεμβράνης (κεφαλή τυμπάνου) από τον παραγόμενο ήχο, με χρήση νευρωνικών δικτύων βαθιάς μάθησης.

Inverse physical parameter estimation of vibrating membranes (drum heads) from sound using deep neural networks.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σταράκης Μιχάλης, ΑΜ: mta10

Επιβλέποντες καθηγητές:

Αλεξανδράκη Χρυσούλα

Ζέρβας Παναγιώτης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 2022.

.....

Αλεξανδράκη Χρυσούλα, Επίκουρη Καθηγήτρια, ΕΛ.ΜΕ.ΠΑ.

.....

Ζέρβας Παναγιώτης, Επίκουρος Καθηγητής, Πανεπιστημίου Πελοποννήσου.

Ρέθυμνο, Φεβρουάριος 2022

.....

ΣΤΑΡΑΚΗΣ ΜΙΧΑΛΗΣ

Πτυχιούχος Μαθηματικός, Πανεπιστήμιο Κρήτης

Copyright © Σταράκης Μιχάλης, 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved. 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 ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Ελληνικού Μεσογειακού Πανεπιστημίου.

Ευχαριστίες

Ευχαριστώ τους επιβλέποντες καθηγητές μου κ. Χρυσούλα Αλεξανδράκη και κ. Παναγιώτη Ζέρβα, για την εμπιστοσύνη που μου έδειξαν και τη στήριξη τους καθόλη τη διάρκεια της συνεργασίας μας. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή κ. Rolf Bader για την κρίσιμη και πολύτιμη συνεισφορά του και τέλος, ένα μεγάλο ευχαριστώ, στην οικογένειά μου, και κυρίως στη Μαρία, που με στήριξε στις επιλογές μου, σε κάθε εύκολη ή δύσκολη στιγμή.

Στους γονείς μου Νίκο και Μαρία.

Περίληψη

Τα αντίστροφα προβλήματα, στο πεδίο της ακουστικής, μπορούν να αποτελέσουν πολύτιμα εργαλεία για την ανάκτηση πληροφοριών μέσω του παραγόμενου ήχου. Παράλληλα, η ταχεία ανάπτυξη των τεχνικών βαθιάς μηχανικής μάθησης και οι αξιοσημείωτες επιδόσεις των τεχνητών νευρωνικών δικτύων, σε μια πληθώρα εφαρμογών, υποδηλώνουν την χρησιμότητά τους σε διάφορους επιστημονικούς τομείς. Η δημιουργία δεδομένων μέσω της επίλυσης φυσικών μοντέλων και η χρησιμοποίηση αυτών ως δεδομένα εισόδου για την εκπαίδευση νευρωνικών δικτύων, με σκοπό την εκτίμηση των φυσικών παραμέτρων του μοντέλου, αποτελεί μια πολλά υποσχόμενη μεθοδολογία. Στο πλαίσιο αυτό, η παρούσα εργασία έχει σκοπό τη μελέτη και την υλοποίηση μια νέας προσέγγισης για την επίλυση ενός αντίστροφου προβλήματος, που αφορά στην εκτίμηση φυσικών παραμέτρων του ακουστικού μοντέλου μιας δονούμενης μεμβράνης, όταν είναι γνωστός ο ήχος που παράγει. Η εκτίμηση αυτών των παραμέτρων επιχειρείται με τη χρήση νευρωνικών δικτύων βαθιάς μάθησης. Για την επίτευξη αυτού του στόχου, δημιουργήθηκε μια βάση δεδομένων, η οποία περιλαμβάνει ηχητικά δείγματα παραγόμενα από μεμβράνες, καθώς και τις παραμέτρους που καθορίζουν τα κατασκευαστικά και γεωμετρικά χαρακτηριστικά αυτών. Η δημιουργία των ηχητικών δειγμάτων βασίστηκε στην επίλυση της κυματικής εξίσωσης της μεμβράνης με τη μέθοδο των Πεπερασμένων Διαφορών Χρονικού Τομέα, επιλύοντας έτσι το ευθύ πρόβλημα. Για την αντιμετώπιση του αντίστροφου προβλήματος, τα ηχητικά δείγματα χρησιμοποιήθηκαν για την εκπαίδευση ενός Συνελκτικού Νευρωνικού Δικτύου, μέσω του οποίου έγινε προσπάθεια εκτίμησης των χαρακτηριστικών των μεμβρανών που παράγουν τους συγκεκριμένους ήχους. Καθώς διαφορετικοί συνδυασμοί φυσικών παραμέτρων της μεμβράνης μπορούν να παράγουν παρόμοιους ήχους, το συνελκτικό μοντέλο δεν μπόρεσε να εκπαιδευτεί επαρκώς. Το γεγονός αυτό οδήγησε στον επαναπροσδιορισμό του αντίστροφου προβλήματος, θέτοντας ως νέους στόχους, την εκτίμηση, με βάση τον παραγόμενο ήχο, του μοτίβου (σχήματος) και της πρόσθετης μάζας, αποσβεστικού υλικού που μπορεί να εφαρμοστεί στην επιφάνεια της μεμβράνης, ώστε να χρησιμοποιηθεί για το κούρδισμά της. Η πειραματική αξιολόγηση του μοντέλου τεκμηριώνει τη δυναμική της μεθοδολογίας που ακολουθήθηκε, καθώς το μοντέλο παρείχε πολύ ενθαρρυντικά αποτελέσματα φτάνοντας σε ακρίβεια 97% για την ταξινόμηση των μοτίβων και βαθμολογία r_square 83% για την εκτίμηση της πρόσθετης μάζας.

Λέξεις Κλειδιά

μεμβράνες, framedrum, μέθοδος πεπερασμένων διαφορών χρονικού τομέα, εκτίμηση φυσικών παραμέτρων, αντίστροφο πρόβλημα, βαθιά μηχανική μάθηση, συνελκτικά νευρωνικά δίκτυα

Abstract

Inverse problems in the field of acoustics can be valuable tools for retrieving information through the produced sound. At the same time, the rapid development of deep machine learning techniques and the remarkable performance of artificial neural networks, in a variety of applications, suggest their usefulness in various scientific fields. Generating data by solving physical models and using them as input data for the training of artificial neural networks, in order to estimate the physical parameters of the model, is a very promising methodology. In this context, the present work aims at studying and implementing a new approach to solving an inverse problem, i.e. the estimation of physical parameters of the acoustic model of a vibrating membrane, from the sound they generate. The estimation of these parameters is attempted using deep neural networks. To achieve this goal, a database was created, which consists of sound samples produced from membranes, associated with parameters defining their structural and geometric characteristics. The sound samples were generated by a numerical solution of the wave equation of the membrane, using Finite Time Domain Differences method, thus solving the direct problem. To address the inverse problem, the audio samples were used to train a Convolutional Neural Network, through which an attempt was made to estimate the characteristics of the membranes that produce these sounds. However, it was soon found out that as different combinations of physical membrane parameters can produce highly similar sounds, it was not possible to efficiently train the convolutional model. This fact required a redefinition of the inverse problem, so that instead of structural and geometric characteristics the convolutional model would estimate patterns (shapes) and additional mass of damping material that can be applied to the membrane surface for tuning purposes. The evaluation of the model verified the validity of the methodology that was followed, as the model provided very encouraging results, reaching an accuracy of 97% for the patterns classification and a r_square score of 83% for the estimation of the additional mass.

Keywords

membranes, framedrum, finite differences time domain method, physical parameters estimation, inverse problem, deep machine learning, convolutional neural networks

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1	Εισαγωγή.....	1
1.1	Αντικείμενο διπλωματικής & ορισμός προβλήματος.....	2
1.2	Σκοπός & συνεισφορά εργασίας.....	4
1.3	Σχετική έρευνα.....	5
1.4	Οργάνωση κειμένου.....	6
	Μέρος Α' Θεωρητικό υπόβαθρο.....	8
2	Μεμβράνες & Μεμβρανόφωνα.....	9
2.1	Κατηγορίες μεμβρανόφωνων.....	9
2.2	Χαρακτηριστικά μεμβρανόφωνων.....	10
2.3	Frame drums, τύμπανα & μπεντίρ.....	10
2.4	Ήχος από μεμβράνες, φυσική περιγραφή.....	11
2.5	Μαθηματική περιγραφή δόνησης μεμβρανόφωνων.....	13
2.5.1	Κυματική εξίσωση τετραγωνικής μεμβράνης.....	13
2.5.2	Κυματική εξίσωση κυκλικής μεμβράνης.....	17
2.5.3	Ανάλυση ιδιοσυχνοτήτων & ιδιομορφών.....	21
2.6	Πραγματικές συνθήκες, φορτίο αέρα & κέλυφος.....	21
2.7	Κατασκευαστικά χαρακτηριστικά κεφαλών τυμπάνων.....	23
2.8	Κούρδισμα κεφαλών τυμπάνων.....	25
2.8.1	Τρόποι κουρδίσματος & έλεγχος απόσβεσης.....	25
2.8.2	Μηχανισμοί απόσβεσης.....	27
3	Αριθμητικές Μέθοδοι & Μοντελοποίηση.....	29
3.1	Επικρατέστερες οικογένειες υπολογιστικών μεθόδων.....	30
3.1.1	Μέθοδος Πεπερασμένων Διαφορών (FDM).....	31
3.1.2	Μέθοδος Πεπερασμένων Διαφορών Χρονικού Τομέα (FDTD).....	32
3.1.3	Μέθοδος Πεπερασμένων Στοιχείων (FEM).....	35
3.1.4	Μέθοδος Συνοριακών Στοιχείων (BEM).....	36
3.1.5	Ιδιότητες & Σφάλματα μεθόδων.....	38
3.2	Μέθοδος πεπερασμένων διαφορών για την κυματική εξίσωση.....	40
3.2.1	Προσέγγιση παραγώγων με σχήματα πεπερασμένων διαφορών.....	40
3.2.2	Ρητές & πεπλεγμένες μέθοδοι επίλυσης.....	42
3.2.3	Διατύπωση πεπερασμένων διαφορών μέσω πίνακα.....	43
3.2.4	Δημιουργία αριθμητικού πλέγματος & διακριτοποίηση.....	44
3.3	Μοντελοποίηση & εφαρμογή FDTD μεθόδου.....	48
3.3.1	Κυματική εξίσωση δονούμενης μεμβράνης.....	48

3.3.2	Διακριτοποίηση κυματικής εξίσωσης & περασμένες διαφορές	51
3.3.3	Προσέγγιση κυματικής εξίσωσης μέσω της μεθόδου FDTD	53
3.3.4	Περιγραφή NSV-leapfrog αλγόριθμου	56
3.3.5	Ιξωδοελαστικό μοντέλο μεμβράνης	57
4	Μηχανική Μάθηση & Τεχνητά Νευρωνικά Δίκτυα.....	60
4.1	Μηχανική Μάθηση.....	60
4.1.1	Κατηγορίες Μηχανικής Μάθησης.....	61
4.1.2	Τύποι αλγόριθμων Μηχανικής Μάθησης.....	63
4.1.3	Βαθιά Μηχανική Μάθηση.....	65
4.1.4	Ανάγκη για Deep Networks & πλεονεκτήματα.....	68
4.2	Τεχνητά Νευρωνικά Δίκτυα	69
4.2.1	Βιολογικοί νευρώνες.....	69
4.2.2	Τα μοντέλα McCulloch-Pitts & Perceptron.....	70
4.2.3	Μαθηματικό μοντέλο τεχνητού νευρώνα.....	71
4.3	Αρχιτεκτονική ΤΝΔ.....	72
4.3.1	Περιγραφή δομής & λειτουργίας ΤΝΔ.....	72
4.3.2	Πρόσθιας & οπίσθιας τροφοδότησης ΤΝΔ	73
4.3.3	Πολυεπίπεδα ΤΝΔ	74
4.3.4	Συναρτήσεις Ενεργοποίησης	75
4.4	Εκπαίδευση ΤΝΔ.....	78
4.4.1	Αλγόριθμοι εκπαίδευσης	78
4.4.2	Αλγόριθμος Back-Propagation	80
4.4.3	Συναρτήσεις Απώλειας.....	80
4.4.4	Αλγόριθμοι βελτιστοποίησης	84
4.4.5	Εκπαίδευση ΤΝΔ με Back-Propagation	86
4.4.6	Μέτρα αξιολόγησης μοντέλων	87
4.4.7	Υπερ-εκπαίδευση και Υπο-εκπαίδευση.....	88
4.4.8	Απόρριψη.....	90
4.4.9	Βελτιστοποίηση Υπερπαραμέτρων.....	91
4.5	Συνελκτικά νευρωνικά δίκτυα (CNN).....	92
4.5.1	Αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων.....	93
4.5.2	Συνελκτικό Επίπεδο.....	93
4.5.3	Επίπεδο συγκέντρωσης.....	97
4.5.4	Πλήρως συνδεδεμένο επίπεδο	98
4.5.5	Διαστάσεις εξόδου & πλήθος παραμέτρων	99
4.5.6	Συνελκτικά δίκτυα σε εφαρμογές ήχου	100

Μέρος Β' Πειραματική διαδικασία.....	102
5 Υλοποίηση ευθέως προβλήματος	103
5.1 Αρχικός Πειραματισμός.....	103
5.2 Επαναπροσδιορισμός στόχου	106
5.2.1 Προγραμματιστικά εργαλεία για την παραγωγή των αρχείων ήχου.....	107
5.2.2 Αρχιτεκτονική CUDA	108
5.3 Παραγωγή ηχητικών δειγμάτων	109
5.4 Φυσικές παράμετροι & παράμετροι μοτίβων πάστας.....	111
5.5 Ανάπτυξη συνόλου δεδομένων (dataset)	117
6 Υλοποίηση αντίστροφου προβλήματος	119
6.1 Προγραμματιστικά εργαλεία για την ανάπτυξη των CNN μοντέλων.....	119
6.2 Προετοιμασία δεδομένων	123
6.2.1 Αναπαραστάσεις Εισόδου	123
6.2.2 Δημιουργία συνόλων εκπαίδευσης & επικύρωσης.....	125
6.3 Ανάπτυξη CNN μοντέλου.....	128
6.3.1 Καθορισμός βασικής αρχιτεκτονικής	129
6.3.2 Βελτιστοποίηση υπερπαραμέτρων	130
6.4 Εκπαίδευση CNN μοντέλου	134
6.5 Αξιολόγηση μοντέλου & Αποτελέσματα.....	137
6.5.1 Αποτελέσματα ταξινόμησης μοτίβων πάστας (classification).....	137
6.5.2 Αποτελέσματα εκτίμησης πρόσθετης μάζας (regression)	140
6.5.3 Συνολική αξιολόγηση αποτελεσμάτων μοντέλου (classification & regression)	143
7 Συμπεράσματα	146
7.1 Σύνοψη & Συμπεράσματα	146
7.2 Μελλοντικές επεκτάσεις & Εφαρμογές.....	147
Αναφορές	149

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 2-1: Αρχαίο Ελληνικό τύμπανο. Μπεντίρ. Διάφορα είδη frame drums.	11
Εικόνα 2-2: Απειροστό τμήμα μεμβράνης.	13
Εικόνα 2-3: Οι αρχικοί (12) κανονικοί τρόποι δόνησης μιας ορθογώνιας μεμβράνης.	16
Εικόνα 2-4: Απειροστό τμήμα όγκου σε κυλινδρικές συντεταγμένες.	17
Εικόνα 2-5: Συναρτήσεις Bessel πρώτου είδους.	18
Εικόνα 2-6: Οι αρχικοί (11) κανονικοί τρόποι δόνησης μιας κυκλικής μεμβράνης.	19
Εικόνα 2-7: Κανονικοί τρόποι δόνησης κυκλικής μεμβράνης (πακτωμένης στα άκρα)	19
Εικόνα 2-8: Οι πρώτοι 12 τρόποι δόνησης κυκλικής μεμβράνης σε 2 διαστάσεις.	20
Εικόνα 2-9: Συζευγμένες μεμβράνες. Ζεύγη κανονικών τρόπων δόνησης (0,1) και (1,1). .	23
Εικόνα 2-10: Εμπορικά προϊόντα για έλεγχο απόσβεσης και κουρδίσματος.	26
Εικόνα 2-11: Τυπική τοποθέτηση τζελ.	28
Εικόνα 2-12: Τυπική τοποθέτηση δακτυλίων.	28
Εικόνα 3-1: Στάδια αριθμητικής επίλυσης.	29
Εικόνα 3-2: Σχηματική απεικόνιση χωρικής διακριτοποίησης αριθμητικών μεθόδων.	30
Εικόνα 3-3: Το πλέγμα του αλγόριθμου Yee σε τρεις διαστάσεις.	34
Εικόνα 3-4: Διάφοροι τρόποι διακριτοποίησης του συνόρου μιας επιφάνειας.	38
Εικόνα 3-5: Σφάλματα διακριτοποίησης (ή θέσης).	39
Εικόνα 3-6: Γραφική απεικόνιση πρόδρομης και ανάδρομης προσέγγισης της παραγώγου σε ένα σημείο A.	41
Εικόνα 3-7: Ρητές μέθοδοι (αριστερά) - Πεπλεγμένες (δεξιά)	43
Εικόνα 3-8: Αριθμητικό πλέγμα - δίκτυο (2D) σημείων στο χώρο.	45
Εικόνα 3-9: Ομοιόμορφο τετραγωνικό δίκτυο κόμβων.	52
Εικόνα 4-1: Διαφορετικές αναπαραστάσεις δεδομένων (καρτεσιανές - πολικές)	63
Εικόνα 4-2: Έννοιες που εμπεριέχονται στον κλάδο της τεχνητής νοημοσύνης.	65
Εικόνα 4-3: Βασικά βήματα για τη δημιουργία ενός μοντέλου βαθιάς μάθησης.	67
Εικόνα 4-4: Μέρη βιολογικού νευρώνα.	69
Εικόνα 4-5: Αρχιτεκτονική του μοντέλου Perceptron.	71
Εικόνα 4-6: Μαθηματικό μοντέλο του νευρώνα.	72
Εικόνα 4-7: Παράδειγμα ΤΝΔ πρόσθιας τροφοδότησης και οπίσθιας τροφοδότησης.	74
Εικόνα 4-8: Πολυεπίπεδο ΤΝΔ ενός κρυφού επιπέδου πρόσθιας τροφοδότησης 2 κρυμμένων επιπέδων με πλήρως συνδεδεμένους κόμβους.	74
Εικόνα 4-9: Παραδείγματα βασικών συναρτήσεων ενεργοποίησης ΤΝΔ.	77
Εικόνα 4-10: Στάδια ολοκλήρωσης ΤΝΔ.	78
Εικόνα 4-11: Πολύ μικρός ρυθμός μάθησης → πολλές επαναλήψεις (αριστερά), Πολύ μεγάλος μικρός ρυθμός μάθησης → Δε συγκλίνει στο ελάχιστο (δεξιά).	84

Εικόνα 4-12: Εφαρμογή της μεθόδου κατάβασης κλίσης σε μία συνάρτηση απώλειας.....	85
Εικόνα 4-13: Λειτουργία Backpropagation αλγορίθμου.....	86
Εικόνα 4-14: Αναπαραστάσεις αποτελεσμάτων για μοντέλα με Underfitting και Overfitting.	89
Εικόνα 4-15: Νευρωνικό δίκτυο χωρίς και με εφαρμογή Dropout.	91
Εικόνα 4-16: Εικόνα RGB ως τριδιάστατος πίνακας, 4x4 εικονοστοιχείων, με 3 τιμές χρωμάτων για κάθε ένα (αριστερά). Εικόνα σε Grayscale (δεξιά).	93
Εικόνα 4-17: Διαδικασία Συνέλιξης - Σάρωση φίλτρου.	94
Εικόνα 4-18: Διαδικασία συνέλιξης & χάρτης χαρακτηριστικών.....	95
Εικόνα 4-19: Διαφορετικά είδη φίλτρων (πυρήνων) και εξαγωγή διαφορετικών χαρακτηριστικών ανάλογα με τις τιμές του φίλτρου.	95
Εικόνα 4-20: Συνέλιξη και διαστάσεις χάρτη χαρακτηριστικών.	95
Εικόνα 4-21: Εφαρμογή 128 φίλτρων 3x3x3 σε είσοδο 7x7x3 (3D).....	96
Εικόνα 4-22: Same padding: Ο πίνακας 5x5x1 που αντιστοιχεί στην εικόνα εισόδου, συμπληρώνεται με 0 για να δημιουργηθεί ένας πίνακας 6x6x1.	96
Εικόνα 4-23: Ένας τρόπος απεικόνισης των χαρακτηριστικών που αναζητούν τα διαφορετικά κρυφά επίπεδα σε ένα νευρωνικό δίκτυο.	97
Εικόνα 4-24: Max Pooling με φίλτρο 2x2 και βήμα 2 πάνω σε ένα 4x4 convolved feature (αριστερά). Το επίπεδο Flattening (δεξιά).	98
Εικόνα 4-25: Αρχιτεκτονική CNN για ταξινόμηση χειρόγραφων ψηφίων. Αναγράφονται οι διαστάσεις των πινάκων εξόδου και το πλήθος των παραμέτρων σε κάθε επίπεδο.	99
Εικόνα 5-1: Κυματομορφές και spectrograms δυο δειγμάτων, που ενώ προέρχονται από μεμβράνες με διαφορετικό συνδυασμό φυσικών παραμέτρων, ακούγονται το ίδιο.	105
Εικόνα 5-2: Τα επαναληπτικά βήματα του NSV-leapfrog αλγόριθμου.....	109
Εικόνα 5-3: Ομοιόμορφο, τετραγωνικό πλέγμα $n_x * n_y = 104 * 104 = 10816$ κελιά.	110
Εικόνα 5-4: Οι 7 περιπτώσεις σχημάτων εφαρμογής αποσβεστικού υλικού.	112
Εικόνα 5-5: Κατανομές τιμών κάθε παραμέτρου για τις περιπτώσεις των μοτίβων πάστας.	116
Εικόνα 6-1: Η αρχιτεκτονική και οι διαστάσεις κάθε επιπέδου του αποδοτικότερου CNN μοντέλου, όπως αυτό καθορίστηκε από το KerasTuner, σε μορφή δένδρου.....	133
Εικόνα 6-2: Η αρχιτεκτονική και ο αριθμός των μονάδων και το μέγεθος των φίλτρων κάθε επιπέδου του τελικού CNN μοντέλου, σε μορφή διαγράμματος.....	133
Εικόνα 6-3: Ιστορικό εκπαίδευσης CNN μοντέλου. Άθροισμα απωλειών ταξινόμησης και παλινδρόμησης (πάνω). Ακρίβεια (accuracy) για την ταξινόμηση (κέντρο). R-squared για την παλινδρόμηση (κάτω).....	136
Εικόνα 6-4: Πίνακας Σύγχυσης (Confusion matrix) των αποτελεσμάτων του μοντέλου για την ταξινόμηση μοτίβων.....	138

Εικόνα 6-5: Γραφήματα διασποράς με προσαρμογή πολυωνυμικής παλινδρόμησης, της ταξινόμηση μοτίβων για το σύνολο εκπαίδευσης και το σύνολο επικύρωσης.	139
Εικόνα 6-6: Προβλέψεις CNN μοντέλου για την ταξινόμηση μοτίβων πάστας.	140
Εικόνα 6-7: Γραφήματα διασποράς με προσαρμογή πολυωνυμικής παλινδρόμησης, της εκτίμησης της πρόσθετης μάζας για το σύνολο εκπαίδευσης (αριστερά) και το σύνολο επικύρωσης (δεξιά).	141
Εικόνα 6-8: Προβλέψεις CNN μοντέλου για (α) ταξινόμηση μοτίβων πάστας και (β) εκτίμηση πρόσθετης μάζας. Μπλε κουκκίδες: πραγματικές τιμές - κόκκινες κουκκίδες: προβλέψεις (για το σύνολο επικύρωσης).	142

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 2-1: Οι 12 πρώτες συχνότητες συντονισμού ως πολλαπλάσια της θεμελιώδους για μια ιδανική κυκλική μεμβράνη.	20
Πίνακας 4-1: Διαφορές Μηχανικής Μάθησης και Βαθιάς Μηχανικής Μάθησης.	66
Πίνακας 5-1: Φυσικές παράμετροι του μοντέλου της δονούμενης μεμβράνης.	104
Πίνακας 5-2: Οι σταθερές φυσικές παράμετροι του μοντέλου δονούμενης μεμβράνης, για τη δημιουργία των ηχητικών δειγμάτων.	111
Πίνακας 5-3: Παράμετροι που χρησιμοποιήθηκαν για το κάθε μοτίβο πάστας.	114
Πίνακας 5-4: Τα σύνολα τιμών των παραμέτρων για τα μοτίβα πάστας, που χρησιμοποιήθηκαν στην παραγωγή των ηχητικών δειγμάτων.	115
Πίνακας 5-5: Τιμές πλάτους αρχείων ήχου και τιμές φυσικών παραμέτρων μεμβράνης. ..	117
Πίνακας 5-6: Τιμές παραμέτρων μοτίβων πάστας.	118
Πίνακας 6-1: Συγκεντρωτικός πίνακας προγραμματιστικών εργαλείων.	122
Πίνακας 6-2: (α) Σύνολα δεδομένων εισόδου, (β) σύνολο δεδομένων στόχου για τα μοτίβα πάστας, (γ) σύνολο δεδομένων στόχου για την πρόσθετη μάζα από την εφαρμογή πάστας.	126
Πίνακας 6-3: Τα σύνολα δεδομένων εκπαίδευσης και επικύρωσης για τα μοτίβα πάστας (pat) και για την πρόσθετη μάζα (mass).	127
Πίνακας 6-4: (α) Δομή του μοντέλου με την μικρότερη απώλεια επικύρωσης. Επίπεδα και αριθμός παραμέτρων. (β) Βέλτιστες τιμές παραμέτρων για το αποδοτικότερο μοντέλο. Βέλτιστο μέγεθος μονάδων, φίλτρων και άλλων παραμέτρων για κάθε επίπεδο.	132
Πίνακας 6-5: Αρχικός και τελικός κύκλος εκπαίδευσης.	135
Πίνακας 6-6: Αποτελέσματα αξιολόγησης για την ταξινόμηση των μοτίβων πάστας (classification task).	137
Πίνακας 6-7: Αποτελέσματα αξιολόγησης για την το σύνολο εκπαίδευσης.	138
Πίνακας 6-8: Αποτελέσματα αξιολόγησης για την εκτίμηση της πρόσθετης μάζας λόγω προσθήκης πάστας (regression task).	141
Πίνακας 6-9: Αποτελέσματα αξιολόγησης για την το σύνολο εκπαίδευσης.	141
Πίνακας 6-10: Ενδεικτικά παραδείγματα προβλέψεων από το σύνολο επικύρωσης, για την ταξινόμηση των μοτίβων πάστας και για την εκτίμηση της πρόσθετης μάζας.	144
Πίνακας 6-11: Αποτελέσματα συνολικής αποτίμησης (classification & regression) στο σύνολο επικύρωσης.	145

1

Εισαγωγή

Ο όρος τεχνητή νοημοσύνη αναφέρεται στον κλάδο της πληροφορικής ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων, τα οποία «μιμούνται» στοιχεία της ανθρώπινης συμπεριφοράς. Τα συστήματα αυτά λαμβάνουν δεδομένα, τα επεξεργάζονται και ανταποκρίνονται βάσει αυτών. Με αυτό τον τρόπο γίνονται ικανά να «κατανοούν» το περιβάλλον τους, να προσαρμόζουν τη συμπεριφορά τους αναλύοντας τις συνέπειες προηγούμενων ενεργειών και να επιλύουν προβλήματα ενεργώντας προς την επίτευξη ενός συγκεκριμένου στόχου.

Η μηχανική μάθηση, προέρχεται από το ευρύ πεδίο της τεχνητής νοημοσύνης και χρησιμοποιεί μεθόδους στατιστικής ανάλυσης, με σκοπό τον προγραμματισμό τέτοιων συστημάτων. Η αλματώδης ανάπτυξη της τεχνολογίας τα τελευταία χρόνια, αλλά και η αφθονία δεδομένων που προσφέρει το διαδίκτυο, έφερε σημαντικές εξελίξεις στο πεδίο της μηχανικής μάθησης και ιδιαίτερα στον κλάδο της βαθιάς μηχανικής μάθησης και των τεχνητών νευρωνικών δικτύων. Τα δίκτυα βαθιάς μάθησης μπορούν να προσαρμόζονται σε καινούργια προβλήματα με ελάχιστη ανθρώπινη επέμβαση και έχουν βρει εφαρμογή σε πλήθος προβλημάτων διαφόρων επιστημονικών πεδίων, όπως στην εκτίμηση των παραμέτρων μαθηματικών μοντέλων που περιγραφούν ένα φυσικό φαινόμενο.

Ένα μαθηματικό μοντέλο αποτελεί την περιγραφή ενός φαινομένου, ενός συστήματος ή μιας διαδικασίας χρησιμοποιώντας μαθηματικούς όρους. Οι στόχοι της μαθηματικής μοντελοποίησης ενός συγκεκριμένου συστήματος (φαινομένου), όπως για παράδειγμα μιας δονούμενης μεμβράνης, είναι η μελέτη και η κατανόηση της συμπεριφοράς πολύπλοκων συστημάτων, η χρήση και η ανάπτυξη νέων εργαλείων για τη επίλυσή τους και η πρόβλεψη των συμπεριφορών και των ιδιοτήτων τους. Ένα μαθηματικό μοντέλο μπορεί να εκφραστεί είτε ως συνεχές μοντέλο, μέσω διαφορικών εξισώσεων, είτε ως διακριτό, μέσω ενός συστήματος αλγεβρικών εξισώσεων και να επιλυθεί μέσω αριθμητικών μεθόδων.

Ως ευθύ πρόβλημα ονομάζεται εκείνο στο οποίο ζητείται ο προσδιορισμός ενός μεγέθους, συνήθως με τη μορφή μιας συνάρτησης, όταν είναι γνωστές τόσο οι παράμετροι που διέπουν το πρόβλημα, όσο και οι εξισώσεις που περιγράφουν τη συμπεριφορά του. Στο αντίστροφο πρόβλημα ζητείται συνήθως εκτίμηση των παραμέτρων ενός φυσικού φαινομένου ή ενός μαθηματικού προβλήματος, όταν είναι γνωστή μόνο η συμπεριφορά του μοντέλου. Σε αυτή την εργασία εξετάζεται το κατά πόσο είναι δυνατή η εκτίμηση των φυσικών παραμέτρων ενός μαθηματικού μοντέλου δονούμενης μεμβράνης, χρησιμοποιώντας τεχνικές μηχανικής μάθησης (αντίστροφο πρόβλημα), αφού πρώτα έχουν προσδιοριστεί και περιγραφεί οι παράμετροι που διέπουν το μοντέλο αυτό (ευθύ πρόβλημα).

1.1 Αντικείμενο διπλωματικής & ορισμός προβλήματος

Η παρούσα διπλωματική εργασία διενεργήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών με τίτλο «Τεχνολογίες Ήχου και Μουσικής» του Τμήματος Μουσικής Τεχνολογίας και Ακουστικής του Ελληνικού Μεσογειακού Πανεπιστημίου. Το γνωστικό του αντικείμενο εστιάζει στις σύγχρονες επιστημονικές τάσεις που αφορούν στον ήχο και πώς αυτός αξιοποιείται ως κύμα, ως σήμα και ως φορέας πληροφορίας και περιλαμβάνει στοιχεία από διάφορα επιστημονικά πεδία, όπως της Ακουστικής, της Επεξεργασίας Σήματος και της Εφαρμοσμένης Μηχανικής Μάθησης.

Κύριο ερευνητικό αντικείμενο της εργασίας αποτελεί η εκτίμηση των φυσικών παραμέτρων ενός αντικείμενου, με βάση τον ήχο που παράγει όταν διεγείρεται. Πιο συγκεκριμένα, το βασικό κίνητρο αυτής της έρευνας είναι να δοθεί απάντηση στο ερώτημα: *«Είναι δυνατόν να εκτιμηθούν τα φυσικά χαρακτηριστικά μιας δονούμενης μεμβράνης, για παράδειγμα μιας κεφαλής ενός τυμπάνου, με βάση τον ήχο που παράγει;»*.

Το παραπάνω ερώτημα εμπεριέχει πολλές παραμέτρους που πρέπει να διερευνηθούν και ανήκουν σε διάφορα επιστημονικά πεδία, όπως στο πεδίο της Ακουστικής, της Αριθμητικής Ανάλυσης και της Τεχνητής Νοημοσύνης. Αυτές οι παράμετροι περιλαμβάνουν τον τρόπο που πραγματοποιείται η μαθηματική μοντελοποίηση του φυσικού μοντέλου της δονούμενης μεμβράνης, τις αριθμητικές μεθόδους που θα επιλεγθούν για την επίλυση της κυματικής εξίσωσης, και τέλος, την εφαρμογή τεχνικών μηχανικής μάθησης, όπως τη χρήση νευρωνικών δικτύων βαθιάς μάθησης.

Προς το παρόν, τέτοια προβλήματα, μετά τη μοντελοποίησή τους, διερευνώνται μέσω μηχανικής μάθησης και βαθέων νευρωνικών δικτύων. Αυτές οι προσεγγίσεις απαιτούν

συνήθως πολλά δεδομένα για την εκπαίδευση των μαθηματικών μοντέλων, προκειμένου να γίνουν εκτιμήσεις με βάση διάφορα στατιστικά μέτρα.

Επομένως, το βασικό πρόβλημα της παρούσας εργασίας, διαχωρίστηκε σε δύο μέρη, στο ευθύ πρόβλημα και στο αντίστροφο πρόβλημα:

- Το ευθύ πρόβλημα (direct problem) αναφέρεται στη δημιουργία (τεχνητών) δειγμάτων ήχου, χρησιμοποιώντας επιλεγμένες (γνωστές) φυσικές παραμέτρους, για το μοντέλο της δονούμενης μεμβράνης.
- Το αντίστροφο πρόβλημα (inverse problem) αναφέρεται στην εκτίμηση αυτών των φυσικών παραμέτρων με βάση τα ηχητικά δείγματα της μεμβράνης, χρησιμοποιώντας βαθιά νευρωνικά δίκτυα.

Ο κύριος λόγος για την επιλογή των μεμβρανών (κεφαλών frame drums), αντί για οποιοδήποτε άλλο δονούμενο αντικείμενο, είναι το γεγονός ότι οι μεμβράνες αντιπροσωπεύονται από ένα σχετικά απλό μοντέλο, το οποίο μάλιστα επιδέχεται περαιτέρω απλοποίηση χωρίς βλάβη της γενικότητάς του. Η μελέτη του μοντέλου είναι απλούστερη συγκριτικά με άλλα δονούμενα συστήματα καθώς, οι φυσικές παράμετροι που το περιγράφουν έχουν μικρό πλήθος.

Ένας δεύτερος λόγος για την επιλογή του συγκεκριμένου μοντέλου ήταν η συνεργασία που έχει το Τμήμα με τον Καθηγητή Rolf Bader¹ από το Πανεπιστήμιο του Αμβούργου, ο οποίος έχει διεξάγει εμπειριστατωμένη έρευνα στις δονήσεις μεμβρανόφωνων [1] [2][3]. Ο Καθηγητής Rolf Bader διέθεσε για τους σκοπούς της εργασίας αυτής, την υλοποίηση ενός προγράμματος, το οποίο υπολογίζει τον ήχο που παράγεται από μία μεμβράνη, με χρήση της μεθόδου Πεπερασμένων Διαφορών Χρονικού Τομέα (Finite Difference Time Domain - FDTD), η οποία αναλύεται στην ενότητα 3.1.2. Η χρησιμοποίηση αυτού του προγράμματος κατέστησε εφικτή τη δημιουργία μεγάλου αριθμού δειγμάτων ήχου, που αντιστοιχούν σε γνωστές φυσικές παραμέτρους δονούμενων μεμβρανών, διευκολύνοντας έτσι την επίλυση του ευθέως προβλήματος.

Για την επίλυση του αντίστροφου προβλήματος, επιλέχθηκαν τα τεχνητά νευρωνικά δίκτυα, και συγκεκριμένα τα βαθιά συνελκτικά δίκτυα [4 - 8]. Η επιλογή αυτή έγινε διότι τα δίκτυα αυτά, παρουσιάζουν σημαντικά πλεονεκτήματα σε σχέση με άλλες

¹ Ο καθηγητής Rolf Bader σπούδασε Συστηματική Μουσικολογία, Φυσική, Εθνολογία και Ιστορική Μουσικολογία στο Πανεπιστήμιο του Αμβούργου, όπου απέκτησε το διδακτορικό του. Είναι Καθηγητής Συστηματικής Μουσικολογίας στο Ινστιτούτο Συστηματικής Μουσικολογίας (Systematic Musicology at the Institute of Systematic Musicology) του Πανεπιστημίου του Αμβούργου από το 2007. Προσωπική σελίδα: <https://rolfbader.de/>.

προσεγγίσεις, όπως για παράδειγμα το γεγονός ότι βελτιώνουν την απόδοσή τους καθώς τροφοδοτούνται με περισσότερα δεδομένα. Επιπλέον, μπορούν να δημιουργούν πρόσθετα χαρακτηριστικά, χωρίς ανθρώπινη παρέμβαση, ακόμα και σε μη δομημένα δεδομένα, εντοπίζοντας πολύπλοκες μη γραμμικές σχέσεις μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών.

Κατά τη διάρκεια των πειραματικών διαδικασιών, προέκυψε ένα σημαντικό εμπόδιο για την επίλυση του αντίστροφου προβλήματος, λόγω του γεγονότος ότι, διαφορετικών παραμέτρων μεμβράνες μπορεί να παράγουν σχεδόν ίδιους ήχους. Το γεγονός αυτό οδήγησε στον επαναπροσδιορισμό του αντίστροφου προβλήματος και στον ορισμό νέων στόχων. Ως νέοι στόχοι ορίστηκαν η ταξινόμηση του μοτίβου (σχήματος) και η εκτίμηση της πρόσθετης μάζας ενός υλικού απόσβεσης, όταν αυτό τοποθετηθεί στην επιφάνεια της μεμβράνης, με σκοπό να χρησιμοποιηθεί για το κούρδισμα της.

1.2 Σκοπός & συνεισφορά εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η χρήση τεχνικών βαθιάς μάθησης (Deep Learning), για την επίλυση ενός αντίστροφου προβλήματος που αφορά στην εκτίμηση των φυσικών παραμέτρων ενός δονούμενου αντικειμένου, μέσω του παραγόμενου ήχου. Συγκεκριμένα, η εργασία στοχεύει στη μελέτη και την υλοποίηση συνελκτικών νευρωνικών δικτύων (Convolutional Neural Network - CNN), τα οποία θα χρησιμοποιηθούν για την εκτίμηση των φυσικών παραμέτρων μιας δονούμενης μεμβράνης, μέσω του ήχου που παράγει όταν αυτή διεγερθεί.

Για το σκοπό αυτό, δημιουργήθηκε μια βάση δεδομένων, η οποία περιλαμβάνει αρχεία ήχου που αντιστοιχούν σε ήχους παραγόμενους από μεμβράνες, καθώς και τις παραμέτρους που καθορίζουν τα κατασκευαστικά και γεωμετρικά χαρακτηριστικά αυτών. Η δημιουργία των ηχητικών δειγμάτων βασίζεται στην επίλυση της κυματικής εξίσωσης της μεμβράνης, με τη μέθοδο των Πεπερασμένων Διαφορών Χρονικού Τομέα (FDTD), η οποία υλοποιείται με ένα leapfrog αλγόριθμο στη μονάδα γραφικής επεξεργασίας (GPU) ενός H/Y με χρήση της αρχιτεκτονικής CUDA, επιλύοντας έτσι το ευθύ πρόβλημα. Για την αντιμετώπιση του αντίστροφου προβλήματος, τα παραπάνω ηχητικά δείγματα χρησιμοποιούνται για την εκπαίδευση ενός Συνελκτικού Νευρωνικού Δικτύου (CNN), μέσω του οποίου γίνεται προσπάθεια εκτίμησης των γεωμετρικών και κατασκευαστικών χαρακτηριστικών των μεμβρανών που παράγουν τους συγκεκριμένους ήχους.

Τα αποτελέσματα του εκπαιδευμένου μοντέλου αξιολογήθηκαν και προέκυψαν χρήσιμα συμπεράσματα, τα οποία στη συνέχεια οδήγησαν στη αλλαγή του αρχικού στόχου της εργασίας, επαναπροσδιορίζοντας το αντίστροφο πρόβλημα και διαμορφώνοντας νέες παραμέτρους για το ευθύ. Το νέο αντίστροφο πρόβλημα διερευνά τρόπους, αρχικά, για την εκτίμηση του μοτίβου/σχήματος (πρόβλημα ταξινόμησης) του αποσβεστικού υλικού που εφαρμόζεται στην επιφάνεια της μεμβράνης, με βάση τον παραγόμενο ήχο και δευτερεύοντος, για την εκτίμηση της επιπλέον μάζας (πρόβλημα παλινδρόμησης) λόγω αυτής της προσθήκης αποσβεστικού υλικού. Οι παράμετροι του ευθέως προβλήματος αναδιαμορφώθηκαν, ώστε να περιλαμβάνουν την προσθήκη αποσβεστικού υλικού σε μορφή πάστας και το αντίστροφο πρόβλημα αντιμετωπίστηκε με επιτυχία χρησιμοποιώντας ένα νέο συνελκτικό μοντέλο.

1.3 Σχετική έρευνα

Στην παρούσα εργασία χρησιμοποιήθηκε το φυσικό μοντέλο δονούμενης μεμβράνης όπως αυτό περιγράφεται στις μελέτες Bader [1, 2]. Συγκεκριμένα, η κυκλική μεμβράνη ενός τυμπάνου (κεφαλή ενός frame drum) διαμορφώθηκε ως βισκοελαστικό (ιξωδοελαστικό) μοντέλο πεπερασμένων διαφορών χρονικού τομέα, με παρόμοιο τρόπο όπως έγινε για την μοντελοποίηση γεωμετριών κιθάρας, βιολιού και άλλων μουσικών οργάνων [3]. Επιπλέον, για την επίλυση του ευθέως προβλήματος, δηλαδή για τη δημιουργία των ηχητικών δειγμάτων, χρησιμοποιήθηκε η υλοποίηση σε κώδικα, που παραχωρήθηκε από τον Καθηγητή R. Bader. Ο κώδικας αυτός μέσω ενός αλγορίθμου leapfrog [3, 9] επιλύει την κυματική εξίσωση που περιγράφει την κίνηση της δονούμενης μεμβράνης, εξάγοντας τιμές μετατόπισης, οι οποίες μετέπειτα προστίθενται για να υπολογίσουν τον ήχο που αναμένεται να παραχθεί σε κάποια απόσταση από τη μεμβράνη. Ο ήχος αυτός υπολογίζεται ως ψηφιακό σήμα το οποίο χρησιμοποιήθηκε στην διαδικασία επίλυσης του αντίστροφου προβλήματος. Στις παραπάνω μελέτες του Bader, αλλά και στην έρευνα των Worland και Miyahira [10], στο φυσικό μοντέλο της μεμβράνης, προστέθηκε υλικό απόσβεσης με ιξωδοελαστικές ιδιότητες και μελετήθηκε η δονητική συμπεριφορά του συστήματος. Στα πλαίσια όμως της παρούσας εργασίας και για λόγους απλοποίησης της πειραματικής διαδικασίας, το υλικό της μεμβράνης θεωρήθηκε ομοιογενές, ενώ για την απόσβεση χρησιμοποιήθηκε ένας σταθερός όρος, όπως θα περιγραφεί στη συνέχεια.

Συναφείς εργασίες για τη μελέτη του ευθέως και για την αντιμετώπιση του αντίστροφου, δηλαδή της εκτίμησης φυσικών παραμέτρων από ήχο, με τη χρήση τεχνικών βαθιάς μηχανικής μάθησης, έχουν δημοσιευθεί στην πρόσφατη επιστημονική βιβλιογραφία, για

παράδειγμα από τους Gabrielli et al. [4, 5, 6], Olivieri et al. [7] και Gonzalez et al. [8]. Στην εργασία των Gabrielli et al. [4] περιγράφεται μια γενική και ολοκληρωμένη προσέγγιση για το πρόβλημα εκτίμησης των παραμέτρων ενός φυσικού μοντέλου μέσω του παραγόμενου ήχου, η οποία βασίζεται σε ένα επιτηρούμενο μοντέλο συνελκτικής μηχανικής μάθησης (supervised convolutional machine learning). Μέσω του συνελκτικού μοντέλου, οι παράμετροι «μαθαίνονται» από δεδομένα που δημιουργούνται από το ίδιο το μοντέλο, χωρίς να απαιτείται καμία προετοιμασία και επισήμανση του συνόλου των δεδομένων εκπαίδευσης. Ως δεδομένα εισόδου το μοντέλο χρησιμοποιεί ηχητικά δείγματα, τα οποία πρώτα μετασχηματίστηκαν μέσω ενός γρήγορου μετασχηματισμού Fourier (Short-Time Fourier Transform). Η μέθοδος αυτή εφαρμόστηκε σε ένα φυσικό μοντέλο κυματοδηγών εκκλησιαστικού οργάνου (waveguide pipe organ model), αποδίδοντας πολύ ελπιδοφόρα αποτελέσματα. Η μελέτη του M. Olivieri et al. [7], εστιάζει σε εναλλακτικές προσεγγίσεις για την ανάκτηση πληροφοριών και τη δονητική, ακουστική και ηχητική ανάλυση των μουσικών οργάνων, χρησιμοποιώντας μηχανική ευφυΐα, αμβλύνοντας το χάσμα μεταξύ αντικειμενικών και υποκειμενικών περιγραφών, αλλά και περιορίζοντας τις δυσκολίες που παρουσιάζει η χρήση εξειδικευμένων εργαστηριακών μετρήσεων. Τέλος, στην εργασία των Gonzalez et al. [8] υποστηρίζεται ότι οι κανονικοί τρόποι δόνησης και οι ιδιοσυχνότητες του κελύφους ενός βιολιού μπορούν να προβλεφθούν από γεωμετρικές παραμέτρους, με χρήση τυπικών εργαλείων στατιστικής εκμάθησης και ότι η τεχνητή νοημοσύνη μπορεί να εφαρμοστεί με επιτυχία στην παραδοσιακή κατασκευή βιολιών. Στη μελέτη αυτή εξήχθησαν μέσω προσομοίωσης οι ιδιοσυχνότητες του κελύφους βιολιών, οι οποίες χρησιμοποιήθηκαν για την εκπαίδευση νευρωνικών δικτύων, τα οποία έπειτα προέβλεψαν με μεγαλύτερη ακρίβεια τις ιδιοσυχνότητες, σε σχέση με την προσομοίωση Finite Elements Method (FEM).

1.4 Οργάνωση κειμένου

Η παρουσίαση της εργασίας χωρίζεται σε δύο μέρη, στο Θεωρητικό και στο Πειραματικό μέρος. Στο 1^ο κεφάλαιο, μετά από μια σύντομη εισαγωγή, παρουσιάζεται το κύριο αντικείμενο και ο σκοπός της έρευνας καθώς και οι δημοσιεύσεις που μελετήθηκαν ως πλέον συναφείς για την επίτευξη του συγκεκριμένου σκοπού. Έπειτα ακολουθεί το Θεωρητικό μέρος της εργασίας, στο οποίο γίνεται παρουσίαση όλων των βασικών θεωρητικών εννοιών με τις οποίες ασχολήθηκε η έρευνα και περιλαμβάνει τα εξής κεφάλαια:

- 2^ο κεφάλαιο, όπου γίνεται περιγραφή των χαρακτηριστικών των κεφαλών των μεμβρανόφωνων οργάνων και του τρόπου με τον οποίο παράγεται ο ήχος, καθώς επίσης και η μαθηματική περιγραφή της δονούμενης μεμβράνης.
- 3^ο κεφάλαιο, όπου παρουσιάζονται οι αριθμητικές μέθοδοι που χρησιμοποιούνται για την επίλυση της κυματικής εξίσωσης, εστιάζοντας στη μέθοδο των Πεπερασμένων Διαφορών Χρονικού Τομέα, που χρησιμοποιήθηκε για την επίλυση του ευθέως προβλήματος, και έπειτα περιγράφεται το μαθηματικό μοντέλο της δονούμενης μεμβράνης στη διακριτή του μορφή.
- 4^ο κεφάλαιο, όπου γίνεται αναφορά στις τεχνικές Βαθιάς Μηχανικής Μάθησης και συγκεκριμένα στα Συνελκτικά Νευρωνικά Δίκτυα, που επιλέχθηκαν για την επίλυση του αντίστροφου προβλήματος.

Στη συνέχεια ακολουθεί το Πειραματικό μέρος της εργασίας, το οποίο περιλαμβάνει τη μεθοδολογία της έρευνας που ακολουθήθηκε, τον τρόπο υλοποίησης των πειραματικών διαδικασιών καθώς και τα αποτελέσματα. Συγκεκριμένα, το Πειραματικό μέρος περιλαμβάνει τα εξής κεφάλαια:

- 5^ο κεφάλαιο, όπου εστιάζει στην αριθμητική μέθοδο για την επίλυση του ευθέως προβλήματος, δηλαδή της εκτίμησης του ήχου από την περιγραφή του αντικειμένου. Σε αυτό το κεφάλαιο παρουσιάζονται ο τρόπος δημιουργίας των ηχητικών δειγμάτων της μεμβράνης, η επιλογή των φυσικών παραμέτρων του μοντέλου της και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια περιγράφεται ο επαναπροσδιορισμός του αντίστροφου προβλήματος και ο λόγος για τον οποίο αυτό ήταν αναγκαίο, όπως επίσης και η εκ νέου διαμόρφωση των φυσικών παραμέτρων των οποίων οι τιμές θα εκτιμηθούν κατά την υλοποίηση του αντίστροφου προβλήματος.
- 6^ο κεφάλαιο, όπου παρουσιάζονται οι προσεγγίσεις μηχανικής μάθησης που επιλέχθηκαν για την επίλυση του (νέου) αντίστροφου προβλήματος, καθώς και τα προγραμματιστικά εργαλεία μηχανικής μάθησης που χρησιμοποιήθηκαν. Ακολουθεί η περιγραφή της δημιουργίας των συνόλων δεδομένων για την εκπαίδευση και την επικύρωση των Συνελκτικών μοντέλων, όπως επίσης και η διαδικασία υλοποίησής τους. Τέλος γίνεται η παρουσίαση του τελικού συνελκτικού δικτύου και των αποτελεσμάτων του.

Τέλος, στο 7^ο κεφάλαιο σημειώνονται τα πιο σημαντικά θέματα με τα οποία ασχολήθηκε η εργασία και αναφέρονται τα προβλήματα που προέκυψαν. Επίσης, συνοψίζονται τα τελικά συμπεράσματα και προτείνονται θέματα για μελλοντική μελέτη και επεκτάσεις, ενώ στο τέλος, παρουσιάζεται η βιβλιογραφία που χρησιμοποιήθηκε για τη διεκπεραίωση αυτής της διπλωματικής εργασίας.

Μέρος Α΄
Θεωρητικό υπόβαθρο

2

Μεμβράνες & Μεμβρανόφωνα

Τα ιδιόφωνα και τα μεμβρανόφωνα, ή γενικότερα τα κρουστά, όπως γενικά έχει επικρατήσει να ονομάζονται, αποτελούν τα αρχαιότερα μουσικά όργανα. Το κτύπημα αντικείμενων με τα χέρια ή και μεταξύ τους, παράγει ήχο μέσω των δονήσεων των αντικειμένων, δημιουργώντας έτσι αυτόματα ένα ιδιόφωνο όργανο. Με τον όρο μεμβρανόφωνα, στην επιστήμη της οργανολογίας θεωρούμε τα μουσικά όργανα, στα οποία η παραγωγή του ήχου οφείλεται στην ταλάντωση μιας τεντωμένης μεμβράνης. Τα μεμβρανόφωνα και οι τρόποι δόνησής τους περιγράφονται σε διάφορα εξειδικευμένα συγγράμματα όπως για παράδειγμα στα Fletcher and Rossing [11], Baines [12] και Michels [13]. Οι παρακάτω υποενότητες παραθέτουν μία γενική περιγραφή που συνοψίζει τα ζητήματα που ενδιαφέρουν στην παρούσα εργασία.

2.1 Κατηγορίες μεμβρανόφωνων

Τα μεμβρανόφωνα διαχωρίζονται σε τρεις κατηγορίες: στα κρουστά (struck), στα τριβής (friction) και στα μεμβρανόφωνα πνεύσης (singing). Στα κρουστά μεμβρανόφωνα, η μεμβράνη (κεφαλή - drumhead) βρίσκεται τεντωμένη εντός πλαισίου (στεφάνη) και η κρούση της γίνεται είτε με τα χέρια (τουμπελέκι, τάμπλα, bass-drum κ.α.), είτε με μπαγκέτες (τύμπανο, νταούλι κ.α.). Μπορούν να κατηγοριοποιηθούν με βάση τη μορφή του σώματός τους (κέλυφος), σε λεβητοειδή, σωληνοειδή και στεφανοειδή, με βάση το υλικό κατασκευής τους, σε ξύλινα, μεταλλικά κ.α., ενώ επιπλέον, διαχωρίζονται σε αυτά που κουρδίζονται και σε αυτά που δεν κουρδίζονται.

Ως δονητικά συστήματα τα μεμβρανόφωνα διαχωρίζονται σε αυτά με δύο μεμβράνες και σε αυτά με μια. Στα μεμβρανόφωνα με δύο μεμβράνες δημιουργείται στο εσωτερικό τους ένας κλειστός χώρος, μέσα στον οποίο ο αέρας θέτει σε συζευγμένη παλμική κίνηση τις δυο μεμβράνες, ενισχύοντας έτσι τον ήχο και επηρεάζοντας το ηχόχρωμα με την προσθήκη πλουσιότερων αρμονικών συχνοτήτων. Στα μεμβρανόφωνα με μια μεμβράνη

λείπει το χαρακτηριστικό της συζευγμένης κίνησης, δημιουργώντας έτσι έναν ακαθόριστος τονικότητας ήχο.

2.2 Χαρακτηριστικά μεμβρανόφωνων

Τα μεμβρανόφωνα αποτελούνται από μια υπό τάση μεμβράνη, συνήθως κυκλική, η οποία είναι στερεωμένη στο πλαίσιο του οργάνου και είναι κατασκευασμένη είτε από δέρματα ζώων, είτε από συνθετικά πλαστικά δέρματα. Ο ήχος παράγεται από την ταλάντωση της μεμβράνης, η οποία προκαλείται μέσω κρούσης, τριβής ή φυσήματος, ενώ ενισχύεται μέσω του σώματος (κελύφους) του οργάνου, το οποίο λειτουργεί ως αντηχείο, όπως στα περισσότερα μουσικά όργανα. Γενικά, τα μεμβρανόφωνα παράγουν ήχους ακαθόριστου τονικού ύψους και συνήθως δεν υπάρχει δυνατότητα να μεταβληθεί άμεσα η τονικότητά τους, εξαιρουμένων των κουρδιζόμενων μεμβρανόφωνων. Σε γενικές γραμμές, κρούοντας τη μεμβράνη προς το κέντρο της, ο ήχος γίνεται πιο μπάσος, ενώ κοντά στη στεφάνη γίνεται πιο υψηλός. Βασικοί παράγοντες που επηρεάζουν, την τονικότητα και το ηχόχρωμα του παραγόμενου ήχου, είναι η τάση που ασκείται στην μεμβράνη, το μέγεθος της, η μάζα ανά μονάδα επιφάνειας αλλά και το σχήμα, το μέγεθος και το υλικό του αντηχείου-κελύφους.

2.3 Frame drums, τύμπανα & μπεντίρ

Η Ανατολίτικη, Αφρικανική, Ασιατική, Ινδική, Ιρλανδική αλλά και η Λατιναμερικανική μουσική χρησιμοποιούν διάφορα είδη frame drums [14][15] (Εικόνα 2-1). Τα τύμπανα αυτά αποτελούνται από ένα στενό πλαίσιο, ξύλινο ή μεταλλικό, πάνω στο οποίο τεντώνεται μια μεμβράνη, παραδοσιακά φτιαγμένη από δέρμα. Παράγουν μια μεγάλη ποικιλία τόνων από βαθιές νότες μπάσων (στο κέντρο της κεφαλής) έως και υψηλές νότες πλούσιες σε αρμονικές (πλησιέστερα στην περιφέρεια).

Το σημαντικότερο μεμβρανόφωνο που συναντάται στην Αρχαία Ελλάδα είναι το τύμπανο, το οποίο είναι παρόμοιο με το ανατολίτικο μπεντίρ (bendir). Αποτελείται από ένα ανοιχτό και ρηχό κύλινδρο (κυλινδρικό ξύλινο πλαίσιο), με διάμετρο 30-50 cm, στην μία ανοιχτή πλευρά του οποίου, είναι στερεωμένη μια τεντωμένη δερμάτινη μεμβράνη. Στην ελληνική παραδοσιακή μουσική, ως εξέλιξη του τυμπάνου μπορεί να θεωρηθεί το νταούλι, αλλά και οποιοδήποτε άλλο σύγχρονο μεμβρανόφωνο όργανο, όπως το ημισφαιρικό τύμπανο (timpani) της συμφωνικής ορχήστρας, το μεγάλο και μικρό

ταμπούρο κ.α. Σ' αυτή την κατηγορία, αλλά με την προσθήκη κυμβάλων (ζίλια), ανήκουν και τα διάφορα είδη ντεφιών με μεμβράνη, όπως το ηπειρώτικο ντέφι, οι νταϊρέδες κ.α.

Το μπεντίρ συναντιέται κυρίως στις αραβικές χώρες της βόρειας Αφρικής αλλά και στην οθωμανική μουσική παράδοση. Δεν διαθέτει ζίλια και σε πολλές περιπτώσεις, εφαρμόζεται πάνω στη μεμβράνη μια χορδή από πλαστικό ή άλλο υλικό, η οποία του προσδίδει ένα χαρακτηριστικό ήχο «τριξίματος». Το ξύλινο πλαίσιο πάνω στο οποίο είναι στερεωμένη η μεμβράνη είναι συνήθως γύρω στα δύο ή τρία χιλιοστά. Αν και τα παραδοσιακά μπεντίρ δεν κουρδίζονται, κυκλοφορούν σήμερα στο εμπόριο μπεντίρ από συνθετικό δέρμα, τα οποία κουρδίζονται με μεταλλικά κλειδιά. Συνήθως ένα μπεντίρ διαμέτρου 40-50 cm, είναι καλό για γενική χρήση, ενώ υπάρχουν και μεγαλύτερης διαμέτρου, κατάλληλα για πιο εντυπωσιακούς, μεγαλύτερης διάρκειας ήχους. Στα κρουστά όργανα τύπου Bendir, λόγω κατασκευής, η πίσω τους περιοχή είναι κατά βάση περιοχή της εξόδου ενός αντηχείου ή περιοχή κοιλότητας συντονισμού, με έντονη την παρουσία χαμηλών συχνοτήτων.



Εικόνα 2-1: Αρχαίο Ελληνικό τύμπανο (αριστερά). Μπεντίρ (κέντρο). Διάφορα είδη frame drums (δεξιά).

2.4 Ήχος από μεμβράνες, φυσική περιγραφή

Όταν μια εύκαμπτη τεντωμένη μεμβράνη, όπως αυτή ενός τύμπανου, διεγερθεί κτυπώντας την σε ένα σημείο, τότε δημιουργείται ένας δισδιάστατος παλμός (μια δισδιάστατη κατανομή πλάτους δόνησης), που ξεκινάει από το σημείο διέγερσης και κατευθύνεται προς την περιφέρειά της [11, 16]. Η περιφέρεια της μεμβράνης, όντας στερεωμένη, παραμένει ακίνητη και επομένως η ενέργεια του παλμού ανακλάται πλήρως και επιστρέφει πίσω προς το κέντρο. Έτσι, αν κάποιο σημείο της μεμβράνης δονείται εξαναγκασμένα, δημιουργούνται συρμοί κυμάτων που κατευθύνονται προς την περιφέρεια και ανακλώνται, δημιουργώντας δισδιάστατα στάσιμα κύματα, τα οποία παρουσιάζουν δεσμό στην ακίνητη περιφέρεια.

Το κάθε ένα από αυτά τα στάσιμα κύματα έχει μια χαρακτηριστική συχνότητα (συχνότητα συντονισμού), που ονομάζεται ιδιοσυχνότητα (φυσική συχνότητα - natural frequency) της μεμβράνης. Η μικρότερη από αυτές ονομάζεται θεμελιώδης και οι επόμενες υπερτόνοι (overtones). Οποιαδήποτε ελεύθερη κίνηση της μεμβράνης αποτελεί μια υπέρθεση της θεμελιώδης συχνότητας με κάποιους από τους υπερτόνους, διεγείροντας έτσι ηχητικά κύματα με ανάλογο φάσμα συχνοτήτων. Η ιδιότητα που κάνει αναγνωρίσιμο και ξεχωριστό τον ήχο κάθε οργάνου ονομάζεται ηχόχρωμα και καθορίζεται τόσο από τη σχέση υπερτόνων, όσο και από την ποσοστιαία αναλογία της συμμετοχής των διαφόρων αρμονικών στο συνολικό παραγόμενο ήχο. Επισημαίνεται εδώ, ότι συγκριτικά με τα μελωδικά μουσικά όργανα, τα κρουστά δεν έχουν αρμονικό φάσμα. Δηλαδή οι κορυφές του φάσματος δεν είναι ακέραια πολλαπλάσια της θεμέλιου, για αυτό άλλωστε και δεν έχουν ξεκάθαρο τονικό ύψος (pitch).

Κατά τη δόνηση των μεμβρανών, διακρίνονται ακίνητες γραμμές, που ονομάζονται δεσμικές γραμμές (δεσμοί). Οι δεσμοί αυτοί μπορεί να είναι ευθείες γραμμές (διαμετρικοί δεσμοί), που είναι διάμετροι της εξωτερικής περιφέρειας, αλλά και κύκλοι (κυκλικοί δεσμοί) ομόκεντροι με τη περιφέρεια. Οι δεσμοί χωρίζουν τη μεμβράνη σε συμμετρικά μέρη, που το καθένα κινείται σε αντίθετη κατεύθυνση με τα γειτονικά του. Ανάλογα με το είδος και τον αριθμό των δεσμών, που παρουσιάζονται στη δονούμενη μεμβράνη, έχουμε και διαφορετικό τρόπο δόνησης (μοτίβο κίνησης - mode shape), ο οποίος ονομάζεται κανονικός τρόπος δόνησης (normal mode) και επισημαίνεται με ένα ζευγάρι (m, n) ακέραιων αριθμών. Ο πρώτος αριθμός (m) δίνει τον αριθμό των διαμετρικών δεσμών και ο δεύτερος (n) τον αριθμό των κυκλικών δεσμών. Οι κανονικοί τρόποι δόνησης διεγείρονται ανεξάρτητα και είναι ορθογώνιοι μεταξύ τους, επομένως κανένας κανονικός τρόπος δόνησης δεν μπορεί να δημιουργηθεί από γραμμικό συνδυασμό άλλων τρόπων. Καθεμία από τις ιδιοσυχνότητες αντιστοιχεί σε ένα κανονικό τρόπο, κάθε ένας από τους οποίους ξεκινά με διαφορετική ένταση και εξασθενεί (αποσβένει) με διαφορετικό ρυθμό. Ως αποτέλεσμα, το φάσμα συχνοτήτων του παραγόμενου ήχου μεταβάλλεται με την πάροδο του χρόνου και εξαρτάται από το ποιοι κανονικοί τρόποι δόνησης διεγείρονται.

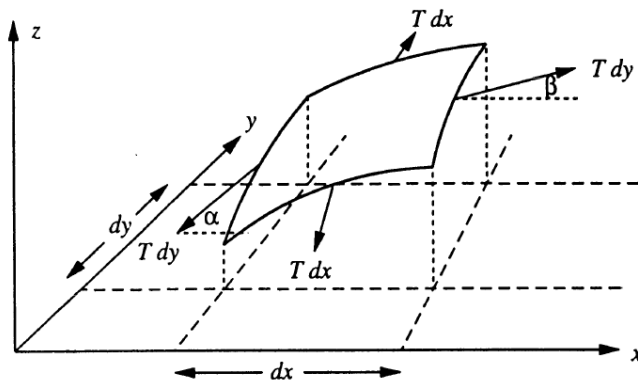
Κατά την κρούση για παράδειγμα μιας μεμβράνης ενός τυμπάνου, οι κανονικοί τρόποι δόνησης που διεγείρονται πιο έντονα είναι αυτοί που παρουσιάζουν κοιλίες (anti-nodes - τοπικά μέγιστα πλάτους δόνησης) στο σημείο όπου κρούεται η κεφαλή [16, 17, 18]. Ο θεμελιώδης τρόπος εμφανίζεται επομένως περισσότερο χτυπώντας προς το κέντρο του τυμπάνου, όπου η δόνηση έχει το μεγαλύτερο πλάτος, ενώ οι συντονισμοί υψηλότερων συχνοτήτων εμφανίζονται περισσότερο όταν η κεφαλή χτυπηθεί μακριά από αυτό. Οι συχνότητες αυτές δεν είναι σε χρήσιμα μουσικά διαστήματα και δεν δημιουργούν μια αρμονική σειρά, δηλαδή δεν παρουσιάζουν μια αισθητή τονικότητα.

2.5 Μαθηματική περιγραφή δόνησης μεμβρανόφωνων

Σε αυτή την ενότητα, ξεκινώντας από την περιγραφή της κυματικής εξίσωσης στην περίπτωση της τετραγωνικής μεμβράνης, δίνεται μια μαθηματική θεώρηση [11, 16, 17, 19] του τρόπου δόνησης της κυκλικής μεμβράνης, η οποία χρησιμοποιείται στο πειραματικό μέρος της εργασίας και στη συνέχεια αναλύονται οι ιδιοσυχρότητες και οι κανονικοί τρόποι δόνησής της.

2.5.1 Κυματική εξίσωση τετραγωνικής μεμβράνης

Μια μεμβράνη είναι μία τέλεια εύκαμπτη λεπτή πλάκα ή έλασμα που υπόκειται σε τάση (αλλιώς ένταση). Έχει αμελητέα αντίσταση στις δυνάμεις διάτμησης ή κάμψης και οι δυνάμεις επαναφοράς προκύπτουν αποκλειστικά από τις δυνάμεις προέντασης ή παραμόρφωσης εντός του επιπέδου (in-plane stretching). Η διάσταση z της μεμβράνης, δηλαδή το πάχος της, θεωρείται αμελητέο σε σχέση με τις άλλες δύο διαστάσεις, μήκος και πλάτος, και γι' αυτό η μεμβράνη εξετάζεται ως δισδιάστατη οντότητα στον τρισδιάστατο ευκλείδειο χώρο.



Εικόνα 2-2: Απειροστό τμήμα μεμβράνης.

Έστω προεντεταμένη ορθογώνια μεμβράνη με διαστάσεις L_x και L_y , πακτωμένη (clamped) σε όλες τις πλευρές της, με προένταση T , ομοιόμορφη σε όλη την επιφάνειά της. Θεωρούμε απειροστό τμήμα της μεμβράνης με επιφανειακή πυκνότητα μ , όπως αυτό απεικονίζεται στην Εικόνα 2-2. Διεγείροντας τη μεμβράνη και δημιουργώντας στοιχειώδη κατακόρυφη μετατόπιση dz , η επιφανειακή τάση T της μεμβράνης αντιδρά, προσπαθώντας να την επαναφέρει στην αρχική της κατάσταση ισορροπίας.

Για μικρές γωνίες α και β ισχύει:

$$\left(\frac{\partial z}{\partial y}\right)_y = \tan \alpha \cong \sin \alpha \quad \text{και} \quad \left(\frac{\partial z}{\partial y}\right)_{y+dy} = \tan \beta \cong \sin \beta$$

και αποδεικνύεται ότι η δύναμη στην ακμή dx κατά την διεύθυνση z είναι:

$$F_x = T dy \frac{\partial^2 z}{\partial x^2} dx \quad (2.1)$$

ενώ στην ακμή dy κατά την διεύθυνση z είναι:

$$F_y = T dx \frac{\partial^2 z}{\partial y^2} dy \quad (2.2)$$

Επομένως η συνολική δύναμη που ασκείται στο στοιχείο $dx dy$ κατά την διεύθυνση z είναι:

$$F_x + F_y = F \Rightarrow T dx dy \left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} \right) = \mu dx dy \frac{\partial^2 z}{\partial t^2}$$

δηλαδή:

$$\frac{\partial^2 z}{\partial t^2} = \frac{T}{\mu} \left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} \right) = c^2 \nabla^2 z \quad (2.3)$$

Η εξίσωση (2.3) αποτελεί την κυματική εξίσωση των εγκάρσιων κυμάτων με ταχύτητα:

$$c = \sqrt{\frac{T}{\mu}} \quad (2.4)$$

Για την επίλυση της κυματικής εξίσωσης μπορεί να χρησιμοποιηθεί η μέθοδος διαχωρισμού των μεταβλητών, εκφράζοντας την μετατόπιση $z(x, y, t)$ ως γινόμενο τριών συναρτήσεων, μια για κάθε μεταβλητή.

Οπότε η εξίσωση (2.3) γίνεται:

$$\frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial t^2} = \frac{c^2}{X} \frac{\partial^2 X}{\partial x^2} + \frac{c^2}{Y} \frac{\partial^2 Y}{\partial y^2} \quad (2.5)$$

Η εξίσωση (2.5) ισχύει μόνο αν και τα δύο μέλη ισούνται με μία σταθερά, η οποία ορίζεται ως ω^2 , οπότε και προκύπτουν οι εξισώσεις:

$$\frac{d^2 \Phi}{dt^2} + \omega^2 \Phi = 0$$

με λύσεις της μορφής:

$$\Phi(t) = E \sin \omega t + F \cos \omega t \quad (2.6)$$

και

$$\frac{1}{X} \frac{\partial^2 X}{\partial x^2} + \frac{\omega^2}{c^2} = -\frac{1}{Y} \frac{\partial^2 Y}{\partial y^2} \quad (2.7)$$

Ομοίως, η εξίσωση (2.7) είναι αληθής μόνο αν και τα δύο μέλη ισούνται με μία σταθερά την οποία ορίζουμε ως k^2 , και εξισώνοντας προκύπτουν οι εξισώσεις:

$$\frac{\partial^2 X}{\partial x^2} + \left(\frac{\omega^2}{c^2} - k^2 \right) X = 0,$$

με λύσεις:

$$X(x) = A \sin\left(\sqrt{\left(\frac{\omega^2}{c^2} - k^2\right)x}\right) + B \cos\left(\sqrt{\left(\frac{\omega^2}{c^2} - k^2\right)x}\right) \quad (2.8)$$

και

$$\frac{\partial^2 Y}{\partial y^2} + k^2 Y = 0$$

με λύσεις:

$$Y(y) = C \sin ky + D \cos ky \quad (2.9)$$

Για τη συγκεκριμένη γεωμετρία μεμβράνης, δηλαδή για ορθογώνια μεμβράνη με διαστάσεις L_x , L_y , πακτωμένη και στις τέσσερις ακμές της, οι συνοριακές συνθήκες απαιτούν ότι:

$$z = 0 \text{ για } x = 0, x = L_x, y = 0 \text{ και } y = L_y \quad (2.10)$$

Με βάση τις συνθήκες (2.8), η εξίσωση (2.9) παίρνει την μορφή:

$$Y(y) = C \sin \frac{n\pi y}{L_y}, \text{ με } n = 1, 2, \dots \quad (2.11)$$

και η εξίσωση (2.8) παίρνει την μορφή:

$$Y(y) = C \sin \frac{n\pi y}{L_y}, \text{ με } n = 1, 2, \dots \quad (2.12)$$

Επομένως, με τη χρήση των εξισώσεων (2.10), (2.11) και (2.12), η εξίσωση (2.5) δίνει την λύση της κυματικής εξίσωσης:

$$z_{mn} = A \sin \frac{m\pi x}{L_x} C \sin \frac{n\pi y}{L_y} (E \sin \omega t + F \cos \omega t) \Rightarrow$$

$$z_{mn} = \sin \frac{m\pi x}{L_x} \sin \frac{n\pi y}{L_y} (M \sin \omega t + N \cos \omega t), \text{ με } m, n = 1, 2, \dots \quad (2.13)$$

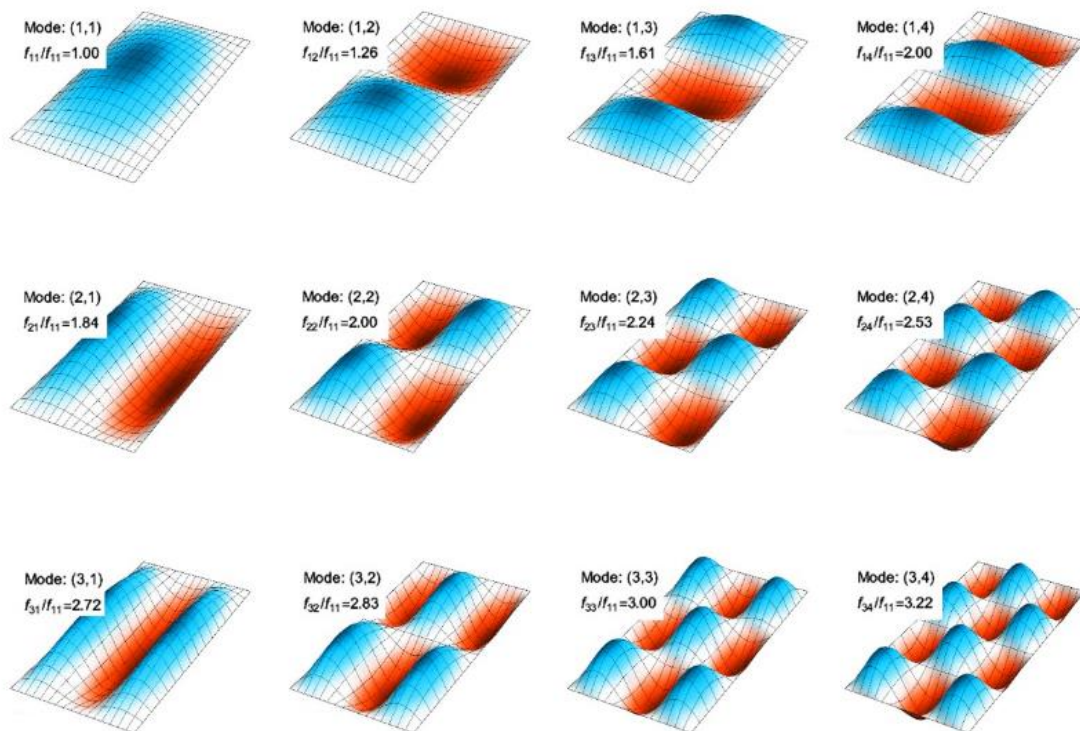
Η εξίσωση για τον προσδιορισμό των ιδιοσυχνοτήτων του συστήματος δίνεται επιλύοντας ως προς ω την παρακάτω εξίσωση:

$$\sqrt{\frac{\omega^2}{c^2} - k^2} = \frac{m\pi}{L_x} \Rightarrow \omega^2 = \left(\frac{m\pi}{L_x}\right)^2 c^2 + k^2 c^2 = \left(\frac{m\pi}{L_x}\right)^2 c^2 + \left(\frac{n\pi}{L_y}\right)^2 c^2$$

και αντικαθιστώντας την κυκλική συχνότητα $\omega=2\pi f$ προκύπτει:

$$f_{mn} = \frac{1}{2} \sqrt{\frac{T}{\mu}} \sqrt{\frac{m^2}{L_x^2} + \frac{n^2}{L_y^2}}, \text{ με } m, n = 1, 2, \dots \quad (2.14)$$

Για μια δεδομένη ορθογώνια μεμβράνη, η δυάδα των ακεραίων αριθμών (m,n) χαρακτηρίζει μονοσήμαντα τον κάθε κανονικό τρόπο δόνησης, αλλά και την αντίστοιχη ιδιοσυχνότητα. Στην Εικόνα 2-3 διακρίνονται κάποιοι από τους αρχικούς κανονικούς τρόπους δόνησης μιας ορθογώνιας μεμβράνης, καθώς και οι αντίστοιχες τιμές των συντελεστών m και n που τους χαρακτηρίζουν.



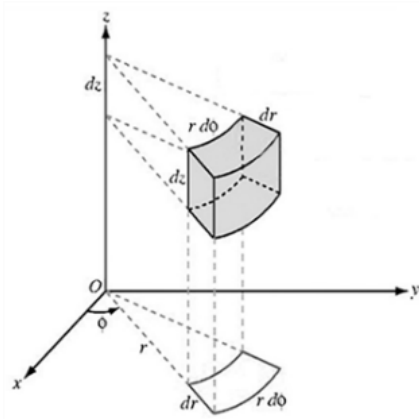
Εικόνα 2-3: Οι αρχικοί (12) κανονικοί τρόποι δόνησης μιας ορθογώνιας μεμβράνης.

Πηγή: [17]

2.5.2 Κυματική εξίσωση κυκλικής μεμβράνης

Στην περίπτωση της κυκλικής μεμβράνης [11, 16, 17, 19], η εξίσωση (2.3) για την ορθογώνια μεμβράνη πρέπει να γραφεί με τη μορφή πολικών συντεταγμένων (Εικόνα 2-4), θέτοντας $x = r \cos \varphi$ και $y = r \sin \varphi$ ως εξής:

$$\frac{\partial^2 z}{\partial t^2} = c^2 \left(\frac{\partial^2 z}{\partial r^2} + \frac{1}{r} \frac{\partial z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 z}{\partial \varphi^2} \right) \quad (2.15)$$



Εικόνα 2-4: Απειροστό τμήμα όγκου σε κυλινδρικές συντεταγμένες.

Όπως και προηγουμένως, για την επίλυση της κυματικής εξίσωσης χρησιμοποιείται η μέθοδος διαχωρισμού μεταβλητών με λύση της μορφής.

$$z(r, \varphi, t) = R(r)\Phi(\varphi)e^{j\omega t} \quad (2.16)$$

Η εξίσωση (2.16) με την προϋπόθεση ότι το κάθε μέλος της είναι ίσο με σταθερά m , οδηγεί στις παρακάτω εξισώσεις:

$$\frac{\partial^2 R}{\partial r^2} + \frac{1}{r} \frac{\partial R}{\partial r} + \left(\frac{\omega^2}{c^2} - \frac{m^2}{r^2} \right) R = 0 \quad (2.17)$$

και

$$\frac{\partial^2 \Phi}{\partial \varphi^2} + m^2 \Phi = 0 \quad (2.18)$$

Η λύση της εξίσωσης (2.18) είναι της μορφής:

$$\Phi(\varphi) = A e^{\pm jm\varphi} \quad (2.19)$$

Η εξίσωση (2.17) είναι μια εξίσωση Bessel της μορφής:

$$\frac{\partial^2 y}{\partial x^2} + \frac{1}{x} \frac{\partial y}{\partial x} + \left(1 - \frac{m^2}{x^2}\right) y = 0 \quad (2.20)$$

όπου: $y=R$ και $x = kr = (\frac{\omega}{c})r$

Οι λύσεις των παραπάνω εξισώσεων είναι συναρτήσεις Bessel πρώτου τύπου, τάξης m ($J_0(x), J_1(x), \dots, J_m(x)$), κάθε μία από τις οποίες μηδενίζεται για αρκετές τιμές του ορίσματος x . Για παράδειγμα όπως φαίνεται και στην Εικόνα 2-5, για τις πρώτες συναρτήσεις Bessel ισχύει: $J_0(x)=0$, για $x = 2.405, 5.520, 8.654, \dots$ και $J_1(x)=0$, για $x = 0, 3.83, 7.02, \dots$ κλπ.

Οι ιδιοσυχνότητες στην περίπτωση της κυκλικής μεμβράνης δίνονται από τον τύπο:

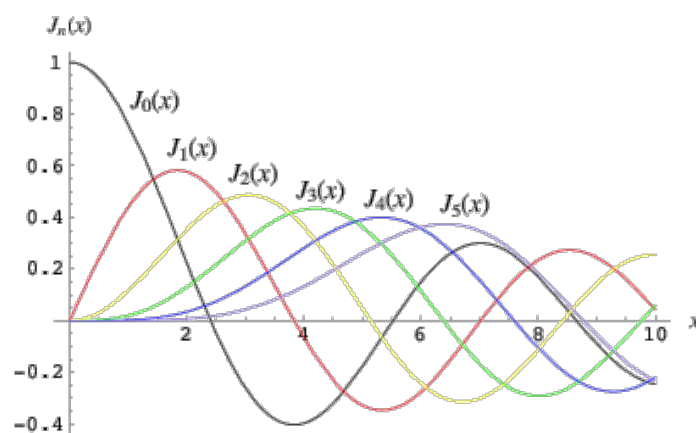
$$f_{mn} = \frac{J_{m,n} \cdot c}{2\pi r} \quad (2.21)$$

όπου, $J_{m,n}$: η n -οστή ρίζα της συνάρτησης $J_m(x)$, για $x = kr = (\frac{\omega}{c})r$

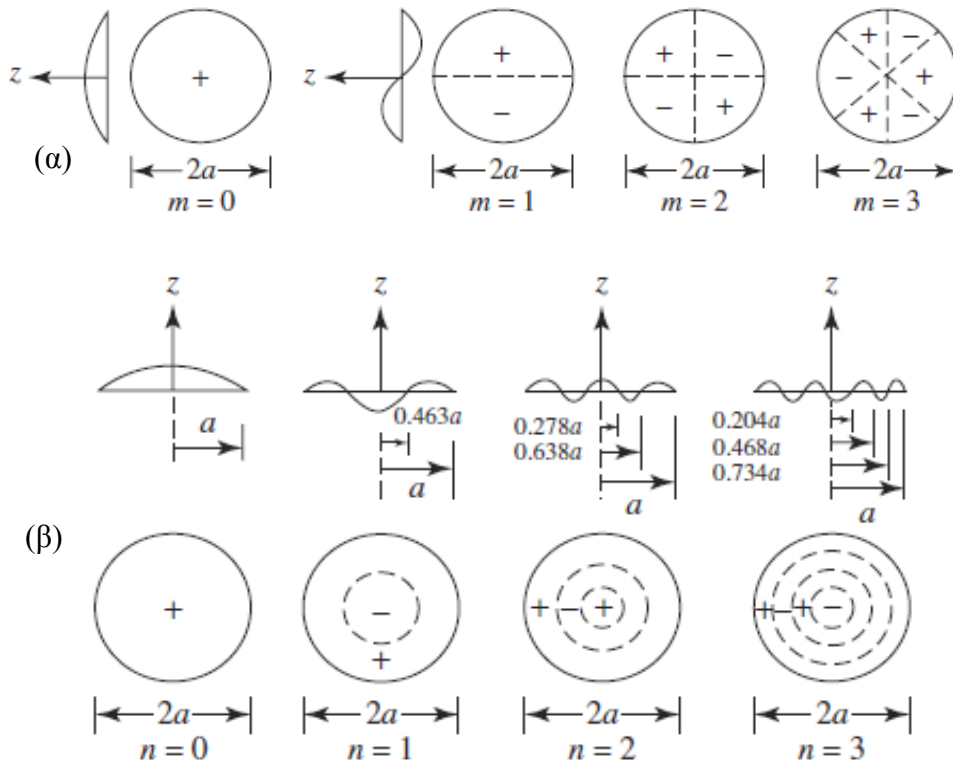
και με βάση την εξίσωση (2.4) της ταχύτητας του ήχου στη μεμβράνη η παραπάνω εξίσωση καταλήγει στη μορφή:

$$f_{mn} = \frac{1}{2\pi r} \cdot \sqrt{\frac{T}{\mu}} \cdot J_{m,n} \quad (2.22)$$

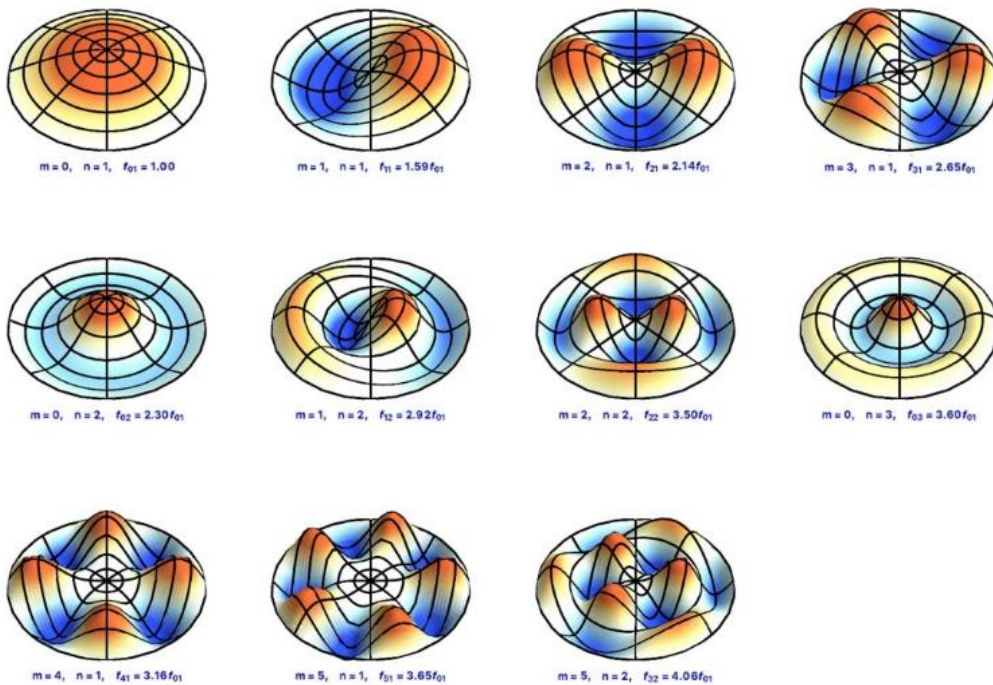
Η συχνότητα $f_{m,n}$ είναι η συχνότητα του τρόπου (m,n) , ο οποίος έχει m διαμετρικούς και n κυκλικούς δεσμούς (συμπεριλαμβανομένου και ενός στο σύνορο). Οι αρχικοί τρόποι δόνησης μιας κυκλικής μεμβράνης παρουσιάζονται στην Εικόνα 2-6, ενώ στην Εικόνα 2-7, διακρίνονται τα σχήματα των κανονικών τρόπων και οι αποστάσεις των δεσμών σε σχέση με την ακτίνα της μεμβράνης.



Εικόνα 2-5: Συναρτήσεις Bessel πρώτου είδους.



Εικόνα 2-6: Οι αρχικοί (11) κανονικοί τρόποι δόνησης μιας κυκλικής μεμβράνης.



Εικόνα 2-7: Κανονικοί τρόποι δόνησης κυκλικής μεμβράνης (πακτωμένης στα άκρα)
 α: ακτίνα μεμβράνης, m: πλήθος διαμετρικών δεσμών (α), n: πλήθος κυκλικών δεσμών (β).
 Τα +,- δηλώνουν διαφορά φάσης 180ο

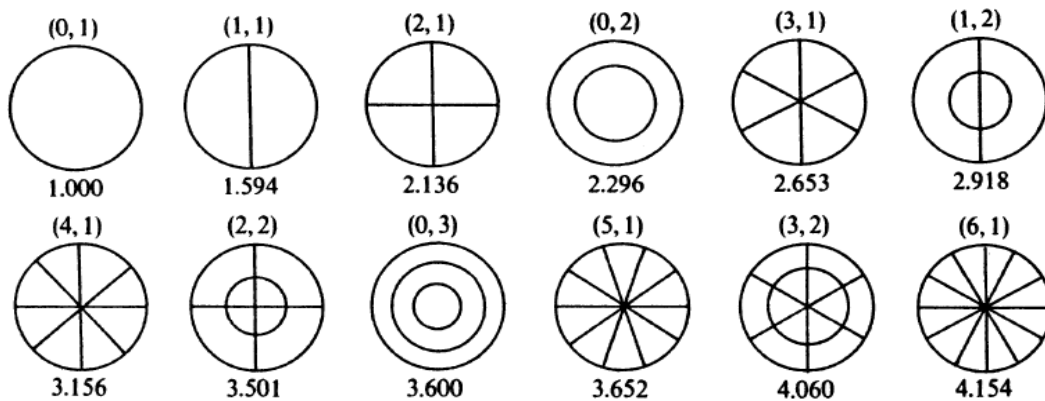
Πηγή: [17]

Στην Εικόνα 2-8, παρουσιάζονται οι 12 τρόποι δόνησης κυκλικής μεμβράνης σε 2 διαστάσεις, ενώ οι 12 πρώτες συχνότητες συντονισμού ως πολλαπλάσια της θεμελιώδους συχνότητας για μια ιδανική κυκλική μεμβράνη δίνονται στον Πίνακα 2-1.

Για να γίνει ο υπολογισμός των ιδιοσυχνοτήτων, απαιτείται ο πολλαπλασιασμός των τιμών αυτών με τον παράγοντα:

$$\left(\frac{2.405}{2\pi r} \right) \sqrt{\frac{T}{\mu}}$$

όπου r η ακτίνα της μεμβράνης (m), T η τάση της (Nt) και μ η επιφανειακή πυκνότητά της (kg/m^2). Από τον ίδιο πίνακα φαίνεται καθαρά ότι η ιδιοσυχνότητα κάθε κανονικού τρόπου ταλάντωσης δεν είναι ακέραιο πολλαπλάσιο της θεμελιώδους, δηλαδή δεν παρουσιάζουν αρμονική σχέση και επομένως ο ήχος που παράγει μια δονούμενη μεμβράνη δεν έχει συγκεκριμένο τονικό ύψος.



Εικόνα 2-8: Οι πρώτοι 12 τρόποι δόνησης κυκλικής μεμβράνης σε 2 διαστάσεις.

Πίνακας 2-1: Οι 12 πρώτες συχνότητες συντονισμού ως πολλαπλάσια της θεμελιώδους για μια ιδανική κυκλική μεμβράνη.

(m,n)	$f_{m,n}$	(m,n)	$f_{m,n}$	(m,n)	$f_{m,n}$
(0, 1)	1.000	(3, 1)	2.653	(0, 3)	3.600
(1, 1)	1.594	(1, 2)	2.918	(5, 1)	6.652
(2, 1)	2.136	(4, 1)	3.156	(3, 2)	4.060
(0, 2)	2.296	(2, 2)	3.501	(6, 1)	4.154

Πηγή: [11]

2.5.3 Ανάλυση ιδιοσυχνοτήτων & ιδιομορφών

Στον θεμελιώδη τρόπο $(m,n)=(0,1)$ ολόκληρη η κυκλική μεμβράνη δονείται σε φάση, ενώ διεγείρεται για κρούσεις σε οποιαδήποτε θέση στη μεμβράνη. Σε αυτό τον τρόπο, η μεμβράνη λειτουργεί σαν μια μονοπολική πηγή, διαδίδει τον ήχο πολύ αποτελεσματικά, μετατρέποντας γρήγορα τη δόνησή της σε ακτινοβολούμενη ηχητική ενέργεια, έχοντας έτσι μικρή διάρκεια. Αυτό σημαίνει ότι ο τρόπος $(0,1)$ δεν συμβάλλει σημαντικά στην ποιότητα του παραγόμενου ήχου, αλλά αντίθετα, παράγει ένα «χτύπημα» που διασπάται γρήγορα και χωρίς συγκεκριμένο τονικό ύψος. Στον κανονικό τρόπο δόνησης $(1,1)$, η ακριβής θέση του διαμετρικού δεσμού εξαρτάται από την ομοιογένεια της μεμβράνης και τις αρχικές συνθήκες όταν ξεκινά η δόνηση. Όταν δονείται στη λειτουργία $(1,1)$, η μεμβράνη λειτουργεί σαν μια διπολική πηγή. Αντί να ωθεί τον αέρα μακριά από τη μεμβράνη όπως κάνει ο τρόπος $(0,1)$, στον $(1,1)$ το ένα μισό της μεμβράνης ωθεί τον αέρα προς τα πάνω, ενώ το άλλο μισό απορροφά τον αέρα προς τα κάτω. Ως αποτέλεσμα, ήχος μεταδίδεται λιγότερο αποτελεσματικά και χρειάζεται περισσότερο χρόνο για να αποσβεστεί, συμβάλλοντας έτσι σημαντικά στο τονικό ύψος του παραγόμενου ήχου. Ομοίως, ο τρόπος $(2,1)$, συμβάλλει ακόμα περισσότερο στη διαμόρφωση του ήχου, καθώς η μεμβράνη λειτουργεί σαν τετραπολική πηγή, μεταφέροντας πολύ πιο αργά την ηχητική ενέργεια. Γενικά, οι κανονικοί τρόποι δόνησης που καθορίζουν σημαντικά την ποιότητα του παραγόμενου ήχου είναι οι τρόποι $(1,1)$, $(2,1)$, $(3,1)$, $(4,1)$ και $(5,1)$ [11, 16, 17, 19]. Τέλος, αξίζει να σημειωθεί ότι μπορεί να παρατηρηθούν εκφυλισμένοι κανονικοί τρόποι δόνησης σε παραπλήσιες συχνότητες, που είναι οπτικά όμοιοι μεταξύ τους αλλά διαφέρουν σε φάση.

Οι διαφορετικοί τρόποι στήριξης, φόρτισης και διέγερσης των μεμβρανών, περιπλέκουν τη μελέτη των κανονικών τρόπων δόνησης και των ιδιοσυχνοτήτων τους, καθώς προϋποθέτουν διαφορετικές συνοριακές συνθήκες και προσθήκες επιπλέον όρων για τη μαθηματική τους περιγραφή. Η μελέτη αυτών των περιπτώσεων δεν άπτονται των σκοπών της παρούσας εργασίας.

2.6 Πραγματικές συνθήκες, φορτίο αέρα & κέλυφος

Τα φαινόμενα που περιγράφηκαν στην προηγούμενη ενότητα, αφορούν μια ιδανική κυκλική μεμβράνη με ομοιόμορφη κατανομή πυκνότητας και τάσης, σε ιδανικές συνθήκες, η οποία είναι στερεωμένη στην περιφέρειά της, χωρίς να συνυπολογίζονται οι αλληλεπιδράσεις της με το πλαίσιο του τυμπάνου και τον αέρα.

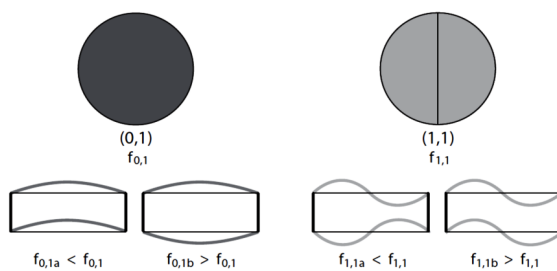
Στην πραγματικότητα όμως, και σύμφωνα με την βιβλιογραφία [11, 16, 17, 19, 20], οι συχνότητες συντονισμού των κυκλικών μεμβρανών αποκλίνουν από αυτές του Πίνακα 2-1. Τα κύρια φαινόμενα που συντελούν προς αυτή την απόκλιση είναι το φορτίο του αέρα (air loading), η καμπτική ακαμψία (bending stiffness) και η διατμητική ακαμψία (shear stiffness) της μεμβράνης και του πλαισίου-κελύφους. Γενικά, το φορτίο αέρα μειώνει τις ιδιοσυχνότητες, ενώ τα άλλα δύο φαινόμενα τείνουν να τις αυξάνουν. Η επίδραση του φορτίου αέρα στις συχνότητες εξαρτάται από τις σχετικές ταχύτητες των κυμάτων στη μεμβράνη και στον αέρα, καθώς και από το αν ο αέρας είναι περιορισμένος κατά οποιοδήποτε τρόπο. Ένας περιορισμένος όγκος αέρα, θα αυξήσει τη συχνότητα των αξονοσυμμετρικών κανονικών τρόπων δόνησης, ιδιαίτερα του τρόπου (0, 1). Αντίθετα όταν η μεμβράνη δονείται σε ελεύθερο αέρα, όλες οι ιδιοσυχνότητες μειώνονται, με τη μεγαλύτερη μείωση να παρουσιάζεται στις χαμηλότερες.

Επιπλέον, όπως έχει αναφερθεί προηγουμένως, υπάρχουν κρουστά μεμβρανόφωνα, που φέρουν μια μεμβράνη προσαρμοσμένη σε στενό κυκλικό πλαίσιο, με ανοικτή την πίσω πλευρά (ανοικτή αέρινη κοιλότητα), αλλά και μεμβρανόφωνα τα οποία φέρουν δυο μεμβράνες, μια στο πάνω μέρος, η οποία είναι αυτή που κρούεται (κρούσης - batter) και μια στο κάτω μέρος του κελύφους του οργάνου (αντήχησης - resonant), σχηματίζοντας έτσι κλειστή αέρινη κοιλότητα. Για τα μεμβρανόφωνα με ανοικτή κοιλότητα, ο τρόπος λειτουργίας, μπορεί να αντιστοιχηθεί με τον τρόπο λειτουργίας ενός αντηχείου Helmholtz². Προϋπόθεση για την δημιουργία του ήχου, αποτελούν οι δονήσεις που προκύπτουν λόγω της ελαστικότητας του αερίου. Σε μια κοιλότητα όπου έχει μόνο ένα άνοιγμα, αν συμπιεστεί το αέριο που βρίσκεται εσωτερικά, θα αυξηθεί η πίεση του και θα τείνει να επιστρέψει στην αρχική του κατάσταση. Λόγω εμπέδησης όμως της ατμοσφαιρικής πίεσης με το πεπιεσμένο εσωτερικό, η μάζα θα δυσκολευτεί να επανέλθει πλήρως κι έτσι μέρος του πεπιεσμένου αέρα θα γυρίσει ξανά στην κοιλότητα και θα δημιουργηθεί μια αποσβεστική εξαναγκασμένη ταλάντωση, όπως συμβαίνει σε ένα σύστημα ελατηρίου μάζας. Στα τύμπανα με κλειστή κοιλότητα, οι δύο μεμβράνες αλληλεπιδρούν μέσω του κελύφους και του αέρα μεταξύ αυτών και αυτή η αλληλεπίδραση, οδηγεί στο σχηματισμό ζευγών τρόπων δόνησης. Για παράδειγμα, οι μεμβράνες παράγουν μαζί δύο τρόπους τάξης (0,1), όπου στην πρώτη περίπτωση κινούνται σε φάση, ενώ στη δεύτερη με αντίθετη φάση (Εικόνα 2-9).

Τέλος, σε πραγματικές συνθήκες θα πρέπει να ληφθούν υπόψη και οι τρόποι δόνησης του κελύφους. Ως κέλυφος χαρακτηρίζεται μια δομή (κοίλοι κύλινδροι, σφαίρες ή τμήματα αυτών) στην οποία η μια διάσταση του υλικού από το οποίο είναι κατασκευασμένο είναι

²Helmholtz, H.L., 1885. On the sensations of tone as a physiological basis for the theory of music.

πολύ μικρότερη από τις άλλες διαστάσεις του. Στα κρουστά μεμβρανόφωνα όργανα, τα κελύφη λειτουργούν ως σχετικά άκαμπτες υποστηρικτικές δομές για τις μεμβράνες και σύμφωνα με τη μελέτη των Love (1888)³ και Rayleigh (1894)⁴, παρατηρούνται δυο ειδών τρόποι δόνησης: οι χαμηλότεροι εκτεταμένοι τρόποι (extensional) και οι μη εκτεταμένοι ή καμπτικοί τρόποι (non-extensional ή flexural) δόνησης. Συνηθίζεται η μεμβράνη του τύμπανου να κρούεται σε ένα σημείο μακριά από το κέντρο του, ώστε να διεγείρονται οι υψηλότερες αρμονικές και μη συχνότητες καθώς και οι αντίστοιχοι τρόποι τους. Οι δονήσεις αυτές των τοιχωμάτων δε συμβάλλουν αισθητά στον παραγόμενο ήχο, εκτός αν πρόκειται για τύμπανα χωρίς μεμβράνη, στα οποία το κέλυφος διεγείρετε άμεσα χτυπώντας το με μπαγκέτες.



Εικόνα 2-9: Συζευγμένες μεμβράνες. Ζεύγη κανονικών τρόπων δόνησης (0,1) και (1,1).

Πηγή: [19]

Η μελέτη των φαινομένων σε μη ιδανικές συνθήκες, οι αλληλεπιδράσεις των συζευγμένων μεμβρανών και δονούμενου κελύφους είναι εκτός των σκοπών αυτή της εργασίας. Αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας, αποτελεί η περίπτωση της ομοιόμορφης κυκλικής μεμβράνης, σε ιδανικές συνθήκες, η οποία είναι σταθερή στην περιφέρεια της, χωρίς να λαμβάνονται οι επιδράσεις του κελύφους και του αέρα.

2.7 Κατασκευαστικά χαρακτηριστικά κεφαλών τυμπάνων

Σε αυτό το σημείο κρίνεται σκόπιμο να αναφερθούν τα γεωμετρικά και κατασκευαστικά χαρακτηριστικά των κεφαλών τυμπάνων, βάσει των οποίων διαμορφώθηκαν οι τιμές των φυσικών παραμέτρων της δονούμενης μεμβράνης που χρησιμοποιήθηκε κατά την επίλυση του ευθέως και του αντίστροφου προβλήματος.

³ Love, A., 1888. The Small Free Vibrations and Deformation of a Thin Elastic Shell. Philosophical Transactions of The Royal Society B: Biological Sciences. 179. 491-546.

⁴ Rayleigh, J. W. S., 1894. The theory of sound. London: Macmillan.

Οι συνθετικές μεμβράνες, που κυκλοφορούν στο εμπόριο είναι κατασκευασμένες από διάφορα πολυμερή με κυριότερα τα Mylar (πολυεστερικό φιλμ) και Kevlar, ενώ για κάποια κρουστά όργανα, όπως bongos, congas κ.α., χρησιμοποιούνται δέρματα ζώων (μοσχαριού ή κατσίκας) [15]. Τα κύρια πλεονεκτήματα των συνθετικών μεμβρανών είναι η εξαιρετική αντοχή τους κατά τη χρήση, η ανοχή τους στις μεταβολές θερμοκρασίας και υγρασίας και το μικρότερο κόστος κατασκευής. Επιπλέον είναι σταθερές κατά το κούρδισμα και λόγω των πολλών διαφορετικών ειδών πλαστικών δερμάτων, κάθε μεμβράνη μπορεί να έχει διαφορετικό ηχόχρωμα. Η θερμοκρασία αποτελεί σημαντική παράμετρο που συντελεί στο ηχόχρωμα και στην τονικότητα των κρουστών, καθώς η μεταβολή της αλλάζει τις ιδιότητες του υλικού, όπως την επιφανειακή πυκνότητα και επομένως επηρεάζει την ταχύτητα διάδοσης του ήχου σε αυτό.

Οι μοντέρνες πλαστικές κεφαλές τυμπάνων διατίθενται σε διάφορες παραλλαγές [18, 20]. Μπορούν να έχουν μια ή και δύο στρώσεις μεμβρανών, να είναι διαφανείς ή/και επικαλυμμένες (coated), να είναι γεμισμένες με λάδι (oil-filled), αλλά και να φέρουν ενσωματωμένο σύστημα απόσβεσης (dampening system). Η επικάλυψη σε μια κεφαλή λειτουργεί ως ένα είδος αποσβεστήρα που μειώνει τη «διαύγεια» αλλά προσθέτει ένα πιο «ζεστό» χαρακτηριστικό. Λόγω της επικάλυψης, η μεμβράνη έχει πιο τραχιά επιφάνεια, και οι κραδασμοί διασκορπίζονται άνισα. Το γεγονός αυτό μειώνει το συνολικό χρόνο απελευθέρωσης (sustain), αλλάζει τον τόνο του παραγόμενου ήχου και μειώνει το συντονισμό των τυμπάνων. Οι κεφαλές διπλής στρώσης (2-Ply heads) περιλαμβάνουν δύο διαφορετικά δέρματα που είναι συνήθως κολλημένα, με κύριο πλεονέκτημα το μεγαλύτερο έλεγχο των υπερτόνων (overtones). Τέλος, οι κεφαλές με ενσωματωμένο μηχανισμό απόσβεσης, συνήθως ένα μικρό ενσωματωμένο «δακτύλιο απόσβεσης» προσαρμοσμένο στην περιφέρεια της κεφαλής, προσφέρουν λεπτομερή έλεγχο του κουρδίσματος του τυμπάνου, χωρίς να αλλοιώνεται ο χρόνος απελευθέρωσης, δίνοντας στο τύμπανο μεγάλο χρόνο απόσβεσης, αυξημένη τονικότητα και ελαφρώς βαθύτερο ήχο. Αυτές οι παραλλαγές κατά την κατασκευή των μεμβρανών, βοηθούν στον έλεγχο της περιβάλλουσας, της διάρκειας και του χρόνου ενεργοποίησης (attack) του παραγόμενου ήχου, όπως επίσης και στο ηχόχρωμα του καλύπτοντας γενικότερα διάφορες μουσικές ανάγκες.

Οι κεφαλές κατασκευάζονται σε διάφορα μεγέθη, με διαμέτρους, που κυμαίνονται από 6in έως 34in (0.15 - 0.86m) και πάχος μεταξύ 0.1in και 0.5in (0.025 - 0.013m). Η πυκνότητα σε φυσιολογικές συνθήκες υγρασίας και θερμοκρασίας είναι συνήθως μεταξύ 200 και 340 kg/m³. Κάποια τυπικά χαρακτηριστικά για το υλικό Mylar (Type A) είναι: Young's modulus: 0.60-0.78×10⁶psi, Surface Roughness (επιφανειακή τραχύτητα): 38nm, Density (πυκνότητα μάζας): 1.39g/cc, Viscosity (ιξώδες): 0.56dL/g και Yield (κάλυψη ταινίας ανά μονάδα βάρους): 21,000in²/lb [22, 25].

Όσο πιο παχιά είναι η κεφαλή, δηλαδή όσο μεγαλύτερη μάζα χρησιμοποιείται, τόσο περισσότερη ενέργεια διαδίδεται και έτσι οδηγεί σε μεγαλύτερο χρόνο απελευθέρωσης, οπότε και σε πιο έντονο συντονισμό, με αποτέλεσμα, να παράγονται πιο πλούσιοι υπερτόνοι. Αντίθετα, μια λεπτότερη κεφαλή έχει λιγότερη μάζα, διαδίδει τον ήχο γρηγορότερα, οπότε συντονίζεται λιγότερο και έτσι ξεχωρίζει η θεμελιώδης συχνότητα. Ως αποτέλεσμα, παρουσιάζει γρήγορο αισθητό τόνο (pronounced tone), μεγαλύτερο χρόνο εκκίνησης και μικρότερο χρόνο απελευθέρωσης.

2.8 Κούρδισμα κεφαλών τυμπάνων

Οι μουσικοί ιστορικά εφάρμοζαν αυτοσχέδια, κάποιο υλικό, στην επιφάνεια των κεφαλών των κρουστών μεμβρανόφωνων, σε μορφή ταινίας, λωρίδας υφάσματος, πάστας ή και αυτοκόλλητων μαξιλαριών, με σκοπό το κούρδισμα του μουσικού οργάνου. Με την προσθήκη τέτοιων υλικών μπορεί να ελεγχθεί η απόσβεση της δόνησης και έτσι να διαμορφωθεί το ηχόχρωμα και η ένταση του παραγόμενου ήχου και γενικότερα να κουρδιστεί με ακρίβεια το μουσικό όργανο. Σήμερα, για το σκοπό αυτό, πωλούνται πολλά εμπορικά προϊόντα προσφέροντας περισσότερο έλεγχο σε σχέση με τις αυτοσχέδιες μεθόδους. Τα υλικά αυτά τοποθετούνται σε κάποιο ή κάποια σημεία στην κεφαλή του τυμπάνου, σε διάφορα μεγέθη και σχηματισμούς (μοτίβα).

2.8.1 Τρόποι κουρδίσματος & έλεγχος απόσβεσης

Στο εμπόριο υπάρχουν διάφορες προσεγγίσεις για το έλεγχο της απόσβεσης στα τύμπανα [18, 20, 22, 23, 10]. Ο έλεγχος της απόσβεσης μπορεί να γίνει:

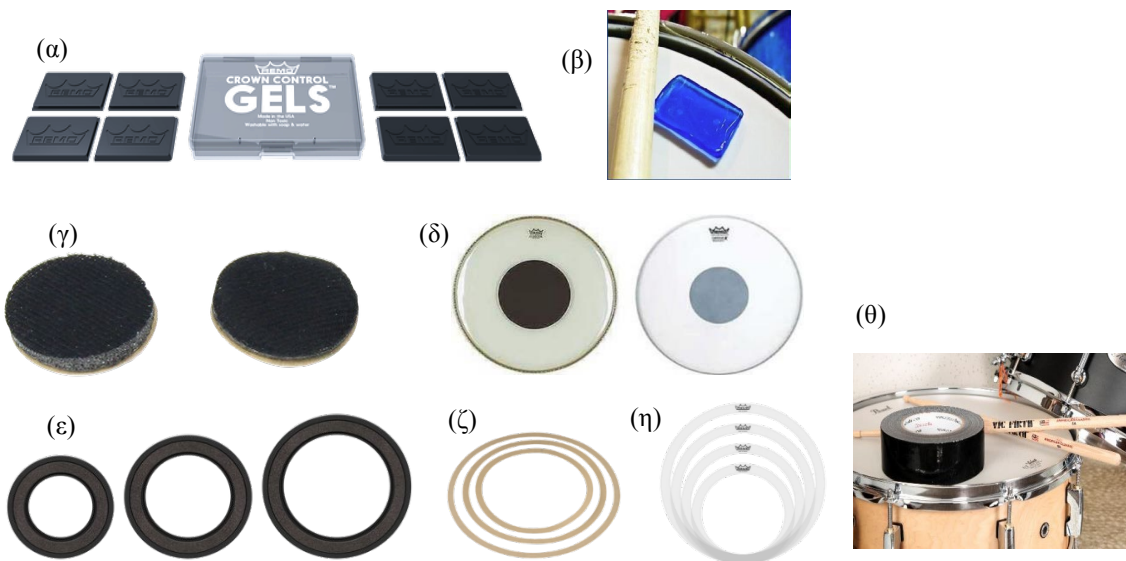
- Εσωτερικά στην κεφαλή, χρησιμοποιώντας είτε δακτυλίους ελέγχου τοποθετημένους στο εσωτερικό της κεφαλής του τυμπάνου, είτε κεφαλές γεμισμένες με λάδι (oil-filled), το οποίο παίζει ενεργό ρολό ως αποσβεστήρας.
- Μέσα στο κέλυφος, τοποθετώντας αφρολέξ μεταξύ του κελύφους και της μεμβράνης του τυμπάνου, ή τοποθετώντας μαξιλαράκια από τσόχα, τα οποία πιέζονται στην κεφαλή μέσω ρυθμιζόμενου μηχανισμού.
- Πάνω στην κεφαλή, με την τοποθέτηση διάφορων υλικών πάνω στη μεμβράνη.

Διάφορα δημοφιλή εμπορικά προϊόντα (Εικόνα 2-10) περιλαμβάνουν:

- Τζελ (Control Gels, Moongels): Μικρά ορθογώνια κομμάτια τζελ που τοποθετούνται πάνω στη μεμβράνη, μεμονωμένα ή σε συνδυασμό, και μπορούν να κοπούν στο

επιθυμητό μέγεθος. Η τοποθέτηση τους καθορίζεται από την επιθυμητή ακτινική απόσταση.

- Μαξιλαράκια από αφρολέξ (Dampening foam dots): Μικρά, εύκαμπτα, αυτοκόλλητα μαξιλαράκια από αφρολέξ, διαμέτρου συνήθως 1 ίντσας με αυτοκόλλητη βάση.
- Κυκλικοί δίσκοι (Tablatone dots, patches): Στρογγυλά πλαστικά «μπαλώματα», σε μορφή κυκλικών δίσκων με αυτοκόλλητη βάση, σε διάφορα μεγέθη, που τοποθετούνται σε επιλεγμένα σημεία πάνω (dot) ή και στο εσωτερικό (reverse dot) της κεφαλής του τυμπάνου.
- Δαχτυλίδια ελέγχου (Muffle Rings ή O-Rings): Λεπτοί πλαστικοί ή χάρτινοι δακτύλιοι, πλάτους περίπου 2 εκατοστών, που τοποθετούνται ελεύθερα ομοκεντρικά της κεφαλής. Διατίθενται σε σετ για τα πιο συνηθισμένα μεγέθη κεφαλών, ενώ μπορούν να κατασκευαστούν από Mylar ή και από κομμάτια παλιών κεφαλών.
- Ταινία Gaffer: Αυτοκόλλητη υφασμάτινη ταινία, που κολλιέται πάνω στη μεμβράνη, σε διάφορα μεγέθη και πυκνότητες.



Εικόνα 2-10: Εμπορικά προϊόντα για έλεγχο απόσβεσης και κουρδίσματος. (α) Crown Control Gels (β) Moongels (γ) Dampening Foam Dots (δ) Tablatone Dots (ε) Muffl Control Ring (ζ) O Self Adhesive Ring Set (η) RemO Sound Control Rings (θ) Ταινία Gaffer.

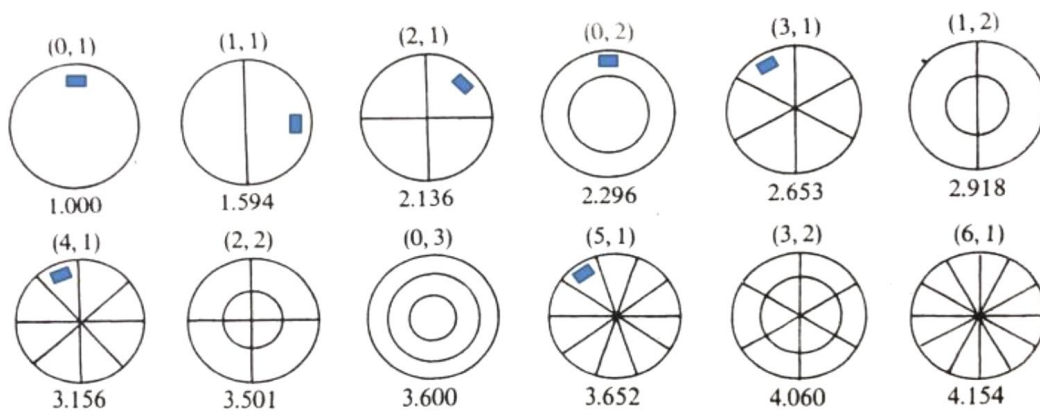
Πηγή: [23]

2.8.2 Μηχανισμοί απόσβεσης

Στη βιβλιογραφία έχουν περιγραφεί μηχανισμοί απόσβεσης για διάφορα τύμπανα, συμπεριλαμβανομένων kettledrums [26, 27, 28], τυμπάνων μπάσων [29, 31], frame drums [30, 1, 2] και snare drums [32]. Με βάση τη μελέτη των Rossing, et al. [32], σε ένα τύμπανο χωρίς πρόσθετους αποσβεστήρες, οι μηχανισμοί απόσβεσης περιλαμβάνουν την ακουστική ακτινοβολία (acoustic radiation), τη διάχυση ενέργειας στο εσωτερικό της μεμβράνης (internal membrane dissipation), καθώς και τις αλληλεπιδράσεις της μεμβράνης με το πλαίσιο στήριξης. Εκτός από κατευθυντική, η διάδοση της ενέργειας ποικίλλει ανάλογα με το σχήμα και τη συχνότητα του κανονικού τρόπου δόνησης. Οι τρόποι δόνησης που περιέχουν μόνο κυκλικούς δεσμούς (π.χ. ο θεμελιώδης) παρουσιάζουν μη μηδενική μετατόπιση όγκου (volume) και τείνουν να διαδίδονται αποτελεσματικά, ενώ αυτοί που περιέχουν μόνο διαμετρικούς δεσμούς έχουν μηδενική μετατόπιση όγκου και καμία μονοπολική ροπή [10]. Η διασταυρούμενη ροή αέρα (cross-flow), παράλληλα με την επιφάνεια της μεμβράνης, μειώνει τη διάδοση καθώς αυξάνεται ο αριθμός των διαμετρικών δεσμών και σχετίζεται με μεγαλύτερο χρόνο απόσβεσης. Οι εσωτερικές απώλειες που σχετίζονται με την κάμψη της μεμβράνης, τείνουν να αυξάνονται με τη συχνότητα και να μειώνουν τους χρόνους απόσβεσης. Επομένως, όλοι οι παραπάνω παράγοντες δρουν συγχρόνως και καθορίζουν το χρόνο απόσβεσης για τον κάθε κανονικό τρόπο δόνησης. Επίσης, σε τύμπανα χωρίς πρόσθετους αποσβεστήρες, το φάσμα συχνοτήτων του παραγόμενου ήχου ποικίλλει ανάλογα με την τάση της μεμβράνης, τον τρόπο διέγερσης, τον χρόνο επαφής και την ακτινική θέση του σημείου διέγερσης. Γενικά, η εμφάνιση των τρόπων δόνησης υψηλότερων συχνοτήτων ευνοείται όταν η μεμβράνη βρίσκεται σε υψηλή τάση και το κτύπημα (διέγερση) είναι έντονο, με μικρό χρόνο επαφής, κοντά στην περιφέρεια του τυμπάνου [19]. Στους κανονικούς τρόπους δόνησης με πολλούς διαμετρικούς δεσμούς, οι κοιλίες (antinodes) εμφανίζονται πιο κοντά στην περιφέρεια της κεφαλής, οπότε διεγείρονται όταν το σημείο κρούσης βρίσκεται κοντά στην περιφέρεια και επομένως αποσβένονται πιο εύκολα τοποθετώντας εκεί το αποσβεστικό υλικό.

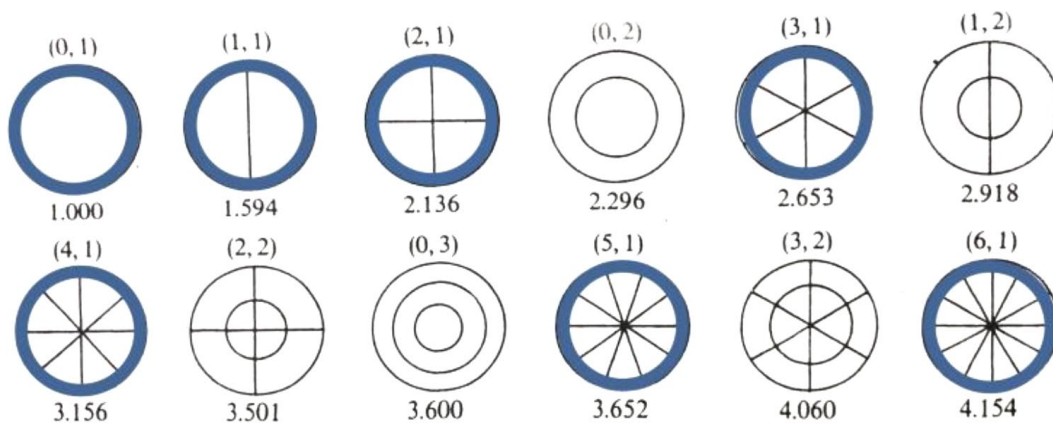
Τα ιξωδοελαστικά υλικά [33], που χρησιμοποιούνται στα προϊόντα απόσβεσης (viscoelastic damping), όπως διάφορα τζελ και μαξιλαράκια, εμφανίζουν μια καθυστέρηση στην παραμόρφωση λόγω της δόνησης, με ένα ποσό ενέργειας να διαχέεται με τη μορφή θερμότητας κατά τη διάρκεια κάθε κύκλου της ταλάντωσης. Τοποθετούνται σε κάποιο ή κάποια σημεία (Εικόνα 2-11) και τεντώνονται, συμπιέζονται και λυγίζουν, ως συνέπεια της επιτάχυνσης και της καμπυλότητας του μέσου στο οποίο προσαρμόζονται. Σε αντίθεση με τα ιξωδοελαστικά μαξιλάρια, τα δακτυλίδια ελέγχου απόσβεσης (από Mylar, πλαστικό ή χαρτί) διατηρούν την κυκλική συμμετρία του

τυμπάνου και για αυτό το λόγο, το παραγόμενο ηχητικό φάσμα, εξαρτάται μόνο από την ακτινική θέση του σημείου διέγερσης. Κατά τη διάρκεια της δόνησης της μεμβράνης, η επιφάνεια της μεμβράνης κάμπτεται και οι δακτύλιοι αυτοί, χάνουν την επαφή τους με την κεφαλή, δημιουργώντας μια air-layer τύπου απόσβεσης (squeeze-film ή gas pumping) [33]. Αυτό συμβαίνει κοντά στις κοιλίες των τρόπων δόνησης και στους διαμετρικούς δεσμούς, όπου οι κινήσεις της μεμβράνης είναι εκτός φάσης μεταξύ τους. Για δακτυλίους κοντά στην περίμετρο της κεφαλής του τυμπάνου, αυτές οι κινήσεις μπορεί να προκαλέσουν επιλεκτική απόσβεση των τρόπων δόνησης που έχουν μεγαλύτερο αριθμό διαμετρικών δεσμών, και των οποίων οι κοιλίες είναι πιο κοντά στην περίμετρο (Εικόνα 2-12). Αντιθέτως, έχουν μικρή επίδραση στον (0,1) θεμελιώδη τρόπο δόνησης, καθώς το μεγαλύτερο μέρος της κίνησης είναι κοντά στο κέντρο της κεφαλής.



Εικόνα 2-11: Τυπική τοποθέτηση τζελ (μπλε χρώμα) σε σημείο, σε σχέση με κάποια σχήματα κανονικών τρόπων δόνησης. Τοποθετώντας το υλικό κοντά στο όριο της μεμβράνης, βρίσκεται πιο κοντά στους αντι-δεσμούς των τρόπων (3,1), (4,1), (5,1) από ό,τι των (0,1), (1, 1), (2,1).

Πηγή: [10]



Εικόνα 2-12: Τυπική τοποθέτηση δακτυλίων (μπλε χρώμα) περιμετρικά της στεφάνης.

Οι τρόποι με μεγαλύτερο αριθμό διαμετρικών κόμβων επηρεάζονται περισσότερο από την προσθήκη του υλικού κοντά στην περίμετρο.

Πηγή: [10]

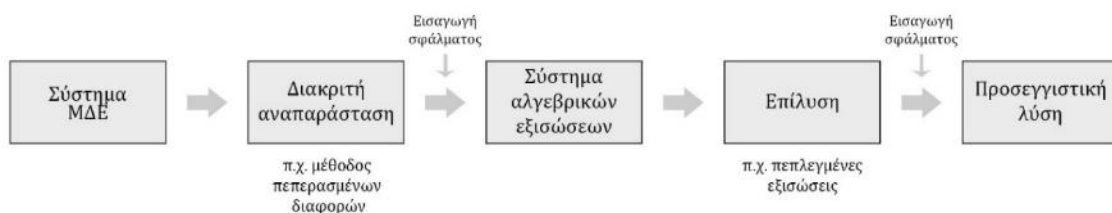
3

Αριθμητικές Μέθοδοι & Μοντελοποίηση

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικές αριθμητικές μέθοδοι που χρησιμοποιούνται για την επίλυση της κυματικής εξίσωσης, εστιάζοντας στη μέθοδο των Πεπερασμένων Διαφορών Χρονικού Τομέα (FDTD).

Η κυματική εξίσωση $\partial^2 u / \partial t^2 = c^2 \nabla^2 u$, όπου ∇^2 ο χωρικός τελεστής Laplace και c μια σταθερά, είναι μια υπερβολική μερική διαφορική εξίσωση, η οποία λειτουργεί ως μαθηματικό μοντέλο για κυματικά φαινόμενα σε συνεχές μέσο. Η επίλυσή της βοηθά τη μελέτη ακουστικών προβλημάτων και τη προσομοίωση φυσικών συστημάτων, όπως για παράδειγμα τη διάδοση των ηχητικών κυμάτων σε μια δονούμενη μεμβράνη. Ένα μαθηματικό πρότυπο μπορεί να εκφραστεί είτε ως συνεχές, μέσω διαφορικών εξισώσεων και να επιλυθεί με αναλυτικές μεθόδους, όπως η μέθοδος χωρισμού μεταβλητών, είτε ως διακριτό, μέσω ενός συστήματος αλγεβρικών εξισώσεων. Για την επίλυση του διακριτού μοντέλου έχουν αναπτυχθεί αριθμητικές μέθοδοι, με το βασικό πλεονέκτημα αυτών να αποτελεί το χαμηλό υπολογιστικό κόστος μέσω της χρήσης H/Y.

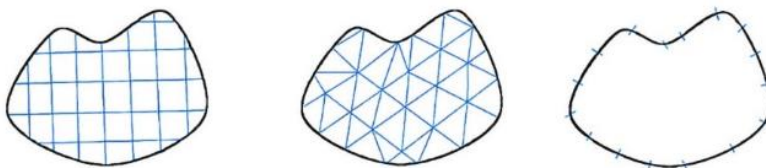
Οι υπολογισμοί και η επίλυση των κυματικών εξισώσεων, σε διακριτή μορφή, πραγματοποιούνται είτε στο πεδίο του χρόνου, είτε της συχνότητας (ή κάποιας άλλης παραμέτρου σχετιζόμενης με την συχνότητα). Η προσομοίωση των προβλημάτων και ακολούθως η αριθμητική επίλυσή τους (Εικόνα 3-1) θεωρητικά, μπορεί να είναι αρκετά λεπτομερής και ακριβής. Ωστόσο, οι απαιτήσεις σε υπολογιστική ισχύ είναι πολλές φορές περιοριστικές για το μέγεθος των προς επίλυση προβλημάτων. Με την εμφάνιση όμως ολοένα και πιο βελτιωμένων αλγόριθμων επίλυσης αλλά και με την περεταίρω ανάπτυξη των H/Y, οι περιορισμοί αυτοί είναι αντιμετωπίσιμοι.



Εικόνα 3-1: Στάδια αριθμητικής επίλυσης.

3.1 Επικρατέστερες οικογένειες υπολογιστικών μεθόδων

Σήμερα, οι επικρατέστερες οικογένειες υπολογιστικών μεθόδων [35 - 40] που χρησιμοποιούνται για την επίλυση των κυματικών εξισώσεων είναι οι μέθοδοι: Πεπερασμένων Διαφορών (Finite Difference Method - FDM), Πεπερασμένων Στοιχείων (Finite Element Method - FEM) και Συνοριακών Στοιχείων (Boundary Element Method - BEM). Για κάθε μια από τις παραπάνω μεθόδους, που στα πλαίσια της υπολογιστικής ακουστικής αποκαλούνται και κυματικές μέθοδοι (wave-based methods), έχουν αναπτυχθεί ξεχωριστές υποκατηγορίες με συγκεκριμένα χαρακτηριστικά αλλά και καταλληλότητα για επιμέρους προβλήματα. Η καταλληλότητα μιας αριθμητικής μεθόδου εξαρτάται από διάφορους παράγοντες όπως, η εύκολη υλοποίηση του αλγορίθμου, το υπολογιστικό κόστος, η ακρίβεια της, η ευστάθειά και η αναστρεψιμότητά της. Η βασική διαφοροποίηση των παραπάνω μεθόδων σχετίζεται κυρίως με την χωρική διακριτοποίηση (πλεγματοποίηση) του ακουστικού μέσου (Εικόνα 3-2). Συνοπτικά, στη FDM το μέσο (ακουστικό πεδίο) χωρίζεται σε ένα κανονικό δίκτυο⁵ (πλέγμα), στους κόμβους του οποίου προσεγγίζονται οι παράγωγοι της διαφορικής εξίσωσης. Στη FEM, το μέσο χωρίζεται σε ένα σύνολο στοιχείων (elements), πάνω στα οποία προσεγγίζονται μέσω συναρτήσεων παρεμβολής τα άγνωστα μεγέθη της διαφορικής εξίσωσης, ενώ η διαφορική εξίσωση αυτή καθαυτή μετασχηματίζεται σε μια ολοκληρωματική μορφή γνωστή ως ασθενής μορφή. Τέλος, στη BEM χρησιμοποιείται μια ολοκληρωματική μορφή εκφρασμένη στο σύνορο του πεδίου, το οποίο είναι και το τμήμα που διακριτοποιείται σε επιμέρους στοιχεία, πάνω στα οποία προσεγγίζονται τα άγνωστα μεγέθη μέσω συναρτήσεων παρεμβολής.



Εικόνα 3-2: Σχηματική απεικόνιση χωρικής διακριτοποίησης αριθμητικών μεθόδων. Αριστερά: μέθοδος FDM με εσωτερικούς κόμβους, Κέντρο: μέθοδος FEM με τριγωνικά στοιχεία, Δεξιά: BEM με συνοριακά στοιχεία.

Άλλες δημοφιλείς μέθοδοι είναι η μέθοδος πεπερασμένων όγκων και οι φασματικές μέθοδοι, ενώ εκτός από την περιγραφή των ακουστικών φαινομένων με βάση το

⁵ Αριθμητικό πλέγμα-δίκτυο (grid ή mesh) ή διακριτοποίηση: Κάθε πεπερασμένο σύνολο σημείων σε ένα διάστημα [a, b]. Μία συνάρτηση που ορίζεται στα σημεία του δικτύου, καλείται συνάρτηση δικτύου.

κυματικό μοντέλο, υπάρχει η δυνατότητα περιγραφής της διάδοσης του ήχου με βάση τη θεωρία σωματιδίων και τη εφαρμογή γεωμετρικών μεθόδων, όπως είναι η μέθοδος της ακτινικής ανίχνευσης (Ray Tracing Method) και η μέθοδος των εικονικών πηγών (Image Source Method) [41], αλλά και με απευθείας προσέγγιση μεγεθών ενέργειας, δηλαδή με μεθόδους στατιστικής ανάλυσης ενέργειας (Statistical energy analysis) [42, 43].

Για τους σκοπούς της παρούσας εργασίας, θα γίνει συνοπτική αναφορά στις επικρατέστερες αριθμητικές μεθόδους και κυρίως στη μέθοδο των πεπερασμένων διαφορών στο χρονικό πεδίο, η οποία χρησιμοποιήθηκε για την παραγωγή των ηχητικών δειγμάτων κατά την πειραματική διαδικασία.

3.1.1 Μέθοδος Πεπερασμένων Διαφορών (FDM)

Η μέθοδος Πεπερασμένων Διαφορών (Finite-Difference Method - FDM) [35, 39, 44] αποδίδεται στον Euler (1707-1783), ο οποίος αναζητούσε προσεγγιστικές λύσεις διαφορικών εξισώσεων και αποτελεί μια από τις πρώτες αριθμητικές μεθόδους επίλυσης που χρησιμοποιήθηκαν για προβλήματα ακουστικής. Αναπτύχθηκε συστηματικά μετά το 1945 όταν οι Η/Υ έγιναν πλέον διαθέσιμοι και συνεχίζει να είναι ιδιαίτερα διαδεδομένη, λόγω της αμεσότητας και της σχετικής ευκολίας της.

Σύμφωνα με την FDM γίνεται αντικατάσταση κάθε παραγώγου της μερικής διαφορικής εξίσωσης με προσεγγιστικά πηλίκα διαφορών, έτσι ώστε το πρόβλημα να μεταβληθεί σε ένα σύστημα αλγεβρικών εξισώσεων. Οι εκφράσεις αυτές ονομάζονται εκφράσεις πεπερασμένων διαφορών και προκύπτουν χρησιμοποιώντας σειρές Taylor ή πολυωνυμική παρεμβολή. Συγκεκριμένα, στην FDM αρχικά διαχωρίζεται (διακριτοποιείται) η επιφάνεια του ακουστικού μέσου (πεδίου), μέσω ενός συνήθως ομοιόμορφου δικτύου (grid ή αριθμητικού πλέγματος - mesh), σε επιμέρους κλιμακωτά διαχωρισμένες περιοχές που ονομάζονται κελιά (cells). Στους κόμβους (nodes) αυτού του δικτύου προσεγγίζονται οι τιμές της άγνωστης συνάρτησης επί της οποίας δρα ο διαφορικός τελεστής. Σύμφωνα με την FDM, κάθε παράγωγος της μερικής διαφορικής εξίσωσης προσεγγίζεται και αντικαθίσταται, με βάση τα αναπτύγματα Taylor περιορισμένων όρων, από πηλίκα διαφορών διακριτών τιμών, που αναφέρονται σε αντίστοιχες θέσεις κόμβων του δικτύου. Προσεγγίζοντας με παρόμοιο τρόπο και τις παραγώγους ως προς την μεταβλητή του χρόνου, προκύπτει η μέθοδος των πεπερασμένων διαφορών στο πεδίο του χρόνου (Finite-Difference Time-Domain method - FDTD). Το τελικό σύστημα εξισώσεων, που αναφέρεται και σε κάθε χρονικό διακριτό σημείο, δημιουργείται από τις επιμέρους εξισώσεις που αντιστοιχούν σε κάθε κόμβο του δικτύου διακριτοποίησης. Το σύστημα των εξισώσεων μπορεί να εκφραστεί

με μια εξίσωση πινάκων (όπου ο πίνακας συντελεστών του συστήματος είναι ένας αραιός, τριδιαγωνικός πίνακας) για την οποία έχουν διαμορφωθεί κατάλληλοι αλγόριθμοι ταχείας επίλυσης.

Η FDM δεν είναι η πιο ακριβής και ευσταθής μέθοδος παρόλα αυτά είναι ευκολότερη στην εφαρμογή της σε χωρία με σχετικά απλή γεωμετρία και αποτελεί τη βάση για την ανάπτυξη πιο περίπλοκων μεθόδων. Σε σύγκριση με άλλες μεθόδους η FDM χαρακτηρίζεται από απλότητα, κυρίως λόγω της ευκολίας προγραμματισμού της χρησιμοποιώντας αλγοριθμικά σχήματα πολύ υψηλής υπολογιστικής αποδοτικότητας και γρήγορης επίλυσης. Ωστόσο, η FDM παρουσιάζει δυσκολίες όταν εφαρμόζεται σε επιφάνειες με πολύπλοκες γεωμετρίες, εξαιτίας της κλιμακωτής χωρικής προσέγγισης του πεδίου. Μπορεί να εφαρμοστεί σε προβλήματα μεταβλητών συντελεστών ή μη γραμμικά προβλήματα, με μικρή υπολογιστική προσπάθεια, παρέχοντας ποσοτικές προσεγγίσεις, αλλά δεν παρέχει ποιοτικές πληροφορίες για τη λύση και είναι δύσκολο να εκτιμηθεί η απόκλιση ανάμεσα στην προσεγγιστική και την ακριβή λύση. Επιπλέον, όταν υπάρχει γρήγορη εναλλαγή τιμών των μεταβλητών, η μέθοδος δεν ανταποκρίνεται καλά λόγω της δυσκολίας που παρουσιάζει στην επιλογή διαφορετικού μεγέθους κελιών.

3.1.2 Μέθοδος Πεπερασμένων Διαφορών Χρονικού Τομέα (FDTD)

Όπως αναφέρεται στη μελέτη των Navarro et al [45], στην αριθμητική μοντελοποίηση ακουστικών δομών, δεδομένου ότι οι οριακές συνθήκες είναι πολλές φορές περίπλοκες, είναι δύσκολο να χρησιμοποιηθεί ένα απλό μοντέλο. Στη βιβλιογραφία είναι διαθέσιμες διάφορες αριθμητικές μέθοδοι για τον εντοπισμό για παράδειγμα συχνοτήτων αποκοπής και κατανομών τρόπων δόνησης (field mode distributions) [46 - 49]. Αυτές όμως οι μέθοδοι έχουν συχνά υψηλό υπολογιστικό κόστος και απαιτούν μεγάλη προσπάθεια προγραμματισμού. Στην πλειοψηφία τους επιλύουν την κυματική εξίσωση μεταφέροντας το πρόβλημα από το πεδίο του χρόνου στο πεδίο συχνοτήτων. Στη συνέχεια, αφού επιλυθεί το πρόβλημα στον τομέα των συχνοτήτων, το αποτέλεσμα πρέπει να μεταφερθεί ξανά πίσω στον τομέα του χρόνου. Σε αντίθεση με αυτές τις μεθόδους, η μέθοδος Πεπερασμένων Διαφορών στον Τομέα του Χρόνου (FDTD) υπολογίζει τη λύση απευθείας στον τομέα χρόνου, ως δεδομένα πραγματικού χρόνου, επιτρέποντας βαθύτερη κατανόηση σε πολύπλοκα προβλήματα.

Η μέθοδος FDTD παρουσιάστηκε για πρώτη φορά από τον Yee το 1966 [50], για τη μελέτη της διάδοσης ηλεκτρομαγνητικών κυμάτων, ως τεχνική απευθείας επίλυσης των εξισώσεων του Maxwell. Η ουσιαστική εφαρμογή της ξεκίνησε στις αρχές της δεκαετίας του '80 με την εισαγωγή των απορροφητικών οριακών συνθηκών από τον Muir το 1981

[51], που έκανε τη μέθοδο κατάλληλη για την επίλυση ενός μεγάλου φάσματος προβλημάτων. Οι Taflove [52], Holland [53] και Umashankar [54], επέκτειναν τον αλγόριθμο και η μέθοδος FDTD εφαρμόστηκε με επιτυχία σε προβλήματα ηλεκτρομαγνητισμού, όπου μέχρι στιγμής, αποτελεί την πιο δημοφιλή τεχνική [55]. Παράλληλα η προσπάθεια για την ανάπτυξη περισσότερο αποτελεσματικών απορροφητικών οριακών συνθηκών κατέληξε στη θεωρία του «τέλεια προσαρμοσμένου στρώματος» (Perfectly Matched Layer - PML) το 1994 [56].

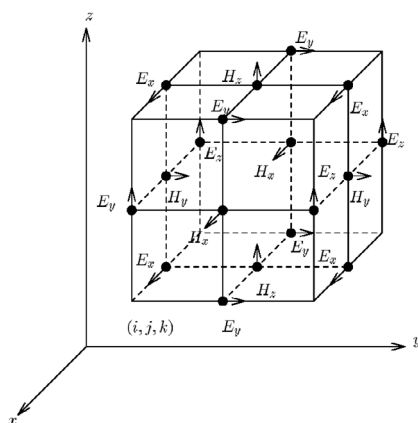
Για την υλοποίηση της μεθόδου FDTD γίνεται αρχικά διακριτοποίηση στο χώρο και στο χρόνο [50] και έπειτα, οι χωρικές και χρονικές παράγωγοι των εξισώσεων του Maxwell προσεγγίζονται από εξισώσεις πεπερασμένων διαφορών, δημιουργώντας ένα σύνολο εξισώσεων πεπερασμένων διαφορών για τις έξι συνιστώσες του ηλεκτρομαγνητικού πεδίου στο τρισδιάστατο χώρο. Οι προκύπτουσες σχέσεις των συνιστωσών του πεδίου, επιλύονται για κάθε «επόμενη» χρονική στιγμή με βάση τις «γνωστές» τιμές των «προηγούμενων» χρονικών βημάτων. Δηλαδή, για τον υπολογισμό του πεδίου σε όλο το χωροχρόνο, δοθέντων των αρχικών οριακών συνθηκών, χρησιμοποιείται ένας αλγόριθμος της μορφής «leapfrog» (τεχνική άλματος - leapfrog technique), γνωστός και ως αλγόριθμος «leapfrog» του Yee (Εικόνα 3-3). Για κάθε χρονικό βήμα οι τιμές του ηλεκτρικού πεδίου υπολογίζονται χρησιμοποιώντας τις τιμές του μαγνητικού πεδίου μισό χρονικό βήμα νωρίτερα και αντίστοιχα για το μαγνητικό. Η διαδικασία αυτή συνεχίζεται μέχρι να ολοκληρωθούν τα χρονικά βήματα που ορίστηκαν στους υπολογισμούς. Οι πεπερασμένες διαφορές που εμφανίζονται κατά την εφαρμογή της μεθόδου είναι κεντρικές και έχουν ακρίβεια δεύτερης τάξης [52].

Μια παρόμοια με την παραπάνω τεχνική χρησιμοποιήθηκε για τη διάδοση ελαστικών κυμάτων κατά την οποία τα ηλεκτρικά και μαγνητικά πεδία αντικαταστάθηκαν από την επιβάρυνση (stress) και την ταχύτητα (velocity) [57, 58]. Η εφαρμογή της FDTD δηλαδή στην ακουστική διάδοση μπορεί να θεωρηθεί ως ειδική περίπτωση της προηγούμενης τεχνικής, για την ανάλυση καταπόνησης και ταχύτητας σε ελαστικά υλικά [59, 60] και στον τομέα της ακουστικής, αντιπροσωπεύει την τεχνική άλματος (leapfrog scheme) που χρησιμοποιείται στη μοντελοποίηση της πίεσης (pressure) και της ταχύτητας [91, 117]. Η FDTD έχει χρησιμοποιηθεί επίσης στην ακουστική σκέδαση [61, 62], στη μελέτη της ακουστικής δωματίων [63, 64, 73], αλλά και για τη μοντελοποίηση της διάδοσης υπερήχων σε ομοιογενές θερμοϊξώδες ρευστό [65].

Η επιλογή της FDTD ως μέθοδος επίλυσης αριθμητικών προβλημάτων παρουσιάζει αρκετά πλεονεκτήματα. Πέρα από την ευκολία της λύσης απευθείας στον τομέα χρόνου, βασικό πλεονέκτημα αποτελεί η ευελιξία σε τροποποιήσεις που προσφέρει η απλότητα των εξισώσεων της μεθόδου FDTD, η οποία επιτρέπει τη συμπερίληψη ιξώδους και μη

γραμμικής επίδρασης [66, 67]. Επίσης, προσφέρει ευκολία στην προγραμματιστική υλοποίηση των σχηματιζόμενων εξισώσεων, αλλά και δυνατότητα προσομοίωσης πολύπλοκων γεωμετριών και χειρισμό αυθαίρετων διεγέρσεων, χωρίς την ανάγκη ειδικής διαχείρισης στα σύνορα του υπολογιστικού πεδίου. Παρέχει ακρίβεια για χαμηλές συχνότητες και γίνεται πιο ακριβής για υψηλότερες συχνότητες όταν χρησιμοποιούνται μικρότερα βήματα κατά τη δημιουργία του δικτύου. Επιπλέον, η FDTD χρησιμοποιεί κεντρικές χωρικές και χρονικές διαφορές για να προσεγγίσει τα παράγωγα, παρέχοντας έτσι ακρίβεια δεύτερης τάξης, σε σύγκριση με άλλες μεθόδους με ακρίβεια πρώτης τάξης.

Σε αντιδιαστολή με τα παραπάνω, η FDTD μπορεί να αποδειχθεί αναποτελεσματική σε προβλήματα πολύ μεγάλων διαστάσεων, λόγω του υψηλού υπολογιστικού κόστους και της εφαρμογής οριακών συνθηκών. Το κόστος αυτό καθορίζεται από την απαιτούμενη χωρική διακριτικότητα, με βάση την οποία οι διαστάσεις των κελιών του πλέγματος, πρέπει να είναι σχετικά μικρές ως προς το μήκος λ του κύματος, συνήθως από $\lambda/20$ έως $\lambda/10$. Το μικρό μήκος συμβάλει ώστε τα σφάλματα αριθμητικής διασποράς να είναι αμελητέα. Επίσης στο υπολογιστικό κόστος, συμβάλει και η επιλογή του χρονικού βήματος, το οποίο πρέπει να είναι μικρότερο από τη συνθήκη Courant, ώστε η άμεση διαδικασία υπολογισμού να είναι ευσταθής. Για την αντιμετώπιση του κόστους και των αδυναμιών της FDTD, έχουν αναπτυχθεί αλγόριθμοι, όπως οι αλγόριθμοι υποδικτύωσης (subgridding algorithm) [69, 70], η ψευδο-φασματική μέθοδος στο πεδίο του χρόνου (Pseudospectral Time-Domain - PSTD) [71] και η μέθοδος πολλαπλής αναλυτικότητας στο πεδίο του χρόνου (Multi Resolution Time-Domain - MRTD) [72]. Επίσης, σε περιπτώσεις όπου η υπολογιστική ισχύς και η μνήμη ενός H/Y δεν επαρκούν, μπορεί εύκολα να υλοποιηθεί παράλληλος υπολογισμός για τον αλγόριθμο FDTD, διαιρώντας το πλέγμα σε τόσα μικρότερα μπλοκ κελιών όσοι είναι οι διαθέσιμοι επεξεργαστές [73]. Τέλος, όπως θα περιγραφεί σε επόμενη ενότητα υπάρχει η δυνατότητα υλοποίησης της FDTD σε αρχιτεκτονική CUDA (Compute Unified Device Architecture) [74, 75], η οποία χρησιμοποιεί την ισχύ των καρτών γραφικών (GPU) της NVIDIA.



Εικόνα 3-3: Το πλέγμα του αλγόριθμου Yee σε τρεις διαστάσεις.

Η μέθοδος FDTD και ο τρόπος που εφαρμόζεται στην παρούσα εργασία, στην περίπτωση της κυματικής εξίσωσης, αναλύεται εκτεταμένα στην ενότητα 3.3 τόσο σε θεωρητικό όσο και σε μαθηματικό επίπεδο.

3.1.3 Μέθοδος Πεπερασμένων Στοιχείων (FEM)

Η μέθοδος των Πεπερασμένων Στοιχείων (Finite Element Method - FEM) εμφανίστηκε στις αρχές τις δεκαετίας του '40 και σημείωσε εντυπωσιακή άνοδο, η οποία οφείλεται κυρίως στην ανάπτυξη των H/Y και των συστημάτων CAD. Χρησιμοποιήθηκε για πρώτη φορά από τον Hrenikoff (1941) σε προβλήματα αεροναυπηγικής και από τον Courant (1943) για την προσεγγιστική επίλυση προβλημάτων τριγωνικής στρέψης. Η καθιέρωσή της οφείλεται στους Turner, Clough, Martin & Topp (1956), Argyris & Kesley (1960) αλλά και στους Zienkiewicz και Chung που δημοσίευσαν το πρώτο βιβλίο πεπερασμένων στοιχείων το 1967. Στη δεκαετία του '60 η FEM χρησιμοποιήθηκε κυρίως, για την προσεγγιστική επίλυση προβλημάτων αντοχής υλικών, μηχανικής ρευστών και μετάδοσης θερμότητας. Στη συνέχεια και μέχρι σήμερα, η FEM έχει δεχθεί βελτιώσεις για την αποτελεσματικότερη εφαρμογή της σε διάφορα προβλήματα [44, 77].

Σύμφωνα με την FEM οποιαδήποτε συνεχής ποσότητα μπορεί να προσεγγιστεί από ένα διακριτό μοντέλο αποτελούμενο από τμηματικά συνεχείς συναρτήσεις που ορίζονται σε έναν αριθμό πεπερασμένων στοιχείων. Συγκεκριμένα, αρχικά γίνεται χωρική διακριτοποίηση του ακουστικού μέσου στο οποίο αναζητείται η λύση μιας διαφορικής εξίσωσης. Διαχωρίζεται δηλαδή η επιφάνεια του πεδίου σε ένα πεπερασμένο σύνολο μη-αλληλεπικαλυπτομένων στοιχείων (elements), απλού συνήθως γεωμετρικού σχήματος, (π.χ. τριγωνικό) και για ένα αριθμό κόμβων και αντίστοιχων βαθμών ελευθερίας. Πάνω σε αυτά τα στοιχεία, που μπορούν να προσαρμόζονται στο σχήμα του πεδίου και των συνόρων που το περικλείουν, γίνεται η προσέγγιση μέσω συναρτήσεων παρεμβολής των άγνωστων μεγεθών της διαφορικής εξίσωσης. Οι τιμές των άγνωστων συναρτήσεων του προβλήματος στους κόμβους των στοιχείων αποτελούν τους βαθμούς ελευθερίας της υπολογιστικής προσομοίωσης και με βάση αυτούς σχηματίζεται η προσεγγιστική έκφραση για το άγνωστο μέγεθος. Η έκφραση αυτή είναι συνήθως μιας χαμηλής τάξης πολυωνυμική προσέγγιση στο εσωτερικό του κάθε στοιχείου (π.χ. γραμμική) και λόγω του ότι δε μπορεί να εισαχθεί απευθείας στη διαφορική εξίσωση, πρέπει να μετασχηματιστεί σε μια ολοκληρωματική σχέση γνωστή και ως ασθενής μορφή. Η μορφή αυτή προκύπτει είτε με την εύρεση μιας συναρτησιακής (functional), η ελαχιστοποίηση της οποίας καταλήγει στη διαφορική εξίσωση που μας ενδιαφέρει (μέθοδος Rayleigh-Ritz), είτε, συνηθέστερα, με την απευθείας εφαρμογή μιας διατύπωσης σταθμισμένων υπολοίπων (weighted residual). Τέλος, στη διατύπωση του προβλήματος εισάγεται η

προσεγγιστική έκφραση η οποία καταλήγει σε ένα γραμμικό σύστημα εξισώσεων, ως προς τους αγνώστους του προβλήματος, το οποίο επιλύεται με διάφορες τεχνικές όπως με τριγωνική παραγοντοποίηση ή με επαναληπτικές μεθόδους [78].

Η FEM [35, 37] χρησιμοποιείται στη μηχανική κατασκευών, στην ρευστοδυναμική, στη θερμοδυναμική, στην κβαντομηχανική, στον ηλεκτρομαγνητισμό και στην ακουστική, όπου εφαρμόζεται πολύ συχνά για την ανάλυση της συμπεριφοράς των ηχητικών κυμάτων σε κλειστούς χώρους. Η FEM, είναι ευκολά εφαρμόσιμη σε περίπλοκες γεωμετρίες, μπορεί να συμπεριλάβει απορρόφηση στις γύρω επιφάνειες και είναι κατάλληλη για την επίλυση διαφορικών εξισώσεων είτε στο πεδίο του χρόνου ή είτε στο πεδίο της συχνότητας, αποτελώντας πολυ καλή αριθμητική προσέγγιση για την επίλυση ιδιοπροβλημάτων.

3.1.4 Μέθοδος Συνοριακών Στοιχείων (BEM)

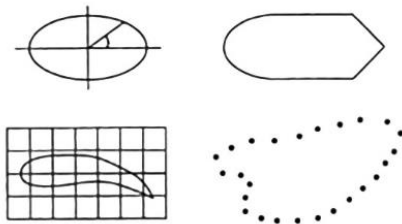
Η μέθοδος των Συνοριακών Στοιχείων (Boundary Element Method - BEM) [35, 79, 80] αποτελεί ισχυρό εργαλείο υπολογισμού, με επάρκεια στην ακρίβεια των αποτελεσμάτων της. Τα πρώτα βήματα για την δημιουργία της μεθόδου έγιναν από τους Somigliana (1886) και Fredholm (1903), ενώ αργότερα με τη διάδοση των H/Y, ο Rizzo (1967) εισήγαγε τον έμμεσο τρόπο υπολογισμού, ο οποίος ακολουθείται μέχρι σήμερα, και ο Cruse (1969) έλυσε πρόβλημα άμεσου τρόπου για τις τρεις διαστάσεις, εφαρμόζοντας διακριτοποίηση της επιφάνειας. Την περίοδο 1967-1972, δημοσιεύθηκαν μια σειρά εργασιών από τους Rizzo, Cruse, Swedlow, Shippy και Van Buren, οι οποίες επέκτειναν τους τομείς υλοποίησης της μεθόδου για διάφορα προβλήματα όπως για προβλήματα με μη ομογενή υλικά, για ελαστοστατικά και ελαστοπλαστικά προβλήματα, για προβλήματα δυναμικού σε ανισότροπα σώματα, καθώς και για προβλήματα θραύσεως σε τρισδιάστατα αντικείμενα [79].

Στα προβλήματα ακουστικής, η BEM εφαρμόζεται για την αριθμητική επίλυση του προβλήματος διατυπωμένου όμως στην ολοκληρωματική του μορφή. Η διαφορική εξίσωση που εκφράζει το σύνορο μετατρέπεται σε ένα ολοκλήρωμα το οποίο καθορίζεται στο σύνορο της επιφάνειας (Boundary Integral Equation) και σε ένα ολοκλήρωμα το οποίο συνδέει τα αποτελέσματα που παράγονται πάνω στο σύνορο με οποιοδήποτε σημείο εντός της προς εξέταση περιοχής. Οι άγνωστες τιμές των μεταβλητών, συνήθως όροι πίεσης ή ταχύτητας, περιορίζονται στα σύνορα της επιφάνειας του ακουστικού μέσου. Όταν το ηχητικό πεδίο είναι γνωστό στα σύνορα, μπορεί να ανακτηθεί, ως διαδικασία μετά-επεξεργασίας (post-processing), σε κάθε σημείο του μέσου και για κάθε χρονική στιγμή ή συχνότητα. Ο προσδιορισμός του συνόρου (Εικόνα 3-4) μπορεί να γίνει

με την παραμετροποίηση συναρτήσεων κωνικών τομών, με χρήση στοιχείων γεωμετρίας που αφορούν διάφορα τμήματα του συνόρου, με μετρήσεις μέσω του μηχανικού σχεδίου του χωρίου κ.α. Στην διακριτοποίηση συναντώνται τόσο δισδιάστατα όσο και τρισδιάστατα μοντέλα, ενώ η απλούστερη μορφή σύνδεσης των κομβικών σημείων ενός συνόρου θεωρείται η γραμμική, κατά την οποία η σύνδεση γίνεται με ευθείες. Οι συνοριακές επιφάνειες, οι οποίες είναι καμπύλες γραμμές στην δυσδιάστατη περίπτωση, διαχωρίζονται σε ένα πεπερασμένο αριθμό στοιχείων (συνοριακά στοιχεία), πάνω στα οποία εκφράζεται η κατανομή των άγνωστων μεταβλητών. Τα στοιχεία διακρίνονται, ανάλογα με τον αριθμό των κόμβων που χρησιμοποιούνται. Η πιο απλή μορφή στοιχείου στις τρεις διαστάσεις είναι το τριγωνικό, αποτελούμενο κατ' ελάχιστο από τρία κομβικά σημεία. Σε κάθε στοιχείο γίνονται δύο προσεγγίσεις, μια για τη γεωμετρία του συνόρου στο συγκεκριμένο σημείο και μια για την κατανομή της άγνωστης συνοριακής ποσότητας στο στοιχείο. Στη συνέχεια υπολογίζονται οι συντελεστές αλληλεπίδρασης μεταξύ των στοιχείων, με βάση τις εξισώσεις που διέπουν το πρόβλημα αλλά και των συνοριακών συνθηκών που θα πρέπει να ικανοποιούνται. Έτσι δημιουργείται ένα σύστημα εξισώσεων για κάθε διακριτή χρονική στιγμή ή για κάθε διακριτή συχνότητα και η επίλυση του προσδιορίζει τα άγνωστα μεγέθη στα σύνορα. Οι πιο γνωστοί τρόποι για να παραχθούν οι απαιτούμενες εξισώσεις είναι η άμεση και η έμμεση μέθοδος υπολογισμού [82, 83]. Η αύξηση των άγνωστων μεγεθών (βαθμών ελευθερίας), οδηγεί σε σημαντική αύξηση της απαιτούμενης υπολογιστικής ισχύος και αυτό προκύπτει όταν η χωρική διακριτοποίηση γίνεται ολοένα και πιο πυκνή ώστε να επιτευχθεί η σύγκλιση των αποτελεσμάτων. Χρησιμοποιώντας τα γνωστά πλέον συνοριακά μεγέθη γίνεται η αναπαραγωγή του ακουστικού πεδίου σε οποιοδήποτε χωρικό σημείο της επιφάνειάς του και για οποιαδήποτε διακριτή χρονική στιγμή ή συχνότητα.

Η BEM μπορεί να διατυπωθεί τόσο στο πεδίο του χρόνου όσο και σε αυτό της συχνότητας. Εφαρμόζεται σε προβλήματα εσωτερικού ή εξωτερικού τύπου και ανάλογα τις τιμές που δίνονται γνωστές, για την ταχύτητα ή την πίεση σε κάθε σημείο του συνόρου, μπορεί να διακριθεί σε υποπεριπτώσεις (προβλήματα τύπου Dirichlet, Neumann, μικτού τύπου και Robin) [80]. Βασικό μειονέκτημα της BEM, έναντι των FEM και FDM, αποτελεί το γεγονός ότι η μετατροπή του ακουστικού προβλήματος από τη διαφορική στην ολοκληρωματική του μορφή, μπορεί να πραγματοποιηθεί μόνο για συγκεκριμένες τάξεις διαφορικών εξισώσεων, σε αντίθεση με τις άλλες δυο μεθόδους που μπορούν να προσαρμοσθούν. Στα προβλήματα όμως που μπορεί να εφαρμοστεί η BEM, παρουσιάζει σημαντικά πλεονεκτήματα όπως το γεγονός ότι η διακριτοποίηση περιορίζεται στο σύνορο, απλοποιώντας έτσι τη διαδικασία, καθώς υπάρχει μείωση κατά μία των διαστάσεων (μείωση αγνώστων-βαθμών ελευθερίας). Αυτό μεταφράζεται σε μείωση του πλήθους δεδομένων και κατ' επέκταση σε μικρότερους χρόνους

υπολογισμού. Επίσης, για απείρως εκτεινόμενα χωρία, περιπτώσεις που είναι συχνές σε προβλήματα ακουστικής, το πρόβλημα μπορεί να διατυπωθεί ως εξωτερικό χωρίς να παρουσιάζει ιδιαίτερη δυσκολία. Επιπλέον, επιτρέπει τον υπολογισμό της λύσης και των παραγώγων της σε οποιοδήποτε σημείο, χρησιμοποιώντας την ολοκληρωτική παράσταση της λύσης ως μαθηματικό τύπο, ενώ προσφέρεται και για την επίλυση προβλημάτων σε χωρία με γεωμετρικές ανωμαλίες, όπως ρωγμές [37, 38, 39, 40].



Εικόνα 3-4: Διάφοροι τρόποι διακριτοποίησης του συνόρου μιας επιφάνειας.

3.1.5 Ιδιότητες & Σφάλματα μεθόδων

Εφαρμόζοντας τις παραπάνω αριθμητικές μεθόδους επίλυσης παρατηρείται ότι για ορισμένες τιμές χωρικού και χρονικού βήματος, τα αποτελέσματα αποκλίνουν αρκετά από την ακριβή λύση και κάποιες φορές δεν επιτυγχάνεται η προσέγγιση. Αυτό είναι αποτέλεσμα φυσικών περιορισμών αλλά και περιορισμών λόγω H/Y , που πρέπει να λαμβάνονται υπόψη [37, 79, 87].

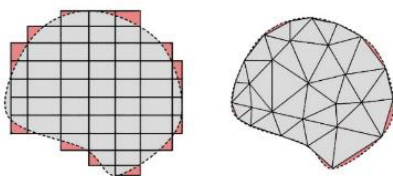
Η διαδικασία της διακριτοποίησης εφαρμόζεται στον χρόνο και στον χώρο και η ακρίβεια των αποτελεσμάτων εξαρτάται σε μεγάλο βαθμό από την πυκνότητα του πλέγματος διακριτοποίησης και από τις ιδιότητες του αλγορίθμου ή του αριθμητικού σχήματος διακριτοποίησης. Το υπολογιστικό πλέγμα καλύπτει ένα ορισμένο χώρο, στα όρια του οποίου απαιτείται ο καθορισμός οριακών συνθηκών. Η επιλογή του αριθμητικού σχήματος-μεθόδου και η επιτυχία της επίλυσης ενός προβλήματος, βασίζεται σε ορισμένες ιδιότητες που πρέπει να ικανοποιεί όπως είναι η ευστάθεια (stability), η συνέπεια (consistency), η σύγκλιση (convergence) και άλλες. Οι ιδιότητες αυτές δεν είναι ανεξάρτητες μεταξύ τους, για παράδειγμα η σύγκλιση της προσεγγιστικής λύσης στην ακριβή λύση, προϋποθέτει τη συνέπεια και την ευστάθεια της μεθόδου (Lax και Richtmyer 1950).

Οι αριθμητικές μέθοδοι δίνουν αποτελέσματα που αποτελούν πάντα προσεγγίσεις κάποιας τάξης της πραγματικής λύσης για το συγκεκριμένο πρόβλημα. Η απόκλιση αυτή

ονομάζεται σφάλμα της αριθμητικής μεθόδου και διαχωρίζεται σε διάφορες κατηγορίες όπως:

- τα σφάλματα προσέγγισης ή αποκοπής, που εμφανίζονται λόγω του πεπερασμένου αριθμού βημάτων που ακολουθούνται σε ένα υπολογισμό. Οφείλεται δηλαδή σε προσεγγίσεις, όπως η παραγραφή όρων που προκύπτουν από την ανάπτυξη του θεωρήματος Taylor.
- τα σφάλματα διακριτοποίησης (ή θέσης) (Εικόνα 3-5), λόγω του ότι κάθε νέα τιμή που υπολογίζεται, βασίζεται σε προηγούμενες προσεγγιστικές τιμές. Τα σφάλματα αυτά εξαρτώνται από το σφάλμα αποκοπής της μεθόδου και μειώνονται με την ελάττωση του μήκους των στοιχείων που χρησιμοποιούνται στη διακριτοποίηση.
- τα σφάλματα επανάληψης, τα οποία παρουσιάζονται όταν χρησιμοποιούνται επαναληπτικές μέθοδοι για την λύση των αλγεβρικών γραμμικών συστημάτων εξισώσεων.
- τα σφάλματα στρογγύλευσης, λόγω του περιορισμένου αριθμού ψηφίων που χρησιμοποιεί ο υπολογιστής στην εκτέλεση πράξεων. Τα σφάλματα αυτά υπάρχουν πάντα και είναι τυχαία.
- τα σφάλματα μοντελοποίησης, που οφείλονται σε απλοποιήσεις της γεωμετρίας της προς εξέτασης επιφάνειας.

Τα σφάλματα προσέγγισης και θέσης τείνουν να εξαλειφθούν όσο αυξάνεται ο αριθμός των διακριτών σημείων του πλέγματος, αλλά σε αυτήν την περίπτωση αυξάνεται το σφάλμα στρογγυλοποίησης καθώς και ο υπολογιστικός χρόνος. Επομένως η επιλογή των διαστημάτων του πλέγματος είναι τελικά προϊόν συμβιβασμού μεταξύ του σφάλματος προσέγγισης και του σφάλματος στρογγυλοποίησης. Επιπλέον, ο αλγόριθμος που επιλέχθηκε και τα αριθμητικά αποτελέσματα, πρέπει να περάσει από μια διαδικασία επαλήθευσης (verification) και πιστοποίησης (validation), ώστε να ελεγχθεί αν όντως επιλύει τις εξισώσεις κίνησης που εξετάζονται αλλά και να γίνει πιστοποίηση της ακρίβειας των αποτελεσμάτων.



Εικόνα 3-5: Σφάλματα διακριτοποίησης (ή θέσης).

3.2 Μέθοδος πεπερασμένων διαφορών για την κυματική εξίσωση

Σε αυτή την ενότητα παρουσιάζονται κάποιες βασικές μαθηματικές έννοιες [35, 37, 44, 88, 89] σχετικά με τις μεθόδους πεπερασμένων διαφορών (FDM), στην περίπτωση της κυματικής εξίσωσης, καθώς και ο τρόπος διακριτοποίησης του πεδίου σε αυτή την περίπτωση.

3.2.1 Προσέγγιση παραγώγων με σχήματα πεπερασμένων διαφορών

Στην περίπτωση ενός δονούμενου δυσδιάστατου μέσου, το σύστημα των εξισώσεων του οποίου θα παρουσιαστεί η αριθμητική λύση είναι το εξής:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + f(x, y, t), \quad x \in (0, L_x), \quad y \in (0, L_y), \quad t \in (0, T] \quad (3.2.1\alpha)$$

$$u(x, y, 0) = I_u(x, y) \quad (3.2.1\beta)$$

$$\frac{\partial u}{\partial t}(x, y, 0) = I_v(x, y) \quad (3.2.1\gamma)$$

$$u(0, 0, t) = d_0(t) \quad \text{ή} \quad u_x(0, 0, t) = n_0(t) \quad (3.2.1\delta)$$

$$u(L_x, y, t) = d_{L_x}(t), \quad u(x, L_y, t) = d_{L_y}(t) \quad \text{ή} \quad u_x(L_x, y, t) = n_{L_x}(t), \quad u_x(x, L_y, t) = n_{L_y}(t) \quad (3.2.1\epsilon)$$

όπου c : η ταχύτητα διάδοσης, (L_x, L_y) : οι διαστάσεις του δυσδιάστατου μέσου, T : ο συνολικός χρόνος απόκρισης, $I_u(x, y)$ & $I_v(x, y)$: οι γνωστές συναρτήσεις των αρχικών συνθηκών, καθορισμένες σε όλο το μήκος x και y , $d_0(t)$, $d_L(t)$ και $n_0(t)$, $n_L(t)$: οι γνωστές στο χρόνο συνοριακές συνθήκες, που αναφέρονται σε συνθήκες τύπου Dirichlet ή Neumann, αντίστοιχα.

Η μερική παράγωγος μιας συνάρτησης $u(x, y, t)$ ως προς τις μεταβλητές x , y και ως προς την μεταβλητή του χρόνου t , αντίστοιχα, ορίζονται από τα ολοκληρώματα:

$$u_x(x, y, t) = \frac{\partial u}{\partial x}(x, y, t) = \lim_{h \rightarrow 0} \frac{u(x+h, y, t) - u(x, y, t)}{h} = u_x \quad (3.2.2)$$

$$u_y(x, y, t) = \frac{\partial u}{\partial y}(x, y, t) = \lim_{h \rightarrow 0} \frac{u(x, y+h, t) - u(x, y, t)}{h} = u_y \quad (3.2.3)$$

$$u_t(x, y, t) = \frac{\partial u}{\partial t}(x, y, t) = \lim_{h \rightarrow 0} \frac{u(x, y, t+h) - u(x, y, t)}{h} = u_t = \dot{u} \quad (3.2.4)$$

Μια προσέγγιση των παραπάνω παραγώγων θα μπορούσε να γίνει αν αντί του απειροστού βήματος h γίνει αντικατάσταση με τα πεπερασμένου μικρού μήκους Δx και

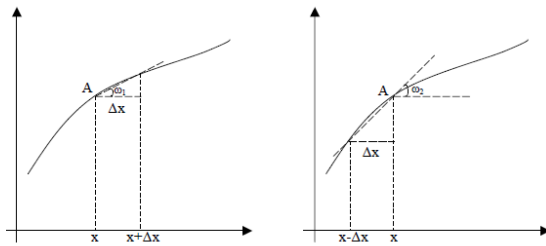
Δy για τη περίπτωση της χωρικής παραγώγου και με το Δt για τη περίπτωση της χρονικής παραγώγου. Η διακριτή εκδοχή τότε των παραπάνω παραγώγων θα είναι:

$$\frac{\partial u}{\partial x} \approx u_x^+ = \frac{u(x + \Delta x, y, t) - u(x, y, t)}{\Delta x} \quad (3.2.5)$$

$$\frac{\partial u}{\partial y} \approx u_y^+ = \frac{u(x, y + \Delta y, t) - u(x, y, t)}{\Delta y} \quad (3.2.6)$$

$$\frac{\partial u}{\partial t} \approx u_t^+ = \frac{u(x, y, t + \Delta t) - u(x, y, t)}{\Delta t} \quad (3.2.7)$$

Τα σχήματα πεπερασμένων διαφορών των εξισώσεων (3.2.5-3.2.7), ονομάζονται πρόδρομες διαφορές (forward difference) (1ης τάξης) (Εικόνα 3-6, αριστερά). Η απόκλιση μεταξύ του πηλίκου διαφοράς και της παραγώγου που προσεγγίζει, αποτελεί το σφάλμα προσέγγισης (truncation error), το οποίο βασίζεται στο ανάπτυγμα της συνάρτησης σε σειρά Taylor, όπου για παράδειγμα, για μικρές τιμές του Δx , οι όροι ανώτερης τάξης ($\Delta x^2, \Delta x^3, \dots$) είναι πολύ μικροί και μπορούν να αγνοηθούν.



Εικόνα 3-6: Γραφική απεικόνιση πρόδρομης (αριστερά) και ανάδρομης (δεξιά) προσέγγισης της παραγώγου u_x σε ένα σημείο A.

Με παρόμοιο τρόπο μπορούν να προκύψουν και να χρησιμοποιηθούν οι ανάδρομες διαφορές (backward difference) (1ης τάξης) (Εικόνα 3-6, δεξιά):

$$\frac{\partial u}{\partial x} \approx u_x^- = \frac{u(x, y, t) - u(x - \Delta x, y, t)}{\Delta x} \quad (3.2.9)$$

$$\frac{\partial u}{\partial y} \approx u_y^- = \frac{u(x, y, t) - u(x, y - \Delta y, t)}{\Delta y} \quad (3.2.10)$$

$$\frac{\partial u}{\partial t} \approx u_t^- = \frac{u(x, y, t) - u(x, y, t - \Delta t)}{\Delta t} \quad (3.2.11)$$

Μια άλλη προσέγγιση της πρώτης παραγώγου είναι μέσω κεντρικής διαφοράς (central difference) για την παράγωγο ως προς t και ως προς x και y αντίστοιχα:

$$\frac{\partial u}{\partial x} \approx u_x = \frac{u(x + \Delta x, y, t) - u(x - \Delta x, y, t)}{2\Delta x} \quad (3.2.12)$$

$$\frac{\partial u}{\partial y} \approx u_y = \frac{u(x, y + \Delta y, t) - u(x, y - \Delta y, t)}{2\Delta y} \quad (3.2.13)$$

$$\frac{\partial u}{\partial t} \approx u_t = \frac{u(x, y, t + \Delta t) - u(x, y, t - \Delta t)}{2\Delta t} \quad (3.2.14)$$

Τέλος, ακολουθώντας παρόμοια λογική, μπορεί να γίνει η εξαγωγή προσεγγιστικών σχημάτων πεπερασμένων διαφορών για τις παραγώγους μεγαλύτερης τάξης, για παράδειγμα 2^{ης} τάξης, μέσω κεντρικής διαφοράς, ως προς t, x και y αντίστοιχα:

$$\frac{\partial^2 u}{\partial x^2} \approx u_{xx} = \frac{u(x + \Delta x, y, t) - 2u(x, y, t) + u(x - \Delta x, y, t)}{\Delta x^2} \quad (3.2.15)$$

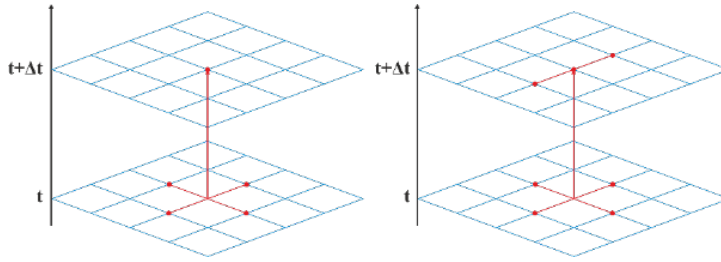
$$\frac{\partial^2 u}{\partial y^2} \approx u_{yy} = \frac{u(x, y + \Delta y, t) - 2u(x, y, t) + u(x, y - \Delta y, t)}{\Delta y^2} \quad (3.2.16)$$

$$\frac{\partial^2 u}{\partial t^2} \approx u_{tt} = \frac{u(x, y, t + \Delta t) - 2u(x, y, t) + u(x, y, t - \Delta t)}{\Delta t^2} \quad (3.2.17)$$

Όπως αναφέρθηκε στην ενότητα 3.1.2, προσεγγίζοντας τις διαφορετικές εξισώσεις με πεπερασμένες διαφορές τόσο ως προς τις μεταβλητές του χώρου όσο και ως προς τη μεταβλητή του χρόνου, προκύπτει η μέθοδος των Πεπερασμένων διαφορών στο πεδίο του χρόνου (FDTD). Η επιλογή του τρόπου προσέγγισης (πρόδρομες, ανάδρομες ή κεντρικές διαφορές, ενός ή μισού βήματος) εξαρτάται από την περίπτωση του προβλήματος που εξετάζεται. Στη συνέχεια, στην ενότητα 3.3, θα γίνει αναλυτική περιγραφή του τρόπου που υλοποιήθηκε στην παρούσα εργασία.

3.2.2 Ρητές & πεπλεγμένες μέθοδοι επίλυσης

Μπορούν να χρησιμοποιηθούν διαφορετικά προσεγγιστικά σχήματα πεπερασμένων διαφορών στο χώρο και στο χρόνο, ανάλογα την περίπτωση, με διαφορετικά σφάλματα προσέγγισης [35, 37, 44, 88, 89]. Η επιλογή τους εξαρτάται άμεσα με τη φυσική και τον τύπο του προβλήματος που εξετάζεται. Οι προσεγγίσεις στις οποίες σε κάθε τύπο υπάρχει μόνο ένας όρος που πρέπει να υπολογιστεί στο επόμενο χρονικό βήμα, με βάση τους διαθέσιμους όρους των προηγούμενων χρονικών επιπέδων, ονομάζονται ρητές (explicit - άμεσες) ως προς το χρόνο (Εικόνα 3-7). Για παράδειγμα ο συνδυασμός πρόδρομης διαφοράς ως προς το χώρο και το χρόνο ή πρόδρομης διαφοράς ως προς το χρόνο και ανάδρομης ως προς χώρο κ.α. Αντίθετα, πεπλεγμένες (έμμεσες - implicit ή μη ρητές ή compact schemes) μέθοδοι επίλυσης, ονομάζονται οι προσεγγίσεις όπου σε κάθε τύπο περιέχονται περισσότεροι από ένα όροι που πρέπει να υπολογιστούν στο επόμενο χρονικό βήμα. Δηλαδή χρησιμοποιούν τόσο το επόμενο σημείο όσο και το προηγούμενο για να υπολογίσουν μια τιμή σε ένα σημείο του πλέγματος. Τυπικά είναι πιο ευσταθείς από τις ρητές μεθόδους αλλά υλοποιούνται πιο δύσκολα. Τα μη ρητά σχήματα σε αντίθεση με τα ρητά απαιτούν την λύση συστήματος εξισώσεων για την προσέγγιση παραγώγου με πεπερασμένες διαφορές. Πεπλεγμένες μέθοδοι επίλυσης αποτελούν ο συνδυασμός ανάδρομης διαφοράς ως προς το χώρο και το χρόνο ή ανάδρομης διαφοράς ως προς το χρόνο και κεντρικής ως προς το χώρο κ.α.



Εικόνα 3-7: Ρητές μέθοδοι (αριστερά) - Πεπλεγμένες (δεξιά)

3.2.3 Διατύπωση πεπερασμένων διαφορών μέσω πίνακα

Οι μέθοδοι πεπερασμένων διαφορών μπορούν να εκφραστούν εναλλακτικά χρησιμοποιώντας πίνακες [35, 37]. Η διατύπωση αυτή διευκολύνει ιδιαίτερα τις πεπλεγμένες μεθόδους. Με την εφαρμογή της διακριτοποιημένης εξίσωσης σε όλους του κόμβους του δικτύου προκύπτει ένα σύστημα αλγεβρικών εξισώσεων που είναι εύχρηστο να εκφραστεί μέσω πινάκων. Η μορφή μιας πεπλεγμένης μεθόδου πεπερασμένων διαφορών χρησιμοποιώντας πίνακες είναι:

$$AU^{j+1} = D \quad (3.2.18)$$

όπου, U^{j+1} : πίνακας στήλη που αποτελείται από τις ζητούμενες προσεγγιστικές λύσεις, A : πίνακας συντελεστών των αγνώστων, D : πίνακας που περιλαμβάνει τις γνωστές τιμές της συνάρτησης στα προηγούμενα χρονικά επίπεδα και τους σταθερούς όρους.

Για ένα διακριτοποιημένο πρόβλημα που αποτελείται από N κόμβους ως προς τη χωρική διάσταση, οι πίνακες U^{j+1} και D θα είναι $N \times 1$ πίνακες, ενώ ο πίνακας A θα είναι $N \times N$ πίνακας.

Λόγω του πλήθους των κόμβων που απαιτούνται για επαρκή ακρίβεια της μεθόδου, το παραπάνω σύστημα θα ήταν περίπλοκο να επιλυθεί. Αυτό όμως δεν ισχύει γιατί ο πίνακας A είναι αραιός, δηλαδή περιλαμβάνει σχετικά λίγα μη μηδενικά στοιχεία και ειδικά για διαφορικές εξισώσεις 2^{15} τάξης αποδεικνύεται ότι είναι τριδιαγωνικός, δηλαδή τα μη μηδενικά στοιχεία του συγκεντρώνονται μόνο στη κύρια διαγώνιο και τις δύο παράλληλες της γραμμές πάνω και κάτω από αυτή, όπως φαίνεται παρακάτω:

$$A = \begin{bmatrix} b_1 & c_1 & & & & & \\ a_2 & b_2 & c_2 & & & & \\ & a_3 & b_3 & c_3 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & a_{N-1} & b_{N-1} & c_{N-1} & \\ & & & & a_N & b_N & \end{bmatrix} \quad (3.2.19)$$

Αυτό έχει ως αποτέλεσμα τη δέσμευση λιγότερης μνήμης αφού αποθηκεύονται μόνο $3 \times N$ στοιχεία αντί για $N \times N$ κατά τον ορισμό του πίνακα A .

Η επίλυση ενός τέτοιου συστήματος γίνεται είτε με απευθείας μεθόδους είτε με επαναληπτικές μεθόδους. Χρησιμοποιώντας απευθείας μεθόδους, όπως η μέθοδος απαλοιφής Gauss ή παραλλαγές της όπως η μέθοδος Cramer και η ανάλυση LU, η λύση προσεγγίζεται μετά από ορισμένο αριθμό βημάτων με αποτέλεσμα να υπεισέρχονται σφάλματα στρογγυλοποίησης. Οι επαναληπτικές μέθοδοι, όπως η μέθοδος Jacobi, η μέθοδος Gauss-Jordan και η μέθοδος Gauss-Seidel, βασίζονται στην διαδοχική βελτίωση των αρχικών προσεγγίσεων της λύσης [88].

Σε συστήματα που οι πίνακες αποτελούνται από πολλούς όρους, οι άμεσες μέθοδοι δεν προτιμώνται λόγω του σφάλματος στρογγυλοποίησης, της μειωμένης ταχύτητας και του μεγάλου ποσού μνήμης που απαιτείται.

3.2.4 Δημιουργία αριθμητικού πλέγματος & διακριτοποίηση

Με την εφαρμογή της μεθόδου πεπερασμένων διαφορών στις μερικές διαφορικές εξισώσεις, υπολογίζονται οι λύσεις μόνο σε ορισμένα σημεία του πεδίου τιμών, έτσι ώστε το πρόβλημα να επιλύεται με πεπερασμένη διαδικασία. Επομένως για την επίλυση ενός τέτοιου προβλήματος ορίζεται στο επίπεδο xyt ένα αριθμητικό πλέγμα (grid), (Εικόνα 3-8), δηλαδή ένα σύνολο σημείων, το οποίο αποτελεί τη διακριτοποιημένη αναπαράσταση του γεωμετρικού πεδίου, πάνω στο οποίο θα επιλυθεί το πρόβλημα [37, 88]. Η διαδικασία αντικατάστασης της αναλυτικής διαφορικής εξίσωσης και του συνεχούς πεδίου ορισμού της με ένα σύστημα εξισώσεων πεπερασμένων διαφορών που ορίζονται στους κόμβους του πλέγματος ονομάζεται διακριτοποίηση. Επομένως, για να πραγματοποιηθεί η διακριτοποίηση, το χρονικό διάστημα $[0, T]$ στο οποίο θα υπολογιστεί αριθμητικά η λύση της εξίσωσης, εκφράζεται ως μια σειρά πεπερασμένου αριθμού διακριτών χρονικών σημείων:

$$0 = t_0 < t_1 < \dots < t_{N_t-1} < t_{N_t} = T \quad \text{ή αλλιώς } t_j \in [0, T], \text{ με } j=0, 1, \dots, N_t \quad (3.2.20)$$

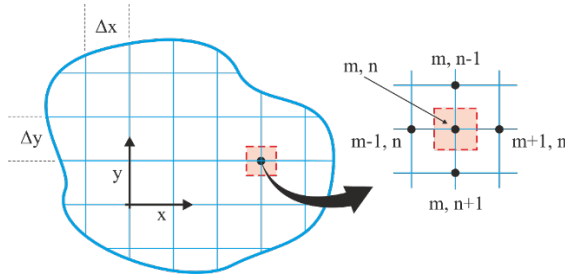
Με όμοιο τρόπο, το συνεχές χωρικό πεδίο του δυσδιάστατου μέσου, αντικαθίσταται από:

$$0 = x_0 < x_1 < \dots < x_{N_x-1} < x_{N_x} = L_x \quad \text{ή αλλιώς } x_m \in [0, L_x], \text{ με } m=0, 1, \dots, N_x \quad (3.2.21)$$

$$0 = y_0 < y_1 < \dots < y_{N_y-1} < y_{N_y} = L_y \quad \text{ή αλλιώς } y_n \in [0, L_y], \text{ με } n=0, 1, \dots, N_y \quad (3.2.22)$$

Επομένως, προκύπτει ένα δίκτυο σημείων στο χώρο και στο χρόνο, κάθε ένα από τα οποία δίνονται από:

$$(x_m, y_n, t_j), \text{ με } m=0, 1, \dots, N_x, n=0, 1, \dots, N_y \text{ και } j=0, 1, \dots, N_t.$$



Εικόνα 3-8: Αριθμητικό πλέγμα - δίκτυο (2D) σημείων στο χώρο.

Πάνω σε αυτό το δίκτυο εκφράζονται όλες οι μεταβλητές αλλά και η άγνωστη συνάρτηση που θα αναπαρίσταται από τη διακριτή της προσέγγιση. Για παράδειγμα η $u_h(x_m, y_n, t_j)$ αποτελεί την προσέγγιση της u στο χωρικό σημείο (x_m, y_n) και στον χρόνο t_j . Σημειώνεται επίσης ότι για $m=0$ ή N_x και $n=0$ ή N_y ισχύουν οι συνοριακές συνθήκες που συνοδεύουν την κυματική εξίσωση και για $j=0$ ισχύουν οι αντίστοιχες αρχικές συνθήκες.

Ένα ομοιόμορφο δίκτυο κόμβων αποτελείται από ισαπέχοντα διαδοχικά σημεία, που σημαίνει ότι για τη χρονική μεταβλητή θα ισχύει:

$$t_{j+1} - t_j = \Delta t = k, \text{ σταθερό για κάθε } j \geq 0 \text{ με } j=0, 1, \dots, N_t \quad (3.2.23)$$

ή αλλιώς $t_j = j\Delta t = j \cdot k$: ο χρόνος μετά από j χρονικά βήματα

και $T = N_t \Delta t \Leftrightarrow \Delta t = \frac{N_t}{T}$., όπου N_t : το πλήθος χρονικών βημάτων

Ομοίως για τις χωρικές μεταβλητές:

$$x_{m+1} - x_m = \Delta x, \text{ σταθερό για κάθε } m \geq 0, m=0, 1, \dots, N_x \quad (3.2.24)$$

$$y_{n+1} - y_n = \Delta y, \text{ σταθερό για κάθε } n \geq 0, n=0, 1, \dots, N_y \quad (3.2.25)$$

ή αλλιώς $x_m = m\Delta x$ και $y_n = n\Delta y$.

$$\text{και } L_x = N_x \Delta x \Leftrightarrow \Delta x = \frac{N_x}{L_x}, \quad L_y = N_y \Delta y \Leftrightarrow \Delta y = \frac{N_y}{L_y}$$

όπου N_x, N_y : πλήθος χωρικών βημάτων στους x, y

Κάθε σημείο (x_m, y_n, t_j) ονομάζεται κόμβος, με (x_0, y_0, t_0) το σημείο αναφοράς. Τα θετικά $\Delta x, \Delta y, \Delta t$ αποτελούν τα διαστήματα του πλέγματος (χωρικά και χρονικά βήματα) στους άξονες x, y και t .

Η επιλογή του μεγέθους του χωρικού και του χρονικού βήματος είναι πολύ σημαντική για την ακρίβεια αλλά και για τη σταθερότητα της μεθόδου που θα εφαρμοστεί, στην προκειμένη περίπτωση της μεθόδου FDTD [44, 89]. Για τα χωρικά βήματα ($\Delta x, \Delta y$), προκειμένου να έχουμε ακριβή αποτελέσματα στην προσομοίωση, πρέπει να γίνει επιλογή του μεγέθους τους ώστε να είναι πολύ μικρότερα από το μικρότερο μήκος κύματος λ που χρειάζεται να μελετηθεί. Επομένως για τη χωρική δειγματοληψία

απαιτείται να υπάρχουν αρκετά σημεία πλέγματος, αρκετά κοντά μεταξύ τους, ώστε το κύμα να μπορεί να δειγματίσει σε τουλάχιστον 5 σημεία. Οι συνηθισμένες επιλογές κυμαίνονται από 1/5λ έως 1/10λ. Όσο μικρότερο είναι το μέγεθος, τόσο πιο ακριβή θα είναι τα αποτελέσματα αλλά με κόστος στην υπολογιστική ταχύτητα.

Επίσης οι τιμές της χωρικής δειγματοληψίας (Δx , Δy) και της περιόδου δειγματοληψίας (Δt) (χρονικό βήμα), πρέπει να ικανοποιούν τη συνθήκη Courant⁶. Σύμφωνα με τα αποτελέσματα της ανάλυσης του Neumann⁷, ο τύπος που εξασφαλίζει τη σταθερότητα σε ένα μοντέλο FDTD, γνωστός ως συνθήκη Courant (στην περίπτωση του δυσδιάστατου μέσου) είναι:

$$c\Delta t \leq \frac{1}{\sqrt{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2}} \quad (3.2.26)$$

όπου c : η ταχύτητα του ήχου στο μέσο.

Ενώ αν $\Delta x = \Delta y$ τότε:

$$\Delta t \leq \frac{\Delta x}{c \cdot \sqrt{2}} \quad (3.2.27)$$

Μετά τη διακριτοποίηση, το νέο πεδίο ορισμού του προβλήματος αποτελεί το υπολογιστικό πλέγμα, του οποίου δομικά στοιχεία είναι τα επιλεγέντα κομβικά σημεία. Για κάθε κόμβο του, διατυπώνεται μια αλγεβρική εξίσωση που περιλαμβάνει την τιμή της εξαρτημένης μεταβλητής στο σημείο και σε γειτονικά σημεία του εντός του πλέγματος. Η αλγεβρική αυτή εξίσωση ονομάζεται εξίσωση πεπερασμένων διαφορών και αποτελεί προσέγγιση της μερικής διαφορικής εξίσωσης στο συγκεκριμένο κομβικό σημείο. Η ακρίβεια της μεθόδου, δηλαδή η διαφορά ανάμεσα στην υπολογιστική (αριθμητική) και πραγματική (αναλυτική αν υπάρχει) λύση εξαρτάται από τη μεθοδολογία των πεπερασμένων διαφορών που θα χρησιμοποιηθεί και μπορεί να αξιολογηθεί μελετώντας την σύγκλιση, την ευστάθεια και την συνοχή του αριθμητικού σχήματος.

Ενδεικτικά, εφαρμόζοντας προσεγγίσεις κεντρικών διαφορών ως προς το χρόνο και το χώρο, η δυσδιάστατη κυματική εξίσωση στη διακριτή της γενική μορφή γράφεται:

$$\frac{\partial^2 u}{\partial t^2}(x_m, y_n, t_j) = c^2 \left(\frac{\partial^2 u}{\partial x^2}(x_m, y_n, t_j) + \frac{\partial^2 u}{\partial y^2}(x_m, y_n, t_j) \right) + f(x_m, y_n, t_j) \quad (3.2.28)$$

⁶ Courant, R.; Friedrichs, K.; Lewy, H. (March 1967) [1928], "On the partial difference equations of mathematical physics", IBM Journal of Research and Development, 11 (2): 215–234, Bibcode:1967IBMJ...11..215C, doi:10.1147/rd.112.0215, MR 0213764, Zbl 0145.40402.

⁷ Analysis of Numerical Methods by E. Isaacson, H. B. Keller

και για κάθε σημείο (x_m, y_n, t_j) θα ισχύει:

$$\frac{u_{m,n}^{j+1} - 2u_{m,n}^j + u_{m,n}^{j-1}}{\Delta t^2} = c^2 \left(\frac{u_{m+1,n}^j - 2u_{m,n}^j + u_{m-1,n}^j}{\Delta x^2} + \frac{u_{m,n+1}^j - 2u_{m,n}^j + u_{m,n-1}^j}{\Delta y^2} \right) + f_{m,n}^j \quad (3.2.29)$$

Στην περίπτωση ύπαρξης απόσβεσης [35], πρέπει να συμπεριληφθεί ένας όρος πρώτης παραγώγου ως προς το χρόνο $(D\partial u/\partial t)$ στην κυματική εξίσωση και έπειτα να γίνουν οι προσεγγίσεις και να διαμορφωθούν κατάλληλα οι εξισώσεις που αφορούν τις συνοριακές και αρχικές συνθήκες. Επίσης, υπάρχουν περιπτώσεις όπου οι διαφορικές εξισώσεις έχουν μεταβλητούς συντελεστές παραγώγων. Για παράδειγμα, περιπτώσεις μη ομογενούς μέσου, δηλαδή περιπτώσεις με μεταβλητή πυκνότητα και άρα και μεταβλητή ταχύτητα διάδοσης του κύματος. Η ταχύτητα τότε θα είναι συνάρτηση της χωρικής συντεταγμένης, δηλαδή $c = c(x, y)$. Σε αυτές τις περιπτώσεις διαμορφώνονται κατάλληλα σχήματα πεπερασμένων διαφορών, εκφράζοντας και τους γνωστούς μεταβλητούς συντελεστές, και πιθανώς κάποιες παραγώγους αυτών, με τις διακριτές τιμές πάνω στο δίκτυο των κόμβων. Επιπλέον, στη περίπτωση «ανοικτών» συνόρων, θα πρέπει να πληρούνται οι συνθήκες ακτινοβολίας (radiation conditions) στα σύνορα, δηλαδή σε αυτά τα σημεία θα πρέπει τα κύματα να εγκαταλείπουν το υπολογιστικό πεδίο και να μην ανακλώνται επιστρέφοντας σε αυτό.

Τέλος, για την εφαρμογή πεπερασμένων διαφορών στην περίπτωση της διάδοσης κύματος σε κυκλική μεμβράνη, η κυματική εξίσωση πρέπει πρώτα να εκφραστεί στο (r, ϕ) επίπεδο, με πολικές συντεταγμένες, ως εξής:

$$\frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \phi^2} \quad (3.2.30)$$

Το βασικό επαναληπτικό μοτίβο που προκύπτει από την προσέγγιση της διαφορικής εξίσωσης με χρήση κεντρικών πεπερασμένων διαφορών δίνεται από την εξίσωση:

$$u_{m,n}^{j+1} = c^2 \Delta t^2 \left[\frac{1}{2r_{m,n} \Delta r} (u_{m+1,n}^j - u_{m-1,n}^j) + \frac{1}{\Delta r^2} (u_{m-1,n}^j + u_{m+1,n}^j - 2u_{m,n}^j) + \frac{1}{r_{m,n}^2 \Delta \phi^2} (u_{m,n-1}^j + u_{m,n+1}^j - 2u_{m,n}^j) \right] - u_{m,n}^{j-1} + 2u_{m,n}^j \quad (3.2.31)$$

Ο αριθμός Courant σε αυτή τη περίπτωση είναι: $C = 2c\Delta t / \Delta r \Delta \phi$

και η αντίστοιχη συνθήκη ευστάθειας του αλγόριθμου θα είναι: $C \leq 1 \Rightarrow \Delta t \leq \Delta r \Delta \phi / 2c$

3.3 Μοντελοποίηση & εφαρμογή FDTD μεθόδου

Σε αυτή την ενότητα περιγράφεται η διαδικασία με την οποία προκύπτει η κυματική εξίσωση της κίνησης της δονούμενης μεμβράνης στο μοντέλο που εφαρμόστηκε στην παρούσα εργασία και αναλύονται οι αρχικές και συνοριακές συνθήκες που την συνοδεύουν. Στη συνέχεια περιγράφεται η διακριτοποίηση της κυματικής εξίσωσης και η εφαρμογή της FDTD μεθόδου χρησιμοποιώντας ένα NSV-leapfrog αλγόριθμο.

3.3.1 Κυματική εξίσωση δονούμενης μεμβράνης

Με βάση τη βιβλιογραφία [35, 37, 44, 88] και την έρευνα του Bader [1, 2, 3], κατά την εφαρμογή της δύναμης κρούσης F_{total} στο αδρανές σύστημα της κυκλικής, ομοιόμορφης, σταθερής στην περιφέρεια μεμβράνης, από το 2ο νόμο του Newton ισχύει:

$$\begin{aligned} F_{total} &= m \cdot a \\ F_R - T_{friction} &= m \cdot a \end{aligned} \quad (3.3.1)$$

όπου F_R : η δύναμη επαναφοράς, $T_{friction}$: η τριβή, m : η μάζα της μεμβράνης και a : η επιτάχυνση

Για τη δύναμη επαναφοράς ισχύει:

$$F_R = c^2(a_x + a_y) \quad (3.3.2)$$

όπου c : ταχύτητα του ηχητικού κύματος και a_x, a_y : τα δυναμικά στο x και y άξονα.

Τα a_x, a_y , είναι οι δεύτεροι παράγωγοι μετατόπισης ως προς το χώρο, τα οποία στο εξής θα αναφέρονται ως δυναμικά (potentials ή curvatures). Τα δυναμικά συνεισφέρουν στη δύναμη επαναφοράς, ενεργώντας στην κατεύθυνση z (κάθετη στην επιφάνεια της μεμβράνης) και επομένως αθροίζονται για να προκύψει η δύναμη επαναφοράς F_R .

Για την ταχύτητα c του ηχητικού κύματος στη μεμβράνη ισχύει:

$$c = \sqrt{\frac{T}{\mu}} \quad (3.3.3)$$

όπου $T(x,y)$: η τάση (προένταση) της μεμβράνης (Nt) και $\mu(x,y)$: η επιφανειακή της πυκνότητα (kg/m^2), για την οποία ισχύει: $\mu(x,y) = m(x,y)/A$, με m : τη μάζα μεμβράνης (kg) και A : το εμβαδόν της μοναδιαίας επιφάνειας (m^2).

Για την τριβή ισχύει:

$$T_{friction} = D \cdot v \quad (3.3.4)$$

όπου D : η σταθερά απόσβεσης και v : η ταχύτητα της κίνησης.

Η απόσβεση της δόνησης της μεμβράνης οφείλεται στην τριβή, η οποία εξαρτάται από την ταχύτητα της κίνησης.

Επομένως, με βάση τα παραπάνω προκύπτει:

$$\begin{aligned} F_{\text{total}} &= m \cdot a \\ F_R - T_{\text{friction}} &= m \cdot a \\ c^2(a_x + a_y) - D \cdot v &= m \cdot a \\ \frac{T}{\mu}(a_x + a_y) &= m \cdot a + D \cdot v \end{aligned}$$

Δηλαδή:

$$\frac{T(x,y)}{\mu(x,y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} \quad (3.3.5)$$

όπου: $u(x,y,t)$: μετατόπιση (m), $T(x,y)$: τάση (προένταση) (Nt), $\mu(x,y)$: επιφανειακή πυκνότητα (kgf/m^2) και D : σταθερά απόσβεσης (ο όρος απόσβεσης εξαρτάται από τη ταχύτητα: $D \partial u / \partial t$)

Η εξίσωση (3.3.5) αποτελεί την αναλυτική μορφή της κυματικής εξίσωσης για το μοντέλο της μεμβράνης που χρησιμοποιείται στην παρούσα εργασία.

Επίσης ισχύουν: $x \in [0, L_x]$, $y \in [0, L_y]$, $t \in [0, T]$, όπου T : ο συνολικός χρόνος απόκρισης και L_x, L_y : οι διαστάσεις του δυοδιάστατου μέσου, ενώ για την κυκλική μεμβράνη επίσης ισχύει: $x^2 + y^2 = R^2$ και $L_x = L_y = R$

Τέλος, για την εφαρμογή της μεθόδου FDTD, η κυματική εξίσωση δεν χρειάζεται μετατροπή σε πολικές συντεταγμένες καθώς η εξίσωση (3.3.5) στη διακριτή της μορφή εφαρμόζεται σε ορθογώνιο πλέγμα, όπως περιγράφεται στη συνέχεια στην ενότητα 3.3.2.

Η εξίσωση (3.3.5) είναι μια διαφορική υπερβολική εξίσωση 2ης τάξης, επομένως συνοδεύεται από δύο αρχικές συνθήκες, όπου επιπλέον της εξαρτημένης μεταβλητής ορίζεται και η παράγωγός της ως προς το χρόνο, σε κάποια χρονική στιγμή.

Γενικά, οι αρχικές συνθήκες δίνονται μέσω των καθορισμένων σε κάθε σημείο της μεμβράνης συναρτήσεων:

$$u(x,y,0) = I_u(x,y) \quad (3.3.6)$$

$$v(x,y,0) = \frac{\partial u}{\partial t}(x,y,0) = I_v(x,y) \quad (3.3.7)$$

και οι συνοριακές συνθήκες μέσω της:

$$u(x_b, y_b, t) = d_b(t) \quad (3.3.8)$$

για κάθε $t \in [0, T]$, όπου x_b, y_b : όλες οι συνοριακές θέσεις-σημεία

Το κτύπημα-κρούση της μεμβράνης μπορεί να περιγραφεί με μια ώθηση Gauss. Για μικρές μεμβράνες ο παλμός αυτός είναι πιο κοφτός-απότομος, ενώ για μεγαλύτερες είναι πιο μεγάλος-πλατύς. Στον τομέα συχνοτήτων, ο μετασχηματισμός Fourier του παλμού, αντιστοιχεί στην εξαρτώμενη από τη συχνότητα ενέργεια που εισάγεται στο σύστημα. Ένας μεγάλος παλμός έχει μικρότερες εντάσεις στις υψηλές συχνότητες, οπότε ο ήχος θα ακουστεί πιο θαμπός. Λόγω του ότι στη περίπτωση που εξετάζετε θέλουμε τη μεγαλύτερη δυνατή εισαγωγή ενέργειας στο σύστημα, ο παλμός αυτός μπορεί να περιγραφεί με μια συνάρτηση Kroneker Delta (ή Dirac):

$$\delta_i = \begin{cases} 0, & i \neq 0 \\ 1, & i = 0 \end{cases} \quad (3.3.9)$$

Χρησιμοποιώντας την παραπάνω συνάρτηση προκύπτουν όλες οι συχνότητες με μέγιστο πλάτος, αν και αυτό ισχύει μόνο θεωρητικά, για βήμα διακριτοποίησης $\Delta t \rightarrow 0$. Για πεπερασμένα βήματα δημιουργείται ένα διαμορφωμένο φάσμα συχνοτήτων, χωρίς να έχουν δηλαδή όλες οι συχνότητες το ίδιο πλάτος.

Επομένως στην περίπτωση που εξετάζουμε δηλαδή για κυκλική μεμβράνη σταθερή στην περιφέρεια έχουμε:

- Αρχικές συνθήκες:

Για όλα τα σημεία, εκτός του σημείου διέγερσης (κρούσης):

$$u(x, y, 0) = 0 \quad (3.3.10)$$

$$v(x, y, 0) = \frac{\partial u}{\partial t}(x, y, 0) = 0 \quad (3.3.11)$$

Ενώ στο σημείο διέγερσης (x_p, y_p):

$$u(x_p, y_p, 0) = 1 \quad (3.3.12)$$

- Συνοριακές συνθήκες (ομογενείς συνοριακές συνθήκες τύπου Dirichlet):

$$u(x_b, y_b, t) = 0 \quad (3.3.13)$$

με x_b, y_b : όλες οι συνοριακές θέσεις-σημεία

Τέλος, όπως έχει ήδη αναφερθεί για λόγους απλοποίησης της διαδικασίας, δε λαμβάνεται υπόψιν ο αέρας, καθώς θεωρείται αμελητέος στην περίπτωση τυμπάνου με μια μεμβράνη, με στενό, ανοικτό κέλυφος (frame drum), και για τους ίδιους λόγους, δεν λαμβάνεται υπόψιν οι συζευγμένες δονήσεις του πλαισίου-κελύφους του τυμπάνου.

3.3.2 Διακριτοποίηση κυματικής εξίσωσης & περασμένες διαφορές

Παρακάτω περιγράφεται η διαδικασία διακριτοποίησης της κυματικής εξίσωσης και επιπλέον δίνεται η σημειογραφία που ακολουθείται στην παρούσα εργασία.

Για την επίλυση ενός προβλήματος με πεπερασμένες διαφορές, πρέπει να οριστεί ένα αριθμητικό πλέγμα (δίκτυο) στο επίπεδο xy , μέσω του οποίου θα γίνει η διακριτοποίηση της κυματικής εξίσωσης [35, 37, 88, 89]. Στην περίπτωση που εξετάζει η παρούσα εργασία, δηλαδή της κυκλικής μεμβράνης (ακτίνας R , μάζας m , με τάση T και επιφανειακή πυκνότητα μ), θεωρώντας ότι το υλικό της είναι συμμετρικό και ισοτροπικό, καθορίστηκε ένα ομοιόμορφο, τετραγωνικό πλέγμα (Εικόνα 3-9), όπου για το χρονικό βήμα ισχύει: $\Delta t = k$ και για το χωρικό: $\Delta x = \Delta y = h$, με h, k σταθερά και $k \leq h/c\sqrt{2}$ (συνθήκη Courant). Το χρονικό διάστημα $[0, T]$ στο οποίο θα υπολογιστεί αριθμητικά η λύση της εξίσωσης εκφράζεται με την πεπερασμένη σειρά διακριτών χρονικών σημείων:

$$0 = t_0 < t_1 < \dots < t_{N_t-1} < t_{N_t} = T \quad \text{ή αλλιώς } t_j \in [0, T], \text{ με } j=0, 1, \dots, N_t \quad (3.3.14)$$

οπότε για τη χρονική μεταβλητή θα ισχύει:

$$t_{j+1} - t_j = \Delta t = k, \text{ για κάθε } j \geq 0 \text{ με } j=0, 1, \dots, N_t$$

ή αλλιώς: $t_j = j\Delta t = j \cdot k$: ο χρόνος μετά από j χρονικά βήματα

και $T = N_t \Delta t \Leftrightarrow \Delta t = \frac{N_t}{T}$: το μέγεθος του χρονικού βήματος,

όπου N_t : πλήθος χρονικών βημάτων

Το συνεχές χωρικό πεδίο x, y του δυσδιάστατου μέσου αντικαθίσταται από τις πεπερασμένες σειρές διακριτών χρονικών σημείων:

$$0 = x_0 < x_1 < \dots < x_{N_x-1} < x_{N_x} = L_x \quad \text{ή αλλιώς } x_m \in [0, L_x], \text{ με } m=0, 1, \dots, N_x \quad (3.3.15)$$

$$0 = y_0 < y_1 < \dots < y_{N_y-1} < y_{N_y} = L_y \quad \text{ή αλλιώς } y_n \in [0, L_y], \text{ με } n=0, 1, \dots, N_y \quad (3.3.16)$$

όπου στην περίπτωση της κυκλικής μεμβράνης σε τετραγωνικό πλέγμα ισχύει:

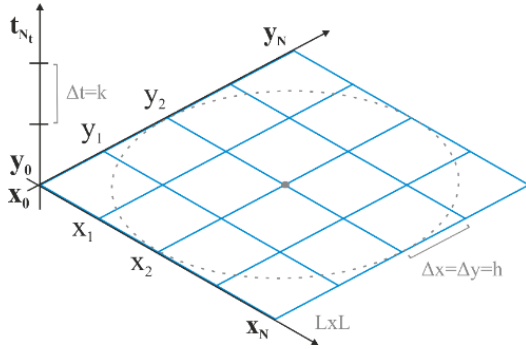
$$L_x = L_y = L = R \text{ (διαστάσεις 2D επιπέδου) και } N_x = N_y = N \text{ (πλήθος χωρικών βημάτων)}$$

Με άλλα λόγια στο ομοιόμορφο, τετραγωνικό πλέγμα $L \times L$, τα χωρικά βήματα στους άξονες x, y , είναι ίσα σε μέγεθος και σε πλήθος. Επομένως, για τις χωρικές μεταβλητές θα ισχύει:

$$x_{m+1} - x_m = \Delta x = h \text{ και } y_{n+1} - y_n = \Delta y = h, \text{ για κάθε } m, n = 0, 1, \dots, N$$

$$\text{ή αλλιώς: } x_m = m\Delta x = mh \text{ και } y_n = n\Delta y = nh ..$$

$$\text{και } L = N\Delta x \Leftrightarrow \Delta x = h = \frac{N}{L}$$



Εικόνα 3-9: Ομοιόμορφο τετραγωνικό δίκτυο κόμβων

Έστω η παρακάτω σημειογραφία και οι συμβολισμοί, για τα μεγέθη που περιλαμβάνει η κυματική εξίσωση (3.3.5) στην αναλυτική της μορφή:

- μετατόπιση (displacement): $u = u(x, y, t)$
- ταχύτητα (velocity): $v = v(x, y, t) = \frac{\partial u}{\partial t} = u_t$
- επιτάχυνση (acceleration) (ως προς το χρόνο): $a = a(x, y, t) = \frac{\partial^2 u}{\partial t^2} = u_{tt}$
- δυναμικό (curvature) κατά τον άξονα x: $a_x = a_x(x, y, t) = \frac{\partial^2 u}{\partial x^2} = u_{xx}$
- δυναμικό (curvature) κατά τον άξονα y: $a_y = a_y(x, y, t) = \frac{\partial^2 u}{\partial y^2} = u_{yy}$

Μετά τη δημιουργία του παραπάνω αριθμητικού πλέγματος και χρησιμοποιώντας τους δείκτες m και n για τη χωρική και το δείκτη j για τη χρονική διαμέριση, η μετατόπιση στη διακριτοποιημένη της μορφή, μπορεί να γραφεί ως:

$$u_{m,n}^j = u(x_m, y_n, t_j)$$

όπου: $t_j = j \cdot k$, με $j = 0, 1, \dots, N_t$, και $x_m = mh$, $y_n = nh$ με $m, n = 0, 1, \dots, N$

Οπότε, χρησιμοποιώντας προσεγγίσεις πεπερασμένων διαφορών και την αντίστοιχη σημειογραφία προκύπτουν:

- μετατόπιση: $u(x, y, t) = u_{m,n}^j$

- ταχύτητα: $v(x,y,t) = u_t = \frac{\partial u}{\partial t} = \frac{u_{m,n}^{j+1} - u_{m,n}^{j-1}}{2\Delta t} = \frac{u_{m,n}^{j+1} - u_{m,n}^{j-1}}{2k}$ (κεντρικές διαφορές)

$$\text{ή } v(x,y,t) = u_t = \frac{\partial u}{\partial t} = \frac{u_{m,n}^{j+1} - u_{m,n}^j}{\Delta t} = \frac{u_{m,n}^{j+1} - u_{m,n}^j}{k} = v_{m,n}^j \text{ (πρόδρομες διαφορές)}$$

- επιτάχυνση (ως προς το χρόνο):

$$a(x,y,t) = u_{tt} = \frac{\partial^2 u}{\partial t^2} = \frac{u_{m,n}^{j+1} - 2u_{m,n}^j + u_{m,n}^{j-1}}{\Delta t^2} = \frac{u_{m,n}^{j+1} - 2u_{m,n}^j + u_{m,n}^{j-1}}{k^2} = a_{m,n}^j$$

(με κεντρικές διαφορές στο χρόνο με βάση τη μετατόπιση)

$$\text{ή } a(x,y,t) = v_t = \frac{\partial v}{\partial t} = \frac{v_{m,n}^{j+1} - v_{m,n}^j}{k} = a_{m,n}^j$$

(με πρόδρομες διαφορές στο χρόνο με βάση την ταχύτητα)

- δυναμικό κατά τον άξονα x και y:

$$a_x(x,y,t) = u_{xx} = \frac{\partial^2 u}{\partial x^2} = \frac{u_{m+1,n}^j - 2u_{m,n}^j + u_{m-1,n}^j}{\Delta x^2} = \frac{u_{m+1,n}^j - 2u_{m,n}^j + u_{m-1,n}^j}{h^2} = a_{x,m,n}^j$$

$$a_y(x,y,t) = u_{yy} = \frac{\partial^2 u}{\partial y^2} = \frac{u_{m,n+1}^j - 2u_{m,n}^j + u_{m,n-1}^j}{\Delta y^2} = \frac{u_{m,n+1}^j - 2u_{m,n}^j + u_{m,n-1}^j}{h^2} = a_{y,m,n}^j$$

(με κεντρικές διαφορές στο χώρο)

Στη συνέχεια περιγράφεται η προσέγγιση της κυματικής εξίσωσης μέσω πεπερασμένων διαφορών χρονικού τομέα (FDTD).

3.3.3 Προσέγγιση κυματικής εξίσωσης μέσω της μεθόδου FDTD

Μπορούν να χρησιμοποιηθούν διαφορετικά προσεγγιστικά σχήματα πεπερασμένων διαφορών στο χώρο και στο χρόνο, (ρητές ή πεπλεγμένες μέθοδοι επίλυσης, είτε κάποιος συνδυασμός) ανάλογα την περίπτωση, με διαφορετικά σφάλματα προσέγγισης (ενότητα 3.2.2). Εδώ χρησιμοποιήθηκε μια μέθοδος πεπερασμένων διαφορών χρονικού τομέα (FDTD), όπως περιγράφεται στα [1, 2, 3] και αναλύεται παρακάτω.

Ένα σύνθητες προσεγγιστικό σχήμα που μπορεί να χρησιμοποιηθεί, με βάση το πλέγμα που περιγράφηκε στην προηγούμενη ενότητα και την αντίστοιχη σημειογραφία, είναι η προσέγγιση της κυματικής εξίσωσης (3.3.5) με μερικές παραγώγους κεντρικών διαφορών για το χώρο και το χρόνο [35, 88, 89]. Με αυτό τον τρόπο προκύπτει η διακριτοποιημένη μορφή της εξίσωσης:

$$\frac{T(x,y)}{\mu(x,y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} \Rightarrow T \left/ \mu \right. (u_{xx} + u_{yy}) = u_{tt} + Du_t \Rightarrow$$

$$\begin{aligned} \frac{T}{\mu} \left(\frac{u(x+\Delta x, y, t) - 2u(x, y, t) + u(x-\Delta x, y, t)}{\Delta x^2} + \frac{u(x, y+\Delta y, t) - 2u(x, y, t) + u(x, y-\Delta y, t)}{\Delta y^2} \right) = \\ = \frac{u(x, y, t+\Delta t) - 2u(x, y, t) + u(x, y, t-\Delta t)}{\Delta t^2} + D \frac{u(x, y, t+\Delta t) - u(x, y, t-\Delta t)}{2\Delta t} \Leftrightarrow \end{aligned}$$

$$\frac{T}{\mu} \left(\frac{u_{m+1,n}^j - 2u_{m,n}^j + u_{m-1,n}^j}{h^2} + \frac{u_{m,n+1}^j - 2u_{m,n}^j + u_{m,n-1}^j}{h^2} \right) = \frac{u_{m,n}^{j+1} - 2u_{m,n}^j + u_{m,n}^{j-1}}{k^2} + D \frac{u_{m,n}^{j+1} - u_{m,n}^{j-1}}{2k}$$

και επιλύοντας ως προς την μετατόπιση $u_{m,n}^{j+1}$ προκύπτει η παρακάτω εξίσωση (3.3.17):

$$u_{m,n}^{j+1} = \underbrace{\left(-\frac{2-kD}{2+kD} \right) u_{m,n}^{j-1}}_{\text{λύση στο χρόνο } t_{j-1}} + \underbrace{\frac{-4(2k^2s-h^2)}{h^2(2+kD)} u_{m,n}^j + \frac{2k^2s}{h^2(2+kD)} (u_{m+1,n}^j + u_{m-1,n}^j) + \frac{2k^2s}{h^2(2+kD)} (u_{m,n+1}^j + u_{m,n-1}^j)}_{\text{λύση στο χρόνο } t_j}$$

Η εξίσωση (3.3.17) αποτελεί ένα σύστημα εξισώσεων, όπου χρησιμοποιώντας τις τιμές της μετατόπισης $u_{m,n}^{j-1}$ και $u_{m,n}^j$ των δύο προηγούμενων χρονικών επιπέδων (εμφανίζονται οι τιμές της εξαρτημένης μεταβλητής σε τρία χρονικά επίπεδα), μπορούν να υπολογιστούν οι μετατοπίσεις σε όλα τα κομβικά σημεία του χωρικού πεδίου και για κάθε χρονικό επίπεδο.

Επιπλέον, αν: $\lambda_1 = \frac{-4(2k^2s-h^2)}{h^2(2+kD)}$, $\lambda_2 = \frac{2k^2s}{h^2(2+kD)}$ και $\lambda_3 = -\frac{2-kD}{2+kD}$

η παραπάνω εξίσωση μπορεί να διατυπωθεί μέσω πινάκων ως εξής:

$$\begin{aligned} \mathbf{u}^{-j+1} = \begin{pmatrix} u_{11}^{j+1} \\ \vdots \\ u_{1N}^{j+1} \\ u_{21}^{j+1} \\ \vdots \\ u_{2N}^{j+1} \\ \vdots \\ \vdots \\ u_{NN}^{j+1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 & 0 & 0 & 0 & \dots & 0 \\ \lambda_2 & \lambda_1 & \lambda_2 & 0 & \lambda_2 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \lambda_1 & 0 & 0 & \lambda_2 & 0 & \dots & 0 \\ \lambda_2 & 0 & 0 & \lambda_1 & \lambda_2 & 0 & \lambda_2 & \dots & 0 \\ 0 & \lambda_2 & 0 & \lambda_2 & \lambda_1 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_2 & 0 & \lambda_2 & \lambda_1 & 0 & \dots & \lambda_2 \\ 0 & 0 & 0 & \lambda_2 & 0 & 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \lambda_2 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & 0 & \dots & \lambda_1 \end{pmatrix} \begin{pmatrix} u_{11}^j \\ \vdots \\ u_{1N}^j \\ u_{21}^j \\ \vdots \\ u_{2N}^j \\ \vdots \\ \vdots \\ u_{NN}^j \end{pmatrix} + \lambda_3 \begin{pmatrix} u_{11}^{j-1} \\ \vdots \\ u_{1N}^{j-1} \\ u_{21}^{j-1} \\ \vdots \\ u_{2N}^{j-1} \\ \vdots \\ \vdots \\ u_{NN}^{j-1} \end{pmatrix} \Rightarrow \\ \Rightarrow \mathbf{u}^{-j+1} = \mathbf{A} \mathbf{u}^{-j} + \lambda_3 \mathbf{u}^{-j-1} \end{aligned}$$

Στην παρούσα εργασία χρησιμοποιήθηκε μια παραλλαγή [1, 2, 3] αυτού του προσεγγιστικού σχήματος ως εξής: Η κυματική εξίσωση (3.3.5) προσεγγίστηκε με μερικές παραγώγους κεντρικών διαφορών για το χώρο όπως και προηγουμένως αλλά με μερικές παραγώγους πρόδρομων διαφορών για το χρόνο. Επιπλέον, η επιτάχυνση

$a(x, y, t) = u_{tt}$ προσεγγίστηκε με πρόδρομες διαφορές 1^{ης} τάξης με βάση την ταχύτητα (και όχι 2^{ης} τάξης με βάση την μετατόπιση):

$$\text{Δηλαδή: } u_t = \frac{\partial u}{\partial t} = \frac{u_{m,n}^{j+1} - u_{m,n}^j}{k} \quad \text{και} \quad u_{tt} = v_t = \frac{\partial v}{\partial t} = \frac{v_{m,n}^{j+1} - v_{m,n}^j}{k}$$

Με βάση τα παραπάνω η κυματική εξίσωση (3.3.5) γίνεται:

$$\begin{aligned} \frac{T(x, y)}{\mu(x, y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) &= \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} \Rightarrow T/\mu (a_x + a_y) = v_t + Du_t \Rightarrow T/\mu (u_{xx} + u_{yy}) = v_t + Du_t \Rightarrow \\ T/\mu \left(\frac{u(x + \Delta x, y, t) - 2u(x, y, t) + u(x - \Delta x, y, t)}{\Delta x^2} + \frac{u(x, y + \Delta y, t) - 2u(x, y, t) + u(x, y - \Delta y, t)}{\Delta y^2} \right) &= \\ = \frac{v(x, y, t + \Delta t) - v(x, y, t)}{\Delta t} + D \frac{u(x, y, t + \Delta t) - u(x, y, t)}{\Delta t} &\Leftrightarrow \\ T/\mu \left(\frac{u_{m+1,n}^j - 2u_{m,n}^j + u_{m-1,n}^j}{h^2} + \frac{u_{m,n+1}^j - 2u_{m,n}^j + u_{m,n-1}^j}{h^2} \right) &= \frac{v_{m,n}^{j+1} - v_{m,n}^j}{k} + D \frac{u_{m,n}^{j+1} - u_{m,n}^j}{k} \quad (3.3.18) \end{aligned}$$

Αυτό το προσεγγιστικό σχήμα είναι 1^{ης} τάξης στο χρόνο και 2^{ης} τάξης στο χώρο ($O[\Delta t, \Delta x^2]$) και αποδεικνύεται με ανάλυση Fourier ότι είναι πάντα ευσταθές για οποιοδήποτε χρονικό βήμα Δt . Με αυτό τον τρόπο υπολογίζονται οι μετατοπίσεις σε όλα τα κομβικά σημεία του χωρικού πεδίου και για κάθε χρονικό επίπεδο του χρονικού διαστήματος που έχει επιλεγεί, χρησιμοποιώντας τιμές της μετατόπισης και της ταχύτητας ($u_{m,n}^j$ και $v_{m,n}^j$), μόνο του προηγούμενου χρονικού επιπέδου. Επίσης, για ευκολία προγραμματισμού, πολλές φορές χρησιμοποιείται η διαμόρφωση μισού βήματος, δηλαδή $h/2$ αντί h και $k/2$ αντί k για την προσέγγιση της παραγώγου [92].

Τα παραπάνω σχήματα καθώς και οι παραλλαγές τους είναι γνωστά στη βιβλιογραφία με διάφορες ονομασίες, όπως μέθοδος Verlet (στη μοριακή δυναμική), μέθοδος Störmer (στην κίνηση ιονισμένων σωματιδίων στο μαγνητικό πεδίο της γης) και μέθοδος Encke (στις πλανητικές τροχιές). Στο πλαίσιο των μερικών διαφορικών εξισώσεων για τη διάδοση κυμάτων, το προσεγγιστικό σχήμα που περιγράφηκε προηγουμένως (καθώς και η παραλλαγή του) είναι γνωστό και ως αλγόριθμος Newton-Störmer-Verlet (NSV) ή μέθοδος Leapfrog (αλλιώς symplectic Euler, semi-explicit Euler, Euler-Cromer) [9, 91].

Τέλος, με βάση τα προηγούμενα, οι συνοριακές και αρχικές συνθήκες (3.3.10-3.3.13) διαμορφώνονται ως εξής:

- Συνοριακές συνθήκες (ομοιόμορφο, τετραγωνικό πλέγμα $L \times L$):

$$u(x_b, y_b, t) = 0 \Rightarrow \begin{cases} u(x_0, y_n, t_j) = u(0, y_n, t_j) = u_{0,n}^j = 0 \\ u(x_N, y_n, t_j) = u(L, y_n, t_j) = u_{L,n}^j = 0 \\ u(x_m, y_0, t_j) = u(x_m, 0, t_j) = u_{m,0}^j = 0 \\ u(x_m, y_N, t_j) = u(x_m, L, t_j) = u_{m,L}^j = 0 \end{cases} \quad (3.3.19)$$

- Αρχικές συνθήκες (οριακές συνθήκες Dirichlet):

Για $m, n = 0, 1, \dots, N$, με $m, n \neq p$:

$$u(x_m, y_n, t_0) = u(x_m, y_n, 0) = u_{m,n}^0 = 0 \text{ και } v(x_m, y_n, t_0) = \frac{\partial u}{\partial t}(x_m, y_n, 0) = v_{m,n}^0 = 0 \quad (3.3.20)$$

Ενώ για το σημείο διέγερσης: (x_p, y_p) με $u(x_p, y_p, 0) = u_{p,p}^0 = 1$

3.3.4 Περιγραφή NSV-leapfrog αλγόριθμοι

Τα επαναληπτικά βήματα για τη λύση της κυματικής εξίσωσης στη μορφή (3.3.18), με βάση τη χρονικά εξαρτημένη μέθοδο NSV, μπορούν να περιγραφούν με τον παρακάτω αλγόριθμο [3, 9, 91, 92]:

- Υπολογισμός της επιτάχυνσης του συστήματος σε κάθε κόμβο για ένα χρονικό επίπεδο (αυτό αποτελεί τη λύση του αριστερού μέρους της εξίσωσης).
- Υπολογισμός της (νέας) ταχύτητας ενός κόμβου, για το επόμενο χρονικό επίπεδο t_{j+1} , χρησιμοποιώντας την τιμή της -παλιάς- ταχύτητας αυτού το κόμβου στο χρόνο t_j .
- Υπολογισμός της (νέας) μετατόπισης αυτού του κόμβου, για το επόμενο χρονικό επίπεδο t_{j+1} , χρησιμοποιώντας την τιμή της -νέας- ταχύτητας στο χρόνο t_{j+1} και της μετατόπισης στο χρόνο t_j για αυτό τον κόμβο (τα βήματα 2 και 3 αποτελούν τη λύση του δεξιού μέρους της εξίσωσης).
- Αποθήκευση της νέας ταχύτητας και της νέας μετατόπισης σε διαφορετικό διάνυσμα από αυτό που χρησιμοποιήθηκε για τον υπολογισμό.
- Αφού έχουν υπολογιστεί όλα τα κομβικά σημεία σε ένα χρονικό επίπεδο, γίνεται αντικατάσταση των παλιών ταχυτήτων και μετατοπίσεων από τις νέες.
- Επιστροφή στο βήμα 1.

Η επαναληπτική διαδικασία επαναλαμβάνεται μέχρι να ολοκληρωθεί ο υπολογιστικός χρόνος που έχει επιλεγεί.

Μαθηματικά ο παραπάνω αλγόριθμος μπορεί να διατυπωθεί με συμβολικό τρόπο:

$$a = \frac{T}{\mu}(u_{xx} + u_{yy})$$

$$v_{\text{new}} = v_{\text{old}} + a\Delta t \Rightarrow v(x,y,t+1) = v(x,y,t) + a \cdot \Delta t \quad (3.3.21)$$

$$u_{\text{new}} = u_{\text{old}} + v_{\text{new}} \Delta t \Rightarrow u(x,y,t+1) = u(x,y,t) + v(x,y,t+1) \cdot \Delta t$$

Η «επιτάχυνση» a του συστήματος, αποτελεί το άθροισμα των δυναμικών (potentials ή curvatures), όπως έχει αναφερθεί στην ενότητα 3.3.1.

Τέλος, λαμβάνεται υπόψιν η απόσβεση του συστήματος, προσθέτοντας στη (νέα) ταχύτητα ένα όρο απόσβεσης D , μετά το 2^ο βήμα ως εξής:

$$v_{\text{new}} = v_{\text{new}} \cdot D \Rightarrow v(x,y,t+1) = v(x,y,t+1) \cdot D \quad (3.3.22)$$

3.3.5 Ιξωδοελαστικό μοντέλο μεμβράνης

Σε αυτό το σημείο πρέπει να αναφερθεί ότι η υλοποίηση του προγράμματος που παραχωρήθηκε από το Καθηγητή Rolf Bader, για τους σκοπούς της παρούσας εργασίας, εφαρμόζει τη μέθοδο πεπερασμένων διαφορών χρονικού τομέα, όπως περιγράφεται στα [3, 9], πάνω σε ένα βισκοελαστικό (ιξωδοελαστικό) μοντέλο κυκλικής μεμβράνης. Λαμβάνει δηλαδή υπόψιν τις βισκοελαστικές ιδιότητες της μεμβράνης και του όποιου αποσβεστικού υλικό προστεθεί πάνω σε αυτή.

Σύμφωνα με την έρευνα του Bader [1, 2], το μοντέλο υλοποιεί την κυματική εξίσωση σε μια μεμβράνη υπό τάση (προένταση) $T(x,y)$, επιφανειακής πυκνότητας μ , σταθερά απόσβεσης D και μετατόπισης u , ως εξής:

$$\frac{T(x,y)}{\mu(x,y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} \quad (3.3.23)$$

Η τάση T εκφράζεται σε Newton (Nt), η μετατόπιση u σε (m) και το μέτρο ελαστικότητας του Young E (Young's modulus) είτε σε Pascal (Pa), είτε σε δύναμη ανά εμβαδόν περιοχή (Nt/m²). Στη βιβλιογραφία χρησιμοποιείται κυρίως το μέτρο ελαστικότητας του Young, ως η αναλογία επιβάρυνσης (stress) προς καταπόνηση (strain) (αναλογία stress-strain). Η διαφορική εξίσωση της μεμβράνης μπορεί να ερμηνευθεί επίσης ως η σχέση stress-strain, ενώ και το ιξωδοελαστικό μοντέλο μπορεί να μελετηθεί χρησιμοποιώντας το μέτρο ελαστικότητας του Young. Η επιφανειακή πυκνότητα $\mu(x,y) = m(x,y)/A$, εκφράζεται ως το πηλίκο της μάζας m (kgr) διαιρούμενη με το εμβαδόν A (m²) της μεμβράνης και εξαρτάται από τη θέση στην οποία μετριέται. Αυτό συμβαίνει γιατί η μάζα της μεμβράνης μπορεί να ποικίλλει ανάλογα με τη θέση που μελετάτε πάνω στην επιφάνεια (κατά μήκος των x και y), καθώς οι μεμβράνες δεν έχουν πάντα ομοιόμορφο πάχος. Επιπλέον, και η τάση $T(x,y)$ έχει χωρική κατανομή, αφού πολλές φορές δεν είναι

δυνατό να ρυθμιστεί ομοιόμορφα η κατανομή τάσης της κεφαλής ενός τυμπάνου σε ολόκληρη την επιφάνειά της. Με αυτό τον τρόπο, το υλοποιημένο μοντέλο επιτρέπει εύκολα πολύπλοκες κατανομές επιφανειακής πυκνότητας και τάσης. Επιπρόσθετα, η επιφανειακή πυκνότητα μπορεί να έχει χωρική κατανομή (spatial distribution) ανάλογα και με την ύπαρξη ή όχι πάστας (tuning paste) ή γενικότερα κάποιου πρόσθετου υλικού απόσβεσης στην επιφάνεια της. Στο [1] λαμβάνεται υπόψιν η ύπαρξη πάστας, από ιξωδοελαστικό υλικό και γίνεται η υπόθεση ότι η μάζα διπλασιάζεται με την προσθήκη πάστας σε ένα σημείο, σε σύγκριση με την περίπτωση χωρίς πάστα ως εξής:

$$\mu(x,y) = m(x,y) / A = \begin{cases} \mu_0 & \text{for } \sqrt{(x-x_0)^2 + (y-y_0)^2} > 0.1nr \\ 2\mu_0 & \text{for } \sqrt{(x-x_0)^2 + (y-y_0)^2} < 0.1nr \end{cases} \quad (3.3.24)$$

με $n=0,1,2,\dots,10$ και κέντρο μεμβράνης το σημείο (x_0, y_0) .

Έτσι, διαμορφώνεται μια κυκλική περιοχή πάστας που μπορεί να καλύπτει τη μεμβράνη από μηδενική ακτίνα (χωρίς κάλυψη) έως πλήρη ακτίνα (πλήρη κάλυψη), διαιρούμενη σε δέκα βήματα ίσης ακτίνας. Στο κεφάλαιο 5, γίνεται αναλυτική περιγραφή του τρόπου που συμπεριλήφθηκε η περίπτωση προσθήκης υλικού απόσβεσης (πάστας) στην επιφάνεια της μεμβράνης, κατά το πειραματικό μέρος αυτής της εργασίας.

Η διαφορική εξίσωση (3.3.23) περιλαμβάνει έναν όρο απόσβεσης D , ο οποίος οδηγεί σε εκθετική απόσβεση τις ιδιοσυχνότητες του τυμπάνου, τόσο στο χρόνο όσο και στη συχνότητα. Έτσι, κάθε μερική συχνότητα (partial) φθίνει εκθετικά και το φάσμα του ήχου παρουσιάζει εκθετική πτώση στις υψηλότερες συχνότητες. Ωστόσο οι πειραματικές μετρήσεις δεν παρουσιάζουν τόσο απλή συμπεριφορά, επιβεβαιώνοντας τη βιβλιογραφία που δείχνει ότι η ιξωδοελαστική απόσβεση έχει ως αποτέλεσμα ένα φασματικό εύρος ζώνης (spectral band-gap) και οι πολλαπλές αποσβεσμένες ζώνες καταλήγουν σε ένα πολύπλοκο φάσμα πλάτους. Αυτό οφείλεται κυρίως λόγω της ύπαρξης εσωτερικής απώλειας ενέργειας εντός του υλικού της μεμβράνης αλλά και της εξωτερικής απόσβεσης που προκαλείται από απώλεια ενέργειας λόγω ακτινοβολίας. Επιπλέον, υπάρχει μια χρονική υστέρηση μεταξύ επιβάρυνσης και καταπόνησης, όπως και ένα είδος «μνήμης» καθώς οι προηγούμενες δονήσεις επηρεάζουν την τωρινή.

Για να ληφθούν τα παραπάνω υπόψιν, σύμφωνα με τον Bader [1, 2], η εσωτερική απόσβεση μπορεί να εκφραστεί ως ένα μοντέλο ενός σύνθετου και εξαρτώμενου από τη συχνότητα συντελεστή ελαστικότητας Young $E(s)$. Λόγω του ότι στην περίπτωση της μεμβράνης δεν υπάρχει μέτρο ελαστικότητας Young, μεταφέρεται η ιδέα, αντικαθιστώντας την καταπόνηση με τη χωρική διαφοροποίηση της μεμβράνης και το συντελεστή Young, με τη δύναμη πάνω στην επιφανειακή πυκνότητα. Επομένως και

σύμφωνα με τα παραπάνω, η εξίσωση της ιξωδοελαστικής μεμβράνης που προκύπτει είναι:

$$\int_{\tau=0}^T h(\tau) \frac{T}{\mu(x,y)} \left(\frac{\partial^2 u(x,y,t-\tau)}{\partial x^2} + \frac{\partial^2 u(x,y,t-\tau)}{\partial y^2} \right) = \frac{\partial^2 u(x,y,t)}{\partial t^2} + D \frac{\partial u(x,y,t)}{\partial t} \quad (3.2.34)$$

όπου $h(\tau)$ είναι μια συνάρτηση που αντιπροσωπεύει το χρονικό πεδίο του μιγαδικού συντελεστή Young $E(s)$.

Για τους σκοπούς της παρούσας εργασίας και για να απλοποιηθούν οι πειραματικές διαδικασίες, το υλικό της μεμβράνης θεωρήθηκε ομοιογενές και για την απόσβεση χρησιμοποιήθηκε ένας σταθερός όρος απόσβεσης D και η πρώτη χρονική παράγωγος μετατόπισης (εξίσωση 3.3.23). Στο πειραματικό μέρος της εργασίας, μετά τον επαναπροσδιορισμό του αντίστροφου προβλήματος, η όποια προσθήκη υλικού απόσβεσης στην επιφάνεια της μεμβράνης υλοποιήθηκε απλά ως αύξηση της μάζας της μεμβράνης στο συγκεκριμένο σημείο, χωρίς δηλαδή βισκοελαστικές ιδιότητες. Επιπλέον, για λόγους απλοποίησης, δε λαμβάνεται υπόψιν η επίδραση του αέρα, αλλά και η επίδραση του δονούμενου κελύφους, καθώς δεν αποτελούν αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας.

4

Μηχανική Μάθηση & Τεχνητά Νευρωνικά Δίκτυα

4.1 Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning - ML), προέρχεται από το ευρύ πεδίο της τεχνητής νοημοσύνης [93]. Χρησιμοποιεί ένα σύνολο μεθόδων στατιστικής ανάλυσης και έχει σκοπό τον προγραμματισμό υπολογιστικών συστημάτων με τέτοιο τρόπο ώστε να μπορούν να προσομοιώνουν ανθρώπινες λογικές λειτουργίες. [94, 209]. Η ιδέα της μηχανικής μάθησης και συγκεκριμένα της επαγωγικής μηχανικής μάθησης (Inductive Learning algorithms), βασίζεται στο γεγονός ότι οι «αισθήσεις» ενός συστήματος μπορεί να χρησιμοποιηθούν όχι μόνο για να διαμορφώσουν μια άμεση αντίδραση αλλά και για να συμβάλουν στη βελτίωση της ικανότητάς του να ενεργεί αποτελεσματικά στο μέλλον [210]. Για τα συστήματα που ανήκουν στη συμβολική Τεχνητή Νοημοσύνη (TN), η δυνατότητα μάθησης προσδιορίζεται ως η ικανότητα απόκτησης επιπλέον γνώσης, η οποία επιφέρει μεταβολές στην υπάρχουσα καταχωρημένη γνώση, είτε αλλάζοντας κάποια χαρακτηριστικά της, είτε αυξομειώνοντάς την. Στην περίπτωση των συστημάτων που ανήκουν στη Μη Συμβολική TN, όπως στην περίπτωση των Τεχνητών Νευρωνικών Δικτύων, ως δυνατότητα μάθησης προσδιορίζεται η δυνατότητα που διαθέτουν τα συστήματα στο να μετασχηματίζουν την εσωτερική τους δομή, παρά στο να μεταβάλλουν τη γνώση που έχει καταχωρηθεί μέσα σε αυτά κατά το σχεδιασμό τους.

Η βασικότερη φάση κάθε αλγόριθμου μηχανικής μάθησης είναι η εκπαίδευση, όπου ο αλγόριθμος χρησιμοποιεί ως είσοδο, ένα σύνολο δεδομένων εκπαίδευσης προς επίτευξη του σκοπού του, δηλαδή τη δημιουργία νέας γνώσης. Έπειτα ακολουθεί η φάση της πιστοποίησης της παραγόμενης νέας γνώσης, η οποία πραγματοποιείται από τον ίδιο τον αλγόριθμο, με τη βοήθεια δεδομένων ελέγχου, μέσω διαδικασιών ανάκλησης (recall).

Τέλος, η νέα γνώση δίνεται προς χρήση σε εφαρμογές, οι οποίες την χρησιμοποιούν για να λυθούν πραγματικά προβλήματα.

Σήμερα, οι αλγόριθμοι μηχανικής μάθησης βρίσκουν εφαρμογή σε ένα πλήθος προβλημάτων όπως στην αναγνώριση ομιλίας και γραπτού λόγου (Speech & Text Recognition), στην αναγνώριση αντικειμένων σε εικόνες (Object Recognition & Detection), στην επεξεργασία, κατανόηση και μοντελοποίηση φυσικών γλωσσών (Natural Language Processing, Understanding & Generation), στις μηχανές αναζήτησης (Search Engines), στην αναπαράσταση γνώσης (Knowledge Representation), στην αυτόματη πλοήγηση, στην ιατρική, στη βιοπληροφορική, στη ρομποτική κ.α.

4.1.1 Κατηγορίες Μηχανικής Μάθησης

Στα συστήματα ML, ο προσδιορισμός της φύσης του μαθησιακού προβλήματος αποτελεί βασικό παράγοντα στη σχεδίαση ενός στοιχείου μάθησης. Κατά συνέπεια, τα προβλήματα της ML, αλλά και οι αλγόριθμοι που χρησιμοποιούνται στις αντίστοιχες περιπτώσεις, διαχωρίζονται σε κατηγορίες ανάλογα με το επιθυμητό αποτέλεσμα ή/και τον τύπο εισόδου που χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης. Οι πιο δημοφιλείς κατηγορίες [95, 96] είναι:

- Επιτηρούμενη ή επιβλεπόμενη Μάθηση (Supervised Learning): Στο υπολογιστικό σύστημα δίνονται παραδείγματα εισόδου και της επιθυμητής εξόδου-αποτελέσματος (σύνολο εκπαίδευσης - training set). Τα παραδείγματα εισόδου αποτελούν δεδομένα στα οποία έχουν προηγουμένως ανατεθεί ετικέτες (labels). Ο στόχος είναι η εξαγωγή ενός γενικού κανόνα αντιστοίχισης των εισόδων στις επιθυμητές εξόδους. Με αυτή τη διαδικασία χαρτογράφησης, εκτιμάται μια συνάρτηση πρόγνωσης (predictor function), όπου κάθε ζεύγος παραδειγμάτων-αποτελεσμάτων, αποτελείται από ένα αντικείμενο (διάνυσμα) εισόδου και μια επιθυμητή τιμή εξόδου. Ο απώτερος στόχος είναι η γενίκευση της συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο (σύνολο ελέγχου). Να μπορεί δηλαδή ο αλγόριθμος να εκτιμήσει σωστά την επιθυμητή έξοδο από περιπτώσεις παραδειγμάτων που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση.

Για τη συνάρτηση πρόγνωσης ισχύουν τα εξής:

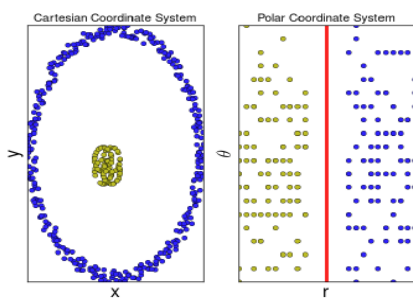
- Κάθε είσοδος, δεδομένη ή όχι, που δέχεται η συνάρτηση χαρακτηρίζεται ως στιγμιότυπο (παράδειγμα ή δείγμα - instance ή example). Το σύνολο των εισόδων δημιουργούν ένα σύνολο στιγμιότυπων.

- Οι είσοδοι περιγράφονται με βάση τα γνωρίσματα τους (attributes ή χαρακτηριστικά - features), τα οποία έχουν χαρακτηριστεί ως σημαντικά από την αρχή της μελέτης του προς επίλυση προβλήματος.
 - Οι είσοδοι που θεωρούνται ως δεδομένοι, συλλέγονται μέσω παρατηρήσεων και αποτελούν το σύνολο εκπαίδευσης (training set), το οποίο είναι υποσύνολο του συνόλου στιγμιότυπων. Το υπόλοιπο μέρος του συνόλου στιγμιότυπων αποτελεί το σύνολο ελέγχου (test set), το οποίο θα χρησιμοποιηθεί κατά τη φάση της επικύρωσης (validation).
 - Η συνάρτηση που αντιστοιχίζει μια είσοδο από το σύνολο εκπαίδευσης στη γνωστή της έξοδο καλείται συνάρτηση στόχου (goal function). Η τιμή που επιστρέφει η συνάρτηση στόχου για ένα στιγμιότυπο, δίνεται σε μια μεταβλητή η οποία καλείται μεταβλητή στόχου (goal variable ή τάξη κατηγοριοποίησης - classification class).
 - Στην επιβλεπόμενη μάθηση, η συμπεριφορά της συνάρτησης στόχου βελτιώνεται μέσω διαδικασιών εκπαίδευσης, με τη βοήθεια της συνάρτησης απώλειας (συνάρτηση λάθους - error function ή loss function), η οποία εντοπίζει τη διαφορά μεταξύ της μεταβλητής στόχου και της επιθυμητής εξόδου.
- Η επιτηρούμενη επαγωγική μάθηση εφαρμόζεται κυρίως σε προβλήματα ταξινόμησης (classification), όπου η έξοδος παίρνει διακριτές (discrete) τιμές, π.χ. «ναι», «όχι», «ίσως» και σε προβλήματα παλινδρόμησης (regression), όπου η έξοδος παίρνει συνεχείς τιμές, π.χ. πραγματικές τιμές. Παράδειγμα εφαρμογής επιτηρούμενης μάθησης είναι η αναγνώριση προκαθορισμένων αντικειμένων σε εικόνες. Σε αυτή την κατηγορία μάθησης, εντάσσονται και τα Τεχνητά Νευρωνικά Δίκτυα τα οποία χρησιμοποιούνται στο πειραματικό μέρος της εργασίας.
 - Μη επιτηρούμενη μάθηση ή χωρίς επίβλεψη (Unsupervised Learning): Στα δεδομένα εισόδου δεν έχουν ανατεθεί ετικέτες και η μάθηση προτύπων εισόδου πραγματοποιείται χωρίς να παρέχονται συγκεκριμένες τιμές εξόδου. Βασικές μέθοδοι αποτελούν η Ανάλυση σε Κύριες Συνιστώσες (Principal Component Analysis - PCA) και η Ανάλυση Συστάδων (Cluster Analysis).
 - Ενισχυτική μάθηση ή δια ανταμοιβής (Reinforcement Learning): Η Ενισχυτική μάθηση ασχολείται με τον τρόπο που τα συστήματα πρέπει να ενεργήσουν σε ένα δυναμικό περιβάλλον, ώστε να μεγιστοποιήσουν κάποια πιθανότητα ανταμοιβής. Όπως και η μη επιτηρούμενη μάθηση, δεν χρειάζεται συγκεκριμένες τιμές εξόδου. Υλοποιείται τυπικά με την μορφή αλυσίδων αποφάσεων Markov. Παράδειγμα εφαρμογής αποτελούν τα συστήματα αυτόματης πλοήγησης οχημάτων.

Άλλες κατηγορίες μηχανικής μάθησης αποτελούν οι: Ημι-επιβλεπόμενη μάθηση (Semi-supervised learning), η Μεταγωγή ή μεταβιβαστική συμπερασματολογία (Transduction ή transductive inference), η Αναπτυξιακή μάθηση (Developmental learning) κ.α. Τέλος, πολλά προβλήματα μπορεί να αποτελούν συνδυασμό των παραπάνω.

4.1.2 Τύποι αλγόριθμων Μηχανικής Μάθησης

Όλοι οι αλγόριθμοι Μηχανικής Μάθησης διαχειρίζονται τη γνώση αναπαριστώντας την με κάποιον τρόπο. Ορισμένοι αλγόριθμοι δέχονται ως είσοδο μόνο παρατηρήσεις και άλλοι λαμβάνουν υπόψη τους την προϋπάρχουσα γνώση. Έχουν αναπτυχθεί διάφοροι (επαγωγικοί) αλγόριθμοι ML που διαφέρουν ως προς την αναπαράσταση του χώρου των δυνατών υποθέσεων και κατά συνέπεια και ως προς τον τρόπο που οργανώνουν την αναζήτηση σε αυτόν τον χώρο. Διαφορετικές αναπαραστάσεις (Εικόνα 4-1) είναι κατάλληλες για τη μάθηση διαφορετικών ειδών συναρτήσεων στόχου [94, 97].



Εικόνα 4-1: Διαφορετικές αναπαραστάσεις δεδομένων (καρτεσιανές - πολικές)

Η μορφή της αναπαράστασης των δεδομένων αποτελεί σημαντικό παράγοντα στην απόδοση των αλγορίθμων ML. Η επιλογή των κατάλληλων χαρακτηριστικών (features), τα οποία αποτελούν την αναπαράσταση των δεδομένων σε ένα πιο υψηλό και αφαιρετικό επίπεδο αντίληψης για τους υπολογιστές, είναι ένα αρκετά δύσκολο και χρονοβόρο έργο. Επομένως, οι αλγόριθμοι ML μπορεί να ομαδοποιηθούν ανάλογα με την ομοιότητα τους σε σχέση με την λειτουργία που εκτελούν, την επιθυμητή μορφή της εξόδου τους, τη μορφή της αναπαράστασης των δεδομένων εισόδου και γενικότερα το είδος του προβλήματος που επιλύουν. Μερικοί από τους πιο δημοφιλείς αλγορίθμους Επαγωγικής Μάθησης, κατηγοριοποιημένοι [95, 98] με βάση την λειτουργία τους είναι:

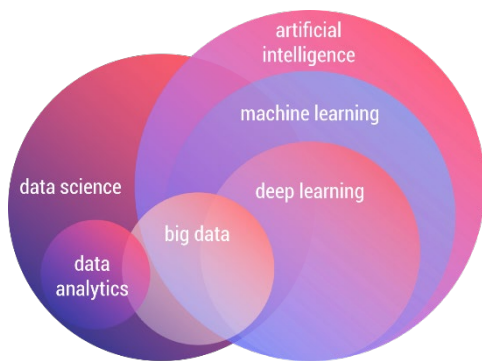
- Regression (Παλινδρόμηση): Ασχολούνται με τη μοντελοποίηση της σχέσης μεταξύ των τιμών των μεταβλητών του προβλήματος. Χρησιμοποιούν ένα μέτρο σφάλματος για τις προβλέψεις που γίνονται από το μοντέλο. Παραδείγματα αποτελούν οι: Ordinary Least Squares Regression (OLSR), Regression: Linear, Logistic και Stepwise, Multivariate Adaptive Regression, Splines (MARS) κ.α.

- Instance-based (Βάση Παραδειγμάτων): Δημιουργούν μία βάση με παραδείγματα δεδομένων και συγκρίνουν τις νέες εισόδους με αυτές που έχουν καταχωρηθεί στην βάση. Χρησιμοποιούν ένα μέτρο ομοιότητας, με σκοπό την εύρεση της καλύτερης αντιστοιχίας. Παραδείγματα: k-Nearest Neighbour (kNN), Learning Vector Quantization (LVQ), Self-Organizing Map (SOM) κ.α.
- Regularization (Κανονικοποίηση): Χρησιμοποιούνται ως επεκτάσεις άλλων μεθόδων, «τιμωρώντας» πολύπλοκα μοντέλα και «ευνοώντας» απλούστερα, τα οποία συνήθως είναι καλύτερα στην γενίκευση της επίλυσης. Παραδείγματα: Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Least-Angle Regression (LARS), Elastic Net κ.α.
- Dimensionality Reduction (Μείωση Διαστάσεων): Χρησιμοποιούνται για την αφαίρεση της μη σημαντικής πληροφορίας από τα δεδομένα. Χρησιμοποιούνται και σαν επεκτάσεις σε regression και classification μοντέλα. Παραδείγματα: Principal Component Analysis (PCA), Discriminant Analysis: Linear, Mixture, Quadratic και Flexible, Principal Component Regression (PCR) κ.α.
- Decision Trees (Δέντρα Αποφάσεων): Χρησιμοποιούνται για την κατασκευή μοντέλων λήψης αποφάσεων, τα οποία χρησιμοποιούν τις πραγματικές τιμές των χαρακτηριστικών των δεδομένων. Παραδείγματα: Classification και Regression Tree (CART), Conditional Decision Trees, M5 κ.α.
- Bayesian: Εφαρμόζουν το θεώρημα Bayes για την επίλυση regression και classification προβλημάτων. Παραδείγματα: Naive Bayes, Gaussian Naive Bayes, Bayesian Network (BN), Bayesian Belief Network (BBN) κ.α.
- Clustering (Ομαδοποίηση): Περιγράφουν τις κλάσεις του προβλήματος. Παραδείγματα: k-Means, k-Medians, Hierarchical Clustering κ.α.
- Artificial Neural Networks - ANN (Τεχνητά Νευρωνικά Δίκτυα - ΤΝΔ): Μοντέλα εμπνευσμένα από τη δομή και τη λειτουργία των βιολογικών νευρωνικών δικτύων. Χρησιμοποιούνται στην επίλυση regression και classification προβλημάτων. Παραδείγματα: Perceptron, Back-Propagation, Radial Basis Function Network (RBFN) κ.α.
- Deep Learning - DL (Βαθιά Μηχανική Μάθηση): Αποτελούν τη σύγχρονη επέκταση των Τεχνητών Νευρωνικών Δικτύων και εκμεταλλεύονται την μεγάλη επεξεργαστική ισχύ των σύγχρονων υπολογιστικών συστημάτων. Παραδείγματα: Autoencoder, Multilayer Perceptron (MLP), Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN), Stacked Auto-Encoders, Recurrent Neural Networks (RNN) κ.α.

Στις επόμενες υποενότητες γίνεται αρχικά μια συνοπτική αναφορά στη Βαθιά Μηχανική Μάθηση και στη συνέχεια περιγράφονται αναλυτικά τα Τεχνητά Νευρωνικά Δίκτυα και ο τρόπος υλοποίησής τους στα πλαίσια της Βαθιάς Μηχανικής Μάθησης.

4.1.3 Βαθιά Μηχανική Μάθηση

Ο τομέας της τεχνητής νοημοσύνης εμπεριέχει τις μεθόδους της μηχανικής μάθησης και βαθιάς μηχανικής μάθησης, με την δεύτερη να είναι υποκατηγορία της πρώτης (Εικόνα 4-2). Τα τελευταία χρόνια η δημοτικότητα των ΤΝΔ έχει ανέβει κατακόρυφα, λόγω της ανάπτυξης μιας νέας οικογένειας αλγορίθμων μηχανικής μάθησης που συγκεντρωτικά ονομάζονται Βαθιά Μηχανική Μάθηση (Deep Learning - DL) (G. Zaccanne 2016) και συνδέονται άμεσα με τα νευρωνικά δίκτυα πολλών επιπέδων (Deep Neural Networks). Η αφθονία δεδομένων που προσφέρει το διαδίκτυο αλλά και η δραματική αύξηση των επιδόσεων των Η/Υ, βοήθησε τη Βαθιά Μηχανική Μάθηση να αναπτυχθεί ταχύτατα.



Εικόνα 4-2: Έννοιες που εμπεριέχονται στον κλάδο της τεχνητής νοημοσύνης.

Πηγή: [100]

Ο Goodfellow [100] και άλλοι όρισαν την DL ως εξής: «Η βαθιά μάθηση είναι ένας συγκεκριμένος τύπος μηχανικής μάθησης, που επιτυγχάνει μεγάλη δύναμη και ευελιξία μαθαίνοντας να αναπαριστά τον κόσμο ως μια ενσωματωμένη ιεραρχία εννοιών, όπου η κάθε έννοια ορίζεται μέσω απλούστερων εννοιών και οι αφηρημένες αναπαραστάσεις υπολογίζονται σε σχέση με λιγότερο αφηρημένες έννοιες.» Η Βαθιά Μηχανική Μάθηση ανήκει στην επιβλεπόμενη μηχανική μάθηση, αφορά μάθηση πολλαπλών επιπέδων αναπαράστασης και γενικών κανόνων και μπορεί να επιλύσει και σύνθετα μη γραμμικά προβλήματα. Προσθέτοντας περισσότερα στρώματα (επίπεδα - layers) και περισσότερα στοιχεία σε κάθε στρώμα, ένα βαθύ δίκτυο (deep network) μπορεί να αναπαραστήσει όλο και πιο πολύπλοκες συναρτήσεις. Από μαθηματικής άποψης, οι αλγόριθμοι deep learning είναι προσεγγιστικές συναρτήσεις, οι οποίες προσπαθούν να ανακαλύψουν μια σχέση μεταξύ αριθμητικών περιγραφών αντικειμένων και των ετικετών τους. Οι αλγόριθμοι

DL, μπορούν να αντικαταστήσουν τα παραδοσιακά «expert systems», τα οποία βασίζονται σε προκατασκευασμένους κανόνες που ορίζονται από ειδήμονες ανθρώπους, κάτι μη εφικτό για δύσκολα προβλήματα ταξινόμησης. Ενώ ένας άνθρωπος θα μπορούσε εύκολα να διαχωρίσει την εικόνα μιας γάτας από την εικόνα ενός σκύλου, είναι πολύ δύσκολο να εκφράσει στη μορφή μίας λίστας από κανόνες, δηλαδή ενός αλγορίθμου, τη διαδικασία που ακολουθεί ο εγκέφαλός του για να κάνει αυτήν την ταξινόμηση. Ο αλγόριθμος ψάχνει μέσα στο χώρο όλων των πιθανών τιμών των παραμέτρων με στόχο να προσδιορίσει τη μαθηματική συνάρτηση που με μεγαλύτερη ακρίβεια ταιριάζει με τα ετικετοποιημένα παραδείγματα και στη συνέχεια θα χρησιμοποιήσει αυτή τη συνάρτηση για να ταξινομήσει άγνωστες εικόνες. Στην ουσία, η κύρια διαφορά μεταξύ Μηχανικής Μάθησης και Βαθιάς Μηχανικής Μάθησης είναι ο τρόπος παρουσίασης των δεδομένων. Οι αλγόριθμοι ML απαιτούν σχεδόν πάντα την χρήση δομημένων δεδομένων, δηλαδή δεδομένων που έχουν κατηγοριοποιηθεί προηγουμένως. Αντίθετα στα νευρωνικά δίκτυα βαθιάς μάθησης η επεξεργασία και η ταξινόμηση των τροφοδοτούμενων δεδομένων γίνεται αυτόματα. Κάποιες άλλες διαφορές μεταξύ ML και DL παρουσιάζονται στον Πίνακα 4-1.

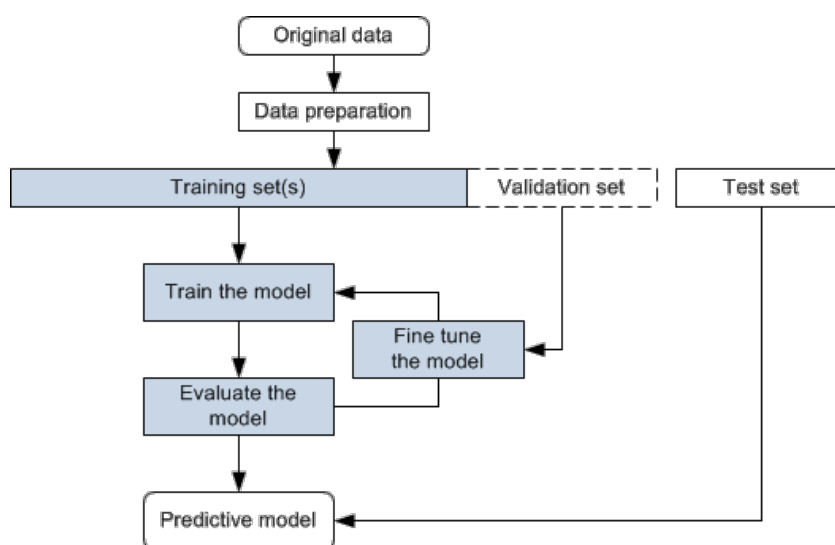
Χρησιμοποιώντας μεθόδους λογιστικής ταξινόμησης (logistic classification) και στοχαστικών αλγορίθμων βελτιστοποίησης (stochastic optimization) και σε συνδυασμό με αρκετά δείγματα εκπαίδευσης (πολλά δεδομένα - Big Data), οι εφαρμογές Βαθιάς Μηχανικής Μάθησης μπορούν να καταστήσουν δυνατή την κατανόηση [102] του περιεχομένου εικόνων (π.χ. της τεχνοτροπίας ενός πίνακα [103]), την κατανόηση ήχων, τη μετάφραση ενός κειμένου [104] κ.α. Γενικότερα, με τεχνικές DL, όπως τα Convolutional Neural Networks (Συνελκτικτικά νευρωνικά δίκτυα), στην τομέα της όρασης υπολογιστών (Computer Vision), μπορούν να εξαχθούν χαρακτηριστικά της εικόνας, τα οποία είναι ικανά να την ταξινομήσουν. Τα CNN θα αναλυθούν σε βάθος στη ενότητα 4.5 της εργασίας.

Πίνακας 4-1: Διαφορές Μηχανικής Μάθησης και Βαθιάς Μηχανικής Μάθησης.

	ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	ΒΑΘΙΑ ΜΑΘΗΣΗ
Διαχωρισμός και συσχέτιση μεταξύ των δεδομένων.	Απαιτείται ώστε να είναι δυνατή η κατανόηση των χαρακτηριστικών τους.	Δεν είναι απαραίτητο.
Πλήθος δεδομένων.	Αποδίδει καλύτερα σε μικρά και μεσαία σύνολα δεδομένων.	Αποδίδει καλύτερα σε μεγάλα σύνολα δεδομένων.
Υπολογιστική ισχύ.	Δεν απαιτείται μεγάλη.	Απαιτείται μεγάλη.
Χρόνος εκπαίδευσης.	Μπορεί να διαφέρει από λεπτά σε ώρες.	Μπορεί να διαφέρει από μέρες σε εβδομάδες.

Για την επίλυση προβλημάτων χρησιμοποιώντας βαθιά μάθηση και ΤΝΔ απαιτούνται σε γενικές γραμμές τα παρακάτω βασικά βήματα (Εικόνα 4-3):

- Συλλογή δεδομένων (data collection): Καθορισμός προβλήματος και συγκέντρωση δεδομένων για την εκπαίδευση και την αξιολόγηση του μοντέλου.
- Προετοιμασία δεδομένων (pre-process data): Καθαρισμός δεδομένων, διόρθωση σφαλμάτων και τροποποίηση σε ενιαία μορφή.
- Επιλογή χαρακτηριστικών (feature selection): Επιλογή των χαρακτηριστικών (εισόδου) και στόχου (εξόδου).
- Δημιουργία συνόλων εκπαίδευσης & αξιολόγησης (creation of training & evaluation sets): Διαχωρισμός δεδομένων στα σύνολα για την εκπαίδευση και για την αξιολόγηση του μοντέλου.
- Επιλογή & δημιουργία μοντέλου (define & compile model): Επιλογή του είδους και της αρχιτεκτονικής του μοντέλου.
- Εκπαίδευση μοντέλου (training model): Το μοντέλο τροφοδοτείται με τα δεδομένα εκπαίδευσης με σκοπό να «εκπαιδευτεί» και να μάθει να αναλύει τέτοιου τύπου δεδομένα, βελτιώνοντας την ακρίβεια και μειώνοντας τη συνολική απώλειά του.
- Αξιολόγηση μοντέλου (model evaluation): Έλεγχος της απόδοσης του μοντέλου βάσει των δεδομένων αξιολόγησης, ώστε να δοκιμαστεί αν μπορεί να ανταποκριθεί σε νέες εισόδους και έτσι να γίνουν οι τελικές διορθώσεις (βελτιστοποίηση παραμέτρων).
- Προβλέψεις & ανάπτυξη μοντέλου (make predictions & deploying model): Πραγματοποίηση προβλέψεων, χρησιμοποιώντας το εκπαιδευμένο μοντέλο σε άγνωστα δεδομένα. Αυτό είναι το τελικό στάδιο όπου θεωρούμε ότι το μοντέλο είναι έτοιμο για πρακτικές εφαρμογές.



Εικόνα 4-3: Βασικά βήματα για τη δημιουργία ενός μοντέλου βαθιάς μάθησης.

4.1.4 Ανάγκη για Deep Networks & πλεονεκτήματα

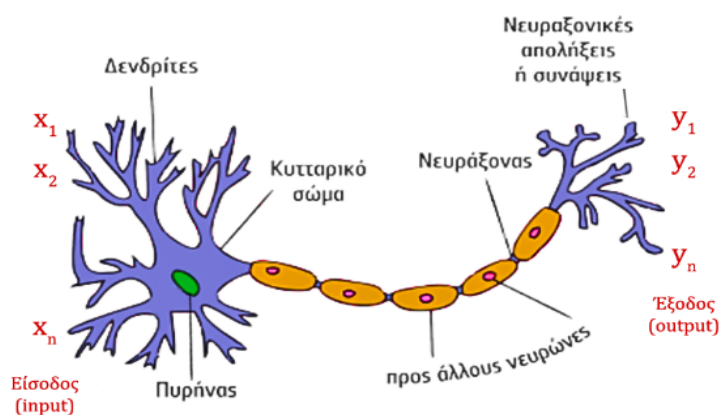
Τα νευρωνικά δίκτυα έχουν κάποιες μοναδικές ιδιότητες. Καταρχήν είναι εξαιρετικά ευέλικτα, ώστε να μπορούν μαθηματικά να αποδώσουν την προσέγγιση οποιασδήποτε συνάρτησης, χρησιμοποιώντας μεγάλο αριθμό απλών συναρτήσεων που συνδέονται με ποικίλους τρόπους. Οι αλγόριθμοι DL, διαφοροποιούνται από παλαιότερες γενιές νευρωνικών δικτύων, λόγω των διαφορετικών αρχιτεκτονικών που χρησιμοποιούν και κυρίως από τα πολλαπλά επίπεδα από τα οποία αποτελούνται. Τα «ενδιάμεσα» επίπεδα, γνωστά και ως κρυφά επίπεδα (hidden layers) δεν επιβλέπονται. Δηλαδή, δεν υπάρχουν ετικέτες που να επιβεβαιώνουν ή να διαψεύδουν τις προβλέψεις των επιπέδων αυτών. Τα DL δίκτυα αναλαμβάνουν εξολοκλήρου μόνα τους να συνάγουν ποιες δομές είναι χρήσιμες στη συνολική ταξινόμηση των δεδομένων, και ποιες όχι, και να φροντίσουν να τις αναπαραστήσουν στα κρυφά επίπεδά τους. Με αυτόν τον τρόπο, κάνουν αυτόματα αυτό που οι data scientists χρειάζεται ακόμη και σήμερα να κάνουν χειροκίνητα: δηλαδή, τον μετασχηματισμό αδόμητων συνόλων δεδομένων (raw data) σε δομές που μπορούν ευκολότερα να συσχετιστούν με μια ετικέτα από τον πραγματικό κόσμο, όπως για παράδειγμα «σκύλος» ή «γάτα».

Επίσης, ένα συχνό πρόβλημα σε εφαρμογές τεχνητής νοημοσύνης είναι η εύρεση και εξαγωγή χαρακτηριστικών υψηλότερου επιπέδου. Τα δίκτυα DL, χρησιμοποιώντας πολλαπλά στρώματα, δίνουν λύσεις σε αυτό το πρόβλημα εκμάθησης αναπαραστάσεων, με την εισαγωγή χαρακτηριστικών τα οποία εκφράζονται με βάση άλλες, απλούστερες αναπαραστάσεις. Το κάθε επίπεδο του μοντέλου, βασισμένο στην πληροφορία που προέρχεται από το προηγούμενο επίπεδο, την κωδικοποιεί, μέσω μιας πιο αφαιρετικής αναπαράστασης (abstract representation). Προσεγγίζει έτσι πιο σύνθετες έννοιες από απλούστερες και επιτρέπει την εξαγωγή περίπλοκων δομών με ιεραρχικό τρόπο [105]. Αν η σχέση μεταξύ εισόδου και εξόδου είναι απλή, τότε μπορεί να ληφθούν συμπαγείς αναπαραστάσεις εισόδου-εξόδου, χρησιμοποιώντας απλά συναρτησιακά μοντέλα ή νευρωνικά δίκτυα με λιγότερα από τρία επίπεδα νευρώνων (ρηγά δίκτυα). Ωστόσο, για πιο πολύπλοκα προβλήματα, η βαθύτερη εξάρτηση μεταξύ εισόδου και εξόδου είναι πολύπλοκη και συνεπώς δε μπορεί να εκφραστεί αναλυτικά με απλό τρόπο. Αν και δεν υπάρχουν ισχυρές θεωρητικές αποδείξεις, προκύπτει ότι χρησιμοποιώντας μοντέλα με περισσότερα επίπεδα, επιτυγχάνονται πιο συμπαγείς αναπαραστάσεις και ως εκ τούτου αυτή η ιεραρχικού τύπου αναπαράσταση των deep networks οδηγεί σε ισχυρότερη ικανότητα γενίκευσης και πιο συμπαγή αποτελέσματα.

4.2 Τεχνητά Νευρωνικά Δίκτυα

4.2.1 Βιολογικοί νευρώνες

Η ιδέα των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) ξεκινά από τις πρώτες προσπάθειες κατανόησης του τρόπου με τον οποίο δομείται και λειτουργεί ο ανθρώπινος εγκέφαλος, δηλαδή την κατανόηση της νοημοσύνης [99]. Ο ανθρώπινος εγκέφαλος είναι ένας πολύπλοκος, μη γραμμικός, παράλληλος επεξεργαστής που οργανώνει τα δομικά του στοιχεία με τέτοιο τρόπο ώστε να εκτελούν πολύ γρήγορα υπολογισμούς, όπως αναγνώριση προτύπων, αντίληψη, έλεγχος, κίνηση κ.α. Το δομικό στοιχείο του εγκεφάλου είναι ο νευρώνας (neuron). Κάθε νευρώνας συνδέεται με άλλους νευρώνες μέσω συνδέσεων που ονομάζονται συνάψεις (synapses) (Εικόνα 4-4) και δέχεται είσοδο από άλλους νευρώνες μέσω των δενδριτών (dendrites) του. Στη συνέχεια επεξεργάζεται το σήμα που λαμβάνει στην είσοδο και στέλνει το αποτέλεσμα στον νευράξονα. Οι χημικές συνάψεις μετατρέπουν ηλεκτρικούς παλμούς, παραγόμενους από κάποιο νευρώνα, σε χημικά σήματα και στη συνέχεια πάλι σε ηλεκτρικά. Ανάλογα τον παλμό εισόδου, μια σύναψη είναι αντίστοιχα ενεργή ή όχι. Μέσω των συνδέσεων αυτών, κάθε νευρώνας είναι συνδεδεμένος με άλλους νευρώνες σε μια ιεραρχική δομή που αποτελείται από διαφορετικά στρώματα. Το νευρικό σύστημα του ανθρώπου είναι το σύστημα που ρυθμίζει και ελέγχει τη λειτουργία των οργάνων του ανθρωπίνου σώματος, καθώς επίσης και τη μεταξύ τους συνεργασία, συμβάλλοντας στην αντίληψη του περιβάλλοντος από τον άνθρωπο και στη δημιουργία κανόνων συμπεριφοράς (εμπειρία).



Εικόνα 4-4: Μέρη βιολογικού νευρώνα: Δενδρίτης (dendrites): Δέχεται είσοδο από άλλους νευρώνες. Σώμα κυττάρου (cell body): Εξάγει συμπεράσματα, με βάση τις εισόδους.

Νευράξονας (axon): Συνδέει την έξοδο που λαμβάνεται από το σώμα με τις απολήξεις.

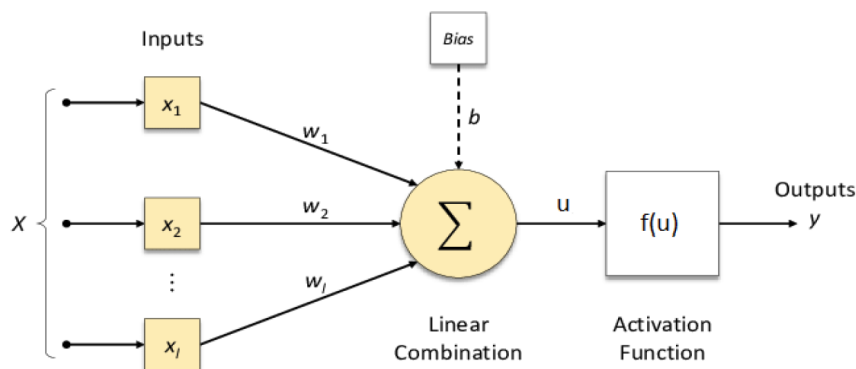
Νευραξονικές απολήξεις (συνάψεις): Συνδέει τον νευράξονα του εκάστοτε νευρώνα με τους τερματικούς κόμβους, από όπου και μεταφέρεται η πληροφορία στην είσοδο άλλων νευρώνων.

4.2.2 Τα μοντέλα McCulloch-Pitts & Perceptron

Η ιστορία των ΤΝΔ ξεκινάει το 1943, με την ανάπτυξη ενός υπολογιστικού μοντέλου για τον βασικό νευρώνα, από τους McCulloch και Pitts [206], συνδυάζοντας τη νευρολογία με τα μαθηματικά. Συγκεκριμένα, έδειξαν ότι ρυθμίζοντας κατάλληλα τις συνδέσεις μεταξύ ενός ικανού αριθμού νευρώνων, όπου κάθε μία αναπαριστάται από ένα βάρος (weight), μπορεί θεωρητικά να υπολογιστεί οποιαδήποτε υπολογίσιμη συνάρτηση. Στη συνέχεια, ο Rosenblatt, [207], χρησιμοποιώντας το μοντέλο των McCulloch και Pitts, ανέπτυξε το σύστημα αναγνώρισης προτύπων Perceptron και έπειτα, με την εισαγωγή του αλγόριθμου Back-Propagation από τον Werbos [208], αλλά και του μοντέλου Perceptron πολλών στρωμάτων (Multi-Layer Perceptron ή MLP) τη δεκαετία του '80, τα νευρωνικά δίκτυα αποτελούν πλέον, μηχανές μάθησης αποτελούμενες από μεγάλο αριθμό νευρώνων συνδεδεμένους μεταξύ τους σε στρώματα (layers). Η εκμάθηση επιτυγχάνεται ρυθμίζοντας τα άγνωστα συναπτικά βάρη με σκοπό την ελαχιστοποίηση μιας προεπιλεγμένης συνάρτησης κόστους. Η απόδοση των Η/Υ εκείνη την περίοδο καθιστούσε απαγορευτική την εκμετάλλευση τέτοιων μοντέλων, αλλά στη σημερινή εποχή, η αλματώδης ανάπτυξη της τεχνολογίας τα έφερε και πάλι στο προσκήνιο. Τα τελευταία χρόνια, υπάρχει μια μεγάλη αύξηση του ενδιαφέροντος για τα νευρωνικά δίκτυα στα πλαίσια της Βαθιά Μηχανική Μάθησης (deep learning).

Το μοντέλο του τεχνητού νευρώνα McCulloch-Pitts [206], μιμείται τη λειτουργικότητα ενός βιολογικού νευρώνα και αποτελεί το θεμελιώδες δομικό στοιχείο των ΤΝΔ. Δέχεται δυαδικές εισόδους και παράγει μια δυαδική έξοδο με βάση μια συγκεκριμένη τιμή κατωφλίου, η οποία μπορεί να προσαρμοστεί. Λαμβάνοντας διάφορες εισόδους, η συνάρτηση τις συγκεντρώνει και παίρνει αποφάσεις με βάση το άθροισμά τους. Εάν το άθροισμα υπερβαίνει το όριο, η έξοδος είναι 1, διαφορετικά είναι 0.

Μια ελαφρώς τροποποιημένη έκδοση του μοντέλου McCulloch-Pitts είναι το μοντέλο Perceptron [207, 99]. Το μοντέλο αυτό μπορεί να λειτουργήσει με μη δυαδικές τιμές, όπου κάθε είσοδος σχετίζεται με ένα βάρος. Η συνάρτηση υπολογίζει το σταθμισμένο άθροισμα και με βάση την τιμή κατωφλίου που παρέχεται, δίνει μια δυαδική έξοδο. Ο νευρώνας Perceptron είναι ένα απλό πρόσθιας τροφοδότησης δίκτυο (Εικόνα 4-5) και αποτελεί ένα δυαδικό (γραμμικό) ταξινομητή (linear classifier), δηλαδή μία συνάρτηση η οποία απεικονίζει την είσοδο (ένα διάνυσμα με πραγματικές τιμές) σε μία τιμή εξόδου (μία και μοναδική δυαδική τιμή). Αν και ο όρος Perceptron χρησιμοποιήθηκε αρχικά για το στοιχειώδες ΤΝΔ πολλών εισόδων και μόνο μιας εξόδου, έχει επικρατήσει να χαρακτηρίζονται ως Perceptrons όλα τα ΤΝΔ πρόσθιας τροφοδότησης που δεν περιέχουν στην αρχιτεκτονική τους κρυφά επίπεδα.



Εικόνα 4-5: Αρχιτεκτονική του μοντέλου Perceptron.

4.2.3 Μαθηματικό μοντέλο τεχνητού νευρώνα

Στα βιολογικά νευρωνικά δίκτυα, όπως αναφέρθηκε προηγουμένως, κάθε νευρώνας, δέχεται είσοδο από τους συνδεδεμένους νευρώνες, επεξεργάζεται αυτό το σήμα και στέλνει το αποτέλεσμα μέσω του νευράξονα, σε άλλους νευρώνες. Στα ΤΝΔ και συγκεκριμένα στα μοντέλα Perceptron, η παραπάνω λειτουργία, αντιστοιχίζεται με ένα μαθηματικό μοντέλο νευρώνα, όπως αυτό παρουσιάζεται στην Εικόνα 4-6 [105]. Η πληροφορία (είσοδος) που μεταφέρεται από τις συνάψεις (x_0, x_1, \dots, x_n) , προτού μεταφερθεί στους δενδρίτες των επόμενων νευρώνων, αλληλεπιδρά πολλαπλασιαστικά με τις συνάψεις $(w_0 \times x_0)$. Οι παράγοντες πολλαπλασιασμού w_n ονομάζονται βάρη (weights) και αποτελούν τις παραμέτρους ενός νευρώνα, οι οποίες ρυθμίζουν την επίδραση μεταξύ των νευρώνων. Τα αποτελέσματα αθροίζονται μέσω της συνάρτησης αθροίσματος (summation function) f :

$$f = \sum_{i=0}^N w_i x_i \quad (4.2.1)$$

Ο τεχνητός νευρώνας δίνει έξοδο (δηλαδή ενεργοποιείται) μέσω της συνάρτησης ενεργοποίησης f (activation function ή συνάρτηση μετάβαση - transfer function), μόνο όταν το άθροισμα των εισόδων είναι μεγαλύτερο μιας ορισμένης τιμής κατωφλίου θ (threshold value), διαφορετικά παραμένει αδρανής, δηλαδή όταν:

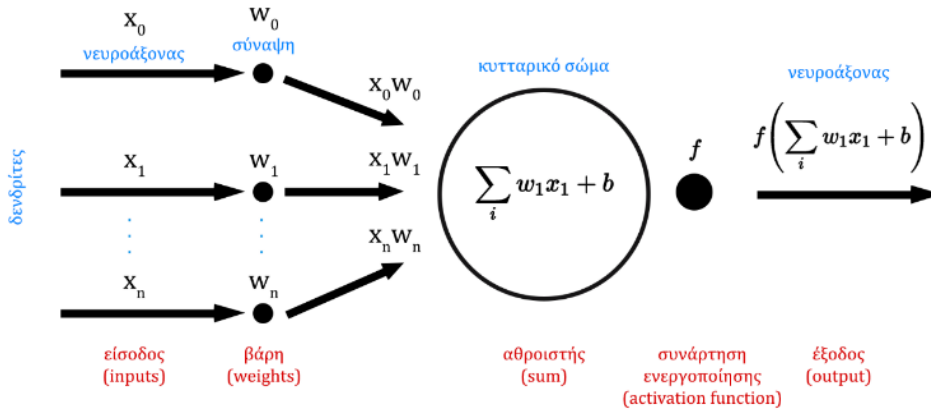
$$u = \sum_{i=0}^N w_i x_i - \theta > 0 \quad (4.2.2)$$

Με άλλα λόγια, η συνάρτηση ενεργοποίησης ελέγχει το αν θα διέρχεται ή όχι, η πληροφορία προς στους συνδεδεμένους νευρώνες. Υπάρχουν πολλές μορφές συναρτήσεων ενεργοποίησης, η επιλογή των οποίων δεν είναι τυχαία, όπως θα αναλυθεί στην ενότητα 4.3.4.

Επίσης, ο νευρώνας μπορεί να είναι και πολωμένος (biased - b), οπότε η έξοδος από ένα νευρώνα υπολογίζεται από τη σχέση:

$$\alpha = f\left(\sum_{i=0}^N w_i x_i + b\right) \quad (4.2.3)$$

Η πόλωση (bias) ανάλογα με το αν είναι θετική ή αρνητική, έχει ως αποτέλεσμα την αύξηση ή μείωση της δικτυακής διέγερσης της συνάρτησης ενεργοποίησης.



Εικόνα 4-6: Μαθηματικό μοντέλο του νευρώνα.

4.3 Αρχιτεκτονική ΤΝΔ

4.3.1 Περιγραφή δομής & λειτουργίας ΤΝΔ

Γενικά, κάθε ΤΝΔ οργανώνεται σε επίπεδα (layers) ή αλλιώς στρώματα. Τα ενδιάμεσα επίπεδα καλούνται κρυφά επίπεδα (hidden layers). Τα επίπεδα αποτελούνται από ένα αριθμό κόμβων (nodes ή μονάδες - units). Οι κόμβοι συνδέονται μεταξύ τους, έτσι ώστε κάθε κόμβος να έχει συνδέσμους με άλλους κόμβους του ίδιου ή άλλου επιπέδου, οι οποίοι με τη σειρά τους μπορούν να επιδρούν σε άλλους κόμβους ενεργοποιώντας ή απενεργοποιώντας τους. Για να επιτευχθεί αυτό, ο κόμβος λαμβάνει το σταθμισμένο άθροισμα όλων των εισόδων, μέσω των συνδέσμων που καταλήγουν σε αυτόν και, αν το άθροισμα υπερβαίνει μία τιμή κατωφλίου, παράγει μέσω της συνάρτησης ενεργοποίησης μία μοναδική έξοδο. Οι εισοδοί παρουσιάζονται στο δίκτυο μέσω του επιπέδου εισόδου (input layer), το οποίο επικοινωνεί με έναν ή περισσότερα κρυμμένα επίπεδα. Τα κρυμμένα επίπεδα συνδέονται με το επίπεδο εξόδου (output layer) από το οποίο εξάγεται το αποτέλεσμα. Ο αριθμός των επιπέδων ή αλλιώς το μέγιστο μήκος του

μονοπατιού που ακολουθεί η πληροφορία από την είσοδο μέχρι την έξοδο, ορίζει το βάθος (depth) του ΤΝΔ (χωρίς να μετρίεται το επίπεδο εισόδου). Τέλος, ο αριθμός των κόμβων σε ένα επίπεδο ορίζει το πλάτος (width) του συγκεκριμένου επιπέδου [105].

Τα βασικά στοιχεία της αρχιτεκτονικής των ΤΝΔ, τα οποία πρέπει να καθοριστούν κατά τη δημιουργία τους είναι:

- Ο αριθμός των ενδιάμεσων κρυφών επιπέδων.
- Ο αριθμός των κόμβων (νευρώνων) ανά επίπεδο.
- Ο τρόπος διασύνδεσης των κόμβων και οι τιμές των αρχικών βαρών.
- Η μορφή της συνάρτησης ενεργοποίησης και η τιμή κατωφλίου.
- Ο αλγόριθμος εκπαίδευσης και οι κανόνες εκμάθησης που χρησιμοποιεί.
- Οι συναρτήσεις απώλειας και οι αλγόριθμοι βελτιστοποίησης.
- Ο ρυθμός μάθησης και το μέγεθος παρτίδας (batch size) κατά την εκπαίδευση

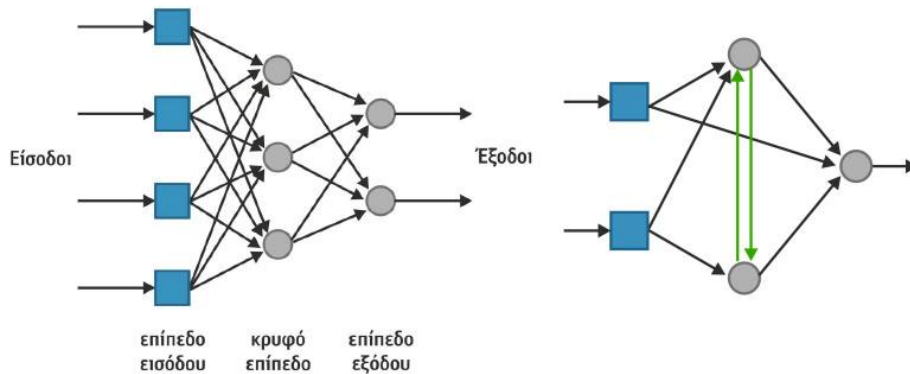
Τα παραπάνω στοιχεία αποτελούν τις υπερπαραμέτρους του ΤΝΔ και καθορίζουν την αρχιτεκτονική ολόκληρου του συστήματος.

4.3.2 Πρόσθιας & οπίσθιας τροφοδότησης ΤΝΔ

Υπάρχουν δυο βασικοί τρόποι με τους οποίους συνδέονται μεταξύ τους οι κόμβοι σε ένα ΤΝΔ, αποτελώντας τις δυο βασικές αρχιτεκτονικές (Εικόνα 4-7):

- Δίκτυα Πρόσθιας Τροφοδότησης (Feed Forward), στα οποία οι κόμβοι είναι οργανωμένοι σε διαφορετικά επίπεδα, ώστε οι κόμβοι του ενός επιπέδου να τροφοδοτούν τους κόμβους του επόμενου επιπέδου, μέχρις ότου τροφοδοτηθούν και οι κόμβοι του τελευταίου επιπέδου. Η πληροφορία ρέει προς μία μόνο κατεύθυνση. Στα Feed-Forward δίκτυα ανήκουν και οι Αυτόματοι κωδικοποιητές (Autoencoders) [106], αλλά και τα Συνελκτικά νευρωνικά δίκτυα (Convolutional neural networks) τα οποία θα αναλυθούν εκτενώς στην ενότητα 4.5 [107].
- Δίκτυα Οπίσθιας Τροφοδότησης (Feed Backward ή Ανατροφοδοτούμενα - Recurrent - RNN), στα οποία η έξοδος ενός κόμβου ενός επιπέδου μπορεί να αποτελεί είσοδο σε ένα κόμβο του ίδιου ή προηγούμενου επιπέδου. Στα RNN δεν υπάρχουν συνήθως άνω του ενός κρυφά επίπεδα. Χρησιμοποιούν διαδοχικά δεδομένα (sequential data), δηλαδή δεδομένα που εξαρτώνται μεταξύ τους. Έχουν βρει εφαρμογή σε μεταφράσεις κειμένων, σε επεξεργασία και αναγνώριση ομιλίας κ.α. Παραδείγματα αποτελούν τα Νευρωνικά δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory - LSTM) [108], οι Επαναλαμβανόμενες μονάδες με πύλη (Gated recurrent units -

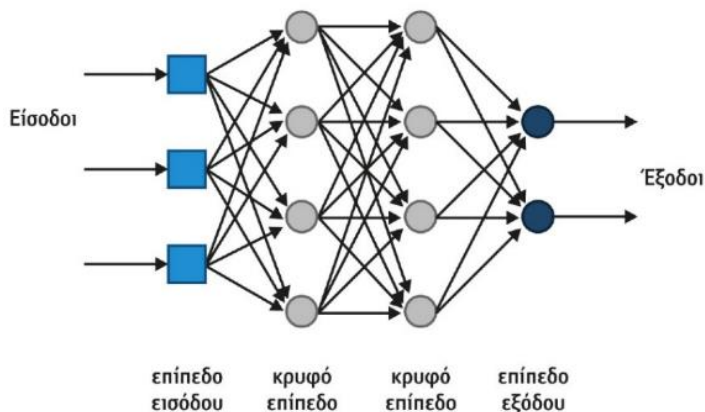
GRUs) [109], τα Αμφίδρομα επαναλαμβανόμενα νευρωνικά δίκτυα (Bidirectional recurrent neural networks - BRNN) [110] κ.α.



Εικόνα 4-7: Παράδειγμα ΤΝΔ πρόσθιας τροφοδότησης (αριστερά), Παράδειγμα ΤΝΔ οπίσθιας τροφοδότησης (δεξιά).

4.3.3 Πολυεπίπεδα ΤΝΔ

Τα αρχικά δίκτυα Perceptrons δεν μπορούσαν να επιλύσουν προβλήματα με μη γραμμικά διαχωρίσιμες τιμές εξόδου. Ο λόγος αυτός οδήγησε στην ανάγκη προσθήκης κρυφών επιπέδων στη δομή των Perceptrons και στη δημιουργία πολυεπίπεδων ΤΝΔ (Multilayer ANN ή αλλιώς Multi-layer Perceptron - MLP). Επομένως, τα πολυεπίπεδα ΤΝΔ διαθέτουν τουλάχιστον ένα κρυφό επίπεδο και οι κόμβοι των διάφορων επιπέδων μπορεί να είναι είτε μερικώς συνδεδεμένοι (partially connected), είτε πλήρως συνδεδεμένοι (fully connected) (Εικόνα 4-8). Πλήρως συνδεδεμένο επίπεδο σημαίνει ότι ο κόμβος, σε αυτό το επίπεδο του δικτύου, συνδέεται με όλους τους κόμβους του προηγούμενου επιπέδου και με την σειρά του συνδέει όλους τους κόμβους του επόμενου επιπέδου.



Εικόνα 4-8: Πολυεπίπεδο ΤΝΔ ενός κρυφού επιπέδου πρόσθιας τροφοδότησης 2 κρυμμένων επιπέδων με πλήρως συνδεδεμένους κόμβους.

Η έξοδος από το εκάστοτε επίπεδο ενός πολυεπίπεδου ΤΝΔ μπορεί να εκφραστεί ως

$$A_{i+1} = f_i (A_i \cdot W_i + B_i) \quad (4.3.1)$$

όπου: A : πίνακας που περιέχει τις εισόδους (x_i), με διαστάσεις $M \times N$, W : πίνακας βαρών (w_i) των νευρώνων του εκάστοτε επιπέδου, με διαστάσεις $K \times M$, B : ο πίνακας με τις τιμές πόλωσης (b_i), με διαστάσεις $K \times N$ και i : η τιμή που αναφέρεται στον αριθμό του εκάστοτε επιπέδου του ΤΝΔ.

Τα κρυφά επίπεδα λειτουργούν ως ανιχνευτές χαρακτηριστικών (features detectors). Καθώς προχωρά η διαδικασία μάθησης οι κρυφοί νευρώνες, εκτελούν ένα μη γραμμικό μετασχηματισμό στα σημεία εισόδου σε ένα χώρο που αποκαλείται χώρος χαρακτηριστικών και με αυτό τον τρόπο ανακαλύπτουν νέα χαρακτηριστικά που χαρακτηρίζουν τα δεδομένα εκπαίδευσης. Με άλλα λόγια, οι μετασχηματισμοί στην αναπαράσταση των δεδομένων που συντελούνται σε κάθε επίπεδο ενός πολυεπίπεδου ΤΝΔ αποτελούν μια «αντιστοίχιση» (mapping). Η αντιστοίχιση αυτή ιδανικά περιέχει περισσότερη πληροφορία από το προηγούμενο επίπεδο, ώσπου τελικά, στο τελευταίο επίπεδο, το πρόβλημα να έχει απλοποιηθεί τόσο ώστε να προκύψει η επιθυμητή λύση.

4.3.4 Συναρτήσεις Ενεργοποίησης

Η Συνάρτηση Ενεργοποίησης (Activation Function ή Συνάρτηση Μετάβασης - Transfer Function) [111, 112] «αποφασίζει» εάν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι., δηλαδή εάν η είσοδος του νευρώνα στο δίκτυο είναι σημαντική ή όχι στη διαδικασία πρόβλεψης. Ο ρόλος της συνάρτησης ενεργοποίησης είναι να εξάγει μια έξοδο από ένα σύνολο τιμών εισόδου που τροφοδοτούνται σε έναν κόμβο. Η επιλογή της κατάλληλης συνάρτησης ενεργοποίησης, έχει μεγάλη σημασία για την εκπαίδευση των ΤΝΔ και γίνεται συνήθως με βάση τον τύπο του προβλήματος που καλείται να επιλύσει το δίκτυο αλλά και με βάση το είδος των τιμών εισόδου. Τέλος, αρκετές φορές η επιλογή προκύπτει από διαδικασία δοκιμής και σφάλματος (trial and error) μέσω της αξιολόγησης της απόδοσης του μοντέλου στο σύνολο επικύρωσης.

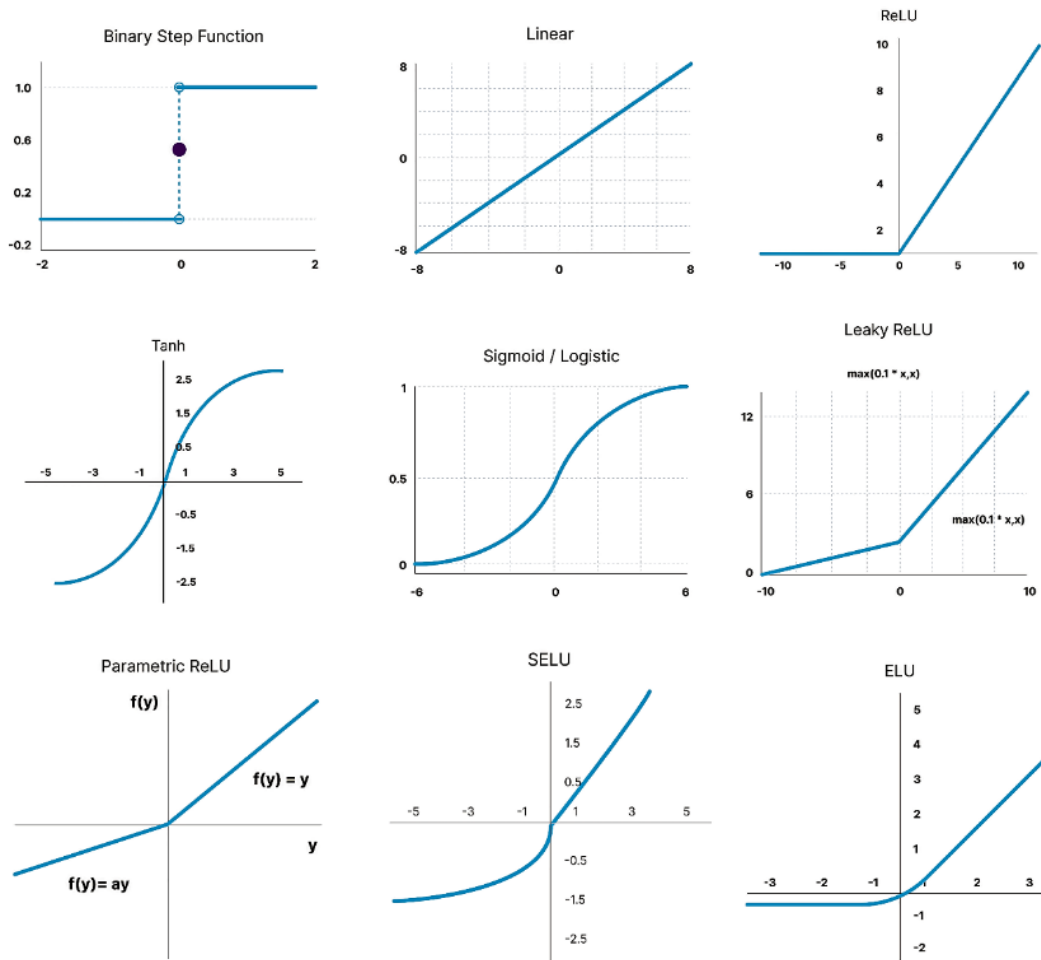
Οι συναρτήσεις ενεργοποίησης (Εικόνα 4-9) μπορούν να διαχωριστούν σε:

- Γραμμικές, όπως είναι οι βηματικές συναρτήσεις, οι συναρτήσεις κατωφλίου (threshold functions), οι συναρτήσεις προσήμου (sign), βηματικής μεταβολής (hard limiter), αναρρίχησης (ramping) κ.α.

- Μη γραμμικές, όπως είναι οι σιγμοειδείς συναρτήσεις (sigmoid functions) και οι Γκαουσιανές συναρτήσεις (Gaussian functions) κ.α. Αυτές οι συναρτήσεις προσφέρουν μη γραμμικότητα στο μοντέλο.

Κάποιες από τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης, είναι οι εξής:

- Γραμμική (Linear Activation Function): Στην συνάρτηση γραμμικής ενεργοποίησης η ενεργοποίηση είναι ανάλογη της εισόδου: $f(x) = x$. Δεν αλλάζει το σταθμισμένο άθροισμα της εισόδου και επιστρέφει απευθείας την τιμή, η οποία μπορεί να είναι μια οποιαδήποτε πραγματική τιμή. Χρησιμοποιείται σχεδόν αποκλειστικά σε προβλήματα παλινδρόμησης (regression).
- Σιγμοειδής - Λογιστική (Sigmoid - Logistic): Είναι μια μη γραμμική συνάρτηση με τη μορφή: $f(x) = 1 / (1 + e^{-x})$. Παίρνει οποιαδήποτε πραγματική τιμή ως είσοδο και εξάγει τιμές στο διάστημα $[0, 1]$.
- Υπερβολική Εφαπτομένη (Tanh Function - Hyperbolic Tangent): Παίρνει είσοδο έναν θετικό αριθμό και τον κανονικοποιεί στο διάστημα $[-1, 1]$, χρησιμοποιώντας την σχέση: $f(x) = (e^x - e^{-x}) / (e^x + e^{-x})$.
- Ράμπας ή Ανορθωτής (ReLU - Rectified Linear Unit): Η ReLU είναι η πιο δημοφιλής συνάρτηση ενεργοποίησης και μαθηματικά μπορεί να αναπαρασταθεί ως: $f(x) = \max(0, x)$. Διατηρεί την ενεργοποίηση οριοθετημένη στο μηδέν και είναι υπολογιστικά αποδοτική. Το κύριο μειονέκτημά της είναι ότι κατά την διάρκεια της εκπαίδευσης, τα βάρη μπορεί να ανανεώνονται με τέτοιο τρόπο, ώστε κάποιοι νευρώνες να μην ενεργοποιηθούν ποτέ (πρόβλημα Dying ReLU).
- Leaky ReLU (Parametric ReLU): Η συνάρτηση Leaky ReLU είναι μια βελτιωμένη έκδοση της ReLU για την επίλυση του προβλήματος Dying ReLU. Έχει μια μικρή θετική κλίση a στην αρνητική ($x < 0$) περιοχή: $f(x) = \max(a \cdot x, x)$. Η τιμή της σταθεράς a μπορεί να δοθεί σαν παράμετρος στον εκάστοτε νευρώνα.
- Softmax: Η συνάρτηση Softmax περιγράφεται ως ένας συνδυασμός πολλαπλών σιγμοειδών: $f(z_j) = e^{z_j} / \sum_k e^{z_k}$. Υπολογίζει τις σχετικές πιθανότητες, δηλαδή επιστρέφει την πιθανότητα κάθε κλάσης. Δέχεται είσοδο ένα διάνυσμα πραγματικών τιμών και το «συμπιέζει» σε ένα διάνυσμα με τιμές στον πιθανοτικό χώρο $[0, 1]$ (το άθροισμα των «νέων» τιμών είναι 1). Χρησιμοποιείται πιο συχνά ως συνάρτηση ενεργοποίησης για το τελευταίο επίπεδο του νευρωνικού δικτύου στην περίπτωση ταξινόμησης πολλαπλών κλάσεων.

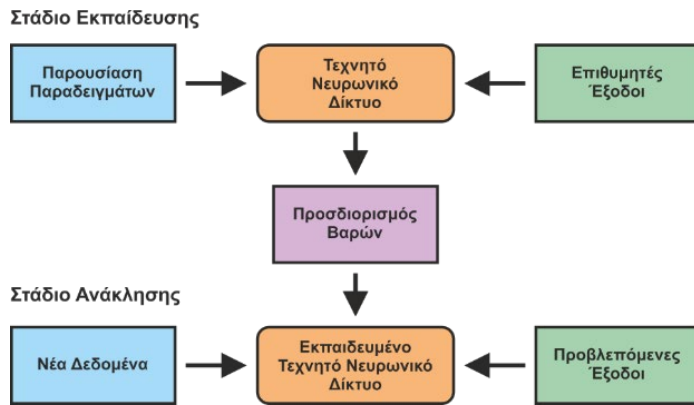


Εικόνα 4-9: Παραδείγματα βασικών συναρτήσεων ενεργοποίησης ΤΝΔ.

Κατά κανόνα, ανάλογα με τον τύπο του προς επίλυση προβλήματος χρησιμοποιούνται:

- για προβλήματα παλινδρόμησης (regression) - Γραμμική συνάρτηση ενεργοποίησης
- για δυαδικής ταξινόμησης (binary classification) - Σιγμοειδής
- για ταξινόμηση πολλαπλών κλάσεων (multiclass classification) - Softmax
- για ταξινόμηση πολλαπλών ετικετών (multilabel classification) - Σιγμοειδής

Αφού καθοριστεί η αρχιτεκτονική ενός ΤΝΔ, όσον αφορά τον τύπο τροφοδότησης και την εσωτερική του δομή, ακολουθούν δυο βασικά στάδια για την ολοκλήρωσή του: η εκπαίδευση (training) και η ανάκληση (recall) (Εικόνα 4-10). Κατά το στάδιο της εκπαίδευσης, τα συνοπτικά βάρη του δικτύου «μαθαίνουν» προσαρμόζοντας τα βάρη τους και έπειτα, στο στάδιο της ανάκλησης, με τη βοήθεια ενός νέου συνόλου δειγμάτων, τα οποία δεν είχαν πάρει μέρος στη διαδικασία εκπαίδευσης (validation set), γίνεται ο έλεγχος και η αξιολόγηση του δικτύου.



Εικόνα 4-10: Στάδια ολοκλήρωσης ΤΝΔ.

4.4 Εκπαίδευση ΤΝΔ

4.4.1 Αλγόριθμοι εκπαίδευσης

Η εκπαίδευση έχει ως στόχο την εύρεση του τρόπου με τον οποίο πρέπει να μεταβληθούν τα συνοπτικά βάρη των κόμβων του ΤΝΔ, ώστε να αλλάξει τη γενική συμπεριφορά του και να αυξηθεί η ικανότητά του να παρέχει την επιθυμητή έξοδο μετά από μία δεδομένη είσοδο. Βασικό στοιχείο της αρχιτεκτονικής ενός ΤΝΔ είναι ο τρόπος ελέγχου της αλλαγής των βαρών κατά την εκπαίδευση, δηλαδή ο αλγόριθμος εκπαίδευσης (training algorithm). Κάποιοι δημοφιλείς αλγόριθμοι εκπαίδευσης είναι ο Αλγόριθμος Οπισθοδιάδοσης Λάθους (Back-Propagation), η Ανταγωνιστική Μάθηση (Competitive Learning) και η Τυχαία μάθηση (Random Learning).

Κάθε αλγόριθμος εκπαίδευσης χαρακτηρίζεται από τους κανόνες εκμάθησης (learning rules) που χρησιμοποιεί, οι οποίοι υπολογίζουν τα σφάλματα και προσαρμόζουν τα συνοπτικά βάρη των εσωτερικών νευρώνων του δικτύου. Ο αλγόριθμος εκπαίδευσης που χρησιμοποιείται κατά κόρον για την εκπαίδευση νευρωνικών δικτύων πολλών επιπέδων είναι ο αλγόριθμος Back-Propagation (αλλιώς Γενικευμένος κανόνας Δέλτα), ο οποίος ανήκει στην κατηγορία των επιτηρούμενων δικτύων και θα αναλυθεί στην επόμενη υποενότητα.

Ο απλούστερος αλγόριθμος επιτηρούμενης εκπαίδευσης είναι αυτός του στοιχειώδους Perceptron, που μπορεί να εφαρμοστεί και σε οποιονδήποτε απλό νευρώνα ενός ΤΝΔ. Η βασική ιδέα της εκπαίδευσης ενός Perceptron περιλαμβάνει τα ακόλουθα βήματα:

- Έλεγχος αν η έξοδος του νευρώνα, για ένα συγκεκριμένο δείγμα εισόδου, είναι η αναμενόμενη.
- Εάν είναι, τότε: η διαδικασία εκπαίδευσης προχωρά στο επόμενο δείγμα.
- Εάν όχι, τότε:
 - Αν η σωστή έξοδος είναι μεγαλύτερη από αυτήν που υπολόγισε ο νευρώνας, ο κανόνας εκμάθησης αυξάνει τα βάρη των εισόδων που είναι θετικές και μειώνει τα βάρη των εισόδων που είναι αρνητικές.
 - Διαφορετικά, αν η έξοδος είναι μικρότερη, μειώνει τα βάρη των θετικών εισόδων και αυξάνει τα βάρη των αρνητικών εισόδων.
- Η διαδικασία αυτή επαναλαμβάνεται, μέχρις ότου ο νευρώνας να απαντά σωστά σε όλα τα δείγματα ή να μη βελτιώνει σημαντικά την απόδοσή του.

Τα βάρη κατά την εκκίνηση συνήθως αρχικοποιούνται σε μικρές αυθαίρετες τιμές. Αν γίνει αρχικοποίηση με μεγάλες τιμές επέρχεται κορεσμός (μικρή μεταβολή των συναπτικών βαρών), όχι γιατί έχει φτάσει σε καλό επίπεδο το νευρωνικό δίκτυο αλλά γιατί η έξοδος των νευρώνων είναι κοντά στο όριο της συνάρτησης ενεργοποίησης.

Υπάρχουν δυο τρόποι αλλαγής των βαρών κατά την παρουσίαση στο δίκτυο των παραδειγμάτων εκπαίδευσης:

- Αυξητική εκπαίδευση (incremental training), όπου για κάθε τιμή στην είσοδο υπολογίζονται και ταυτόχρονα εφαρμόζονται οι αλλαγές στα βάρη.
- Μαζική εκπαίδευση (batch training), όπου εισάγονται μια φορά όλες οι τιμές εκπαίδευσης στην είσοδο, υπολογίζονται οι αλλαγές στα βάρη για κάθε δείγμα και εφαρμόζονται ταυτόχρονα, αφού παρουσιαστούν όλα τα παραδείγματα.

Η παρουσίαση όλων των παραδειγμάτων από μια φορά (ανεξαρτήτως του τρόπου αλλαγής των βαρών) αντιστοιχεί σε έναν κύκλο εκπαίδευσης (ή εποχή - training epoch). Η αλλαγή του βάρους εισόδου από τον ένα νευρώνα στον επόμενο, εξαρτάται από το σφάλμα του προηγούμενου νευρώνα, την έξοδο του επόμενου νευρώνα και το ρυθμό μάθησης (learning rate). Ο τερματισμός της διαδικασίας εκπαίδευσης πραγματοποιείται, μετά από αρκετούς κύκλους εκπαίδευσης, όταν το συνολικό σφάλμα, για όλα τα παραδείγματα και για όλους τους νευρώνες εξόδου, πέσει κάτω από μια μικρή, προκαθορισμένη τιμή [105].

4.4.2 Αλγόριθμος Back-Propagation

Ο αλγόριθμος οπισθοδρομικής διάδοσης ή οπισθοδιάδοση (Back-Propagation Algorithm - BP) [96, 208], είναι μια αλγοριθμική διαδικασία, στη λογική της Gradient Descent, ενός αλγόριθμου βελτιστοποίησης, που υπολογίζει αναδρομικά τα συναπτικά βάρη του δικτύου, ώστε να επιτυγχάνεται η ελαχιστοποίηση της επιλεγμένης συνάρτησης απώλειας. Ο αλγόριθμος Gradient Descent περιγράφεται στην ενότητα 4.4.4.

Ο αλγόριθμος Back-Propagation περιλαμβάνει τα εξής βήματα:

- Επιλογή κατάλληλης συνάρτησης απώλειας, για παράδειγμα τη συνάρτηση MSE. Η επιλογή γίνεται λαμβάνοντας υπόψιν τις επιθυμητές και τις πραγματικές αποκρίσεις του δικτύου, για τα διαθέσιμα διανύσματα εκπαίδευσης.
- Επιλογή αλγόριθμου για τη βελτιστοποίηση της συνάρτησης σφάλματος, ως προς τα συναπτικά βάρη, για παράδειγμα μέσω του αλγόριθμου Gradient descent.

Σκοπός του αλγόριθμου Back-Propagation είναι να ελαχιστοποιήσει το σφάλμα, χρησιμοποιώντας τον αλγόριθμο Gradient Descent [99] ή και παραλλαγές του (SGD, AdaGrad, Adam κ.α.). Η διαδικασία ξεκινάει από το τελευταίο επίπεδο του δικτύου και κατευθύνεται προς τα προηγούμενα, προσαρμόζοντας κατάλληλα τα βάρη και τις πολώσεις του δικτύου, ώστε να επιτευχθεί η μέγιστη ακρίβεια του. Πρακτικά, εφαρμόζει τον κανόνα της αλυσιδωτής παραγωγής (gradient chain rule) για τον υπολογισμό των μερικών παραγώγων της συνάρτησης απώλειας. Σήμερα, ο αλγόριθμος Back-Propagation χρησιμοποιείται κατά κόρον για την εκπαίδευση μεγάλων νευρωνικών δικτύων με εκατομμύρια παραμέτρους. Η περαιτέρω ανάλυση του Back-Propagation ξεφεύγει από τα πλαίσια της παρούσας εργασίας.

4.4.3 Συναρτήσεις Απώλειας

Βασικός στόχος της διαδικασίας εκπαίδευσης ενός ΤΝΔ είναι η ελαχιστοποίηση της συνάρτησης απώλειας. Το πρόβλημα εκπαίδευσης του μοντέλου μπορεί επομένως να ερμηνευτεί ως ένα πρόβλημα βελτιστοποίησης στο οποίο γίνεται εύρεση του μέγιστου ή ελαχίστου μίας συνάρτησης. Κάθε αλγόριθμος βελτιστοποίησης χρησιμοποιεί μια συνάρτηση για την αξιολόγηση μιας πιθανής λύσης. Η συνάρτηση αυτή καλείται αντικειμενική συνάρτηση, και η βέλτιστη λύση είναι αυτή που την ελαχιστοποιεί σε προβλήματα ελαχιστοποίησης (ή το αντίθετο σε προβλήματα μεγιστοποίησης) [113].

Κατά τη διαδικασία εκπαίδευσης, αρχικά υπολογίζεται το σφάλμα του δικτύου μέσω της επιλεγμένης συνάρτησης απώλειας (συνάρτηση κόστους ή σφάλματος - loss function)⁸. Το σφάλμα αυτό υπολογίζεται συγκρίνοντας το αποτέλεσμα του με την επιθυμητή έξοδο, δηλαδή:

$$E_k = (a_{k,p} - o_{k,p}) \quad (4.4.1)$$

όπου: E_k το σφάλμα της εξόδου ενός νευρώνα k , $o_{k,p}$ η τιμή της εξόδου του νευρώνα k , $a_{k,p}$ η τιμή της επιθυμητής εξόδου, για ένα συγκεκριμένο δείγμα p .

Επίσης, μπορεί να οριστεί το συνολικό σφάλμα για όλα τα παραδείγματα εκπαίδευσης για ένα νευρώνα, ως το μέσο τετραγωνικό σφάλμα, δηλαδή:

$$E = \frac{1}{P} \sum_p (a_{k,p} - o_{k,p})^2 \quad (4.4.2)$$

Τέλος, μπορεί να οριστεί το μέσο τετραγωνικό σφάλμα για όλους τους νευρώνες:

$$E = \frac{1}{P * K} \sum_{p=1}^P \sum_{k=1}^K (a_{k,p} - o_{k,p})^2 \quad (4.4.3)$$

Στη συνέχεια για να επιτευχθεί η μεγαλύτερη δυνατή ακρίβεια του μοντέλου, η συνάρτηση απώλειας πρέπει να ελαχιστοποιηθεί και ιδανικά να μηδενιστεί.

Οι συναρτήσεις απώλειας (κόστους) υπολογίζουν το σφάλμα του δικτύου και έχουν στόχο να «τιμωρήσουν» τις λανθασμένες αποφάσεις που λαμβάνονται στην έξοδο του κατά την διάρκεια της εκπαίδευσης. Το σφάλμα αυτό, υπολογίζεται με βάση το σύνολο εκπαίδευσης και ονομάζεται σφάλμα εκπαίδευσης (training error). Η επιλογή της συνάρτησης απώλειας γίνεται κυρίως με βάση το είδος του προς επίλυση προβλήματος, αλλά καθορίζεται και από άλλους παράγοντες.

Τα ΤΝΔ χρησιμοποιούνται για την επίλυση προβλημάτων ταξινόμησης και προβλημάτων παλινδρόμησης, επομένως, χρησιμοποιούνται αντίστοιχα, δύο βασικές κατηγορίες συναρτήσεων: οι συναρτήσεις απώλειας ταξινόμησης και οι συναρτήσεις απώλειας παλινδρόμησης. Στόχος των συναρτήσεων απώλειας ταξινόμησης, είναι η πρόβλεψη του αποτελέσματος από διαφορετικές τιμές ομάδων, δηλαδή μίας διακριτής τιμής, η οποία αντιπροσωπεύει την κατηγορία που ανήκουν τα δεδομένα εισόδου. Τα προβλήματα ταξινόμησης διακρίνονται σε προβλήματα δυαδικής ταξινόμησης και σε προβλήματα ταξινόμησης πολλαπλών κλάσεων. Στη δυαδική ταξινόμηση τα δεδομένα

⁸ Το σφάλμα που παρουσιάζεται σε ένα μεμονωμένο παράδειγμα εκπαίδευσης, προκύπτει από τη συνάρτηση απώλειας (loss function). Το σφάλμα σε ολόκληρο το σύνολο δεδομένων ή την παρτίδα δεδομένων που χρησιμοποιείται προκύπτει από τη συνάρτηση κόστους (cost function).

διαχωρίζονται σε δύο κατηγορίες και οι τιμές εξόδου του μοντέλου μπορούν να λάβουν ένα εύρος τιμών από 0 έως 1. Στην περίπτωση της ταξινόμησης πολλαπλών κλάσεων, οι κατηγορίες στις οποίες έχουν ομαδοποιηθεί τα δεδομένα είναι περισσότερες των δύο και τα αποτελέσματα των προβλέψεων του μοντέλου συνήθως είναι ακέραιοι αριθμοί. Τα μοντέλα παλινδρόμησης είναι υπεύθυνα για την πρόβλεψη μίας ποσότητας πραγματικών αριθμών, οπότε χρησιμοποιείται μια συνάρτηση απώλειας παλινδρόμησης, της οποίας ο στόχος είναι η πρόβλεψη μιας συνεχούς τιμής, όπως για παράδειγμα η πρόβλεψη των καιρικών συνθηκών της επόμενης μέρας ή η μελλοντική τιμή μιας μετοχής.

Οι πιο δημοφιλείς συναρτήσεις απώλειας που χρησιμοποιούνται στα ΤΝΔ είναι:

A. Συναρτήσεις απώλειας παλινδρόμησης (Regression loss functions):

- Απόλυτου Μέσου Σφάλματος (Mean Absolute Error Loss - MAE ή L1 loss):

Η MAE, επιστρέφει το άθροισμα των απόλυτων διαφορών μεταξύ των επιθυμητών και των προβλεπόμενων τιμών, διαιρεμένο με το πλήθος των προβλέψεων που πραγματοποιήθηκαν από το μοντέλο:

$$MAE = \frac{1}{n} \sum_{t=0}^n |y_i - \hat{y}_i| \quad (4.4.4)$$

όπου: n είναι το σύνολο των προβλέψεων, y τα επιθυμητά αποτελέσματα και \hat{y} τα αποτελέσματα των προβλέψεων. Το εύρος των τιμών της κυμαίνεται από 0 μέχρι ∞ . Είναι υπολογιστικά γρήγορος τρόπος, αλλά ως γραμμική μέθοδος βαθμολόγησης λαμβάνει όλα τα σφάλματα στην ίδια κλίμακα.

- Μέσου Απόλυτου Ποσοστού Σφάλματος (MAPE):

Η MAPE είναι παραλλαγή της MAE και υπολογίζει το σφάλμα ως ποσοστό:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% \quad (4.4.5)$$

Είναι ανεξάρτητη από την κλίμακα των μεταβλητών. Η απώλεια υπολογίζεται κανονικοποιώντας όλα τα σφάλματα σε μια κοινή κλίμακα, αλλά δεν υπολογίζεται για μηδενικές επιθυμητές τιμές, ενώ επηρεάζεται και από το μέγεθός τους.

- Μέσου Τετραγωνικού Σφάλματος (Mean Squared Error Loss - MSE ή L2 loss):

Η MSE δίνει πάντα θετικές τιμές και υπολογίζεται από το μέσο όρο του τετραγώνου της διαφοράς μεταξύ των επιθυμητών τιμών και των τιμών που έχουν προβλεφθεί από το μοντέλο:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.4.6)$$

Αναπαριστώντας τις τιμές ως σημεία στο επίπεδο, η MSE, δίνει μια ευθεία η οποία διαμορφώνεται ώστε να ελαχιστοποιεί τις αποστάσεις των σημείων από την ευθεία. Για μικρά σφάλματα, η MSE βοηθά στη σύγκλιση, αλλά είναι ευαίσθητη σε ακραίες τιμές.

- Μέσου Τετραγωνικού Λογαριθμικού Σφάλματος (MSLE):

Η MSLE, είναι μία παραλλαγή της MSE που λαμβάνει υπόψη τη σχετική διαφορά μεταξύ προβλεπόμενης και πραγματικής τιμής:

$$\text{MSLE} = \frac{1}{n} \sum_{i=0}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2 \quad (4.4.7)$$

- Σφάλματος Ρίζας Μέσου Τετραγώνου (Root Mean Squared Error - RMSE):

Η RMSE είναι η τετραγωνική ρίζα της MSE και δίνει συγκριτικά μεγαλύτερη βαρύτητα σε μεγαλύτερα σφάλματα:

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (4.4.8)$$

Παρουσιάζει λιγότερες ακραίες απώλειες ακόμα και για μεγαλύτερες τιμές και είναι πιο ευαίσθητη σε ακραίες τιμές από το MAE.

B. Συναρτήσεις απώλειας ταξινόμησης (Classification loss functions):

Στα προβλήματα ταξινόμησης χρησιμοποιούνται συναρτήσεις απώλειας οι οποίες είναι παραλλαγές της εγκάρσιας εντροπίας (cross entropy). Η εγκάρσια εντροπία είναι η διαφορά μεταξύ δύο κατανομών πιθανοτήτων για μία τυχαία μεταβλητή ή για ένα σύνολο δεδομένων.

- Εγκάρσια Εντροπία Πολλαπλών Κλάσεων (Multiclass Cross Entropy Loss):

Χρησιμοποιείται στα προβλήματα ταξινόμησης πολλαπλών κατηγοριών:

$$\text{Loss} = -\sum_j y_{i,j} \log(\hat{y}_{i,j}) \quad (4.4.9)$$

όπου: i ο δείκτης του δείγματος στο σύνολο των δεδομένων εκπαίδευσης του μοντέλου και j ο δείκτης που αντιπροσωπεύει τις τιμές των επιθυμητών αποτελεσμάτων. Οι μεταβλητές y και \hat{y} εκφράζουν πιθανότητες.

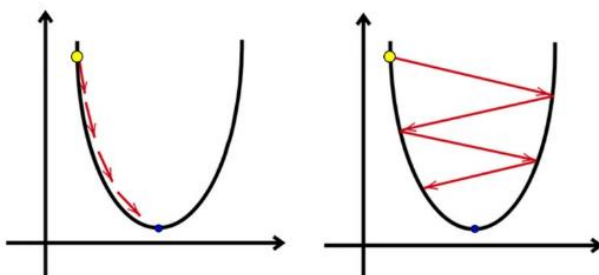
- Δυαδική Εγκάρσια Εντροπία (Binary Cross Entropy Loss / Log Loss):

Παραλλαγή της εγκάρσιας εντροπίας, χρησιμοποιείται για προβλήματα δυαδικής ταξινόμησης:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^n y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (4.4.10)$$

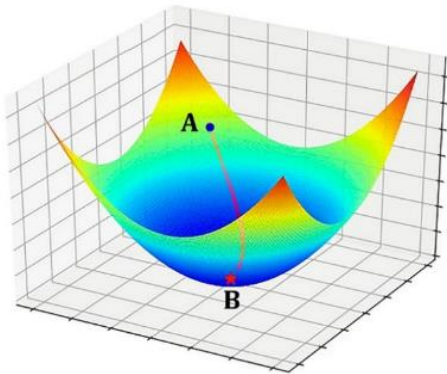
4.4.4 Αλγόριθμοι βελτιστοποίησης

Όπως περιγράφηκε στις προηγούμενες ενότητες, το πρόβλημα της εκπαίδευσης ενός ΤΝΔ αντιμετωπίζεται ως ένα πρόβλημα βελτιστοποίησης για τη εύρεση του μέγιστου ή ελαχίστου μίας συνάρτησης. Η συνάρτηση απώλειας ελαχιστοποιείται με την βοήθεια των αλγορίθμων βελτιστοποίησης (βελτιστοποιητές - optimizers). Ο απλούστερος και συγχρόνως ο πιο δημοφιλής αλγόριθμος βελτιστοποίησης είναι αλγόριθμος Gradient Descent (Κατάβαση Κλίσης ή αλγόριθμος σύγκλισης με ελάττωση της παραγώγου). Χρησιμοποιείται ευρέως σε προβλήματα γραμμικής παλινδρόμησης και ταξινόμησης, όπως επίσης και σε αλγόριθμους Οπισθοδιάδοσης (Back-Propagation) στα ΤΝΔ. Ο αλγόριθμος Gradient Descent εξαρτάται από την παράγωγο πρώτης τάξης (κλίση) της συνάρτησης απώλειας και υπολογίζει τον τρόπο με τον οποίο θα πρέπει να τροποποιηθούν τα βάρη ώστε να ελαχιστοποιηθεί η συνάρτηση. Μέσω της οπισθοδιάδοσης, η απώλεια μεταφέρεται από το ένα επίπεδο στο άλλο και οι παράμετροι του μοντέλου μεταβάλλονται αναλόγως με σκοπό την ελαχιστοποίηση της απώλειας. Το βήμα της κατάβασης ή ρυθμός μάθησης (learning rate) όπως συνήθως αναφέρεται είναι η κλίση η οποία υπολογίζεται από την μερική παράγωγο της συνάρτησης απώλειας σε σχέση με τα βάρη. Μικρός ρυθμός μάθησης συνεπάγεται ομαλή κάθοδο προς το τοπικό ελάχιστο, αλλά με περισσότερες επαναλήψεις, ενώ μεγάλος ρυθμός μάθησης συνεπάγεται ταχύτερη κάθοδο (μεγαλύτερα βήματα, λιγότερες επαναλήψεις), αλλά με αυξημένη πιθανότητα εμφάνισης ταλαντώσεων γύρω από το σημείο ελαχίστου, με αποτέλεσμα να το προσπεράσει. Όταν η τιμή του ρυθμού μάθησης έχει επιλεγθεί κατάλληλα τότε η τιμή της συνάρτησης απώλειας θα πρέπει να μειώνεται. Αν δεν παρατηρείται περαιτέρω μείωση της τιμής της τότε είναι πιθανό να έχει βρεθεί το επιθυμητό ελάχιστο σημείο (Εικόνα 4-11). Ένα παράδειγμα εφαρμογής της μεθόδου Gradient Descent παρουσιάζεται στην Εικόνα 4-12. Το σημείο Α είναι ένα τυχαίο σημείο που αντιπροσωπεύει την τιμή της συνάρτησης απώλειας πριν την εκπαίδευση του ΤΝΔ, ενώ το σημείο Β είναι το ελάχιστο σημείο της συνάρτησης, το οποίο θα προσεγγιστεί εφαρμόζοντας τον αλγόριθμο Gradient Descent.



Εικόνα 4-11: Πολύ μικρός ρυθμός μάθησης → πολλές επαναλήψεις (αριστερά), Πολύ μεγάλος μικρός ρυθμός μάθησης → Δε συγκλίνει στο ελάχιστο (δεξιά).

Αρχικά για να βρεθεί το ελάχιστο σημείο της συνάρτησης πρέπει να οριστεί ένα τυχαίο σημείο, αρχικοποιώντας αυθαίρετα τις τιμές των συναπτικών βαρών και των πολώσεων του δικτύου. Στη συνέχεια με την βοήθεια του Gradient Descent βρίσκεται το ολικό ελάχιστο της συνάρτησης απώλειας, κάνοντας βήματα προς την πιο απότομη κλίση (κόκκινη διαδρομή). Το μέγεθος του βήματος καθορίζεται από την τιμή του ρυθμού μάθησης.



Εικόνα 4-12: Εφαρμογή της μεθόδου κατάβασης κλίσης σε μία συνάρτηση απώλειας.

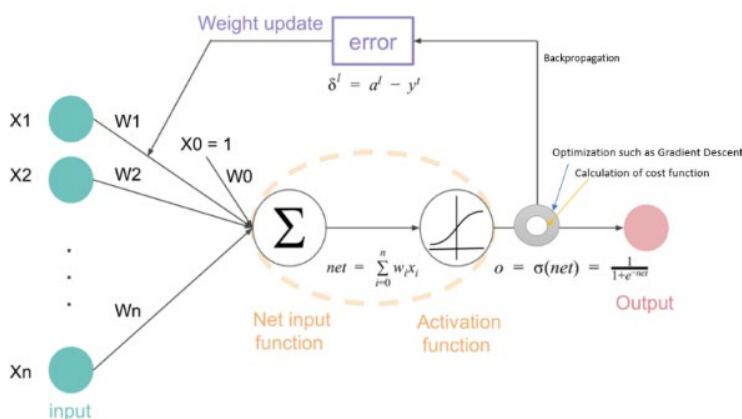
Υπάρχουν αρκετοί αλγόριθμοι βελτιστοποίησης με τους περισσότερους να αποτελούν παραλλαγές της Gradient Descent:

- Κατάβαση Κλίσης Μιας Παρτίδας (Batch Gradient Descent): Υπολογίζει το σφάλμα για κάθε παράδειγμα του συνόλου εκπαίδευσης, προσαρμόζοντας τις παραμέτρους του μοντέλου, μόνο αφού ολοκληρωθεί ένα πέρασμα από όλα τα δεδομένα.
- Στοχαστική Κατάβαση Κλίσης (Stochastic Gradient Descent - SGD): Αποτελεί παραλλαγή της Gradient Descent, η οποία ενημερώνει συχνότερα τις παραμέτρους του μοντέλου, τροποποιώντας τις μετά τον υπολογισμό απώλειας σε κάθε παράδειγμα εκπαίδευσης.
- Κατάβαση Κλίσης Μικρών Παρτίδων (Mini-Batch Gradient Descent): Συνδυασμός των παραπάνω αλγορίθμων. Το σύνολο των δειγμάτων προς εκπαίδευση διαχωρίζεται σε μικρότερα σύνολα τα οποία ονομάζονται παρτίδες (batches) και στη συνέχεια πραγματοποιείται η ενημέρωση των παραμέτρων για κάθε παρτίδα. Ο αριθμός των δειγμάτων σε κάθε παρτίδα (batch size) μπορεί να ρυθμιστεί.
- Προσαρμοστικός Αλγόριθμος Κλίσης (Adaptive Gradient - Adagrad): Προσαρμόζει το ρυθμό μάθησης των παραμέτρων, πραγματοποιώντας μικρότερες ενημερώσεις για παραμέτρους που αφορούν χαρακτηριστικά με υψηλή συχνότητα και μεγαλύτερες για παραμέτρους που αφορούν λιγότερα συχνά χαρακτηριστικά [114].
- RMS-Prop: Αποτελεί παραλλαγή του αλγορίθμου AdaGrad. Τροποποιεί το βήμα της εκπαίδευσης, ελαττώνοντάς το όταν υπάρχουν μεγάλες κλίσεις και αυξάνοντάς το όταν η κλίση είναι μικρή [115].

- Προσαρμοστική Εκτίμηση Στιγμής (Adaptive Moment Estimation - ADAM) [120, 121]: Είναι ένας αλγόριθμος κλίσης πρώτης τάξης που βασίζεται σε προσαρμοστικές εκτιμήσεις ροπών χαμηλότερης τάξης [124]. Συνδυάζει τα πλεονεκτήματα των αλγορίθμων AdaGrad και RMSProp και αποτελεί μια από τις πιο σύγχρονες και προσαρμοστικές μεθόδους βελτιστοποίησης.

4.4.5 Εκπαίδευση ΤΝΔ με Back-Propagation

Στα πολυεπίπεδα ΤΝΔ, ο συνηθέστερος τρόπος επιτηρούμενης μάθησης είναι η μάθηση με οπισθοδιάδοση σφάλματος (error back-propagation). Τα μοντέλα αυτά καλούνται Backpropagation ΤΝΔ (Backpropagation ANN). Κατά την εκπαίδευση των μοντέλων αυτών, για κάθε είσοδο που δίνεται στο δίκτυο, υπολογίζονται οι έξοδοι, σε κάθε κόμβο κρυφού ή εξωτερικού επιπέδου, με την εφαρμογή των συναρτήσεων ενεργοποίησης (Εικόνα 4-13). Για κάθε κόμβο εξωτερικού επιπέδου λαμβάνονται υπόψη οι διαφορές μεταξύ του υπολογιζόμενου και του επιθυμητού αποτελέσματος και διαδίδονται προς τα πίσω στους κόμβους των κρυφών επιπέδων, έτσι ώστε να καθορίσουν τις απαραίτητες αλλαγές στα βάρη σύνδεσης μεταξύ των κόμβων. Οι αλλαγές αυτές γίνονται με βάση τον κανόνα εκμάθησης οπισθοδιάδοσης και ως σκοπό έχουν τη μείωση του εμφανιζόμενου στην έξοδο σφάλματος, δηλαδή τη βελτιστοποίηση της επιλεγμένης συνάρτησης σφάλματος. Η συμμετοχή ενός κόμβου στα σφάλματα των κόμβων του επόμενου επιπέδου είναι ανάλογη της εισόδου του και των συντελεστών βαρύτητας που τον συνδέουν με τους κόμβους του επόμενου επιπέδου. Το δίκτυο στη συνέχεια εφαρμόζει εκ νέου τις συναρτήσεις ενεργοποίησης, για να υπολογίσει το νέο σφάλμα. Η διαδικασία της εκπαίδευσης περιλαμβάνει πλήθος τέτοιων κύκλων διόρθωσης σφάλματος, που ονομάζονται κύκλοι εκπαίδευσης (training epochs) και λήγει είτε με τη μείωση του σφάλματος κάτω από ένα επιθυμητό όριο, είτε με την πραγματοποίηση ενός συγκεκριμένου αριθμού κύκλων εκπαίδευσης.



Εικόνα 4-13: Λειτουργία Backpropagation αλγορίθμου.

4.4.6 Μέτρα αξιολόγησης μοντέλων

Μετά την εκπαίδευση του μοντέλου ακολουθεί η ανάκληση και η αξιολόγησή του. Ο έλεγχος της ποιότητας των αποτελεσμάτων και γενικά η απόδοση των μοντέλων ΤΝΔ γίνεται με τον υπολογισμό και την εκτίμηση του σφάλματος, χρησιμοποιώντας διάφορα μέτρα αξιολόγησης:

A. Για τα μοντέλα ταξινόμησης (Classification models) [125, 126]:

- Πίνακας Σύγχυσης (Confusion matrix): Με δεδομένες m κατηγορίες, ο πίνακας σύγχυσης είναι ένας $m \times m$ πίνακας, όπου το κάθε (i,j) στοιχείο του ισούται με το πλήθος των σημείων που ενώ προέρχονται από την κλάση i , καταχωρούνται στην κλάση j . Δίνει πληροφορίες σχετικά με το αν κάποιες κλάσεις έχουν τη τάση να συγχέονται με άλλες. Τα στοιχεία εντός διαγώνιου υποδηλώνουν σωστή πρόβλεψη για την κάθε κλάση, ενώ τα εκτός δηλώνουν λανθασμένη.
- Ακρίβεια Ταξινόμησης (Classification Accuracy): Το ποσοστό των σωστά ταξινομημένων δειγμάτων στο σύνολο των δειγμάτων.
- Ακρίβεια (Precision): Η ακρίβεια ως προς (κάθε) μια κατηγορία ορίζεται ως το πηλίκο των πραγματικά θετικών προς το σύνολο των πραγματικά θετικών και εσφαλμένων θετικών (της συγκεκριμένης κατηγορίας).
- Ανάκληση (Recall): Η ανάκληση ως προς (κάθε) μια κατηγορία ορίζεται ως το πηλίκο των πραγματικά θετικών προς τον αριθμό των συνολικά θετικών παραδειγμάτων (της συγκεκριμένης κατηγορίας).
- F1 score: Αποτελεί το αρμονικό μέσο ακρίβειας και ανάκλησης και είναι ένα συγκεντρωτικό μέτρο για την ικανότητα ταξινόμησης της (κάθε) κλάση. Συγκλίνει προς το μικρότερο εκ των δύο μέτρων λαμβάνοντας πιο αξιόπιστη βαθμολογία για το μοντέλο, ειδικά όταν τόσο η ανάκληση όσο και η ακρίβεια είναι σημαντικές.

Άλλα δημοφιλή μέτρα αξιολόγησης μοντέλων ταξινόμησης είναι τα: Specificity, Sensitivity, Συντελεστής Kappa, ROC Curve, Area Under The Curve (AUC).

B. Για τα μοντέλα παλινδρόμησης (Regression models):

Τα μέτρα που χρησιμοποιούνται για την αξιολόγηση των μοντέλων παλινδρόμησης πρέπει να μπορούν να λειτουργούν σε ένα σύνολο συνεχών τιμών. Παρότι τα μέτρα αξιολόγησης είναι διαφορετικά από τις συναρτήσεις απώλειας, πολλές από τις συναρτήσεις απώλειας μπορούν να χρησιμοποιηθούν και ως μέτρα αξιολόγησης. Τα μέτρα αξιολόγησης δεν χρειάζεται να είναι διαφοροποιήσιμα σε αντίθεση με τις συναρτήσεις απώλειας.

- Απόλυτο Μέσο Σφάλμα (Mean Absolute Error - MAE)
- Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error Loss - MSE)
- Σφάλμα Ρίζας Μέσου Τετραγώνου (Root Mean Squared Error - RMSE)
Λεπτομέρειες για τα παραπάνω μέτρα έχουν δοθεί στην υποενότητα 4.4.3.
- R-squared (R^2): Αποτελεί μια κανονικοποίηση της MSE:

$$R^2 = 1 - \frac{SSE}{TSS} \quad (4.4.11)$$

όπου: $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = n \cdot MSE$ και $TSS = \sum_{i=1}^n (y_i - \bar{y})^2$, με $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

Ενώ η RMSE προβλέπει την επιθυμητή τιμή σε απόλυτες τιμές (συγκρίσιμες με τις τιμές της μεταβλητής), το R-squared (R^2 score) προβλέπει την τιμή σε ποσοστιαίους όρους, δηλαδή δεν εξαρτάται από την κλίμακα των δεδομένων.

4.4.7 Υπερ-εκπαίδευση και Υπο-εκπαίδευση

Όπως αναφέρθηκε και προηγουμένως, μετά την εκπαίδευση ενός μοντέλου ΤΝΔ, ακολουθεί η αξιολόγησή του. Η αξιολόγηση του δικτύου πραγματοποιείται μετρώντας την απόδοσή του στο άγνωστο για το δίκτυο σύνολο δοκιμών, το σύνολο επικύρωσης (validation set), το οποίο δημιουργήθηκε ξεχωριστά από το σύνολο εκπαίδευσης. Το σφάλμα που προκύπτει χρησιμοποιώντας το σύνολο επικύρωσης ονομάζεται σφάλμα γενίκευσης (generalization error) και πρέπει να είναι χαμηλό [127]. Το σφάλμα γενίκευσης αναμένεται να είναι μεγαλύτερο ή ίσο του σφάλματος εκπαίδευσης, καθώς το δίκτυο έχει «μάθει» πάνω στο σύνολο εκπαίδευσης. Στο πλαίσιο αυτής της διαδικασίας εκπαίδευσης-επικύρωσης, οι παράγοντες που καθορίζουν πόσο καλά θα αποδώσει ένας μοντέλο ΤΝΔ, είναι η ικανότητά του να ελαχιστοποιεί το σφάλμα εκπαίδευσης, αλλά και να μειώνει τη διαφορά μεταξύ του σφάλματος εκπαίδευσης και του σφάλματος γενίκευσης.

Αυτοί οι δύο παράγοντες αντιστοιχούν σε δύο βασικά «εμπόδια» που εμφανίζονται κατά τη διαδικασία εκπαίδευσης των μοντέλων ΤΝΔ. Τα προβλήματα αυτά είναι η Υπερεκπαίδευση (Υπερπροσαρμογή - Overfitting) και η Υποεκπαίδευση (Underfitting) (Εικόνα 4-14).

- Υποεκπαίδευση (Underfitting): Η Υποεκπαίδευση συμβαίνει όταν η διαφορά μεταξύ σφάλματος εκπαίδευσης και σφάλματος γενίκευσης είναι πολύ μεγάλη. Προκύπτει όταν το μοντέλο δεν είναι σε θέση να καταγράψει τη μεταβλητότητα των δεδομένων

εκπαίδευσης και δε μπορεί να τα χαρτογραφήσει σωστά. Ένας τρόπος αντιμετώπισης της Υποεκπαίδευσης είναι η συλλογή περισσότερων και ποιοτικότερων δεδομένων.

- Υπερεκπαίδευση (Υπερπροσαρμογή - Overfitting): Η Υπερεκπαίδευση συμβαίνει όταν το μοντέλο δεν είναι σε θέση να αποκτήσει μια αρκετά χαμηλή τιμή σφάλματος στο σύνολο εκπαίδευσης. Δημιουργείται εξαιτίας της απομνημόνευσης των δεδομένων εκπαίδευσης και συνήθως οδηγεί σε πολύ καλή απόδοση στο σύνολο εκπαίδευσης αλλά σε κακή απόδοση στο σύνολο επικύρωσης. Η απώλεια της γενίκευσης του δικτύου μπορεί να οφείλεται σε πολλά ζητήματα, όπως η χωρητικότητα του δικτύου ή η φύση του ίδιου του συνόλου εκπαίδευσης. Το Overfitting χαρακτηρίζεται συνήθως από υψηλή διακύμανση (variance) στις τιμές των παραμέτρων και χαμηλά bias.

	Underfitting	Just right	Overfitting
Symptoms	<ul style="list-style-type: none"> • High training error • Training error close to test error • High bias 	<ul style="list-style-type: none"> • Training error slightly lower than test error 	<ul style="list-style-type: none"> • Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Deep learning illustration			
Possible remedies	<ul style="list-style-type: none"> • Complexify model • Add more features • Train longer 		<ul style="list-style-type: none"> • Perform regularization • Get more data

Εικόνα 4-14: Αναπαραστάσεις αποτελεσμάτων για μοντέλα με Underfitting και Overfitting.
Πηγή: [129]

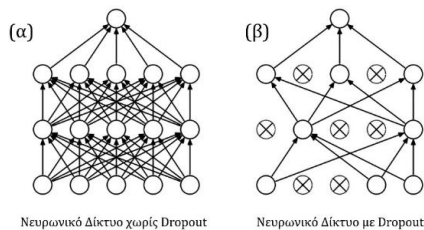
Για την αντιμετώπιση του προβλήματος της Υπερεκπαίδευσης έχουν παρουσιαστεί αρκετές προτάσεις στη βιβλιογραφία [113, 127, 130]. Παρακάτω αναφέρονται συνοπτικά κάποιες από αυτές τις τεχνικές γενίκευσης:

- **Αύξηση Συνόλου Δεδομένων:** Συλλογή περισσότερων και ποιοτικότερων δεδομένων. Όταν αυτό δεν είναι εφικτό, χρησιμοποιούνται τεχνικές επαύξησης δεδομένων (Data augmentation), οι οποίες αυξάνουν τον όγκο των δεδομένων με την προσθήκη ελαφρώς τροποποιημένων αντιγράφων των ήδη συλλεχθέντων δεδομένων.
- **Πρόωρη διακοπή (Early Stopping):** Διακοπή της εκπαίδευσης του δικτύου όταν η απόδοση στο σύνολο εκπαίδευσης σταματά να βελτιώνεται για έναν προκαθορισμένο αριθμό κύκλων εκπαίδευσης. Η διακοπή πολλές φορές γίνεται στο σημείο της εκπαίδευσης όπου ξεκινάει η απόκλιση των σφαλμάτων εκπαίδευσης και γενίκευσης.
- **Απόρριψη (Dropout):** Πραγματοποιεί «προσωρινή», απομάκρυνση κόμβων. Η τεχνική αναλύεται στην επόμενη ενότητα.
- **Κανονικοποίηση Βάρους (Ομαλοποίηση - Regularization):** Κατά τη διάρκεια της εκπαίδευσης, όταν το μοντέλο αρχίζει να υπερπροσαρμόζεται, τα βάρη των νευρώνων αρχίζουν να παίρνουν μεγάλες τιμές. Η ομαλοποίηση λειτουργεί με την επιβολή μιας παραμέτρου στη συνάρτηση απώλειας (προσθέτει μια ποινή βάρους), η οποία αναγκάζει τα βάρη να λαμβάνουν μόνο μικρές τιμές (Weight Regularization). Με αυτόν τον τρόπο τίθενται περιορισμοί στην πολυπλοκότητα (πλήθος και κατανομή των παραμέτρων) ενός δικτύου. Υπάρχουν δυο βασικές μέθοδοι Regularization: η L1 (LASSO Regression) και η εξέλιξη της L2 (Ridge Regression). Τέλος, μια τεχνική που επιφέρει παρόμοια αποτελέσματα ομαλοποιώντας τα δεδομένα, μειώνοντας τους απαιτούμενους κύκλους εκπαίδευσης και γενικότερα το σφάλμα γενίκευσης είναι η Κανονικοποίηση Παρτίδας (Batch Normalization). Η τεχνική αυτή κανονικοποιεί τις διαχωρισμένες σε παρτίδες εισόδους, δηλαδή τις μετασχηματίζει ώστε να διατηρήσει την μέση τιμή και την τυπική απόκλιση των ενεργοποιήσεων του προηγούμενου επιπέδου στις τιμές 0 και 1 αντίστοιχα.

4.4.8 Απόρριψη

Η Απόρριψη (Dropout) [131] είναι μια τεχνική αποτροπής της Υπερπροσαρμογής και είναι αποτελεσματική για σε όλες σχεδόν τις αρχιτεκτονικές ΤΝΔ. Η τεχνική αυτή πραγματοποιεί «προσωρινή», απομάκρυνση κόμβων, μαζί με όλες τις εισερχόμενες και εξερχόμενες συνδέσεις, από ένα ή περισσότερα επίπεδα ενός ΤΝΔ (Εικόνα 4-15). Αν και οι κόμβοι που θα απομακρυνθούν επιλέγονται τυχαία, το πλήθος τους μπορεί να

ρυθμίζεται ποσοστιαία σε κάθε επίπεδο. Η απόρριψη υλοποιείται μόνο για το σύνολο των δεδομένων εκπαίδευσης, ενώ για το σύνολο επικύρωσης τα δεδομένα διέρχονται από όλους τους κόμβους. Η εφαρμογή της παραμέτρου απόρριψης σε ένα νευρωνικό δίκτυο οδηγεί στη δειγματοληψία ενός αραιωμένου δικτύου σε σχέση με το αρχικό, το οποίο αποτελείται από τους νευρώνες εκείνους που «επιβίωσαν» της παραμέτρου απόρριψης. Έχει παρατηρηθεί πειραματικά ότι η εκπαίδευση ενός νευρωνικού δικτύου με χρησιμοποίηση Dropout οδηγεί σε σημαντική μείωση του σφάλματος γενίκευσης σε μια πληθώρα προβλημάτων.



Εικόνα 4-15: Νευρωνικό δίκτυο χωρίς και με εφαρμογή Dropout.

4.4.9 Βελτιστοποίηση Υπερπαραμέτρων

Ο σχεδιασμός ενός ΤΝΔ έγκειται ουσιαστικά στην επιλογή των υπερπαραμέτρων του. Δεν υπάρχει μια καθιερωμένη τυποποιημένη διαδικασία που να επιτρέπει τον αποτελεσματικό σχεδιασμό ενός δικτύου χωρίς ανθρώπινη παρέμβαση. Ως εκ τούτου ο αποτελεσματικός σχεδιασμός ενός ΤΝΔ αποτελεί μια απαιτητική διαδικασία, η οποία μέχρι και σήμερα αντιμετωπίζεται κυρίως ως μια «τέχνη», όπου η εμπειρία και η ικανότητα του σχεδιαστή του παίζει πολύ σημαντικό ρόλο. Η επιλογή του κατάλληλου συνδυασμού υπερπαραμέτρων είναι εξαιρετικά δύσκολη υπόθεση, λόγω του ιδιαίτερα μεγάλου αριθμού τους. Πρόσφατες μελέτες στα βαθιά δίκτυα αποδεικνύουν ότι πολλές φορές η προσεκτική επιλογή των υπερπαραμέτρων ενός μοντέλου, επιφέρει καλύτερα αποτελέσματα από τη χρήση νέων μοντέλων [113, 132].

Λόγω των παραπάνω, έχουν αναπτυχθεί αλγόριθμοι αυτόματης βελτιστοποίησης των υπερπαραμέτρων, οι οποίοι επιτρέπουν την επιλογή υπερπαραμέτρων χωρίς την ανθρώπινη παρέμβαση. Αυτοί οι αλγόριθμοι δεν ενδιαφέρονται για την εσωτερική λειτουργία του μοντέλου, αλλά μόνο για την επιρροή που ασκούν οι υπερπαραμέτροι στην απόδοση του. Υπάρχουν δύο είδη αλγορίθμων βελτιστοποίησης υπερπαραμέτρων: οι αλγόριθμοι αναζήτησης και οι αλγόριθμοι βελτιστοποίησης.

- Αλγόριθμοι αναζήτησης: Αναζητούν τη λύση χωρίς την χρήση κάποιας ευρετικής μεθόδου που να βελτιώνει την ποιότητα των προτεινόμενων λύσεων σταδιακά. Αυτό

το χαρακτηριστικό τους καθιστά μη αποδοτικούς για μεγάλο αριθμό υπερπαραμέτρων. Παρέχουν όμως ευκολία στη υλοποίηση τους. Παραδείγματα αποτελούν ο αλγόριθμος Τυχαίας Αναζήτησης (Random Search) και η Αναζήτηση Πλέγματος (Grid Search).

- Αλγόριθμοι βελτιστοποίησης: Η επιλογή της κάθε προτεινόμενης λύσης πραγματοποιείται με βάση προηγούμενες πληροφορίες που έχει συλλέξει ο αλγόριθμος. Οι αλγόριθμοι αυτοί είναι πιο σύνθετοι και η υλοποίηση τους πιο πολύπλοκη. Παραδείγματα αποτελούν η βελτιστοποίηση Bayesian [133], οι Γενετικοί αλγόριθμοι [134] και η βελτιστοποίηση Σμήνους Σωματιδίων (Particle Swarm Optimization) [135].

4.5 Συνελικτικά νευρωνικά δίκτυα (CNN)

Στην παρούσα εργασία χρησιμοποιήθηκαν Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) [113, 136], τα οποία αποτελούν μια κατηγορία βαθέων νευρωνικών δικτύων. Το πρώτο συνελικτικό μοντέλο προτάθηκε από τον Fukushima (1980) [137], οποίος στηρίχθηκε στο μοντέλο που πρότειναν οι Hubbel και Weisel (1958) [138], μετά από τη μελέτη τους πάνω στον οπτικό φλοιό διάφορων ζώων. Τα CNN έχουν εφαρμοστεί με επιτυχία σε διάφορες εφαρμογές υπολογιστικής όρασης (Computer Vision) και ειδικά στην αναγνώριση και στον εντοπισμό αντικειμένων σε εικόνες [137, 140, 141, 142]. Τα συνελικτικά δίκτυα είναι πολυεπίπεδα δίκτυα πρόσθιας τροφοδότησης και είναι σε θέση να αποτυπώσουν τις χωρικές και χρονικές εξαρτήσεις των δεδομένων εισόδου, συνήθως εικόνων, μέσω της εφαρμογής σχετικών φίλτρων. Τα διάφορα επίπεδα των CNN εξάγουν όλο και πιο περίπλοκα χαρακτηριστικά, μεγαλύτερα στοιχεία-τμήματα, για παράδειγμα ενός αντικειμένου σε μια εικόνα, ξεκινώντας από τα πιο απλά και αφαιρετικά, όπως ακμές και χρώματα. Καθώς τα δεδομένα εικόνας διαδίδονται δια μέσω των επιπέδων, το δίκτυο διαμορφώνει τα βάρη και τις πολώσεις σε κάθε επίπεδο και τελικά μαθαίνει να αναγνωρίζει και να ταξινομεί το αντικείμενο που απεικονίζεται, στην εικόνα που δέχεται στη είσοδο. Η προεπεξεργασία που απαιτείται σε ένα CNN είναι πολύ χαμηλότερη σε σύγκριση με άλλους αλγόριθμους. Παρέχουν καλύτερα αποτελέσματα από άλλα νευρωνικά δίκτυα ειδικά στις περιπτώσεις όπου η είσοδος είναι στη μορφή σήματος εικόνας, ομιλίας ή ήχου, καθώς αξιοποιούν αρχές γραμμικής άλγεβρας, όπως τον πολλαπλασιασμό πινάκων, ώστε να εντοπίζουν τα διάφορα μοτίβα μέσα στα δεδομένα εισόδου. Σε αρκετές όμως περιπτώσεις τα CNN έχουν μεγάλο υπολογιστικό κόστος, απαιτώντας μονάδες επεξεργασίας γραφικών (GPU) για τη γρήγορη εκπαίδευση των μοντέλων τους.

4.5.1 Αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων

Τα CNN έχουν τρεις βασικούς τύπους επιπέδων:

A. Συνελκτικό επίπεδο (Επίπεδο Συνέλιξης - Convolutional layer)

B. Επίπεδο συγκέντρωσης (Επίπεδο Υπόδειγματοληψίας - Pooling layer)

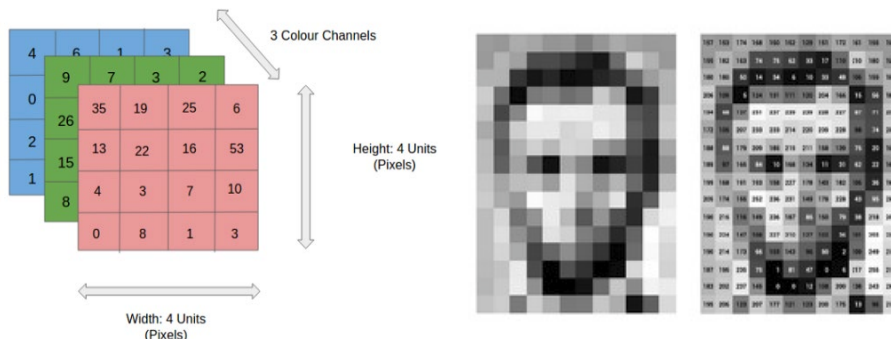
Γ. Πλήρως συνδεδεμένο επίπεδο (Fully-connected layer - FC ή Dense)

Το Συνελκτικό επίπεδο αποτελεί το πρώτο επίπεδο ενός CNN. Έπειτα, μπορεί να ακολουθούν πρόσθετα συνελκτικά επίπεδα αλλά και στρώματα συγκέντρωσης, αυξάνοντας έτσι την πολυπλοκότητά του. Το τελευταίο επίπεδο είναι ένα πλήρως συνδεδεμένο επίπεδο.

Παίρνοντας ως παράδειγμα εφαρμογής, την περίπτωση όπου η είσοδος του δικτύου είναι μια έγχρωμη εικόνα, η οποία αναπαρίσταται με ένα αριθμητικό πίνακα εικονοστοιχείων (pixels) τριών διαστάσεων (Εικόνα 4-16), μπορεί να γίνει αναλυτική περιγραφή της δομής και του τρόπου λειτουργίας κάθε ενός από τους βασικούς τύπους επιπέδων ενός CNN.

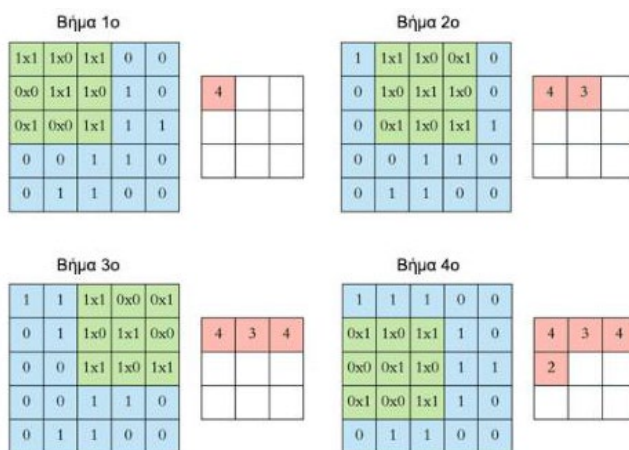
4.5.2 Συνελκτικό Επίπεδο

Το Συνελκτικό Επίπεδο (Convolutional layer) είναι το βασικό δομικό στοιχείο ενός CNN, καθώς εκεί λαμβάνει χώρα η πλειονότητα των υπολογισμών. Απαιτεί δεδομένα εισόδου (input data), χρησιμοποιεί ένα φίλτρο (filter ή πυρήνας - kernel) και εξάγει έναν χάρτη χαρακτηριστικών (feature map).



Εικόνα 4-16: Εικόνα RGB ως τρισδιάστατος πίνακας, 4x4 εικονοστοιχείων, με 3 τιμές χρωμάτων για κάθε ένα (αριστερά). Εικόνα σε Grayscale (δεξιά).

Το φίλτρο (filter ή πυρήνας - kernel) είναι ένας ανιχνευτής χαρακτηριστικών που σαρώνει σε περιοχές της εικόνας, ελέγχοντας εάν υπάρχει το χαρακτηριστικό. Αυτή η διαδικασία είναι γνωστή ως συνέλιξη. Η συνέλιξη είναι μία μαθηματική πράξη μεταξύ δύο συναρτήσεων, η οποία έχει ως αποτέλεσμα μία νέα συνάρτηση που δείχνει κατά πόσο επικαλύπτονται οι δύο αρχικές συναρτήσεις. Κύριος σκοπός της συνέλιξης σε ένα CNN, είναι η εξαγωγή χαρακτηριστικών. Η συνέλιξη διατηρεί τη χωροταξική σχέση μεταξύ των pixels, μαθαίνοντας τα χαρακτηριστικά της εικόνας. Συγκεκριμένα, ο ανιχνευτής χαρακτηριστικών είναι μια δισδιάστατη (στη συγκεκριμένη περίπτωση που εξετάζουμε) διάταξη βαρών (weights), δηλαδή ένας πίνακας μικρότερων διαστάσεων από την εικόνα, με μέγεθος συνήθως από 2x2 έως 7x7. Το φίλτρο εφαρμόζεται σε μια περιοχή της εικόνας και υπολογίζεται το εσωτερικό γινόμενο (dot product) μεταξύ των εικονοστοιχείων εισόδου. Δηλαδή το γινόμενο του πίνακα που αντιστοιχεί σε μια περιοχή της εικόνας και του φίλτρου, δηλαδή του πίνακα του ανιχνευτή. Στη συνέχεια, το φίλτρο μετατοπίζεται (Εικόνα 4-17) κατά ένα βήμα (stride), επαναλαμβάνοντας τη διαδικασία μέχρις ότου το φίλτρο σαρώσει ολόκληρη την εικόνα. Η απόσταση, ο αριθμός δηλαδή των pixels κατά τον οποίο μετακινείται το φίλτρο πάνω από τον πίνακα εισόδου, ονομάζεται βήμα μετατόπισης (stride) και αποτελεί επίσης μία υπερπαράμετρο των επιπέδων συνέλιξης. Η τελική έξοδος, μετά από μια σειρά τέτοιων γινομένων, είναι μια νέα εικόνα, ένας νέος πίνακας, που ονομάζεται χάρτης χαρακτηριστικών (feature map ή χάρτης ενεργοποίησης - activation map ή συνελιγμένο χαρακτηριστικό - convolved feature).

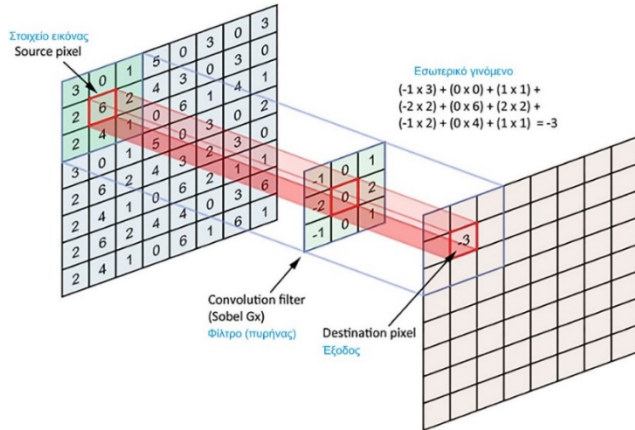


Εικόνα 4-17: Διαδικασία Συνέλιξης - Σάρωση φίλτρου.

Πηγή: [143]

Ο χάρτης που παράγεται είναι αποτέλεσμα της συνέλιξης και είναι μικρότερος σε διαστάσεις από την αρχική εικόνα. Στο χάρτη αυτό «αποτυπώνονται» με μεγαλύτερες τιμές τα χαρακτηριστικά τα οποία είναι πιο «ευαίσθητα» στο συγκεκριμένο φίλτρο που εφαρμόστηκε. Στη Εικόνα 4-18, φαίνονται οι υπολογισμοί που γίνονται κατά τη διαδικασία της συνέλιξης (πολλαπλασιασμός στοιχείο προς στοιχείο του φίλτρου με το

αντίστοιχο τμήμα της εικόνας και έπειτα άθροιση των στοιχείων) για την παραγωγή ενός μοναδικού στοιχείου. Είναι προφανές ότι ανάλογα με τις τιμές του φίλτρου θα παραχθεί και διαφορετικός χάρτης χαρακτηριστικών (Εικόνα 4-19). Η μείωση των διαστάσεων χάρτη χαρακτηριστικών σε σχέση με την αρχική εικόνα, οφείλεται στον τρόπο με τον οποίο εκτελείται η πράξη της συνέλιξης (Εικόνα 4-20).



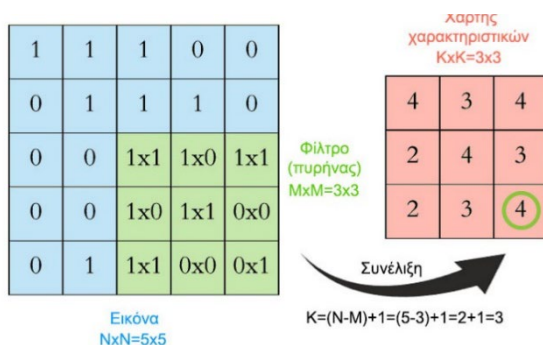
Εικόνα 4-18: Διαδικασία συνέλιξης & χάρτης χαρακτηριστικών.

Πηγή: [143]

Operation	Filter	Convolved Image	Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$		Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$		Box blur (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$		Gaussian blur (approximation)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$				

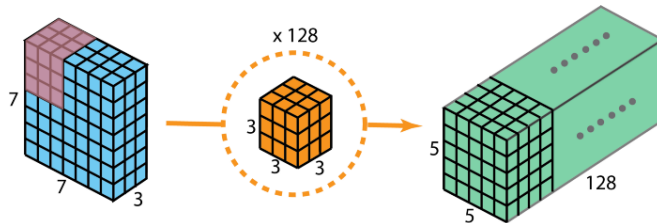
Εικόνα 4-19: Διαφορετικά είδη φίλτρων (πυρήνων) και εξαγωγή διαφορετικών χαρακτηριστικών ανάλογα με τις τιμές του φίλτρου.

Πηγή: [144]



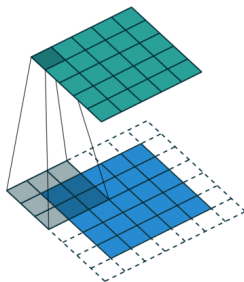
Εικόνα 4-20: Συνέλιξη και διαστάσεις χάρτη χαρακτηριστικών.

Η τιμή του βάθους στην έξοδο ενός επιπέδου συνέλιξης αντιστοιχεί στον αριθμό των φίλτρων που εφαρμόζονται στον πίνακα της εικόνα εισόδου. Δηλαδή ο αριθμός των χαρτών ενεργοποίησης αντιστοιχεί στον αριθμό των φίλτρων. Για παράδειγμα αν η είσοδος είναι $7 \times 7 \times 3$ και εφαρμοστούν 128 φίλτρα συνέλιξης διαστάσεων $3 \times 3 \times 3$, τότε ο πίνακας εξόδου θα είναι $5 \times 5 \times 128$. Ο αριθμός των φίλτρων αποτελεί μία υπερπαραμέτρο των επιπέδων συνέλιξης (Εικόνα 4-21).



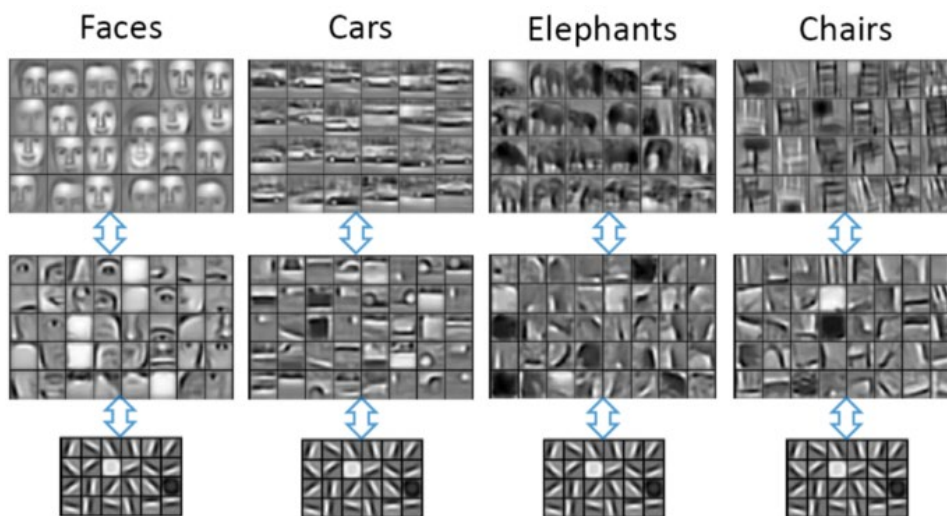
Εικόνα 4-21: Εφαρμογή 128 φίλτρων $3 \times 3 \times 3$ σε είσοδο $7 \times 7 \times 3$ (3D).

Επίσης, ένα συχνό πρόβλημα που εμφανίζεται στα μοντέλα CNN με μεγάλο αριθμό επιπέδων είναι η γρήγορη μείωση των διαστάσεων (μήκους και πλάτους) του πίνακα εισόδου, το οποίο είναι αποτέλεσμα της διαδοχικής εφαρμογής διαδικασιών συνέλιξης, περιορίζοντας έτσι, τις διαστάσεις των φίλτρων που μπορούμε να χρησιμοποιήσουμε σε κάθε συνελκτικό επίπεδο. Για την αποφυγή αυτού του προβλήματος χρησιμοποιείται μια διαδικασία που ονομάζεται Μηδενική συμπλήρωση (Zero-padding), με την οποία επεκτείνονται οι διαστάσεις του πίνακα εισόδου, προσθέτοντας μηδενικά στα σύνορά του. Χρησιμοποιείται επίσης όταν τα φίλτρα δεν «ταιριάζουν» στην εικόνα εισόδου, ώστε να αυξηθεί ή να μειωθεί η διάσταση του συνελκμένου χαρακτηριστικού, μηδενίζοντας όλα τα στοιχεία που βρίσκονται εκτός του πίνακα εισόδου (Εικόνα 4-22). Το μέγεθος του συνόρου που προστίθεται αποτελεί και αυτό μία υπερπαραμέτρο των επιπέδων συνέλιξης. Υπάρχουν τρεις τύποι συμπλήρωσης: Valid Padding (Εγκυρη πλήρωση), στην οποία ο χάρτης χαρακτηριστικών είναι μειωμένος σε διαστάσεις σε σύγκριση με την είσοδο, Same Padding (Ίδια πλήρωση), η διάσταση είτε αυξάνεται είτε παραμένει η ίδια και Full padding (Πλήρης επένδυση), όπου το μέγεθος της εξόδου αυξάνει προσθέτοντας μηδενικά στα όρια της εισόδου.



Εικόνα 4-22: Same padding: Ο πίνακας $5 \times 5 \times 1$ που αντιστοιχεί στην εικόνα εισόδου, συμπληρώνεται με 0 για να δημιουργηθεί ένας πίνακας $6 \times 6 \times 1$.

Μετά από κάθε διαδικασία συνέλιξης, το CN συνήθως χρησιμοποιεί μια συνάρτηση ενεργοποίησης Rectified Linear Unit (ReLU), εφαρμόζοντας έναν μετασχηματισμό στον χάρτη χαρακτηριστικών και εισάγοντας έτσι μη γραμμικότητα στο μοντέλο. Η ReLU αντικαθιστά όλες τις αρνητικές τιμές του χάρτη με 0. Εάν ακολουθεί και άλλο επίπεδο συνέλιξης, η δομή του δικτύου γίνεται ιεραρχική, καθώς τα μεταγενέστερα επίπεδα, δέχονται τα απλούστερα χαρακτηριστικά των προηγούμενων επιπέδων και εντοπίζουν πιο πολύπλοκα χαρακτηριστικά του σήματος εισόδου (της εικόνας). Τελικά, το συνελκτικό επίπεδο μετατρέπει την εικόνα σε αριθμητικές τιμές, επιτρέποντας στο νευρωνικό δίκτυο να ερμηνεύει και να εξάγει σχετικά μοτίβα (Εικόνα 4-23).



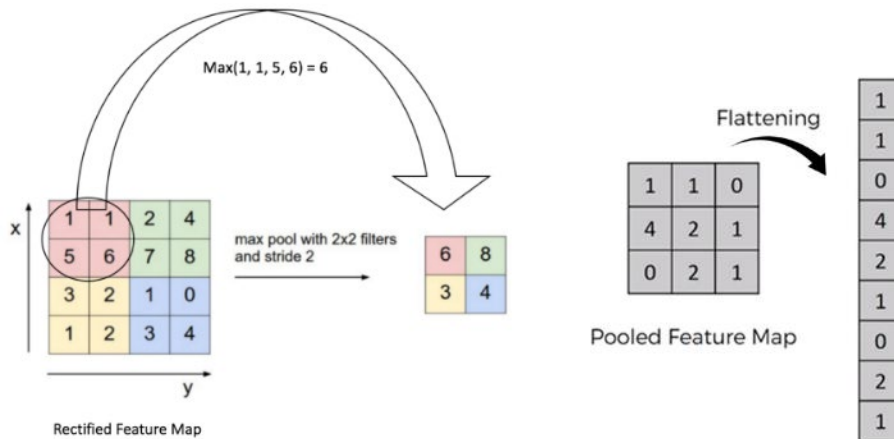
Εικόνα 4-23: Ένας τρόπος απεικόνισης των χαρακτηριστικών που αναζητούν τα διαφορετικά κρυφά επίπεδα σε ένα νευρωνικό δίκτυο.

4.5.3 Επίπεδο συγκέντρωσης

Το Επίπεδο συγκέντρωσης (Pooling layer), γνωστό και ως downsampling (υποδειγματοληψία), πραγματοποιεί μείωση διαστάσεων, ελαττώνοντας τον αριθμό των παραμέτρων στην είσοδο, άρα και των απαιτούμενων υπολογισμών. Αν και στο επίπεδο συγκέντρωσης χάνεται αρκετή πληροφορία, η εφαρμογή του συμβάλλει στη μείωση της πολυπλοκότητας του δικτύου, στη βελτίωση της απόδοσής του και στον περιορισμό του κινδύνου υπερβολικής προσαρμογής (Overfitting).

Με παρόμοια λειτουργία όπως αυτή του συνελκτικού επιπέδου, η λειτουργία συγκέντρωσης χρησιμοποιεί ένα φίλτρο που σαρώνει ολόκληρη την είσοδο, αλλά χωρίς την προσθήκη βαρών. Αντίθετα, ο πυρήνας εφαρμόζει μια συνάρτηση συνάθροισης στις τιμές της περιοχής που ελέγχει, συμπληρώνοντας τον πίνακα εξόδου (Εικόνα 4-24,

αριστερά). Υπάρχουν δύο κύριοι τύποι συγκέντρωσης: Max pooling (Μέγιστη συγκέντρωση), κατά την οποία το φίλτρο μετακινείται κατά μήκος της εισόδου και επιλέγει το εικονοστοιχείο με τη μέγιστη τιμή για αποστολή στον πίνακα εξόδου, και Average pooling (Μέση συγκέντρωση), στην οποία το φίλτρο υπολογίζει και αποστέλλει τη μέση τιμή.



Εικόνα 4-24: Max Pooling με φίλτρο 2x2 και βήμα 2 πάνω σε ένα 4x4 convolved feature (αριστερά). Το επίπεδο Flattening (δεξιά).

Μετά την ολοκλήρωση των προηγούμενων βημάτων, ακολουθεί το επίπεδο ισοπέδωσης (flattening layer), στο οποίο ο χάρτης των συγκεντρωμένων χαρακτηριστικών «ισοπεδώνεται» σε μια στήλη (1D διάνυσμα), προκειμένου να τροφοδοτήσει στην έξοδο του το Πλήρως Συνδεδεμένο Επίπεδο (Fully-Connected Layer) που ακολουθεί (Εικόνα 4-24, δεξιά).

4.5.4 Πλήρως συνδεδεμένο επίπεδο

Στο Πλήρως συνδεδεμένο επίπεδο (Fully-Connected Layer - FC) κάθε κόμβος του συνδέεται απευθείας με έναν κόμβο του προηγούμενου επιπέδου, όπως γίνεται στα απλά πολυεπίεδα ΤΝΔ και συνήθως έχει τόσους νευρώνες όσες και οι κλάσεις της πρόβλεψης. Το FC επίπεδο εκτελεί το έργο της ταξινόμησης με βάση τα χαρακτηριστικά που εξάγονται μέσω των προηγούμενων επιπέδων και της εφαρμογής των διαφορετικών φίλτρων τους. Ενώ τα επίπεδα συνέλιξης και συγκέντρωσης τείνουν να χρησιμοποιούν συναρτήσεις ReLu, τα επίπεδα FC χρησιμοποιούν συνήθως είτε μια συνάρτηση ενεργοποίησης Softmax για την ταξινόμηση των εισόδων, εξάγοντας μια πιθανότητα από 0 έως 1, είτε μια γραμμική συνάρτηση ενεργοποίησης (Linear) για προβλήματα παλινδρόμησης, εξάγοντας μια μόνο συνεχή πραγματική τιμή.

4.5.5 Διαστάσεις εξόδου & πλήθος παραμέτρων

Σε αυτό το σημείο είναι σκόπιμο να γίνει μια περιγραφή του τρόπου που υπολογίζονται οι διαστάσεις του πίνακα εξόδου σε κάθε επίπεδο συνέλιξης και το πλήθος των παραμέτρων (βάρη και πλώσεις) που δημιουργούνται [145].

Επομένως έστω ότι: $I \times I \times D_i$: οι διαστάσεις του πίνακα εισόδου, $F \times F \times D_f$: το μέγεθος του φίλτρου (filter/kernel), S : το βήμα μετατόπισης (Stride), P : το μέγεθος συμπλήρωσης (Padding) και D αριθμός φίλτρων (βάθος χάρτη χαρακτηριστικών):

Τότε οι διαστάσεις του πίνακα εξόδου θα είναι:

$$\text{Convolution Output dimension} = [(I - F + 2 * P) / S] + 1 \times D \quad (4.5.1)$$

και ο αριθμός των παραμέτρων (με num_bias όρους πλώσης) θα είναι:

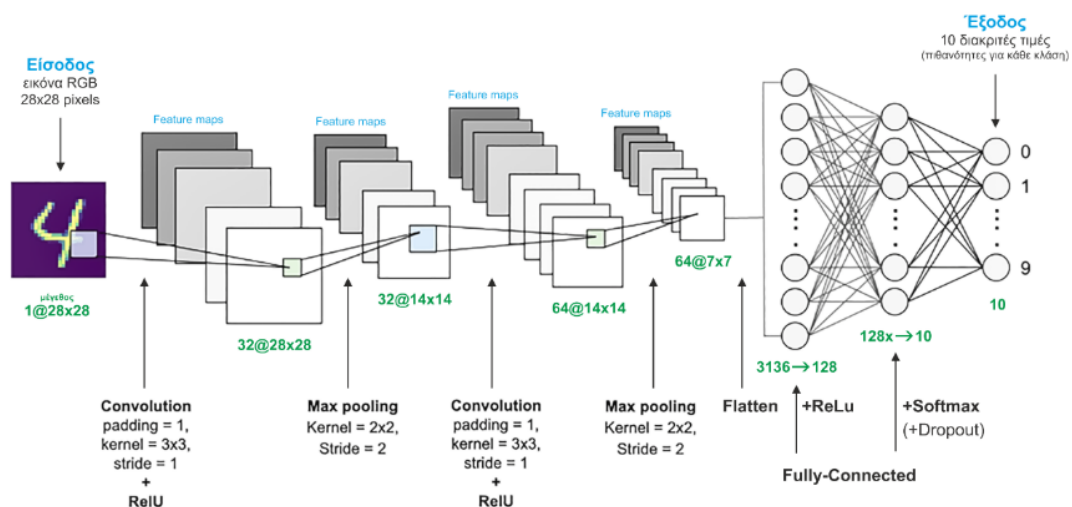
$$\text{Number of Parameter} = (F * F * D_i + \text{num_bias}) * D \quad (4.5.2)$$

Επίσης, όταν ο πίνακας εξόδου ενός επίπεδου συνέλιξης, περάσει από ένα επίπεδο συγκέντρωσης (Max-Pooling) τότε ο πίνακας εξόδου του θα είναι:

$$\text{Pooling Output dimension} = [(I - F) / S] + 1 \times D \quad (4.5.3)$$

και δεδομένου ότι η λειτουργία συγκέντρωσης είναι μια σταθερή συνάρτηση, δεν εισάγει πρόσθετες παραμέτρους.

Στην Εικόνα 4-25, παρουσιάζεται μια τυπική αρχιτεκτονική CNN μοντέλου για ταξινόμηση χειρόγραφων ψηφίων, όπου αναγράφονται οι διαστάσεις των πινάκων εξόδου και το πλήθος των παραμέτρων σε κάθε επίπεδο.



Εικόνα 4-25: Αρχιτεκτονική CNN για ταξινόμηση χειρόγραφων ψηφίων. Αναγράφονται οι διαστάσεις των πινάκων εξόδου και το πλήθος των παραμέτρων σε κάθε επίπεδο.

4.5.6 Συνελκτικά δίκτυα σε εφαρμογές ήχου

Τα CNN έχουν σχεδιαστεί με έμπνευση το ανθρώπινο οπτικό σύστημα και παρέχουν εξαιρετικά αποτελέσματα στην επεξεργασία και ταξινόμηση εικόνας (Computer Vision). Ωστόσο, πρόσφατα τα μοντέλα CNN έχουν χρησιμοποιηθεί και για άλλους τύπους προβλημάτων όπως ταξινόμηση χρονοσειρών και παλινδρόμηση. Παράγουν καλά αποτελέσματα όταν τα δεδομένα εισόδου αντιμετωπίζονται με παρόμοιο τρόπο όπως με τα δεδομένα εικόνας. Δηλαδή, όπως οι ψηφιακές εικόνες αντιστοιχούν σε πίνακες με αριθμητικές τιμές και τροφοδοτούνται ως είσοδος στο CNN, έτσι και κάθε «παρόμοιος» πίνακας που περιέχει για παράδειγμα, τις διαδοχικές ακατέργαστες τιμές (raw samples values) ενός αρχείου ήχου, μπορεί να χρησιμοποιηθεί ως είσοδος στο CNN μοντέλο.

Γενικότερα, τα Συνελκτικά νευρωνικά δίκτυα έχουν βρει εφαρμογή μεταξύ άλλων στην αποθορυβοποίηση ήχου και ομιλίας (Audio & Speech Denoising), στην ταξινόμηση ήχων (Audio Classification) και στην ανάκτηση μουσικής πληροφορίας (Music Information Retrieval - MIR).

Στην αποθορυβοποίηση ήχου και ειδικότερα στη βελτίωση ομιλίας (Speech Enhancement - SE), πρωταρχικός στόχος είναι η βελτίωση του βαθμού κατανόησης και της ποιότητας των σημάτων ομιλίας, μειώνοντας τα στοιχεία θορύβου. Για το σκοπό αυτό τα τελευταία χρόνια, έχουν προταθεί μέθοδοι SE που βασίζονται στη βαθιά μάθηση, όπως η αποθορυβοποίηση χρησιμοποιώντας αυτοκωδικοποιητές (denoising autoencoders) [146, 147], καθώς και προσεγγίσεις που χρησιμοποιούν CNN μοντέλα [148, 149, 150]. Αυτά τα μοντέλα CNN λόγω του πλεονεκτήματός τους να χειρίζονται διδιάστατες αναπαραστάσεις χρόνου-συχνότητας ενός ήχου, για παράδειγμα μιας ομιλίας, παρουσιάζουν καλύτερες επιδόσεις από τα συμβατικά μοντέλα SE. Επιπλέον, πειραματικά αποτελέσματα έδειξαν ότι ένας συνδυασμός βαθέων TND και CNN, με είσοδο ακουστικά αλλά και οπτικά χαρακτηριστικά (audio-visual multimodality) παρουσίασαν βελτιωμένες επιδόσεις στην αποθορυβοποίηση ομιλίας [151, 152, 153]. Τέλος, οι Kuleshov et. al [154] πρότειναν μια νέα τεχνική επεξεργασίας ήχου, με την οποία αυξάνεται ο ρυθμός δειγματοληψίας των ηχητικών σημάτων, χρησιμοποιώντας βαθιά συνελκτικά νευρωνικά δίκτυα. Στην τεχνική αυτή το μοντέλο εκπαιδεύεται σε ζεύγη ακουστικών παραδειγμάτων, χαμηλής και υψηλής ποιότητας και προβλέπει δείγματα που «λείπουν» σε ένα σήμα χαμηλής ανάλυσης, χρησιμοποιώντας μια διαδικασία παρεμβολής.

Τα βαθιά συνελκτικά δίκτυα έχουν επίσης χρησιμοποιηθεί σε ένα μεγάλο εύρος εργασιών που αφορούν στην ταξινόμηση ήχου, όπως στην ταξινόμηση ειδών μουσικής (Music Genre Classification) [155, 156, 157] και στην ταξινόμηση ήχων περιβάλλοντος

(Environment Sound Classification) [158, 159, 160] Επιπλέον έχουν προταθεί διάφορες τεχνικές, όπου τα CNN τροφοδοτούνται είτε με ακατέργαστες τιμές πλάτους αρχείων ήχου [161, 162, 163], είτε με φασματογραφήματα (Mel-spectrograms & MFCCs) [164 - 168]. Σε πολλές προσεγγίσεις χρησιμοποιούνται σε συνδυασμό με RNN μοντέλα, ώστε να βελτιωθεί ή κατανοηθεί η χρονική εξάρτηση των ηχητικών δεδομένων [169, 170, 171]. Επίσης, όπως περιγράφεται στο [172], προβλήματα ταξινόμησης ήχου αντιμετωπίστηκαν με επιτυχία, τροφοδοτώντας εικόνες spectrograms σε μοντέλο με τυπική αρχιτεκτονική AlexNet [173], το οποίο ήταν προεκπαιδευμένο στη βάση δεδομένων εικόνων ImageNet [174]. Τέλος, οι Palanisamy et. al. [175], χρησιμοποιώντας επίσης προεκπαιδευμένα μοντέλα με τυπικές αρχιτεκτονικές δικτύων, όπως Inception [176], ResNet [177] και DenseNet [178], και ένα ενιαίο σύνολο χαρακτηριστικών εισόδου όπως τα Mel-spectrograms, κατάφεραν εξαιρετικά αποτελέσματα σε διάφορα γνωστά ηχητικά σύνολα δεδομένων, όπως είναι τα ESC-50 [179] και UrbanSound8k [180].

Πρόσφατες μελέτες στον τομέα της ανάκτησης μουσικής πληροφορίας (MIR) και ειδικότερα στην ταξινόμηση ειδών μουσικής (Music Genre Classification - MGC), επιβεβαιώνουν ότι τα CNN μοντέλα με είσοδο spectrograms, έχουν εφαρμοστεί με επιτυχία σε ένα πλήθος περιπτώσεων [181]. Σε κάποιες από αυτές τις εργασίες προτάθηκε ο συνδυασμός ακουστικών και πολυεπίπεδων οπτικών χαρακτηριστικών που αντικατοπτρίζουν τις υφές του φασματογράμματος και τις χρονικές τους μεταβολές [182, 183]. Επιπλέον, οι Sènac et. al [184], κατάφεραν αυξημένες επιδόσεις στην ταξινόμηση μουσικών ειδών, χρησιμοποιώντας, αντί για spectrograms, επιλεγμένα μουσικά χαρακτηριστικά ως στοιχεία εισόδου ενός CNN, το οποίο ήταν βασισμένο στην τοπολογία των Zhang & al. [185]. Τέλος, οι Adiyansjah et. al [186] με βάση τις μελέτες των Choi et al. [187, 188] για την ανάπτυξη συστήματος ταξινόμησης μουσικών ειδών, πρότειναν τη χρήση συνελκτικών επαναλαμβανόμενων νευρωνικών δικτύων (CRNN) τα οποία λαμβάνουν υπόψη τόσο τα συχνοτικά όσο και τα χρονικά χαρακτηριστικά του σήματος παρέχοντας έτσι συνολικά καλύτερη απόδοση.

Μέρος Β΄
Πειραματική διαδικασία

5

Υλοποίηση ευθέως προβλήματος

Το ευθύ πρόβλημα αναφέρεται στη δημιουργία δειγμάτων ήχου, χρησιμοποιώντας γνωστές φυσικές παραμέτρους (κατασκευαστικά και γεωμετρικά χαρακτηριστικά), για το μοντέλο της δονούμενης μεμβράνης. Τα ηχητικά δείγματα στη συνέχεια θα τροφοδοτήσουν το συνελκτικό μοντέλο για την επίλυση του αντίστροφου προβλήματος και την εκτίμηση των φυσικών παραμέτρων που χρησιμοποιήθηκαν για την παραγωγή τους. Η λύση του ευθέως προβλήματος βασίστηκε στην επίλυση της κυματικής εξίσωσης της μεμβράνης, με τη μέθοδο των Πεπερασμένων Διαφορών Χρονικού Τομέα (FDTD).

Σε αυτό το κεφάλαιο πραγματοποιείται μια επισκόπηση της προγραμματιστικής υλοποίησης της μεθόδου FDTD, για την παραγωγή των ηχητικών δειγμάτων, καθώς και ο τρόπος που έγινε η επιλογή των φυσικών παραμέτρων που στη συνέχεια θα εκτιμηθούν από το ΤΝΔ. Συγκεκριμένα, περιγράφεται ο αρχικός πειραματισμός και ο λόγος για τον οποίο πραγματοποιήθηκε ο επαναπροσδιορισμός του αντίστροφου προβλήματος και η εκ νέου διαμόρφωση των παραμέτρων για το ευθύ. Στη συνέχεια, αφού περιγραφούν συνοπτικά τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν, παρουσιάζεται αναλυτικά η διαδικασία παραγωγής των ηχητικών δειγμάτων και του συνόλου δεδομένων το οποίο θα χρησιμοποιηθεί για την επίλυση του αντίστροφου προβλήματος.

5.1 Αρχικός Πειραματισμός

Όπως περιγράφηκε στην ενότητα 1.1, οι αρχικοί στόχοι αυτής της εργασίας ήταν η δημιουργία δειγμάτων ήχου, χρησιμοποιώντας κάποιες γνωστές φυσικές παραμέτρους, για το μοντέλο της δονούμενης μεμβράνης (επίλυση ευθέως προβλήματος) και έπειτα η εκτίμηση αυτών των παραμέτρων με χρήση συνελκτικών δικτύων βαθιάς μάθησης (επίλυση αντίστροφου προβλήματος).

Για το σκοπό αυτό δημιουργήθηκε ένα σύνολο δεδομένων που αποτελείται από 87205 αρχεία ήχου. Ορισμένες από τις φυσικές παραμέτρους της δονούμενης μεμβράνης

επιλέχθηκαν ως μεταβλητές παράμετροι και κάποιες παρέμειναν σταθερές σε όλη την πειραματική διαδικασία, για όλα τα παραγόμενα ηχητικά δείγματα. Ως μεταβλητές παράμετροι επιλέχθηκαν οι τιμές του πάχους, της ακτίνας και της τάσης της μεμβράνης, οι οποίες και αποτέλεσαν τις παραμέτρους που επιχειρήθηκε να εκτιμηθούν μέσω ενός μοντέλου CNN. Οι τιμές των φυσικών παραμέτρων του μοντέλου μεμβράνης (σταθερές και μεταβλητές) που επιλέχθηκαν κατά την πειραματική διαδικασία παρουσιάζονται στον Πίνακα 5-1.

Με βάση αυτό το σύνολο δεδομένων, δημιουργήθηκαν διαφορετικές αναπαραστάσεις για τα δεδομένα εισόδου όπως ανεπεξέργαστες τιμές πλάτους των αρχείων ήχου, ηχητικά χαρακτηριστικά και εικόνες spectrograms. Πραγματοποιώντας αρκετούς πειραματισμούς, χρησιμοποιώντας αυτές τις αναπαραστάσεις ως δεδομένα εισόδου, τα αποτελέσματα που προέκυψαν ήταν παρόμοια. Οπότε επιλέχθηκε τελικά ως σύνολο εισόδου δεδομένων, το σύνολο των ανεπεξέργαστων τιμών πλάτους των αρχείων ήχου (raw sample values), λόγω του μικρότερου υπολογιστικού κόστους.

Παράλληλα με το σύνολο δεδομένων εισόδου, δημιουργήθηκε και το σύνολο δεδομένων εξόδου (στόχου), το οποίο αποτελείται από τις τιμές των παραμέτρων πάχους, ακτίνας και τάσης που χρησιμοποιήθηκαν για την παραγωγή του κάθε ηχητικού δείγματος.

Πίνακας 5-1: Φυσικές παράμετροι του μοντέλου της δονούμενης μεμβράνης.

Φυσικές παράμετροι με <u>σταθερές</u> τιμές	
Πυκνότητα μεμβράνης (density)	300 kgr/m ³
Σημείο διέγερσης (point of excitation)	κέντρο μεμβράνης
Απόσταση σημείου ακρόασης (ear-listening distance)	0.7 m
Σταθερά απόσβεσης (damping constant)	0.9999
Βισκοελαστική απόσβεση (viscoelasticity damping)	όχι
Ταχύτητα ήχου (sound of speed)	343 m/sec
Διάρκεια ηχητικού δείγματος (duration)	1.0 sec
Ρυθμός δειγματοληψίας (sampling rate)	96000 Hz

Φυσικές παράμετροι με <u>μεταβλητές</u> τιμές			
	από	έως	πλήθος μοναδικών τιμών
Πάχος (thickness) (m)	0.0005	0.023	194
Ακτίνα (radius) (m)	0.0401	0.5028	199
Τάση (tension) (Nt)	11.0	5971.0	200
Σύνολο συνδυασμών (πλήθος αρχείων ήχου): 87205			

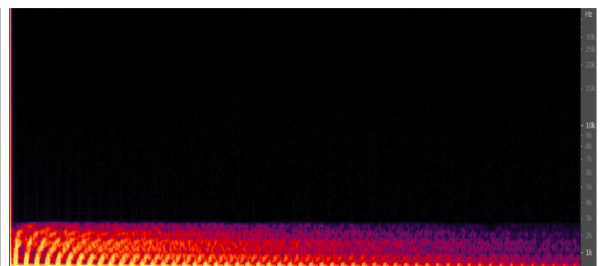
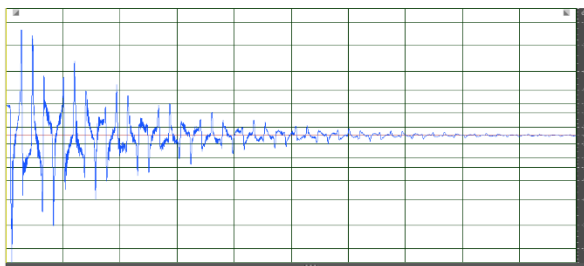
Στη συνέχεια, χρησιμοποιώντας τα παραπάνω σύνολα δεδομένων εισόδου και στόχου, αναπτυχθήκαν, εκπαιδεύτηκαν και αξιολογήθηκαν πολλές διαφορετικές αρχιτεκτονικές CNN μοντέλων, χρησιμοποιώντας διάφορες προσεγγίσεις και διερευνώντας πολλούς συνδυασμούς για τις τιμές των υπερπαραμέτρων.

Τα αποτελέσματα όμως αυτών των μοντέλων, σε κάθε περίπτωση, δεν ήταν ικανοποιητικά. Μετά από περαιτέρω ανάλυση, έγινε φανερός ο λόγος για τον οποίο δεν έγινε εφικτό να αναπτυχθεί ένα τέτοιο μοντέλο νευρωνικού δικτύου το οποίο να είναι σταθερά ικανό να εκτιμά τις συγκεκριμένες φυσικές παραμέτρους της μεμβράνης.

Ο λόγος είναι ότι συνδυασμοί διαφορετικών τιμών παραμέτρων πάχους, ακτίνας και τάσης, μπορούν να δημιουργήσουν σχεδόν τον ίδιο, πανομοιότυπο ήχο. Με άλλα λόγια, όταν διεγερθεί μια μεμβράνη με μεγάλο πάχος που είναι τεντωμένη με μεγαλύτερη τάση, μπορεί να παράγει ήχο, σε πολύ μεγάλο βαθμό ίδιο, με μια άλλη μεμβράνη με πιο μικρό πάχος και με μικρότερη τάση. Για παράδειγμα, τα παρακάτω αρχεία ήχου, οι κυματομορφές των οποίων διακρίνονται στην Εικόνα 5-1 (αριστερά), ακούγονται ακριβώς το ίδιο, παρότι το 1^ο ηχητικό δείγμα (πάνω) προέρχεται από μια μεμβράνη με πάχος 4,3 mm, ακτίνα 15,7 cm και τάση 382 Nt, ενώ το 2^ο δείγμα (κάτω) αντιστοιχεί σε μεμβράνη πάχους 13,1 mm, ακτίνας 5,68cm και τάσης 150 Nt. Στο ίδιο συμπέρασμα μπορεί κάποιος να καταλήξει μελετώντας το φάσμα συχνοτήτων στα spectrograms τα οποία εξήχθησαν από τα συγκεκριμένα ηχητικά δείγματα, που όπως φαίνεται στην Εικόνα 5-1 (δεξιά) είναι πανομοιότυπα.

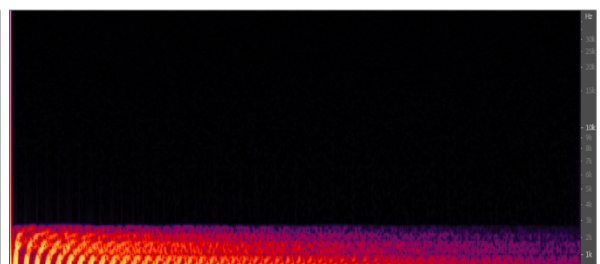
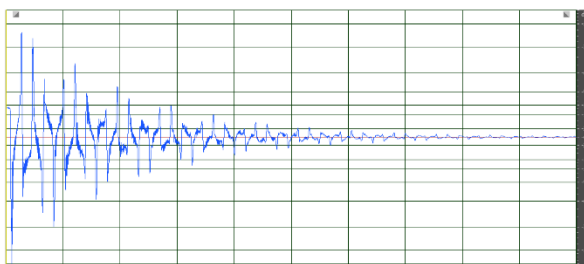
01-filename:

memb_H0.0043-R0.157-T382.0-D1.0-X52-Y52-A0.7.wav



02-filename:

memb_H0.0131-R0.0568-T150.0-D1.0-X52-Y52-A0.7.wav



Εικόνα 5-1: Κυματομορφές και spectrograms δυο δειγμάτων, που ενώ προέρχονται από μεμβράνες με διαφορετικό συνδυασμό φυσικών παραμέτρων, ακούγονται το ίδιο.

5.2 Επαναπροσδιορισμός στόχου

Το πρόβλημα που περιγράφηκε στην προηγούμενη ενότητα και ο λόγος για τον οποίο παρουσιάστηκε οδήγησαν στην αλλαγή του αρχικού στόχου της εργασίας, επαναπροσδιορίζοντας το αντίστροφο πρόβλημα και διαμορφώνοντας νέες παραμέτρους για το ευθύ.

Επομένως, για τον νέο στόχο, οι φυσικές παράμετροι του μοντέλου της δονούμενης μεμβράνης παρέμειναν σταθερές και το νέο αντίστροφο πρόβλημα μπορεί να περιγραφεί ως εξής:

*Ταξινόμηση των μοτίβων πάστας,
η οποία έχει εφαρμοστεί στην επιφάνεια της μεμβράνης,
με βάση τον ήχο που παράγεται από τη μεμβράνη, μέσω ΤΝΔ βαθιάς μάθησης.*

Όπως αναφέρθηκε στην ενότητα 2.8, συχνά οι μουσικοί εφαρμόζουν διάφορα υλικά, στην επιφάνεια των κεφαλών των τυμπάνων, προκειμένου να ρυθμίσουν το χρόνο απόσβεσης, την τονικότητα και γενικότερα για να τα κουρδίσουν. Τα υλικά αυτά μπορεί να είναι ειδικά μαξιλαράκια απόσβεσης, δακτύλιοι απόσβεσης, κομμάτια κολλητικής ταινίας, αποσβεστική πάστα, ακόμα και κομμάτια άλλων μεμβρανών, που προσαρμόζονται πάνω στην επιφάνεια της κεφαλής του τυμπάνου.

Σε αυτή τη μελέτη, προσομοιώνεται η εφαρμογή αποσβεστικού υλικού πάνω στη μεμβράνη (χωρίς ιξωδοελαστικές ιδιότητες), σε μορφή πάστας (tuning paste), με μια τοπική αύξηση της μάζας της μεμβράνης σε συγκεκριμένα σημεία εφαρμογής, με παρόμοιο τρόπο όπως περιγράφεται στα [1, 2]. Τα σημεία της μεμβράνης όπου εφαρμόστηκε η πάστα, επιλέχθηκαν έτσι ώστε να διαμορφώνουν διάφορα σχήματα (μοτίβα) στην επιφάνειά της. Η τοπική αύξηση της μάζας της μεμβράνης στα σημεία αυτά, προκαλεί και την αύξηση της επιφανειακής της πυκνότητας. Οπότε η αρχικά ομοιόμορφη επιφανειακή πυκνότητα της μεμβράνης, πλέον, αποκτά χωρική κατανομή (spatial distribution) ανάλογα με την ύπαρξη ή όχι πάστας σε κάποια από τα σημεία της επιφάνειάς της (υποενότητα 3.3.5).

Εκτός από το παραπάνω πρόβλημα ταξινόμησης, ορίστηκε ένας επιπλέον στόχος, ως πρόβλημα παλινδρόμησης, το οποίο μπορεί να περιγραφεί ως εξής:

*Εκτίμηση της πρόσθετης μάζας από την εφαρμογή πάστας,
με βάση τον ήχο που παράγεται από τη μεμβράνη,
μέσω ΤΝΔ βαθιάς μάθησης.*

Λόγω του επαναπροσδιορισμού του αντίστροφου προβλήματος και τον ορισμό νέων στόχων οι παράμετροι του ευθέως προβλήματος αναδιαμορφώθηκαν, κρατώντας τις φυσικές παραμέτρους του μοντέλου της μεμβράνης σταθερές και περιλαμβάνοντας αυτή τη φορά, την προσθήκη αποσβεστικού υλικού κατά τη δημιουργία των νέων ηχητικών δειγμάτων. Οι νέες μεταβλητές παράμετροι είναι το μοτίβο πάστας και η πρόσθετη μάζα λόγω της προσθήκης πάστας και είναι αυτές που θα εκτιμηθούν κατά την επίλυση του νέου αντίστροφου προβλήματος. Στην συνέχεια, στην ενότητα 5.3 παρουσιάζεται αναλυτικά όλη η παραπάνω διαδικασία.

5.2.1 Προγραμματιστικά εργαλεία για την παραγωγή των αρχείων ήχου

Για την παραγωγή των ηχητικών δειγμάτων και την επίλυση του ευθέως προβλήματος, χρησιμοποιήθηκε το υλοποιημένο σε γλώσσα C# πρόγραμμα που παραχωρήθηκε από τον Καθηγητή R. Bader. Το πρόγραμμα αυτό εφαρμόζει τη μέθοδο FDTD, μέσω ενός αλγορίθμου `leapfrog` και επιλύει την κυματική εξίσωση που περιγράφει την κίνηση της δονούμενης μεμβράνης, εξάγοντας τιμές μετατόπισης, οι οποίες μετέπειτα αθροίζονται για να παραχθούν τα ηχητικά δείγματα. Για την επίλυση της FDTD, αξιοποιείται η μονάδα επεξεργασίας γραφικών (GPU) του H/Y, με χρήση της αρχιτεκτονικής CUDA [123], μειώνοντας δραματικά το χρόνο που χρειάζεται για τη δημιουργία των αρχείων ήχου.

Εκτός από την υλοποίηση του R. Bader, για την παραγωγή ήχων, απαιτήθηκε ο προγραμματισμός κάποιων επιπρόσθετων λειτουργιών, για τις οποίες χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python v3.7.9 [189]. Οι λειτουργίες που υλοποιήθηκαν σε Python αφορούν: στον προσδιορισμό του συνδυασμού των τιμών για τις φυσικές παραμέτρους, στην κλήση του προγράμματος του R. Bader με τις συγκεκριμένες τιμές παραμέτρων, στην αποθήκευση σε αρχεία ήχου με την αντίστοιχη ονοματοδοσία, καθώς και στη δημιουργία και στην οπτικοποίηση της βάσης δεδομένων. Η υλοποίηση των Python scripts πραγματοποιήθηκε στο ολοκληρωμένο περιβάλλον ανάπτυξης Visual Studio Code v1.6 [190] αλλά και στο Visual Studio 2019 v16.7 [191], το οποίο χρησιμοποιήθηκε και για τις τροποποιήσεις του προγράμματος του R. Bader.

Το σύστημα H/Y που χρησιμοποιήθηκε για την παραγωγή των αρχείων ήχου περιλαμβάνει επεξεργαστή (CPU) Intel i7-4790k 4GHz, μνήμη (RAM) 32GB, μονάδα επεξεργασίας γραφικών (GPU) NVIDIA GeForce GTX 970 4GB (CUDA-enabled) και λειτουργικό Windows 7 64bit.

Τα ηχητικά δείγματα, αποθηκευτήκαν σε wav μορφή αρχείων (Waveform Audio File [155]), στα 16 bit PCM format, με 44.1KHz ρυθμό δειγματοληψίας. Η παραγωγή του κάθε αρχείου ήχου διήρκησε περίπου 3-4 δευτερόλεπτα στο παραπάνω σύστημα, με ενεργοποιημένη την CUDA αρχιτεκτονική.

Οι κυριότερες βιβλιοθήκες της Python που χρησιμοποιήθηκαν σε αυτό το σημείο της πειραματικής διαδικασίας είναι:

- Pandas [193]: για δημιουργία, χειρισμό και ανάλυση των βάσεων δεδομένων.
- Numpy [194]: για αριθμητικούς υπολογισμούς και για χρήση γραμμικής άλγεβρας.
- Scipy [197]: για υπολογισμό ολοκληρωμάτων, γραμμικής άλγεβρας, βελτιστοποίησης, παρεμβολής και στατιστικών, αλλά και για αποθήκευση των ήχων.
- Matplotlib [194] και Seaborn [196]: για δημιουργία γραφημάτων και για οπτικοποίηση των δεδομένων.
- Librosa [198], openSMILE [199] και Essentia [200]: για ανάλυση, εξαγωγή χαρακτηριστικών και επεξεργασία των ήχων.

5.2.2 Αρχιτεκτονική CUDA

Ο όρος CUDA (Compute Unified Device Architecture) αναφέρεται σε μια αρχιτεκτονική υλικού και λογισμικού που παρέχει στους προγραμματιστές πρόσβαση στα παράλληλα υπολογιστικά στοιχεία των μονάδων επεξεργασίας γραφικών (GPU - graphics processing unit) της NVIDIA [120]. Δημιουργήθηκε από τη NVIDIA το 2006 και έχει σχεδιαστεί για να λειτουργεί με γλώσσες υψηλού επιπέδου όπως C, Fortran, OpenCL και DirectCompute αλλά και C++ [123, 129].

Οι μονάδες GPU, ως εξειδικευμένοι επεξεργαστές, χρησιμοποιήθηκαν αρχικά για τις υπολογιστικές απαιτήσεις της απόδοσης των γραφικών σε πραγματικό χρόνο. Σταδιακά όμως εξελίχθηκαν σε ισχυρά συστήματα πολλαπλών πυρήνων που χρησιμοποιούν παράλληλη αρχιτεκτονική με δυνατότητα εκτέλεσης χιλιάδων νημάτων (threads) ταυτόχρονα. Ο εξειδικευμένος αυτός σχεδιασμός δίνει μεγάλο πλεονέκτημα στις μονάδες GPU έναντι της κεντρικής μονάδας επεξεργασίας γενικής χρήσης (CPU), ειδικά σε περιπτώσεις παράλληλης επεξεργασίας μεγάλου όγκου δεδομένων. Οι σύγχρονοι επεξεργαστές διαθέτουν αυτήν τη στιγμή δεκάδες υπολογιστικούς πυρήνες, ενώ μια GPU μπορεί να έχει χιλιάδες, όπως για παράδειγμα, η σειρά της NVIDIA Ampere A100 με πάνω από 8000 πυρήνες.

Επομένως, οι υπολογισμοί κατά την υλοποίηση της FDTD είναι πολύ πιο γρήγοροι όταν εκτελούνται σε GPU με δυνατότητα CUDA (CUDA-enabled NVIDIA GPUs). Υπάρχει πλούσια βιβλιογραφία [124, 202, 203, 204, 205], σχετικά με το θέμα της επιτάχυνσης της μεθόδου FDTD χρησιμοποιώντας την αρχιτεκτονική CUDA, η οποία αν και περίπλοκη στην εφαρμογή της, επιτυγχάνει πολύ σημαντικά οφέλη.

Τέλος, και τα νευρωνικά δίκτυα βαθιάς μάθησης, τα οποία αποτελούνται από ένα μεγάλο αριθμό όμοιων νευρώνων που εκτελούν από τη φύση τους παράλληλα πράξεις, ωφελούνται από τη χρήση GPU με δυνατότητα CUDA, αυξάνοντας σημαντικά την ταχύτητα των υπολογισμών.

5.3 Παραγωγή ηχητικών δειγμάτων

Όπως αναφέρθηκε στην υποενότητα 5.2.1, για την παραγωγή των ηχητικών δειγμάτων, χρησιμοποιήθηκε το πρόγραμμα που παραχωρήθηκε από τον Καθηγητή R. Bader. Με βάση την υλοποίηση αυτή, η δονούμενη μεμβράνη μοντελοποιείται βάσει της εξίσωσης (3.3.5):

$$\frac{T(x,y)}{\mu(x,y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} \quad (3.3.5)$$

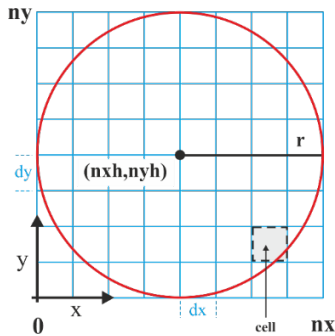
όπου: $u(x,y,t)$: μετατόπιση (m), $T(x,y)$: τάση (προένταση) (Nt), $\mu(x,y)$: επιφανειακή πυκνότητα (kg/m^2) και D : σταθερά απόσβεσης

και η μέθοδος FDTD, υλοποιήθηκε εφαρμόζοντας τα επαναληπτικά βήματα του αλγόριθμου NSV-leapfrog (Εικόνα 5-2), που περιγράφηκαν στη ενότητα 3.3.4.

1. $a = \frac{T}{\mu} (u_{xx} + u_{yy})$
2. $v_{new} = v_{new} \cdot D$
3. $v_{new} = v_{old} + a \Delta t$
4. $u_{new} = u_{old} + v_{new} \Delta t$

Εικόνα 5-2: Τα επαναληπτικά βήματα του NSV-leapfrog αλγόριθμου.

Για τη διακριτοποίηση της κυματικής εξίσωσης χρησιμοποιήθηκε ένα ομοιόμορφο, τετραγωνικό πλέγμα $104 \times 104 = 10816$ κομβικών σημείων, από το οποίο χρησιμοποιήθηκαν μόνο τα σημεία πλέγματος που αντιστοιχούν στην κυκλική περιοχή της μεμβράνης (Εικόνα 5-3).



Εικόνα 5-3: Ομοιόμορφο, τετραγωνικό πλέγμα $n_x \cdot n_y = 104 \times 104 = 10816$ κελιά.

Ο ρυθμός δειγματοληψίας ορίστηκε στα 96 kHz και οι υπολογισμοί για κάθε παραγόμενο ηχητικό δείγμα, πραγματοποιήθηκαν για χρόνο 1 sec, καθώς κρίθηκε ότι είναι ικανός χρόνος για την πλήρη απόσβεση του ηχητικού σήματος.

Για την υλοποίηση των βημάτων του αλγόριθμου NSV-leapfrog, όπως περιγράφεται στα [1, 2], το χωρικό πλέγμα εφαρμόστηκε στη GPU χρησιμοποιώντας τρία διανύσματα μήκους $104 \times 104 = 10816$ καταχωρήσεων, ένα για τη μετατόπιση (displacement), ένα για τη ταχύτητα (velocity) και ένα για την αποθήκευση μετατόπισης (displacement storage), ενώ η μνήμη επιτάχυνσης (acceleration memory) υλοποιήθηκε ως ένα διάνυσμα $104 \times 104 \times 1000 = 10816000$ καταχωρήσεων. Όλοι οι υπολογισμοί για την επιτάχυνση και την ταχύτητα για κάθε ένα χρονικό σημείο πραγματοποιήθηκαν παράλληλα (με χρήση της αρχιτεκτονικής CUDA) και για τα 10816000 κομβικά σημεία και οι νέες επιταχύνσεις προστέθηκαν στο τέλος ενός κυκλικού διανύσματος μνήμης επιτάχυνσης (circular acceleration memory vector), μετατοπίζοντας το δείκτη του κατά μία διανυσματική καταχώριση (vector entry). Αφού εκτελεστεί ένα βήμα, οι νέες υπολογισμένες μετατοπίσεις μεταφέρονται στο διάνυσμα αποθήκευσης μετατόπισης και έπειτα μέσω αυτού υπολογίζονται οι νέες επιταχύνσεις στο επόμενο χρονικό βήμα κ.ο.κ., μέχρι να ολοκληρωθεί ο υπολογιστικός χρόνος που ορίστηκε. Η υλοποίηση αυτή συμπεριλαμβάνει και την ιξωδοελαστική απόσβεση, αλλά όπως έχει ήδη αναφερθεί για τους σκοπούς αυτής της εργασίας δε χρησιμοποιήθηκε απόσβεση με ιξωδοελαστικές ιδιότητες.

Το πρόγραμμα του Καθηγητή R. Bader τροποποιήθηκε ώστε κατά την παραγωγή των αρχείων ήχου, να συμπεριλαμβάνεται η περίπτωση προσθήκης αποσβεστικού υλικού (σε μορφή πάστας) στην επιφάνεια της μεμβράνης. Η προσθήκη πάστας σε επιλεγμένα σημεία της μεμβράνης υλοποιήθηκε με τοπική αύξηση της μάζας της, σε αυτά τα σημεία, όπως περιγράφηκε στην ενότητα 5.2. Επίσης ορίστηκαν οι σταθερές τιμές των φυσικών παραμέτρων του μοντέλου της μεμβράνης και τέλος, το πρόγραμμα τροποποιήθηκε ώστε να δέχεται ως ορίσματα τις μεταβλητές τιμές των παραμέτρων για τις περιπτώσεις των μοτίβων πάστας.

Στην επόμενη ενότητα, περιγράφεται η επιλογή των τιμών των φυσικών παραμέτρων της μεμβράνης καθώς και οι παράμετροι και οι περιπτώσεις εφαρμογής πάστας που χρησιμοποιήθηκαν για την παραγωγή των ηχητικών δειγμάτων.

5.4 Φυσικές παράμετροι & παράμετροι μοτίβων πάστας

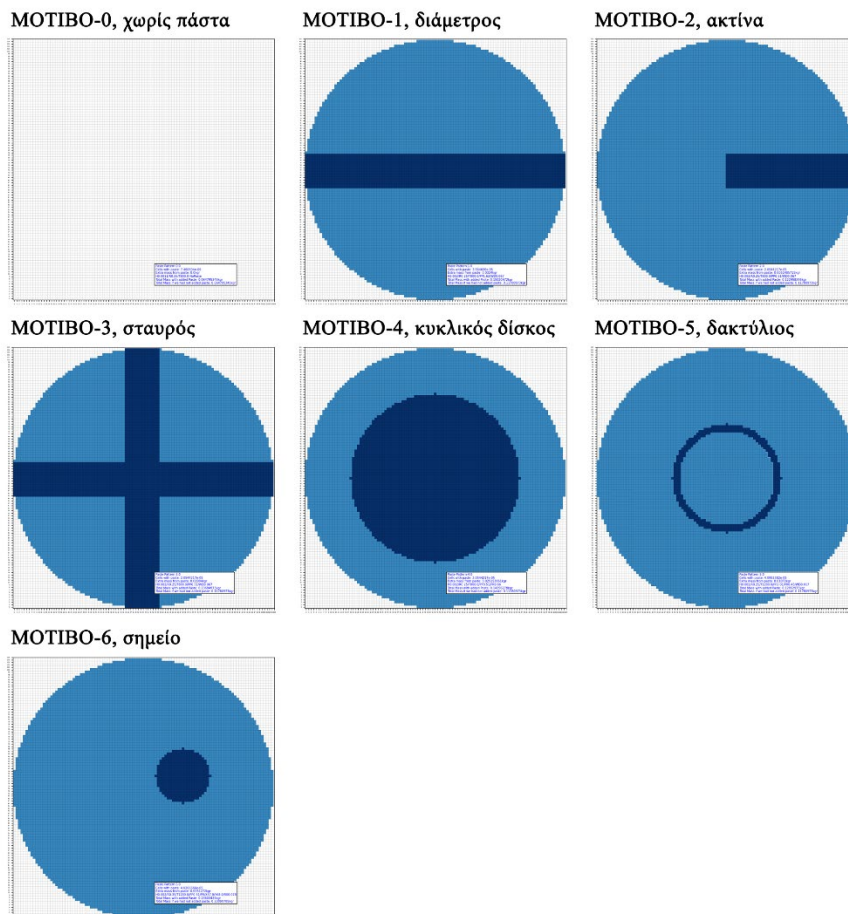
Κατά τη διαδικασία παραγωγής των ηχητικών δειγμάτων, μέσω του προγράμματος του R. Bader, οι τιμές των φυσικών παραμέτρων του μοντέλου της δονούμενης μεμβράνης παρέμειναν σταθερές, όπως παρουσιάζονται στον Πίνακα 5-2, ενώ ορίστηκαν και οι 7 περιπτώσεις μοτίβων πάστας, όπως φαίνεται στην Εικόνα 5-4.

Πίνακας 5-2: Οι σταθερές φυσικές παράμετροι του μοντέλου δονούμενης μεμβράνης, για τη δημιουργία των ηχητικών δειγμάτων.

Φυσικές παράμετροι (σταθερές τιμές)	
Πάχος (thickness)	0.003 m
Ακτίνα (radius)	0.250 m
Τάση (tension)	800.0 Nt
Πυκνότητα (volume density)	300 kg/m ³
Σημείο διέγερσης (point of excitation)	κέντρο μεμβράνης
Απόσταση καταγραφής ήχου (ear-listening distance)	0.7 m
Σταθερά απόσβεσης (damping constant)	0.9999
Βισκοελαστική απόσβεση (viscoelasticity damping)	όχι
Ταχύτητα ήχου (speed of sound)	343 m/sec
Διάρκεια ηχητικού δείγματος (duration)	1.0 sec
Ρυθμός δειγματοληψίας (sampling rate)	96000 Hz

Όπως φαίνεται στον Πίνακα 5-2, χρησιμοποιήθηκε μια μεμβράνη με πάχος 0.0003m, ακτίνα 0.25m και τάση 800Nt. Οι τιμές αυτές επιλέχθηκαν ως τυπικές τιμές για μια εμπορική κεφαλή τυμπάνου (ενότητα 2.7). Επίσης, καθώς οι τυπικές κεφαλές έχουν πυκνότητα μεταξύ 200 έως 340 kg/m³, για το συγκεκριμένο μοντέλο ορίστηκε πυκνότητα 300 kg/m³. Επιπλέον, επιλέχθηκε το κέντρο της μεμβράνης ως σημείο διέγερσης και το σημείο ακρόασης, σε απόσταση 0.7 m από το κέντρο της (παραπλήσιες

αποστάσεις δεν επέφεραν σημαντικές μεταβολές στον ήχο). Η σταθερά απόσβεσης ορίστηκε στο 0.9999, τιμή η οποία κρίθηκε ικανοποιητική για σύντομη και ομαλή απόσβεση του ήχου, ο οποίος είχε διάρκεια 1 sec και ρυθμό δειγματοληψίας 96 kHz. Τέλος, ορίστηκε 343 m/sec η ταχύτητα του ήχου και όπως ειπώθηκε δε χρησιμοποιήθηκε βισκοελαστική απόσβεση.



Εικόνα 5-4: Οι 7 περιπτώσεις σχημάτων εφαρμογής αποσβεστικού υλικού (7 μοτίβα πάστας).

Οι 7 περιπτώσεις εφαρμογής αποσβεστικού υλικού, στο εξής μοτίβα πάστας, που διακρίνονται στην Εικόνα 5-4, είναι:

- Μοτίβο-0, Χωρίς πάστα (No paste): Χωρίς προσθήκη αποσβεστικού υλικού (πάστας).
- Μοτίβο-1, Διάμετρος (Diameter): Πάστα τοποθετημένη σε μια διάμετρο της μεμβράνης, σαν μια ταινία απόσβεσης, σε διάφορες διαστάσεις (πλάτη).
- Μοτίβο-2, Ακτίνα (Radius): Όπως στο Μοτίβο-1, αλλά τοποθετημένη σε μια ακτίνα.
- Μοτίβο-3, Σταυρός (Cross): Όπως στο Μοτίβο-1, αλλά τοποθετημένη σε σχήμα σταυρού (σε 2 κάθετες διαμέτρους).

- Μοτίβο-4, Κυκλικός δίσκος (Circular): Πάστα τοποθετημένη σε κυκλική επιφάνεια με ίδιο κέντρο με το κέντρο της μεμβράνης.
- Μοτίβο-5, Δακτύλιος (Ring): Πάστα σε σχήμα δακτυλίου, με το ίδιο κέντρο με το κέντρο της μεμβράνης, σε διάφορα μεγέθη και πλάτη.
- Μοτίβο-6, Περιοχή σημείου (Point Area): Πάστα τοποθετημένη σε μια μικρή κυκλική περιοχή γύρω από ένα συγκεκριμένο σημείο της μεμβράνης.

Τα παραπάνω μοτίβα επιλέχθηκαν με βάση τους πιο δημοφιλείς τρόπους εφαρμογής υλικού απόσβεσης, όπως περιγράφηκαν στην υποενότητα 2.8.1, δηλαδή κυκλικοί δίσκοι (Tablstone dots, patches), δακτύλιοι ελέγχου (Muffle Rings, O-Rings) και ταινία Gaffer, Τζελ (Control Gels, Moongels).

Για κάθε περίπτωση μοτίβου πάστας, ορίστηκαν οι αντίστοιχες παράμετροι, οι οποίες διαμορφώνουν το σχήμα αλλά και την ποσότητα πάστας που εφαρμόζεται στη μεμβράνη. Συγκεκριμένα:

- Για τις περιπτώσεις 1, 2, και 3, δηλαδή για τα ορθογώνια μοτίβα (Diameter, Radius, Cross), χρησιμοποιήθηκαν οι ακόλουθες παράμετροι:
 - Ποσοστό πάστας (Paste Percentage, %): Η επιφανειακή πυκνότητα⁹ της μεμβράνης στο σημείο που έχει εφαρμοστεί το αποσβεστικό υλικό, ως ποσοστό της αρχικής επιφανειακής πυκνότητας της μεμβράνης σε αυτό το σημείο.
 - Πλάτος πάστας (Paste Width, m): Πλάτος πάστας σε μέτρα (σαν να ήταν ταινία).
- Για το Μοτίβο-4, δηλαδή για τον κυκλικό δίσκο (Circular):
 - Ποσοστό πάστας (Paste Percentage, %): Όπως περιγράφηκε παραπάνω.
 - Ακτίνα πάστας (Radius Circle Percentage, %): Η ακτίνα του κυκλικού δίσκου ως ποσοστό της ακτίνας της μεμβράνης.
- Για το Μοτίβο-5, δηλαδή για το δακτύλιο (Ring):
 - Ποσοστό πάστας (Paste Percentage, %): Όπως περιγράφηκε παραπάνω.
 - Ακτίνα μεγάλη πάστας (Radius Ring Percentage, %): Η μεγάλη ακτίνα του δακτυλίου, ως ποσοστό της ακτίνας της μεμβράνης.
 - Πλάτος πάστας (Paste Width, m): Πλάτος πάστας σε μέτρα (σαν να ήταν ταινία).

⁹ Επιφανειακή πυκνότητα (kg/m^2) = πυκνότητα όγκου (kg/m^3) x πάχος (m).

- Για την Μοτίβο-6, δηλαδή για την περιοχή γύρω από ένα σημείο (Point Area):
 - Ποσοστό πάστας (Paste Percentage, %): Όπως περιγράφηκε παραπάνω.
 - Κέντρο σημείου (Point - x_0, y_0): Το κέντρο της μικρής κυκλικής περιοχής.
 - Πλάτος/ακτίνα πάστας (Paste radius/width, m): Πλάτος/ακτίνα της παραπάνω κυκλικής περιοχής σε μέτρα.

Για την περίπτωση χωρίς πάστα (Μοτίβο-0), χρησιμοποιήθηκε μεταβλητό πάχος (thickness) μεμβράνης, ώστε να δημιουργηθούν διάφορες παραλλαγές μεμβρανών χωρίς προσθήκη αποσβεστικού υλικού. Αυτό έγινε προκειμένου αυτά τα ηχητικά δείγματα να λειτουργήσουν ως μια μορφή augmentation (υποενότητα 4.4.7) και το μοντέλο να μπορεί να διακρίνει αν έχει γίνει ή όχι προσθήκη πάστας. Τέλος, το ποσοστό πάστας (Paste Percentage) χρησιμοποιήθηκε για όλες τις περιπτώσεις. Στον Πίνακα 5-3, παρουσιάζονται συνοπτικά ποιες παράμετροι ορίστηκαν για το κάθε μοτίβο.

Πίνακας 5-3: Παράμετροι που χρησιμοποιήθηκαν για το κάθε μοτίβο πάστας.

Παράμετροι μοτίβων πάστας (Paste Patterns Parameters)	0: No paste	1: Diameter	2: Radius	3: Cross	4: Circular	5: Ring	6: Point
Ποσοστό πάστας (Paste Percentage)	x	✓	✓	✓	✓	✓	✓
Πλάτος πάστας (Paste Width)	x	✓	✓	✓	x	x	x
Ακτίνα πάστας (Radius Circle Percentage)	x	x	x	x	✓	x	x
Ακτίνα μεγάλη πάστας (Radius Ring Percentage)	x	x	x	x	x	✓	x
Πλάτος πάστας (Paste Width)	x	x	x	x	x	✓	x
Κέντρο σημείου (Point)	x	x	x	x	x	x	✓
Πλάτος/ακτίνα πάστας (Paste radius/width)	x	x	x	x	x	x	✓

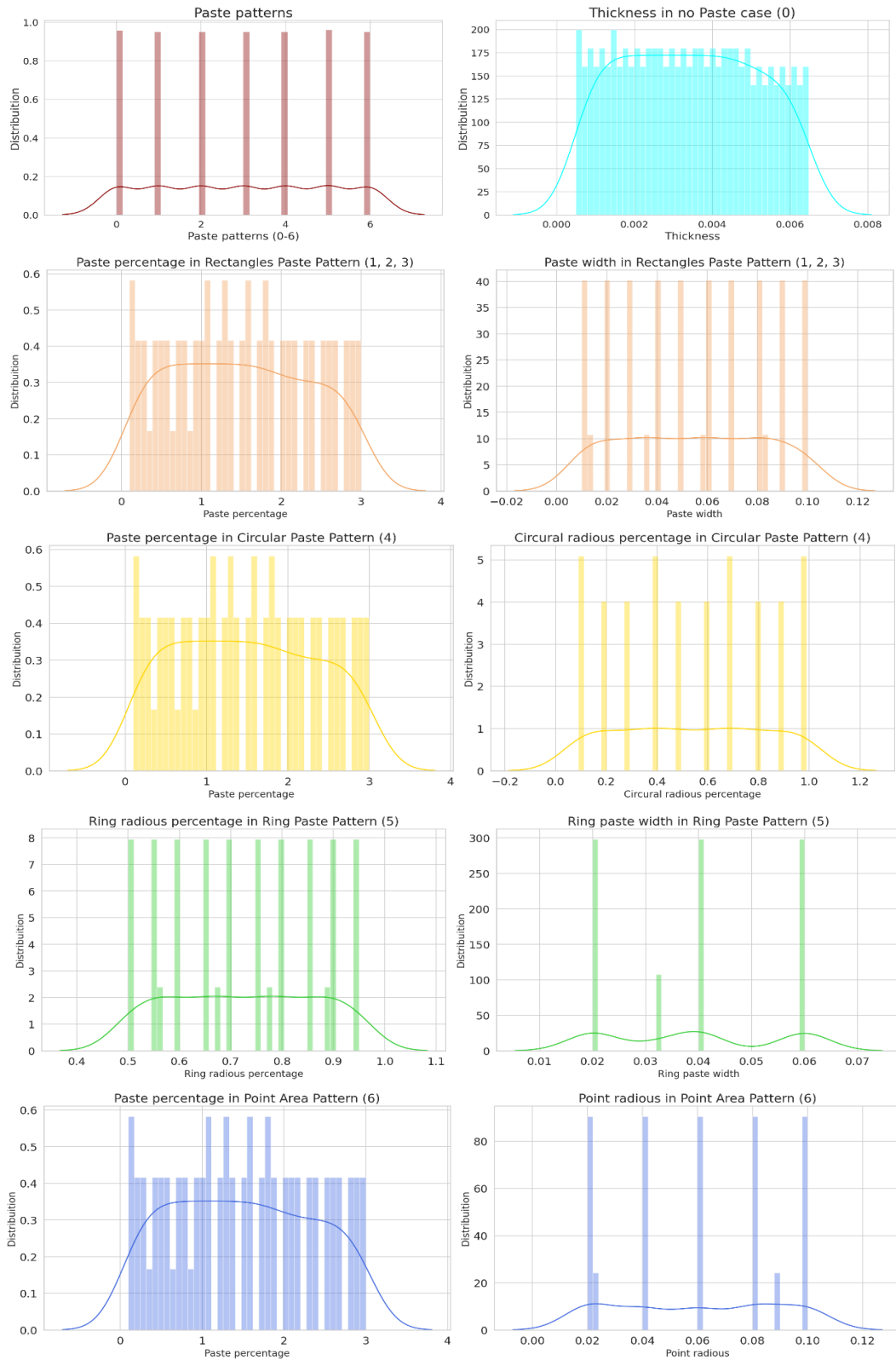
Οι τιμές των παραπάνω παραμέτρων, για τα μοτίβα πάστας, οι οποίες χρησιμοποιήθηκαν στην παραγωγή των νέων ηχητικών δειγμάτων, παρουσιάζονται στον Πίνακα 5-4. Οι διάφοροι συνδυασμοί των τιμών αυτών επιλέχθηκαν με τέτοιο τρόπο, ώστε αφενός να αντιστοιχούν στην πραγματικότητα, δηλαδή στον τρόπο με τον οποίο οι μουσικοί εφαρμόζουν αποσβεστικό υλικό στις κεφαλές τυμπάνων και αφετέρου, ώστε,

να λαμβάνεται ένα αντιπροσωπευτικό και ομοιογενές σύνολο τιμών για κάθε περίπτωση. Δημιουργώντας έτσι ένα σύνολο με όσο το δυνατόν κανονικές κατανομές (Εικόνα 5-5), καθώς αυτό αποτελεί σημαντική προϋπόθεση για την αποτελεσματική εκπαίδευση ενός μοντέλου ΤΝΔ.

Πίνακας 5-4: Τα σύνολα τιμών των παραμέτρων για τα μοτίβα πάστας, που χρησιμοποιήθηκαν στην παραγωγή των ηχητικών δειγμάτων.

Τιμές παραμέτρων για τα μοτίβα πάστας				
ΜΟΤΙΒΟ	ΤΙΜΕΣ			
0 - Χωρίς Πάστα	Πλάτος μεμβράνης (Thickness)			
	από 0.0005 m	έως 0.00648 m		
	Σύνολο τιμών: 335			
1 - Διάμετρος 2 - Ακτίνα 3 - Σταυρός	Ποσοστό πάστας (Paste Percentage) (%)		Πλάτος πάστας (m)	
	από 0.100 m	έως 3.000 m	από 0.01 m	έως 0.1 m
	Πλήθος τιμών: 332*3 = 996			
4 - Κυκλικός Δίσκος	Ποσοστό πάστας (Paste Percentage) (%)		Ακτίνα πάστας (Radius Circle Percentage) (%)	
	από 0.100	έως 3.000	από 0.090	έως 0.990
	Πλήθος τιμών: 332*3 = 996			
5 - Δακτύλιος	Ποσοστό πάστας (Paste Percentage) (%)		Ακτίνα μεγάλη πάστας (Radius Ring Percentage) (%)	
	από 0.100	έως 2.950	από 0.500	έως 0.950
	Πλάτος πάστας (Ring) (m)			
	από 0.020 m	έως 0.060 m		
Πλήθος τιμών: 336				
6 - Σημείο	Ποσοστό πάστας (Paste Percentage) (%)		Θέση σημείου (x₀, y₀) στο 104x104 πλέγμα	
	από 0.100	έως 3.000	(65,75)	(75,85)
	Πλάτος/ακτίνα πάστας (Paste radius/width) (m)		(63,79)	(70,83)
	από 0.020 m	έως 0.100 m		
Πλήθος τιμών: 332				
Συνολικοί συνδυασμοί τιμών: 2331				

Επομένως όπως φαίνεται και στον παραπάνω πίνακα, χρησιμοποιήθηκαν περίπου 330 μοναδικοί συνδυασμοί τιμών για κάθε μια από τις 7 περιπτώσεις των μοτίβων πάστας, δηλαδή 2331 συνδυασμοί συνολικά.



Εικόνα 5-5: Κατανομές τιμών κάθε παραμέτρου για τις περιπτώσεις των μοτίβων πάστας.

5.5 Ανάπτυξη συνόλου δεδομένων (dataset)

Χρησιμοποιώντας το πρόγραμμα του Καθηγητή R. Bader και τις τιμές των παραμέτρων που περιγράφηκαν στην προηγούμενη ενότητα, παράχθηκαν 2331 αρχεία ήχου.

Η παραγωγή του κάθε αρχείου ήχου διήρκεσε περίπου 4 δευτερόλεπτα, ενώ για το σύνολο των αρχείων χρειάστηκαν περίπου 3 ώρες, στο σύστημα Η/Υ που αναφέρθηκε στην υποενότητα 5.2.1, με ενεργοποιημένη την CUDA αρχιτεκτονική.

Μετά την παραγωγή των ηχητικών δειγμάτων και την αποθήκευσή τους σε αρχεία ήχου (waveform audio file, με συχνότητα δειγματοληψίας 44.1 KHz), δημιουργήθηκε μια βάση δεδομένων (dataset) η οποία περιλαμβάνει τις ανεπεξέργαστες τιμές πλάτους των αρχείων ήχου (raw sample values), καθώς και τις τιμές των φυσικών παραμέτρων της μεμβράνης και των παραμέτρων των μοτίβων πάστας, που χρησιμοποιήθηκαν για την παραγωγή τους. Στους Πίνακες 5-5 και 5-6, παρουσιάζεται ένα τμήμα της βάσης δεδομένων που δημιουργήθηκε.

Πίνακας 5-5: Τιμές πλάτους αρχείων ήχου και τιμές φυσικών παραμέτρων μεμβράνης.

index	όνομα αρχείου ήχου	raw samples values					φυσικές παράμετροι μεμβράνης (σταθερές)						
		0	1	...	44098	44099	Πάχος	Ακτίνα	Τάση	Διάρκεια	Σημείο κρούσης	Απόσταση ακρόασης	
101	m_21-06-27-042728_1_H0.00644-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT0-PP0.5-WD0.02-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0.0653	0.1170	...	0.0011	0.0012	0.006	0.250	800.0	1.000	52,52	0.70	
501	m_21-06-27-043047_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.2-WD0.08-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0.0663	0.1189	...	0.0027	0.0029	0.003	0.250	800.0	1.000	52,52	0.70	
901	m_21-06-27-054428_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT2-PP1.7-WD0.04-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0.0623	0.1117	...	-0.0021	-0.0023	0.003	0.250	800.0	1.000	52,52	0.70	
1201	m_21-06-27-061349_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT3-PP0.5-WD0.07-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0.0670	0.1202	...	-0.0029	-0.0031	0.003	0.250	800.0	1.000	52,52	0.70	
1501	m_21-06-27-065734_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT4-PP0.3-WD0.02-RP0.79-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0.0544	0.0975	...	-0.0020	-0.0021	0.003	0.250	800.0	1.000	52,52	0.70	
1701	m_21-06-27-082057_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT5-PP2.8-WD0.02-RP0.5-RR0.5-WR0.04-PX65-PY75-PR0.02-NB1.wav	0.0670	0.1202	...	-0.0029	-0.0031	0.003	0.250	800.0	1.000	52,52	0.70	
2001	m_21-06-27-090626_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT6-PP2.7-WD0.02-RP0.5-RR0.8-WR0.02-PX75-PY85-PR0.06-NB1.wav	0.0544	0.0975	...	-0.0020	-0.0021	0.003	0.250	800.0	1.000	52,52	0.70	

Πίνακας 5-6: Τιμές παραμέτρων μοτίβων πάστας.

index	όνομα αρχείου ήχου	παραμέτροι μοτίβων πάστας								
		paste pattern	paste perc	paste width	circural radius perc	ring radius perc	ring paste width	X0	Y0	point radius
101	m_21-06-27-042728_1_H0.00644-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT0-PP0.5-WD0.02-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	0	0.500	x	x	x	x	x	x	x
501	m_21-06-27-043047_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.2-WD0.08-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	1	0.200	0.080	x	x	x	x	x	x
901	m_21-06-27-054428_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT2-PP1.7-WD0.04-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	2	1.700	0.040	x	x	x	x	x	x
1201	m_21-06-27-061349_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT3-PP0.5-WD0.07-RP0.5-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	3	0.500	0.070	x	x	x	x	x	x
1501	m_21-06-27-065734_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT4-PP0.3-WD0.02-RP0.79-RR0.8-WR0.02-PX65-PY75-PR0.02-NB1.wav	4	0.300	x	0.790	x	x	x	x	x
1701	m_21-06-27-082057_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT5-PP2.8-WD0.02-RP0.5-RR0.5-WR0.04-PX65-PY75-PR0.02-NB1.wav	5	2.800	x	x	0.500	0.040	x	x	x
2001	m_21-06-27-090626_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT6-PP2.7-WD0.02-RP0.5-RR0.8-WR0.02-PX75-PY85-PR0.06-NB1.wav	6	2.700	x	x	x	x	75.000	85.000	0.060

Στο επόμενο κεφάλαιο γίνεται μια αναλυτική περιγραφή της μεθοδολογίας που ακολουθήθηκε για την επίλυση του αντίστροφου προβλήματος. Θα περιγραφεί αρχικά η διαδικασία ανάπτυξης, εκπαίδευσης και επικύρωσης του συνελκτικού μοντέλου, το οποίο θα εκτιμήσει το μοτίβο πάστας και την πρόσθετη μάζα, λόγω της προσθήκης πάστας, και έπειτα θα παρουσιαστούν τα τελικά αποτελέσματα.

6

Υλοποίηση αντίστροφου προβλήματος

Το αντίστροφο πρόβλημα, αναφέρεται στην ταξινόμηση των μοτίβων και στην εκτίμηση της πρόσθετης μάζας από την εφαρμογή πάστας στην επιφάνεια της μεμβράνης, με βάση τον ήχο που παράγεται μέσω ΤΝΔ βαθιάς μάθησης. Σε αυτό το κεφάλαιο θα περιγραφεί αναλυτικά η υλοποίηση του συνελκτικού μοντέλου που αναπτύχθηκε για την επίλυση του αντίστροφου προβλήματος και έπειτα θα παρουσιαστούν τα τελικά αποτελέσματα. Πριν την περιγραφή της μεθοδολογίας, κρίνεται σκόπιμο να γίνει αναφορά στα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν για το σκοπό αυτό, καθώς και στους λόγους για τους οποίους επιλέχθηκαν.

6.1 Προγραμματιστικά εργαλεία για την ανάπτυξη των CNN μοντέλων

Τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν είναι κυρίως βιβλιοθήκες λογισμικού που παρέχουν συναρτήσεις οι οποίες υλοποιούν τους αλγορίθμους που συζητήθηκαν στα προηγούμενα κεφάλαια, αλλά και ολοκληρωμένα περιβάλλοντα στα οποία μπορεί να πειραματιστεί κάποιος χωρίς την ανάγκη εγκατάστασης δεκάδων προγραμμάτων και τη χρήση εξοπλισμού υψηλού κόστους. Για την υλοποίηση του συνελκτικού μοντέλου χρησιμοποιήθηκαν η πλατφόρμα TensorFlow [213] και η βιβλιοθήκη Keras [214].

Το TensorFlow είναι μια ολοκληρωμένη (end-to-end) πλατφόρμα ανοιχτού κώδικα για μηχανική μάθηση, η οποία παρέχει ένα ευέλικτο οικοσύστημα εργαλείων, βιβλιοθηκών και άλλων πόρων, χρησιμοποιώντας διεπαφή (API - Automatic Programming Interface) υψηλού επιπέδου. Χρησιμοποιεί γράφους ροής δεδομένων (dataflow graphs) για να αναπαραστήσει αριθμητικούς υπολογισμούς, οι κόμβοι των οποίων αναπαριστούν μαθηματικούς υπολογισμούς, ενώ οι ακμές τους πολυδιάστατους πίνακες δεδομένων (tensors). Ο πυρήνας της βιβλιοθήκης (back end) είναι γραμμένος σε C++, ενώ για το

χειρισμό της παρέχεται API σε Python και άλλες γλώσσες προγραμματισμού (front end). Επίσης, υποστηρίζει την υλοποίηση υπολογισμών σε πολυπύρηνους επεξεργαστές (multicore CPU), σε κάρτες γραφικών (GPU) και σε ειδικά σχεδιασμένα ολοκληρωμένα κυκλώματα (Tensor Processing Units - TPU). Γενικά, το TensorFlow υλοποιεί πληθώρα αλγορίθμων μηχανικής μάθησης για την επίλυση προβλημάτων κάθε τομέα, από αναγνώριση εικόνων μέχρι επεξεργασία φυσικής γλώσσας και επικεντρώνεται στα βαθιά ΤΝΔ, όπου προσφέρει πολλές λύσεις, με ιδιαίτερα υψηλή αποδοτικότητα και απλοποιημένες διαδικασίες. Το Tensorflow επιτρέπει την εισαγωγή διαφόρων βιβλιοθηκών υψηλού επιπέδου, όπως τη βιβλιοθήκη Keras.

Το Keras, είναι μια βιβλιοθήκη νευρωνικών δικτύων υψηλού επιπέδου που εκτελείται «πάνω» στο TensorFlow. Η χρήση του Keras στη βαθιά μηχανική μάθηση επιτρέπει την εύκολη και γρήγορη δημιουργία πρωτοτύπων, καθώς και την απρόσκοπτη λειτουργία σε CPU και GPU. Υποστηρίζει τη δημιουργία περίπλοκων μοντέλων για την κατασκευή συνελκτικών και ανατροφοδοτούμενων νευρωνικών δικτύων. Βασίζεται στη δημιουργία ενός σειριακού (Sequential) μοντέλου, δηλαδή μιας σειριακής δομής επιπέδων, τα οποία μπορεί να είναι συνελκτικά επίπεδα, συναρτήσεις ενεργοποίησης και κόστους, καθώς και τεχνικές βελτιστοποίησης και dropout συναρτήσεις. Μετά τον ορισμό των επιπέδων και τον τρόπο εκτέλεσής τους, ορίζεται το μέγεθος των πακέτων δεδομένων που θα εκπαιδεύονται σε κάθε κύκλο εκπαίδευσης (batch size), καθώς και το πλήθος των κύκλων εκπαίδευσης (epochs). Το προγραμματιστικό πλαίσιο του Keras είναι γραμμένο σε κώδικα Python, ο οποίος παρέχει ευκολία στον εντοπισμό σφαλμάτων και στην επεκτασιμότητα. Τα δύο συστήματα TensorFlow και Keras αλληλοσυμπληρώνονται, δίνοντας έτσι μεγάλη ευελιξία στο επίπεδο προσαρμογής που μπορεί να έχει ένας χρήστης. Το γεγονός αυτό αποτελεί και τον κύριο λόγο για την επιλογή τους.

Συγκεκριμένα, για την αναζήτηση των βέλτιστων παραμέτρων (hyperparameter tuning - βελτιστοποίηση υπερπαραμέτρων) του CNN μοντέλου, χρησιμοποιήθηκε το KerasTuner [214], ένα κλιμακούμενο προγραμματιστικό πλαίσιο βελτιστοποίησης υπερπαραμέτρων (scalable hyperparameter optimization framework) της βιβλιοθήκης Keras. Το KerasTuner είναι ένας πολύ καλός τρόπος για τον έλεγχο διαφορετικών συνδυασμών υπερπαραμέτρων, όπως τον αριθμό των επιπέδων και των κόμβων σε κάθε επίπεδο, τον αριθμό και το μέγεθος των φίλτρων, των συναρτήσεων ενεργοποίησης κ.α. Χρησιμοποιείται για την εύρεση των βέλτιστων παραμέτρων για το μοντέλο βαθιάς μάθησης, οι συνδυασμοί των οποίων επιτυγχάνουν την υψηλότερη ακρίβεια. Το KerasTuner προτιμήθηκε, έναντι κάποιου άλλου αυτόματου αλγορίθμου ανάπτυξης και βελτιστοποίησης, όπως είναι για παράδειγμα το AutoKeras [215], λόγω της αμεσότητας και της απλότητας που παρέχει στον έλεγχο της αρχιτεκτονικής του μοντέλου.

Υπάρχουν πολλοί τρόποι αναζήτησης υπερπαραμέτρων στο Keras, οι οποίοι ονομάζονται tuners. Ο κάθε ένας χρησιμοποιεί διαφορετικό αλγόριθμο για την αναζήτηση, όπως είναι για παράδειγμα οι αλγόριθμοι HyperBand, Bayesian Optimization και RandomSearch [216]. Σε αυτή την εργασία, χρησιμοποιήθηκε ο RandomSearch, ο οποίος είναι ένας πιο «διαισθητικός» τρόπος για hyperparameter tuning. Με αυτόν τον αλγόριθμο αναζήτησης δοκιμάζονται τυχαία δείγματα συνδυασμών υπερπαραμέτρων. Έπειτα από κάθε δοκιμή, το σύνολο αυτών των συνδυασμών μπορεί να επαναρυθμιστεί καταλήγοντας σε αυτούς με τα πιο ικανοποιητικά αποτελέσματα, μέσω μιας διαδικασίας δοκιμών και ελέγχου.

Το KerasTuner παρέχει έναν πρακτικό τρόπο για τη δημιουργία ενός μοντέλου TNΔ και τον καθορισμό του χώρου αναζήτησης των (υπερ)παραμέτρων που θα αναζητήσει ο επιλεγμένος tuner. Αρχικά, μέσω μιας συνάρτησης δημιουργίας μοντέλων (build_model function), καθορίζεται η αρχιτεκτονική του μοντέλου και ορίζεται ο χώρος αναζήτησης των παραμέτρων, δηλαδή ποιες παράμετροι πρέπει να βελτιστοποιηθούν και σε ποιο εύρος. Η συνάρτηση αυτή δέχεται ως όρισμα ένα αντικείμενο του Keras, το hp, μέσω του οποίου γίνεται η δειγματοληψία των υπερπαραμέτρων από το KerasTuner. Για παράδειγμα η γραμμή:

```
hp.Int('units', min_value=32, max_value=512, step=32)
```

δηλώνει ότι η τιμή της συγκεκριμένης παραμέτρου θα είναι ένα ακέραιος αριθμός μεταξύ των τιμών 32 και 512 με βήμα 32.

Στην ίδια συνάρτηση, ορίζονται και οι υπόλοιπες παράμετροι όπως η συνάρτηση κόστους (loss function), τα μέτρα αξιολόγησης (metrics), ο αλγόριθμος βελτιστοποίησης (optimizer) κ.α., για τις οποίες υπάρχει δυνατότητα να συμπεριληφθούν στο χώρο αναζήτησης των υπερπαραμέτρων.

Η βιβλιοθήκη Keras προσφέρει υλοποιήσεις [217] για όλα τα επίπεδα που απαιτούνται για την ανάπτυξη ενός συνελκτικού δικτύου (ενότητα 4.5). Συγκεκριμένα, τα επίπεδα καθώς και οι βασικές παράμετροι που χρησιμοποιήθηκαν περιγράφονται συνοπτικά παρακάτω:

- InputLayer: Επίπεδο εισόδου του μοντέλου, με παράμετρο τις διαστάσεις του συνόλου εισόδου (tensor shape).
- Convolution1D: Επίπεδο συνέλιξης, με παραμέτρους τη μορφολογία εισόδου, τον αριθμό και τις διαστάσεις των φίλτρων συνέλιξης και τη συνάρτηση ενεργοποίησης.
- MaxPooling1D: Επίπεδο υποδειγματοληψίας, με παραμέτρους τις διαστάσεις του φίλτρου και το βήμα μετατόπισης (stride).
- ZeroPadding1D: Επίπεδο συμπλήρωσης, με παράμετρο τις διαστάσεις του φίλτρου.
- Activation: Επίπεδο ενεργοποίησης, με παράμετρο τη συνάρτηση ενεργοποίησης.

- Dropout: Επίπεδο πρόληψης υπερπροσαρμογής, με παράμετρο το ποσοστό απόρριψης.
- Dense: Πλήρες συνδεδεμένο επίπεδο, με παραμέτρους τις διαστάσεις του συνόλου εισόδου και εξόδου και τη συνάρτηση ενεργοποίησης.
- Flatten: Επίπεδο ισοπέδωσης.
- BatchNormalization: Επίπεδο ομαλοποίησης-κανονικοποίησης.

Επίσης, χρησιμοποιήθηκε το TensorBoard [220], ένα σετ εργαλείων του TensorFlow, για την παροχή μετρήσεων και απεικονίσεων που απαιτούνται κατά τις διαδικασίες μηχανικής μάθησης. Επιτρέπει την παρακολούθηση των μέτρων αξιολόγησης των πειραμάτων, την οπτικοποίηση του γραφήματος του μοντέλου, την προβολή ιστογραμμάτων για τα βάρη και τις πολώσεις, καθώς αυτά αλλάζουν με την πάροδο του χρόνου, κ.α.

Τέλος, η υλοποίηση των CNN μοντέλων πραγματοποιήθηκε με κώδικα Python, τοπικά στο Visual Studio Code v1.6, στο σύστημα Η/Υ που περιγράφηκε στην ενότητα 5.2.1 και εξ αποστάσεως (remotely) σε σημειωματάρια (notebooks) του Google Colab [219], μέσω του περιηγητή Google Chrome. Το Google Colab είναι ένα (δωρεάν, με κάποιους περιορισμούς) περιβάλλον ανάπτυξης Python που εκτελείται σε προγράμματα περιήγησης (browser) χρησιμοποιώντας το Google Cloud [221] και παρέχει δυνατότητα χρήσης ισχυρών GPU όπως την Tesla K80 της NVidia με αρχιτεκτονική CUDA.

Στον Πίνακα 6-1 παρουσιάζονται συγκεντρωτικά τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν για το πειραματικό μέρος αυτής της εργασίας, συμπεριλαμβανομένου και αυτών που περιγράφηκαν στην υποενότητα 5.2.1.

Πίνακας 6-1: Συγκεντρωτικός πίνακας προγραμματιστικών εργαλείων.

Γλώσσες προγραμματισμού:	Βιβλιοθήκες Python:	
C# v9.0	Pandas v1.35	Librosa v0.81
Python v3.79	Numpy v1.19	openSMILE v2.41
Περιβάλλοντα ανάπτυξης:	Scipy v1.41	Essentia v2.1
Visual Studio 2019 v16.7	Matplotlib v3.22	TensorFlow v2.7
Visual Studio Code v1.6	Seaborn v0.11	Keras v2.7
Google Colab (notebooks)		

Στη συνέχεια περιγράφεται αρχικά, η προετοιμασία του συνόλου δεδομένων εισόδου και στόχου, τα οποία θα χρησιμοποιηθούν στο συνελκτικό μοντέλο και έπειτα, αναλύεται η διαδικασία ανάπτυξης, εκπαίδευσης και επικύρωσής του.

6.2 Προετοιμασία δεδομένων

Όπως περιγράφηκε στην ενότητα 5.5 κατά την επίλυση του ευθέως προβλήματος, δημιουργήθηκε μια βάση δεδομένων (Πίνακες 5-5 και 5-6,) με τις ανεπεξέργαστες τιμές πλάτους των αρχείων ήχου (raw sample values), τις (σταθερές) τιμές των φυσικών παραμέτρων του μοντέλου της μεμβράνης και τις τιμές των παραμέτρων των μοτίβων πάστας που χρησιμοποιήθηκαν για την παραγωγή τους.

Επίσης, με βάση το παραπάνω σύνολο δεδομένων, έγιναν παράλληλα και κάποιοι υπολογισμοί διαφόρων ποσοτήτων, μέσω κώδικα Python, ώστε να γίνει εφικτός ο υπολογισμός της επιπλέον μάζας από την προσθήκη πάστας. Οι ποσότητες αυτές περιλαμβάνουν το αριθμό των κελιών του υπολογιστικού πλέγματος που χρησιμοποιήθηκε στην εφαρμογή της μεθόδου FDTD, το μήκος και τη μάζα του κάθε κελιού, την επιφάνεια που εφαρμόστηκε η πάστα σε πλήθος κελιών, την αρχική μάζα της μεμβράνης, την πρόσθετη μάζα της πάστας, τη συνολική μάζα μεμβράνης-πάστας κ.α.

Επομένως, μέσω της βάσης δεδομένων και των επιπλέον υπολογισμένων ποσοτήτων, δημιουργήθηκαν τα σύνολα δεδομένων εκπαίδευσης και επικύρωσης, όπως περιγράφονται στη συνέχεια.

6.2.1 Αναπαραστάσεις Εισόδου

Για τη δημιουργία του συνόλου εκπαίδευσης, που θα τροφοδοτήσει το συνελκτικό μοντέλο επεξεργάστηκαν, δημιουργήθηκαν και δοκιμάστηκαν διαφορετικές αναπαραστάσεις των δεδομένων εισόδου. Οι διαφορετικές αναπαραστάσεις του σήματος εισόδου, που δοκιμάστηκαν καθώς και λεπτομέρειες για αυτές παρουσιάζονται παρακάτω:

- Ανεπεξέργαστες τιμές πλάτους των αρχείων wav (raw sample values):
 - Μορφή αρχείων wav: PCM, Κανάλια: 1 (Mono)
 - Βάθος δειγματοληψίας (Bit-depth): 16 bit
 - Ρυθμός δειγματοληψίας (Sampling Rate): 22050 Hz
Τα αρχεία ήχου μειώθηκαν στα 22050 Hz, για περιορισμό του υπολογιστικού κόστους.
 - Μέγεθος συνόλου εκπαίδευσης: (2331, 22050)
 - 2331 αρχεία ήχου x 22050 δείγματα (1 sec διάρκεια)

- Ηχητικά χαρακτηριστικά (Audio features):
 - Χρησιμοποιώντας τις βιβλιοθήκες openSMILE, Essentia και Librosa εξήχθησαν αρκετά ηχητικά χαρακτηριστικά όπως:
Zero Crossing Rate, Spectral Energy, Spectral Centroid, Spectral Spread, Entropy, Flux & Rolloff, MFCCs Mean, HPCP, HPCP Entropy, Tuning Frequency, Spectral Contrast Coeffs., Spectral Contrast Valleys, Loudness EBU128 Integrated, Spectral RMS, Loudness, Pitch classes (semitone spectrum), Voice quality (Jitter, Shimmer), Chroma & Cens features κ.α.
Αρχικά εξήχθησαν συνολικά περισσότερες από 170 τιμές ηχητικών χαρακτηριστικών και έπειτα από ταξινόμηση συσχέτισης μέσω Correlation και P-value μειώθηκαν στις 40 διαφορετικές τιμές.
 - Μέγεθος συνόλου εκπαίδευσης: (2331, 177) και μετά την μείωση: (2331, 40)
 - Χρόνος επεξεργασίας και εξαγωγής: 30 min
(επιπλέον χρόνος σε σχέση με την πρώτη προσέγγιση των raw sample values)

- Εικόνες γραφικών αναπαραστάσεων και spectrograms:
 - Μέσω των βιβλιοθηκών openSMILE, Essentia και Librosa, για κάθε αρχείο ήχου δημιουργήθηκαν και αποθηκεύτηκαν ως εικόνες διάφορες γραφικές αναπαραστάσεις & spectrograms (φασματογράμματα), όπως:
Waveform & Phase-Space Portrait, Phase Spectrogram & Phase Unwrapped Spectrogram, Mag Linear Freq Spectrogram (& Log), MFCCs (& Log), Mel-scaled Spectrogram (& Log), Chromagram and Constant-Q Power (& Log).
 - Δημιουργήθηκαν συνολικά 13 σελίδες εικόνων spectrograms, οι οποίες μετατράπηκαν σε κλίμακα του γκρι (grayscale). Κάθε εικόνα αντιστοιχεί σε έναν δισδιάστατο πίνακα, με 256 τιμές του γκρι (0-255) σε κάθε κελί, ο οποίος έπειτα μετατρέπεται σε 1D διάνυσμα για χρήση από το μοντέλο.
 - Τα καλύτερα αποτελέσματα από τις εικόνες spectrograms λήφθηκαν χρησιμοποιώντας MFCCs spectrograms (στην κλίμακα Log/dB), επιβεβαιώνοντας την αντίστοιχη βιβλιογραφία [221 - 232].
 - Μέγεθος συνόλου εκπαίδευσης: (2331, 80x87) (για τα MFCCs spectrograms)
 - Χρόνος επεξεργασίας: >1 h (επιπλέον χρόνος σε σχέση με τα raw sample values)

Λεπτομερείς πληροφορίες για τα audio features, τα spectrograms και τον τρόπο εξαγωγής τους μπορεί να αναζητήσει κάποιος στα [198, 199, 200].

6.2.2 Δημιουργία συνόλων εκπαίδευσης & επικύρωσης

Μετά από αρκετές δοκιμές, μέσω εκπαίδευσης και επικύρωσης πολλών μοντέλων CNN διαφορετικής δομής και μεγέθους, έγινε φανερό ότι τα αποτελέσματα που προέκυψαν για το πρόβλημα ταξινόμησης και παλινδρόμησης, χρησιμοποιώντας και τις 3 αναπαραστάσεις δεδομένων εισόδου, που περιγράφηκαν στην προηγούμενη υποενότητα, ήταν παραπλήσια.

Ως είσοδος λοιπόν επιλέχθηκε η πρώτη προσέγγιση, το σύνολο δηλαδή των ανεπεξέργαστων τιμών πλάτους των αρχείων (raw sample values), λόγω του ότι η δημιουργία αυτού του συνόλου ήταν σαφώς ταχύτερη, χωρίς την απαίτηση των επιπλέον υπολογισμών και της επεξεργασίας που χρειάζεται η εξαγωγή των audio features και των spectrograms και αφετέρου λόγω του ότι οι ακατέργαστες τιμές περιλαμβάνουν όλη την «πληροφορία» του ήχου, γεγονός που μπορεί να αποδειχθεί κομβικό στη συνέχεια των πειραμάτων.

Επομένως, το σύνολο δεδομένων εισόδου αποτελείται από 22050 δείγματα (1 sec ήχου) για καθένα από τα 2331 δείγματα ήχου.

Εκτός του συνόλου δεδομένων εισόδου (input dataset), η διαδικασία εκπαίδευσης του CNN μοντέλου χρειάζεται επίσης το σύνολο δεδομένων εξόδου-στόχου (target dataset), το οποίο είναι η πραγματική-επιθυμητή έξοδος για ένα δεδομένο σύνολο εισόδου. Δημιουργήθηκαν επομένως δυο σύνολα εξόδου (στόχων), τα οποία περιέχουν τις πραγματικές, επιθυμητές τιμές εξόδου, οι οποίες χρησιμοποιήθηκαν για τον υπολογισμό των ηχητικών δειγμάτων κατά την επίλυση του ευθέως προβλήματος.

Συγκεκριμένα δημιουργήθηκαν τα εξής σύνολα:

- το σύνολο δεδομένων εισόδου (input dataset): το οποίο περιλαμβάνει τις ανεπεξέργαστες τιμές πλάτους των αρχείων ήχου (raw sample values), κανονικοποιημένες (normalized) στο διάστημα $[-1, 1]$, με διαστάσεις: (2331, 22050) και συνολικού μεγέθους 196 MB.
- το σύνολο εξόδου-στόχου για το πρόβλημα ταξινόμησης (paste pattern output dataset): το οποίο περιλαμβάνει τις ακέραιες τιμές (από 0 έως 6) που αντιστοιχούν στις 7 περιπτώσεις του μοτίβου πάστας (ενότητα 5.4), με διαστάσεις: (2331, 1) και συνολικού μεγέθους: 28 KB.
- το σύνολο εξόδου-στόχου για το πρόβλημα παλινδρόμησης (extra mass output dataset): το οποίο περιλαμβάνει τις τιμές της πρόσθετης μάζας (από 0 έως 0.5 kgr),

λόγω της εφαρμογής πάστας στην επιφάνεια της μεμβράνης, με διαστάσεις: (2331, 1), και συνολικού μεγέθους: 28 KB.

Η επεξεργασία, η ανάλυση και γενικότερα η διαχείριση όλων των συνόλων των δεδομένων έγινε με τη χρήση dataframes της βιβλιοθήκης Pandas, ενώ η αποθήκευση και η ανάκτηση των δεδομένων και των μοντέλων, με τη χρήση της ενσωματωμένης βιβλιοθήκης pickle [233] της Python.

Στον Πίνακα 6-2, παρουσιάζονται τμήματα αυτών των συνόλων και συγκεκριμένα: (α) Τμήμα του συνόλου δεδομένων εισόδου, με τις raw sample values. (β) Τμήμα του συνόλου δεδομένων στόχου, με τις τιμές που αντιστοιχούν στα 7 μοτίβα πάστας και (γ) Τμήμα του συνόλου δεδομένων στόχου, με τις τιμές της πρόσθετης μάζας.

Πίνακας 6-2: (α) Σύνολα δεδομένων εισόδου, (β) σύνολο δεδομένων στόχου για τα μοτίβα πάστας, (γ) σύνολο δεδομένων στόχου για την πρόσθετη μάζα από την εφαρμογή πάστας.

(α) index	audio sample filename	0	1	2	...	22048	22049
0	m_21-06-27-041122_1_H0.0045-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT0-PP0.5-WD0.02-...	0.0653	0.1170	0.1020	...	0.0011	0.0012
1	m_21-06-27-084000_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT6-PP1.0-WD0.02-R...	0.0663	0.1189	0.1037	...	0.0027	0.0029
2	m_21-06-27-044244_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.9-WD0.09-R...	0.0623	0.1117	0.0974	...	-0.0021	-0.0023
3	m_21-06-27-043249_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.3-WD0.1-RP...	0.0670	0.1202	0.1048	...	-0.0029	-0.0031
4	m_21-06-27-044526_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP1.1-WD0.05-R...	0.0544	0.0975	0.0850	...	-0.0020	-0.0021

(β) index	audio sample filename	paste_pattern
0	m_21-06-27-041122_1_H0.0045-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT0-PP0.5-WD0.02-...	0
1	m_21-06-27-084000_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT6-PP1.0-WD0.02-R...	6
2	m_21-06-27-044244_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.9-WD0.09-R...	1
3	m_21-06-27-043249_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.3-WD0.1-RP...	1
4	m_21-06-27-044526_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP1.1-WD0.05-R...	1

(γ) index	audio sample filename	extra_mass_from_adding_paste
0	m_21-06-27-041122_1_H0.0045-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT0-PP0.5-WD0.02-...	0.0000
1	m_21-06-27-084000_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT6-PP1.0-WD0.02-R...	0.0010
2	m_21-06-27-044244_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.9-WD0.09-R...	0.0346
3	m_21-06-27-043249_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP0.3-WD0.1-RP...	0.0134
4	m_21-06-27-044526_1_H0.003-R0.25-T800.0-D1.0-X52-Y52-A0.7-PT1-PP1.1-WD0.05-R...	0.0236

Στη συνέχεια, το κάθε ένα από τα παραπάνω σύνολα δεδομένων διαχωρίστηκε σε δύο διαφορετικά σύνολα, το σύνολο δεδομένων εκπαίδευσης (train set) και το σύνολο δεδομένων επικύρωσης (validation set), σε αναλογία 87% και 13% αντίστοιχα. Επομένως, χρησιμοποιήθηκαν 2027 τιμές για την εκπαίδευση (train) και 304 τιμές για την επικύρωση (validation). Στον Πίνακα 6-3 αναφέρονται οι διαστάσεις των συνόλων αυτών.

Πίνακας 6-3: Τα σύνολα δεδομένων εκπαίδευσης και επικύρωσης για τα μοτίβα πάστας (pat) και για την πρόσθετη μάζα (mass).

Σύνολο	Διαστάσεις	για την ταξινόμηση των μοτίβων	Σύνολο	Διαστάσεις	για την εκτίμηση της πρόσθετης μάζας
Xtrain_pat	(2027,22050)	σύνολο εκπαίδευσης (εισόδου X)	Xtrain_mass	(2027,2050)	σύνολο εκπαίδευσης (εισόδου X)
Ytrain_pat	(2027,1)	σύνολο εκπαίδευσης (στόχου Y)	Ytrain_mass	(2027,1)	σύνολο εκπαίδευσης (στόχου Y)
Xval_pat	(304,22050)	σύνολο επικύρωσης (εισόδου X)	Xval_mass	(304,22050)	σύνολο επικύρωσης (εισόδου X)
Yval_pat	(304,1)	σύνολο επικύρωσης (στόχου Y)	Yval_mass	(304, 1)	σύνολο επικύρωσης (στόχου Y)

Κατά τη διαδικασία εκπαίδευσης, το σύνολο δεδομένων εκπαίδευσης διαχωρίζεται εκ νέου σε train set και test set, σε αναλογία 80% και 20% (με μια διαδικασία cross-validation), ώστε να ελέγχεται η απόδοση σε κάθε κύκλο εκπαίδευσης.

Πρέπει να διευκρινιστεί επίσης ότι τα σύνολα δεδομένων εκπαίδευσης εισόδου (`Xtrain_pat`, `Xtrain_mass`) είναι πανομοιότυπα καθώς και τα δυο περιέχουν τις raw sample values των αρχείων ήχου. Το ίδιο ισχύει και για τα σύνολα δεδομένων επικύρωσης εισόδου (`Xval_pat`, `Xval_mass`). Τα «αντίτυπα» αυτά δημιουργήθηκαν ώστε να τροφοδοτήσουν ξεχωριστά την είσοδο του μοντέλου που θα επιλύσει, ταυτόχρονα, το πρόβλημα της ταξινόμησης (classification task) και το πρόβλημα της παλινδρόμησης (regression task), όπως θα περιγράψει στη συνέχεια. Τέλος, αξίζει να σημειωθεί ότι τα σύνολα επικύρωσης παρέμειναν εντελώς άγνωστα στο μοντέλο μέχρι το τέλος της εκπαίδευσής του.

Στη συνέχεια όλα τα σύνολα δεδομένων μετατράπηκαν σε διανύσματα και η σειρά των τιμών τους τυχαιοποιήθηκε (randomize), ώστε να μην επηρεάσει την εκπαίδευση του μοντέλου. Επίσης, οι τιμές των `Xtrain_mass` και `Yval_mass` κανονικοποιήθηκαν (normalized) στο διάστημα $[0, 1]$, για μικρότερο υπολογιστικό κόστος.

Μετά τη διαμόρφωση των συνόλων εκπαίδευσης και επικύρωσης, ακολουθεί η ανάπτυξη του μοντέλου CNN και ο καθορισμός της δομής και των παραμέτρων του. Η διαδικασία αυτή περιγράφεται αναλυτικά στην επόμενη ενότητα.

6.3 Ανάπτυξη CNN μοντέλου

Το επόμενο βήμα στην ανάπτυξη του μοντέλου, περιλαμβάνει τη δημιουργία του μοντέλου, τον καθορισμό της αρχιτεκτονικής του, αλλά και του χώρου αναζήτησης των υπερπαραμέτρων του, ώστε να βελτιστοποιηθούν. Αυτή η διαδικασία πραγματοποιείται με τον ορισμό της συνάρτησης δημιουργίας του μοντέλου (`build_model`), η οποία δέχεται ως όρισμα το αντικείμενο `hp` της βιβλιοθήκης Keras, μέσω του οποίου καθορίζεται και ο χώρος αναζήτησης των υπερπαραμέτρων του.

Η μεθοδολογία που ακολουθήθηκε στην παρούσα εργασία περιλαμβάνει την πραγματοποίηση μιας σειράς αρχικών δοκιμών, χρησιμοποιώντας επαναληπτικές δομές και τον αλγόριθμο αναζήτησης RandomSearch του KerasTuner. Οι δοκιμές αυτές αποσκοπούν αφενός στο να προκύψει μια βασική αρχιτεκτονική για το μοντέλο, η οποία να δίνει σταθερά καλά αποτελέσματα και αφετέρου, στο να περιοριστεί ο χώρος αναζήτησης των υπερπαραμέτρων, ώστε έπειτα να γίνει πιο σύντομα και αποτελεσματικά η βελτιστοποίησή τους.

6.3.1 Καθορισμός βασικής αρχιτεκτονικής

Μέσω διαδοχικών δοκιμών ανάπτυξης, εκπαίδευσης και επικύρωσης, δοκιμάστηκαν πολλά διαφορετικά μοντέλα με διάφορους συνδυασμούς αρχιτεκτονικών και παραμέτρων. Συγκεκριμένα, μέσω του KerasTuner και του αλγόριθμου αναζήτησης RandomSearch, πραγματοποιήθηκε μια πρώτη προσπάθεια αναζήτησης των καλύτερων παραμέτρων, ώστε να καθοριστεί η βασική αρχιτεκτονική και να περιοριστεί ο χώρος αναζήτησης των υπερπαραμέτρων του μοντέλου. Χρησιμοποιώντας επίσης επαναληπτικές δομές για τη δημιουργία επιπλέον επιπέδων (συνελκτικών, πλήρως συνδεδεμένων, συγκέντρωσης και απόρριψης), δοκιμάστηκαν πολλά CNN μοντέλα με διαφορετικά πλάτη και βάθη.

Με παρόμοια διαδικασία δοκιμάστηκαν επίσης διαφορετικά είδη συναρτήσεων ενεργοποίησης (όπως ReLu, Tanh και Linear), συναρτήσεων κόστους (όπως MAE, MSE, Cross Entropy Loss), βελτιστοποιητών (όπως SGD, Adagrad και ADAM), μέτρων αξιολόγησης (όπως MAE, MSE, RMSE, R²score και Accuracy), αλλά και διαφορετικοί ρυθμοί εκπαίδευσης. Επιπλέον, δοκιμάστηκαν μοντέλα «αλυσίδας» (chain models) στα οποία η έξοδος της ταξινόμησης χρησιμοποιήθηκε ως πρόσθετη είσοδος για την εργασία παλινδρόμησης ενώ, όπως έχει ήδη αναφερθεί στην υποενότητα 6.2.1, δοκιμάστηκαν και μοντέλα που τροφοδοτήθηκαν με ηχητικά χαρακτηριστικά και με εικόνες spectrograms, χωρίς όμως σημαντικά πλεονεκτήματα.

Καθώς η διαδικασία δοκιμών και ελέγχου προχωρούσε, μετά από κάθε βήμα δοκιμών επιλέγονταν η βασική αρχιτεκτονική (πλήθος και είδος επιπέδων) και οι τιμές των βασικών παραμέτρων (μέγεθος κάθε επιπέδου και μέγεθος φίλτρων) με τα καλύτερα αποτελέσματα. Έπειτα η διαδικασία ξεκινούσε από την αρχή, με όλο και μικρότερο εύρος για το χώρο αναζήτησης. Αυτή η διαδικασία ήταν και η πιο χρονοβόρα, καθώς σε αρκετές περιπτώσεις η εκπαίδευση και η αξιολόγηση του κάθε υπό δοκιμή διαφορετικού μοντέλου διαρκούσε από αρκετά λεπτά μέχρι πολλές ώρες.

Εφαρμόζοντας, λοιπόν την παραπάνω μεθοδολογία, σταδιακά, προέκυψε μια αρχιτεκτονική συνελκτικού μοντέλου, που έχοντας ως είσοδο τις ανεπεξέργαστες τιμές πλάτους των αρχείων ήχου, ταξινομεί το μοτίβο πάστας και ταυτόχρονα εκτιμά την πρόσθετη μάζα από την εφαρμογή της πάστας στη μεμβράνη (Εικόνα 6-1).

Η τελική αρχιτεκτονική που διερευνήθηκε αποτελείται ουσιαστικά από 2 παράλληλους ανεξάρτητους κλάδους (branches), οι οποίοι έχουν κοινή είσοδο (raw sample values) και καταλήγουν σε διπλή έξοδο (multiple target), μια για την ταξινόμηση των μοτίβων (ακέραια τιμή από 0 έως 6) και μια για την παλινδρόμηση (εκτίμηση πρόσθετης μάζας,

1 πραγματική τιμή σε kg). Ο κάθε κλάδος περιλαμβάνει 2 συνελκτικά επίπεδα (Conv1D), ακολουθούμενα από τα αντίστοιχα επίπεδα συγκέντρωσης (MaxPooling). Έπειτα ακολουθεί ένα επίπεδο απόρριψης (Dropout) και το επίπεδο ισοπέδωσης (Flatten). Η έξοδος του επιπέδου Flatten τροφοδοτεί ένα πλήρως συνδεδεμένο επίπεδο (Dense), που έπειτα από ένα ακόμα επίπεδο Dropout καταλήγει τελικά στο τελευταίο πλήρως συνδεδεμένο επίπεδο εξόδου.

Για την ενεργοποίηση όλων των επιπέδων χρησιμοποιήθηκαν συναρτήσεις ReLu, εκτός των τελευταίων επιπέδων εξόδου, όπου χρησιμοποιήθηκε η Softmax για την έξοδο της ταξινόμησης, και η Linear για την έξοδο της παλινδρόμησης (υποενότητα 4.3.4).

Ως συνάρτηση κόστους για το πρόβλημα ταξινόμησης των μοτίβων, επιλέχθηκε η Αραιή Εγκάρσια Εντροπία για ποιοτικές μεταβλητές (Sparse Categorical Cross Entropy - SCCE) [218], ενώ για το πρόβλημα παλινδρόμησης της πρόσθετης μάζας επιλέχθηκε το μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE) (υποενότητα 4.4.3). Η Sparse Categorical Cross Entropy χρησιμοποιήθηκε λόγω του είδους της ταξινόμησης, δηλαδή της ταξινόμησης πολλών κλάσεων (multi-class classification) ποιοτικών τιμών, όπως επίσης και λόγω του ότι οι κλάσεις (κατηγορίες) των μοτίβων πάστας είναι αμοιβαία αποκλειόμενες, δηλαδή κάθε δείγμα ανήκει ακριβώς σε μια κατηγορία.

Στη διαδικασία εκπαίδευσης του συνολικού μοντέλου, χρησιμοποιήθηκε το άθροισμα και των δύο απωλειών επικύρωσης (validation loss) ταξινόμησης και παλινδρόμησης, ενώ για τον αλγόριθμο βελτιστοποίησης (optimizer) χρησιμοποιήθηκε ο δημοφιλής Adam optimizer (υποενότητα 4.4.4). Τέλος, ως μέτρα αξιολόγησης επιλέχθηκαν η ακρίβεια (classification accuracy) για την ταξινόμηση και τα MAE, MSE, RMSE και R-squared για το πρόβλημα παλινδρόμησης (υποενότητα 4.4.6).

6.3.2 Βελτιστοποίηση υπερπαραμέτρων

Μετά τον καθορισμό της βασικής αρχιτεκτονικής του μοντέλου και τον καθορισμό του τελικού, μικρότερου πλέον χώρου αναζήτησης των παραμέτρων του, χρησιμοποιήθηκε ξανά το KerasTuner, ώστε να πραγματοποιηθεί η τελική διαδικασία βελτιστοποίησης των υπερπαραμέτρων (fine-tuning hyperparameters). Το KerasTuner σε συνδυασμό με τον αλγόριθμο αναζήτησης RandomSearch, δημιουργεί αυτόματα διάφορα μοντέλα με βάση το χώρο αναζήτησης, όπως αυτός καθορίστηκε προηγουμένως και τα εκπαιδεύει για ένα συγκεκριμένο αριθμό κύκλων εκπαίδευσης (epochs), μεγέθους παρτίδας (batch size) και ρυθμού μάθησης (learning rate), εξάγοντας και αποθηκεύοντας τις βέλτιστες παραμέτρους για το μοντέλο με τα καλύτερα αποτελέσματα.

Ταυτόχρονα μπορούν να οριστούν διάφορες συναρτήσεις επανάκλησης (callbacks) [234, 235], οι οποίες χρησιμοποιούνται από το KerasTuner για να βοηθήσουν στη διαδικασία εκπαίδευσης, μειώνοντας τον απαιτούμενο χρόνο και αντιμετωπίζοντας προβλήματα υπερ-εκπαίδευσης και υπο-εκπαίδευσης (Overfitting & Underfitting, υποενότητα 4.4.7).

Οι συναρτήσεις callbacks που χρησιμοποιήθηκαν για την υποβοήθηση της εκπαίδευσης του μοντέλου είναι:

- ReduceLROnPlateau [236]: μεταβάλλει το ρυθμό μάθησης (learning rate) κατά τη διάρκεια της εκπαίδευσης. Η συνάρτηση ορίζει μια αρχική τιμή για το ρυθμό και σταδιακά το μειώνει (με συγκεκριμένο τρόπο) σε μια ελάχιστη τιμή, με βάση την απώλεια επικύρωσης.
- EarlyStopping [237]: παρακολουθεί την απώλεια επικύρωσης και τερματίζει την εκπαίδευση, αν η απώλεια δε βελτιωθεί μετά από συγκεκριμένους κύκλους εκπαίδευσης.
- CSVLogger [238]: αποθηκεύει τα αποτελέσματα κάθε κύκλου εκπαίδευσης σε αρχείο της μορφής csv, ώστε να μπορούν να αναλυθούν αργότερα.
- Tensorboard: συνάρτηση για την οπτικοποίηση και την ανάλυση των αποτελεσμάτων κατά τη διάρκεια της εκπαίδευσης των μοντέλων που διερευνώνται.

Μετά το τέλος της διαδικασίας βελτιστοποίησης, το KerasTuner επιστρέφει τις βέλτιστες τιμές των υπερπαραμέτρων που βρέθηκαν, καθώς και το μοντέλο που παρουσίασε το καλύτερο αποτέλεσμα χρησιμοποιώντας αυτές τις τιμές. Το βέλτιστο μοντέλο, μαζί με τις τιμές των βαρών, των πολώσεων και των παραμέτρων του, αποθηκεύεται ώστε να μπορεί να ανακτηθεί άμεσα για τη διαδικασία της εκπαίδευσής του, η οποία ακολουθεί.

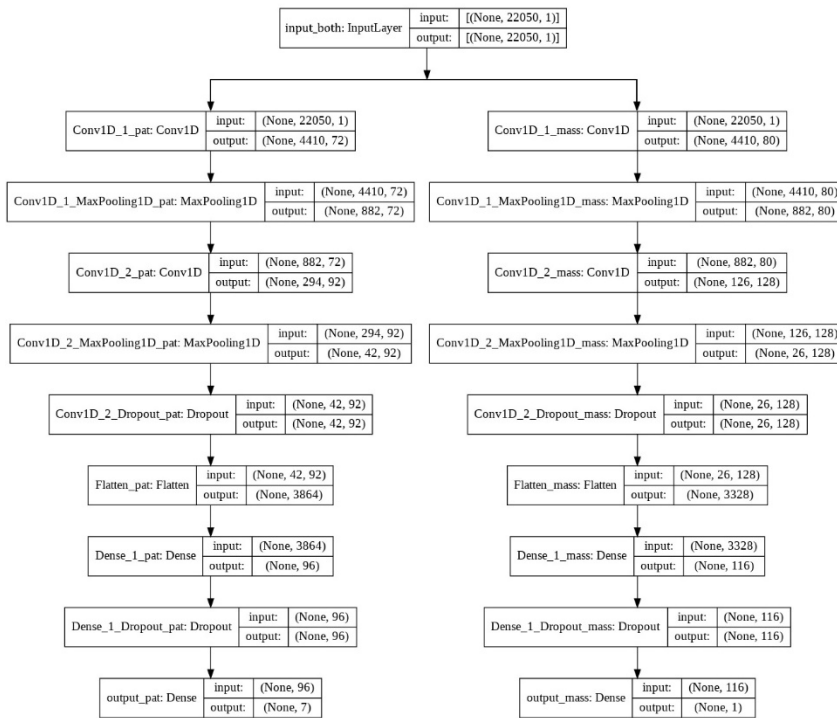
Η τελική διαδικασία fine-tuning hyperparameters, πραγματοποιήθηκε στο Google Colab και σε χρόνο 52 min, επέστρεψε τις βέλτιστες τιμές των υπερπαραμέτρων και το αποδοτικότερο μοντέλο, με τη μικρότερη απώλεια επικύρωσης. Η δομή του μοντέλου και οι τιμές των παραμέτρων παρουσιάζονται Πίνακα 6-4, ενώ η δομή του σε μορφή δένδρου στην Εικόνα 6-1.

Τέλος, η τελική αρχιτεκτονική του αποδοτικότερου CNN μοντέλου, το πλήθος και το είδος κάθε επιπέδου από τα οποία αποτελείται, καθώς και ο αριθμός των μονάδων κάθε επιπέδου και το μέγεθος των φίλτρων του, παρουσιάζεται στην Εικόνα 6-2, σε μορφή διαγράμματος.

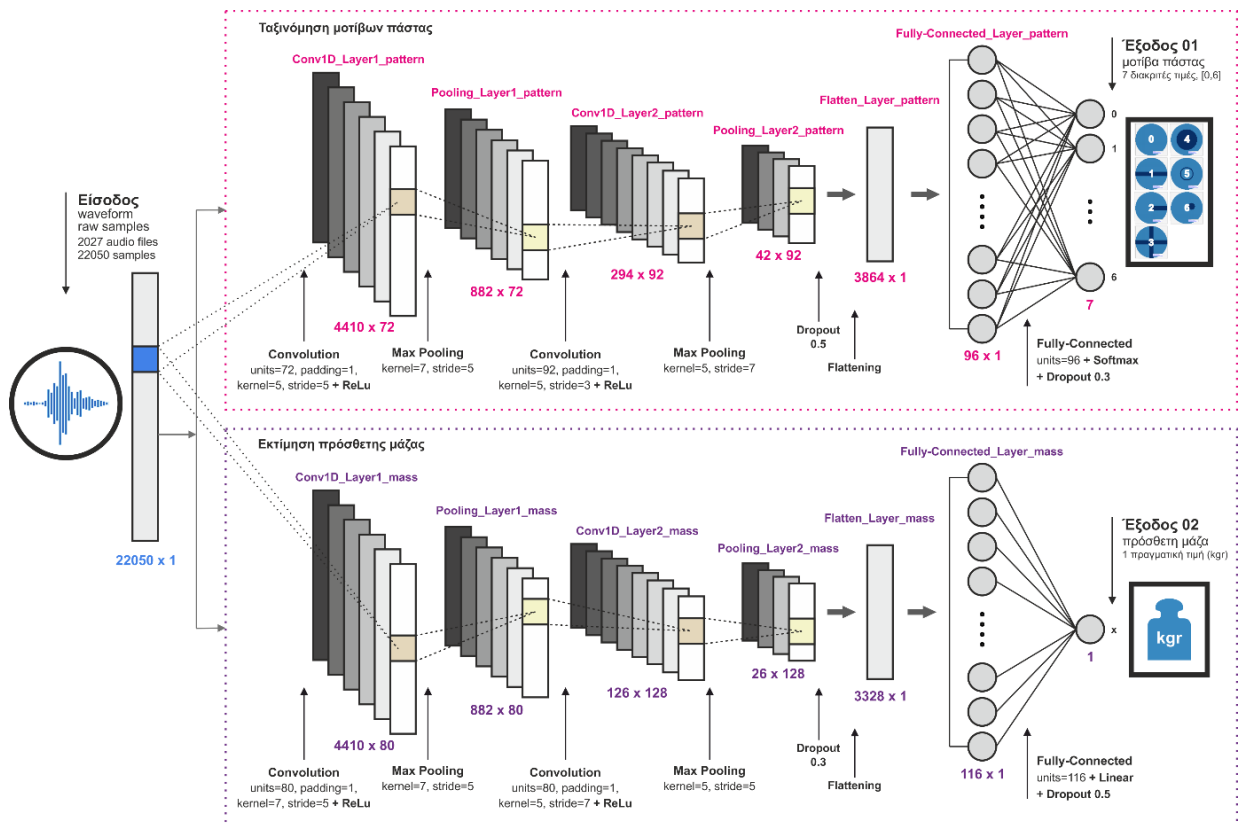
Πίνακας 6-4: (α) Δομή του μοντέλου με την μικρότερη απώλεια επικύρωσης. Επίπεδα και αριθμός παραμέτρων. (β) Βέλτιστες τιμές παραμέτρων για το αποδοτικότερο μοντέλο. Βέλτιστο μέγεθος μονάδων, φίλτρων και άλλων παραμέτρων για κάθε επίπεδο.

(α) Μοντέλο με την μικρότερη απώλεια επικύρωσης			
Best val_loss So Far: 0.12427633255720139			
Total elapsed time: 00 h 52 m 29 s			
> Best Model Layers:			
Layer (type)	Output Shape	Param #	Connected to
input_both (InputLayer)	[(None,22050,1)]	0	
Conv1D_1_pat (Conv1D)	(None, 4410, 72)	432	input_pat[0][0]
Conv1D_1_mass (Conv1D)	(None, 4410, 80)	640	input_mass[0][0]
Conv1D_1_MaxPooling1D_pat	(None, 882, 72)	0	Conv1D_1_pat (Conv1D)
Conv1D_1_MaxPooling1D_mass	(None, 882, 80)	0	Conv1D_1_mass (Conv1D)
Conv1D_2_pat (Conv1D)	(None, 294, 92)	33212	Conv1D_1_MaxPooling1D_pat[0][0]
Conv1D_2_mass (Conv1D)	(None, 126, 128)	51328	Conv1D_1_MaxPooling1D_mass[0][0]
Conv1D_2_MaxPooling1D_pat	(None, 42, 92)	0	Conv1D_2_pat (Conv1D)
Conv1D_2_MaxPooling1D_mass	(None, 26, 128)	0	Conv1D_2_mass (Conv1D)
Conv1D_2_Dropout_pat (Dropout)	(None, 42, 92)	0	Conv1D_2_MaxPooling1D_pat[0][0]
Conv1D_2_Dropout_mass (Dropout)	(None, 26, 128)	0	Conv1D_2_MaxPooling1D_mass[0][0]
Flatten_pat (Flatten)	(None, 3864)	0	Conv1D_2_MaxPooling1D_pat[0][0]
Flatten_mass (Flatten)	(None, 3328)	0	Conv1D_2_MaxPooling1D_mass[0][0]
Dense_1_pat (Dense)	(None, 96)	371040	Flatten_pat (Flatten)
Dense_1_mass (Dense)	(None, 116)	386164	Flatten_mass (Flatten)
Dense_1_Dropout_pat (Dropout)	(None, 96)	0	Dense_1_pat (Dense)
Dense_1_Dropout_mass (Dropout)	(None, 116)	0	Dense_1_mass (Dense)
output_pat (Dense)	(None, 7)	679	Dense_1_Dropout_pat (Dropout)
output_mass (Dense)	(None, 1)	117	Dense_1_Dropout_mass (Dropout)

(β) Βέλτιστες τιμές παραμέτρων για το αποδοτικότερο μοντέλο	
Objective (name='val_loss', direction='min ')	
Score: 0.12427633255720139	
> Hyperparameters:	
Conv1D_1_pat_units: 72	Conv1D_1_mass_units: 80
Conv1D_1_pat_strides: 5	Conv1D_1_mass_strides: 5
Conv1D_1_pat_kernel: 5	Conv1D_1_mass_kernel: 7
Conv1D_1_pat_maxnorm: 5	Conv1D_1_mass_maxnorm: 7
Conv1D_1_pat_pool_size: 7	Conv1D_1_mass_pool_size: 7
Conv1D_1_pat_pool_strides: 5	Conv1D_1_mass_pool_strides: 5
Conv1D_2_pat_units: 92	Conv1D_2_mass_units: 128
Conv1D_2_pat_strides: 3	Conv1D_2_mass_strides: 7
Conv1D_2_pat_kernel: 5	Conv1D_2_mass_kernel: 5
Conv1D_2_pat_maxnorm: 7	Conv1D_2_mass_maxnorm: 5
Conv1D_2_pat_pool_size: 5	Conv1D_2_mass_pool_size: 5
Conv1D_2_pat_pool_strides: 7	Conv1D_2_mass_pool_strides: 5
Conv1D_2_Dropout_pat_rate: 0.5	Conv1D_2_Dropout_mass_rate: 0.3
Dense_1_pat_units: 96	Dense_1_mass_units: 116
Dense_1_pat_maxnorm: 5	Dense_1_mass_maxnorm: 7
Dense_1_Dropout_pat_rate: 0.3	Dense_1_Dropout_mass_rate: 0.5



Εικόνα 6-1: Η αρχιτεκτονική και οι διαστάσεις κάθε επιπέδου του αποδοτικότερου CNN μοντέλου, όπως αυτό καθορίστηκε από το KerasTuner, σε μορφή δένδρου.



Εικόνα 6-2: Η αρχιτεκτονική και ο αριθμός των μονάδων και το μέγεθος των φίλτρων κάθε επιπέδου του τελικού CNN μοντέλου, σε μορφή διαγράμματος.

6.4 Εκπαίδευση CNN μοντέλου

Αφού καθορίστηκαν οι βέλτιστες τιμές υπερπαραμέτρων και η γενικότερη αρχιτεκτονική του αποδοτικότερου CNN μοντέλου, η διαδικασία υλοποίησης συνεχίζεται με την εκπαίδευσή του, μέσω της μεθόδου `fit`.

Ο μέγιστος αριθμός κύκλων εκπαίδευσης καθορίστηκε στις 300 και το μέγεθος της παρτίδας στα 64. Όπως και προηγουμένως, χρησιμοποιήθηκαν οι callback συναρτήσεις `ReduceLRonPlateau`, `EarlyStopping`, `CSVLogger` και `TensorBoard`, καθώς επίσης και η callback συνάρτηση `History` [239], η οποία καταγράφει τις μετρήσεις εκπαίδευσης για κάθε κύκλο. Οι αποθηκευμένες αυτές μετρήσεις μπορεί να χρησιμοποιηθούν για τη δημιουργία γραφημάτων και την ανάλυση της διαδικασίας εκπαίδευσης.

Εφαρμόζοντας τη διαδικασία εκπαίδευσης με τις παραπάνω παραμέτρους, σε περιβάλλον `Google Colab` (GPU `Tesla T4`), η εκπαίδευση τερματίστηκε μετά από 125 κύκλους εκπαίδευσης, λόγω της callback συνάρτησης `EarlyStopping` και ο ρυθμός μάθησης μειώθηκε σταδιακά από 0.001 σε 0.000001. Η διάρκεια της εκπαίδευσης ήταν περίπου 3 min. Όπως αναφέρθηκε στην υποενότητα 6.3.1, στη διαδικασία εκπαίδευσης του μοντέλου, ως συνολική απώλεια επικύρωσης (`validation loss`) χρησιμοποιήθηκε το άθροισμα και των δύο απωλειών επικύρωσης, δηλαδή της ταξινόμησης (`Sparse Categorical Cross Entropy`) και της παλινδρόμησης (`MAE`).

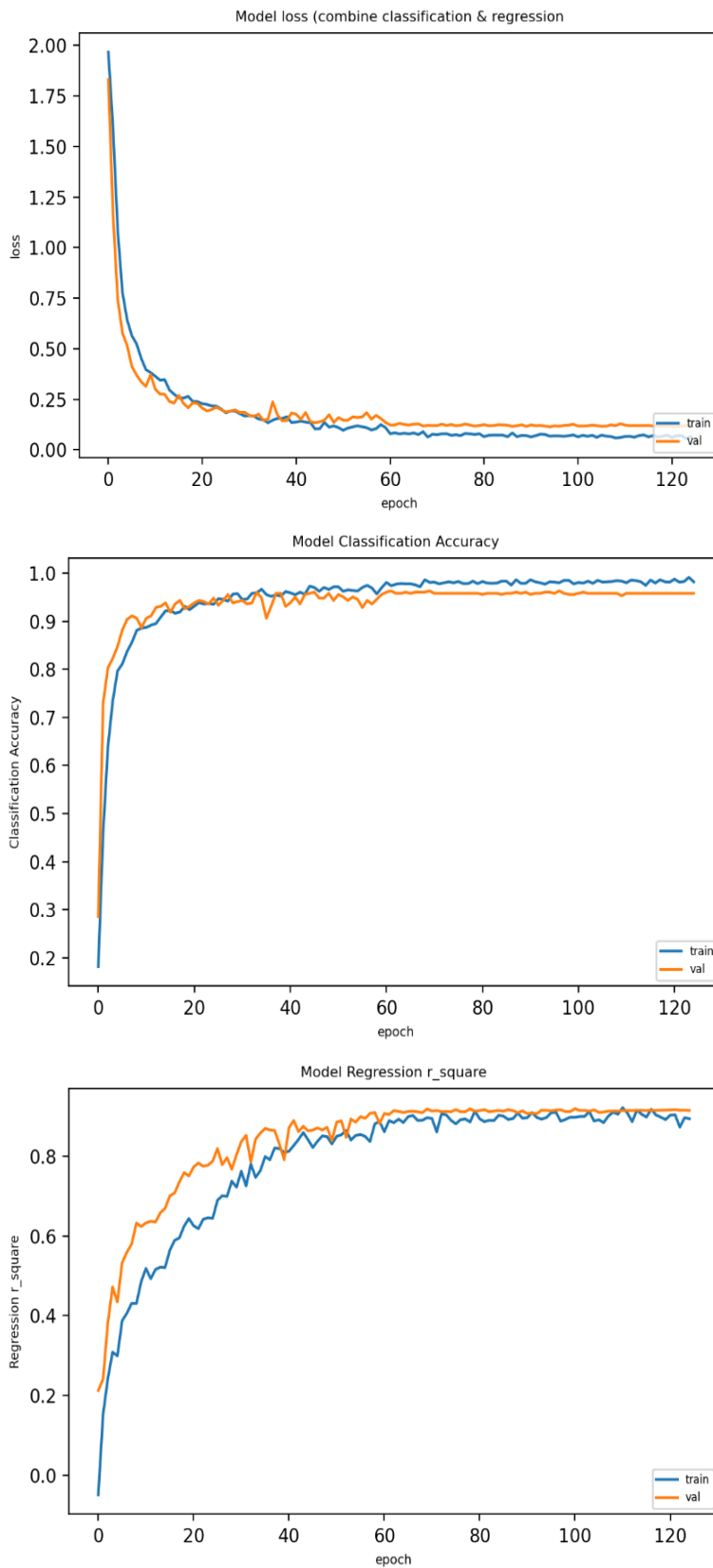
Στον Πίνακα 6-5, παρουσιάζονται το αρχικό και το τελικό βήμα (κύκλοι εκπαίδευσης) της διαδικασίας εκπαίδευσης. Η απώλεια `val_loss` είναι η τιμή της συνάρτησης κόστους για τα `cross-validation`¹⁰ [240] δεδομένα (διασταυρούμενη επικύρωση δεδομένων), που λαμβάνει χώρα σε κάθε κύκλο εκπαίδευσης, ενώ η απώλεια `loss` είναι η τιμή της συνάρτησης κόστους για τα δεδομένα εκπαίδευσης.

¹⁰ Σε κάθε εποχή, η κάθε παρτίδα δεδομένων εκπαίδευσης διαχωρίζονται σε `train data` για την εκπαίδευση και `test data` για τον έλεγχο της απόδοσης.

Πίνακας 6-5: Αρχικός και τελικός κύκλος εκπαίδευσης.

<p>Epoch 1/300 26/26 [=====] - 17s 114ms/step - loss: 2.0072 - output_pat_loss: 1.9269 - output_mass_loss: 0.0803 - output_pat_accuracy: 0.1640 - output_mass_r_square: -0.1497 - output_mass_rmse: 0.0803 - output_mass_mse: 0.0197 - val_loss: 1.8313 - val_output_pat_loss: 1.7760 - val_output_mass_loss: 0.0553 - val_output_pat_accuracy: 0.2857 - val_output_mass_r_square: 0.2126 - val_output_mass_rmse: 0.0553 - val_output_mass_mse: 0.0101 ...</p>
<p>Epoch 125/300 26/26 [=====] - 1s 44ms/step - loss: 0.0634 - output_pat_loss: 0.0447 - output_mass_loss: 0.0186 - output_pat_accuracy: 0.9846 - output_mass_r_square: 0.9033 - output_mass_rmse: 0.0186 - output_mass_mse: 0.0014 - val_loss: 0.1183 - val_output_pat_loss: 0.1044 - val_output_mass_loss: 0.0140 - val_output_pat_accuracy: 0.9581 - val_output_mass_r_square: 0.9147 - val_output_mass_rmse: 0.0140 - val_output_mass_mse: 0.0012</p>
<p>Epoch 00125: ReduceLROnPlateau reducing learning rate to 1.0000000656873453e-06. Epoch 00125: early stopping time: 2min 39s</p>

Στην Εικόνα 6-3 παρουσιάζονται τα διαγράμματα απώλειας (accuracy για την ταξινόμηση και R-squared για την παλινδρόμηση) κατά τη διάρκεια των κύκλων εκπαίδευσης, για τα σύνολα εκπαίδευσης (train) και επικύρωσης (validation). Όπως διακρίνεται στα διαγράμματα, οι καμπύλες της απώλειας για το σύνολο εκπαίδευσης και για το σύνολο επικύρωσης συγκλίνουν και σταθεροποιούνται στην ελάχιστη (accuracy) - μέγιστη (R-squared) τιμή τους μετά τον 120^ο κύκλο εκπαίδευσης. Επίσης η καμπύλη απώλειας του συνόλου εκπαίδευσης είναι υψηλότερα (και αντίστοιχα χαμηλότερα στην περίπτωση της παλινδρόμησης) της καμπύλης του συνόλου επικύρωσης, γεγονός που επιβεβαιώνει, ως έναν βαθμό, τη «σωστή» εκπαίδευση του μοντέλου, χωρίς προβλήματα overfitting. Παρατηρείται, επίσης, ότι μετά τον 60^ο κύκλο εκπαίδευσης, η απόδοση του μοντέλου για την ταξινόμηση παραμένει σταθερή, αλλά η απόδοση για την παλινδρόμηση (και η γενική απόδοση) βελτιώνεται λίγο μέχρι τον τερματισμό της εκπαίδευσης στον 125^ο κύκλο (λόγω της EarlyStopping).



Εικόνα 6-3: Ιστορικό εκπαίδευσης CNN μοντέλου. Άθροισμα απωλειών ταξινόμησης και παλινδρόμησης (πάνω). Ακρίβεια (accuracy) για την ταξινόμηση (κέντρο). R-squared για την παλινδρόμηση (κάτω).

6.5 Αξιολόγηση μοντέλου & Αποτελέσματα

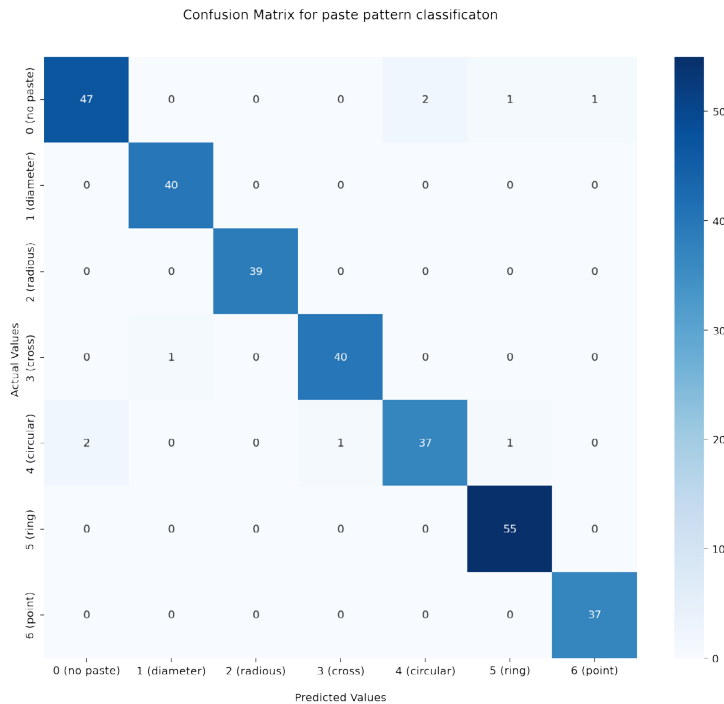
Στο τελευταίο βήμα της υλοποίησης του συνελκτικού μοντέλου πραγματοποιείται η αξιολόγησή του (model evaluation), δοκιμάζοντας την απόδοσή του σε ένα άγνωστο, για το μοντέλο, σύνολο δεδομένων, το σύνολο επικύρωσης (validation set). Το σύνολο αυτό δημιουργήθηκε πριν την έναρξη της διαδικασίας εκπαίδευσης και περιλαμβάνει 304 τιμές, όπως περιγράφηκε στην υποενότητα 6.2.2. Χρησιμοποιώντας το σύνολο επικύρωσης, το οποίο παρέμενε κρυφό από το μοντέλο μέχρι τώρα, το εκπαιδευμένο πλέον CNN μοντέλο έδωσε τα αποτελέσματα που παρουσιάζονται στις επόμενες υποενότητες.

6.5.1 Αποτελέσματα ταξινόμησης μοτίβων πάστας (classification)

Για την ταξινόμηση των μοτίβων πάστας (paste pattern classification task), όπως φαίνεται στον Πίνακα 6-6, το μοντέλο παρείχε πολύ καλές εκτιμήσεις του μοτίβου πάστας φτάνοντας σε συνολική ακρίβεια 97,04% για την ταξινόμηση των μοτίβων. Ενώ πέτυχε επίσης πολύ καλές επιδόσεις στα μέτρα αξιολόγησης Precision, Recall και F1-score για κάθε ξεχωριστή περίπτωση μοτίβου. Τέλος, στην Εικόνα 6-4, παρουσιάζεται ο Πίνακας Σύγχυσης (Confusion matrix) των αποτελεσμάτων.

Πίνακας 6-6: Αποτελέσματα αξιολόγησης για την ταξινόμηση των μοτίβων πάστας (classification task).

Αποτελέσματα αξιολόγησης για την ταξινόμηση μοτίβων πάστας (paste pattern classification) στο σύνολο επικύρωσης				
Total accuracy (συνολική ακρίβεια)	0.97039		97.04%	
Accuracy per class (ακρίβεια ανά περίπτωση μοτίβου - κλάση):				
Περιπτώσεις μοτίβων	Ακρίβεια (Precision)	Ανάκληση (Recall)	F1-score	Πλήθος δειγμάτων ανά μοτίβο
0 - Χωρίς Πάστα	0.95918	0.92157	0.94000	51
1 - Διάμετρος	0.97561	1.00000	0.98765	40
2 - Ακτίνα	1.00000	1.00000	1.00000	39
3 - Σταυρός	0.97561	0.97561	0.97561	41
4 - Κυκλικός Δίσκος	0.94872	0.90244	0.92500	41
5 - Δακτύλιος	0.96491	1.00000	0.98214	55
6 - Σημείο	0.97368	1.00000	0.98667	37



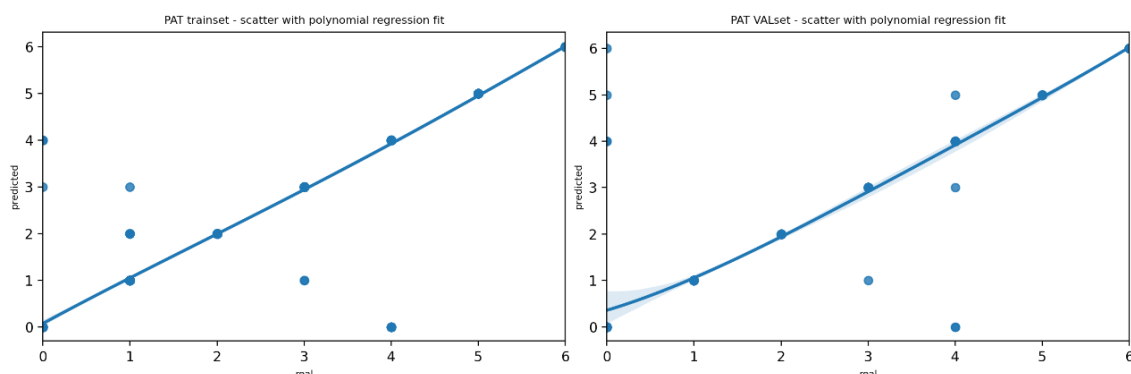
Εικόνα 6-4: Πίνακας Σύγχυσης (Confusion matrix) των αποτελεσμάτων του μοντέλου για την ταξινόμηση μοτίβων.

Για σύγκριση, παρουσιάζονται επίσης στον Πίνακα 6-7 και τα αποτελέσματα της ταξινόμησης για το σύνολο εκπαίδευσης:

Πίνακας 6-7: Αποτελέσματα αξιολόγησης για την το σύνολο εκπαίδευσης.

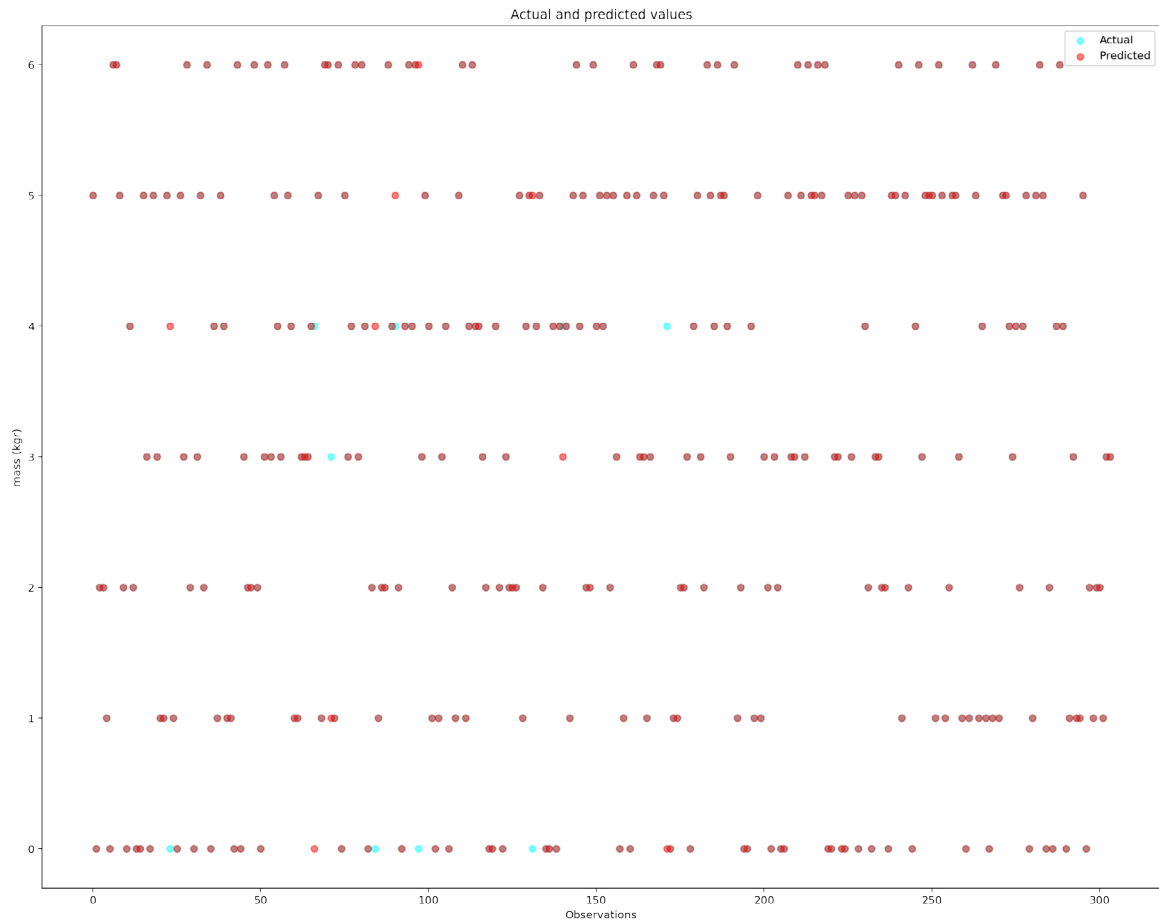
Αποτελέσματα αξιολόγησης για την ταξινόμηση μοτίβων πάστας (paste pattern classification) στο σύνολο εκπαίδευσης				
Total accuracy (συνολική ακρίβεια)	0.98668		98.67%	
Accuracy per class (ακρίβεια ανά περίπτωση μοτίβου - κλάση):				
Περιπτώσεις μοτίβων	Ακρίβεια (Precision)	Ανάκληση (Recall)	F1-score	Πλήθος δειγμάτων ανά μοτίβο
0 - Χωρίς Πάστα	0.96194	0.97887	0.97033	284
1 - Διάμετρος	0.99301	0.97260	0.98270	292
2 - Ακτίνα	0.97667	1.00000	0.98820	293
3 - Σταυρός	0.99313	0.99313	0.99313	291
4 - Κυκλικός Δίσκος	0.98246	0.96220	0.97222	291
5 - Λακτύλιος	1.00000	1.00000	1.00000	281
6 - Σημείο	1.00000	1.00000	1.00000	295

Επιπλέον, στην Εικόνα 6-5, παρουσιάζονται τα διαγράμματα διασποράς (scatter plots) των αποτελεσμάτων για την ταξινόμηση, με προσαρμογή πολυωνυμικής παλινδρόμησης (polynomial regression fit), για το σύνολο εκπαίδευσης (αριστερά) και το σύνολο επικύρωσης (δεξιά).



Εικόνα 6-5: Γραφήματα διασποράς, με προσαρμογή πολυωνυμικής παλινδρόμησης, της ταξινόμησης μοτίβων για το σύνολο εκπαίδευσης (αριστερά) και το σύνολο επικύρωσης (δεξιά).

Τέλος, ένας πρακτικός τρόπος ελέγχου των προβλέψεων της ταξινόμησης μοτίβων πάστας σε σχέση με τις πραγματικές τιμές παρουσιάζεται στην Εικόνα 6-6, όπου ο κατακόρυφος άξονας αντιστοιχεί στις 7 περιπτώσεις των μοτίβων πάστας (0 έως 6) και ο οριζόντιος άξονας αντιστοιχεί στον αύξοντα αριθμό της παρατήρησης. Κάθε μια από τις μπλε κουκκίδες αντιστοιχεί στην πραγματική περίπτωση μοτίβου πάστας (actual values) για τη συγκεκριμένη παρατήρηση, ενώ κάθε κόκκινη κουκκίδα αντιστοιχεί στην περίπτωση που πρόβλεψε το μοντέλο ότι ανήκει (predicted values). Στην ιδανική περίπτωση, εάν το μοντέλο ήταν 100% ακριβές, όλες οι μπλε και κόκκινες κουκκίδες θα πρέπει να συμπίπτουν. Όπως διακρίνεται στην εικόνα, το μοντέλο έδωσε πολύ λίγες λανθασμένες προβλέψεις. Υπάρχουν 3 μπλε κουκκίδες στην 4^η περίπτωση μοτίβου πάστας, 1 στην 5^η περίπτωση και 4 στην περίπτωση 0^η, δηλαδή στην περίπτωση χωρίς προσθήκη πάστας. Επομένως παρουσίασε 8 λάθη σε ένα σύνολο 304 παρατηρήσεων.



Εικόνα 6-6: Προβλέψεις CNN μοντέλου για την ταξινόμηση μοτίβων πάστας. Μπλε κουκκίδες: πραγματικές τιμές, κόκκινες κουκκίδες: προβλέψεις (για το σύνολο επικύρωσης).

6.5.2 Αποτελέσματα εκτίμησης πρόσθετης μάζας (regression)

Για την εκτίμηση της πρόσθετης μάζας λόγω της προσθήκης πάστας (extra mass regression task), όπως φαίνεται στον Πίνακα 6-7, το μοντέλο παρείχε σχετικά καλές εκτιμήσεις για την πρόσθετη μάζα φτάνοντας σε συνολικό R-squared score 82,5%. Επίσης, στα μέτρα αξιολόγησης Mean Squared Error (MSE), Root Mean Squared Error (RMSE) και Mean Absolute Error (MAE) κατάφερε καλή βαθμολογία.

Πίνακας 6-8: Αποτελέσματα αξιολόγησης για την εκτίμηση της πρόσθετης μάζας λόγω προσθήκης πάστας (regression task).

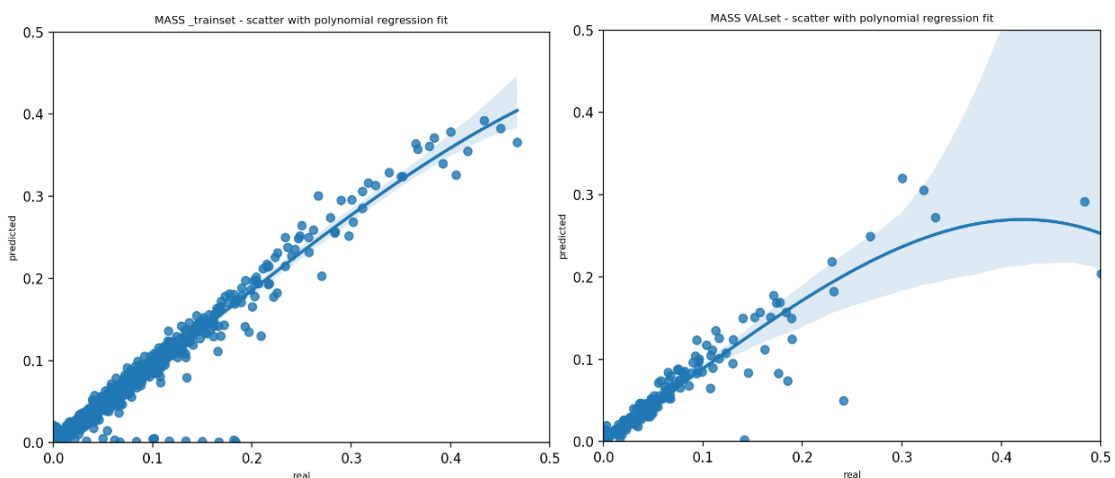
Αποτελέσματα αξιολόγησης για την εκτίμηση της πρόσθετης μάζας (extra mass regression task) στο σύνολο επικύρωσης		
MAE	0.00899	
R-squared	0.82499	82.5%
RMSE	0.02792	
MSE	0.00078	

Για σύγκριση, παρουσιάζονται επίσης στον Πίνακα 6-7 και τα αποτελέσματα της ταξινόμησης για το σύνολο εκπαίδευσης:

Πίνακας 6-9: Αποτελέσματα αξιολόγησης για την το σύνολο εκπαίδευσης.

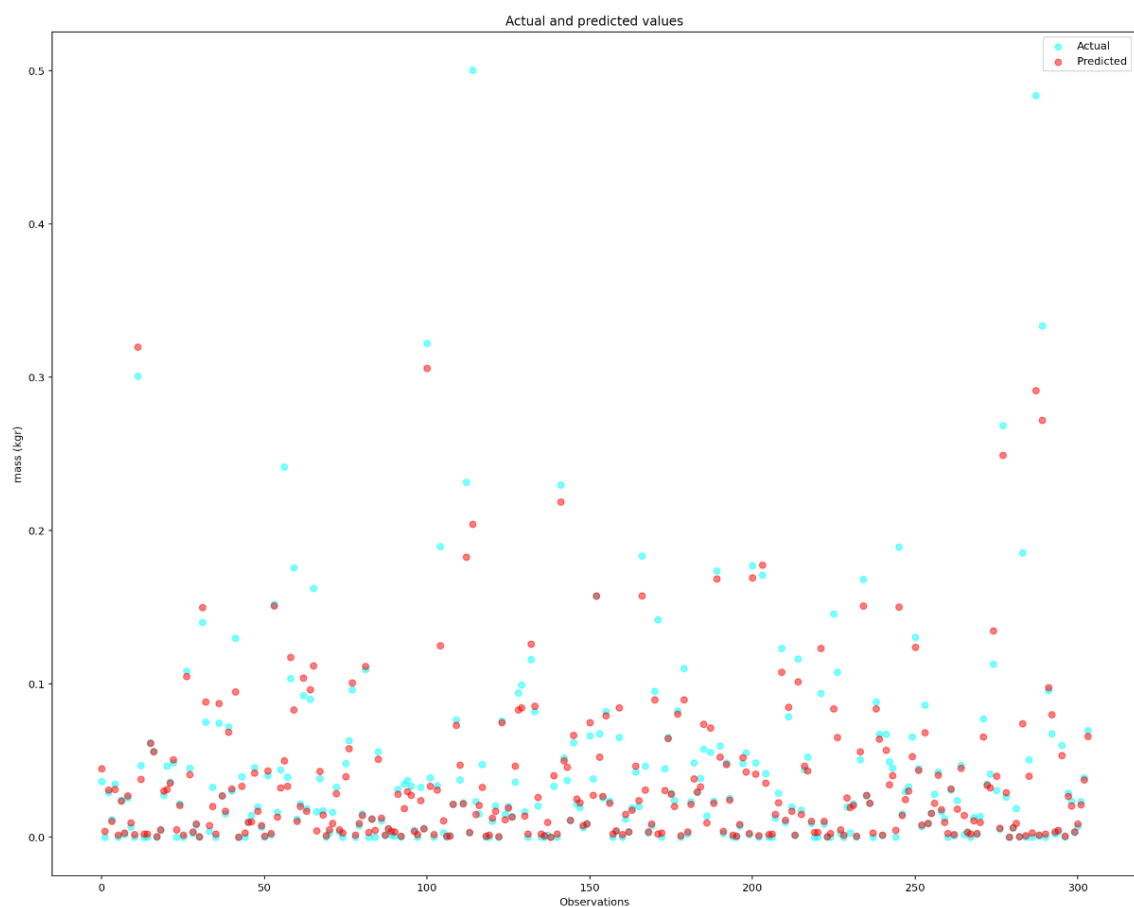
Αποτελέσματα αξιολόγησης για την εκτίμηση της πρόσθετης μάζας (extra mass regression task) στο σύνολο εκπαίδευσης		
MAE	0.00341	
R-squared	0.96321	96.3%
RMSE	0.00953	
MSE	0.00009	

Στην Εικόνα 6-7, παρουσιάζονται τα διαγράμματα διασποράς των αποτελεσμάτων για την εκτίμηση της πρόσθετης μάζας για το σύνολο εκπαίδευσης (αριστερά) και το σύνολο επικύρωσης (δεξιά).



Εικόνα 6-7: Γραφήματα διασποράς με προσαρμογή πολυωνυμικής παλινδρόμησης, της εκτίμησης της πρόσθετης μάζας για το σύνολο εκπαίδευσης (αριστερά) και το σύνολο επικύρωσης (δεξιά).

Τέλος, στην Εικόνα 6-8, παρουσιάζεται το αντίστοιχο διάγραμμα για την εκτίμηση της πρόσθετης μάζας, όπου το μοντέλο φαίνεται να παρουσιάζει περισσότερες «λανθασμένες» προβλέψεις, έχοντας κάποιες μικρές και κάποιες μεγαλύτερες αποκλίσεις μεταξύ των actual και predicted τιμών. Οι αποκλίσεις αυτές οφείλονται, ως ένα βαθμό, στο είδος του προβλήματος, δηλαδή της εκτίμησης πραγματικής τιμής (πρόβλημα παλινδρόμησης). Επίσης, ένας επιπλέον λόγος είναι ότι η αύξηση της μάζας λόγω της προσθήκης πάστας, στην ουσία επιφέρει μεταβολή της συνολικής μάζας της μεμβράνης, γεγονός που μπορεί να οδηγεί σε αντίστοιχο πρόβλημα, όπως αυτό που παρουσιάστηκε στην επίλυση του αρχικού αντίστροφου προβλήματος (ενότητα 5.2), δηλαδή της αδυναμίας εκτίμησης των φυσικών παραμέτρων του μοντέλου της δονούμενης μεμβράνης.



Εικόνα 6-8: Προβλέψεις CNN μοντέλου για (α) ταξινόμηση μοτίβων πάστας και (β) εκτίμηση πρόσθετης μάζας. Μπλε κουκκίδες: πραγματικές τιμές - κόκκινες κουκκίδες: προβλέψεις (για το σύνολο επικύρωσης).

6.5.3 Συνολική αξιολόγηση αποτελεσμάτων μοντέλου (classification & regression)

Όπως περιγράφηκε στην ενότητα 6.3, οι δυο κλάδοι του τελικού συνελικτικού μοντέλου, εκπαιδεύτηκαν στην ουσία ανεξάρτητα, με τον κλάδο της ταξινόμησης να χρησιμοποιεί ως συνάρτηση κόστους τη συνάρτηση Sparse Categorical Cross Entropy (SCCE) και τον κλάδο της παλινδρόμησης τη συνάρτηση Mean Absolute Error (MAE). Το μοντέλο εκπαιδεύτηκε συνολικά βάσει του αθροίσματος των δύο απωλειών επικύρωσης (validation loss) και επομένως παρείχε συνδυαστικά μια απώλεια επικύρωσης $combine_loss = 0.16316$. Η συνολική αυτή απώλεια προέκυψε από το άθροισμα των απωλειών της ταξινόμησης που ήταν $pattern_loss_SCCE = 0.14391$ και της παλινδρόμησης που ήταν $mass_loss_MAE = 0.01924$.

Η τιμή της συνολικής απώλειας επικύρωσης δεν έχει κάποια πρακτική αξία, καθώς δεν έχουν οριστεί κανόνες συσχέτισης του μοντέλου ανάμεσα στην ταξινόμηση των μοτίβων και στην εκτίμηση της μάζας. Αυτό έχει ως αποτέλεσμα να μη μπορεί να εκτιμηθεί άμεσα η συνολική απόδοση του μοντέλου.

Οπότε για τη συνολική αποτίμηση της απόδοσης του μοντέλου, θα μπορούσε να οριστεί μια τιμή κατωφλίου (threshold) για την εκτίμηση της πρόσθετης μάζας και σε συνδυασμό με την επιτυχημένη ή όχι ταξινόμηση του μοτίβου μάζας, να προκύπτει ένα μέτρο αξιολόγησης που θα δείχνει αν η γενική-συνολική πρόβλεψη ήταν σωστή ή όχι.

Επομένως, με βάση τα παραπάνω, ως μια πρώτη προσπάθεια, ορίστηκε μια τιμή κατωφλίου, η οποία αντιστοιχεί στην ποσοστιαία μεταβολή μάζας, ανάμεσα στην εκτιμώμενη και την πραγματική τιμή πρόσθετης μάζας, για ένα συγκεκριμένο παράδειγμα ενός μοτίβου πάστας.

Δηλαδή: $(\text{πραγματική τιμή μάζας} - \text{εκτιμώμενη τιμή μάζας}) / \text{πραγματική} \cdot 100\%$

Η εύρεση μιας αντιπροσωπευτικής τιμής κατωφλίου, που να δίνει μια σωστή συνολική αποτίμηση του μοντέλου είναι θέμα μελλοντικής εργασίας. Για παράδειγμα θα μπορούσε να ερευνηθεί και να βρεθεί μια τιμή κατωφλίου που να είναι σωστή αναφορικά με το ηχητικό αποτέλεσμα που προκαλεί. Για τους σκοπούς, όμως της παρούσας εργασίας επιλέχθηκε αυθαίρετα μια τιμή κατωφλίου $\pm 20\%$.

Στον Πίνακα 6-10, παρουσιάζονται ενδεικτικά 6 παραδείγματα προβλέψεων από το σύνολο επικύρωσης. Σε αυτά τα παραδείγματα φαίνονται οι προβλέψεις για την περίπτωση πάστας και αν είναι σωστές (μπλε χρώμα), καθώς και οι εκτιμήσεις της

πρόσθετης μάζας. Επίσης έχουν υπολογιστεί οι ποσοστιαίες μεταβολές μάζας σε κάθε περίπτωση.

Πίνακας 6-10: Ενδεικτικά παραδείγματα προβλέψεων από το σύνολο επικύρωσης, για την ταξινόμηση των μοτίβων πάστας και για την εκτίμηση της πρόσθετης μάζας.

# 01		# 02	
Pattern Prediction	5	Pattern Prediction	0
Pattern Real	[5] TRUE	Pattern Real	[0] TRUE
Mass Prediction	[0.0445]	Mass Prediction	[0.0038]
Mass Real	[0.03629]	Mass Real	[0.]
Mass Difference	[0.0082]	Mass Difference	[0.0038] %
Percentage change	[22.60683] %	Percentage change	-
# 03		# 04	
Pattern Prediction	2	Pattern Prediction	1
Pattern Real	[2] TRUE	Pattern Real	[1] TRUE
Mass Prediction	[0.03081]	Mass Prediction	[0.03116]
Mass Real	[0.02911]	Mass Real	[0.0346]
Mass Difference	[0.0017]	Mass Difference	[-0.00344]
Percentage change	[5.83882] %	Percentage change	[-9.94139] %
# 05		# 06	
Pattern Prediction	5	Pattern Prediction	6
Pattern Real	[5] TRUE	Pattern Real	[6] TRUE
Mass Prediction	[0.02705]	Mass Prediction	[0.00267]
Mass Real	[0.02575]	Mass Real	[0.0017]
Mass Difference	[0.0013]	Mass Difference	[0.00097]
Percentage change	[5.04564] %	Percentage change	[57.31675] %
<p>Pattern Prediction: Εκτιμώμενη περίπτωση μοτίβου πάστας. Pattern Real: Πραγματική. Mass Prediction: Εκτιμώμενη πρόσθετη μάζα σε kg. Mass Real: Πραγματική. Mass Difference: Διαφορά πραγματικής και εκτιμώμενη μάζας. Percentage change: Ποσοστιαία μεταβολή μάζας.</p>			

Χρησιμοποιώντας, λοιπόν τη σωστή ή λανθασμένη πρόβλεψη του μοτίβου και την ποσοστιαία μεταβολή μάζας, με τιμή κατωφλίου $\pm 20\%$, υλοποιήθηκε με κώδικα Python ένα μέτρο αξιολόγησης, το οποίο θεωρεί σωστή τη γενική πρόβλεψη, όταν για ένα παράδειγμα το μοντέλο προβλέπει σωστά το μοτίβο πάστας και ταυτόχρονα η ποσοστιαία μεταβολή (κατά απόλυτη τιμή) είναι κάτω από 20%. Αντίθετα, θεωρεί λάθος τη γενική πρόβλεψη στην αντίθετη περίπτωση δηλαδή, όταν είτε πρόβλεψη του μοτίβου δεν είναι σωστή, είτε η ποσοστιαία μεταβολή είναι πάνω από 20%. Στις περιπτώσεις όπου το παράδειγμα αντιστοιχεί σε μοτίβο χωρίς πάστα, λόγω του ότι η ποσοστιαία μεταβολή δεν

ορίζεται, το μέτρο θεωρεί σωστή τη γενική πρόβλεψη, εάν η πρόβλεψη του μοντέλου για το μοτίβο πάστας είναι σωστή.

Με βάση το μέτρο αυτό, το μοντέλο αξιολογήθηκε για το σύνολο επικύρωσης και έδωσε τα αποτελέσματα που παρουσιάζονται στον Πίνακα 6-11.

Πίνακας 6-11: Αποτελέσματα συνολικής αποτίμησης (classification & regression) στο σύνολο επικύρωσης.

Αποτελέσματα συνολικής αποτίμησης (classification & regression) στο σύνολο επικύρωσης		
Συνολική ακρίβεια με βάση το νέο μετρό (σωστό μοτίβο και ικανοποιητική εκτίμηση μάζας)	0.6941	69.41%
Accuracy per class (ακρίβεια ανά περίπτωση μοτίβου - κλάση):		
Περιπτώσεις μοτίβων	Ακρίβεια (Precision)	Πλήθος δειγμάτων ανά μοτίβο
0 - Χωρίς Πάστα	0.9216	51
1 - Διάμετρος	0.8250	40
2 - Ακτίνα	0.6923	39
3 - Σταυρός	0.7804	41
4 - Κυκλικός Δίσκος	0.4634	41
5 - Δακτύλιος	0.6364	55
6 - Σημείο	0.4865	37

Όπως βλέπουμε στον παραπάνω πίνακα για τιμή κατωφλίου 20%, το μοντέλο παρείχε μέτρια αποτελέσματα στη συνολική ακρίβεια, βάσει του νέου μέτρου, της τάξης του 69%. Επίσης, το μοντέλο δεν ήταν εφικτό να δώσει μια γενικά σωστή πρόβλεψη (μοτίβου και μάζας) για τις περιπτώσεις των μοτίβων: 4 - Κυκλικός Δίσκος και 6 - Σημείο.

Τέλος, όπως αναφέρθηκε και προηγουμένως η διερεύνηση της τιμής κατωφλίου αλλά και η ανάπτυξη άλλων μέτρων αξιολόγησης για τη συνολική αποτίμηση του μοντέλου, είναι ένα ιδιαίτερα ενδιαφέρον θέμα το οποίο θα αποτελέσει μελλοντική εργασία.

7

Συμπεράσματα

7.1 Σύνοψη & Συμπεράσματα

Φτάνοντας στα συμπεράσματα, μπορεί να γίνει μια σύνοψη του σκοπού και των αποτελεσμάτων της έρευνας. Σε αυτή τη μελέτη επιχειρήθηκε η εκτίμηση των φυσικών παραμέτρων του πάχους, της ακτίνας και της τάσης, μιας δονούμενης μεμβράνης, με βάση τον παραγόμενο ήχο. Τα αρχεία ήχου παρήχθησαν επιλύοντας την κυματική εξίσωση με τη μέθοδο Πεπερασμένων Διαφορών Χρονικού Τομέα κι η εκτίμηση των παραμέτρων πραγματοποιήθηκε με τη χρήση Συνελκτικών Νευρωνικών Δικτύων.

Καθώς διαφορετικοί συνδυασμοί των φυσικών παραμέτρων της μεμβράνης παράγουν παρόμοιους ήχους, το συνελκτικό μοντέλο δεν ήταν εφικτό να εκπαιδευτεί επαρκώς. Επομένως, αφού επαναπροσδιορίστηκε το αντίστροφο πρόβλημα πραγματοποιήθηκε:

α. ο προσδιορισμός του μοτίβου/σχήματος του αποσβεστικού υλικού (πάστας) που εφαρμόζεται στην επιφάνεια της μεμβράνης, το οποίο μπορεί να χρησιμοποιηθεί για το κούρδισμα της (πρόβλημα ταξινόμησης), και

β. η εκτίμηση της πρόσθετης μάζας λόγω αυτή της προσθήκης πάστας (πρόβλημα παλινδρόμησης)

Το τελικό συνελκτικό μοντέλο που προέκυψε μετά από χρονοβόρα διαδικασία βελτιστοποίησης, δοκιμών και ελέγχου, αποτελείται από δυο συνελκτικά επίπεδα, ακολουθούμενα από τα αντίστοιχα επίπεδα συγκέντρωσης. Έπειτα ακολουθεί ένα επίπεδο απόρριψης και το επίπεδο ισοπέδωσης. Η έξοδος του επιπέδου ισοπέδωσης τροφοδοτεί ένα πλήρως συνδεδεμένο επίπεδο, που έπειτα από ένα επιπλέον επίπεδο απόρριψης καταλήγει τελικά στο τελευταίο πλήρως συνδεδεμένο επίπεδο εξόδου.

Το μοντέλο παρείχε πολύ καλές εκτιμήσεις για τον προσδιορισμό του μοτίβου και την πρόσθετη μάζα, φτάνοντας σε ακρίβεια 97% για την ταξινόμηση των μοτίβων και βαθμολογία R^2 score 83% για την εκτίμηση της πρόσθετη μάζας.

Η σχετικά μικρότερη απόδοση στην εκτίμηση της πρόσθετης μάζας ίσως σχετίζεται με το γεγονός ότι η αύξηση της συνολικής μάζας της μεμβράνης λόγω της προσθήκης πάστας, μπορεί να οδηγήσει σε αντίστοιχο πρόβλημα, όπως αυτό που παρουσιάστηκε κατά την επίλυση του αρχικού αντίστροφου προβλήματος. Δηλαδή η μεταβολή της συνολικής μάζας της μεμβράνης, σε συνδυασμό με τις υπόλοιπες παραμέτρους, να οδηγούν στην παραγωγή παρόμοιων ήχων, οπότε και στην αδυναμία εκτίμησης των φυσικών παραμέτρων του μοντέλου, όπως είναι η συνολική μάζα της, και κατ' επέκταση η πρόσθετη μάζα λόγω της προσθήκης πάστας.

Τέλος, για τη συνολική αποτίμηση της απόδοσης του μοντέλου θα πρέπει να πραγματοποιηθεί ευρύτερη έρευνα, καθώς συνδυαστικά το μοντέλο δεν παρείχε ικανοποιητικά αποτελέσματα, ειδικά σε κάποιες από τις περιπτώσεις μοτίβων πάστας.

7.2 Μελλοντικές επεκτάσεις & Εφαρμογές

Στο πλαίσιο της παρούσας εργασίας, χρησιμοποιήθηκαν συγκεκριμένες προσεγγίσεις για τη μοντελοποίηση της δονούμενης μεμβράνης και την αντιμετώπιση του ευθέως προβλήματος. Κάποιες μελλοντικές επεκτάσεις που μπορούν να διερευνηθούν είναι:

- μια διαφορετική προσέγγιση για τη διερεύνηση των φυσικών παραμέτρων της δονούμενης μεμβράνης, διεγείροντας την σε διαφορετικά σημεία κρούσης
- η μελέτη του φυσικού μοντέλου της δονούμενης μεμβράνης και η εκτίμηση των παραμέτρων της, συμπεριλαμβάνοντας τις βισκοελαστικές ιδιότητες για το πρόσθετο υλικό απόσβεσης.
- η μελέτη διαφορετικών μοτίβων και σχημάτων για το αποσβεστικό υλικό και η διερεύνηση του προσανατολισμού του σε σχέση με το σημείο κρούσης.
- η εκτίμηση περισσότερων μεταβλητών παραμέτρων του φυσικού μοντέλου της δονούμενης μεμβράνης.

Κατά τη διαδικασία επίλυσης του αντίστροφου προβλήματος το συνελκτικό μοντέλο αναπτύχθηκε εφαρμόζοντας διάφορες τεχνικές και παραλλαγές παραμέτρων. Ωστόσο, υπάρχουν ακόμη πολλές επιλογές παραμετροποίησης οι οποίες δεν καλύφθηκαν και θα

μπορούσαν να αποτελέσουν αντικείμενο μελλοντικής μελέτης, με σκοπό την περαιτέρω βελτίωση του μοντέλου, η οποία με βάση τα όσα παρατηρήθηκαν κατά το πειραματικό μέρος κρίνεται εφικτή. Συγκεκριμένα:

- Διερεύνηση διαφορετικών τεχνικών ανάπτυξης και αρχιτεκτονικής τεχνητών νευρωνικών δικτύων, όπως των Ανατροφοδοτούμενων Νευρωνικών δικτύων (RNN) ή και συνδυασμό τους με συνελκτικά δίκτυα και σύγκριση αποτελεσμάτων με τις πιο σύγχρονες προσεγγίσεις CNN μοντέλων για ταξινόμηση ήχων.
- Περεταίρω δοκιμή εναλλακτικών μορφών για τα δεδομένα εισόδου, χρησιμοποιώντας διάφορες τεχνικές προ-επεξεργασίας των δεδομένων, αλλά και πιο εξειδικευμένων ηχητικών χαρακτηριστικών (audio features) που θα μπορούσαν να δοθούν ως είσοδος στα νευρωνικά δίκτυα.
- Έλεγχος της απόδοσης των νευρωνικών μοντέλων, με χρήση πραγματικών, ηχογραφημένων ηχητικών δειγμάτων τυμπάνων.
- Δημιουργία αυτόνομης εφαρμογής αναγνώρισης μοτίβων αποσβεστικής πάστας.

Επιπλέον, με αφορμή αυτή την εργασία, διαγράφονται διάφορες κατευθύνσεις για μελλοντική έρευνα, όπως ή μελέτη διαφορετικών φυσικών μοντέλων δονούμενων αντικειμένων, για παράδειγμα πνευστών και έγχορδων οργάνων.

Τέλος, συζητώντας πιθανές εφαρμογές, πρέπει να επισημάνουμε ότι τα αντίστροφα προβλήματα στο πεδίο της ακουστικής και της ανάκτησης πληροφοριών με δεδομένο τον παραγόμενο ήχο, μπορούν να προσφέρουν μερικές πολύ χρήσιμες εφαρμογές για μουσικούς και κατασκευαστές οργάνων, όπως:

- στην κατασκευή οργάνων, βασιζόμενοι στον επιθυμητό ήχο,
- στο κούρδισμα οργάνων σε έναν επιθυμητό ήχο (ηχόχρωμα και τόνο), με βάση την προσθήκη υλικού απόσβεσης, και
- στην προσπάθεια κατασκευής οργάνων που βρίσκονται σε εθνομουσικολογικές ή ιστορικές ηχογραφήσεις.

Αναφορές

- [1] R. Bader, “Finite-Difference model of Mode shape changes of the Myanmar Pat WAIN drum circle using tuning paste,” In Proceedings of Meetings on Acoustics, vol. 29, 2016.
- [2] R. Bader, “Spectrally-Shaping Viscoelastic Finite-Difference Time Domain (FDTD) model of a membrane,” <https://doi.org/10.5281/zenodo.3332874>, 2019
- [3] R. Bader, “Musical instruments,” Nonlinearities and Synchronization in Musical Acoustics and Music Psychology, vol. 2, pp. 157-284, 2013.
- [4] L. Gabrielli, S. Tomassetti, C. Zinato, and F. Piazza, “End-to-end learning for physics-based acoustic modeling,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 2, pp. 160–170, 2018.
- [5] L. Gabrielli, S. Tomassetti, S. Squartini, C. Zinato, and S. Guaiana, “A multi-stage algorithm for acoustic physical model parameters estimation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 8, pp. 1229–1240, 2019.
- [6] L. Gabrielli and S. Tomassetti, “Introducing Deep Machine Learning For Parameter Estimation In Physical Modelling,” in Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17), 2017.
- [7] M. Olivieri, R. Malvermi, M. Pezzoli, M. Zanoni, S. Gonzalez, F. Antonacci, and A. Sarti, “Audio Information Retrieval and musical acoustics,” IEEE Instrumentation; Measurement Magazine, vol. 24, no. 7, pp. 10–20, 2021.
- [8] S. Gonzalez, D. Salvi, D. Baeza, F. Antonacci, and A. Sarti, “A data-driven approach to violin making,” 2021.
- [9] F. Pfeifle and R. Bader, “Real-time finite-difference method physical modeling of musical instruments using field-programmable gate array hardware,” Journal of the Audio Engineering Society, vol. 63, no. 12, pp. 1001-1016, 2016.
- [10] R. Worland and W. Miyahira, “Physics of musical drum head damping using externally applied products,” in Proceedings of Meetings on Acoustics 35, 035004, 2018, doi: 10.1121/2.0001011
- [11] N. H. Fletcher and T. D. Rossing, The physics of musical instruments. New York: Springer, 1998.
- [12] A. C. Baines, The oxford companion to musical instruments. Oxford: Oxford University Press, 1992.
- [13] U. Michels, Ο Άτλας της Μουσικής, Τόμος 1 & 2. Αθήνα: ΝΑΚΑΣ, 1994.
- [14] Γ. Ντέλλας, “Τα διάφορα Μπεντίρ,” *Κρουστά - ένας κόσμος ολόκληρος*, 13-May-2008. [Online]. Available: <http://krousta-cosmos.blogspot.com/>. [Accessed: 07-Aug-2021].
- [15] D. Kuckhermann, “Frame drum comparison,” *World Percussion by David Kuckhermann*, 04-Oct-2015. [Online]. Available: <https://worldpercussion.net/frame-drum-comparison>. [Accessed: 07-Aug-2021].
- [16] Ν. Παπαδογιάννης και Ε. Μπακαρέζος, Κυματική Φυσική Για Μηχανικούς Ήχου και Ακουστικής. Αθήνα: ΣΕΑΒ, 2015.
- [17] S. S. Rao, *Vibration of continuous systems*. Hoboken, NJ: John Wiley ET Sons Ltd, 2019.
- [18] J. Roberts, “About drums,” *Circular Science*, 03-Jan-2019. [Online]. Available: <http://circularscience.com/about-drums>. [Accessed: 07-Aug-2021].

- [19] A. Wagner, "Analysis of drumbeats - interaction between drummer, drumstick and instrument," Master Thesis, Department of Speech and Hearing, Royal Institute of Technology, Stockholm, Sweden, 2006.
- [20] T. D. Rossing, *Science of percussion instruments*. Singapore: World Scientific, 2000.
- [21] D. Ritz, "Understanding Skins - Resonant Drumhead Explained," *The New Drummer*, 28-Sep-2016. [Online]. Available: <https://www.thenewdrummer.com/resonant-head-explained>. [Accessed: 07-Aug-2021].
- [22] Thomann GmbH, "Thomann online Guides accessories drumheads," *Musikhaus Thomann*. [Online]. Available: https://www.thomann.de/gb/onlineexpert_topic_drumheads.html. [Accessed: 28-Oct-2021].
- [23] Remo Inc., "Accessories," *Remo*. [Online]. Available: <https://remo.com/products/accessories>. [Accessed: 28-Oct-2021].
- [24] H. Becker, "Elastic modulus of Mylar Sheet," *Wiley Online Library*, 09-Mar-2003. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/app.1965.070090309>. [Accessed: 10-Aug-2021].
- [25] Laminated Plastics, "Technical data sheet mylar - Laminated Plastics," *Laminated Plastics*, 23-Jan-2014. [Online]. Available: <https://laminatedplastics.com/mylar.pdf>. [Accessed: 27-Oct-2021].
- [26] R. S. Christian, R. E. Davis, A. Tubis, C. A. Anderson, R. I. Mills, and T. D. Rossing, "Effects of air loading on timpani membrane vibrations," *The Journal of the Acoustical Society of America*, vol. 76, no. 5, pp. 1336-1345, 1984.
- [27] L. Rhaouti, A. Chaigne, and P. Joly, "Time-domain modeling and numerical simulation of a kettledrum," *The Journal of the Acoustical Society of America*, vol. 105, no. 6, pp. 3545-3562, 1999.
- [28] T. D. Rossing, "The physics of kettledrums," *Scientific American*, vol. 247, no. 5, pp. 172-178, 1982.
- [29] H. Fletcher and I. G. Bassett, "Some experiments with the bass drum," *The Journal of the Acoustical Society of America*, vol. 64, no. 6, pp. 1570-1576, 1978.
- [30] S. Tiwari and A. Gupta, "Effects of air loading on the acoustics of an Indian musical drum," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2611-2621, 2017.
- [31] R. Bader, "Finite-element calculation of a bass drum," *J. Acoust. Soc. Am.*, vol. 119, pp. 3290, 2006.
- [32] T. D. Rossing, I. Bork, H. Zhao, and D. O. Fystrom, "Acoustics of snare drums," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 84-94, 1992.
- [33] L. Cremer, M. Heckl, and B. A. T. Petersson, *Structure-borne sound*. New York: Springer-Verlag Berlin Heidelberg, 2005.
- [34] M. Campbell and C. A. Greated, *The musician's guide to acoustics*. London: Dent, 1987.
- [35] X. Παναγιωτόπουλος, "Αριθμητικές Μέθοδοι στην Εφαρμοσμένη Ακουστική - Χρήση της Climax στο περιβάλλον εργασίας SDE," στο Σεμινάριο Ακουστικής, 2017.
- [36] H. Igel, *Computational Seismology: A practical introduction*. Oxford: Oxford University Press, 2017.
- [37] Ι. Αικατερινάρης, "Εισαγωγικά αριθμητικών μεθόδων," in *Υπολογιστική μηχανική ρευστών*, 2011.

- [38] Γ. Σταυρουλάκης, Α. Μουράντοβα Κονταδάκη, and Μ. Σταυρουλάκη, *Υπολογιστική μηχανική*. Αθήνα: Εκδόσεις Κάλλιπος, 2015.
- [39] Ε. Παπαμίχος, Αριθμητικές μέθοδοι επίλυσης διαφορικών εξισώσεων με εφαρμογές στη μηχανική. Θεσσαλονίκη: Τζιόλα, 2004.
- [40] A. Kaveh, *Computational structural analysis and finite element methods*. Springer International Pu, 2014.
- [41] J. H. Rindel, "The Use of Computer Modeling in Room Acoustics," *Journal of Vibroengineering*, 2000.
- [42] R. H. Lyon and G. Maidanik, "Power flow between linearly coupled Oscillators," *The Journal of the Acoustical Society of America*, vol. 34, no. 5, pp. 623-639, 1962.
- [43] P. W. Smith, "Response and radiation of structural modes excited by sound," *The Journal of the Acoustical Society of America*, vol. 34, no. 5, pp. 640-647, 1962.
- [44] J. C. Strikwerda, *Finite Difference Schemes and Partial Differential Equations*, 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 2004.
- [45] E. A. Navarro, J. Segura, R. Sanchis, and A. Soriano, "Solving 2D acoustic ducts and membranes by using FDTD method," *Electronic Journal Technical Acoustics ISSN 1819-2408*, Nov. 2008.
- [46] M. Guerich and M. A. Hamdi. A numerical method for vibro-acoustic problems with incompatible finite element meshes using B-spline functions. *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1682-1694, March 1999.3.
- [47] F. Fontana and D. Rocchesso. Physical Modeling of Membranes for Percussion Instruments. *ACUSTICA-acta acustica*, vol. 84, pp. 529-542, 1998.4.
- [48] J. Bretos, C. Santamaria and J. Alonso-Moral. Vibrational patterns and frequency responses of the free plates and box of a violin obtained by finite element analysis. *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1942-1950, March 1999.5.
- [49] A-C. Hladky-Hennion and R. Bossut. Time analysis of immersed waveguides using the finite element method. *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 64-71, Jul. 1998.
- [50] K. S. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302-307, 1966.
- [51] G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-23, no. 4, pp. 377-382, 1981.
- [52] A. Taflove and S. C. Hagness, *Computational electrodynamics: The finite-difference time-domain method*. Boston, Mass: Artech House, 2005.
- [53] R. Holland, L. Simpson and K. Kunz. Finite difference analysis of EMC coupling to lossy dielectric structures. *IEEE Trans. Electromagnetic Compat.* vol. EMC-22, pp. 203-209, 1980.
- [54] K. Umashankar and A. Taflove. A novel method to analyze electromagnetic scattering of complex objects. *IEEE Trans. Electromagnetic Compat.* vol. EMC-24, pp. 397-405, 1982.
- [55] Kurt L. Shlager and John B. Schneider. A selective survey of the Finite-Difference Time-Domain Literature. *IEEE Antennas and Propagation Magazine*, vol. 37, no.4, August 1995, pp.39-56.
- [56] J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *Journal of Computational Physics*, vol. 114, no. 2, pp. 185-200, 1994.
- [57] R. Madariaga. Dynamics of an expanding circular fault. *Bull. Seismol. Soc. Am.*, vol. 66, pp. 639-666, 1976.

- [58] J. Virieux. SH-wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, vol. 49, pp. 1933-1942, 1984.
- [59] Q-H. Liu, E. Schoen, F. Daube, C. Randall, H-L. Liu and P. Lee. A three-dimensional finite difference simulation of sonic logging. *J. Acoust. Soc. Am.*, vol.100, no.1, pp. 72-79, 1996.
- [60] Y-H. Chen, W. C. Chew and Q-H. Liu. A three-dimensional finite difference code for the modeling of sonic logging tools. *J. Acoust. Soc. Am.*, vol. 103, no. 2, pp. 702-712, 1998.
- [61] S. Wang. Finite-difference time-domain approach to underwater acoustic scattering problems. *J. Acoust. Soc. Am.*, vol. 99, no. 4, pt. 1, pp. 1924-1931, Apr.1996.
- [62] F.D. Hastings, J.B. Schneider and S.L. Broschat. A finite-difference time-domain solution to scattering from a rough pressure-release surface. *J. Acoust. Soc. Am.*, vol. 102, no. 6, pp. 3394-3400, 1997.
- [63] D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *J. Acoust. Soc. Am.*, vol. 98, pp. 3302-3308, 1995.
- [64] J. LoVetri, D. Mardare and G. Souldre. Modeling of the seat dip effect using the finite-difference time-domain method. *J. Acoust. Soc. Am.*, vol. 100, pp. 2204-2212, 1996.
- [65] I. M. Hallaj, R. O. Cleveland. FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound. *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. L7-L12, 1999.
- [66] D. Botteldooren. Vorticity and entropy boundary conditions for acoustical finite difference time-domain simulations. *J. Acoust. Soc. Am.*, vol. 102, pp. 170-178, 1997.
- [67] D. Botteldooren. Numerical model for moderately nonlinear sound propagation in three-dimensional structures. *J. Acoust. Soc. Am.*, vol. 100, pp. 1357-1367, 1996.
- [68] X. Zhang and K. K. Mei. Time-domain finite difference approach to the calculation of the frequency-dependent characteristics of microstrip discontinuities. *IEEE Trans.*, 1988, MTT-36, (12), pp. 1775-1787
- [69] J. De Poorter and D. Botteldooren. Acoustical finite-difference time-domain simulations of subwavelength geometries. *J. Acoust. Soc. Am.*, vol. 104, no. 3, pp. 1171-1177, 1998.
- [70] D. Botteldooren. Acoustical finite-difference time-domain simulation in quasi-Cartesian grids. *J. Acoust. Soc. Am.*, vol. 95, no. 5, pp. 2313-2319, 1994.
- [71] Q. H. Liu, "The PSTD algorithm: A TIME-DOMAIN method requiring only two cells per wavelength," *Microwave and Optical Technology Letters*, vol. 15, no. 3, pp. 158-165, 1997.
- [72] M. Krumpholz and L. P. B. Katehi, "MRTD: New time-domain schemes based on multiresolution analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 4, pp. 555-571, 1996.
- [73] J. J. López, J. Escolano, and B. Pueo, "On the Implementation of a Room Acoustics Modeling Software using Finite-Difference Time-Domain Methods," *Audio Engineering Society - 122nd Audio Engineering Society Convention*, 2007.
- [74] NVIDIA, "CUDA toolkit documentation v11.5.0," *CUDA Toolkit Documentation*, 20-Oct-2021. [Online]. Available: <https://docs.nvidia.com/cuda/index.html>. [Accessed: 21-Oct-2021].
- [75] NVIDIA, "Cuda FAQ," *NVIDIA Developer*, 08-Sep-2021. [Online]. Available: <https://developer.nvidia.com/cuda-faq>. [Accessed: 21-Oct-2021].
- [76] Θ. Δ. Τσιμπούκης, "Εισαγωγή στη μέθοδο των πεπερασμένων στοιχείων," στο *Σημειώσεις Υπολογιστικός ηλεκτρομαγνητισμός*, 1994.

- [77] T. Angkaew, M Masanori and N. Kumagai. Finite element analysis of waveguide modes: a novel approach that eliminates spurious modes. *IEEE Trans. Microw. Theory Techniques*, vol. MTT-35, (2), pp. 117-123, 1987.
- [78] Τ. Γιούλτσης, “Η μέθοδος των πεπερασμένων στοιχείων,” στο *Σημειώσεις Μέθοδοι Πεπερασμένων Στοιχείων και Διαφορών*, 2003.
- [79] T. W. Wu, *Boundary element acoustics: Fundamentals and computer codes*. Southampton, UK: WIT Press, 2000.
- [80] Ι. Κατσικαδέλης, *Συνοριακά στοιχεία στην επιστήμη του Μηχανικού. Θεωρία και εφαρμογή*. Αθήνα: Εκδόσεις Συμμεών, 1999.
- [81] A. A. Becker, *The boundary element method in engineering: A complete course*. London: McGraw-Hill, 1992.
- [82] Y. F. Rashed, “Tutorial 1: An introduction to boundary elements,” *Boundary Element Communications*, vol. 12, no. 1, pp. 49-54, 2001.
- [83] Y. F. Rashed, “Tutorial 2: The first step to the direct integral forms,” *Boundary Element Communications*, vol. 12, no. 2-3, pp. 34-43, 2001.
- [84] A. F. Seybert and T. W. Wu, “Acoustic Modeling: Boundary Element Methods,” in *Handbook of Acoustics*, New York, NY: Wiley, 1998, pp. 157-168.
- [85] S. M. Kirkup, *The Boundary Element Method in Acoustics*. 2007.
- [86] C. A. Brebbia, J. R. Silva, and P. W. Partridge, *Boundary Element Methods in Acoustics*. Southampton: Computational Mechanics Publications, 1994.
- [87] Μ. Πλεξουσάκης and Π. Χατζηπαντελίδης, *Αριθμητική Επίλυση Μερικών Διαφορικών Εξισώσεων*. ΣΕΑΒ, 2015.
- [88] Α. Μπράτσος - Μαθήματα Εφαρμοσμένων Μαθηματικών - ΣΕΑΒ 2015
- [89] S. Dong, “Finite Difference Methods for the Hyperbolic Wave Partial Differential Equations,” in *Mathematical Methods for Engineers II*, 2006.
- [90] A. Taflove and M. E. Brodwin, “Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 23, no. 8, pp. 623-630, 1975.
- [91] R. Bader, “Finite-Difference model of Mode shape changes of the Myanmar Pat WAIN drum circle using tuning paste,” In *Proceedings of Meetings on Acoustics*, vol. 29, 2016.
- [92] E. Hairer, C. Lubich, and G. Wanner, “Geometric numerical integration illustrated by the Störmer-verlet method,” *Acta Numerica* 2003, pp. 399-450, 2003.
- [93] S. J. Russell, *Artificial Intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall/Pearson Education, 2010.
- [94] T. M. Mitchell, *Machine learning*. New York: McGraw-Hill, 1997.
- [95] T. O. Ayodele, *Types of machine learning algorithms*. INTECH Open Access Publisher, 2010.
- [96] S. S. Haykin, *Neural networks and learning machines*. Upper Saddle River: Pearson Education, 2009.
- [97] E. Alpaydin, *Introduction to machine learning*. Cambridge, MA: The MIT Press, 2010.
- [98] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Machine Learning*, vol. 6, no. 1, pp. 37-66, 1991.
- [99] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 1 edition. Amsterdam: Elsevier, 2015.

- [100] O. G. Yalçın, “4 intersecting domains that you can easily confuse with artificial intelligence,” Medium, 19-Dec-2020. [Online]. Available: <https://towardsdatascience.com/4-intersecting-domains-that-you-can-easily-confuse-with-artificial-intelligence-2233cb6ad7d1>. [Accessed: 12-Feb-2022].
- [101] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning,” Genetic Programming and Evolvable Machines, vol. 19, no. 1-2, pp. 305-307, 2017.
- [102] L. Deng, “Deep learning: Methods and applications,” 2014.
- [103] L. Gatys, A. Ecker, and M. Bethge, “A neural algorithm of artistic style,” Journal of Vision, vol. 16, no. 12, p. 326, 2016.
- [104] M. Popel, M. Tomkova, J. Tomek, Ł. Kaiser, J. Uszkoreit, O. Bojar, and Z. Žabokrtský, “Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals,” Nature Communications, vol. 11, no. 1, 2020.
- [105] A. Georgouli *Τεχνητή νοημοσύνη*. [ebook] Athens: Hellenic Academic Libraries. Available Online at: <http://hdl.handle.net/11419/3381>, 2015
- [106] J. Jordan, “Introduction to autoencoders.” Jeremy Jordan, 19-Mar-2018. [Online]. Available: <https://www.jeremyjordan.me/autoencoders/>. [Accessed: 25-Jan-2022].
- [107] A. O’Shea, G. Lightbody, G. Boylan, and A. Temko, “Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture,” Neural Networks, vol. 123, pp. 12-25, 2020.
- [108] H. Sak, A. Senior, and F. Beaufays, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling,” Interspeech 2014, 2014.
- [109] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014..
- [110] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, 1997.
- [111] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of trends in Practice and Research for Deep Learning”, arXiv: 1811.03378 [cs], p. 20, Nov. 2018.
- [112] F. Rosenblatt, “The Perceptron: A probabilistic model for information storage and organization in the brain.” Psychological Review, vol. 65, no. 6, pp. 386-408, 1958.
- [113] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. The MIT Press, 2017.
- [114] J. Duchi, E. Hazan., and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” Journal of Machine Learning Research, vol. 12, pp. 2121-2159, 2011
- [115] G. Hinton, N. Srivastava, and K. Swersky, “Neural Networks for Machine Learning,” in Lecture 6a. Available Online at: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf
- [116] D. Botteldooren, “Acoustical finite-difference time-domain simulation in a quasi-cartesian grid,” *The Journal of the Acoustical Society of America*, vol. 95, no. 5, pp. 2313-2319, 1994.
- [117] J. G. Maloney and K. E. Cummings, “Adaptation of FDTD techniques to acoustic modelling,” *11th Ann. Review of Progr. Appl. Comput. Electromag*, vol. 2, pp. 724-731, Monterey, CA, USA, 1995.

- [118] S. Jeyapalina, “Studies on the hydro-thermal and viscoelastic properties of leather,” thesis, University of Leicester, Leicester, 2004.
- [119] A. D. Pierce, “Intrinsic damping, relaxation processes, and internal friction in vibrating systems.,” *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1888-1888, 2010.
- [120] K. Team, “Keras Documentation: Adam,” Keras. [Online]. Available: <https://keras.io/api/optimizers/adam/>. [Accessed: 05-Feb-2022].
- [121] D. Kingma and J. Ba, in Adam: A Method for Stochastic Optimization: Proceedings of the International Conference on Learning Representations, 2014.
- [122] NVIDIA, “World leader in artificial intelligence computing,” *NVIDIA*. [Online]. Available: <https://www.nvidia.com/>. [Accessed: 30-Oct-2021].
- [123] NVIDIA, “About Cuda,” *NVIDIA Developer*, 11-Oct-2018. [Online]. Available: <https://developer.nvidia.com/about-cuda>. [Accessed: 30-Oct-2021].
- [124] D. P. Kingma and J. L. Ba, “Adam: a Method for Stochastic Optimization”, International Conference on Learning Representations, pp. 1-13, 2015
- [125] A. C. Müller and S. Guido, Introduction to machine learning with python: A guide for data scientists. Sebastopol, CA: O'Reilly, 2017.
- [126] C. D. Manning, P. Raghavan, and Schütze Hinrich, Introduction to information retrieval. New York: Cambridge University Press, 2009.
- [127] I. Kandel, M. Castelli, and A. Popovič, “Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images,” *Journal of Imaging*, vol. 6, no. 9, pp. 92, 2020.
- [128] Ong Cen Yen, M. Weldon, S. Quiring, L. Maxwell, M. Hughes, C. Whelan, and M. Okoniewski. Speed it up. *Microwave Magazine*, IEEE, 11(2):70{78, 2010.
- [129] “Underfitting vs just right vs Overfitting in Machine Learning: Data Science and Machine Learning,” Kaggle. [Online]. Available: <https://www.kaggle.com/getting-started/166897>. [Accessed: 13-Feb-2022].
- [130] F. Chollet, Deep learning with python. Shelter Island, NY: Manning Publications Co, 2021.
- [131] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *J. Mach. Learn. Res.*, vol. 15, 2014.
- [132] J S. Bergstra, R. Bardenet, Y. Bengio and B. Kégl, “Algorithms for hyper-parameter optimization”, In *Advances in Neural Information Processing Systems*, vol. 24, 2011.
- [133] J. Mockus, “Global Optimization and the Bayesian approach,” *Mathematics and Its Applications*, pp. 1-3, 1989.
- [134] W. Banzhaf, Genetic programming: An introduction; on the automatic evolution of computer programs and its applications. San Francisco, Calif: Kaufmann, 2009.
- [135] Y. Zhang, S. Wang, and G. Ji, “A comprehensive survey on particle swarm optimization algorithm and its applications,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1-38, 2015.
- [136] D. Stutz, “Introduction to Neural Networks,” in *Human Language Technology and Pattern Recognition*, RWTH Aachen University, 10-Feb-2014.
- [137] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

- [138] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the Cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.
- [139] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [140] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [141] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [142] J. Redmon, "Darknet: Open Source Neural Networks in C," Darknet: Open source neural networks in C. [Online]. Available: <https://pjreddie.com/darknet/>. [Accessed: 26-Jan-2022].
- [143] freeCodeCamp.org, "An intuitive guide to Convolutional Neural Networks," freeCodeCamp.org, 26-Feb-2018. [Online]. Available: <https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/>. [Accessed: 26-Jan-2022].
- [144] Ujjwalkarn, "An intuitive explanation of convolutional neural networks," Ujjwal Karn's blog, 29-May-2017. [Online]. Available: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>. [Accessed: 26-Jan-2022].
- [145] V. Kohir, "Calculating Output Dimensions in a CNN for convolution and pooling layers with keras," Medium, 25-Jun-2020. [Online]. Available: <https://kvirajdatt.medium.com/calculating-output-dimensions-in-a-cnn-for-convolution-and-pooling-layers-with-keras-682960c73870>. [Accessed: 26-Jan-2022].
- [146] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proc. INTERSPEECH*, 2013, pp. 436-440.
- [147] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Ensemble modeling of denoising autoencoder for speech spectrum restoration," in *Proc. INTERSPEECH*, 2014, pp. 885-889.
- [148] S.-W. Fu, Y. Tsao, and X. Lu, "SNR-Aware convolutional neural network modeling for speech enhancement," in *Proc. INTERSPEECH*, 2016.
- [149] S.-W. Fu, Y. Tsao, and X. Lu, "Complex spectrogram enhancement by convolutional neural network with multi-metrics learning," in *Proc. MLSP*, 2017
- [150] Y. Shi, W. Rong and N. Zheng, "Speech Enhancement using Convolutional Neural Network with Skip Connections," 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP), 2018, pp. 6-10, doi: 10.1109/ISCSLP.2018.8706591.
- [151] J.-C. Hou, S.-S. Wang, Y. H. Lai, J.-C. Lin, Y. Tsao, H.-W. Chang, and H.-M. Wang, "Audio-visual speech enhancement using deep neural networks," in *Proc. APSIPA ASC*, 2016.
- [152] Z. Wu, S. Sivadas, Y. K. Tan, B. Ma, and S. M. Goh, "MultiModal hybrid deep neural network for speech enhancement," arXiv:1606.04750, 2016.
- [153] J.-C. Hou, S.-S. Wang, Y.-H. Lai, Y. Tsao, H.-W. Chang, and H.-M. Wang, "Audio-visual speech enhancement using multimodal deep convolutional Neural Networks," arXiv.org, 24-Jan-2018. [Online]. Available: <https://arxiv.org/abs/1703.10893>. [Accessed: 09-Feb-2022].
- [154] Kuleshov, Volodymyr & Enam, S. & Ermon, Stefano. (2017). Audio Super Resolution using Neural Networks.

- [155] M. Dong, “Convolutional neural network achieves human-level accuracy in music genre classification,” 2018.
- [156] K. Choi, G. Fazekas, M. Sandler, and K. Cho, “Convolutional recurrent neural networks for music classification,” 2016.
- [157] W. Zhang, W. Lei, X. Xu, and X. Xing, “Improved music genre classification with convolutional neural networks,” in *Interspeech 2016*, 2016, pp. 3304–3308. [Online]. Available: <http://dx.doi.org/10.21437/Interspeech.2016-1236>
- [158] A. Guzhov, F. Raue, J. Hees, and A. Dengel, “Esresnet: Environmental sound classification based on visual domain models,” 2020.
- [159] Y. Aytar, C. Vondrick, and A. Torralba, “Soundnet: Learning sound representations from unlabeled video,” 2016.
- [160] F. Demir, D. A. Abdullah, and A. Sengur, “A new deep cnn model for environmental sound classification,” *IEEE Access*, vol. 8, pp. 66 529-66 537, 2020.
- [161] J. Lee, J. Park, K. L. Kim, and J. Nam, “Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms,” *arXiv preprint arXiv:1703.01789*, 2017.
- [162] Z. Zhu, J. H. Engel, and A. Hannun, “Learning multiscale features directly from waveforms,” *arXiv preprint arXiv:1603.09509*, 2016.
- [163] Y. Tokozume and T. Harada, “Learning environmental sounds with end-to-end convolutional neural network,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2721–2725.
- [164] K. Choi, G. Fazekas, and M. Sandler, “Automatic tagging using deep convolutional neural networks,” *arXiv preprint arXiv:1606.00298*, 2016.
- [165] S. Dieleman, P. Brakel, and B. Schrauwen, “Audio-based music classification with a pretrained convolutional network,” in *ISMIR*, 2011.
- [166] M.-T. Chen, B.-J. Li, and T.-S. Chi, “Cnn based two-stage multiresolution end-to-end model for singing melody extraction,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1005–1009.
- [167] X. Li, V. Chebiyyam, and K. Kirchhoff, “Multi-stream network with temporal attention for environmental sound classification,” 2019.
- [168] Y. Su, K. Zhang, J. Wang, and K. Madani, “Environment sound classification using a two-stream cnn based on decision-level fusion,” *Sensors*, vol. 19, no. 7, p. 1733, 2019.
- [169] J. Sang, S. Park, and J. Lee, “Convolutional recurrent neural networks for urban sound classification using raw waveforms,” in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2444-2448.
- [170] K. Choi, G. Fazekas, M. Sandler, and K. Cho, “Convolutional recurrent neural networks for music classification,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2392–2396.
- [171] Z. Wang, S. Muknahallipatna, M. Fan, A. Okray, and C. Lan, “Music classification using an improved crnn with multi-directional spatial dependencies in both time and frequency dimensions,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.
- [172] G. Gwardys and D. M. Grzywczak, “Deep image features in music information retrieval,” *International Journal of Electronics and Telecommunications*, vol. 60, no. 4, pp. 321–326, 2014.

- [173] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [174] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [175] Palanisamy, Kamalesh & Singhania, Dipika & Yao, Angela. (2020). Rethinking CNN Models for Audio Classification.
- [176] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [177] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [178] [226]G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [179] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *Proceedings of the 23rd Annual ACM Conference on Multimedia*. ACM Press, 2015, pp. 1015–1018. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2733373.2806390>
- [180] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *22nd ACM International Conference on Multimedia (ACM-MM'14)*, Orlando, FL, USA, Nov. 2014, pp. 1041–1044.
- [181] Yandre M.G. Costa, Luiz S. Oliveira, and Carlos N. Silla Jr. 2017. An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Applied Soft Computing* 52 (2017), 28 – 38. <https://doi.org/10.1016/j.asoc.2016.12.024>
- [182] Loris Nanni, Yandre M.G. Costa, Alessandra Lumini, Moo Young Kim, and Seung Ryul Baek. 2016. Combining Visual and Acoustic Features for Music Genre Classification. *Expert System Application* 45, C (March 2016), 108–117.
- [183] Ming-Ju Wu and Jyh-Shing R. Jang. 2015. Combining Acoustic and Multilevel Visual Features for Music Genre Classification. *ACM Transactions on Multimedia Computing Communications and Applications* 12, 1, Article 10 (Aug. 2015), 17 pages.
- [184] Christine Senac, Thomas Pellegrini, Florian Mouret, and Julien Pinquier. 2017. Music Feature Maps with Convolutional Neural Networks for Music Genre Classification. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing (CBMI '17)*. Association for Computing Machinery, New York, NY, USA, Article 19, 1–5. DOI:<https://doi.org/10.1145/3095713.3095733>
- [185] Weibin Zhang, Wenkang Lei, Xiangmin Xu, and Xiaofeng Xing. 2016. Improved Music Genre Classification with Convolutional Neural Networks. In *Interspeech 2016, USA*, September 8-12. 3304–3308.
- [186] Adiyansjah, Gunawan, A.A., & Suhartono, D. (2019). Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks. *Procedia Computer Science*.
- [187] Choi K, Fazekas G, Sandler M. Automatic Tagging using Deep Convolutional Neural Network. arXiv eprints arXiv:1606.00298. 2016.
- [188] Choi K, Fazekas G, Sandler M. Convolutional Recurrent Neural Networks For Music Classification. 2017

- [189] “Welcome to Python.org,” Python.org. [Online]. Available: <https://www.python.org/>. [Accessed: 02-Feb-2022].
- [190] Microsoft, “Visual studio code - code editing. redefined,” RSS, 03-Nov-2021. [Online]. Available: <https://code.visualstudio.com/>. [Accessed: 02-Feb-2022].
- [191] “IDE and code editor for software developers and teams,” Visual Studio, 07-Jan-2022. [Online]. Available: <https://visualstudio.microsoft.com/>. [Accessed: 02-Feb-2022].
- [192] K. Iqbal, “WAV - waveform audio file format,” WAV - Waveform Audio File Format, 13-Dec-2019. [Online]. Available: <https://docs.fileformat.com/audio/wav/>. [Accessed: 02-Feb-2022].
- [193] “Pandas,” pandas. [Online]. Available: <https://pandas.pydata.org/>. [Accessed: 02-Feb-2022].
- [194] NumPy. [Online]. Available: <https://numpy.org/>. [Accessed: 15-Feb-2022].
- [195] “Visualization with python,” Matplotlib. [Online]. Available: <https://matplotlib.org/>. [Accessed: 02-Feb-2022].
- [196] “Statistical Data Visualization,” seaborn. [Online]. Available: <https://seaborn.pydata.org/>. [Accessed: 02-Feb-2022].
- [197] “Scipy,” A blue circle with a snake in the shape of the letter 'S'. [Online]. Available: <https://scipy.org/>. [Accessed: 02-Feb-2022].
- [198] “Librosa,” librosa. [Online]. Available: <https://librosa.org/doc/latest/index.html>. [Accessed: 02-Feb-2022].
- [199] “OpenSMILE 3.0,” audeERING, 11-Jan-2022. [Online]. Available: <https://www.audeering.com/research/opensmile/>. [Accessed: 05-Feb-2022].
- [200] “Essentia,” Homepage - Essentia 2.1-beta6-dev documentation. [Online]. Available: <https://essentia.upf.edu/>. [Accessed: 05-Feb-2022].
- [201] NVIDIA, “CUDA toolkit documentation v11.5.0,” *CUDA Toolkit Documentation*. [Online]. Available: <https://docs.nvidia.com/cuda/index.html>. [Accessed: 30-Oct-2021].
- [202] P. Micikevicius, “3D finite difference computation on GPUs using Cuda,” Proceedings of 2nd Workshop on General Purpose Processing on Graphics Processing Units - GPGPU-2, 2009.
- [203] J. F. Stack, “Accelerating the Finite Difference Time Domain (FDTD) Method with CUDA,” Applied Computational Electromagnetics Society Conference, 2011.
- [204] A. Valcarce and J. Zhang, “Implementing a 2D FDTD scheme with CPML on a GPU using CUDA.” 2010.
- [205] M. Garland, S. Le Grand, J. Nickolls, J. Anderson, J. Hardwick, S. Morton, E. Phillips, Y. Zhang, and V. Volkov, “Parallel computing experiences with Cuda,” *IEEE Micro*, vol. 28, no. 4, pp. 13-27, 2008.
- [206] W. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity (1943),” *Ideas That Created the Future*, pp. 79-88, 2021.
- [207] F. Rosenblatt, “The Perceptron: A probabilistic model for information storage and organization in the brain.” *Psychological Review*, vol. 65, no. 6, pp. 386-408, 1958.
- [208] P. J. Werbos, *The roots of backpropagation: From ordered derivatives to neural networks and political forecasting*. New York: J. Wiley & Sons, 1994.
- [209] R. S. Michalski, J. G. Carbonell, T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.
- [210] S. Russell and P. Norvig, “A modern, agent-oriented approach to introductory artificial intelligence,” *ACM SIGART Bulletin*, vol. 6, no. 2, pp. 24-26, 1995.

- [211] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533-536, 1986.
- [212] "Tensorflow," TensorFlow. [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 05-Feb-2022].
- [213] K. Team, "Simple. flexible. powerful.," Keras. [Online]. Available: <https://keras.io/>. [Accessed: 05-Feb-2022].
- [214] K. Team, "Keras Documentation: Kerastuner," Keras. [Online]. Available: https://keras.io/keras_tuner/. [Accessed: 05-Feb-2022].
- [215] "Autokeras," AutoKeras. [Online]. Available: <https://autokeras.com/>. [Accessed: 05-Feb-2022].
- [216] K. Team, "Keras Documentation: RandomSearch Tuner," Keras. [Online]. Available: https://keras.io/api/keras_tuner/tuners/random/. [Accessed: 05-Feb-2022].
- [217] K. Team, "Keras Documentation: Core layers," Keras. [Online]. Available: <https://keras.io/layers/core/>. [Accessed: 05-Feb-2022].
- [218] "Tf.keras.losses.SparseCategoricalCrossentropy: Tensorflow core v2.8.0," TensorFlow. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy. [Accessed: 05-Feb-2022].
- [219] Google Colab. [Online]. Available: <https://colab.research.google.com/>. [Accessed: 05-Feb-2022].
- [220] "Tensorboard: TensorFlow," TensorFlow. [Online]. Available: <https://www.tensorflow.org/tensorboard>. [Accessed: 05-Feb-2022].
- [221] "Google Cloud," Google. [Online]. Available: <https://cloud.google.com/>. [Accessed: 14-Feb-2022].
- [222] S. Demircan and H. K. Örnek, "Comparison of the effects of Mel coefficients and spectrogram images via deep learning in emotion classification," *Traitement du Signal*, vol. 37, no. 1, pp. 51–57, 2020.
- [223] J. Li, W. Dai, F. Metze, S. Qu, and S. Das, "A comparison of deep learning methods for environmental sound detection," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
- [224] Y. Chen and H. Jin, "Rare sound event detection using Deep Learning and data augmentation," *Interspeech 2019*, 2019.
- [225] E. Cakir, T. Heittola, H. Huttunen, and T. Virtanen, "Polyphonic sound event detection using multi label Deep Neural Networks," 2015 International Joint Conference on Neural Networks (IJCNN), 2015.
- [226] Y. Liu, J. Tang, Y. Song, and L. Dai, "A capsule based approach for Polyphonic Sound Event Detection," 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2018.
- [227] I. McLoughlin, H. Zhang, Z. Xie, Y. Song, and W. Xiao, "Robust sound event classification using Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 3, pp. 540–552, 2015.
- [228] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using Deep Learning," *Applied Intelligence*, vol. 42, no. 4, pp. 722–737, 2014.
- [229] Y. Badr, P. Mukherjee, and S. Thumati, "Speech emotion recognition using MFCC and hybrid neural networks," *Proceedings of the 13th International Joint Conference on Computational Intelligence*, 2021.

- [230] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy, and Kong-Pang Pun, "An efficient mfcc extraction method in speech recognition," 2006 IEEE International Symposium on Circuits and Systems.
- [231] H. Li, Z. Zhang, F. Jiang, and X. Zhang, "An event recognition method for fiber distributed acoustic sensing systems based on the combination of MFCC and CNN," 2017 International Conference on Optical Instruments and Technology: Advanced Optical Sensors and Applications, 2018.
- [232] A. Mahmood and U. Köse , "Speech recognition based on convolutional neural networks and MFCC algorithm", Advances in Artificial Intelligence Research, vol. 1, no. 1, pp. 6-12, Jan. 2021
- [233] "Pickle - Python object serialization," pickle - Python object serialization - Python 3.10.2 documentation. [Online]. Available: <https://docs.python.org/3/library/pickle.html>. [Accessed: 06-Feb-2022].
- [234] R. Hinno, "Callbacks in neural networks," Medium, 06-Dec-2020. [Online]. Available: <https://towardsdatascience.com/callbacks-in-neural-networks-b0b006df7626>. [Accessed: 06-Feb-2022].
- [235] K. Team, "Keras Documentation: Callbacks API," Keras. [Online]. Available: <https://keras.io/api/callbacks/>. [Accessed: 06-Feb-2022].
- [236] K. Team, "Keras Documentation: ReduceLROnPlateau," Keras. [Online]. Available: https://keras.io/api/callbacks/reduce_lr_on_plateau/. [Accessed: 06-Feb-2022].
- [237] K. Team, "Keras Documentation: EarlyStopping," Keras. [Online]. Available: https://keras.io/api/callbacks/early_stopping/. [Accessed: 06-Feb-2022].
- [238] K. Team, "Keras Documentation: CSVLogger," Keras. [Online]. Available: https://keras.io/api/callbacks/csv_logger/. [Accessed: 06-Feb-2022].
- [239] "Tf.keras.callbacks.history: Tensorflow core v2.8.0," TensorFlow. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/History. [Accessed: 09-Feb-2022].
- [240] J. Brownlee, "Evaluate the performance of deep learning models in Keras," Machine Learning Mastery, 27-Aug-2020. [Online]. Available: <https://machinelearningmastery.com/evaluate-performance-deep-learning-models-keras/>. [Accessed: 09-Feb-2022].